doi:10.3969/j.issn.2097-3853.2024.01.014

改进的 DDeepLabV3+语义分割网络

蔡思静,汪严昱

(福建理工大学 电子电气与物理学院,福建 福州 350118)

摘要:针对语义分割网络在移动智能化终端上存在参数量大、分割精度不足的问题,提出一种改进的 DDeepLabV3+网络算法。首先,采用深度可分离的 MobileNet 结构作为网络的骨干部分,降低网络的参数量和复杂度,从而有效减少了运行时间。其次,引入网络的低级特征,实现多尺度信息融合,减少网络下采样引起的空间信息损失。最后,结合注意力机制设计网络 ASPP 结构,增强特征提取在实验中的利用。优化后的网络结构在保持较高分类准确性的前提下,计算时间显著减少。网络的平均交并比在 Cityscapes 和 Camvid 数据集中分别提升了 2.37%和 2.13%。

关键词:语义分割;SE 注意力机制模块;DeepLabV3+网络

中图分类号:TP391.41　　**文献标志码**:A　　**文章编号**:2097-3853(2024)01-0095-08

Improved DDeepLabV3+ semantic segmentation network

CAI Sijing, WANG Yanyu

(School of Electronics, Electrical and Physics, Fujian University of Technology, Fuzhou 350118, China)

Abstract: Aiming at the problems of too large a number of parameters and insufficient segmentation accuracy of semantic segmentation network on mobile intelligent terminals, an improved DDeepLabV3+ network algorithm was proposed. First, the depth-separable MobileNet structure is used as the backbone of the network to reduce the number of parameters and complexity of the network, thereby effectively reducing the running time. Secondly, low-level features of the network are introduced to achieve multi-scale information fusion and reduce the spatial information loss caused by network downsampling. Finally, the network ASPP structure is designed based on the attention mechanism to enhance the utilization of feature extraction in experiments. The optimized network structure significantly reduces the calculation time while maintaining high classification accuracy. In the Cityscapes data set used in the study, the average intersection and union ratio of the network increased by 2.37%, while in the Camvid dataset, the ratio increased by 2.13%.

Keywords: semantic segmentation; SE attention module; DeeplabV3+ network

语义分割是计算机视觉的重要组成部分之一,其应用广泛,涵盖自动驾驶、无人机图像分割、智慧安防、医学影像等领域[1]。随着分割任务对精度和时效性要求的提升,新的骨干网络应运而生。针对分割模型的速度与准确率等关键性能指标的提升对语义分割的骨干网络开展研究,具有重要的现实意义。

语义分割主要基于 FCN(fully convolutional network)网络的改进,FCN 在卷积得到特征图后进行反卷积的上采样,得到与原图大小一致的分割图[2]。DeepLab 系列网络是对 FCN 网络的优化,最早的 DeepLabV1[3]网络采用 VGG 做为主干网络,并且利用空洞卷积的方法扩大感受野。DeepLabV2[4]在此基础上引入了 ASPP(atrous spatial pyramid pooling)进行多尺度的特征融合,进一步提高了分割精度。DeepLabV3[5]采用并联

收稿日期:2023-09-28

第一作者简介:蔡思静(1983-),女,福建南平人,副教授,博士,研究方向:图像处理、人工智能、计算机视觉、机器学习的工程应用,卷积神经网络模型设计。

或者级联的 ASPP 模块,调整了 ASPP 模块中的参数并舍弃了条件随机场。DeepLabV3+[6] 在之前的基础上,把从深度卷积网络中获取到的低级特征与经过 ASPP 后的高级特征在解码部分进行融合,进一步提高分割精度。

然而,DeepLab 系列网络主要着眼于像素分割的精度的提高,没有解决分割速度问题,难以应用于嵌入式设备。针对自动驾驶与人机交互等任务对分割精度、模型参数量和实时性等的高要求,本研究在 DeepLabV3 网络的基础上,以 MobileNet[7] 结构作为网络的骨干,增加了一条低级特征,并根据 DenseASPP[8] 和深度可分离卷积的思想,提出了一种改进的轻量级语义分割网络,DDeepLabV3+网络,在降低模型参数量的同时提高其精度,以适用于不同场景。

1 DeepLabV3+网络结构设计

1.1 DeepLabV3+网络

DeepLabV3+网络由编解码组成,结构如图 1 所示。在编码部分,原始图像经过深度卷积神经网络进行信息提取;根据不同的骨干网络和任务要求,图像经过骨干网络后下采样 8 倍或者 16 倍;随后,将提取到的信息分别经过 1×1 卷积,空洞率为 6、12、18 的空洞卷积和池化层;将得到的结果在通道维度拼接,再通过 1×1 卷积降低通道数完成编码结构。

解码结构是由骨干网络中的低级特征经过 1×1 卷积降维后与编码层上采样的结果进行通道上的堆叠,再经过 3×3 卷积和上采样操作,返回原图大小,得到最终分割结果。

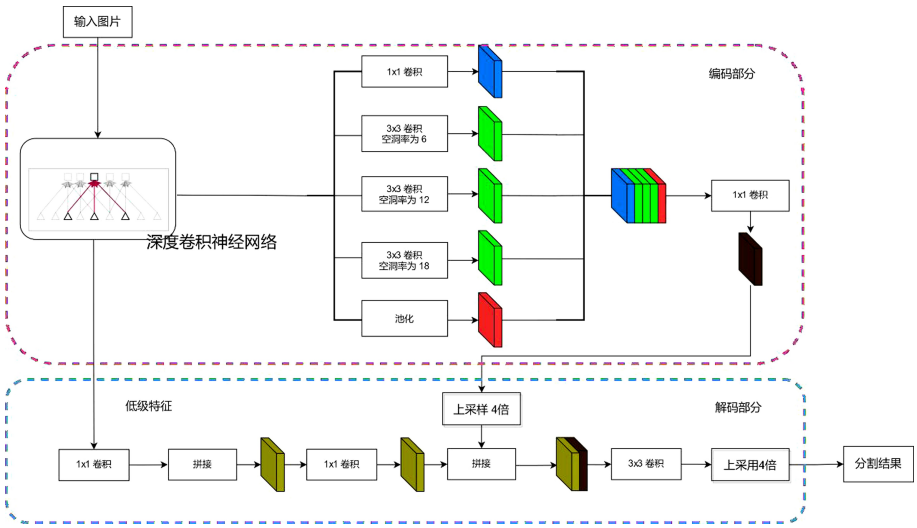

图 1 DeepLabV3+的网络结构

Fig.1 Network structure of DeepLabV3+

1.2 深度可分离卷积

深度可分离卷积将标准的卷积过程拆分为 DW 卷积(depthwise convolution)和 PW 卷积(point-wise convolution),如图 2、图 3。深度可分离卷积与传统卷积相比参数量少,运算成本低。传统卷积的卷积核通道数为输入特征图的通道数,卷积核的个数依靠参数设定,而 DW 卷积的的卷积核通道数始终为 1,卷积核的个数也始终与输入特征图的通道数相同。DW 卷积后再进行 PW 卷积。PW 卷积是大小为 1×1 的普通卷积,可以通过设置卷积核的个数来控制输出的通道数,以融合输入的通道信息。

假定输入特征图的大小为 D_I,通道数为 M,卷积核的大小为 D_k,个数为 N,在普通卷积的运算过程中,卷积的计算量为 $D_I \times D_I \times M \times D_k \times D_k \times N$。而深度可分离卷积的计算量为 $D_I \times D_I \times M \times D_k \times D_k + M \times D_I \times D_I \times N$。深度可分离卷积与普通卷积计算量的比为 $(1/N+1/D_k^2):1$,由于卷积核的大小一般为 3,所以深度可分离卷积与普通卷积的计算量比为 $(1/N+1/9):1$,理论上普通卷积的计算量是深度可分离卷积的 8 到 9 倍。

1.3 MobileNetV2 主干网络

MobileNet 是一种轻量级网络模型,主要采用深度可分离卷积构成。MobileNetV2 通过反向残差结构来提高网络的性能。相对于传统 FCN 模型,选择 MobileNetV2 作为分割模型能够减少大

概 88%的计算资源消耗,同时保持近似分割准确率,更符合低功耗和实时性方面的特点,正好符合研究的目标和需求。因此,DDeepLabV3+选择采 MobileNetV2 作为分割模型。为解决图片分辨率过大的问题,将骨干网络的部分卷积替换为空洞卷积以扩大感受野、减少资源的耗费,对网络的运行速度也有着一定的优化作用。

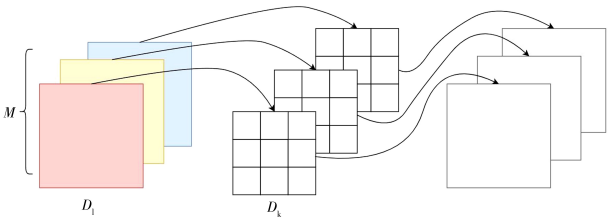

图 2　DW 卷积示意图

Fig.2　Schematic diagram of DW convolution

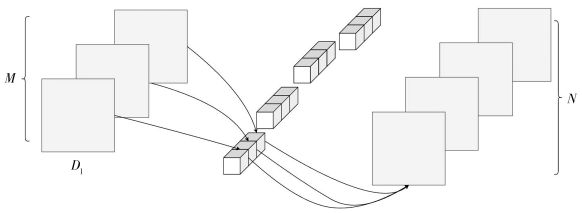

图 3　PW 卷积示意图

Fig.3　Schematic diagram of PW convolution

MobileNetV2 网络共有 7 个线性瓶颈结构,每个线性瓶颈结构由若干个倒残差结构组成,共有 17 个倒残差结构。每个倒残差结构又由深度可分离卷积和 1×1 卷积组成。其中,深度可分离卷积、在 stride=1 且输入特征图与输出特征图大小相同的情况下有跳跃链接。

1.4　SE 注意力机制模块

SE 注意力机制[9]模块对通道维度进行注意力加权操作,让网络自动学习到重要通道的特征,忽略影响较小的特征,使网络在保持模型准确性的同时提高了运行效率。

压缩和激励网络(squeeze-and-excitation networks,SEnet)大致分为压缩、激励、比例相乘 3 个操作,如图 4 所示,其中 C、H、W 分别表示输入特征图的通道数、高和宽度。压缩操作把输入特征图经过一个全局平均池化下采样,实现对每个通道的信息的压缩;激励操作把压缩后的特征图经过两个全连接层,使得模型学会为每个通道动态地分配不同的权重;比例相乘操作把得到的权重与输入特征图相乘,得到输出结果。

表 1　原始的 MobileNetV2 网络层结构

Tab.1　Original MobileNetV2 network layer structure

层号	操作	通道数/个	该层的重复次数/次	卷积通道的扩张率	步距
1	卷积	32	1	—	2
2	瓶颈结构	16	1	1	1
3	瓶颈结构	24	2	6	2
4	瓶颈结构	32	3	6	2
5	瓶颈结构	64	4	6	2
6	瓶颈结构	96	3	6	1
7	瓶颈结构	160	3	6	2
8	瓶颈结构	320	1	6	1
9	1×1 卷积	1 280	1	—	1
10	7×7 池化	—	1	—	—
11	1×1 卷积	k	—	—	—

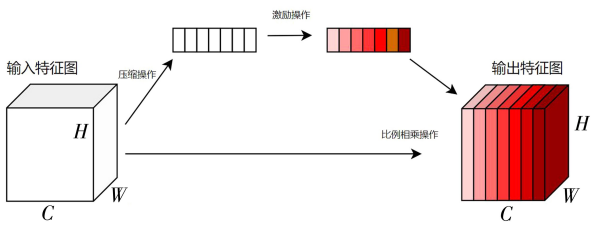

图 4　SE 注意力机制的网络结构

Fig.4　Network structure of SE attention mechanism

2　改进网络 DDeepLabV3+

改进的网络 DDeepLabV3+是由主干网络、DDASPP 和上采样模块组成。首先,用改进的 MobileNetV2 替换了 DeepLabV3+的 Xception 骨干网络,可以极大地减少网络的参数数量。其次,参考了特征融合策略,使 DeepLabV3+网络进一步融合浅层特征让网络保留更多的浅层信息,即在编码器中多提取一条语义分支,在解码端融合多尺度信息。接着,将主干网络的特征经过改进的 ASPP 结构,增加了感受野并提高了主干网络特征的利用率。最后,在多处加上 SE 注意力机制,使得模型可以将更多的注意力集中在具有较高表征能力的通道上。改进后的网络示意图如图 5 所示。

2.1　MobileNetV2 主干网络优化

因 MobileNetV2 从第 9 层开始网络的通道数急剧上升到 1 280,计算量会大量增加。为了减少

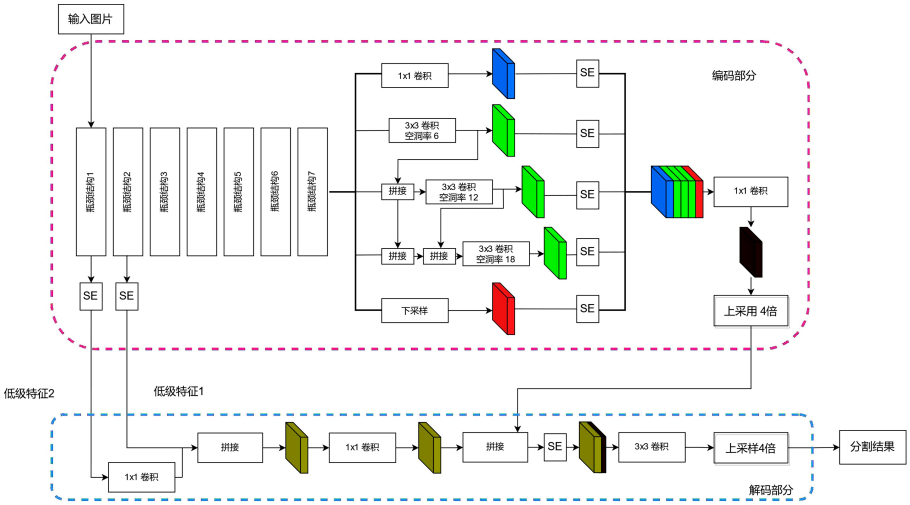

图 5 DDeepLabV3+的网络结构

Fig.5 Network structure of DDeepLabV3+

计算和内存的浪费,DDeepLabV3+仅采用 MobileNetV2 的前 8 层网络。原始 MobileNetV2 网络在默认的情况下是用于图像分类任务,下采样了 32 倍,导致图像的会丢失一些细节。DDeepLabV3+在考虑细节丢失和显存资源的情况下将第 14 个线性瓶颈结构的 stride 改为 1,相当于最后只下采样了 16 倍,使其更好的适用于语义分割任务。下采样的减少同时会导致图像的感受野变小,DDeepLabV3+将第 8~17 倒残差结构的深度可分离卷积替换为带孔洞的深度可分离卷积。其中 15~17 的 dialation 设为 2。改进后的 MobileNetV2 网络各层结构如表 2 所示。其中,新增的空洞率表示空洞卷积各元素的间隔,当空洞率等于 1 时,空洞卷积就是普通卷积。

表 2 改进后的 MobileNetV2 网络层结构

Tab.2 Improved MobileNetV2 network layer structure

层号	操作	通道数	该层重复次数	卷积通道的扩张率	步距	空洞率
1	卷积	32	1	—	2	1
2	瓶颈结构	16	1	1	1	1
3	瓶颈结构	24	2	6	2	1
4	瓶颈结构	32	3	6	2	1
5	瓶颈结构	64	4	6	2	1
6	瓶颈结构	96	3	6	1	1
7	瓶颈结构	160	3	6	1	2
8	瓶颈结构	320	1	6	1	2

2.2 DDASPP 模块

为更好地保留图像细节特征,DDeepLabV3+使用空洞卷积在保持较大感受野的同时有效地捕捉更广范围的信息,在一定程度上缓解了下采样可能引起的细节丢失问题。DeepLabV3+通过 ASPP 模块分别从多个尺度的特征图像中提取信息并融合输出结果。为了解决小物体识别精度较低及因感受野增大导致部分信息丢失等问题,DDeepLabV3+采用根据 DenseNet 网络提出的密集连接结构,将网络中的 ASPP 重构为密集的 ASPP,并将普通卷积替换成深度可分离卷积,在提升分割精度的同时减少了网络的分割时间。将重构的模型命名为 DDASPP(depthwise dense ASPP)。在 DDASPP 中,随着空洞率的增大,低尺度特征信息能够在高尺度卷积过程中被有效地复用,从而使输入图像的特征提取更加密集。该方法不仅提高了高维特征点的利用效率,而且在深度神经网络中充分发掘了浅层信息,从而提高了整个模型的分割性能。

3 实验结果及分析

3.1 数据集介绍

3.1.1 Cityscapes 数据集

Cityscapes 数据集采用双摄像头拍摄立体视频序列,在图像分割中使用的是左摄像头的图片,包含了 50 多个不同城市的视频序列,有精细和粗糙两种标注图像。DDeepLabV3+使用精细标注的图片,使用 2 975 张训练图片进行训练,用 500 张

验证图片进行验证跟测试,图像的分辨率都为 1 024×2 048 像素,DDeepLabV3+使用常用的 19 种语义类别进行实验。

3.1.2　Camvid 数据集

Camvid 数据集是由剑桥大学公开发布的城市道路场景的数据集,该数据集提供了高质量的 30 Hz 视频镜头,有 32 个真实的标签信息。数据集一共有 701 张图片,实验中随机划分 367 张训练图片进行训练,100 张图片用来验证,234 张图像用来测试,图像的分辨率都为 960×720 像素,常用的 11 种语义类别进行实验。

3.2　实验环境与评价指标

实验选用 64 位 Win10 为操作系统,CPU 为 11th Gen Intel(R) Core(TM) i7-11800H @ 2.30 GHz,16 GB 内存,GPU 为 NVIDIA GeForce RTX 3060 Laptop,6 G 显存。开发环境是 pycharm,深度学习框架为 pytorch1.11.0、python 3.8、CUDA 11.3、CUDNN 8.2,其他参数设置如表 3 所示。

表 3　参数设置

Tab.3　Parameter settings

超参数名称	Cityscape 数据集参数	Camvid 数据集参数
crop size	768×768 像素	720×720 像素
validation size	1 024×2 048 像素	960×720 像素
Loss function	Cross entropy	Cross entropy
batch size	2	2
optimizer	SGD	SGD
scheduler	Poly	Poly
max iteration	16 1000	30 000
output stride	16	16

Cityscape 和 Camvid 数据集的 crop size 在设为 513×513 的情况下收敛缓慢,分割精度低,在考虑显存资源的情况下将 Cityscapes 的 crop size 设置为 768×768 像素,将 Camvid 的 crop size 设置为 720×720 像素, batch size 都设置为 2。验证时采用单张图片验证,验证尺寸为原图大小。优化器采用 SGD,动量(Momentum)设为 0.9,采用 poly 的学习策略,学习率可表示为:

$$\mathrm{lr}=\mathrm{base\_lr}\times\left(1-\frac{\mathrm{cur\_itr}}{\mathrm{max\_itr}}\right)^{\mathrm{power}} \tag{1}$$

式中,base_lr 为初始学习率,cur_itr 为当前迭代次数,max_itr 为最大迭代次数,power 是衰减指数。

在语义分割中,mIoU 值是体现分割精度的重要参数,其计算公式可表示为:

$$\mathrm{mIoU}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \tag{2}$$

式中,k+1 表示 k 个语义类别和 1 个背景,p_{ij}为将 i 类别预测为 j 类别。

MPA 是计算每个类别的正确像素比例再求平均,可表示为:

$$\mathrm{MPA}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}} \tag{3}$$

DDeepLabV3+的评价指标为平均交并比(mIoU)、平均像素精度(MPA)、浮点计算量 FLOPs 和参数量。

3.3　实验结果

3.3.1　Cityscapes 数据集消融实验

为了验证网络改进部分和添加注意力机制的效果,DDeepLabV3+设置了不同的对比实验。在同一深度学习框架下,比较不同模块对网络分割精度效果的影响,并在数据集上进行验证,结果如表 4 所示。其中,Base model 为替换了 MobileNetV2 的 DeepLabV3+网络。

表 4　不同方案的消融对比

Tab.4　Ablation comparison of different schemes

模型	mIoU	参数量/MB	浮点运算量/G	帧数
Base model	71.53	5.225	134.35	16.07
Base model+SE	71.88	5.233	134.36	15.98
Base model+SE+Low_feature 2	72.58	5.246	134.55	15.69
Base model+SE+Low_feature 2+DDASPP	73.90	3.492	120.18	16.23

由表 4 可知,与只改变主干网络的 Base model 相比,添加 SE 注意力机制后模型的 mIoU 提升了 0.35%;添加了 SE 注意力机制和低级特征

后模型的 mIoU 提升了 1.05%;在此基础上再加入 DDASPP 模块后模型的 mIoU 比原网络提升了 2.37%。DDeepLabV3+与 Base model 相比,参数数量减少了 1.733 MB,验证时间也有少量的减少。以上结果表明,SE 注意力机制、额外的低级特征提取和 DDASPP 的改进一定程度上可以有效提高模型的分割精度。

从表 5 可以看出,对比基准模型,DDeepLabV3+每个类别的 IoU 都有一定程度的提升,其中对于墙体、栅栏、交通灯、卡车和火车的精度提升较大,验证了 DDeepLabV3+的有效性。

表 5 在 Cityscape 数据集上 19 个类别的分类的交并比(IoU)结果

Tab.5 IoU results of 19 categories classification on Cityscape dataset

类别	IoU	
	基准模型	DDeepLabV3+
道路	97.49	97.62
人行道	80.93	81.82
建筑物	90.82	91.36
墙	46.13	48.29
栅栏	54.41	58.24
栏杆	58.27	60.16
交通灯	59.83	63.46
交通标志	71.23	73.99
植被	91.46	91.73
地形	59.47	62.60
天空	93.85	94.23
行人	77.63	78.87
骑手	54.03	56.59
汽车	93.21	94.06
卡车	66.97	74.28
公共汽车	77.65	80.30
火车	61.09	65.02
摩托车	52.67	57.70
自行车	71.93	73.69

3.3.2 不同网络在 Cityscapes 数据集上的对比

不同网络在城市景观数据集上的不同表现如表 6 所示。从表 6 可见,Deeplabv3+作为高精度的分割网络之一,其分割效果是比较有优势的。DDeepLabV3+在 DeeplabV3+基础上精度提升了 2.37%。从表 6 可以看出,在骨干网络中,以参数大的 VGG 为模型的 FCN-8s 网络的 mIoU 低于 DDeepLabV3+,以 Resnet101 为骨干网络的 DeelabV3+网络的分割精度虽然高于 DDeepLabV3+,但分割时间几乎是后者的 3 倍。考虑时间和网络复杂度,Enet、Cgnet 和 Lednet 是极为轻量级的网络,但是分割精度远低于 DDeepLabV3+网络。

网络在 Cityscapes 数据集的分割结果如图 6 所示,从图 6(a1)的预测结果可见,DDeepLabV3+相较于原始网络对图片左上角的建筑物和右边的行人预测更准确;从图 6(a2)的预测结果可见,原始网络对图像左方摩托车和右上角的树木分割较为粗糙,DDeepLabV3+更为准确;从图 6(a3)的预测结果可见,DDeepLabV3+对于人物边界和栏杆的预测情况更加完善,效果得到了提升。

表 6 不同网络在原图尺寸的测试结果

Tab.6 Test results of different networks

模型	骨干网络	mIoU	浮点运算量/G	参数量/MB	帧数
Enet	—	47.46	17.68	0.336	26.32
Cgnet	—	51.02	27.73	0.491	21.84
Lednet	—	51.30	50.18	2.33	15.62
FCN-8s	VGG-16	62.21	2 569.42	30.036	6.08
Bisenet	Resnet18	67.96	104.07	12.796	26.45
Deeplabv3+	Resnet101	75.62	633.17	58.753	5.18
DDeepLabV3+	MobilenetV2	73.90	120.18	3.492	16.13

3.3.3 Camvid 数据集实验结果

DDeepLabV3+与基准模型的对比如表 7 所示,从表 7 可以看出,DDeepLabV3+的方法在 Camvid 数据集上的 MPA 和 mIoU 也分别提升了 2.14%和 2.13%。表 8 为 DDeepLabV3+与基准模型各类别 IoU 的对比,其中栏杆、栅栏和道路标志的 IoU 提升较大,证明了 DDeepLabV3+网络模型的泛化性和有效性。网络在 Camvid 数据集的结果如图 7 所示。

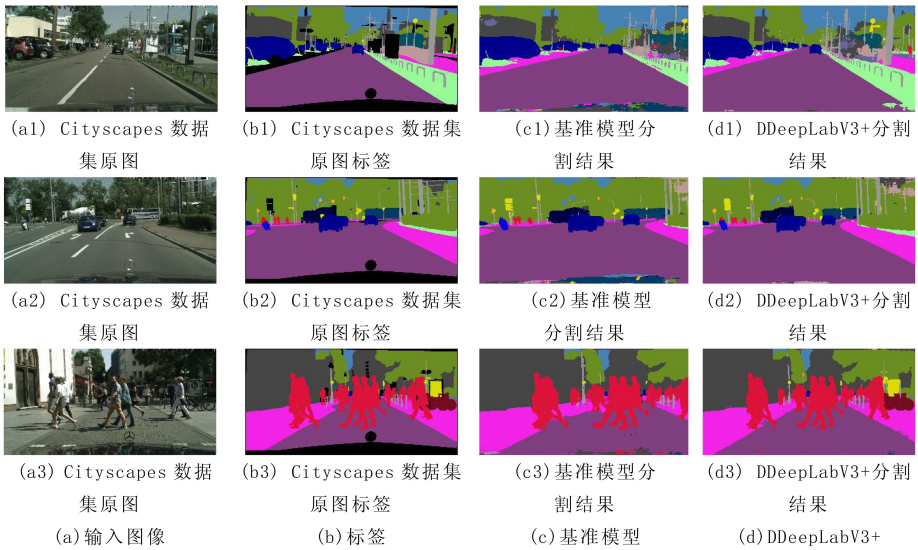

图 6　Cityscapes 数据集的分割结果

Fig.6　Segmentation results of Cityscapes dataset

表 7　与初始方案的对比

Tab.7　Comparison with initial plan

模型	骨干网络	平均像素精度	mIoU	浮点运算量/G	参数量/MB	帧数
Base model	MobilenetV2	83.94	78.00	44.39	5.225	46.61
DDeepLabV3+	MobilenetV2	86.08	80.13	39.72	3.492	47.24

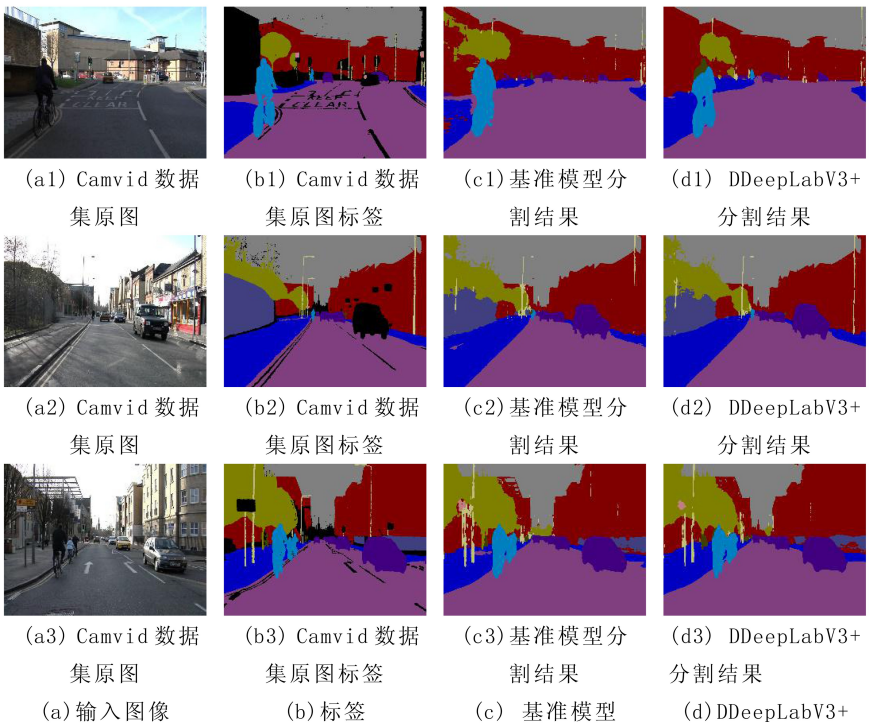

图 7　Camvid 数据集的分割结果

Fig.7　Segmentation results of Camvid dataset

表 8 在 Camvid 数据集上 11 个类别的 IoU

Tab.8 Classification results of 11 categories on Camvid dataset

类别	基准模型	DDeepLabV3+
自行车手	78.98	78.62
建筑物	93.03	93.53
汽车	91.27	92.13
栏杆	43.77	48.59
栅栏	77.80	81.27
行人	64.72	67.13
道路	97.24	97.73
人行道	87.17	89.10
道路标志	45.52	54.06
天空	93.64	93.94
树木	84.89	85.33

从图 7(d1)的预测结果可以看出,DDeepLabV3+相较于(c1)对行人、栏杆和树木的预测更准确;从图 7(d2)的预测结果可以发现,(c2)对车辆旁的行人分割错误,DDeepLabV3+对自行车和树木的分割更为准确;从图 7(d3)的预测结果可以发现,DDeepLabV3+对于建筑和人行道路的预测情况比(c3)更加完善。

4 结束语

针对城市道路的场景数据集,提出了一种改进的 DDeepLabV3+卷积神经网络模型,并在网络中引入 SE 注意力机制,重构了 ASPP 模块让网络的高级特征被重复利用,提高了网络模型的精确度并减少了网络参数,在网络复杂度和精度二者间取得了相对平衡。在解码端融合不同尺度特征,使得网络能更好地获取上下文的信息。所提网络模型在 Cityscape 数据集的 mIoU 提升了 2.37%,在 Camvid 数据集上也提升了 2.17%。然而,DDeepLabV3+模型还是存在一定的局限性,对小目标的图像分割准确度低于当前最佳的语义分割模型,无法同时满足实时性和高准确性。未来的工作方向包括在更多的数据集和更复杂的场景中进行进一步测试和验证,以及进一步结合深度学习的其他技术和方法进行优化和改进。

参考文献:

[1] 王可,沈川贵,罗孟华. 基于深度学习的图像语义分割方法综述[J]. 信息技术与信息化,2022(4):23-30.

[2] LONG J,SHELHAMER E,DARRELL T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).Boston: IEEE,2015:3431-3440.

[3] CHEN L C,PAPANDREOU G,KOKKINOS I,et al.Semantic image segmentation with deep convolutional nets and fully connected CRFs[EB/OL].(2014-12-22)[2021-02-10]arXiv:1412.7062

[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,2018,40(4):834-848.

[5] CHEN L C,PAPANDREOU G,KOKKINOS I,et al.Rethinking atrous convolution for semantic image segmentation[EB/OL].(2017-06-17)[2021-02-10]arXiv:1706.05587.

[6] CHEN L C,ZHU Y K,PAPANDREOU G,et al. Encoder-decoder withatrous separable convolution for semantic image segmentation[C]//European Conference on Computer Vision. Cham:Springer,2018:833-851.

[7] HOWARD A G,ZHU M L,CHEN B,et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. arXiv:1704.04861,2017.

[8] 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. Salt Lake City: IEEE,2018:3684-3692.

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

(责任编辑: 方素华)