面向非结构化道路的轻量化轴向上下文分割网络

曹福[1]，邢雯彬[1]，左勇[2]，张荣辉[1]，陈俊周[1]
（1.中山大学智能工程学院，广东 广州 510275；2.武汉理工大学汽车学院，湖北 武汉 430070）

摘　要：非结构化道路分割是自动驾驶技术环境感知的重要组成部分，面临全局拓扑建模不完整、边界细节难以保持，及模型效率与精度的权衡等挑战。针对这些挑战，设计了一种轻量化轴向上下文网络 AXON-Net。该网络采用编码器-解码器架构，在编码器中引入通道-空间注意力模块（CASAB），通过聚合多维统计信息自适应重标定特征权重，有效抑制环境噪声，以增强复杂背景下的特征区分度；在瓶颈层设计轻量化部分上下文（LightPCT）模块，利用部分通道交互策略降低计算冗余，高效捕获长程依赖关系以修复道路拓扑连通性。在解码器中集成双路径通道融合（DPCF）与轴向细结构增强（TSE）模块，旨在缩小特征语义鸿沟并显式强化轴向几何特征，改善模糊道路边缘的精细化恢复效果。在基于印度驾驶数据集（IDD）与越野空间检测（ORFD）数据集二次构建的非结构化道路数据集上的实验结果表明，AXON-Net 在道路交并比（IoU_Road）分别达到 95.3%、88.1%，参数量仅为 8.49×10^3，实现了分割精度与模型效率的较优平衡。消融实验验证了各模块协同作用的有效性，展示了该网络在非结构化道路感知任务中的应用潜力。

关键词：非结构化道路；轴向上下文；通道-空间注意力；细结构增强；长程依赖
源代码链接：https://github.com/fucao01/AXON-Net
中图分类号：TP391.41　　**文献标志码：**A　　**DOI：**10.19678/j.issn.1000-3428.0253402

Lightweight Axial Context Segmentation Network for Unstructured Road

CAO Fu[1], XING Wenbin[1], ZUO Yong[2], ZHANG Ronghui[1], CHEN Junzhou[1]
(1.School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou 510275, Guangdong, China;
2.School of Automotive Engineering, Wuhan University of Technology, Wuhan 430070, Hubei, China)

【Abstract】Unstructured road segmentation is a crucial component of environmental perception for autonomous driving systems, and it suffers from challenges such as the integrity of global topological modeling, preservation of boundary details, and trade-off between model efficiency and accuracy. To address these challenges, this paper proposes a lightweight axial context network AXON-Net. The network employs an encoder-decoder architecture and introduces a Channel And Spatial Attention Block (CASAB) in the encoder, which adaptively recalibrates feature weights by aggregating multidimensional statistical information to effectively suppress environmental noise, thereby enhancing feature discriminability in complex backgrounds. A Lightweight Partial Context (LightPCT) model is integrated at the bottleneck, utilizing a partial channel interaction strategy to reduce computational redundancy and efficiently capture long-range dependencies for restoring road topological connectivity. In addition, the decoder integrates Dual-Path Channel Fusion (DPCF) and Thin Structure Enhancer (TSE) modules for bridging the feature semantic gap and explicitly enhancing axial geometric features for the refined recovery of blurred road edges. Experimental results on unstructured road datasets constructed from the India Driving Dataset (IDD) and Off-Road Freespace Detection (ORFD) dataset show that AXON-Net achieves road Intersection over Union (IoU_Road) scores of 95.3% and 88.1%, respectively. With only 8.49×10^3 parameters, it achieves a superior balance between segmentation accuracy and model efficiency. Ablation studies further validate the synergistic effectiveness of the proposed modules, demonstrating the potential application of the network for unstructured road perception tasks.

【Key words】unstructured road; axial context; channel and spatial attention; Thin Structure Enhancer (TSE); long-range dependencies

基金项目：国家自然科学基金（52172350，W2421069）；广州市重点研发计划（2024B01W0079）；深圳市科技攻关面上项目（JSGG20210802153412036）。

作者简介：曹福，男，硕士研究生，主研方向为计算机视觉；邢雯彬，硕士研究生；左勇，博士研究生；张荣辉，副教授；陈俊周（通信作者），副教授。

收稿日期：2025-12-08　　**修回日期：**2026-03-17　**E-mail：**chenjunzhou@mail.sysu.edu.cn

0 引言

精准的环境感知能力是自动驾驶安全运行的前提。在众多感知任务中，道路分割作为识别可行驶区域的核心环节，是实现路径规划与决策控制的关键前提。当前，多数研究聚焦于特征清晰的结构化道路，得益于伯克利深度驾驶数据集[1]、城市街景数据集[2](Cityscapes)等大规模结构化道路的分割数据集的支持，结构化道路分割的研究已日趋成熟。然而，在现实场景中驾驶环境是复杂多样的，不仅涵盖结构化道路场景，还涉及广泛存在的非结构化道路(如乡村土路、山间小径等)，因此对非结构化道路场景的研究具有重要的现实意义。非结构化道路因缺乏明确的边界标识、路面纹理多变且极易受光照阴影等环境因素干扰，这对分割算法的鲁棒性与精度提出了更高要求。

早期，采用多样化的方法对非结构化道路分割进行研究。例如，研究人员利用纹理基元块识别来改善分割对象的轮廓模糊问题[3]。也有方法通过融合图像边界与区域分类结果，并结合 Kalman 滤波对道路边界进行跟踪[4]；或基于 Gabor 纹理主方向投票检测道路消失点，以辅助判断道路走向[5]，此外，还有利用共点映射的自适应方法[6]，结合统计特征分析与广度优先搜索的方案[7]，以及融合立体视觉与数字地图的混合视觉系统[8]。然而，这些方法往往依赖于人工设计的特征提取算子或特定的几何假设，在光照变化、纹理缺失或复杂场景下，其鲁棒性受到限制。另一类方法则转向传感器融合，尝试利用雷达数据进行占用网格映射[9]或融合多种传感器信息[10]来识别可行驶区域。尽管多种传感器能提供更丰富的信息，但这类方法通常依赖特定的、有时较为昂贵的传感器组合，并可能面临数据配准、标定和实时处理方面的挑战。

考虑到基于视觉的分割方案仅需单目或双目相机，能大幅降低硬件成本与部署难度，且符合未来车载设备的轻量化需求趋势，发展基于视觉的道路分割模型是一种经济高效的选择。随着深度学习的兴起，研究范式逐渐转向基于卷积神经网络的端到端学习。尽管取得了显著进展，但在应对非结构化道路时，其常用架构也面临一些挑战。1)卷积操作倾向于关注局部信息，这使得标准网络在捕获长距离上下文依赖关系时可能遇到困难，当道路被遮挡或其纹理与周围环境高度相似时，这种特性有时可能导致分割结果中出现不连贯的“断裂”或“空洞”；2)在编码器-解码器结构中，编码器在逐级降采样以提取高级语义特征的过程中，往往会平滑并丢失部分道路边界、细微纹理等高频细节信息，给解码器恢复清晰、锐利的分割边缘带来挑战。

总结当前研究现状，基于图像的非结构化道路分割技术仍面临三大核心挑战：1)现有模型对长距离依赖关系和全局拓扑结构的捕获能力不足，导致分割结果中易出现道路中断或“空洞”；2)在降采样过程中，道路边界、细微纹理等细粒度特征容易丢失，导致边缘分割模糊；3)为追求高精度而设计的复杂模型通常参数量巨大，在计算资源受限的应用场景下面临计算效率的瓶颈。而现有轻量化模型，如为视觉导盲场景设计的 EMSeg[11]或通过分组卷积改造的 G-lite-DeepLabv3+[12]虽然在效率上表现出色，却常以牺牲部分分割精度为代价。针对上述挑战，本文设计了一种面向非结构化道路分割的轻量化轴向上下文网络 AXON-Net，通过定制化模块设计，在精度与效率间实现平衡，为非结构化场景感知任务贡献了一种轻量化且高精度的模型架构。

1 相关工作

1.1 基于深度学习的道路分割

自全卷积网络(FCN)实现端到端的像素级预测以来，基于深度学习的语义分割已成为主流[13]。其中，以 U-Net[14]为代表的编码器-解码器结构，通过对称的下采样和上采样路径和跳跃连接，有效融合了深层语义信息与浅层空间的细节信息，成为道路分割任务的基础架构。后续研究大多围绕如何增强该架构的能力展开。一些工作通过在 U-Net 架构中融合多种成熟模块来提升性能，例如，将密集连接、空洞空间金字塔池化(ASPP)与卷积注意力模块(CBAM)集成到 U-Net 中，以增强特征提取能力[15]；另一些工作则致力于设计更高效的网络单元，如通过设计 MobileNet[16]中的深度可分离卷积来构建轻量化模型，以满足实时性要求。然而，在非结构化道路分割的场景下，这些通用模型依旧面临一定的挑战。

1.2 上下文信息与全局建模

非结构化道路分割的核心难点之一在于其全局拓扑连通性易被破坏。由于缺乏清晰边界，道路常与周围的泥土、沙地等在纹理和颜色上高度相似，加上车辆、阴影的遮挡，标准卷积网络因其固有的局部感受野限制，极易将一条完整的道路错误地分割为多个独立的片段。为克服这些困难，增强模型的上下文信息捕获能力成为一种主流方法。其中，DeepLab 系列使用 ASPP 来聚合多尺度上下文，

PSPNet 则采用金字塔池化模块(PPM)来获取全局信息[17]。这些方法需要面临精度和计算成本上的均衡,虽然效果得到提升,但也增加了计算量。

此外,注意力机制虽然有助于捕获长程依赖关系,但现有的双重注意力策略[18]或 Transformer 架构[19]往往伴随着高昂的计算复杂度与显存开销,难以直接适配轻量化模型。因此,如何在满足实时性约束的前提下,设计高效的全局上下文融合机制以改善道路连通性,仍需进一步探索。

1.3 细粒度特征与边界处理

非结构化道路的另一个显著特征是其边界模糊且形态不规则。标准分割网络在降采样过程中容易丢失这些细粒度的边界信息。一些工作通过设计特定的边界损失函数来加强对边缘像素的监督。另一些工作则设计了专门的模块来增强边界特征,例如通过条带池化来捕获细长目标的轴向特征,这对于保持道路的连通性十分有效。如何在轻量化模型中高效地融合这些细结构增强(TSE)策略,以在不显著增加计算量的前提下提升边界分割质量,是当前研究的一个重要方向。

2 本文算法

本文提出一种面向非结构化道路分割的轻量化网络 AXON-Net。该模型遵循经典的 U-Net 编码器-解码器设计范式,摒弃了高计算冗余的重量级骨干网络,针对性地引入了 3 个核心优化模块:分别是编码器中通道-空间注意力模块(CASAB)、瓶颈层的轻量化部分上下文(LightPCT)模块以及解码器中集成双路径通道融合(DPCF)与轴向细结构增强模块 DPCF-TSE。

具体算法处理流程如下:首先,输入图像经基础卷积层完成浅层特征映射;随后进入编码器,利用各级嵌入的 CASAB 模块实现多尺度特征重标定,以有效抑制非结构化场景噪声;深层特征传输至瓶颈层后,通过 LightPCT 模块高效捕获全局长程依赖关系,修复道路拓扑断裂;在解码阶段,采用逐级上采样策略,并结合 DPCF 模块融合跨层特征与 TSE 模块强化细长结构;最终经 Softmax 函数生成概率图,并在训练时利用复合损失函数进行参数优化。

2.1 AXON-Net 总体架构

AXON-Net 的整体结构如图 1 所示(彩色效果见《计算机工程》官网 HTML 版,下同)。模型主要由注意力增强特征编码器(AF-Encoder)、LightPCT 及结构细化解码器(SR-Decoder)三大核心组成。编码器集成 CASAB 以增强多尺度特征;瓶颈层借由 LightPCT 高效捕获全局上下文信息;解码器则协同上采样与 DPCF-TSE 模块融合跨层特征,精准恢复细长结构的连通性与边界细节信息。

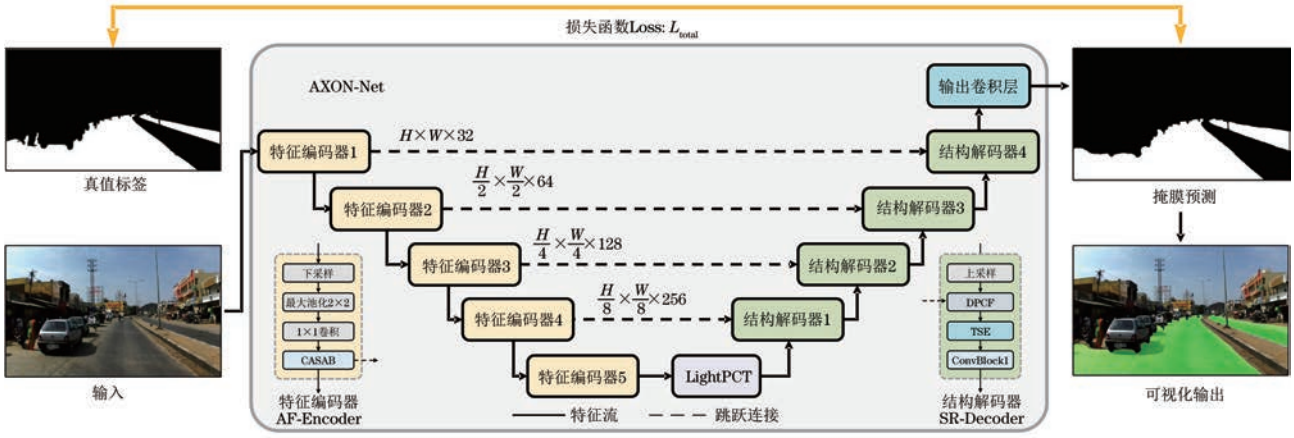

图 1 AXON-Net 模型总架构

Fig. 1 Overall architecture of AXON-Net model

2.2 编码器及注意力增强模块

编码器采用渐进式下采样架构,遵循 U-Net 的编码-解码范式,初始卷积层(AF-Encoder 1)与四级下采样模块(AF-Encoder 2~5)构成层级特征提取路径。初始卷积层通过深度可分离卷积与逐点卷积的级联组合对输入图像进行初始特征映射,生成基础通道数为 C_0 的特征表示,如图 2 所示。

各下采样级采用步长为 2 的最大池化算子实现空间维度减半,随后串联卷积块(ConvBlock)与本文设计的 CASAB,在多个分辨率尺度下协同实现特征冗余抑制与语义判别性增强。CASAB 模块由深度可分离卷积预处理单元、双路径通道注意力机制和多统计量空间注意力机制等模块按并联形式构成。给定输入特征张量 $\boldsymbol{X}\in R^{C\times H\times W}$,首先经过深度可分离卷积进行局部特征 $\boldsymbol{X}'$ 提取:

$$\boldsymbol{X}_1=\mathrm{Conv}_{\mathrm{PW}}(\mathrm{BN}(\mathrm{Conv}_{\mathrm{DW}}(\boldsymbol{X}))) \tag{1}$$

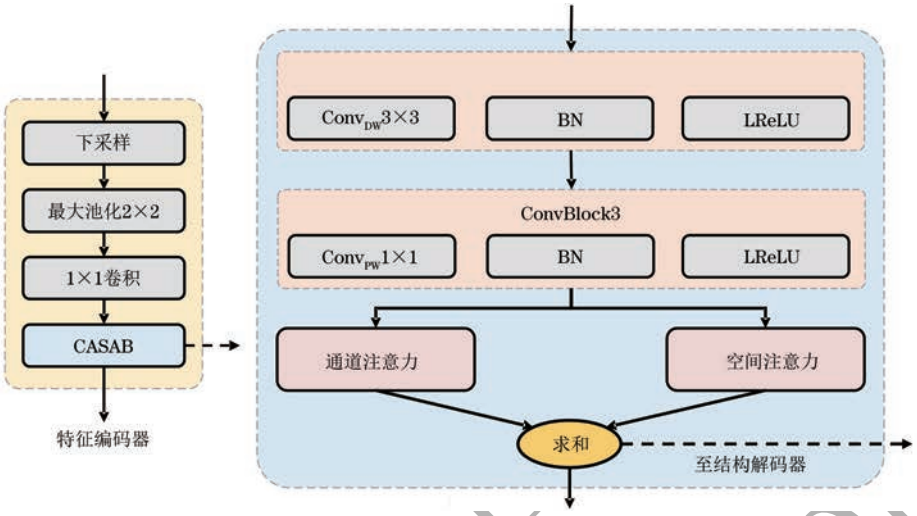

图 2 特征编码器及 CASAB 模块示意图

Fig. 2 Schematic diagram of the feature encoder and CASAB module

式中：$\mathrm{Conv}_{\mathrm{DW}}\in R^{3\times3}$ 表示分组卷积数等于输入通道数的深度卷积；$\mathrm{Conv}_{\mathrm{PW}}\in R^{1\times1}$ 为通道间信息融合的逐点卷积。该分解策略将标准卷积的参数复杂度从 $O(C^2k^2)$ 降至 $O(Ck^2+C^2)$，其中 k 表示卷积核的大小（此处 $k=3$），在保证等效感受野的前提下降低计算负担。

在通道注意力分支，CASAB 并行采用全局平均池化 F_{avg} 和全局最大池化 F_{max}，以提取通道维度的全局上下文统计描述信息。这两路统计量共享一套通道压缩比为 $r=16$ 的多层感知机（MLP），通过瓶颈结构建模通道间的非线性依赖关系，并结合激活函数生成自适应通道权重映射：

$$M_{\mathrm{channel}}=\sigma(\mathrm{MLP}(F_{\mathrm{avg}})+\mathrm{MLP}(F_{\mathrm{max}})) \tag{2}$$

式中：$M_{\mathrm{channel}}\in R^{3\times3}$ 表示生成的自适应通道权重映射；MLP(•)表示共享权重的多层感知机；σ(•)表示 Sigmoid 激活函数。该双路径设计聚合互补统计量，增强通道选择性并突出语义特征。

空间注意力分支则沿通道维度计算均值 F_{mean}、最大值 F_{max}、最小值 F_{min} 与总和 F_{sum} 4 种高阶统计量，以捕获更丰富的空间显著性模式。这四路统计量经通道维数拼接后，依次通过核尺寸为 7×7 的卷积层、SiLU 非线性激活函数及 1×1 卷积层，最后通过 Sigmoid 函数生成归一化的空间权重 M_{spatial} 分布：

$$M_{\mathrm{spatial}}=\sigma(\mathrm{Conv}_{1\times1}(\gamma)) \tag{3}$$

$$\gamma=\mathrm{SiLU}(\mathrm{Conv}_{7\times7}(\mathrm{Concat}(\Delta))) \tag{4}$$

式中：Concat 为通道拼接；γ 表示经过大感受野卷积后的中间特征；Δ 表示 F_{mean}，F_{max}，F_{min}，F_{sum} 的连接。相比仅使用均值与最大值双统计量，该多统计量策略能够更全面地刻画特征分布 F_{max} 特性，尤其在处理非结构化道路场景中的不均匀光照、阴影干扰等复杂因素时具有更强的鲁棒性。CASAB 的最终输出通过加性融合整合通道与空间两个维度的注意力调制结果：

$$Y_{\mathrm{casab}}=(\boldsymbol{X}_1\odot M_{\mathrm{channel}})+(\boldsymbol{X}_2\odot M_{\mathrm{spatial}}) \tag{5}$$

式中：Y_{casab} 为经双重注意力增强的特征；$\boldsymbol{X}_1$，$\boldsymbol{X}_2$ 表示分别送入通道注意力和空间注意力分支；⊙为逐元素乘法。该加性融合策略允许网络根据输入特征的统计特性自适应平衡两类注意力机制的相对贡献，避免乘性融合可能导致的信息抑制效应。

值得指出，CASAB 在架构上与传统注意力模块如 CBAM 存在本质差异。其采用“串联局部卷积＋并联全局注意力”的混合拓扑，先通过卷积增强底层纹理，再并行执行通道与空间重标定，有效规避了串行注意力的顺序偏置。同时，该模块引入包含最小值与求和的多维空间描述符，显著增强了对阴影等低对比度结构的感知鲁棒性，并有机结合深度可分离卷积与多维注意力，在极低参数代价下实现了局部纹理与全局上下文的互补增强。

由于编码器各级均嵌入 CASAB，因此多尺度特征在传递至瓶颈层与解码器之前已完成显著性重标定，形成判别性更强的层级特征金字塔，为后续的 DPCF 与 TSE 的特征重建奠定了高质量的表征基础。此外，CASAB 的轻量设计使得模块集成对模型总体参数量与计算复杂度的增量影响控制在可接受范围内，满足实时语义分割任务对效率的约束要求。

2.3 上下文瓶颈模块

本文优化设计 LightPCT 模块，在不显著增加

计算负担的前提下捕获全局上下文信息。上下文瓶颈模块 LightPCT 结构如图 3 所示。

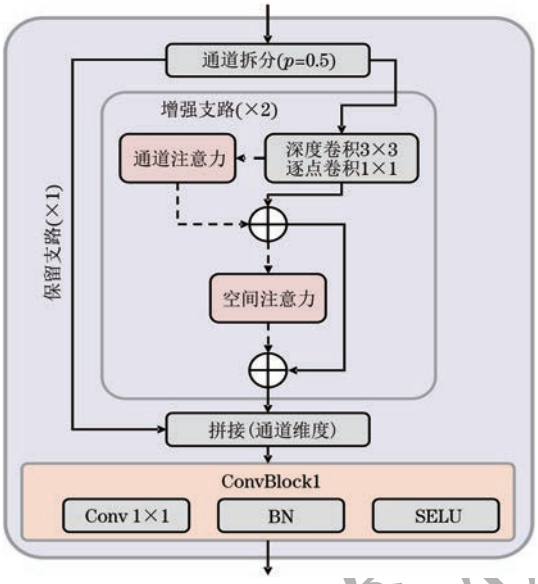

图 3 上下文瓶颈模块 LightPCT 结构

Fig.3 Structure of the context bottleneck module LightPCT

该模块首先将输入的瓶颈层特征按通道比例($p=0.5$)拆分为保留支路与增强支路,仅对增强支路的特征依次执行 3×3 深度可分离卷积、1×1 投影卷积以及通道和空间注意力操作。处理后的增强特征与被保留的原始特征进行拼接,最后通过 1×1 卷积进行通道融合,将维度恢复至原始大小。这种设计使得模型能够以极小的代价实现类 Transformer 的全局上下文建模能力。

2.3.1 通道拆分与保留

输入瓶颈特征按比例划分为保留支路 $\boldsymbol{X}_{\text{pres}}$ 与增强支路 $\boldsymbol{X}_{\text{enh}}$,如式(6)所示:

$$(\boldsymbol{X}_{\text{pres}},\boldsymbol{X}_{\text{enh}})=\text{Split}(\boldsymbol{X},p) \tag{6}$$

式中:$\boldsymbol{X}_{\text{enh}}\in R^{C_1\times H\times W}$,$C_1=p\cdot C$ 表示拆分后的通道数;Split(•,p)表示按通道比例 p 划分,这样后续的复杂计算仅在占比较小的增强支路上进行,从而有效减少计算量。

2.3.2 局部卷积增强

增强支路处理特征 $\boldsymbol{X}_{\text{enh}}$ 经历了一系列处理以提取和增强上下文信息。通过 1 个由 3×3 深度可分离卷积和 1×1 投影卷积组成的复合函数 Φ_{conv} 进行局部特征提取。

随后,通过级联通道与空间注意力建模特征依赖关系,前者利用全局平均与最大池化经 MLP 调制通道权重,后者基于通道统计量通过卷积聚焦空间显著区域,最终输出增强特征 $\hat{\boldsymbol{X}}_{\text{enh}}$。

最后,将经过复杂处理的增强支路特征 $\hat{\boldsymbol{X}}_{\text{enh}}$ 与被直接保留下来的原始支路特征 $\boldsymbol{X}_{\text{pres}}$ 沿通道拼接,并经 1×1 卷积完成特征融合与维度恢复,得到 LightPCT 模块的最终输出 Y_{pct}。Y_{pct} 如式(7)所示:

$$Y_{\text{pct}}=\phi_{1\times1}(\text{Concat}(\boldsymbol{X}_{\text{pres}},\hat{\boldsymbol{X}}_{\text{enh}})) \tag{7}$$

式中:Concat(•,•)为通道维数拼接操作;$\phi_{1\times1}(\cdot)$表示1×1卷积。通过这种“部分处理”的设计,LightPCT 模块能够以较低的成本高效地建模类 Transformer 的长距离上下文依赖关系。

2.4 轴向注意细结构解码器

为提升对道路等细长目标的分割连续性,解码器中集成了 DPCF 与 TSE 模块,如图 4 所示。DPCF 模块通过将来自特征编码器的特征与上采样特征分组并自适应加权,有效缓解了信息类型差异过大而难以直接有效结合的问题。TSE 模块则通过水平和垂直方向的条带池化生成轴向注意力图,并配合深度可分离卷积,显式地强化了道路等细长结构的连通性与边界细节信息。

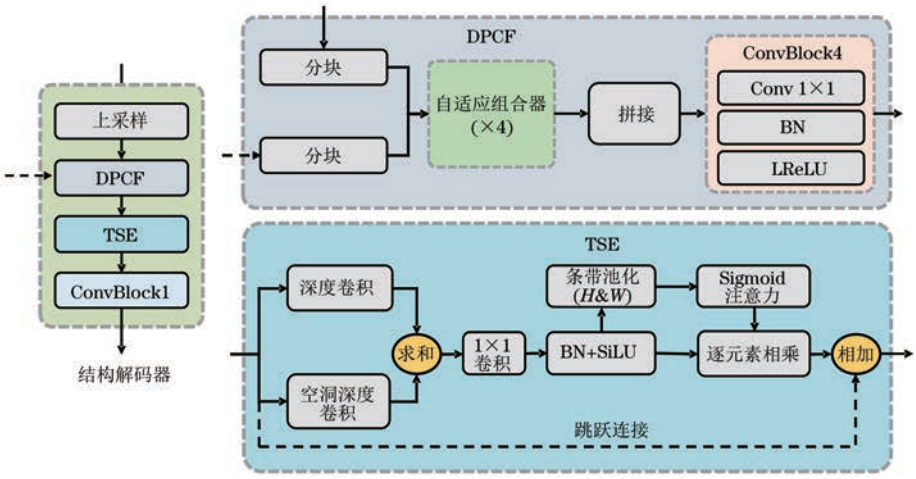

图 4 结构细化解码器示意图

Fig.4 Schematic diagram of the structure refinement decoder

2.4.1 双路径通道融合

为了有效融合解码器中来自上采样路径的高层语义特征 $\boldsymbol{X}_{\text{high}}$ 与来自跳跃连接的低层细节特征 $\boldsymbol{X}_{\text{low}}$，模型采用了 DPCF 模块。该模块旨在通过自适应加权的方式缓解两者之间的“语义鸿沟”。DPCF 模块首先将通道数为 C 的 $\boldsymbol{X}_{\text{high}}$ 与 $\boldsymbol{X}_{\text{low}}$ 沿通道维度均分为 4 组，如式(8)所示：

$$\boldsymbol{X}_{\text{high}}=[H_1,\cdots,H_4],\boldsymbol{X}_{\text{low}}=[L_1,\cdots,L_4] \tag{8}$$

式中：H_k 和 L_k 表示第 k 个通道分组，$k=1,2,3,4$。随后，每一个子块对(H_k，L_k)通过一个可学习的门控机制进行自适应融合。一个标量参数 d_k 经过 Sigmoid 函数生成权重 α_k，用于动态调整跳层信息 L_k 与上采样语义信息 H_k 的融合比例，得到融合后的子块 F_k。

$$F_k=\alpha_k L_k+(1-\alpha_k)H_k \tag{9}$$

最后，所有融合后的子块被重新拼接，并通过 1 个 1×1 的逐点卷积 $\phi_{1\times1}$ 进行最终信息整合，生成既保留了精细细节又与高层语义特征 F 对齐。

$$F=\phi_{1\times1}([F_1;F_2;F_3;F_4]) \tag{10}$$

这种方式可以有效减少因语义不一致导致的分割错检和边缘混乱问题。

2.4.2 薄结构增强

为进一步强化对道路等细长结构的感知，DPCF 模块输出的特征 F 会被送入 TSE 模块。该模块首先利用并行的 3×3 深度卷积和空洞率为 3 的 3×3 深度卷积捕获初步的轴向纹理特征 Y_{tse}。随后，模块的核心组件条带池化能够有效地理解和利用图像中在水平或垂直方向上相距很远的特征之间的关联关系，这对于识别像道路这样具有细长、连续特性的目标非常有帮助，同时使模型保持分割结果的连通性，减少断裂。

具体而言，特征 Y_{tse} 分别沿着宽度和高度进行全局平均池化，生成 2 个一维的上下文向量。这 2 个向量再通过各自独立的 1×1 卷积进行编码，得到水平上下文 $\boldsymbol{A}_h$ 和垂直上下文 $\boldsymbol{A}_w$。这 2 个上下文向量相加后，被广播回原始的 $H\times W$ 尺寸，并通过 Sigmoid 函数生成 1 个同时关注水平与垂直上下文的轴向注意力图 $\boldsymbol{A}_{\text{strip}}$。最后，该注意力图 $\boldsymbol{A}_{\text{strip}}$ 被施加于初步提取的轴向特征 Y_{tse} 上(通过逐元素乘法⊙)，并通过残差连接与模块的原始输入 F 相加，得到最终增强后的解码特征 Z，如式(11)所示：

$$Z=F+\boldsymbol{A}_{\text{strip}}\odot Y_{\text{tse}} \tag{11}$$

TSE 模块利用上述轴向交互显式地建模长距离依赖关系，能够有效提升分割结果中道路的连通性，抑制断裂和锯齿现象。通过 DPCF-TSE 的协同作用，解码阶段既能保持跳层细节信息，又能在道路纵横方向施加强约束，为非结构化道路分割带来更高的连通性与边界质量。

2.5 多级损失函数

为兼顾像素分类精度、区域重叠质量与细边界连续性，AXON-Net 训练阶段引入多级复合损失 L_{total}：

$$L_{\text{total}}=\alpha+\beta \tag{12}$$

$$\alpha=\lambda_{\text{WCE}}L_{\text{WCE}}+\lambda_{\text{Dice}}L_{\text{Dice}} \tag{13}$$

$$\beta=\lambda_{\text{lap}}L_{\text{lap}}+\lambda_{\text{sobel}}L_{\text{sobel}} \tag{14}$$

式中：λ_{WCE}，λ_{Dice}，λ_{lap}，λ_{sobel} 为可调权重，实验默认值均为 1.0；α 和 β 分别表示区域损失项与边界损失项。

像素级交叉熵损失(L_{WCE})是最基础的损失项，旨在保证模型在像素级别的分类精度。通过惩罚模型对每个像素类别的错误预测来引导训练，其定义如下：

$$L_{\text{WCE}}=-\frac{1}{N}\sum_i\sum_c w_c y_{i,c}\log_a p_{i,c} \tag{15}$$

式中：N 是图像中的有效像素总数；w_c 是像素属于类别的独热真值标签；$y_{i,c}$ 是模型预测像素属于类别的概率；$p_{i,c}$ 是为应对数据集中可能存在的类别不平衡问题而引入的可选类别权重，用于调整不同类别在损失计算中的重要性。

Dice 损失(L_{Dice})为优化预测区域与真实区域的重叠度，特别是在处理目标结构连通性时，引入了 Dice 损失。L_{Dice} 直接衡量预测结果和真值之间的集合相似性，如式(16)所示：

$$L_{\text{Dice}}=1-\frac{2\sum_i p_i y_i}{\sum_i p_i+\sum_i y_i+\zeta} \tag{16}$$

式中：p_i 和 y_i 分别是像素 i 的预测概率和真值标签(前景类)；ζ 是防止分母为零的平滑项。根据任务需求，此项可替换为 Tversky 损失或结构化损失等兼顾边界质量或假阳性/假阴性平衡的变种。

边界损失 L_{lap} 和 L_{sobel} 为了进一步提升分割边界的清晰度和连续性，模型还引入了基于图像梯度的边界损失项。L_{lap} 和 L_{sobel} 通过计算预测结果与真值标签在梯度图上的差异，显式地惩罚模糊或错误的边界，引导网络生成更锐利的边缘。

3 实验结果及分析

3.1 实验设置及数据集

本研究以基于印度驾驶数据集(IDD)[20]与越野空间检测(ORFD)[21]数据集为主要数据来源。针

对这两个异构数据集，本研究均进行了数据清洗与重构，以消除标签噪声并统一训练标准。

鉴于原始数据集包含复杂的全景分割标签以及极端气候场景数据，本研究制定了以下客观筛选标准进行数据清洗与重构：在语义完整性方面，利用掩膜统计信息，剔除道路标签缺失、标注严重偏离物理边界的样本，确保模型能够学习到正确的道路拓扑特征并减少标签噪声的负面影响；在环境典型性方面，主要保留气候条件为常态（晴天、多云）的场景，剔除降雨、浓雾或极端低光照导致的视觉语义完全模糊的样本；在非结构化场景定义上，筛选符合无清晰车道线、无路缘石限制、边界由植被或泥土构成的典型非结构化道路场景（如乡间小道等），如图 5 所示。通过上述客观准则，最终构建了更具针对性的 IDD-UR 与 ORFD-AV 数据集用于模型训练与验证。

图 5 保留样本与剔除样本对比

Fig. 5 Comparison between retained samples and excluded samples

3.1.1 IDD 数据预处理

IDD 是公开的非结构化场景全景分割数据集，其 Part I 包含 7 034 张训练图像和 1 055 张验证图像，如植被、其他可行驶区域、行人、轿车等 41 个类别，类别样本占比排名如表 1 所示（部分）。

表 1 IDD 部分类别样本统计

Table 1 IDD partial category sample statistics

类别	计数/个	占比/%
障碍物	149 696	21.11
植被	140 828	19.86
杆/柱	91 727	12.94
摩托车	40 471	5.71
自行车	35 555	5.01
可行驶区域	18 994	2.68
不可行驶区域	9 882	1.39

基于非结构化道路分割任务的特定需求，对原始的 IDD 数据进行针对性的筛选与重构。首先，解析原始 JSON 标注文件，提取“可行驶区域”类别并滤除无关语义标签，生成分辨率为 1 920×1 080 像素的 PNG 格式掩膜。

随后，实施标签二值化处理，将非结构化道路区域划分为前景（像素值为 255），其余为背景（像素值为 0），并存储为 8 位深度图像，如表 2 所示。为确保标注精度，将掩膜与原图叠加进行可视化校验，由人工逐张筛查，剔除了标签缺失、标注严重偏离及非典型道路场景的样本，如图 5 所示。最终，本文构建得到 5 700 张训练集，验证集为 955 张的数据集，为主数据集命名为 IDD-UR。

表 2 模型训练标签映射

Table 2 Label mapping for model training

类别	像素值
道路面	255
背景	0

3.1.2 ORFD 数据预处理

ORFD 数据集包含 12 198 对同步的 40 线 LiDAR 点云和 RGB 图像（分辨率为 1 280×720 像素），采集自中国多种越野场景（林地、农田等场景），覆盖晴天、雨天和各种光照条件，数据集像素标注情况如表 3 所示。

表 3 ORFD 像素分类

Table 3 Pixel classification of ORFD

类别	颜色	像素值
白色	通行区	255
黑色	障碍威胁区	0
灰色	远处无关区	128

本文对 ORFD 数据集实施同等清洗。剔除冗余 LiDAR 及极端天气样本，如图 5 所示，仅保留 RGB 图像。标签生成沿用 IDD 的二值化范式（前景 255/背景 0），并由人工通过可视化进行逐张筛查，剔除劣质数据。最终筛选出训练集 6 178 张，验证集 1 158 张，得到辅助数据集并将该数据集命名为 ORFD-AV。

本实验遵循统一基准原则，所有模型均不使用预训练权重并进行从头复现，以消除预训练偏差。训练阶段开启自动混合精度（AMP）以兼顾效率与稳定性，并基于验证集指标实时更新保存最优模型权重。U-Net，SegNet[22] 等对比模型的学习率为 4.5×10^{-2}，损失函数为 ProbOhemCrossEntropy2d。本文模型的实验参数如表 4 所示。

表 4 本文模型的实验参数

Table 4 The experimental parameters of the proposed model in this paper

参数名称	配置详情
学习率	0.2
损失函数	L_{total}
软件环境	PyTorch, CUDA 11.8
硬件配置	单卡 NVIDIA GeForce RTX 3090 (24 GB)
输入尺寸/像素	256×512
批量大小	8
训练轮次/次	200
优化器	SGD (Momentum 为 0.9, Weight Decay 为 5×10^{-4})
训练方式	从头开始训练

3.2 测评标准

为评估模型的综合性能,本研究重点评估道路区域的像素级分类准确性、区域重叠质量以及精确率与召回率的平衡性。具体采用以下核心指标。

交并比(IoU,R_{IoU})是语义分割任务中核心评价指标,直接衡量模型预测的道路区域与真实标注区域的重叠程度。本文重点关注道路类别的交并比(IoU_Road,$R_{\text{IoU\_Road}}$),其计算公式如下:

$$R_{\text{IoU\_Road}}=\frac{N_{\text{TP}}}{N_{\text{TP}}+N_{\text{FP}}+N_{\text{FN}}} \tag{17}$$

式中:N_{TP} 表示被正确预测的道路像素;N_{FP} 表示被错判为道路的背景像素;N_{FN} 表示被漏判的道路像素。

F1 值(F_1)是精确率(Precision,P)和召回率(Recall,R)的调和平均数,能够综合反映模型在保持高精确度的同时,避免漏检,对于处理类别不平衡问题尤为重要。

$$P=\frac{N_{\text{TP}}}{N_{\text{TP}}+N_{\text{FP}}} \tag{18}$$

$$R=\frac{N_{\text{TP}}}{N_{\text{TP}}+N_{\text{FN}}} \tag{19}$$

$$F_1=2\times\frac{P\times R}{P+R} \tag{20}$$

除此之外,本文还采用像素准确率(PA)、乘加累加操作数(MAC)、推理速度来评估所提模型的性能。

3.3 模型对比

为验证 AXON-Net 的性能表现,本文在 IDD-UR 数据集(主数据集)和 ORFD-AV 数据集(辅助数据集)上,将该模型与 U-Net、SegNet[22] 以及轻量化网络 CGNet[23]、ERFNet[24],以及近几年提出的分割模型 LCNet[25] 和 BEVANet[26] 进行对比实验。2 个数据集的实验结果分别如表 5 和表 6 所示,加粗表示最优结果(下同),下划线表示次优结果。

表 5 不同方法在 IDD-UR 数据集上的实验结果

Table 5 Experimental results among different methods on the IDD-UR dataset

方法	IoU_Road	mIoU	F1 值	Precision	PA	Recall	Parameter/10^3	MAC/10^9	推理速度/(帧·s^{-1})
U-Net[13]	0.939 0	0.957 7	0.968 7	0.964 1	0.982 6	0.981 4	13.40	145.49	58.59
SegNet[22]	0.915 7	0.941 1	0.956 0	0.941 2	0.975 5	0.971 3	29.44	188.29	49.61
ERFNet[24]	0.938 5	0.957 0	0.968 2	0.948 2	0.982 2	**0.989 2**	2.06	17.28	146.41
LinkNet[27]	0.924 3	0.947 0	0.960 6	0.941 3	0.977 9	0.980 8	11.53	14.20	252.60
EDANet[28]	0.944 5	0.961 4	0.971 5	0.958 9	0.984 1	0.984 4	0.68	5.22	128.74
LEDNet[29]	0.913 0	0.939 1	0.954 5	0.934 2	0.974 5	0.975 7	0.92	6.69	66.69
U^2-Net[30]	0.945 6	0.962 4	0.972 0	0.975 4	0.984 7	0.968 7	44.01	58.95	49.96
CGNet[23]	0.946 3	0.962 7	0.972 4	0.964 4	0.984 7	0.980 5	0.49	4.17	74.35
LETNet[31]	0.922 4	0.946 4	0.971 2	0.971 2	0.978 1	0.948 3	0.76	8.33	20.80
LCNet[25]	0.930 4	0.951 5	0.964 0	0.951 9	0.980 0	0.976 3	0.51	15.96	154.09
BEVANet[26]	0.945 8	0.962 4	0.972 1	0.965 5	0.984 7	0.978 9	58.60	35.08	79.55
Ours	**0.953 0**	**0.968 0**	**0.976 0**	**0.978 6**	**0.987 1**	0.982 2	8.49	40.32	38.19

在 IDD-UR 主数据集上,实验结果表明,AXON-Net 在 IoU_Road、mIoU 及 PA 等指标上均优于对比模型。其中,IoU_Road 为 0.953 0,较 U-Net 提升约 1.4 百分点,并高于 SegNet 与 LEDNet[29]。

针对 ERFNet 的指标差异分析,虽然 ERFNet 的 Recall 略高于本文模型,但其 Precision 为 0.948 2,低于本文模型。经过分析认为这种“高召回、低精确”的现象与其使用的空洞卷积在大感受野下产生的边缘平滑效应有关,使得模型倾向于将路边草地等纹理相似的背景区域误分类为道路。相比之下,AXON-Net 通过 CASAB 模块的特征抑制与 TSE 的边界约束,一定程度上减少了此类误判,实现了 Recall 与 Precision 的相对平衡。

表 6 不同方法在 ORFD-AV 数据集上的实验结果

Table 6 Experimental results among different methods on the ORFD-AV dataset

方法	IoU_Road	mIoU	F1 值	Precision	PA	Recall	Parameter/10^3	MAC/10^9	推理速度/(帧·s^{-1})
U-Net[13]	0.797 9	0.862 1	0.887 6	0.937 5	0.942 9	0.842 7	13.40	145.49	58.59
SegNet[22]	0.800 5	0.860 3	0.889 2	0.871 9	0.939 5	0.907 2	29.44	188.29	49.61
ERFNet[24]	0.835 6	0.8867	0.910 4	0.924 3	0.952 8	0.897 0	2.06	17.28	146.41
LinkNet[27]	0.823 9	0.883 6	0.903 5	0.931 9	0.949 9	0.876 7	11.53	14.20	252.60
EDANet[28]	0.830 9	0.891 0	0.907 7	0.944 9	0.960 5	0.873 3	0.68	5.22	128.74
LEDNet[29]	0.845 4	0.893 1	0.916 2	0.918 7	0.955 3	0.913 7	0.92	6.69	66.69
U^2-Net[30]	0.830 0	0.883 6	0.907 1	0.938 8	0.951 9	0.877 4	44.01	58.95	49.96
CGNet[23]	0.849 3	0.894 8	0.918 5	0.897 1	0.955 3	0.941 0	0.49	4.17	74.35
LETNet[31]	0.875 3	0.904 2	0.933 5	0.941 0	0.964 7	0.926 2	0.76	8.33	20.80
LCNet[25]	0.872 1	0.901 2	0.931 7	**0.954 0**	0.964 3	0.910 4	0.51	15.96	154.09
BEVANet[26]	0.873 8	0.913 3	0.932 7	0.944 0	0.964 4	0.921 6	58.60	35.08	79.55
Ours	**0.881 0**	**0.917 0**	**0.937 0**	0.927 0	**0.965 0**	**0.947 0**	8.49	40.32	38.19

此外，实验中观察到不同年份模型性能的差异，2020 年提出的 CGNet 在 IoU_Road 上优于较近年份的 LETNet[31]、LCNet、BEVANet 的模型，这反映了模型架构与数据特征的适配性问题。CGNet 的上下文引导机制在 IDD-UR 这种边界模糊的场景下表现出较好的适应性；而 LETNet、LCNet 虽然参数效率更高，但在处理该特定非结构化数据时，其特征提取优势未得到充分发挥。

在 ORFD-AV 辅助验证集上，AXON-Net 表现出一定的泛化能力。鉴于该部分实验旨在考察模型对跨域可行驶区域的提取效果，故重点关注道路交并比(IoU_Road)指标。由于 ORFD-AV 包含林地、农田等复杂非结构化道路场景，且图像清晰度相对较低，因此各模型的性能指标普遍低于 IDD-UR 数据集。尽管受此影响，模型性能普遍出现下滑，但 AXON-Net 仍取得了 0.881 0 的 IoU_Road。此外，在该数据集上 LETNet 与 LCNet 的表现优于 CGNet，这与 IDD-UR 上的结果相反。这一变化说明不同模型对数据分布的敏感度存在差异，LETNet 在处理跨域差异较大的数据时表现相对稳健，而 CGNet 则在特定上下文特征明显的场景中更具优势。AXON-Net 在两个数据集上均保持了较稳定的性能，表明其架构设计对于不同非结构化场景具有较好的适应性。

值得注意的是，在 ORFD-AV 数据集上 AXON-Net 的 Precision 略低于部分对比模型。分析认为，主要源于模型在跨域场景下对特征敏感度的侧重不同。为了在模糊区域保持较好的道路连通性(Recall 较高)，模型对潜在可行驶区域的响应更为积极，这虽然可能引入少量边界噪声，却有效减少了漏检。相较于倾向抑制模糊边界的对比方法，AXON-Net 虽然精确率略低，但得益于更完整的道路拓扑捕获能力，在综合反映分割质量的 IoU_Road 指标上仍取得了较优表现。特别地，AXON-Net 在 IoU_Road 指标上略微优于 BEVANet。这反映出在复杂的非结构化场景道路中，AXON-Net 凭借 LightPCT 模块倾向于优先保障道路的全局连通性，其 Recall(0.947 0)相对较高。这种对潜在可行驶区域相对积极的响应策略，使得模型在应对非典型路面时保持了较平稳的综合表现。

在模型效率方面，AXON-Net 在精度与速度之间取得了良好的平衡。一方面，虽然 CGNet 参数量更小，只有 0.49×10^3，但其分割精度 IoU_Road 在两个数据集上分别比本文模型低 0.67 和 3.17 百分点。AXON-Net 的参数量为 8.49×10^3(约为 U-Net 的 63%)；另一方面，与推理速度较快的 BEVANet 相比，AXON-Net 的参数量和计算复杂度相对较小。AXON-Net 在单卡 RTX 3090 上的推理速度为 38.19 帧/s，满足自动驾驶通常要求的 30 帧/s 实时阈值，这表明通过引入注意力组件，模型在控制计算成本的同时，有效提升了特征表达能力。

3.4 消融实验

为验证 LightPCT、CASAB 及 DPCF-TSE 模块的协同有效性，本节以轻量化 U-Net(A 组)为基线进行消融实验，以 IDD-UR 数据集为消融实验的训练数据集。消融实验结果如表 7 所示。

实验中首先观察到一个“反常”现象：在基线模型中单独引入 LightPCT(B 组)或 CASAB(C 组)并

表 7 消融实验结果

Table 7 Ablation experimental results

编号	Baseline	LightPCT	CASAB	DPCF-TSE	IoU_Road	mIoU	F1 值	Precision	PA	Recall
A	√	×	×	×	0.939 1	0.957 7	0.968 7	0.961 2	0.982 6	0.981 4
B	√	√	×	×	0.935 2	0.954 1	0.966 5	0.949 2	0.981 8	0.981 1
C	√	×	√	×	0.924 3	0.947 2	0.965 2	0.946 3	0.978 7	0.976 2
D	√	×	×	√	0.941 2	0.959 3	0.975 1	0.959 4	0.983 8	0.981 3
E	√	√	√	×	0.930 0	0.951 5	0.964 5	0.949 8	0.986 6	0.978 4
F	√	√	×	√	0.813 1	0.867 4	0.897 2	0.866 9	0.941 5	0.935 5
G	√	×	√	√	0.944 0	0.961 2	0.971 1	0.961 7	0.984 4	0.982 1
H (ours)	√	√	√	√	**0.953 0**	**0.968 0**	**0.976 0**	**0.978 6**	**0.987 1**	**0.982 2**

未带来预期的性能提升，反而导致 IoU_Road 分别微幅下降至 0.935 2 和 0.924 3。这一结果否定了模块的独立增益假设，其深层机理可归因于“语义-分布错位”。具体而言，LightPCT 位于瓶颈层，强化了全局长程依赖关系，但在缺乏后端 DPCF-TSE 进行细结构恢复配合的情况下，过于强势的全局上下文倾向于抑制局部高频细节，导致基线解码器难以将抽象特征还原为精细边界。同样，CASAB 在编码端执行了强烈的多统计量特征重标定，显著改变了原始特征的统计分布。由于基线解码器未针对这种“锐化”后的特征分布进行适配，导致编解码特征分布失配，使得重标定收益无法体现，甚至可能因解码器对背景噪声的敏感性而产生负面效果。相比之下，单独使用 DPCF-TSE(D 组)解码器带来了小幅提升，IoU_Road 为 0.941 2，验证了其作为独立解码单元的有效性。

在 LightPCT+DPCF-TSE(F 组)与 CASAB+DPCF-TSE(G 组)的性能反差上，这深刻揭示了“噪声级联放大”效应。虽然 G 组的 IoU_Road 为 0.944 0，实现了性能的明显提升，证明 CASAB 生成的高质量特征能被 DPCF 精准解析，但 F 组在移除 CASAB 后性能却出现崩塌式下滑，IoU_Road 为 0.813 1，表明 CASAB 模块在架构中充当了不可或缺的“前置净化器”。在缺失 CASAB 的情况下，原始跳层特征包含大量非结构化场景特有的噪声(如树影、碎石纹理等)。后端 DPCF 模块的自适应门控机制错误地将这些高频噪声识别为有效细节信息进行融合，随后 TSE 进一步对这些错误的纹理响应进行了显式强化。这种“噪声输入—门控误判—增强放大”的级联效应，最终导致分割结果出现严重伪影与性能崩塌。

综上所述，完整模型 H 组的 IoU_Road 为 0.953 0，证实了 AXON-Net 是一个高度耦合的协同系统。实验结果表明，CASAB 负责前端特征的重标定与判别性增强，LightPCT 负责中端全局上下文信息的捕获，而 DPCF-TSE 负责后端细粒度结构的重建。LightPCT 提供的全局上下文必须建立在 CASAB 与 DPCF-TSE 构成的稳定特征编解码通路上，才能发挥最大效能。如图 6 所示的激活热力图，H 组对道路区域的激活最准确、最集中，直观地验证了该协同设计的有效性。

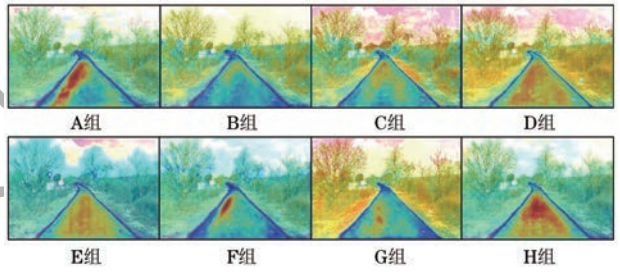

图 6 消融实验模块激活热力图

Fig.6 Activation heatmaps for the ablation experimental module

3.5 可视化结果与分析

为了直观展示 AXON-Net 在非结构化道路分割任务中的综合性能，图 7 提供了本模型与多种基准方法在不同典型场景下的定性对比结果。其中，左侧两列展示了 IDD-UR 主数据集的分割效果，右侧一列展示了 ORFD-AV 辅助数据集的测试结果。

在 IDD-UR 场景的可视化对比中，传统模型(如 U-Net、SegNet)在处理车辆周边区域时，分割边缘粗糙且呈锯齿状，难以精准贴合轮廓，导致道路边缘碎片化。U^2-Net 则存在欠分割现象，未能完整识别道路区域。部分轻量化模型(如 LETNet、LCNet 等)在应对车辆遮挡时，容易丢失遮挡区后方的道路语义信息。BEVANet 则在车辆旁边产生微小的空洞。相比之下，AXON-Net 生成的掩膜边界更平滑，与车辆及障碍物边缘贴合较好，有效改善了边缘粘连与误分类问题。这一表现佐证了 DPCF-TSE

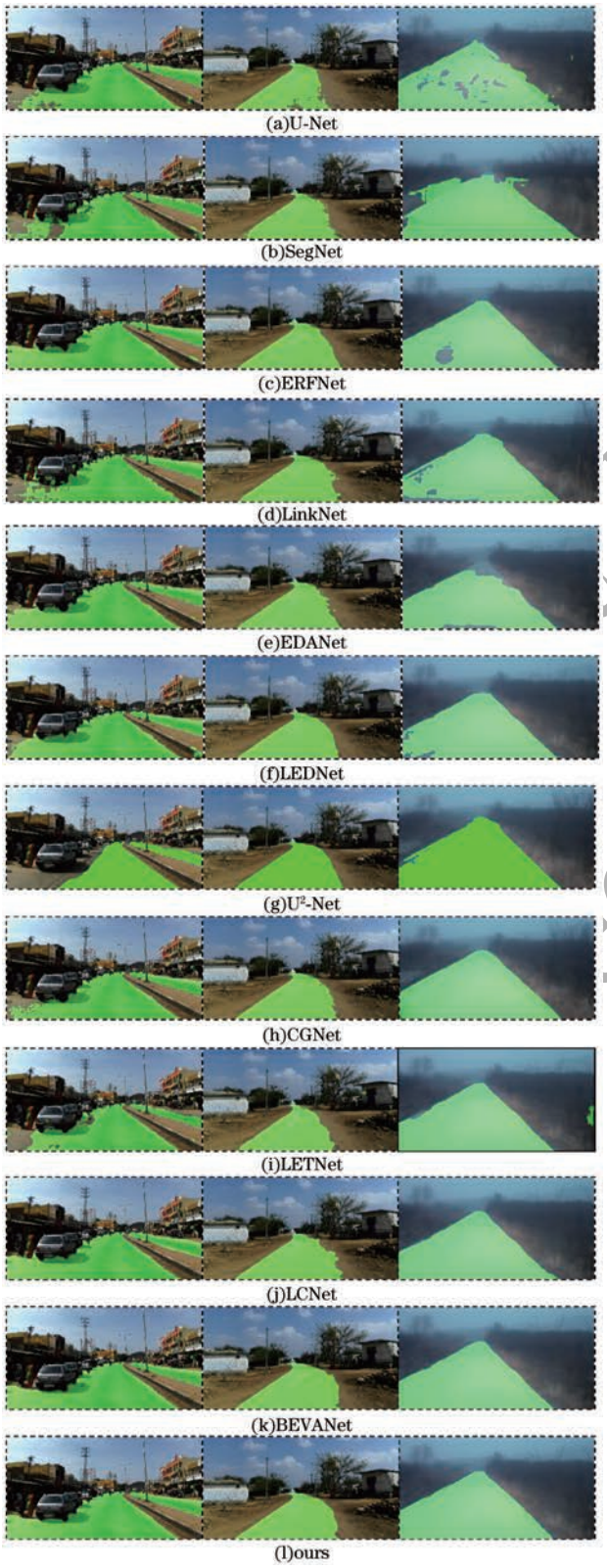

图 7 不同模型分割结果可视化对比

Fig. 7 Visualization comparison of segmentation results from different models

模块在多尺度特征融合与边缘细节增强方面的积极作用。

在全局连通性与背景抑制方面，LinkNet 及部分轻量化模型虽然能定位道路主体，但细节上存在小空洞或错误包含靠近建筑物的非道路区域。相比之下，AXON-Net 的分割区域更连续均匀，有效减少了对背景纹理的误判。这验证了 CASAB 模块在抑制噪声提升区分度，以及 LightPCT 模块利用长程上下文维持拓扑连通性方面的有效性。

在 ORFD-AV 场景的可视化对比中，AXON-Net 体现了在跨域场景下的适应性。面对纹理复杂且路界模糊的越野环境，U-Net、SegNet 等网络在推理时出现了不同程度的空洞断裂或区域识别偏差。而 AXON-Net 相对 LETNet、LCNet 等网络推理的图片边缘更平滑，在该场景中表现出较好的适应性，倾向于在保持道路区域全局连通性的同时，提供较为准确的边界定位，有效改善了误分割情况。这表明模型学习到的特征表示具有一定的抗干扰能力，能够适应从结构化向非结构化场景的迁移。这一视觉特性在量化指标上得到了印证。分析认为，这得益于 LightPCT 对全局上下文的敏锐捕获，使模型在跨域场景下对非典型路面特征保持了高敏感度。这种侧重保障拓扑完整性而非极端抑制边界噪声的平衡策略，虽然牺牲了微幅的精确率，却换取了 IoU_Road 的优势，证实了其在复杂环境下的综合有效性。

4 结束语

针对非结构化道路分割的上下文信息不足、边界模糊及计算资源受限等问题，本文设计一种轻量化轴向上下文网络 AXON-Net。该模型在编码器端集成通道-空间注意力模块 CASAB 以抑制背景噪声，在瓶颈层引入轻量化部分上下文模块 LightPCT 增强长程依赖关系的捕获能力，并在解码器结合双路径通道融合（DPCF）与细结构增强（TSE）策略，旨在保持计算效率的同时，改善全局拓扑感知与细粒度特征的恢复质量。在 IDD-UR 主数据集与 ORFD-AV 辅助数据集上的实验结果表明，AXON-Net 以 8.49×10^3 的参数量分别取得了 95.3%和 88.1%的道路 IoU。与部分经典及轻量化分割模型相比，该模型在分割精度与模型效率之间呈现出较好的平衡。

后续将从 2 个主要方向展开：1）探索模型量化、知识蒸馏等压缩技术，以评估 AXON-Net 在资源受限平台上的部署潜力；2）关注提升模型在环境更复杂条件下的鲁棒性，如通过研究适用于低光照、雨天、雾天等恶劣天气的数据增强方法或领域自适应技术，以期拓展模型在不同场景下的适用性。

参考文献

[1] YU F, CHEN H, WANG X, et al. BDD100K: a diverse driving dataset for heterogeneous multitask learning[EB/OL]. [2025-11-01]. https://arxiv.org/pdf/1805.04687.

[2] CORDTS M, OMRAN M, RAMOS S, et al. The Cityscapes dataset for semantic urban scene understanding[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D.C., USA: IEEE Press, 2016: 3213-3223.

[3] 林畅, 郭伟, 任哲聪, 等. 基于 Transformer 的目标跟踪与分割统一算法[J]. 计算机工程, 2024, 50(9): 130-141.
LIN C, GUO W, REN Z C, et al. Unification algorithm for object tracking and segmentation based on Transformer[J]. Computer Engineering, 2024, 50(9): 130-141. (in Chinese)

[4] 叶宝林, 孙瑞涛, 李灵犀, 等. 基于融合状态预测的深度强化学习 A2C 的交通信号控制[J]. 计算机工程, 2025, 51(5): 33-42.
YE B L, SUN R T, LI L X, et al. Traffic signal control based on deep reinforcement learning A2C integrated with state prediction[J]. Computer Engineering, 2025, 51(5): 33-42. (in Chinese)

[5] 朱思远, 李佳圣, 邹丹平, 等. 基于半监督学习的非结构化道路缺陷检测算法[J]. 计算机工程, 2025, 51(9): 14-24.
ZHU S Y, LI J S, ZOU D P, et al. Unstructured road defect detection algorithm based on semi-supervised learning[J]. Computer Engineering, 2025, 51(9): 14-24. (in Chinese)

[6] LIU Z Y, YU S Y, ZHENG N N. A co-point mapping-based approach to drivable area detection for self-driving cars[J]. Engineering, 2018, 4(4): 479-490.

[7] HSU C M, LIAN F L, HUANG C M, et al. Detecting drivable space in traffic scene understanding[C]//Proceedings of International Conference on System Science and Engineering (ICSSE). Washington D.C., USA: IEEE Press, 2012: 79-84.

[8] FERNÁNDEZ C, FERNÁNDEZ-LLORCA D, SOTELO M A. A hybrid vision-map method for urban road detection[J]. Journal of Advanced Transportation, 2017(1): 7090549.

[9] SLESS L, SHLOMO B E, COHEN G, et al. Road scene understanding by occupancy grid learning from sparse radar clusters using semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Washington D.C., USA: IEEE Press, 2019: 101-106.

[10] MA B, LAKSHMANAN S, HERO A O. Simultaneous detection of lane and pavement boundaries using model-based multisensor fusion[J]. IEEE Transactions on Intelligent Transportation Systems, 2002, 1(3): 135-147.

[11] 沙宇洋, 陆京涛, 杜浩凡, 等. 适用于导盲场景的多尺度特征融合轻量化道路图像分割算法[J]. 计算机工程, 2025, 51(7): 314-325.
SHA Y Y, LU J T, DU H F, et al. Lightweight road image segmentation algorithm with multi-scale feature fusion for blind guidance scenarios[J]. Computer Engineering, 2025, 51(7): 314-325. (in Chinese)

[12] 金汝宁, 赵波, 李洪平. 一种轻量化非结构化道路语义分割神经网络[J]. 四川大学学报(自然科学版), 2023, 60(1): 66-73.
JIN R N, ZHAO B, LI H P. A lightweight neural network for semantic segmentation of unstructured roads[J]. Journal of Sichuan University (Natural Science Edition), 2023, 60(1): 66-73. (in Chinese)

[13] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 39(4): 640-651.

[14] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]// Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin, Germany: Springer, 2015: 234-241.

[15] 栾庆磊, 毛宜东, 陈梓华, 等. 基于 YOLOv8-AFDP 的红外道路目标检测[J]. 光电工程, 2025, 52(8): 250079.
LUAN Q L, MAO Y D, CHEN Z H, et al. Infrared road target detection based on YOLOv8-AFDP[J]. Opto-Electronic Engineering, 2025, 52(8): 250079. (in Chinese)

[16] SINHA D, EL-SHARKAWY M. Thin MobileNet: an enhanced MobileNet architecture[C]//Proceedings of the 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). Washington D.C., USA: IEEE Press, 2019: 280-285.

[17] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D.C., USA: IEEE Press, 2017: 2881-2890.

[18] 安玉, 葛海波, 何文昊, 等. 基于补偿注意力机制的 Siamese 网络跟踪算法[J]. 计算机工程, 2024, 50(4): 187-196.
AN Y, GE H B, HE W H, et al. Siamese network tracking algorithm based on compensated attention mechanism[J]. Computer Engineering, 2024, 50(4): 187-196. (in Chinese)

[19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2017: 5998-6008.

[20] VARMA G, SUBRAMANIAN A, NAMBOODIRI A, et al. IDD: a dataset for exploring problems of autonomous navigation in unconstrained environments[C]//Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV). Washington D.C., USA: IEEE Press, 2019: 1743-1751.

[21] MIN C, JIANG W Z, ZHAO D W, et al. ORFD: a dataset and benchmark for off-road freespace detection[C]// Proceedings of International Conference on Robotics and Automation (ICRA). Washington D.C., USA: IEEE Press, 2022: 6542-6548.

[22] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[23] 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.

[24] ROMERA E, ALVAREZ J M, BERGASA L M, et al. ERFNet: efficient residual factorized ConvNet for real-time semantic segmentation[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 19(1): 263-272.

[25] SHI M, LIN S, YI Q, et al. Lightweight context-aware network using partial-channel transformation for real-time semantic segmentation[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(7): 7401-7416.

[26] HUANG P M, CHAO I T, HUANG P C, et al. BEVANet:

bilateral efficient visual attention network for real-time semantic segmentation[C]//Proceedings of IEEE International Conference on Image Processing (ICIP). Washington D. C., USA: IEEE Press, 2025: 2778-2783.

[27] CHAURASIA A, CULURCIELLO E. LinkNet: exploiting encoder representations for efficient semantic segmentation [EB/OL]. [2025-11-01]. https://arxiv.org/pdf/1707.03718.

[28] LO S Y, HANG H M, CHAN S W, et al. Efficient dense modules of asymmetric convolution for real-time semantic segmentation[EB/OL]. [2025-11-01]. https://dl.acm.org/doi/epdf/10.1145/3338533.3366558.

[29] WANG Y, ZHOU Q, LIU J, et al. LEDNet: a lightweight encoder-decoder network for real-time semantic segmentation[C]//Proceedings of IEEE International Conference on Image Processing (ICIP). Washington D. C., USA: IEEE Press, 2019: 1860-1864.

[30] QIN X B, ZHANG Z C, HUANG C Y, et al. U^2-Net: going deeper with nested U-structure for salient object detection [J]. Pattern Recognition, 2020, 106: 107404.

[31] XU G A, LI J C, GAO G W, et al. Lightweight real-time semantic segmentation network with efficient transformer and CNN[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(12): 15897-15906.

文字编辑 薛晋栋
栏目编辑 赖玉玲