

# 一种基于增强校准策略的自动驾驶多任务分割网络

游玙瑞<sup>1</sup>, 宋春林<sup>1</sup>, 徐旭辉<sup>2</sup>

<sup>1</sup>同济大学信息与通信工程系, 上海

<sup>2</sup>同济大学海洋地质国家重点实验室, 上海

收稿日期: 2026年1月2日; 录用日期: 2026年2月2日; 发布日期: 2026年2月10日

---

## 摘要

在自动驾驶中, 可行驶区域检测与车道线检测是两项语义上紧密相关的关键视觉任务。然而多数方法仍将两项任务独立处理, 未能有效利用任务间的关联, 此外, 大规模网络结构庞大、计算量高, 难以满足实时嵌入式车载系统的应用需求。本文提出一种基于增强 - 校准策略的轻量化多任务语义分割网络GMSANet (Grouped Multi-Scale Attention Network), 用于同时处理可行驶区域检测与车道线检测任务。模型以GSConv-ESP (Grouped Shuffle Convolution-Efficient Spatial Pyramid)结构为编码器基础, 通过增强 - 校准设计, 在保持高精度的同时降低模型复杂度。网络引入分组多尺度注意力(Grouped Multi-Scale Attention, GMSA)模块, 以分组条带卷积提升对方向性与关键区域的响应; 同时提出多尺度动态矩形自校准(Multi-scale Dynamic Rectangular Self-Calibration Module, MD-RCM)模块, 通过多尺度感受野实现对目标区域的自适应校准。实验结果表明, 在BDD100K数据集上, GMSANet以仅2.9M参数和6.45 G FLOPs的复杂度, 实现了92.8%的可行驶区域mIoU、85.1%的车道线准确率与34.0%的车道线IoU, 性能优于YOLOP和A-YOLOM等轻量化方法。模型推理速度可达55 FPS, 具备良好的实时性与嵌入式部署潜力。

## 关键词

语义分割, 多任务学习, 驾驶感知, 可行驶区域检测, 车道线检测

---

# A Multi-Task Segmentation Network Based on an Enhancement-Calibration Strategy for Autonomous Driving

Yurui You<sup>1</sup>, Chunlin Song<sup>1</sup>, Xuhui Xu<sup>2</sup>

<sup>1</sup>Department of Information and Communication Engineering, Tongji University, Shanghai

<sup>2</sup>State Key Laboratory of Marine Geology, Tongji University, Shanghai

Received: January 2, 2026; accepted: February 2, 2026; published: February 10, 2026

## Abstract

Drivable area detection and lane line detection are two related perception tasks in autonomous driving. Many prior works treat them as two separate problems and do not explicitly model their shared features. In many cases, the required computation and model size exceed what typical embedded platforms can support in real time. This paper introduces the Grouped Multi-Scale Attention Network (GMSANet), a lightweight multi-task semantic segmentation network based on an enhancement-calibration strategy, designed for drivable area and lane line detection. The model is built upon the GSConv-ESP (Grouped Shuffle Convolution-Efficient Spatial Pyramid) encoder and uses an enhancement-calibration design, which reduces model complexity while preserving high segmentation accuracy. We integrate a Grouped Multi-Scale Attention (GMSA) module into the network. GMSA applies grouped strip convolutions at multiple scales, which increases the sensitivity to directional features and key regions. Also, we introduce a Multi-scale Dynamic Rectangular Self-Calibration Module (MD-RCM) that calibrates target regions by adjusting receptive fields across multiple scales. Experimental results on the BDD100K dataset show that GMSANet, with only 2.9 M parameters and 6.45 G FLOPs, achieves 92.8% mIoU for drivable area segmentation, 85.1% LaneAccuracy, and 34.0% LaneIoU, outperforming lightweight models such as YOLOP and A-YOLOM. The model further achieves an inference speed of 55 FPS, demonstrating strong real-time capability and suitability for embedded deployment.

## Keywords

Semantic Segmentation, Multi-Task Learning, Driving Perception, Drivable Area Detection, Lane Line Detection

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

视觉感知是智能驾驶系统决策与安全控制的重要任务，可行驶区域检测能够帮助车辆识别路面区域，降低驶入危险区域或人行道的风险。车道线检测使得车辆可以保持在当前车道，并且给变道或转弯等操作提供辅助驾驶的参考。这两项任务都是对路面元素的视觉感知任务，二者相辅相成，为自动驾驶系统的高级功能提供了基础支持。

传统的感知系统在处理这两个任务时，主要依赖直接处理图像的方法，例如颜色划分[1]、边缘检测霍夫变换[2]、图像处理和感兴趣区域操作[3]等等。这类基于预先设定规则的方法在应对复杂环境时，鲁棒性和泛化能力明显不足。随着深度学习技术的不断发展，基于卷积神经网络的图像分割方法得到了广泛关注与研究[4]。UNet[5]、SegNet[6]、PSPNet[7]和DFBNet[8]等网络结构凭借其对空间特征的高效建模能力，可以实际应用于可行驶区域检测任务；SCNN[9]、CLRerNet[10]、FENet[11]等方法则针对车道线的结构化特征进行了专门设计，在车道线检测任务中较传统方法有更好的效果。虽然上述方法比起传统方法有所进步，但它们的不足之处在于仅针对单一任务进行结构设计，忽视了不同视觉任务之间的内在联系和协同作用。

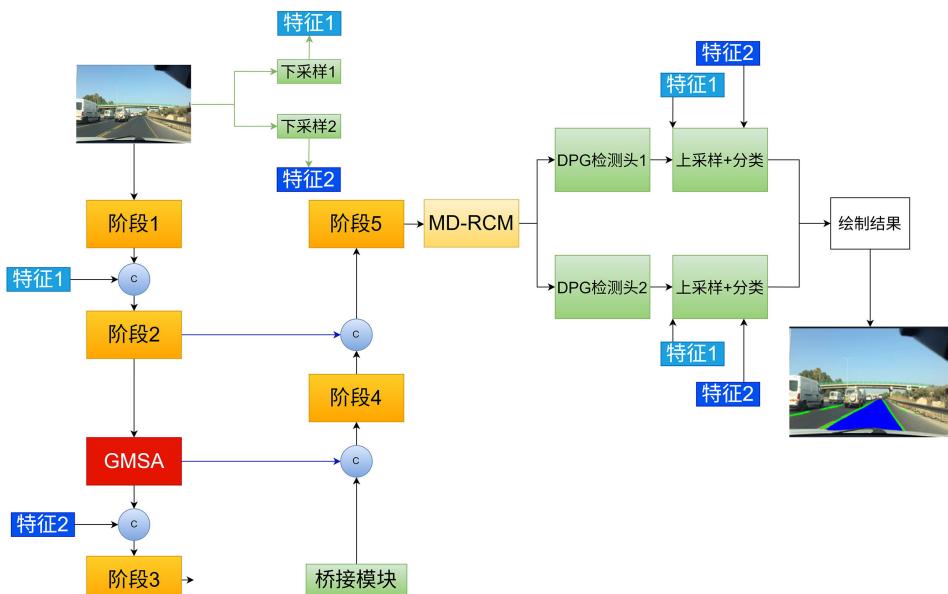
多任务方法近年来被广泛研究，用于集成不同的任务，提升感知系统的效率。YOLOP[12]、YOLOPv2[13]、A-YOLOM[14]和IALaneNet[15]等方法在该方向取得了初步成果。YOLOP系列模型将目标检测、

可行驶区域检测和车道线检测同时进行,提升了系统的整体性; A-YOLOM 和 IALaneNet 引入自适应权重与注意力机制,并且实现了任务之间的协调。然而这些多任务方法仍普遍存在模型较大,效率较低等问题,通常难以满足车载平台对高实时性与低计算资源消耗的实际需求。

本文提出一种轻量化且具有较高计算效率的多任务语义分割网络架构 GMSANet,在保有一定检测精度的同时,尽量降低模型参数与计算成本,更能适应自动驾驶系统对快速实时处理与低资源消耗的要求。网络采用编码-解码结构,通过统一编码器和并行的双分支解码器同时进行可行驶区域检测和车道线检测任务。在共享特征的基础上,为各任务设计了相对独立的解码分支,以减弱任务间的解码干扰并提升整体感知性能。

## 2. 模型整体结构

GMSANet 模型采用编码-增强-解码的结构,整体由共享编码器、MD-RCM 增强模块以及分别用于可行驶区域检测与车道线检测的两个解码器分支组成。模型的整体结果如图 1 所示。



**Figure 1.** Overall architecture of GMSANet model  
**图 1.** GMSANet 模型的整体结构

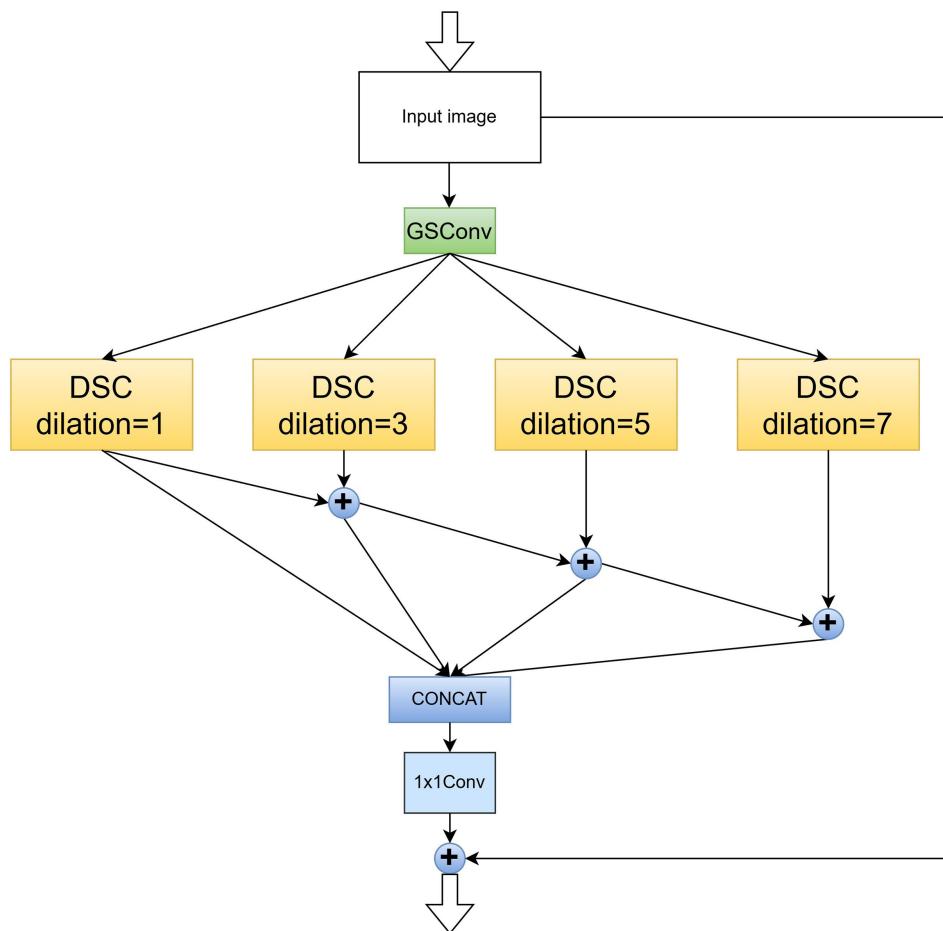
输入图像首先会通过不同尺度的简单下采样得到两组用于后续跳跃连接的中间特征,分别记为特征 1 和特征 2。模型将输入图像输入到编码器中,以提取多层次的语义特征。编码器内部设置了多个由 GSConv-ESP 模块构成的特征提取阶段模块,并引入分组多尺度注意力模块 GMSA,这个模块采用了分组条带卷积提高模型对具有方向性的线性结构的响应强度,并且会将特征图划分为多个通道子组,分别施加不同尺度的卷积操作,它能有效增强模型对关键区域的响应强度,提升空间感知能力,增强模型在复杂场景下的稳定性与鲁棒性。编码器将得到图像的主干特征以及中间特征 1 和中间特征 2。

主干特征将被输入至多尺度动态矩形自校准模块 MD-RCM,这个模块通过结合不同的上下文信息,利用多尺度感受野对特征进行提取,校准模型对关键区域的关注范围。经过校准的特征分别送入两个任务分支。每个解码器包含独立的特征解耦模块和上采样结构,逐步恢复空间分辨率,并结合中间特征 1 和中间特征 2 完成各自的像素级预测输出。整体结构在实现高效解耦的同时,保持特征共享的优势,兼顾精度、速度与部署效率,适用于多任务自动驾驶场景中的道路感知需求。

### 3. 模块介绍

#### 3.1. GSConv-ESP 模块与编码器

本研究所设计的编码器模块主要由五个阶段模块(Stage1 至 Stage5)构成, 每个阶段分别对不同语义层级的信息进行建模与融合。受 ESPNetV2 [16] 的启发, 本文设计 GSESP 模块, 它的整体结构遵循 ESPNetV2 的“降维 - 分支 - 变换 - 融合”设计思想。在保持 ESP 模块多尺度特征提取能力的基础上, 用 GSConv [17] 与逐深度可分离卷积 [18] (Depthwise Separable Convolution, DSC) 两种高效结构代替了普通的空洞卷积, 以实现特征表达能力与模型复杂度之间的平衡。整体的结构图如图 2 所示。



**Figure 2.** Architecture of the GSConv-ESP module  
**图 2.** GSConv-ESP 模块的结构

为在计算效率与表达能力之间取得平衡, 本文在降维阶段引入 GSConv。GSConv 首先通过  $1 \times 1$  卷积提取主要特征  $F_1$ , 实现降维与信息压缩; 然后采用  $5 \times 5$  深度卷积生成辅助特征  $F_2$ , 以补充局部纹理信息; 最后通过通道混洗(Shuffle)实现两者的均匀融合:

$$F_{\text{out}} = \text{Shuffle}(\text{Concat}(F_1, F_2)) \quad (1)$$

“降维 - 深度卷积 - 通道混合”的组合结构在不显著增加计算量与参数量的前提下, 保留了通道之间的联系, 增强了特征的多样性。不过, 若在多尺度分支部分中继续使用 GSConv, 会导致网络参数量和

推理延迟大幅上升。因此, GSESP 模块仅在降维阶段采用 GSConv 压缩高维输入并提升特征提取能力。

在多尺度分支部分, ESPNet 使用的是传统空洞卷积, 通过在卷积核中调节空洞率来改变感受野, 捕获多尺度上下文信息。当输入通道数与输出通道数较大时, 这种方法的计算量与参数开销随之显著增加。在多分支结构中, 高耦合卷积操作还容易引发栅格效应, 使得特征响应不连续、上下文信息缺失。

在 GSESP 模块的多尺度分支中采用 DSC。该结构由逐深度卷积与逐点卷积组成, 逐深度卷积结合不同空洞率( $d=1,3,5,7$ )提取多尺度空间特征, 逐点卷积在通道维度完成信息融合。在相同输入输出条件下, 单分支参数量由 864 降至 123, FLOPs 从 212.3 M 降至 30.2 M。逐点卷积带来的通道融合机制有效缓解了空洞卷积的栅格效应, 增强了多分支间的特征一致性与上下文连贯性。

网络的编码器由不同的 stage 组成, 每个 stage 内部包含若干个 GSESP 模块。通过下采样与通道增强, 逐步提升语义表达能力。多尺度建模与注意力机制更进一步强化了特征的感知, 为后续分支任务提供的特征基础。编码器整体架构如图 3 所示。

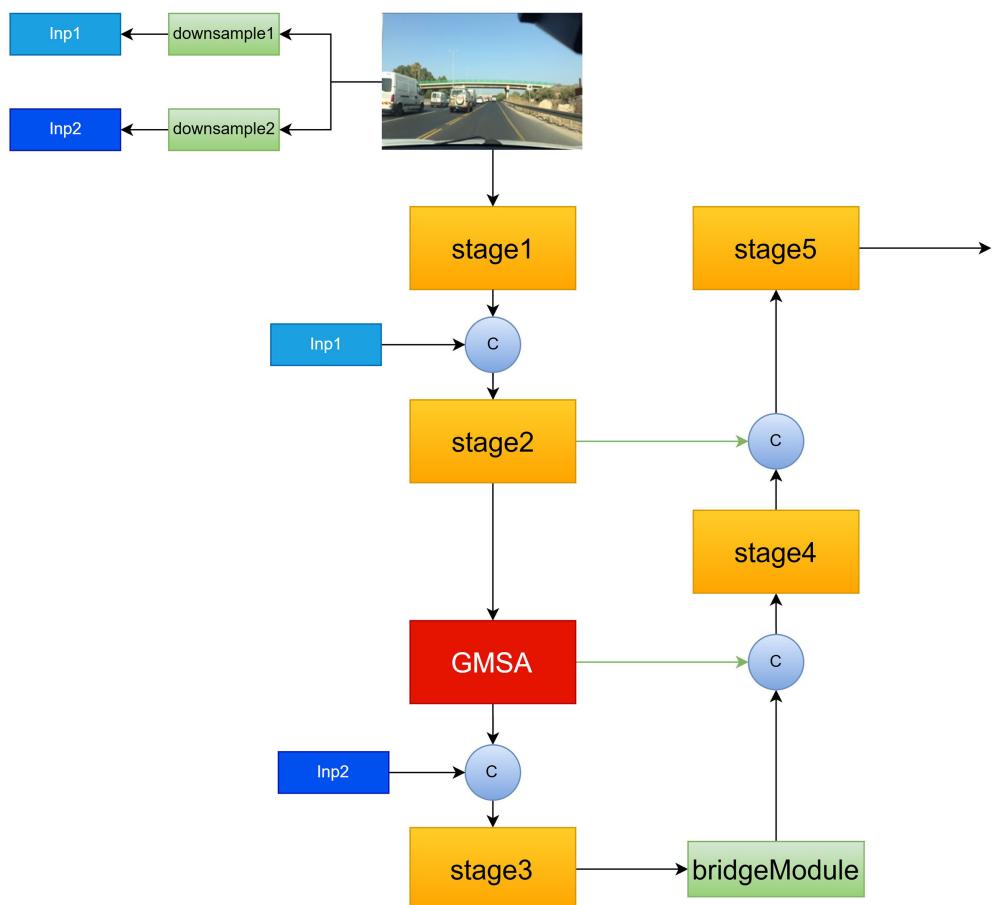


Figure 3. Encoder architecture

图 3. 编码器结构

编码器通过多阶段的下采样特征提取, 建立了从浅层到深层的多层次特征表示。每一个阶段逐步提升通道与语义的表达能力。在卷积堆叠之外, 编码器还引入多尺度卷积和注意力机制, 实现特征的融合与强化。网络也采用了跨阶段连接的结构, 用于缓解多层下采样带来的信息损失, 使特征传递更加连续与稳定。

### 3.2. GMSA 模块

在道路感知任务中, 可行驶区域分割与车道线分割并非相互独立的任务, 而是有一定的内在关联。可行驶区域为车道线提供了明确的范围与约束, 能够减少背景对车道线检测的干扰; 车道线作为道路结构的边界, 有助于提升可行驶区域的定位精度。相比于单一任务模型, 多任务模型可以在共享特征中同时刻画可行驶区域与车道线二者的相关语义信息和结构, 实现两者协同。

为提升模型对不同尺度与形状车道线及可行驶区域的感知能力, 提高模型对关键区域的响应强度, 本文设计了一种分组多尺度注意力模块(Grouped Multi-Scale Attention, GMSA), 加入在编码器中。该模块受到 CBAM [19]与 CPCA [20]的启发, 整体结构由三部分构成: 通道注意力模块、分组条带卷积模块和空间注意力模块。不同于 CBAM 仅在通道与空间维度建模注意力, GMSA 在其框架上引入“通道分组 + 多尺度 + 方向卷积”机制, 使网络能在同一模块内同时建模不同方向与尺度的结构信息, 从而在较低计算开销下获得更丰富的上下文表达。整体结构如图 4 所示。

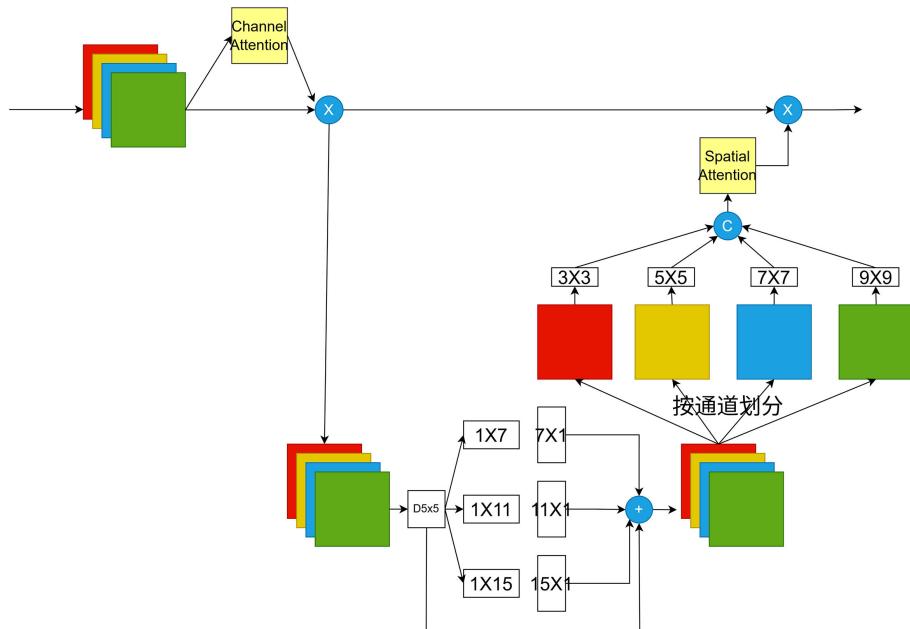


Figure 4. Architecture of the GMSA module

图 4. GMSA 模块的结构

通道注意力模块用于衡量并增强不同通道在特征表达中的重要性。对于输入特征图  $F \in R^{C \times H \times W}$ , 该模块通过全局平均池化和最大池化生成两个通道描述向量。随后由共享参数的两层  $1 \times 1$  卷积网络提取通道间依赖关系。在卷积过程中依次引入 ReLU 和 Sigmoid 激活函数, 用以提升非线性建模能力并生成通道注意力权重图。随后将权重图与输入特征图逐通道相乘, 得到加权重构后的特征。

$$f_{\text{avg}} = \text{GAP}(F), f_{\text{max}} = \text{GMP}(F) \quad (2)$$

$$M_c = \sigma(W_2 \cdot \delta(W_1 \cdot f_{\text{avg}}) + W_2 \cdot \delta(W_1 \cdot f_{\text{max}})) \quad (3)$$

$$F_{\text{out}} = F \otimes M_c \quad (4)$$

该机制使模型能够聚焦于关键通道特征, 从而提升对车道线与可行驶区域的识别能力。

车道线在图像中具有明显的方向性, 并且伴有尺度变化。为增强模型对多尺度与方向性空间结构的

感知能力，增强后的特征输入至多个分组条带卷积模块，分别在横纵方向上设置不同尺度的卷积核，以同时捕获细节特征与全局信息，从而增强模型对方向性特征的敏感性。该模块受 Inception [21] 和 PKINet [22] 多分支设计思想的启发，结合深度可分离卷积与条带卷积的优势，通过多分支结构提取不同感受野下的空间上下文信息。输入特征  $F \in R^{C \times H \times W}$  首先经  $5 \times 5$  Depthwise 卷积获得基础特征  $F_0$ 。

$$F_0 = \text{DWConv}_{5 \times 5}(F) \quad (5)$$

把  $F_0$  并行送入三组串联的  $1 \times i$  与  $i \times 1$  条带卷积分支 ( $i \in \{7, 11, 15\}$ )，以捕获方向上的空间特征。各分支输出与基础分支逐元素加和得到融合特征  $F_s$ 。

$$F_1 = \text{DWConv}_{7 \times 1}(\text{DWConv}_{1 \times 7}(F_0)) \quad (6)$$

$$F_2 = \text{DWConv}_{11 \times 1}(\text{DWConv}_{1 \times 11}(F_0)) \quad (7)$$

$$F_3 = \text{DWConv}_{15 \times 1}(\text{DWConv}_{1 \times 15}(F_0)) \quad (8)$$

$$F_s = F_0 + F_1 + F_2 + F_3 \quad (9)$$

将  $F_s$  按通道划分为四个子特征组： $\{F_s^1, F_s^2, F_s^3, F_s^4\}$ ，并分别通过不同尺度的 Depthwise Separable 卷积 ( $1 \times 1$ 、 $3 \times 3$ 、 $5 \times 5$ 、 $7 \times 7$ ) 进行局部上下文建模，以进一步提取不同感受野下的空间信息。分组策略一方面能够降低卷积计算量，另一方面在不同特征子空间内实现方向性特征的独立建模，避免通道间特征干扰。模块在保持轻量化的同时进一步提升了特征的多样性与表达能力。

$$F_m^1 = \text{DSConv}_{1 \times 1}(F_s^1) \quad (10)$$

$$F_m^2 = \text{DSConv}_{3 \times 3}(F_s^2) \quad (11)$$

$$F_m^3 = \text{DSConv}_{5 \times 5}(F_s^3) \quad (12)$$

$$F_m^4 = \text{DSConv}_{7 \times 7}(F_s^4) \quad (13)$$

经多尺度卷积后，各子特征沿通道维度进行拼接，并进行归一化与非线性激活处理，得到融合特征  $F_m$ 。

$$F_m = \text{Concat}(F_m^1, F_m^2, F_m^3, F_m^4) \quad (14)$$

这样的设计在有效控制参数量的同时，提升了模型对道路场景中具有强方向性的结构的建模能力，并且对尺度的变化能够有较好的鲁棒性。

模块在结构末端引入轻量级空间注意力，继续增强对关键区域的响应能力。该机制在先分别通道维度上施加平均池化与最大池化，把结果在通道维拼接并由  $7 \times 7$  卷积和 Sigmoid 激活生成空间注意力图，输出为加权融合得到的结果。

$$F_s = \text{Concat}(\text{AvgPool}(F_m), \text{MaxPool}(F_m)) \quad (15)$$

$$M_s = \sigma(\text{Conv}_{7 \times 7}(F_s)) \quad (16)$$

$$F_{\text{out}} = F_m \otimes M_s \quad (17)$$

空间注意力强化了车道边缘等关键区域的响应强度，提升了特征表达的能力。

为整合通道与空间两个方面的注意力信息，模块将通道注意力增强后的特征图与空间注意力生成的权重图进行逐元素相乘，得到最终的输出特征图。这样的融合策略兼顾了通道与空间两个方面，提升模型对关键区域结构的响应强度。

$$F_{\text{final}} = F_c \otimes F_s \quad (18)$$

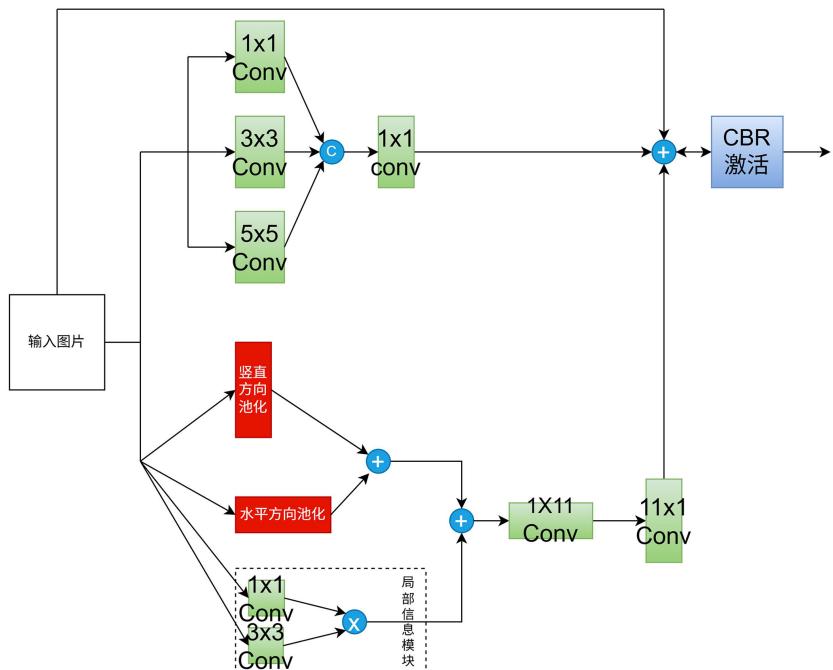
作为编码器中的核心模块之一, GMSA 在特征提取阶段通过多尺度条带卷积与注意力融合机制, 实现了方向性与尺度信息的建模强化。有助于增强模型对关键区域的响应强度。

### 3.3. MD-RCM 模块

复杂道路场景中, 轻量化分割模型的预测区域由于缺乏有效校准, 往往不够稳定。校准的难点在于, 近处道路需要模型通过较大的感受野获取整体形状信息, 远处边界与其他较小的结构则需要小范围的感受野, 以保证定位精度。若采用固定尺度的校准方式, 通常难以同时兼顾全局一致性与局部细节。

本文受 RCM [23] 模块的启发, 改进并引入了多尺度动态矩形自校准模块(Multi-scale Dynamic Rectangular Self-Calibration Module, MD-RCM)以帮助模型校准预测区域的范围。RCM 模块能够在固定尺度大小范围内实现空间自校准, 但其提前固定的卷积尺度降低了模型在复杂场景中的适应能力。MD-RCM 在 RCM 的基础上引入多尺度卷积机制, 通过并行分支实现局部与全局空间特征的同时提取, 提升模型对不同尺度目标的感受和校准。

模块结构如图 5 所示, 由两个并行分支组成: 第一个分支采用多尺度卷积以提取不同感受野下的局部特征, 从而更好地提取空间结构信息, 第二个分支通过水平方向与垂直方向的轴向池化获取全局上下文信息, 再结合动态卷积与条带卷积生成空间注意力图。



**Figure 5.** Architecture of the MD-RCM module  
**图 5.** MD-RCM 模块的结构

第一个分支中, 采用三种不同尺度的深度卷积核( $1 \times 1$ 、 $3 \times 3$ 、 $5 \times 5$ )对输入特征进行并行卷积操作, 以提取多感受野下的局部特征信息, 捕捉不同空间尺度上的特征, 增强对局部变化的感知能力。设输入特征为  $X \in \mathbb{R}^{C \times H \times W}$ , 得到三组卷积后, 为实现多尺度信息的融合, 将三组特征沿通道维度拼接。

$$X_{\text{cat}} = \text{Concat}(f_{1 \times 1}(X), f_{3 \times 3}(X), f_{5 \times 5}(X)) \quad (19)$$

融合后的特征图通过一个  $1 \times 1$  卷积进行维度压缩, 使其通道数与原始输入一致。

$$X_{\text{MS}} = f_{1 \times 1}^{\text{last}}(X_{\text{cat}}) \quad (20)$$

其中  $f_{1 \times 1}^{\text{last}}(\cdot)$  表示  $1 \times 1$  卷积的通道压缩操作, 得到融合后的多尺度局部特征图。多尺度结构使模型能够在同一个模块中同时提取小尺度细节特征和大尺度整体语义信息, 与固定尺度卷积相比, 能更好地适应不同大小和形状的道路目标, 从而提升特征表达的完整性。

第二个分支中, 分别对输入特征图  $X \in \mathbb{R}^{C \times H \times W}$  在水平方向和垂直方向施加自适应平均池化, 生成两个轴向特征图, 通过逐元素相加融合为轴向全局的上下文特征图, 提取具有方向性的全局上下文用于补充空间长依赖信息。

$$X_g = \text{AvgPool}_h(X) + \text{AvgPool}_v(X) \quad (21)$$

引入了动态卷积机制增强对局部变化的响应能力, 通过  $1 \times 1$  卷积和  $3 \times 3$  卷积提取局部特征, 完成对局部信息的动态调整。

$$W_{\text{dyn}} = f_{1 \times 1}(X) \quad (22)$$

$$X_{\text{local}} = f_{3 \times 3}^{\text{dw}}(X) \odot W_{\text{dyn}} \quad (23)$$

其中,  $\odot$  表示逐元素乘法操作,  $f_{3 \times 3}^{\text{dw}}(\cdot)$  表示深度可分离卷积。

模块将轴向池化提取的全局上下文特征  $X_g$  与动态卷积生成的局部增强特征  $X_{\text{local}}$  进行逐元素相加, 形成融合后的特征图:

$$X_f = X_g + X_{\text{local}} \quad (24)$$

为了进一步突出关键区域的响应, 该机制采用两个方向性卷积(横向  $1 \times 11$  与纵向  $11 \times 1$ ), 模拟图像中常见的矩形区域感知方式, 从而提取更符合前景结构的注意力权重图。

$$A = \sigma\left(f_{1 \times 1}\left(\delta\left(\text{BN}\left(f_{1 \times 11}(X_f)\right)\right)\right)\right) \quad (25)$$

其中,  $\text{BN}(\cdot)$  表示批归一化操作,  $\delta(\cdot)$  为 ReLU 激活函数,  $\sigma(\cdot)$  为 Sigmoid 函数。

最终, 将两个分支的输出做逐元素相乘, 完成特征的自校准与加权增强, 高响应区域得到强化, 低响应区域受到抑制, 关键区域响应更突出, 空间结构线索更清晰。

$$\text{Output} = A \cdot X_{\text{MS}} + X \quad (26)$$

相较于 RCM 仅依赖单尺度卷积结构, MD-RCM 在保留其自校准特性的基础上, 新增多尺度卷积路径, 并与动态卷积、方向性卷积协同作用, 从而显著提升了模型对复杂场景中目标尺寸变化的适应能力。

## 4. 实验与分析

本研究中选用 BDD100K [24] 数据集作为模型训练与评估的主要基准数据集。该数据集根据用途被划分为训练集(70,000 张)、验证集(10,000 张)和测试集(20,000 张)。由于测试集的标签尚未公开, 本研究在验证集上对模型性能进行评估。针对可行驶区域检测任务, 采用平均交并比(mean Intersection over Union, mIoU)作为主要指标。针对车道线检测任务, 采用车道线精度(Lane Accuracy, LA)和车道线交并比(Lane IoU)两项指标进行评估。

本实验在 Ubuntu 22.04 系统下进行, 使用 Miniconda 创建运行环境, Python 版本为 3.10.4。深度学习框架为 PyTorch 2.4.1, CUDA 版本为 12.1, cuDNN 版本为 9.1 (90100)。训练设备为 4 张 NVIDIA RTX 4090D (24 GB) GPU, 配备 60 核心的 Intel Xeon Platinum 8474C CPU 及 320 GB 内存, 支持多卡分布式

训练。

#### 4.1. 数据集结果

进行可行驶区检测时, 将 BDD100K 数据集中标注为“area/drivable”与“area/alternative”的区域统一视为“可行驶区域”进行处理, 因此模型只需区分图像中的可行驶区域与背景部分。采用 mIoU 作为主要评估指标, 用于衡量不同模型对于可行驶区域的检测性能。在进行车道线检测时, 评估指标包括 Lane Accuracy (像素级准确率)和 Lane IoU (交并比), 以全面衡量模型的检测能力, 结果如表 1 所示。

**Table 1.** Results of different models on the BDD100K dataset

**表 1.** 不同模型在 BDD100K 数据集的结果

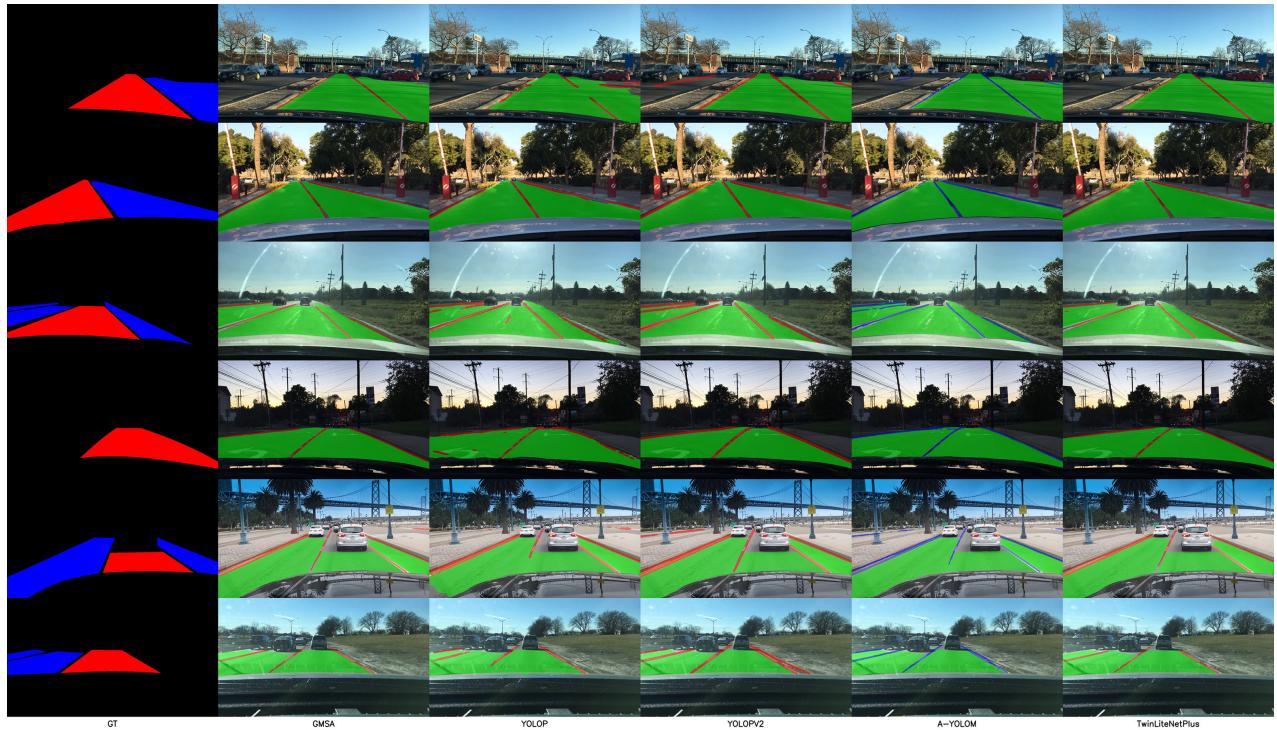
MODEL	Drivable Area mIoU (%)	Lane Accuracy (%)	Lane IoU (%)	FLOPS (G)	Param (M)
YOLOP [12]	91.6	70.5	26.5	8.11	5.53
YOLOPv2 [13]	93.2	87.31	27.25	70.9	38.9
A-YOLOM (s) [14]	91.0	84.9	28.8	19.47	13.61
IALaneNet (ConvNeXt-tiny) [15]	91.29	-	31.48	96.52	18.35
TwinLiteNetPlus [25]	92.9	81.9	34.2	17.58	1.94
DeepLabV3+ [26]	90.9	-	29.8	30.7	15.4
SegFormer [27]	92.3	-	31.7	12.1	7.2
PDPMamba [28]	92.8	89.2	21.6	-	-
U-MobileViT [29]	92.4	79.7	32.5	3.47	2.25
GDMNet [30]	92.2	75.3	26.4	-	-
GMSANET (ours)	92.8	85.1	34.0	6.45	2.9

在可行驶区域检测方面, 所提出的 GMSANet 取得了 92.8% 的 mIoU, 整体性能优于 YOLOP、SegFormer 和 DeepLabV3+ 等主流方法, 仅略低于 YOLOPv2。相较于精度较高但计算开销较大的模型, GMSANet 在参数量仅为 2.9 M、FLOPs 为 6.45 G 的条件下仍能保持较优性能, 体现了良好的轻量化设计与推理效率优势。在车道线检测任务中, 模型在 Lane IoU 指标上取得 34.0%, 优于 YOLOP、YOLOv8 (multi)、SegFormer 等方法, 表明模型在车道线边界定位与结构连续性建模方面具有较强能力。在 Lane Accuracy 指标上, 模型达到 85.1%, 整体表现处于较优水平, 与 A-YOLOM (s) 接近, 虽略低于 YOLOPv2, 但明显优于 YOLOP 和 TwinLiteNetPlus。综合来看, GMSANet 在可行驶区域与车道线两项任务上均取得了较为均衡的性能表现, 在保持较低参数规模和计算复杂度的同时, 实现了对区域型目标与结构型目标的有效感知, 验证了该模型在实际驾驶场景中的应用潜力。

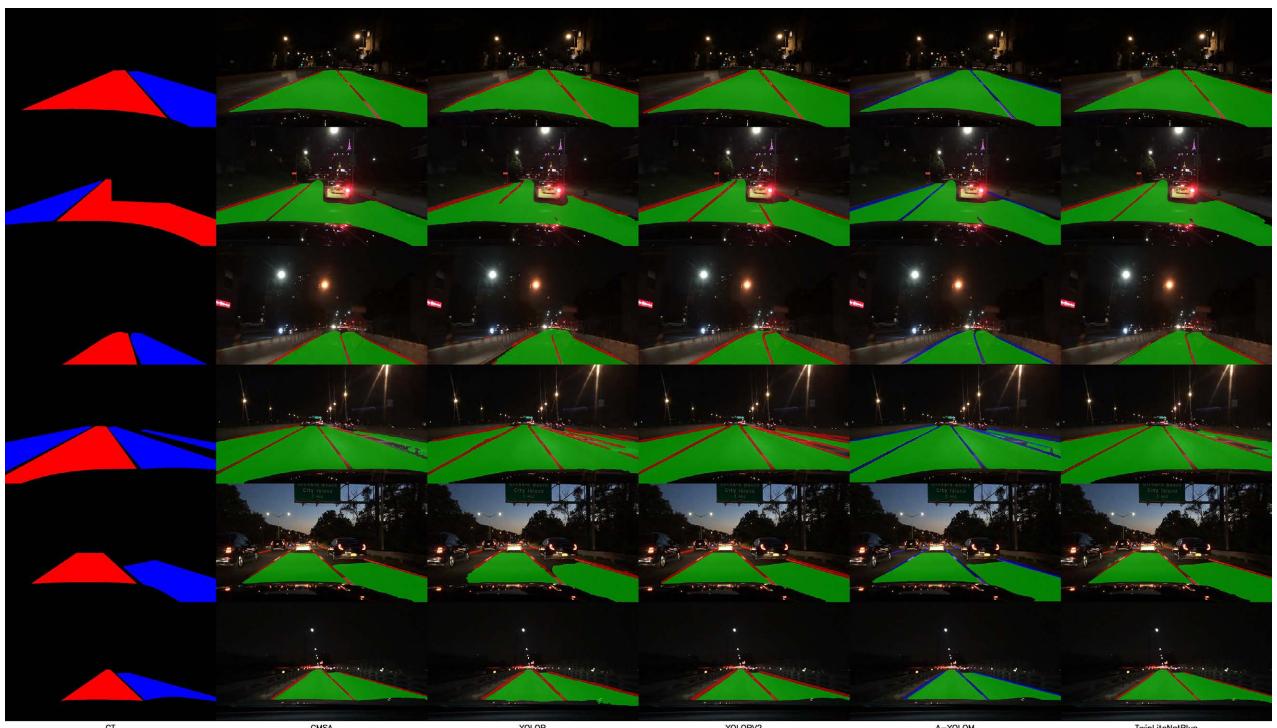
#### 4.2. 可视化结果

本文将所提出的 GMSA 模型与多种代表性方法的预测结果进行了对比分析, 以评估模型在实际道路场景下的表现。图 6 展示了晴天场景下各模型的结果, 从左至右依次为真值(GT)、GMSA、YOLOP、YOLOPv2、A-YOLOM 以及 TwinLiteNetPlus。在晴天条件下, GMSA 模型在车道线与可行驶区域分割中表现更为准确, 预测结果边界清晰、连续性较好。相比之下, YOLOP 与 A-YOLOM 在部分区域易出现车道线断裂或偏移; YOLOPv2 虽具备较好的车道线连续性, 但在复杂场景中仍存在重复或虚假车道线预测; TwinLiteNetPlus 能较好保持整体结构, 但在边界区域存在一定模糊和局部不连续现象。总体而言,

GMSA 模型在晴天场景下展现出更稳定的预测结果。



**Figure 6.** Visual comparison of results on a sunny day of different models  
**图 6.** 不同模型晴天场景结果的对比



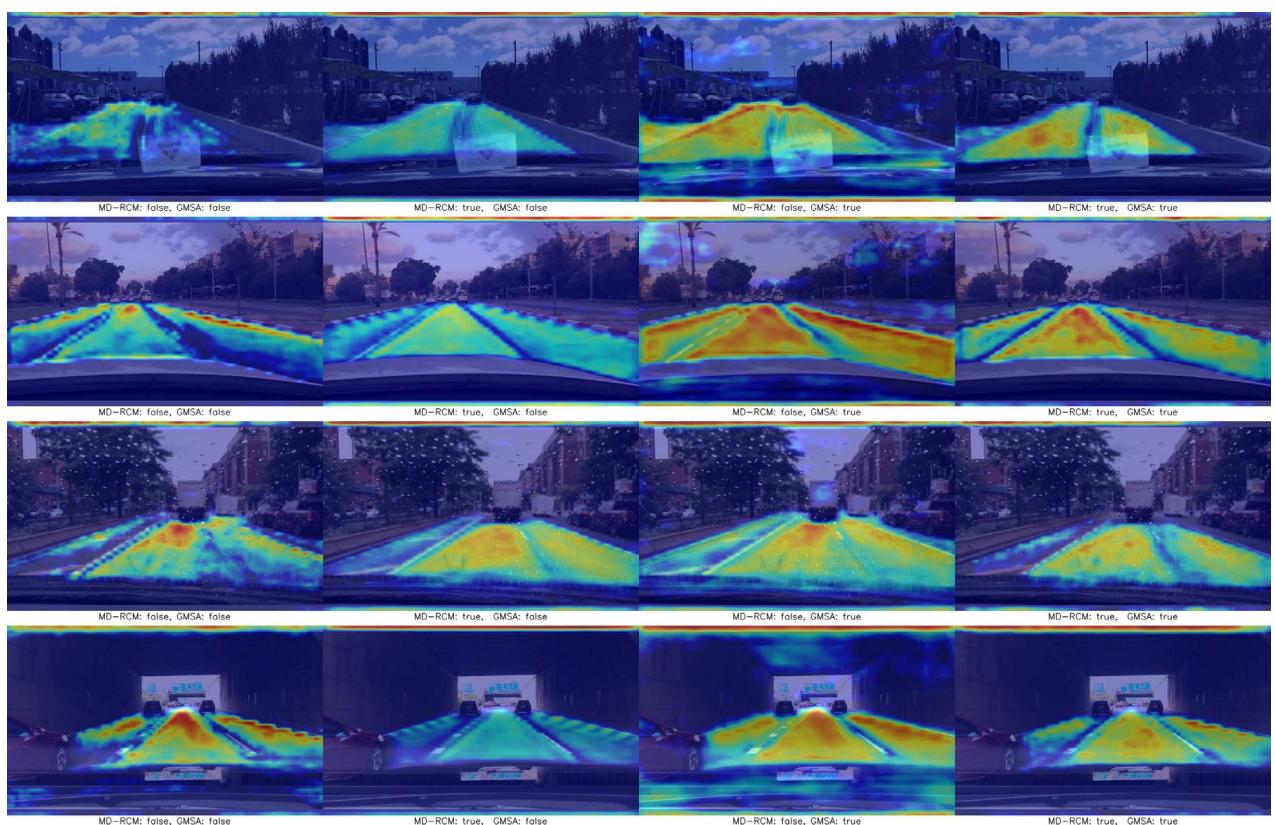
**Figure 7.** Visual comparison of results at night of different models  
**图 7.** 不同模型夜间场景的结果对比

夜间场景下的分割结果如图 7 所示, 从左至右依次为真值(GT)、GMSA、YOLOP、YOLOPv2、A-YOLOM 以及 TwinLiteNetPlus。受光照不足等干扰影响, 图像对比度与可用纹理信息明显降低, 对模型的鲁棒性提出了更高要求。在该场景中, YOLOP 表现不稳定, 车道线易出现断裂或模糊; YOLOPv2 在整体效果上较优, 但仍存在重复检测和虚假响应的问题; A-YOLOM 与 TwinLiteNetPlus 虽能保持车道线的基本完整性, 但在局部区域易出现模糊或粘连现象。相比之下, GMSA 模型在低光照条件下仍能较为准确地分割车道线与可行驶区域, 预测结果清晰, 表现出更好的夜间适应性。

### 4.3. 热力图

本研究采用 Grad-CAM [31]方法对模型在可行驶区域检测与车道线检测两个子任务中的激活特征图进行可视化处理, 生成代表模型对不同区域关注程度的热力图, 分析不同结构模块对模型感知能力的影响。分析了不同的模型结构, 包括“无模块”“仅使用 MD-RCM”“仅使用 GMSA”“两者同时使用”四种, 在不同场景下对热力图进行定性分析, 以评估各模块对模型效果的影响。

模型在可行驶区域任务中的热力图结果如图 8 所示。在图中可以看到, MD-RCM 模块使得模型对目标区域边界进行了校准, 使热力图边界更加清晰, 预测结果更贴合实际道路结构; GMSA 模块则提升了模板区域的响应强度, 使道路中心更加突出。在二者皆有的完整结构下, 模型在目标聚焦能力和边界连续性方面均有明显提升, 整体热力图质量优于不使用模块的基线模型。

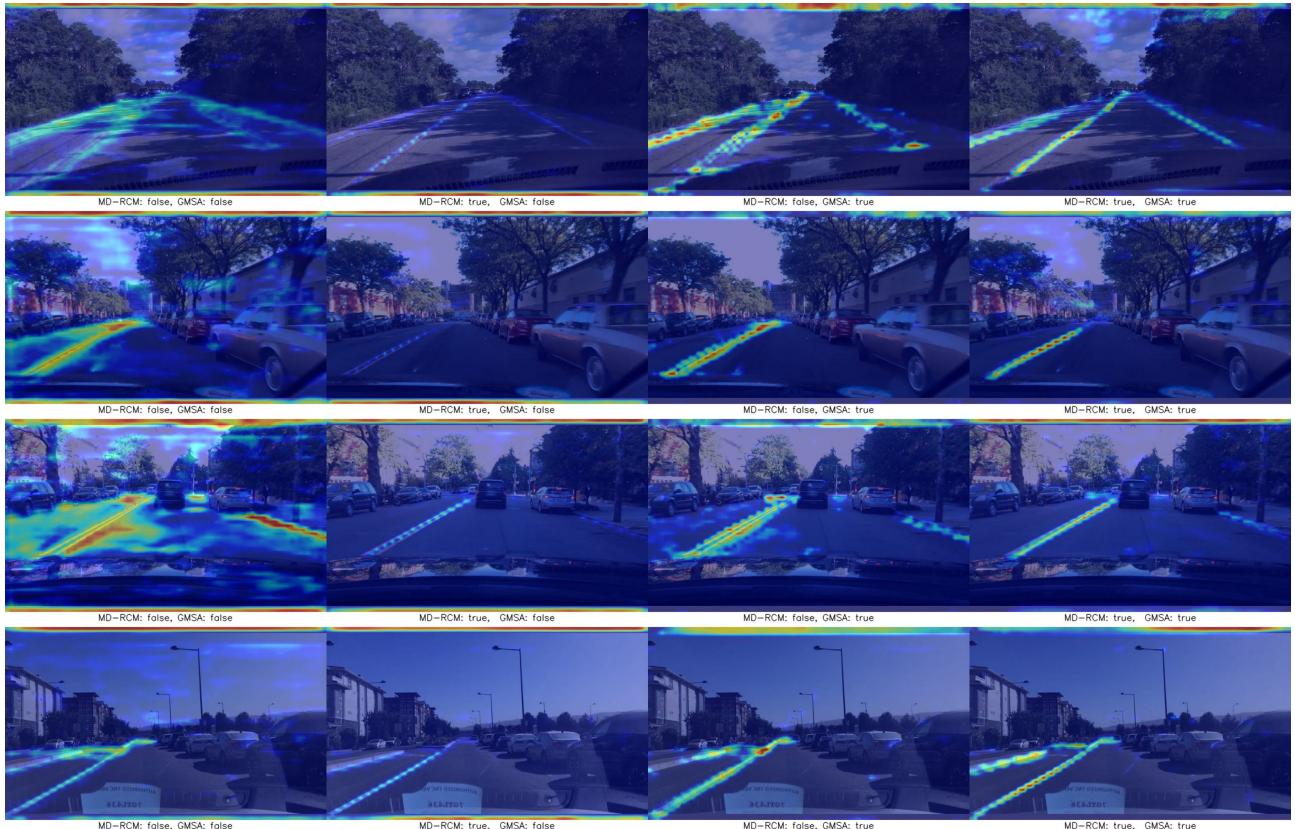


**Figure 8.** Heatmap visualization for drivable area detection

**图 8. 可行驶区域检测的热力图结果**

模型在车道线检测任务中热力图结果如图 9 所示。MD-RCM 模块使得模型在车道线边界、间断及遮挡区域表现更为稳定, 有效增强了对车道线位置的捕捉能力; GMSA 模块强化了模型对细长条状目标的

响应,使目标区域更加明显。同时使用的情况下,模型在低光照和雨天等复杂场景中仍能保持较好的车道线连贯性,而基线模型则更容易出现响应断裂或模糊的问题。



**Figure 9.** Heatmap visualization for lane line detection

图 9. 车道线检测的热力图结果

GMSA 模块侧重于增强目标区域的响应强度,MD-RCM 模块更关注区域校准,两者具有良好的互补性。二者联合使用能够有效提升模型对关键区域的感知能力与预测稳定性。

#### 4.4. 模块消融实验

**Table 2.** Module ablation experiment results

表 2. 模块消融实验的结果

MODEL	Drivable Area mIoU (%)	Lane Accuracy (%)	Lane IoU (%)
YOLOP	91.6	70.5	26.5
A-YOLOM (s)	91.0	84.9	28.8
基础模型	90.7	83.3	31.2
基础模型 + MD-RCM	92.2	84.3	32.5
基础模型 + GMSA	91.9	84.6	33.1
基础模型 + MD-RCM + GMSA	92.8	85.1	34.0

为分析各模块对模型性能的影响,本文以原始模型为基础,分别引入 MD-RCM 模块、GMSA 模块

以及两者的组合进行对比实验，同时与 YOLOP 和 A-YOLOM (s)方法进行对比。实验结果如表 2 所示。从表中可以看出，加入 MD-RCM 模块后，车道线检测的 ACC 从 83.3% 提升至 84.3%，IoU 从 31.2% 提升至 32.5%。这表明该模块能够在一定程度上改善模型检测的范围，提升精度。加入 GMSA 模块后，ACC 和 IoU 分别提升至 84.6% 和 33.1%，说明该模块可以强化模型的响应，提升检测精度。当同时引入 MD-RCM 和 GMSA 模块时，模型性能进一步提升，车道线检测 ACC 达到 85.1%，IoU 达到 34.0%，可行驶区域 mIoU 也提高至 92.8%。结果表明，两模块在全局与局部特征建模方面具有互补效果，共同提升了模型在车道线与可行驶区域检测任务上的整体性能。

#### 4.5. 多任务与单任务的对比

为验证可行驶区域分割与车道线检测两项任务之间的关联是否能够通过联合学习得到有效利用，本文在相同编码器网络与相同训练设置下，对比了单任务模型与多任务模型，对比的结果如表 3 所示。单任务模型分别仅针对可行驶区域或车道线任务训练；多任务模型采用共享编码器与双解码器结构，同时进行两个任务。在可行驶区域分割任务上，单任务模型的 mIoU 为 92.7%，多任务模型为 92.8%，两者基本一致；在车道线检测任务上，单任务模型的 Accuracy 与 IoU 分别为 85.4% 和 34.2%，多任务模型分别为 85.1% 和 34.0%，相较单任务仅有轻微下降，整体差异较小。多任务模型的更适合部署，且共享编码器使两项任务在同一特征表示上学习可行驶区域语义与车道线结构信息，能够复用道路场景的特征，减少重复建模带来的冗余。相比分别训练与部署两个单任务模型，多任务模型可在一次前向推理中同时输出两类结果，模型结构更紧凑，工程集成与维护成本更低。

**Table 3.** Comparison results between single-task and multi-task models  
**表 3.** 单任务模型与多任务模型的对比结果

MODEL	Drivable Area mIoU (%)	Lane Accuracy (%)	Lane IoU (%)
可行驶区域单任务	92.7	-	-
车道线检测单任务	-	85.4	34.2
多任务模型	92.8	85.1	34.0
可行驶区域单任务	92.7	-	-

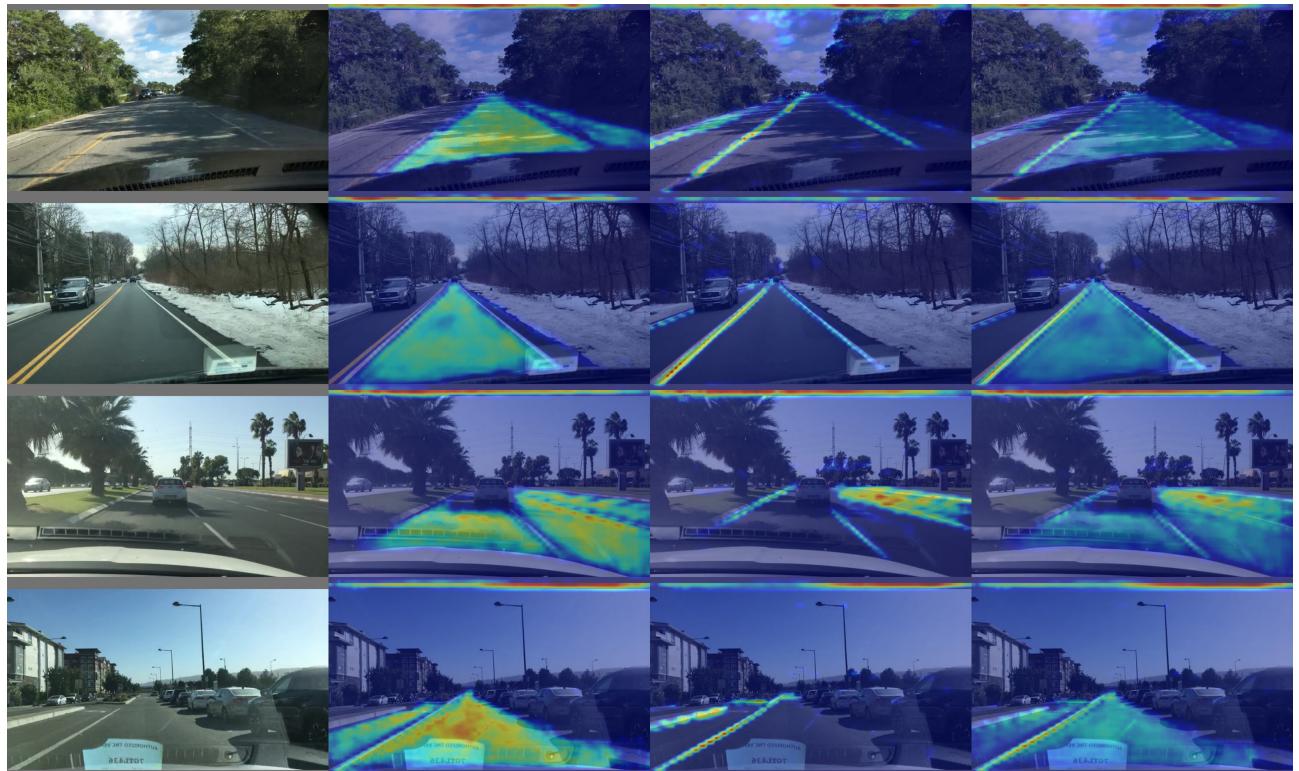
本文使用 Grad-CAM 方法创建了两个单独任务网络与多任务网络的对比热力图，用于分析单任务模型与多任务模型的关注区域。图 10 展示了在相同输入下，单任务与多任务模型的热力图对比。可行驶区域单任务主要关注路面，但易受到背景干扰，车道线检测单任务在车道线区域响应较强，但有时会错误地响应路面纹理和遮挡区域。多任务模型的可行驶区域分支能够抑制背景区域的误激活，车道线检测分支则能在车道线处形成更连续的响应，减少对背景区域的误关注。多任务学习通过共享特征空间增强了两任务间的协同效应，提升了模型在复杂环境中的鲁棒性。

本文设计了不同的损失函数权重，以探讨多任务学习中不同任务对结果的影响。损失函数可以写为：

$$L = w_{da} L_{da} + w_{ll} L_{ll} \quad (27)$$

其中， $L_{da}$  与  $L_{ll}$  分别表示可行驶区域任务与车道线分割任务的损失项， $w_{da}$  与  $w_{ll}$  为对应的权重系数。通过调整权重比  $w_{da} : w_{ll}$ ，多任务模型在相同训练设置下的性能对比结果如表 4 所示。权重的调整会改变任务在共享参数更新中的贡献比例，引发性能的改变，增大某一任务的权重时，相关的指标通常更占优势，另一任务的指标会出现轻微下降。在共享主干上统计 mean\_cos 与 neg\_ratio，其中 mean\_cos 数值越大说

明任务之间优化方向越相似, neg\_ratio 用于反映梯度相互“对抗”的发生频率, 数值越大说明任务之间的冲突越多。多任务学习在训练阶段的大部分时间能够保持较高的一致性, 并且冲突较少, 共享主干上更倾向于推动相近的特征表达, 这说明可行驶区域与车道线的信息能够形成互补, 实现特征复用并受益。但训练阶段中仍有可能出现 mean\_cos 降低且 neg\_ratio 上升的情况, 这代表不同任务之间有轻微的特征冲突, 协同与竞争并存。



**Figure 10.** Comparison of heatmaps between the multi-task model and the single-task model

**图 10. 多任务与单任务模型的热力图对比**

**Table 4.** Comparison of models with different task weights  
**表 4. 不同权重的模型对比**

权重 可行驶区域: 车道线检测	Drivable Area mIoU (%)	Lane Accuracy (%)	Lane IoU (%)
1:2	92.4	85.7	34.2
1:1	92.8	85.1	34.0
2:1	93.0	83.9	32.8

## 5. 总结与展望

本文针对自动驾驶中可行驶区域与车道线检测任务之间关联性不足以及模型结构复杂的问题, 提出了一种基于增强 - 校准策略的轻量化多任务语义分割网络 GMSANet。该模型基于 GSConv-ESP 结构, 并结合“增强 - 几何”的设计理念, 实现了多任务特征的高效表达。网络中引入的分组多尺度注意力(GMSA)模块用于模型对关键区域的响应强度, 而多尺度动态矩形自校准(MD-RCM)模块则增强了模型对关键区

域的自适应校准能力。实验结果验证了所提方法的有效性与可行性。GMSANet 在保证模型轻量化与实时性的同时, 实现了对可行驶区域和车道线的同时检测, 并且有较好的精度。研究结果表明, 该方法在复杂道路环境中表现出较好的鲁棒性与泛化能力, 能够为资源受限条件下的智能驾驶感知系统提供一种高效且可靠的解决方案。

## 基金项目

本文受上海市科委 2024 创新行动计划项目资助, 项目编号: 24511103302。

本文受中央高校基本科研业务费专项资金资助。

## 参考文献

- [1] Chiu, K.Y. and Lin, S.F. (2005) Lane Detection Using Color-Based Segmentation. *Proceedings of the IEEE Intelligent Vehicles Symposium*, Las Vegas, 6-8 June 2005, 706-711. <https://doi.org/10.1109/ivs.2005.1505186>
- [2] Satzoda, R.K., Sathyanarayana, S., Srikanthan, T. and Sathyanarayana, S. (2010) Hierarchical Additive Hough Transform for Lane Detection. *IEEE Embedded Systems Letters*, **2**, 23-26. <https://doi.org/10.1109/les.2010.2051412>
- [3] Zhang, Y., Gong, P., Ji, S. and Xu, Q. (2022) Real-Time Lane Detection Method Based on Region of Interest. 2022 *International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, Hengyang, 26-27 March 2022, 1188-1192. <https://doi.org/10.1109/icitbs55627.2022.00255>
- [4] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/cvpr.2016.90>
- [5] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 234-241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [6] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2481-2495. <https://doi.org/10.1109/tpami.2016.2644615>
- [7] Zhao, H., Shi, J., Qi, X., Wang, X. and Jia, J. (2017) Pyramid Scene Parsing Network. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 6230-6239. <https://doi.org/10.1109/cvpr.2017.660>
- [8] 高程阳, 郁湧, 秦江龙. 面向交通场景的图像分割网络[J]. 计算机科学与应用, 2024, 14(4): 13-23.
- [9] Parashar, A., Rhu, M., Mukkara, A., Puglielli, A., Venkatesan, R., Khailany, B., et al. (2017) SCNN: An Accelerator for Compressed-Sparse Convolutional Neural Networks. *Proceedings of the 44th Annual International Symposium on Computer Architecture*, Toronto, 24-28 June 2017, 27-40. <https://doi.org/10.1145/3079856.3080254>
- [10] Honda, H. and Uchida, Y. (2024) ClrNet: Improving Confidence of Lane Detection with Laneiou. 2024 *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 3-8 January 2024, 1165-1174. <https://doi.org/10.1109/wacv57701.2024.00121>
- [11] Wang, L. and Zhong, H. (2024) FENet: Focusing Enhanced Network for Lane Detection. 2024 *IEEE International Conference on Multimedia and Expo (ICME)*, Niagara Falls, 15-19 July 2024, 1-6. <https://doi.org/10.1109/icme57554.2024.10687857>
- [12] Wu, D., Liao, M., Zhang, W., Wang, X., Bai, X., Cheng, W., et al. (2022) YOLOP: You Only Look Once for Panoptic Driving Perception. *Machine Intelligence Research*, **19**, 550-562. <https://doi.org/10.1007/s11633-022-1339-y>
- [13] Han, C., Zhao, Q., Zhang, S., Chen, Y., Zhang, Z. and Yuan, J. (2022) YOLOPv2: Better, Faster, Stronger for Panoptic Driving Perception. <https://arxiv.org/abs/2208.11434>
- [14] Wang, J., Jonathan Wu, Q.M. and Zhang, N. (2024) You Only Look at Once for Real-Time and Generic Multi-Task. *IEEE Transactions on Vehicular Technology*, **73**, 12625-12637. <https://doi.org/10.1109/tvt.2024.3394350>
- [15] Tian, W., Yu, X. and Hu, H. (2023) Interactive Attention Learning on Detection of Lane and Lane Marking on the Road by Monocular Camera Image. *Sensors*, **23**, Article 6545. <https://doi.org/10.3390/s23146545>
- [16] Mehta, S., Rastegari, M., Shapiro, L. and Hajishirzi, H. (2019) ESPNetv2: A Light-Weight, Power Efficient, and General Purpose Convolutional Neural Network. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 9182-9192. <https://doi.org/10.1109/cvpr.2019.00941>

- [17] Li, H., Li, J., Wei, H., Liu, Z., Zhan, Z. and Ren, Q. (2022) Slim-Neck by GSConv: A Lightweight-Design for Real-Time Detector Architectures. <https://arxiv.org/abs/2206.02424>
- [18] Chollet, F. (2017) Xception: Deep Learning with Depthwise Separable Convolutions. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 1800-1807. <https://doi.org/10.1109/cvpr.2017.195>
- [19] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y. Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 3-19. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)
- [20] Hou, Q., Zhou, D. and Feng, J. (2021) Coordinate Attention for Efficient Mobile Network Design. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 13708-13717. <https://doi.org/10.1109/cvpr46437.2021.01350>
- [21] Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A. (2017) Inception-v4, Inception-Resnet and the Impact of Residual Connections on Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, **31**, 4278-4284. <https://doi.org/10.1609/aaai.v31i1.11231>
- [22] Cai, X., Lai, Q., Wang, Y., Wang, W., Sun, Z. and Yao, Y. (2024) Poly Kernel Inception Network for Remote Sensing Detection. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 27706-27716. <https://doi.org/10.1109/cvpr52733.2024.02617>
- [23] Ni, Z., Chen, X., Zhai, Y., Tang, Y. and Wang, Y. (2024) Context-Guided Spatial Feature Reconstruction for Efficient Semantic Segmentation. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T. and Varol, G., Eds., *Lecture Notes in Computer Science*, Springer, 239-255. [https://doi.org/10.1007/978-3-031-72943-0\\_14](https://doi.org/10.1007/978-3-031-72943-0_14)
- [24] Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., et al. (2020) BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 2633-2642. <https://doi.org/10.1109/cvpr42600.2020.00271>
- [25] Che, Q., Le, D., Pham, M., Nguyen, V. and Lam, D. (2025) TwinLiteNet+: An Enhanced Multi-Task Segmentation Model for Autonomous Driving. *Computers and Electrical Engineering*, **128**, Article 110694. <https://doi.org/10.1016/j.compeleceng.2025.110694>
- [26] Chen, L., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L. (2018) Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 834-848. <https://doi.org/10.1109/tpami.2017.2699184>
- [27] Xie, E., Wang, W., Yu, Z., et al. (2021) SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *Advances in Neural Information Processing Systems*, **34**, 12077-12090.
- [28] Li, Z., Bian, J., Sun, M., Zhao, X., Li, W. and Zhang, L. (2025) PDPMamba: Potential Panoptic Driving Perception via Multi-Task Visual Mamba. 2025 *25th International Conference on Digital Signal Processing (DSP)*, Pylos (Messinia, Southwest Pelo-Ponnes), 25-27 June 2025, 1-5. <https://doi.org/10.1109/dsp65409.2025.11075142>
- [29] Nguyen, P., Nguyen, T., Pham, P. and Bui, Q. (2026) U-MobileViT: A Lightweight Vision Transformer-Based Backbone for Panoptic Driving Segmentation. *Signal Processing: Image Communication*, **142**, Article 117461. <https://doi.org/10.1016/j.image.2025.117461>
- [30] Liu, Y., Ma, H., Zhu, J. and Zhang, Q. (2024) GDMNet: A Unified Multi-Task Network for Panoptic Driving Perception. *Computers, Materials & Continua*, **80**, 2963-2978. <https://doi.org/10.32604/cmc.2024.053710>
- [31] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2017) Grad-Cam: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 618-626. <https://doi.org/10.1109/iccv.2017.74>