

陈雨佳, 宁媛, 唐坤俊, 等. 改进 DeeplabV3plus 的道路场景分割算法[J]. 智能计算机与应用, 2025, 15(11): 87–93. DOI: 10.20169/j. issn. 2095-2163. 251114

改进 DeeplabV3plus 的道路场景分割算法

陈雨佳¹, 宁媛¹, 唐坤俊¹, 刘聂天和²

(1 贵州大学 电气工程学院, 贵阳 550025; 2 贵州电网有限责任公司 贵阳花溪供电局, 贵阳 550025)

摘要: 道路场景分割是自动驾驶感知系统的一项基本任务。为解决 DeeplabV3plus 模型在道路场景分割中精度不够, 特别是针对小目标分割精度不够的问题, 本文提出一种融合注意力机制的改进 DeeplabV3plus 模型, 该模型整体结构为非对称的编码器-解码器架构。编码器方面, 将 Xception 模型替换为 ResNet50 模型, 编码器中使用膨胀卷积增大感受野, 以提高特征提取能力; 解码器方面, 使用自行设计的 UCA 模块, 将深层次特征与浅层特征进行融合并且使用双线性插值进行上采样, 提高特征图边界轮廓清晰度的同时更容易捕获小目标, 同时在 UCA 模块中引入坐标注意力机制, 使得模型能在通道方向和空间位置方面聚焦, 捕捉长距离依赖, 提升模型对重要目标的关注度。在 Cityscape 数据集和自建数据集组成的数据集上的实验结果表明, 平均交并比(mIoU)指标提升了 2.6%, 像素精确度(PA)提升了 0.9%; 针对道路场景中目标较小, 数量较少的行人和非机动车的交并比(IoU)指标分别提高了 3.2% 和 4.3%, 可行驶区域的分割精度提高了 2%, 更加准确的获取了道路信息, 为自动驾驶感知系统提供了更多信息。

关键词: 道路场景分割; 注意力机制; DeeplabV3plus

中图分类号: TP391.41

文献标志码: A

文章编号: 2095-2163(2025)11-0087-07

Improved DeeplabV3plus road scene segmentation algorithm

CHEN Yujia¹, NING Yuan¹, TANG Kunjun¹, LIU Nietianhe²

(1 School of Electrical Engineering, Guizhou University, Guiyang 550025, China;

2 Guiyang Huaxi Power Supply Bureau, Guizhou Power Grid Co., Ltd., Guiyang 550025, China)

Abstract: Road scene segmentation is a basic task of autonomous driving perception system. In order to solve the problem that the accuracy of the DeeplabV3plus model is not enough in road scene segmentation, especially for the lack of accuracy of small target segmentation, an improved DeeplabV3plus model with attention mechanism is proposed, and the overall structure of the model is an asymmetric encoder-decoder architecture. In terms of encoders, the original encoder Xception is replaced with ResNet50, and the encoder uses dilated convolution to increase the receptive field to improve the feature extraction ability. The decoder is a self-designed Upsample-Concatenation-Attention module, which fuses deep features with shallow features and uses bilinear interpolation for upsampling, which improves the clarity of the boundary contour of the feature map and makes it easier to capture small targets. Increase the model's focus on important goals. Experimental results show that on the dataset composed of Cityscape dataset and self-built dataset, the mean Intersection over Union ratio index is increased by 2.6%, and the pixel accuracy is increased by 0.9%. The IoU index for pedestrians and non-motorized vehicles with small targets and small numbers in the road scene is increased by 3.2% and 4.3%, and the segmentation accuracy of the drivable area is increased by 2%, which more accurately obtains road information and provides more information for the autonomous driving perception system.

Key words: road scene segmentation; attention mechanism; DeeplabV3plus

0 引言

随着计算机视觉技术的发展, 自动驾驶感知技

术也取得了飞速进步。道路场景分割作为自动驾驶感知系统的基本任务之一, 其研究方法已从传统方法转向基于深度学习的道路场景分割方法, 得益于

基金项目: 贵州省科技计划基金(黔科合 ZK2022135)。

作者简介: 陈雨佳(1994—), 男, 硕士, 主要研究方向: 机器学习, 图像处理; 唐坤俊(1998—), 男, 硕士, 主要研究方向: 目标检测, 模型部署; 刘聂天和(1998—), 男, 硕士, 主要研究方向: 电网技术, 边缘计算。

通信作者: 宁媛(1968—), 女, 硕士, 教授, 主要研究方向: 计算机视觉, 图像处理。Email: ee.yning@gzu.edu.cn。

收稿日期: 2024-02-29

哈尔滨工业大学主办 ◆ 系统开发与应用

其优异的性能,基于深度学习的方法现已成为自动驾驶感知领域的主流技术^[1]。好的分割结果能为自动驾驶决策系统提供丰富的信息,便于决策系统做出控制决策以保证行车安全。因此,设计精准的道路场景分割方法对于自动驾驶系统具有重要的意义。

Shelhamer 等^[2] 基于全卷积神经网络 FCN (Fully Convolutional Network),将图像分割任务带到了一个全新的发展阶段。基于深度学习的语义分割方法相较于传统检测方法,不仅在精度上实现了显著提升,同时针对复杂环境具有更强的鲁棒性,但仍面临诸多挑战,例如网络感受野不足,上下文信息不充分等等导致的精度不足问题。解决方式一类是以 U-Net, SegNet 为代表的编码器-解码器架构模型,通过融合深层次语义信息与浅层次语义信息使得图像空间信息不至于完全损失,恢复目标上下文信息从而获得较好的分割效果。另一类解决方式是谷歌团队的 Chen 等^[3] 提出的 Deeplab 系列模型,该系列模型在 V1 版本中引入了膨胀卷积,在不额外增加参数量、不降低分辨率的前提下,极大的提高了卷积核的感受野;而在 V2 版本中引入了空洞空间金字塔池化 ASPP (Atrous Spatial Pyramid Pooling) 模块,该模块中并行的嵌入了不同膨胀率的膨胀卷积,使特征图具有不同尺度的信息,增强了网络对于不同尺度目标的预测能力^[4];而 DeeplabV3plus 版本在原模型基础上引入了一个简单的解码器结构,该模型能够融合部分浅层次语义特征,从而获得性能的提升,由于该模型优秀的性能,许多工作仍然围绕其展开^[5-6]。

DeeplabV3plus 模型虽然引入了一个简单的解码器结构融合特征,但该解码器只使用了一个浅层特征进行融合,因此还会造成部分信息丢失,对于部分小目标如行人、被遮挡车辆等,部分可行驶区域分割仍然不够精准,徐聰等^[7] 将部分浅层进行直接融合取得一定的效果,然而浅层特征使用的并不充分,只使用了 8 倍和 16 倍下采样的特征,且 16 倍的下采样倍率过高使得网络损失部分高频信息;若将浅层特征直接融合,张蕊^[8] 等提出这样的融合方式会带来噪声问题;何雪东等^[9] 在改进的 DeeplabV3plus 模型中加入了通道注意力机制取得了不错的效果。

注意力机制代表性的工作有基于通道注意力的 SE (Squeeze-and-Excitation) 注意力机制,基于空间通道注意力的 CBAM (Convolutional Block Attention Module) 注意力机制,高效通道注意力机制 ECA

(Efficient Channel Attention) 等等,这些注意力机制在实际应用中都取得了不错的效果^[10-12]。然而 SE, ECA 注意力机制无法捕捉空间位置信息,CBAM 无法捕捉长距离依赖,自注意力 (Self-Attention, S-A) 机制能够较好的解决长距离依赖问题,但会带来沉重的计算代价^[13]。Hou 等^[14] 提出了一种轻量级的坐标注意力 (Coordinate Attention, CA) 机制,在特征图宽高方向进行池化操作,使得网络能够捕捉长距离依赖关系,不仅使网络更加聚焦于感兴趣的区域,计算代价又小于自注意力机制。

本文将原 DeeplabV3plus 模型的解码器进行了重新设计,将特征提取阶段的所有浅层次特征进行融合,使得网络能够充分利用低层次语义信息,对小目标多尺度信息、图像边界等高频信息予以极大的保留,在还原图像时,结果更加精确,边界更加清晰;在进行深层次语义信息与浅层次语义信息融合时,加入轻量级的 CA 注意力机制,使网络更加聚焦于感兴趣区域,忽略噪声信息的影响。

1 DeeplabV3plus 模型

DeeplabV3plus 模型凭借其出色的分割能力,广泛的应用于各个分割任务中,例如桥梁裂缝检测,卫星遥感图像,滑坡识别等。模型整体为编码器-解码器 (Encoder-Decoder) 架构,如图 1 所示。

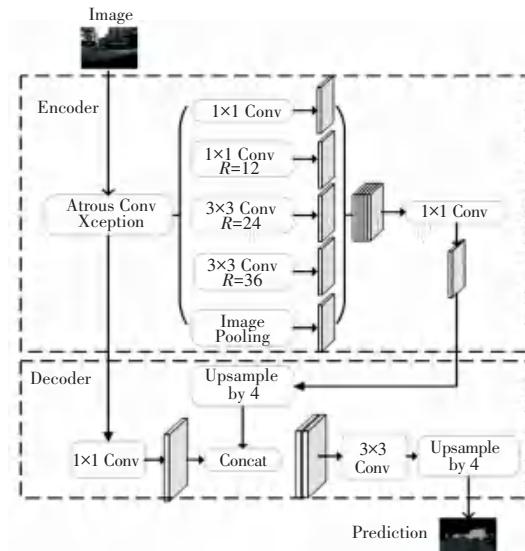


图 1 DeeplabV3plus 模型结构

Fig. 1 Structure of DeeplabV3plus model

DeeplabV3plus 模型编码器使用的是 Xception 模型,加上 ASPP 模块进行特征的多尺度融合,解码器部分引出了特征提取网络中下采样 4 倍后的特征层,与 ASPP 模块输出 4 倍上采样后的结果进行拼

接,融合低语义信息和高级语义信息,最后进行 4 倍上采样还原图片,带来的问题就是网络没有充分利用低语义层次的信息,而且直接进行两次 4 倍上采样会带来极大的信息损失,导致分割精度不够。

2 改进 DeeplabV3plus 模型

改进 DeeplabV3plus 模型整体结构如图 2 所示。编码器方面,将 Xception 模型替换为 ResNet50 模型增强特征提取能力,并在 ResNet50 模型最后两个残差层中引入膨胀卷积,在不增加参数量的情况下进

一步增大感受野,以便于更好的提取图像特征;在 ASPP 模块中,将卷积核膨胀率分别为 [6, 12, 18] 的膨胀卷积核的膨胀率增大为 [12, 24, 36] 进一步提升其特征提取能力;解码器方面,基本单元使用自行设计的 UCA (Upsample-Concat-Attention) 模块,将编码器 4 个下采样层的输出和原图输入进 UCA 模块进行深浅层次特征的充分融合,丰富语义信息,还增加了 CA 注意力机制,进一步帮助网络过滤浅层特征噪声,聚焦有用信息,增加分割性能。

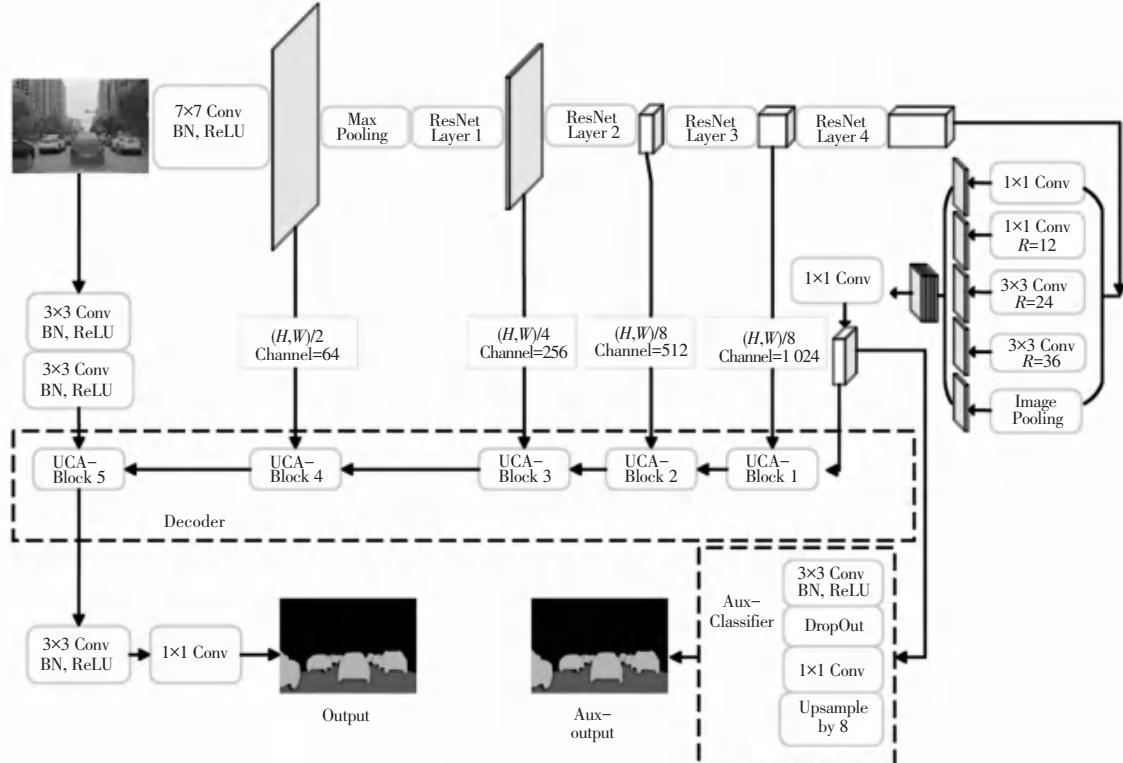


图 2 改进 DeeplabV3plus 模型结构

Fig. 2 Structure of improve DeeplabV3plus model

2.1 编码器部分

一般来说,随着 CNN 网络层数加深,网络可以获得更深层次的图像特征,但是层数加深带来的梯度消失或者爆炸问题使得深层次模型训练困难,He 等^[15]提出的残差网络 ResNet 解决了这一问题,使得特征提取网络可以极大的加深并训练出优异的效果。针对道路场景分割中环境复杂、目标多、部分目标小等问题并综合计算成本,本文选取 ResNet50 作为编码器提取图像特征,其基本残差模块结构如图 3 所示,2 个残差模块组成一个瓶颈层 (BottleNeck-Layer),而多个瓶颈层组成残差层 (ResNet-Layer) 结构。

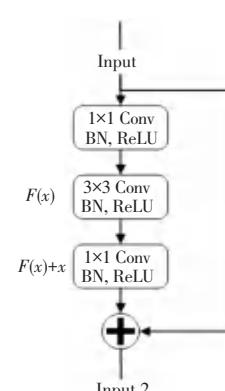


图 3 基本残差模块

Fig. 3 Basic residual module

为了增大编码器网络的感受野,在第 3 和第 4 个 ResNet-Layer 中,还使用了不同膨胀率的膨胀卷积,膨胀卷积是在 DeeplabV1 版本中提出的,其主要目的

是在不额外增加参数的情况下,使得卷积核感受野增大,从而获取到更多的上下文信息,极大的提升了语义分割任务的准确性。具体配置参数见表 1。

表 1 编码器配置

Table 1 Encoder configuration

卷积层	层内配置	输出尺寸	卷积核膨胀率
ResNet-Layer1	BottleNeck-Layer × 3	(H, W)/2	0
ResNet-Layer2	BottleNeck-Layer × 4	(H, W)/4	0
ResNet-Layer3	BottleNeck-Layer × 6	(H, W)/8	[1, 2]
ResNet-Layer4	BottleNeck-Layer × 3	(H, W)/8	[2, 4]

由于输入网络的图片分辨率较高,为了降低下采样后高频信息的损失,ResNet-Layer4 之后,并没有继续下采样,故最终输入到 ASPP 模块的特征图只下采样到 8 倍,没有延续作为分类网络时 ResNet50 下采样到 16 倍。而为了帮助网络训练收敛,还增加了一个训练辅助分类器(Aux-classifier),其结构简单,只是直接将解码器中 ASPP 模块的输出经过一个 3×3 卷积和一个 1×1 卷积进行通道调整,然后通过 8 倍双线性插值进行上采样,并将结果和真实标签计算交叉熵损失值,网络最终损失值由辅助分类器和网络共同组成,计算公式如下:

$$\text{Loss} = \text{Loss}(\text{network}) + 0.5 \times \text{Loss}(\text{aux}) \quad (1)$$

2.2 解码器部分

解码器部分主要由一系列级联的 UCA 模块组成,将编码器中每一个下采样后的特征图输入到 UCA 模块中,通过融合不同语义层次、不同尺寸的特征图,使得网络获取到更多的上下文和目标多尺度信息,从而获得更佳的分割效果,与目标检测中解决目标多尺度的 FPN(Feature Pyramid Networks)特征金字塔网络不同的是,本文在进行特征融合时并没有直接相加,而是进行拼接操作,这样虽然增加了参数量,但是拼接操作并未对特征图进行变换,因此,在训练时损失梯度进行反向传播时,高频信息并不会丢失。UCA 模块结构如图 4 所示。

UCA 模块包含两个输入,其中输入 1 主要由解码器中残差层输入,输入 2 主要由前一级 UCA 模块的输出作为输入,其计算步骤为将残差层输出,首先进行 3×3 卷积降维,经过一个注意力模块过滤后,与前一级输出进行拼接,融合深层次与浅层次语义信息,然后经过两个 3×3 卷积将通道数调整为拼接后的通道数的一半,最后进行输出。本文所使用 5 个 UCA 模块,融合编码器下采样 2、4、8 倍的特征图以及原图,充分利用浅层次特征信息。每个 UCA 模

块的具体配置见表 2。

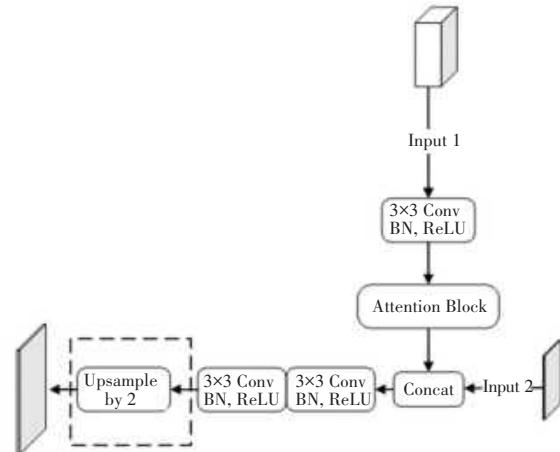


图 4 UCA 模块结构

Fig. 4 Structure of UCA block

表 2 UCA 模块配置

Table 2 UCA module configuration

模块名称	Input1 通道	Input2 通道	Output 通道	上采样
UCA_Block1	1 024	256	128	False
UCA_Block2	512	128	64	False
UCA_Block3	256	64	32	True
UCA_Block4	64	32	16	True
UCA_Block5	32	16	32	True

2.3 引入注意力机制

道路场景检测中,对于目标边界信息和位置信息关注度较高。精确的可行驶区域的分割结果十分影响自动驾驶感知系统的决策,常用的注意力机制比如 SE 注意力机制只关注通道之间的相互依赖信息而忽略空间信息,CBAM 注意力机制虽然分别对空间和通道进行了注意力机制的处理,但是忽略了长距离依赖。针对上述问题,在编码器与解码器之间引入 CA 注意力机制,其结构如图 5 所示。

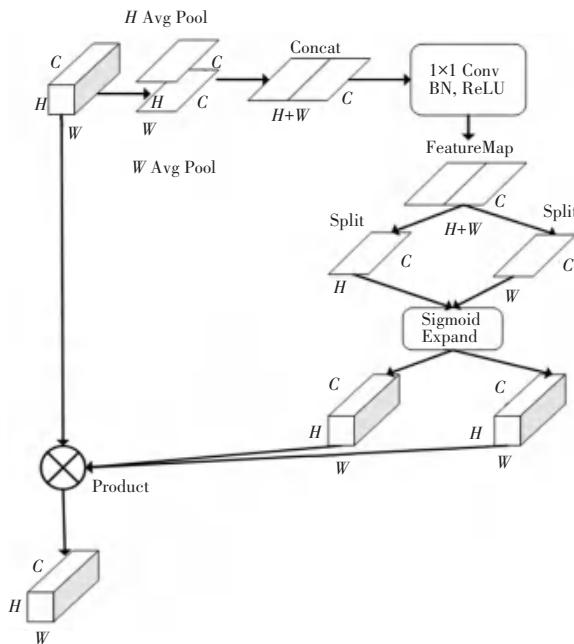


图 5 CA 注意力机制结构示意图

Fig. 5 Structure of CA attention module

CA 注意力机制没有使用全局平均池化,而是对特征图 x 沿着高宽方向进行平均池化,计算公式如下:

$$Z(h, c) = \frac{1}{W} \sum_{0 < i < W} x_c(h, i) \quad (2)$$

$$Z(w, c) = \frac{1}{H} \sum_{0 < i < H} x_c(w, i) \quad (3)$$

其中, H 、 W 为特征图的高、宽; $x_c(h, i)$ 为特征图 x 在 c 通道处 (h, i) 坐标的像素值; $Z(h, i)$ 为输出特征图在 (h, i) 坐标的值。

这样的转换不仅有助于在空间方向上捕捉长距离的依赖关系,在另一个空间方向上还能保留位置信息,有助于网络更准确的定位感兴趣的区域。对得到的两个池化向量进行拼接和卷积变换,再加上非线性激活函数得到新特征图 F ,计算公式如下:

$$F = \delta(\text{Conv2d}(\text{Concat}(Z^w, Z^h))) \quad (4)$$

其中, δ 为非线性激活函数。

接着对新特征图 F 进行分裂(Split)操作,获得两个张量,经过 1×1 卷积操作和 Sigmoid 非线性激活函数后得到两个注意力向量 \mathbf{g}^w , \mathbf{g}^h ,计算公式如下:

$$\mathbf{g}^w, \mathbf{g}^h = \text{Sigmoid}(\text{Conv2d}(\text{Split}(F))) \quad (5)$$

最后,将注意力向量进行维度扩张后与原始特征图进行对应位置相乘,计算如下式:

$$\text{out}_c(i, j) = x_c(i, j) \times \text{pad}(\mathbf{g}^w) \times \text{pad}(\mathbf{g}^h) \quad (6)$$

3 实验和分析

3.1 实验数据集及实验配置

数据集包含两部分,其一为在四川省眉山市采集的道路信息数据,自建数据集,共 1 200 张图片,每张图片分辨率为 1920×1080 ,手动标注为背景、可行驶区域、行人、机动车和非机动车,其中 1 000 张加入训练集,200 张加入验证集;其二为公共数据集 Cityscape,共 5 000 张精细标注图片,为保持与自建数据集类别一致,将其标注化为与自建数据集中一样的类别,数据集样例如图 6 所示。



(a) Cityscape 数据集样例及标注 (b) 自建数据集样例及标注

图 6 数据集样例

Fig. 6 Dataset example

实验平台配置见表 3。为加快训练速度,开启了混合精度训练,数据增强使用随机裁剪、水平翻转、随机缩放等操作,选用随机梯度下降 SGD(Stochastic Gradient Descent)优化算法,动量设置为 0.9,初始学习率为 0.000 1,使用 ResNet 原论文提出的 warmup 学习率策略,所有模型统一训练 200 轮。

表 3 实验平台配置

Table 3 Experimental platform configuration

名称	配置
操作系统	Ubuntu20.04
CPU	i7-13700KF
内存	32 GB
显卡	RTX 4070ti
Pytorch 版本	1.10.0
Cuda 版本	11.3

3.2 评价指标

语义分割任务中,最常用的评价指标为像素精确度 PA(Pixel Accuracy)和平均交并比 mIoU(mean Intersection over Union)。PA 表示分类正确的像素个数占总像素个数的比例,计算公式如下:

$$\text{PA} = \frac{\sum_i n_{ii}}{\sum_i t_i} \quad (7)$$

其中, n_{ii} 表示真实标签为 i 预测结果也为 i 的像素个数, t_i 表示在真实标签中类别为 i 的像素总个数。

mIoU 则表示真实标注的像素区域与预测像素区域之间的重合度, 计算公式如下:

$$\text{mIoU} = \frac{1}{n_{\text{class}}} \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ij} - n_{ii}} \quad (8)$$

其中, n_{ij} 表示真实标签为 i 但是预测结果为 j 的像素总个数。

3.3 不同模型对比实验

为验证本文提出的改进 DeeplabV3plus 模型的性能, 将改进的模型与其他分割模型进行对比实验, 包括较为经典的 Unet、Lraspp、DeeplabV3plus 模型, 实验结果见表 4。可见改进后模型的 mIoU 指标相较于 U-Net、Lraspp、DeeplabV3plus 模型分别高出 12.5%、11.7%、2.6%, PA 指标分别高出 4%、3.5%、0.9%。

表 4 不同模型的检测性能

Table 4 Detection performance of different models

模型	mIoU/%	PA/%	参数量/ 10^6
U-Net	74.3	93.1	17.26
Lraspp	75.1	93.6	3.21
DeeplabV3plus	84.2	96.2	39.75
本文模型	86.8	97.1	44.19

3.4 改进前后模型对比实验

实验中发现 DeeplabV3plus 对于部分小目标如行人、被遮挡车辆等、部分可行驶区域分割仍然并不够精准。语义分割任务中, 网络预测结果的精确度

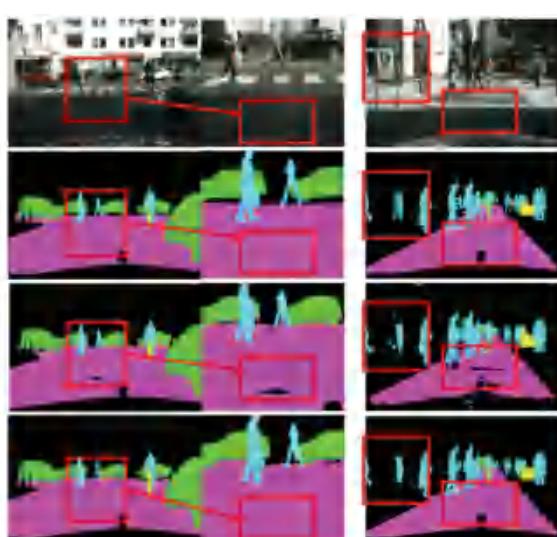
与 mIoU 强相关, 改进前后模型在每一个类别上的精度对比结果见表 5, 可见对于每一个类别, 改进后模型均高于原模型。特别的, 对于目标较小、数量较少的行人与非机动车类别上, 改进后模型分别高出原模型 3.2%, 4.3%。改进后模型的 mIoU 和 PA 指标分别高出原模型 2.6%, 0.9%。

表 5 改进前后模型指标对比

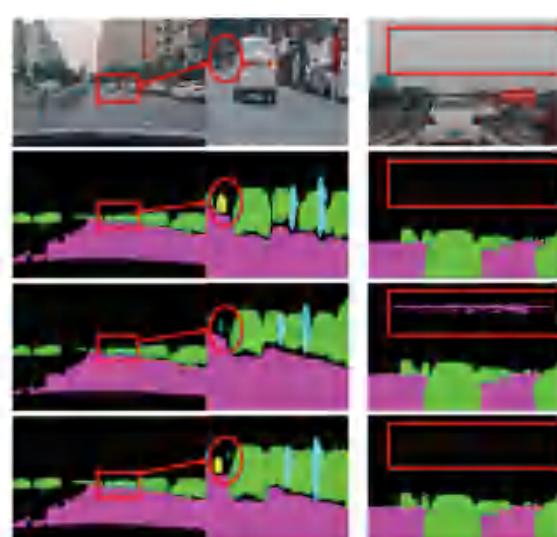
Table 5 Comparison of model metrics before and after improvement

类别	DeeplabV3plus	改进后模型
背景(黑色)	94.4	95.7
可行驶区域(粉色)	91.3	93.3
行人(蓝色)	77	80.2
车辆(绿色)	90.4	92.6
非机动车(黄色)	67.9	72.2
mIoU	84.2	86.8
PA	96.2	97.1

对分割结果进行可视化如图 7 所示。图 7 中第一行图片为原图和局部放大图, 第二行为标注图片, 第三行为 DeeplabV3plus 模型预测结果, 第四行为改进后模型预测结果。用红色框框出明显存在差异的地方, 如果在效果图中不明显的地方, 如远处的小目标等, 则予以放大显示。可见图 7 中第一张和第二张可行驶区域存在漏检, 在第三张中远处非机动车(黄色区域)DeeplabV3plus 模型识别成了车辆, 改进后模型则正确识别, 第四张图片中, DeeplabV3plus 模型将天空背景区域误检测成了可行驶区域。整体来看, 无论是分割精度还是图像细粒度, 改进后的模型均好于原模型。



(a) Cityscape 验证集中图片



(b) 自建数据集图片

图 7 DeeplabV3plus 与改进模型对比

Fig. 7 Comparison between DeeplabV3plus and improved models

3.5 不同注意力机制对比实验

为验证 CA 注意力机制有效性, 分别使用基于通道注意力机制的 SE 注意力机制, 基于空间和通道注意力机制的 CBAM 注意力机制, 无注意力机制和 CA 注意力机制进行对比实验。为方便说明, 将改进后无注意力机制的模型称之为 MyModel, 结果见表 6。

表 6 不同注意力机制对比结果

Table 6 Comparison results of different attention mechanisms

模型	mIoU/%	PA/%	参数量/ 10^6
MyModel	86.3	96.9	44.180
MyModel+SE	86.5	97.0	44.191
MyModel+CBAM	86.5	97.0	44.191
MyModel+CA	86.8	97.1	44.196

由表 6 可见, 对于没有添加注意力机制的 MyModel 模型, 由于低语义特征中存在部分噪声, 在进行深浅语义特征融合时网络中引入了噪声, 其精度稍微低于其他几个模型, 而加入了 CA 注意力机制的模型效果最好, 在 mIoU 指标上高于其余两种注意力机制 0.3%, 而模型参数量相比于未加入注意力的模型只增加了约 0.016 M 的参数量, 训练曲线如图 8 所示。

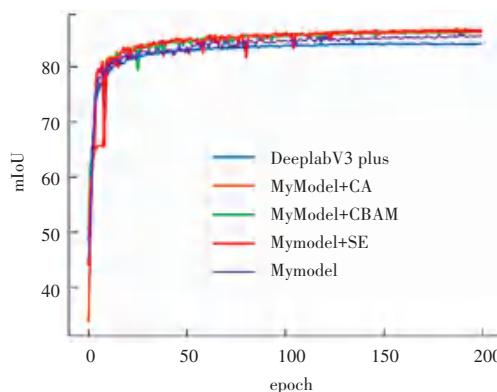


图 8 训练时 mIoU 变化

Fig. 8 Changes in mIoU during training

4 结束语

为解决 DeeplabV3plus 模型在道路场景分割中精度不够的问题, 本文进行了一系列改进, 包括对编码器的改进, 重新设计解码器, 并在解码器中融入 CA 注意力机制, 提升了模型分割精度。实验结果表明, 改进后模型相比于原模型增加了约 11% 的参数量, mIoU 提升了 2.6%, 更容易捕获小目标, 分割精度也有明显的提升; 融合注意力机制后, 相比于未添加注意力机制的改进模型, mIoU 提升约 0.5%, 使用

CA 注意力模块相比于其他注意力机制, 网络的参数量增加很少, 精度提升约 0.3%。

改进后的模型参数量较多, 不利于边缘设备的部署和推理, 后续可以尝试使用移动端友好的轻量级网络, 或者使用大模型对轻量级模型进行知识蒸馏、剪枝等轻量化操作, 以期在边缘设备上获得精度与推理时间的平衡。

参考文献

- [1] 王龙飞, 严春满. 道路场景语义分割综述[J]. 激光与光电子学进展, 2021, 58(12): 44–66.
- [2] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640–651.
- [3] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv preprint arXiv:1412.7062, 2014.
- [4] 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, 2017, 40(4): 834–848.
- [5] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking Atrous Convolution for Semantic Image Segmentation [J]. arXiv preprint arXiv:1706.05587, 2017.
- [6] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]// Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 801–818.
- [7] 徐聪, 王丽. 基于改进 DeepLabV3+ 网络的图像语义分割方法 [J]. 激光与光电子学进展, 2021, 58(16): 225–232.
- [8] 张蕊, 李锦涛. 基于深度学习的场景分割算法研究综述[J]. 计算机研究与发展, 2020, 57(4): 17.
- [9] 何雪东, 宣士斌, 王款, 等. 融合累积分布函数和通道注意力机制的 DeepLabV3+ 图像分割算法 [J]. 计算机应用, 2023, 43(3): 936–942.
- [10] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 7132–7141.
- [11] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C]// Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 3–19.
- [12] LIU T, LUO R, XU L, et al. Spatial channel attention for deep convolutional neural networks [J]. Mathematics, 2022, 10(10): 1750.
- [13] VASWANI A, SHAZER N, PARMAR N, et al. Attention is all you need [C]// Proceedings of the Advances in Neural Information Processing Systems. NIPS, 2017: 30.
- [14] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 13713–13722.
- [15] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770–778.