文章编号　1004-924X(2026)01-0167-11

基于三分支网络的实时图像语义分割

任凤雷[1,2]，高紫阳[1,2]，张　炎[1,2]，周海波[1,2*]，杨　璐[1,2]，秦志昌[1,2]
(1. 天津理工大学 天津市先进机电系统设计与智能控制重点实验室，天津 300384；
2. 天津理工大学 机电工程国家级实验教学示范中心，天津 300384)

摘要：针对自动驾驶环境感知等应用场景对算法准确性和实时性的严苛要求，为了有效平衡语义分割模型的精度与推理速度，提出一种基于三分支网络的实时图像语义分割算法。借鉴PIDNet算法设计三分支网络结构，分别用于提取图像的细节信息、语义上下文信息和边缘信息。在语义分支设计高效金字塔池化模块，用于获取不同尺度的上下文信息，同时增大网络特征感受野。在细节分支和边缘分支设计轻量高效的多尺度通道交互注意力模块，以对提取到的特征进行增强。最后，融合上述三分支提取的图像特征并输出最终的语义分割结果。实验结果表明，所提出的基于三分支网络的实时图像语义分割算法在Cityscapes数据集取得了79.2% mIoU及88.5 frame/s的实时语义分割性能，在CamVid数据集取得了80.5% mIoU及140.1 frame/s的实时语义分割性能。本文提出的算法可以高效地实现图像语义分割任务，实时性和准确性方面均获得了极佳的平衡，语义分割性能显著优于现有基准方法。

关　键　词：语义分割；深度学习；实时性；注意力机制；多尺度特征
中图分类号：TP391.41　　**文献标识码**：A
doi：10.37188/OPE.20263401.0167　　**CSTR**：32169.14.OPE.20263401.0167

Real-time image semantic segmentation based on three-branch network

REN Fenglei[1,2], GAO Ziyang[1,2], ZHANG Yan[1,2], ZHOU Haibo[1,2*], YANG Lu[1,2], QIN Zhichang[1,2]

(1. *Tianjin Key Laboratory for Advanced Mechatronic System Design and Intelligent Control, Tianjin University of Technology, Tianjin* 300384, *China*;
2. *National Demonstration Center for Experimental Mechanical and Electrical Engineering Education, Tianjin University of Technology, Tianjin* 300384, *China*)
* *Corresponding author, E-mail: haibo_zhou*@163.*com*

Abstract: To meet the stringent requirements for both accuracy and real-time performance in applications such as autonomous driving, a real-time image semantic segmentation algorithm based on a triple-branch network is proposed to achieve a favorable balance between segmentation accuracy and inference speed. Inspired by PIDNet, a triple-branch architecture is designed to extract fine-grained detail information, semantic contextual information, and edge cues from the input image, respectively. An efficient pyramid pooling module is integrated into the semantic context branch to capture multi-scale contextual information

收稿日期：2025-09-19；**修订日期**：2025-11-06.
基金项目：国家自然科学基金资助项目(No. 51275209)；天津市技术创新引导专项(基金)企业科技特派员资助项目(No. 23YDTPJC00250)

and enlarge the network receptive field. In addition, a lightweight multi-scale channel interaction attention mechanism is introduced into both the detail and edge branches to enhance feature representations. Features from the three branches are subsequently fused, and a semantic segmentation head is employed to produce the final result. The proposed network achieves 79.2% mIoU at 88.5 frame/s on the Cityscapes dataset and 80.5% mIoU at 140.1 frame/s on the CamVid dataset. Experimental results demonstrate that the proposed method performs semantic segmentation efficiently, providing an effective trade-off between real-time performance and accuracy while significantly outperforming existing baseline methods.

Key words: semantic segmentation; deep learning; real time; attention mechanism; multi-scale features

1 引　言

语义分割作为计算机视觉领域的一项关键任务,其核心目标是对图像的所有像素进行精确分类,并赋予相应的标签[1]。语义分割在智能交通系统、医疗图像分析与机器人导航等诸多领域均有着广泛的应用,如在智能交通系统中,自动驾驶辅助的车道保持、无人驾驶车辆的导航等多个场景均需要语义分割技术的支持[2-3]。

近年来,深度学习技术的持续发展推动了卷积神经网络在图像语义分割领域的广泛应用,并在性能上大大超越了传统基于手工设计特征的方法。自从全卷积网络(Fully Convolutional Networks, FCN)[4]提出将卷积神经网络引入语义分割任务以来,诸多新网络结构相继被提出,并逐步提高了语义分割算法的精度。Chen等提出的DeepLab算法[5]通过使用扩张卷积增大网络的感受野,同时减少残差骨干网络的下采样操作,以维持较高的特征分辨率,该算法取得了优异的语义分割性能。以此网络结构为基准,PSPNet[6],DenseASPP[7],DeepLabV3+[8]等一系列改进算法相继涌现,并显著提升了语义分割算法的准确性。近年来,Transformer结构广泛应用于语义分割任务,使语义分割的准确性得到了进一步提升[9-10]。

对于无人驾驶车辆的环境感知等应用场景,除了算法的准确性外,对算法的实时性也提出了较高的要求。准确且快速的语义分割算法对于保障自动驾驶车辆的安全性是至关重要的。然而,语义分割是一种像素级别的密集预测任务,上述专注于语义分割准确性的算法计算复杂度过高,其处理速度上难以达到实际应用的实时性要求。针对这一挑战,实时语义分割方法应运而生,正日益成为研究的重要方向。ENet[11]为一种轻量级的编码-解码结构,能够显著提升语义分割的推理速度。ICNet[12]为一种基于多分辨率输入的实时语义分割网络,通过级联特征融合策略在速度和精度之间实现高效平衡。BiSeNetV1[13]作为一种双分支结构,分别用于提取空间细节信息和深层语义信息。BiSeNetV2[14]在上述网络的基础上进行了改进,进一步提升了算法的实时性。在上述双分支结构的基础上通过不断地优化网络结构,逐步提升了语义分割的准确性和实时性[15-17]。MSFNet[18]融合了不同尺度的空间特征图来同时保留丰富的空间细节与高级语义信息,有效提升了实时语义分割的精度与边缘质量。CABiNet[19]通过分解与重组多尺度特征,并利用轻量级注意力机制,在极低计算开销下实现了较高的分割精度。STDC[20]构建了短期密集连接模块,高效提取并融合多尺度感受野特征,从而在维持高推理速度的同时实现了优异的语义分割精度。HyperSeg[21]动态生成卷积权重,使主干网络能够自适应地处理不同区域的内容,并实现了高效的语义分割。SFNet[22]将高层语义信息精准对齐并融合至底层特征,从而实现了高效高精度的场景解析。为了更好地平衡语义分割算法的准确性和实时性,PIDNet[23]借鉴控制系统中经典的比例-积分-微分算法,提出一种新颖的三分支架构,分别用于提取图像的空间细节特征、上下文特征和边缘特征,并取得了领先的语义分割性能。

为了实现语义分割算法精确性与实时性的

最佳平衡，本文提出了一个简单高效的框架，用于完成实时且准确的图像语义分割。本文借鉴了PIDNet算法的三分支网络结构并分别用于提取图像的细节信息，语义信息和边缘信息；设计了高效金字塔池化模块用于获取不同尺度的上下文信息，并提出了一种轻量的多尺度通道交互注意力模块，实现了对特征的高效增强。本文算法相较于当前主流算法实现了精度与推理速度之间的更优平衡。

2　相关工作进展

2.1　图像语义分割

语义分割旨在将任意输入图像$X\in\mathbf{R}^{H\times W\times 3}$进行处理并得到分割结果，即$Y=f(X)$，$Y\in\mathbf{R}^{H\times W\times C}$，其中$H\times W$表示图像的空间分辨率，$C$表示预定义的语义类别[24]。实时图像语义分割是指在保证算法运行速度的前提下完成对图像的分割，是自动驾驶环境感知、机器人导航等领域的研究热点。现有的实时语义分割算法主要可以分为基于编码器-解码器结构[25-27]和基于多分支结构[13-17]的两种类型算法。其中，前者通过编码器提取多尺度特征，再通过解码器逐步恢复空间分辨率并融合高低层特征，最终输出分割结果。后者通过设计并行分支处理不同分辨率或不同语义层次的特征，最后融合各分支的特征得到分割结果。图像语义分割任务的示意图如图1所示。

图1　图像语义分割

Fig. 1　Image semantic segmentation

2.2　多尺度特征融合

多尺度特征融合旨在解决语义目标尺度差异的问题，其核心思想是通过借助多层次或多分辨率特征的融合，有效提升模型对多尺度目标的识别与感知性能。PSPNet[6]和DenseASPP[7]等方法通过引入不同尺度的上下文聚合策略或者密集连接的空间金字塔池化模块，增强了对不同尺度物体的识别能力。DeepLabV3+[8]设计了空洞空间金字塔池化模块（Atrous Spatial Pyramid Pooling，ASPP），通过采用多分支空洞卷积结构，利用不同的膨胀率并行捕获多尺度信息。DDRNet[15]设计了深度聚合金字塔池化模块（Deep Aggregation Pyramid Pooling Module，DAPPM）来捕获多尺度信息。PIDNet[23]在此基础上进一步优化改进，提出了PAPPM（Parallel Aggregation Pyramid Pooling Module），轻量化地完成多尺度特征的融合。

2.3　注意力机制

注意力机制（Attention Mechanism，AM）借鉴人类视觉系统中的选择性信息处理特性，通过动态分配计算资源到关键区域，提升计算机视觉任务的性能。其核心思想是抑制无关信息，增强重要特征的表达。近年来，AM已成为深度学习相关的核心组件之一[28]。在图像语义分割领域，CBAM[29]通过串联通道注意力和空间注意力模块来增强卷积神经网络的特征表示能力。DANet[30]同时引入通道注意力和空间注意力模块，分别优化特征图的通道维度和空间维度。基于Transformer结构的图像语义分割算法[31-34]则通过使用自注意力机制直接建模像素间的全局关系，并取得了相对卷积神经网络结构更加优异的性能。

3　原理与算法

本文算法拟借鉴PIDNet语义分割算法的三分支网络结构，并在其基础上进行优化设计。本文所提出的基于三分支网络的实时图像语义分割算法框图如图2所示。其中，输入图像由骨干网络提取图像浅层特征，同时将特征分辨率降采样至输入图像的1/8大小。细节分支、语义分支、边缘分支提取到的特征分别用粉色、黄色、蓝色进行表示（彩图见期刊电子版）。EPPM表示高效金字塔池化模块；MC-IAM表示多尺度通道交互注意力模块；Seg Head表示语义分割头，用于输出语义分割结果。

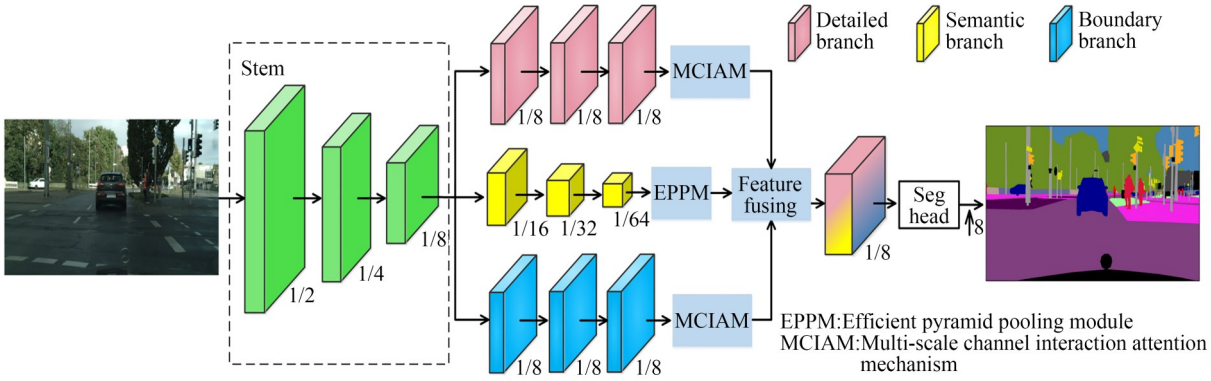

图 2　基于三分支网络的实时图像语义分割算法

Fig. 2　Framework of real-time image semantic segmentation based on three-branch network

3.1　三分支网络结构

PIDNet算法的核心创新在于借鉴控制理论中的PID控制器设计了并行的三分支结构。其中，P分支即比例分支，用于捕获高分辨率特征图中的细节信息；I分支，即积分分支，用于通过多次下采样逐步增大网络的感受野，并提取低分辨率特征图中的语义上下文特征；D分支，即微分分支，用于获取图像高频特征以精准预测边界区域。本文算法选择参考上述PIDNet算法的三分支网络结构进行设计，分别对应细节分支、语义分支和边缘分支。值得一提的是，PIDNet算法通过设计PAG模块以实现积分分支语义上下文特征对比例分支细节特征的引导。然而，PAG模块需要计算高分辨率特征图下的像素级注意力，这会消耗大量的计算资源。鉴于此，本文算法取消了PAG模块，旨在降低网络结构的复杂度，从而更加高效地实现特征的提取。

3.2　高效金字塔池化模块

多尺度上下文信息对于语义分割任务至关重要，自然场景中的物体尺寸差异极大，需要通过融合不同尺度的特征以解决因尺度差异导致的误分割问题。因此，设计了一种高效金字塔池化模块，如图3所示。其中，Avg-Pooling表示平均池化操作，Conv表示卷积操作，upsample表示上采样操作，“⊕”表示相加即分级残差连接操作，Concat表示特征拼接操作，ARS表示注意力优化结构，“⊗”表示相乘操作。

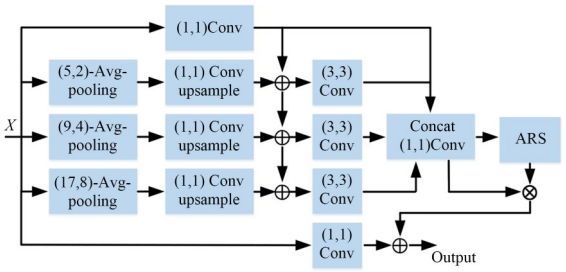

图 3　高效金字塔池化模块

Fig. 3　Efficient pyramid pooling module

高效金字塔池化模块首先使用不同核大小和步长的池化操作对模块输入特征进行处理，以获取不同尺度的上下文信息，同时不断增大特征感受野。在此基础上，对不同尺度的特征分别进行卷积操作和上采样操作恢复特征图至相同空间分辨率，并通过分级残差连接、3×3卷积操作和拼接操作整合所有的多尺度上下文信息。然后，使用1×1卷积操作达到减少通道数的目的，并使用注意力优化结构获取全局感受野，对多尺度特征进行进一步地增强。最后，将上述结果与输入特征进行残差连接，以避免训练过程中出现梯度消失的情况。上述过程如下：

$$y_i=\begin{cases}\mathrm{Conv}_{1\times1}(x), i=1\\ \mathrm{Conv}_{3\times3}(\mathrm{UP}(\mathrm{Conv}_{1\times1}(\mathrm{Pool}_{2^i+1,2^{i-1}}(x)))+y_{i-1}), 1<i<5\end{cases}, \tag{1}$$

$$y_C=\mathrm{Conv}_{1\times1}(\mathrm{Concat}(y_i)), \tag{2}$$

$$\alpha=\mathrm{Sigmoid}(\mathrm{Conv}(\mathrm{Gap}(y_C))), \tag{3}$$

$$y_A=\alpha\times y_C, \tag{4}$$

$$y_{\mathrm{out}}=y_A+\mathrm{Conv}_{1\times1}(x), \tag{5}$$

其中：x表示输入特征，y_i表示多尺度特征，y_C表示拼接多尺度特征后的结果，y_A表示注意力结构

优化后的特征，y_{out} 表示所提出的高效金字塔模块输出的特征，Conv 表示卷积操作，UP 表示上采样操作，Pool 表示池化操作，Concat 表示拼接操作，α 表示注意力权重图，Sigmoid 表示非线性操作，Gap 表示全局平均池化操作。

经过上述过程，高效金字塔模块可以有效提取并聚合多尺度全局上下文信息，从而提升算法对于不同尺度目标的分割性能。值得一提的是，本文算法将所提出的高效金字塔池化模块置于积分分支的末端，如图 2 所示。由于其输入的特征图大小仅为输入图像空间分辨率的 1/64，因此不会引入过多的计算量，确保了算法的实时性能。

3.3　多尺度通道交互注意力模块

通道注意力机制作为一种常见的用于特征增强的注意力机制，其核心思想是以对不同通道的特征根据其重要性进行加权的方式让神经网络学会关注哪些特征通道是重要的。

传统的通道注意力机制首先借助全局平均池化（Global Average Pooling，GAP）对空间维度信息进行聚合，以形成一个通道统计描述向量，然后通过设计两个全连接（Fully Connected，FC）层进行降维的同时捕获跨通道的特征交互。然而，上述降维操作在一定程度上损害了注意力机制的有效性。为了解决此问题，ECANet[35] 使用一维卷积进行相邻通道的特征交互。其中，一维卷积的作用是让每个通道的注意力权重由其 k 个相邻通道共同决定，k 即卷积核大小，而不是像全连接层那样由所有通道决定。这样既捕获了局部的跨通道交互信息，又可以极大地减少模型的参数量，从而在保持高性能的同时显著减少了模型复杂度和参数量。

然而，上述操作仅能实现单一尺度下的跨通道信息交互，为了高效捕获多尺度的跨通道特征交互，设计了一种轻量高效的多尺度通道交互注意力模块（Multi-scale Channel Interaction Attention Mechanism，MCIAM），如图 4 所示。使用 5 个具有相同的核大小（k=3）和不同扩张率（1，3，5，7，9）的并行的一维卷积来捕获不同尺度下的跨通道特征交互，生成通道权重。其中，较小的扩张率表示更多的关注局部相邻通道之间的特征关系，而较大的扩张率则捕获更长范围的通道交互。上述过程对应的代码如表 1 所示。

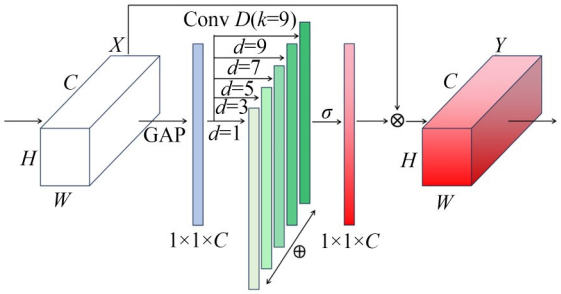

图 4　多尺度通道交互注意力模块

Fig. 4　Multi-scale channel interaction attention module

表 1　多尺度通道交互注意力模块代码

Tab. 1　Pytorch code of multi-scale channel interaction attention module

Algorithm：Multi-scale channel interaction attention mechanism，MCIAM

Input： $X\in\mathbf{R}^{C\times H\times W}\in\mathbf{R}^{C\times H\times W}$　（Feature map）

Output： Y　（Feature map enhanced by our proposed MCIAM）

```
class MCIAM(nn.Module):
def__init__(self,k_size=3):
super(MCIAM,self).__init__()
self.avg_pool=nn.AdaptiveAvgPool2d(1)
self.conv1=nn.Conv1d(1,1,kernel_size=k_size,
padding=k_size//2,dilation=1,bias=False)
self.conv3=nn.Conv1d(1,1,kernel_size=k_size,
padding=k_size//2*3,dilation=3,bias=False)
self.conv5=nn.Conv1d(1,1,kernel_size=k_size,
padding=k_size//2*5,dilation=5,bias=False)
self.conv7=nn.Conv1d(1,1,kernel_size=k_size,
padding=k_size//2*7,dilation=7,bias=False)
self.conv9=nn.Conv1d(1,1,kernel_size=k_size,
padding=k_size//2*9,dilation=9,bias=False)
self.sigmoid=nn.Sigmoid()
def forward(self,x):
y=self.avg_pool(x)
y_transformed=y.squeeze(-1).transpose(-1,-2)
y1=self.conv1(y_transformed)
y3=self.conv3(y_transformed)
y5=self.conv5(y_transformed)
y7=self.conv7(y_transformed)
y9=self.conv9(y_transformed)
y=y1+y3+y5+y7+y9
y=y.transpose(-1,-2).unsqueeze(-1)
y=self.sigmoid(y)
return x*y.expand_as(x)
```

将上述多尺度通道交互注意力模块分别放置于细节分支和边缘分支的尾部以对其特征进行有效增强，鉴于此模块轻量化的特性，可以保证算法整体的实时性。

最后，对上述细节分支、语义分支和边缘分支输出的特征进行有效融合并借助分割头结构输出语义分割结果。其中，特征融合结构选择使用 PIDNet 算法的 Bag 模块，通过利用边缘分支的特征引导细节分支和语义分支的特征进行融合。语义分割头结构设置最终的输出通道数为数据集定义的语义类别总数，即针对 Cityscapes 数据集设置为 19，针对 CamVid 数据集设置为 11，并通过双线性插值操作上采样至输入图像分辨率，以实现语义分割。

4　结果与分析

4.1　数据集简介

本研究的实验部分基于 Cityscapes 与 CamVid 两个广泛使用的图像语义分割数据集展开。其中，Cityscapes 是一个涵盖城市街景的大规模高质量数据集，专注于自动驾驶和城市场景理解任务，已成为语义分割领域的重要基准数据集之一。数据集图像采集自欧洲 50 个城市的街景，涵盖不同季节、天气条件和交通场景，图像及语义标签的空间分辨率为 1 024×2 048，可分为车辆、行人、建筑等 19 个语义类别，本文实验使用数据集中精细标注的 5 000 张图像（训练集 2 975 张，验证集 500 张，测试集 1 525 张）。CamVid 数据集是一个面向自动驾驶和场景理解的语义分割数据集，由剑桥大学于 2008 年发布。该数据集提供了共 701 张（训练集 367 张，验证集 101 张，测试集 233 张）图像及其对应的包含道路、车辆、行人等共 11 个语义类别的标签图，图像空间分辨率为 720×960。

4.2　实验平台及参数设置

实验平台基于 Ubuntu 18.04 OS，NVIDIA RTX 3090 GPU，CUDA 11.2，cuDNN8.0。算法实现使用 PyTorch 深度学习框架完成模型构建与训练过程。参设设置参考 PIDNet 算法，本文使用“Poly”学习率调度策略，使其按照迭代次数依次衰减$\left(1-\frac{\text{iter}}{\text{max_iter}}\right)^{\text{power}}$，其中衰减因子设置值为 0.9，同时应用随机裁剪、水平翻转及尺度缩放等数据增强方法。

4.3　相关评价指标

在算法准确性评估方面，依据数据集评估规范，使用平均交并比（mean Intersection-over-Union，mIoU）作为性能度量指标。具体计算公式如下：

$$\text{IoU}=\frac{P_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}}, \tag{6}$$

$$\text{mIoU}=\frac{\sum_{i=0}^{k}\text{IoU}_i}{k+1}, \tag{7}$$

其中：$k+1$ 表示数据集的语义类别总数，i 表示真实值，j 表示预测值。此外，本文使用帧率，即 Frames Per Second（FPS）作为实时性性能的评估标准。

4.4　消融实验

本文通过大量消融实验来验证所设计的三分支网络结构和各组成模块的有效性。为了直观展示各分支提取到的图像特征，基于 Cityscapes 数据集对其进行了可视化展示，结果如图 5 所示。其中，第一列为输入图像，第二至第四列分别对应细节分支、语义分支和边缘分支提取到的图像特征。由图可知，细节分支、语义分支和边缘分支可分别用于提取丰富的图像细节信息、语义上下文信息和边缘信息。

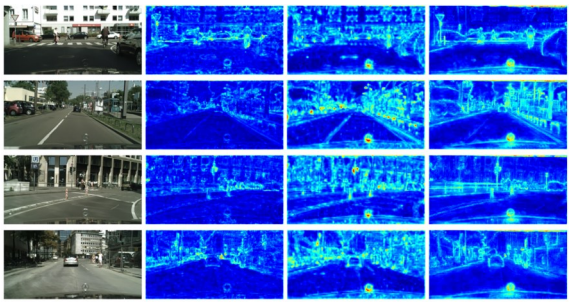

图 5　三分支特征可视化结果

Fig. 5　Three-branch feature visualization results

通过在语义分支设计高效金字塔池化模块用于获取不同尺度的上下文信息，同时达到增大网络特征感受野的目的。在 Cityscapes 数据集进行消融实验，结果如表 2 所示，为了更好地体现所提出的高效金字塔池化模块的有效性，将其与

ASPP[8]，DAPPM[15]，PAPPM[23]进行了对比。由表可知，本文提出的高效金字塔池化模块可以在保持较高计算效率的前提下显著提升了算法的分割准确性。

表 2　高效金字塔池化模块消融实验结果

Tab. 2　Ablation experiment of efficient pyramid pooling module

Method	mIoU/%	FPS
Baseline	76.2	103.5
Baseline＋ASPP[8]	78.9	79.3
Baseline＋DAPPM[15]	79.0	92.3
Baseline＋PAPPM[23]	78.7	98.2
Baseline＋EPPM(Ours)	79.2	88.5

本文在细节分支和边缘分支设计轻量高效的多尺度通道交互注意力模块，对提取到的特征进行有效增强。为了更好地体现所提出的多尺度通道交互注意力模块的有效性，将其分别与经典的注意力机制优化模块 SENet[28]，CBAM[29]，ECANet[35]进行了对比，其实验结果如表 3 所示。此外，对本文多尺度通道交互注意力模块的优化结果进行了可视化展示，结果如图 6 所示。其中，第一列为输入图像，第二列为加入注意力机制前的热力图(汽车语义类别)，第三列为加入本文提出的注意力模块后的热力图(汽车语义类别)。由上述实验结果可知，多尺度通道交互注意力模块可以使特征更好地关注和聚焦于相应区域；且相比较现有注意力机制模型，该模块通过引入多尺度卷积使模型能够同时捕获不同范围的上下文信息，通过并行多分支结构提升了特征表达的丰富性，可以更全面地理解通道间的关系，从而

表 3　多尺度通道交互注意力模块消融实验结果

Tab. 3　Ablation experiment of multi-scale channel interaction attention module

Method	mIoU/%	FPS
Baseline	78.7	93.1
Baseline＋SENet[28]	78.9	85.2
Baseline＋CBAM[29]	79.0	82.8
Baseline＋ECANet[35]	79.0	90.6
Baseline＋MCIAM(Ours)	79.2	88.5

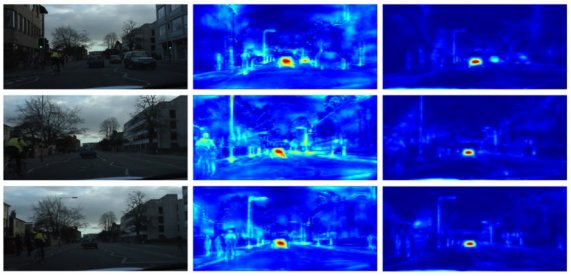

图 6　加入注意力机制前后热力图对比

Fig. 6　Comparison of heatmaps with and without MCIAM

轻量高效地实现对特征的增强，提升算法的语义分割准确性。

本文算法选择使用 PIDNet 算法的 Bag 模块，通过利用边缘分支输出的特征引导细节分支和语义分支输出的特征并实现高效融合。消融实验结果如表 4 所示，其中，Add 表示特征相加操作，Concat 表示特征拼接操作，上述操作为常用的深度学习特征融合方式。由表可知，Bag 模块可以更加高效地融合来自不同分支的特征，从而提升算法的性能。

表 4　特征融合模块消融实验结果

Tab. 4　Ablation experiment of feature fusion module

Feature Fusion	mIoU/%	FPS
Add	78.7	96.6
Concat	78.9	95.7
Bag(Ours)	79.2	88.5

4.5　定性分析实验

基于 Cityscapes 数据集和 CamVid 数据集对所提出的算法进行了性能验证。图 7 与图 8 分别展示了算法在上述两个数据集上的可视化分割结果。为突显本文算法相对于基准模型 PIDNet 的改进，图中对比展示了两种算法的语义分割效果。其中，首列为输入图像，第二、第三列分别为 PIDNet 算法和本文算法的语义分割结果，第四列为数据集提供的真实标注，即具备精确类别标签的参考分割结果。实验结果表明，基于三分支网络的实时语义分割算法能够有效适应多种道路环境场景，并实现准确的像素级预测。与 PIDNet 算法相比，本文方法在分割精度方面表现出更高的性能。

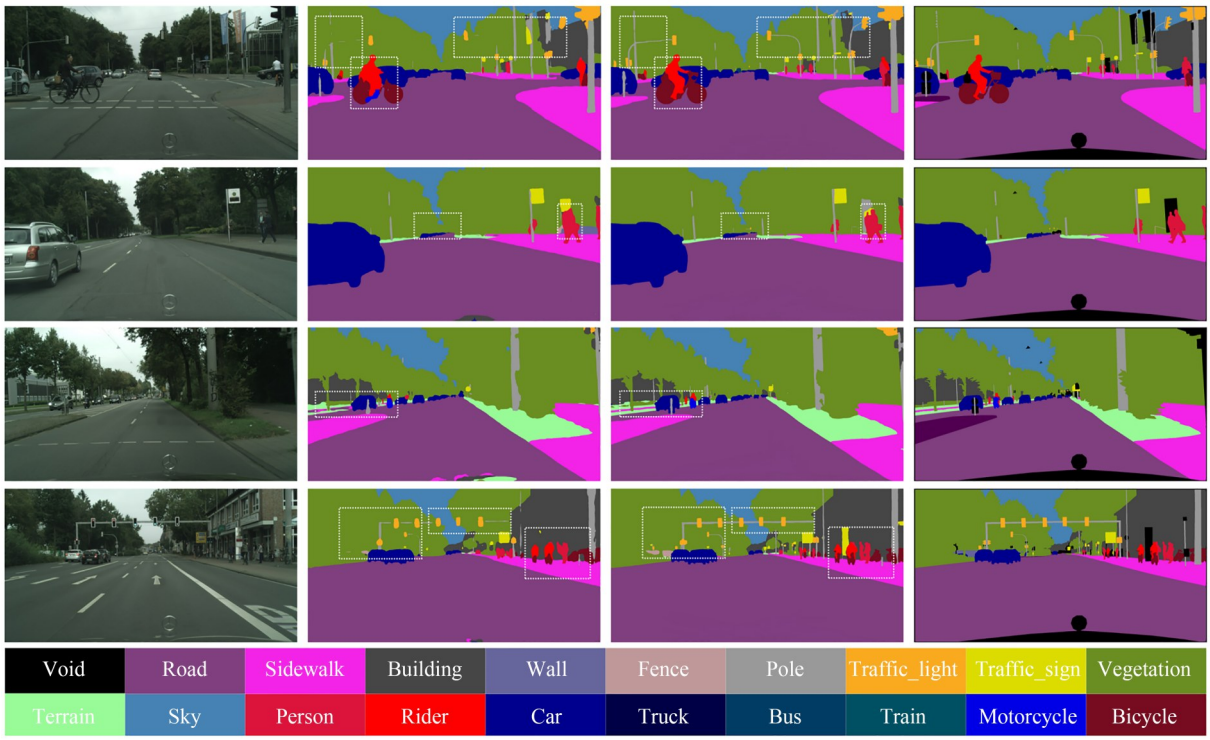

图 7　Cityscapes 数据集定性对比结果

Fig. 7　Qualitative comparison of results on Cityscapes dataset

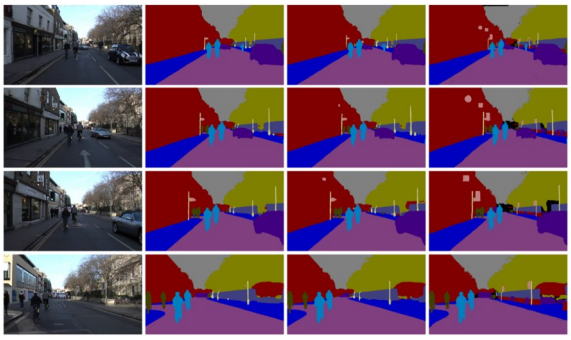

图 8　CamVid 数据集定性对比结果

Fig. 8　Qualitative comparison of results on CamVid dataset

4.6　定量分析实验

对所提出算法与当前主流先进的图像语义分割方法基于 Cityscapes 数据集和 CamVid 数据集进行了定量对比分析，结果如表 5 和表 6 所示。图 9 所示为本文算法与其他现有算法在 Cityscapes 数据集上的语义分割实时性和准确性间的性能对比图（彩图见期刊电子版），其中，红色竖线表示算法能否满足实时性的分界线，即 FPS 超过 30 代表可以基本满足实时性的要求；红色五角星和蓝色圆点分别代表本文算法和其他主流算法。由此可知，与 PIDNet 算法相比，本文算法经过改进使其获得了更优的语义分割性能，在语义分割的准确性和实时性方面，相较于现有算法达到了更理想的平衡。

表 5　Cityscapes 数据集定量对比结果

Tab. 5　Quantitative comparison of results on Cityscapes dataset

Method	Resolution	GPU	mIoU/%	FPS
MSFNet[18]	2 048*1 024	2080Ti	77.1	41
CABiNet[19]	2 048*1 024	2080Ti	75.9	76.5
BiSeNet[13]	1 536*768	1080Ti	74.7	65.5
BiSeNetV2[14]	1 024*512	1080Ti	75.3	47.3
STDC[20]	1 536*768	RTX3090	75.3	74.8
PP-LiteSeg[25]	1 536*768	RTX3090	74.9	96.0
HyperSeg[21]	1 024*512	RTX3090	75.8	59.1
SFNet[22]	2 048*1 024	RTX3090	77.8	87.6
DDRNet[15]	2 048*1 024	RTX3090	77.4	108.1
PIDNet-S[23]	2 048*1 024	RTX3090	78.6	93.2
Our Method	2 048*1 024	RTX3090	79.2	88.5

表 6　CamVid 数据集定量对比结果

Tab. 6　Quantitative comparison of results on CamVid dataset

Method	GPU	mIoU/%	FPS
MSFNet[18]	2080Ti	75.4	91.0
PP-LiteSeg[25]	1080Ti	75.0	154.8
BiSeNetV2[14]	1080Ti	76.7	124.0
HyperSeg[21]	1080Ti	78.4	38.0
DDRNet[15]	RTX3090	78.6	182.4
PIDNet[23]	RTX3090	80.1	153.7
Our Method	RTX3090	80.5	140.1

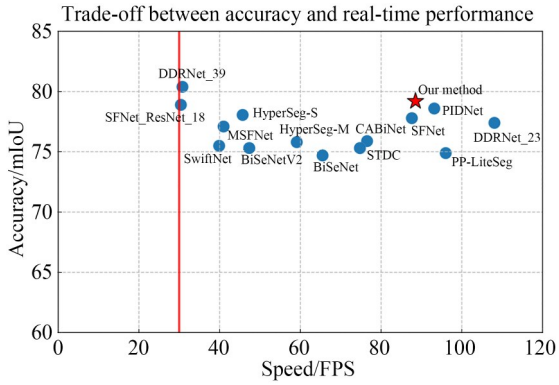

图 9　不同算法的准确性与实时性对比

Fig. 9　Comparison of accuracy and real-time performance of different algorithms

5　结　论

本文围绕平衡语义分割任务中准确性与实时性的核心挑战，提出了一种基于三分支网络的实时图像语义分割算法。该算法通过构建三分支网络架构，并行提取图像的细节信息、语义上下文信息和边缘信息，并在语义分支设计高效金字塔池化模块，有效获取了多尺度上下文信息并显著扩大了网络的特征感受野。同时，在细节分支与边缘分支创新性地提出了轻量高效的多尺度通道交互注意力模块，进一步增强了对关键特征的表征能力。本文算法在 Cityscapes 数据集取得了 79.2% mIoU 及 88.5 frame/s 的语义分割性能，在 CamVid 数据集取得了 80.5% mIoU 及 140.1 frame/s 的语义分割性能，在准确性与实时性的平衡方面显著优于现有主流实时语义分割算法，能够更好地满足自动驾驶等实际应用场景对算法准确性与推理速度的严苛需求。

作者贡献声明：

任凤雷：研究方案设计，论文撰写；

高紫阳、张炎：实验操作，数据分析；

周海波：图表绘制，论文审阅与编辑；

杨璐、秦志昌：实验材料提供。

参考文献：

[1] 白瑞峰，江山，孙海江，等. 基于编码解码结构的微血管减压图像实时语义分割[J]. 中国光学(中英文)，2022，15(5)：1055-1065.
BAI R F，JIANG SH，SUN H J，*et al*. Real-time semantic segmentation of microvascular decompression images based on encoder-decoder structure[J]. *Chinese Optics*，2022，15(5)：1055-1065. (in Chinese)

[2] 任凤雷，杨璐，周海波，等. 基于改进 BiSeNet 的实时图像语义分割[J]. 光学 精密工程，2023，31(8)：1217-1227.
REN F L，YANG L，ZHOU H B，*et al*. Real-time semantic segmentation based on improved BiSeNet[J]. *Opt. Precision Eng.*，2023，31(8)：1217-1227. (in Chinese)

[3] 任凤雷，周海波，杨璐，等. 基于双注意力机制的车道线检测[J]. 中国光学(中英文)，2023，16(3)：645-653.
REN F L，ZHOU H B，YANG L，*et al*. Lane detection based on dual attention mechanism[J]. *Chinese Optics*，2023，16(3)：645-653. (in Chinese)

[4] LONG J，SHELHAMER E，DARRELL T. Fully convolutional networks for semantic segmentation[C]. 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 7-12，2015，*Boston*，*MA*，*USA*. IEEE，2015：3431-3440.

[5] 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.

[6] ZHAO H S，SHI J P，QI X J，*et al*. Pyramid scene parsing network[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition*

(*CVPR*). *July* 21-26, 2017, *Honolulu*, *HI*, *USA*. IEEE, 2017: 6230-6239.

[7] YANG M K, YU K, ZHANG C, *et al*. DenseASPP for semantic segmentation in street scenes [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. *June* 18-23, 2018, *Salt Lake City*, *UT*, *USA*. IEEE, 2018: 3684-3692.

[8] CHEN L C, ZHU Y K, PAPANDREOU G, *et al*. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. *Computer Vision-ECCV* 2018. Cham: Springer, 2018: 833-851.

[9] STRUDEL R, GARCIA R, LAPTEV I, *et al*. Segmenter: transformer for semantic segmentation [C]. 2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *October* 10-17, 2021, *Montreal*, *QC*, *Canada*. IEEE, 2021: 7242-7252.

[10] Li X T, Ding H H, Yuan H B, *et al*. Transformer-based visual segmentation: A survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[11] PASZKE A, CHAURASIA A, KIM S, *et al*. Enet: A deep neural network architecture for real-time semantic segmentation [J]. *arXiv preprint* arXiv:1606.02147, 2016.

[12] ZHAO H S, QI X J, SHEN X Y, *et al*. ICNet for real-time semantic segmentation on high-resolution images [C]. Computer Vision—ECCV 2018. Cham: Springer, 2018: 418-434.

[13] YU C Q, WANG J B, PENG C, *et al*. BiSeNet: bilateral segmentation network for real-time semantic segmentation [C]. *Computer Vision-ECCV* 2018. Cham: Springer, 2018: 334-349.

[14] YU C Q, GAO C X, WANG J B, *et al*. BiSeNet V2: bilateral network with guided aggregation for real-time semantic segmentation[J]. *International Journal of Computer Vision*, 2021, 129(11): 3051-3068.

[15] PAN H H, HONG Y D, SUN W C, *et al*. Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(3): 3448-3460.

[16] REN F L, ZHOU H B, YANG L, *et al*. STDBNet: shared trunk and dual-branch network for real-time semantic segmentation[J]. *IEEE Signal Processing Letters*, 2023, 31: 770-774.

[17] POUDEL R P K, LIWICKI S, CIPOLLA R. Fast-scnn: Fast semantic segmentation network [J]. *arXiv preprint* arXiv:1902.04502, 2019.

[18] SI H, ZHANG Z, LV F, *et al*. Real-time semantic segmentation via multiply spatial fusion network [J]. *arXiv preprint* arXiv:1911.07217, 2019.

[19] KUMAAR S, LYU Y, NEX F, *et al*. CABiNet: efficient context aggregation network for low-latency semantic segmentation[C]. 2021 *IEEE International Conference on Robotics and Automation* (*ICRA*). *May* 30-*June* 5, 2021. *Xi'an*, *China*. IEEE, 2021: 13517-13524.

[20] FAN M Y, LAI S Q, HUANG J S, *et al*. Rethinking BiSeNet for real-time semantic segmentation[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 20-25, 2021, *Nashville*, *TN*, *USA*. IEEE, 2021: 9711-9720.

[21] NIRKIN Y, WOLF L, HASSNER T. HyperSeg: patch-wise hypernetwork for real-time semantic segmentation [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 20-25, 2021. *Nashville*, *TN*, *USA*. IEEE, 2021: 4060-4069.

[22] LI X T, YOU A S, ZHU Z, *et al*. Semantic flow for fast and accurate scene parsing[C]. *Computer Vision-ECCV* 2020. Cham: Springer International Publishing, 2020: 775-793.

[23] XU J C, XIONG Z X, BHATTACHARYYA S P. PIDNet: a real-time semantic segmentation network inspired by PID controllers[C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 17-24, 2023, *Vancouver*, *BC*, *Canada*. IEEE, 2023: 19529-19539.

[24] 闫河, 雷秋霞, 王旭. 融合注意力机制的改进型 DeepLabv3+语义分割[J]. 光学 精密工程, 2025, 33(1): 123-134.
YAN H, LEI Q X, WANG X. Improved DeepLabv3+ semantic segmentation incorporating attention mechanisms [J]. *Opt. Precision Eng.*, 2025, 33(1): 123-134. (in Chinese)

[25] PENG J, LIU Y, TANG S, *et al*. Pp-liteseg: A superior real-time semantic segmentation model [J]. *arXiv preprint* arXiv:2204.02681, 2022.

[26] ROMERA E, ÁLVAREZ J M, BERGASA L M, *et al*. ERFNet: efficient residual factorized

ConvNet for real-time semantic segmentation[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, 19(1): 263-272.

[27] WU T Y, TANG S, ZHANG R, *et al*. CGNet: a light-weight context guided network for semantic segmentation [J]. *IEEE Transactions on Image Processing*, 2020, 30: 1169-1179.

[28] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition. June* 18-23, 2018, *Salt Lake City*, *UT*, *USA*. IEEE, 2018: 7132-7141.

[29] WOO S, PARK J, LEE J Y, *et al*. CBAM: convolutional block attention module [C]. *Computer Vision-ECCV* 2018. Cham: Springer, 2018: 3-19.

[30] FU J, LIU J, TIAN H J, *et al*. Dual attention network for scene segmentation [C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 15-20, 2019, *Long Beach*, *CA*, *USA*. IEEE, 2019: 3141-3149.

[31] ZHENG S X, LU J C, ZHAO H S, *et al*. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 20-25, 2021, *Nashville*, *TN*, *USA*. IEEE, 2021: 6877-6886.

[32] CHENG B W, MISRA I, SCHWING A G, *et al*. Masked-attention mask transformer for universal image segmentation [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 18-24, 2022, *New Orleans*, *LA*, *USA*. IEEE, 2022: 1280-1289.

[33] ZHANG L X, HUANG W T, FAN B. SAR-Former: segmenting anything guided transformer for semantic segmentation [J]. *Neurocomputing*, 2025, 635: 129915.

[34] FATEH A, MOHAMMADI M R, JAHED-MOTLAGH M R. MSDNet: Multi-scale decoder for few-shot semantic segmentation *via* transformer-guided prototyping[J]. *Image and Vision Computing*, 2025, 162: 105672.

[35] WANG Q L, WU B G, ZHU P F, *et al*. ECA-net: efficient channel attention for deep convolutional neural networks[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 13-19, 2020. *Seattle*, *WA*, *USA*. IEEE, 2020: 11531-11539.

作者简介：

任凤雷(1991－)，男，河北沧州人，博士，硕士生导师，2015年于吉林大学获得学士学位，2020年于中国科学院长春光学精密机械与物理研究所获得博士学位，主要从事数字图像处理，自动驾驶视觉环境感知方面的研究。E-mail：renfenglei15@mails. ucas. edu. cn；renfl@email. tjut. edu. cn

通讯作者：

周海波(1973－)，男，黑龙江肇东人，博士，教授，博士生导师，1998年、2005年于佳木斯大学分别获得学士和硕士学位，2009年于吉林大学获得博士学位，主要从事计算机视觉、人工智能、智能机器人技术等方面的研究。E-mail：haibo_zhou@163. com