

面向复杂交通场景的改进DeepLabV3+语义分割算法研究

程伟贤

应急管理大学计算机科学与工程学院，北京

收稿日期：2025年12月29日；录用日期：2026年1月26日；发布日期：2026年2月5日

摘要

针对现有语义分割算法在处理多目标图像时难度大、精度低的问题，改进DeepLabV3+网络提出一种面向复杂交通场景的语义分割算法。该算法改用轻量级骨干特征提取网络，提高计算速度；重构密集空洞空间金字塔模块并引入卷积注意力机制，加强高级特征信息提取；采用多尺度特征融合策略，提高特征恢复精度。在Cityscapes数据集上的验证结果表明，该算法能在保证时效性的同时，准确实现车辆、行人等交通目标的特征提取，为复杂交通环境下的实时分割任务提供可行方案。

关键词

DeepLabV3+，交通目标，语义分割，特征融合

Research on the Improved DeepLabV3+ Semantic Segmentation Algorithm for Complex Traffic Scenarios

Weixian Cheng

School of Computer Science and Engineering, University of Emergency Management, Beijing

Received: December 29, 2025; accepted: January 26, 2026; published: February 5, 2026

Abstract

Addressing the challenges of difficulty and low accuracy in handling multi-objective images with existing semantic segmentation algorithms, an improved DeepLabV3+ network is proposed for a semantic segmentation algorithm tailored to complex traffic scenarios. This algorithm adopts a lightweight backbone feature extraction network to enhance computational speed; reconstructs the

dense dilated spatial pyramid module and introduces a convolutional attention mechanism to strengthen the extraction of high-level feature information; and employs a multi-scale feature fusion strategy to improve feature restoration accuracy. The validation results on the Cityscapes dataset demonstrate that this algorithm can accurately extract features of traffic objects such as vehicles and pedestrians while ensuring timeliness, providing a feasible solution for real-time segmentation tasks in complex traffic environments.

Keywords

DeepLabV3+, Traffic Target, Semantic Segmentation, Feature Fusion

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

精准定位与捕获目标是智能驾驶和交通管理系统中的关键，而以像素级解析为核心的语义分割研究，能对多种交通目标实现精准识别与分割。Chen 等[1]-[4]提出的 DeepLab 系列网络，从 V1 的首次引入空洞卷积，到 V2 以不同扩张率的空洞卷积构造空洞空间金字塔池化模块，再到 V3 的批量归一化处理并加入全局平均池化，最终集成编解码结构的 V3+，凭其优异的性能在语义分割领域中广泛应用。

众多学者在运用 DeepLabV3+ 处理交通场景时进行了多方面的改进。邵玉文等[5]将骨干网络替换为轻量化网络 MobileNetV3，显著提高了推理速度，使其更加适用于移动端设备；郭江等[6]针对城市建筑物分割任务优化了解码器结构，增强了边缘细节的保持能力；李阳等[7]在训练阶段引入多支 DBB 模块，并在推理时转换为单分支结构，既提升了精度也保证了效率；闫河等[8]提出一种融合注意力的改进语义分割网络，在 VOC2012 数据集上实现了较高的交并比；郑红彬[9]针对 DeepLabV3+ 的单目标语义分割进行了改进，提出的算法对多目标城市街景图像具有较好的分割效果；朱俊涛等[10]在编码与解码时分别引入卷积注意力与 Transformer 多头注意力，主要针对非结构化道路的可行驶区域进行分割，为智能车辆自主决策提供理论依据。

上述方法在不同层面均提升了语义分割的性能，但在复杂的交通场景中仍存在目标尺度差异大、遮挡频繁、边缘细节易丢失等问题，车辆、行人等关键目标的分割精度仍有待提升。为此，本文提出了一种面向复杂交通场景的改进 DeepLabV3+ 语义分割算法，选择 MobileNetV2 轻量级网络用于骨干特征提取，优化 ASPP 结构并引入卷积注意力机制，加强高级特征的语义表达能力，采用多尺度融合策略重构解码结构，以此恢复更多的边界特征信息。在公开数据集 Cityscapes 上的实验表明，该算法在保持较低参数量的同时，能够有效提升交通场景中目标的识别精度与边缘质量。

2. 研究方法

2.1. 传统 DeepLabV3+ 网络

传统 DeepLabV3+ 网络的编码器采用 Xception 骨干网络提取特征，空洞空间金字塔池化模块(Atrous Spatial Pyramid Pooling, ASPP)处理深度卷积神经网络(Deep Convolutional Neural Network, DCNN)输出的高级特征，捕获其上下文语义信息；解码器通过对编码器输出的低级与高级特征进行上采样拼接操作，得到与输入图像尺寸相同的特征图，网络整体架构见图 1。经实验证，DeepLabV3+ 网络在 VOC、

Cityscapes 等场景数据集中整体表现出良好性能，但其在复杂交通场景中仍面临计算复杂度高、目标感知有限以及细节特征利用不足等问题。

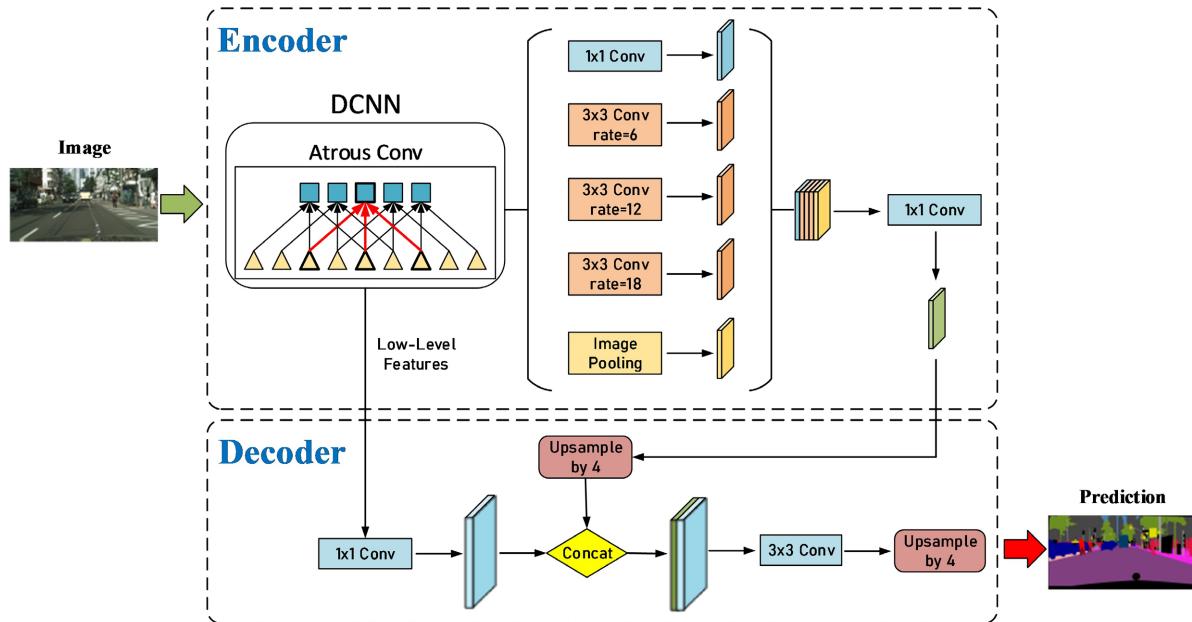


Figure 1. Structure of DeepLabV3+ network

图 1. DeepLabV3+网络结构

2.2. 改进 DeepLabV3+ 网络

针对上述传统 DeepLabV3+ 网络的不足，本文从网络轻量化、加强特征提取和多尺度特征融合三个角度进行改进，使网络能够更适用于复杂交通场景语义分割任务，改进后的网络架构见图 2。选用轻量级网络 MobileNetV2 替换原 Xception 网络作为骨干特征提取网络，将标准卷积分解为深度卷积和逐点卷积，大幅降低计算复杂度。

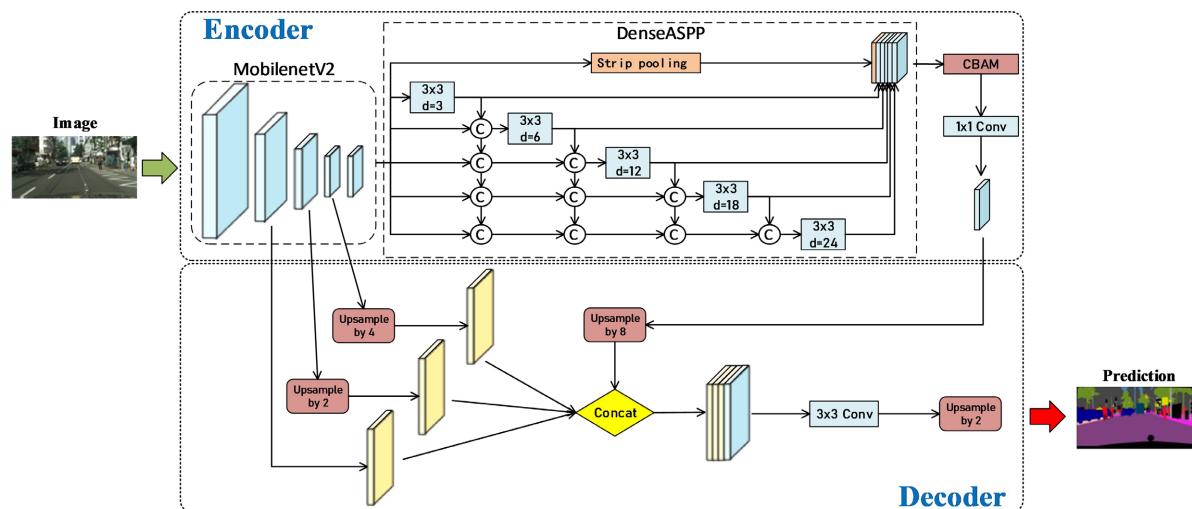


Figure 2. Structure of the improved DeepLabV3+ network

图 2. 改进 DeepLabV3+ 网络架构

在面对包含众多目标的交通场景时，ASPP 模块所覆盖的感受野面积不足，故添加扩张率为 3 和 24 的空洞卷积，采用密集连接构建拥有更大感受野的密集空洞空间金字塔池化模块(DenseASPP)，复用特征来增强语义信息的传递与融合。针对其输出，添加卷积注意力模块(Convolutional Block Attention Module, CBAM)，依次实施通道注意力(Channel Attention)和空间注意力(Spatial Attention)，选择性融合特征图的细节信息和空间结构：通道注意力通过全局平均池化和全连接层生成通道权重，突出重要特征通道；空间注意力则通过最大池化和平均池化聚合空间信息，并通过卷积层生成空间权重图，强化关键区域特征。

解码器部分，以 1/2 尺寸特征图为融合标准，分别对 1/4 尺寸特征图和 1/8 特征图进行 2 倍与 4 倍上采样，同时将尺寸分别为 1/2、1/4 和 1/8 大小的低级特征图与经 CBAM 加权及 8 倍上采样后的高级特征图进行 Concat 特征融合，弥补 DeepLabV3+ 网络特征利用的不足，有效提升对目标和边界区域的关注度与恢复精度。

3. 实验验证

3.1. 实验准备

本研究选用泛化性较好的 Cityscapes 数据集对改进模型进行训练与评估。Cityscapes 数据集专注于城市街景的语义理解，内有来自于 50 个不同城市的交通场景图像。数据集包含 5000 张精细标注图像和 20,000 张粗标注图像，涵盖 30 个视觉类别，选择其中 19 个类用于测试评估。图像分辨率均为 2048×1024 ，内容包含多种光照条件、天气场景和交通密度，非常适用于复杂交通场景语义分割研究。数据集划分为 2975 张训练集、500 张验证集和 1525 张测试集。在训练前，对训练数据需进行预处理与增强操作，以此提升模型的泛化能力，包括随机缩放、水平翻转及标准化等，最终调整至统一尺寸输入网络。

实验的硬件配置如下：13th Gen Intel Core i5-13600KF 处理器、32 G 运行内存、Nvidia GeForce RTX 4060 Ti (8 G) 图形处理器(GPU)。优化器选择带动量的随机梯度下降(SGD)，损失函数选用交叉熵与 Dice 损失的加权组合，以缓解类别不平衡问题。模型训练是在深度学习 Pytorch 框架下进行，其他辅助软件包括 Pycharm 和 Anaconda。参数设置见表 1。

Table 1. Parameter setting of the network

表 1. 网络参数设置

参数	设置
Image size	1024×512
Epoch	500
Batch size	4
Optimizer	Sgd
Learning Rate	7×10^{-3}
Decay Type	cos
Loss Function	Cross Entropy Loss + Dice Loss

3.2. 评价指标

在使用深度学习模型处理图像时，一般为了客观地评估模型的性能，都要根据每个像素点的分类结果绘制混淆矩阵。本文语义分割研究主要将像素点划分为 19 种不同的类，其混淆矩阵如表 2 所示。真正(True Positive, TP)指预测像素结果为正确的情况，即预测为正，实际也为正；真负(True Negative, TN)指

预测像素结果为正确的情况，即预测为负，实际也为负；假正(False Positive, FP)指将其他像素误判为目标像素的情况，即预测为正，实际却为负；假负(False Negative, FN)指将目标像素误判为其他像素的情况，即预测为负，实际却为正。

Table 2. Confusion matrix**表 2.** 混淆矩阵

真实情况	预测结果	
	正	负
正	TP	FN
负	FP	TN

根据混淆矩阵，选择召回率(Recall)、平均交并比(mean Intersection over Union, mIoU)、平均像素精度(mean Pixel Accuracy, mPA)和像素准确率(Accuracy)这四种具有代表性的评价指标来评估模型的性能。Recall 用来衡量模型对正样本的判别能力，交并比(IoU)指预测结果与真实标签之间交集与并集的比值，像素精度(PA)指预测正确的像素数占真实像素数的比例，mIoU、mPA 取它们的平均；Accuracy 则表示预测结果正确的像素数占所有像素数的比例，用来衡量模型整体的分类正确率。计算公式如下：

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (1)$$

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^N \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}} + N_{\text{FP}}}, \quad (2)$$

$$\text{mPA} = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{j=1}^N n_{ij}}{\sum_{i=1}^N \sum_{j=1}^N n_{ij}}, \quad (3)$$

$$\text{Accuracy} = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{TN}} + N_{\text{FN}} + N_{\text{FP}}}, \quad (4)$$

式中， N_{TP} 为真正样本的个数； N_{TN} 为真负样本的个数； N_{FP} 为假正样本的个数； N_{FN} 为假负样本的个数； n_{ij} 是混淆矩阵第 i 行第 j 列的值； N 为像素所分类别数。

4. 结果分析

4.1. 消融实验

为验证交通场景分割模型改进方法的有效性，实验分析了添加每项改进措施后模型在 Cityscapes 数据集上所得各项性能指标(见表 3)。模型 1 为传统 DeepLabV3+ 网络，模型 2 替换了骨干特征提取网络，模型参数量从 54.709M 骤降至 5.813M，但各项指标均有小幅度下降。模型 2 引入 DenseASPP 结构得到模型 3，mIoU 等指标上升显著，验证了扩大感受野策略的有效性。模型 4 在 DenseASPP 后引入 CBAM 注意力机制，空间与通道双向提取高级特征信息。模型 5 融合了多个尺度的特征图像，解决了低级特征利用率低下的问题，使模型能够从低级特征图中提取出更多、更精细的边界信息，提高了预测精度。实验结果表明，改进 DeepLabV3+ 网络在测试集上性能表现有所提升，与模型 1 相比，mIoU 提高 2.29%，为 68.75%，Recall 提高 1.65%，为 76.04%，mPA 提高 1.40%，为 78.63%，Accuracy 提高 1.19%，同时模型的参数量整体下降 72.19%。

Table 3. Evaluation of segmentation performance of different improvements**表 3. 不同改进的分割性能评估**

模型	骨干网络	DenseASPP	CBAM	多尺度特征融合	mIoU/ (%)	Recall/ (%)	mPA/ (%)	Accuracy/ (%)	参数量/M
1	Xception				66.46	74.39	77.23	95.15	54.709
2	MobileNetV2				63.86	71.63	76.46	94.66	5.813
3	MobileNetV2	√			65.73	74.07	76.93	95.12	15.127
4	MobileNetV2	√	√		67.02	74.58	77.91	95.60	15.168
5	MobileNetV2	√	√	√	68.75	76.04	78.63	96.34	15.213

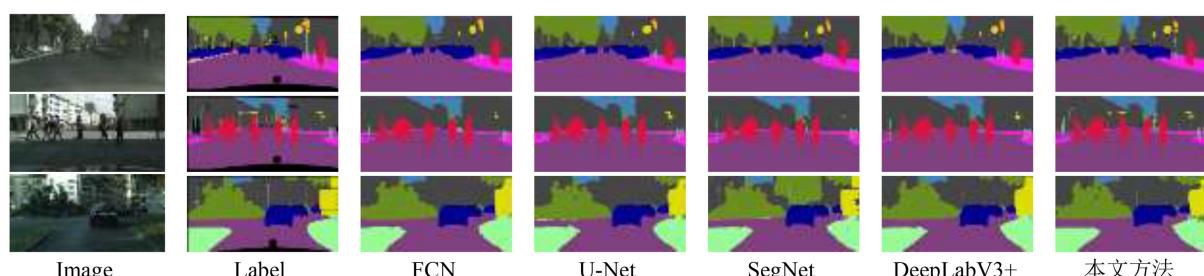
4.2. 对比实验

为了验证改进 DeepLabV3+ 网络的优越性，以 FCN、U-Net、SegNet、DeepLabV3+ 四种语义分割模型为参照做了对比实验，训练参数与改进模型保持一致，几种模型在测试集上的评价指标结果见表 4。据表中可知，改进 DeepLabV3+ 网络的各项性能指标均优于其他模型。

Table 4. Evaluation of segmentation performance of different models**表 4. 不同模型的分割性能评估**

模型	骨干网络	mIoU/ (%)	Recall/ (%)	mPA/ (%)	Accuracy/ (%)
FCN	VGG16	64.54	73.39	74.52	90.14
U-Net	VGG16	65.90	73.83	76.96	94.85
SegNet	VGG16	61.96	70.07	72.49	87.78
DeepLabV3+	Xception	66.46	74.39	77.23	95.15
改进 DeepLabV3+	MobileNetV2	68.75	76.04	78.63	96.34

实验对比结果见图 3，为能更直观地展现路况，将拍摄车辆的前脸部分归于背景。FCN 与 U-Net 在交通场景的分割效果相近，SegNet 效果最差，DeepLabV3+ 在目标边界的精细度逊于 U-Net，但在整体的识别上较好。本文方法的分割效果明显优于其他模型，不仅降低了误检率，还保留了更多的边界特征，结果与标签图像更加接近，表现出更高的精度与更强的抗干扰能力，能够更好地在交通场景中分割出目标特征。

**Figure 3.** Segmentation results of different models in traffic scenes**图 3. 不同模型在交通场景中的分割结果**

5. 结论

本研究基于 DeepLabV3+ 网络提出一种面向交通场景的改进语义分割算法。该算法改换 MobileNetV2

轻量级网络提取特征，有效减少模型参数量；扩充 ASPP 构建 DenseASPP 结构，并引入 CBAM 注意力，捕获更多的高级语义信息；融合了编码器各阶段多个尺度的特征图，恢复更多的特征信息。实验结果表明，改进的 DeepLabV3+ 网络分割精度更高、计算速度更快、抗干扰性更强，在交通场景中的分割结果更接近真实标签，为交通监管、智能驾驶等工作提供可行性方案。

参考文献

- [1] Chen, L.C., Papandreou, G., Kokkinos, I., et al. (2014) Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. <https://doi.org/10.48550/arXiv.1412.7062>
- [2] Chen, L.C., 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>
- [3] Chen, L.C., Papandreou, G., Schroff, F., et al. (2017) Rethinking Atrous Convolution for Semantic Image Segmentation. <https://doi.org/10.48550/arXiv.1706.05587>
- [4] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: *Lecture Notes in Computer Science*, Springer, 833-851. https://doi.org/10.1007/978-3-030-01234-2_49
- [5] 邵玉文. 基于轻量化 DeepLabV3+的街景语义分割算法研究与优化[D]: [硕士学位论文]. 兰州: 西北师范大学, 2025.
- [6] 郭江. 基于 DeepLabV3+的遥感建筑物提取与变化检测[D]: [硕士学位论文]. 西宁: 青海师范大学, 2024.
- [7] 李阳, 李猛, 王中华. 基于改进 DeepLabv3+的室外交通场景识别[J]. 交通科技与管理, 2023, 4(4): 1-3.
- [8] 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.
- [9] 郑红彬. 基于深度学习的城市街景语义分割算法研究[D]: [硕士学位论文]. 西安: 西安工业大学, 2024.
- [10] 朱俊涛, 刘佳琦, 杨璐. 面向非结构化道路的可行驶区域语义分割[J]. 天津理工大学学报, 2025, 41(2): 105-112.