

基于改进 DeepLabv3+ 网络的轻量级语义分割算法

姚燕, 胡立坤*, 郭军

广西大学电气工程学院, 广西 南宁 530004

摘要 由于深度学习中语义分割模型参数量较大且算法耗时较长, 不适合部署到移动端, 针对此问题, 提出了一种基于改进 DeepLabv3+ 网络的轻量级语义分割算法。首先, 采用 MobileNetv3 代替原 DeepLabv3+ 语义分割模型特征提取骨干网络以降低模型复杂度, 加快模型运行速度; 其次, 将空洞空间金字塔池化模块中的标准卷积替换为深度可分离卷积, 提高模型训练效率; 最后, 引入注意力机制模块和组归一化方法, 提升分割精度。所提分割算法在语义分割数据集 Cityscapes 验证集上的平均交并比 (mIoU) 达到 72.94%。实验结果表明, 与常见分割算法 SegNet、Fast-SCNN、ENet 等相比, 所提算法在减少模型参数的同时提高了分割效果。

关键词 图像处理; DeepLabv3+ 模型; MobileNetv3; 轻量级; 空洞空间金字塔池化

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202259.0410015

Improved Lightweight Semantic Segmentation Algorithm Based on DeepLabv3+ Network

Yao Yan, Hu Likun*, Guo Jun

School of Electrical Engineering, Guangxi University, Nanning, Guangxi 530004, China

Abstract Due to the large number of semantic segmentation model parameters and time-consuming algorithm in deep learning, it is not suitable for deployment to mobile terminal. To solve this problem, a lightweight semantic segmentation algorithm based on improved DeepLabv3+ network is proposed. First, MobileNetv3 is used to replace the original DeepLabv3+ semantic segmentation model backbone network for feature extraction to reduce the complexity of the model and speed up the running speed of the model; second, the standard convolution in atrous spatial pyramid pooling module is replaced by depthwise separable convolution to improve the efficiency of model training; finally, the attention mechanism module and group normalization method are introduced to improve the segmentation accuracy. The proposed segmentation algorithm achieves a mean intersection over union (mIoU) of 72.94% on the Cityscapes validation set of semantic segmentation dataset. Experimental results show that compared with common segmentation algorithms such as SegNet, Fast-SCNN, and ENet, the proposed algorithm can improve the segmentation effect while reducing the number of model parameters.

Key words image processing; DeepLabv3+ model; MobileNetv3; lightweight; atrous spatial pyramid pooling

1 引言

语义分割是深度学习的重要研究领域, 对图像

进行语义分割, 能够把图像信息映射成语义信息, 使得计算机对图像有更好的理解。语义分割在自动驾驶、遥感影像分类和医学图像等领域^[1-5]都有应

收稿日期: 2021-02-18; 修回日期: 2021-03-24; 录用日期: 2021-04-14

基金项目: 国家自然科学基金(61863002)

通信作者: hlk3email@163.com

用,同时也是移动机器人行为理论等研究中的关键技术,因此对语义分割进行研究具有广泛的应用前景。

传统图像语义分割算法需要人工设计特征提取方案,进而建立所提取特征和语义信息之间的映射,对不同类型的分割任务需要重新设计方案,泛化性较差。近年来,随着卷积神经网络(CNN)的发展,多种基于全卷积的对称语义分割算法相继出现,解决了传统语义分割算法的不足。Long等^[6]提出一种全卷积神经网络(FCNN),将网络改编为全卷积网络,并通过微调网络,把学习到的特征表示迁移到分割任务中。FCNN将网络后半部分产生的低分辨率高层语义特征图与前半部分产生的高分辨率低层语义特征图融合,实现了更加准确而详细的图像语义分割。其中FCN-8s在PASCAL VOC 2012数据集上获得62.2%的平均交并比(mIoU),对后来的语义分割算法有一定的借鉴意义。在此基础上,Ronneberger等^[7]提出了一种对称语义分割模型UNet,该模型由一个可以捕获上下文信息的收缩通路和一个能够精确定位分割边界的对称扩张通路组成。UNet使用图像切块方式进行训练,因此训练数据远大于输入图像值,在少量样本的情况下也能获得良好的鲁棒性。UNet在生物医学分割应用上具有较好的性能,在2015年ISBI细胞跟踪挑战赛上赢得了冠军。针对FCNN在语义分割中存在的感受野固定和分割物体细节容易丢失等问题,Badrinarayanan等^[8]提出了一种深度卷积神经网络SegNet用于语义像素分割。SegNet的核心部分是编解码器,在编码过程中提取特征,在解码过程将低分辨率特征图映射到全输入分辨率特征图后,进行像素分类。SegNet的创新之处在于解码器对较低分辨率的输入特征图以池化索引方式进行上采样,训练参数数量较少,在推理期间的存储和计算时间方面都极有优势。

针对基于全卷积的对称语义分割算法忽略像素空间一致性的问题,基于全卷积的扩张卷积语义分割算法纷纷涌现。Chen等^[9]提出了一种新的空洞卷积语义分割模型DeepLabv1,该模型采用VGG-16作为骨干网络,并且引入空洞卷积和全连接条件随机场(CRF),提高了模型捕获精细细节的能力,使得模型能够捕获复杂的对象边界。DeepLabv2^[10]把骨干网络替换成ResNet-101,在DeepLabv1的基础上引入了空洞空间金字塔池化(ASPP)模块。

ASPP以多个采样率和有效视场的滤波器探测输入卷积特征层,从而整合了多尺度特征信息。DeepLabv3^[11]为了实现在多个尺度上分割对象,设计了采用级联或并行的空洞卷积模块,通过采用多种空洞率来捕获多尺度上下文。DeepLabv3+^[12]是通过在DeepLabv3的基础上添加一个简单而有效的解码器模块而来的,ASPP模块整合了多尺度特征信息,解码器结构则通过逐渐恢复空间信息来捕获更清晰的对象边界,使分割结果得到优化。针对深度CNN中重复子采样操作导致初始图像分辨率显著降低的问题,Lin等^[13]提出了一种可以使用各个层级特征的高精度语义分割模型RefineNet,该模型类比UNet中的两段通路,分别由ResNet和RefineNet为基础构成。每个独立的RefineNet模块包括残差卷积单元(RCU)、多分辨率融合(MRF)和链残差池(CRP)3个部分。RCU根据一个自适应块组成卷积集,此卷积集主要用来微调预训练的ResNet权重;MRF将所有路径输入融合成高分辨率特征图;将MRF输出的特征映射输入到CRP块中,目的是从一个大的图像区域捕获背景上下文,有效地汇集具有多个窗口大小的特征,并使用可学习的权重将它们融合在一起。RefineNet模型这种级联结构可以有效结合高级语义和低级特征,使得语义分割结果更为精准,在PASCAL VOC 2012数据集上获得83.4%的mIoU,是语义标注技术发展过程中的一个标志。针对复杂场景解析,Zhao等^[14]提出了一种优越的像素级预测框架PSPNet,PSPNet使用预训练网络ResNet对输入图像进行特征提取,然后利用ASPP整合多尺度特征信息,将金字塔特征表示与原始特征表示拼接起来进行卷积,进而生成最终预测分割图像。PSPNet通过对基于不同区域的上下文聚合来获取全局上下文信息,为像素级场景预测提供了一个出色的框架。

上述方法的分割评价更多聚集在像素分割精度指标上,而移动机器人自主移动、人机交互等任务不仅对分割精度有要求,同时也对模型参数量和算法实时性有要求。因此,本文提出了一种基于改进DeepLabv3+网络的轻量级语义分割算法,将原DeepLabv3+特征提取骨干网络替换成MobileNetv3网络^[15],将ASPP模块中卷积方式替换成深度可分离卷积并引入注意力机制模块,并将批归一化(BN)^[16]替换成组归一化(GN)^[17]。这些

改进降低了模型参数量,使算法更好地平衡了分割精度和速度。

2 DeepLabv3+网络结构设计

2.1 改进的 DeepLabv3+网络

改进的 DeepLabv3+网络结构如图 1 所示,分为编码层和解码层。该网络采用 MobileNetv3 网络代替原 Deeplabv3+特征提取骨干网络并运用

迁移学习^[18]加载预训练模型,在降低模型复杂度的同时提高分割性能;对 ASPP 模块进行改进,把模块中标准卷积替换成深度可分离卷积同时引入注意力机制,通过学习特征通道维度的权重信息,对网络性能好的特征通道权重进行提升,抑制对网络性能不好的特征通道权重,提高训练效率;最后采用组归一化,从而获得了更好的分割性能。

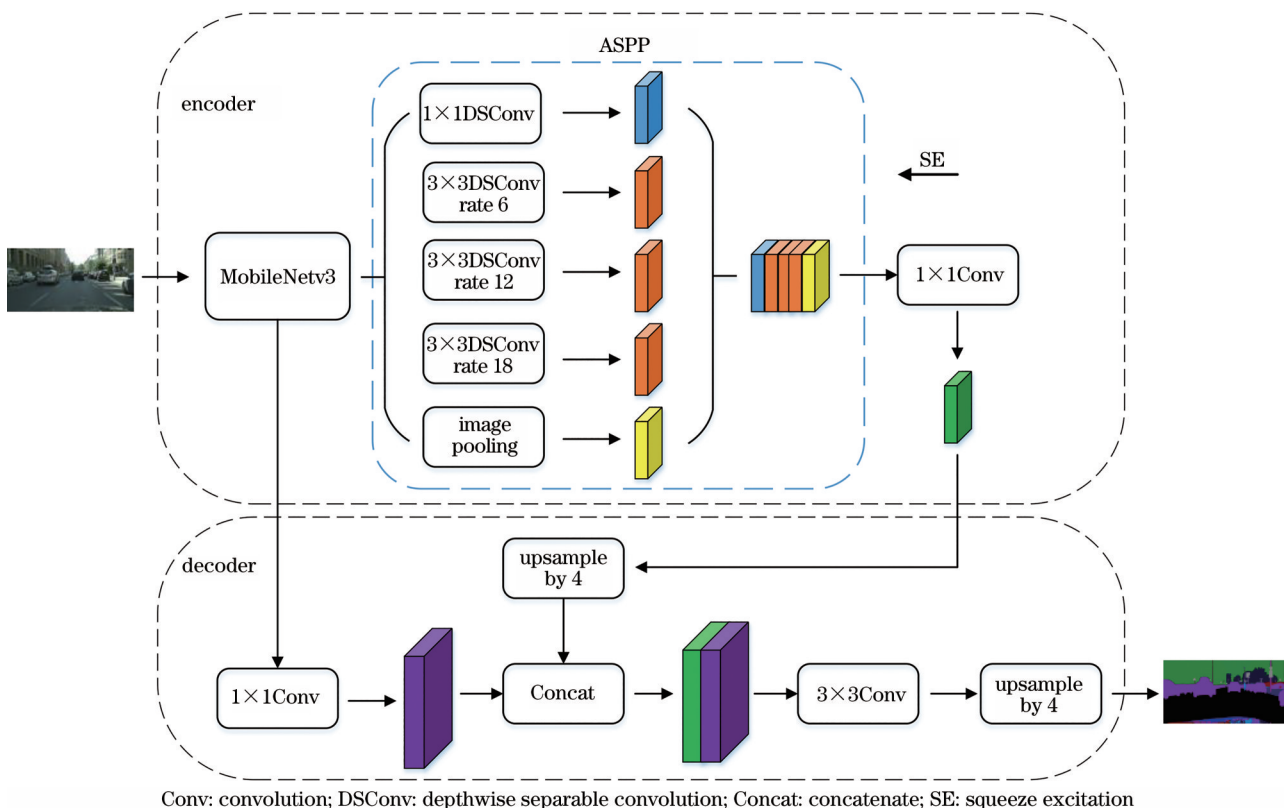


图 1 改进的 DeepLabv3+网络
Fig. 1 Improved DeepLabv3+network

2.2 深度可分离卷积

深度可分离卷积^[19]如图 2 所示,将卷积分为深度卷积和逐点卷积,针对输入特征图的每个通道,分别进行卷积核大小为 3×3 的卷积,输出 3 个通道分离的特征属性,然后对每个通道进行 1×1 卷积得到输出。标准卷积中,网络同时学习输入特征图的空间相关性和通道相关性,通过一步操作即对所有输入进行结合得到新的一系列输出。深度可分离卷积的这种操作,使性能与标准卷积相当的同时,使用更少的参数来进行特征学习,能够有效减少模型参数量。深度可分离卷积的作用主要在图 1 中 ASPP 模块体现。

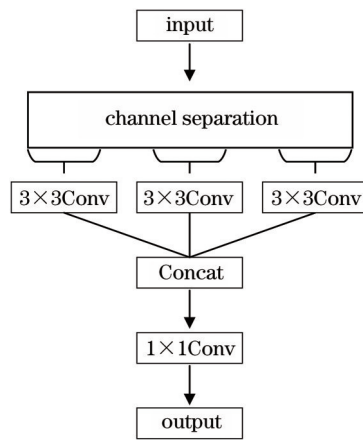


图 2 深度可分离卷积
Fig. 2 Depthwise separable convolution

2.3 SE 注意力机制模块

SE 注意力机制模块^[20]如图 3 所示,SE 模块在特征提取通道维度上引入注意力机制,专注于通道关系,使网络从全局信息出发,选择对网络性能好的特征通道权重进行提升,抑制对网

络性能不好的特征通道权重,即对通道权重进行分配。SE 模块以最小的额外计算成本为现有的深度架构带来了显著的性能提升。SE 注意力机制模块的作用主要在图 1 中 ASPP 模块体现。

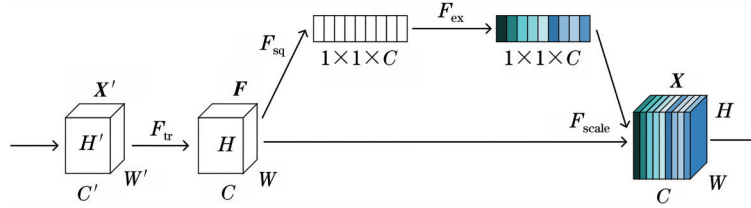


图 3 SE 模块

Fig. 3 Squeeze and excitation module

SE 模块主要包含压缩 (squeeze) 和激励 (excitation) 两种操作,压缩操作输出作为后面激励操作输入,激励操作输出再与原始输入特征相乘得到 SE 模块最后输出。其中压缩操作可表示为

$$z = F_{sq}(f) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W f(i, j), \quad (1)$$

式中: F_{sq} 是压缩操作函数; $f \in \mathbf{R}^{H \times W}$ 是二维特征图集合, $f(i, j)$ 是其中一个元素, H 和 W 分别表示特征图空间信息的高和宽; z 为压缩操作输出。

激励操作可表示为

$$s = F_{ex}(z, W) = \sigma[W_2 \delta(W_1 z)], \quad (2)$$

式中: F_{ex} 是激励操作函数; σ 和 δ 分别表示 Sigmoid 和 ReLU 激活函数; $W_1 \in \mathbf{R}^{r \times c}$, $W_2 \in \mathbf{R}^{c \times r}$, W_1 和 W_2 分别是其中的某个元素, r 是降维系数; s 为激励操作输出。

在激励操作之后,将产生的输出权重与原始输入特征相乘,有

$$x = F_{scale}(f, s) = s \cdot f(i, j), \quad (3)$$

式中: F_{scale} 是 scale 操作; x 是 SE 模块最后输出 \mathbf{X} 中的一个值, $\mathbf{X} = [x_1, x_2, \dots, x_c]$ 。

2.4 归一化方式

深度学习在训练神经网络的时候,参数传递会导致各层的输入分布发生变化,较低的学习率和繁琐的参数初始化减慢了训练速度。针对这一问题,引入归一化并对每个训练小批量进行归一化来汲取其优势,这样在训练深度神经网络的时候就可以使用更高的学习率和随机初始化^[16]。

BN 是深度学习发展中的一项重要技术,使得各种网络都可以进行训练。但是,由于批次统计信息估算不准确,当批尺寸 (batch size) 变小时,BN 的误差会迅速增加,而检测和分割等计算机视觉任务往往需要小批量,因此使用 GN 代替 BN。

BN 和 GN 归一化方式对比如图 4 所示,图中 C 表示特征图通道, N 表示批尺寸。每个图表示一个特征图张量,黑色像素通过相同的均值和方差归一化,该均值和方差是通过汇总这些像素的值而得出的。

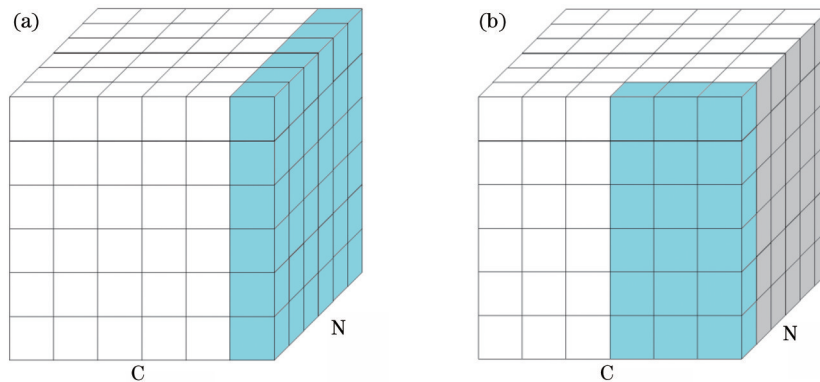


图 4 归一化方式对比。(a) BN; (b) GN

Fig. 4 Comparison of normalization methods. (a) BN; (b) GN

3 实验与分析

3.1 实验设置

实验所用的数据集是 Cityscapes^[21]和 Foggy Cityscapes^[22], Cityscapes 数据集来自 50 个不同城市的街景,共有 19 种语义类别,总共包括 5000 张精细的像素级标注图片,是图像语义分割任务中的重要数据集。Foggy Cityscapes 数据集则在 Cityscapes 的基础上,通过将雾添加到 Cityscapes 中来创建 Foggy Cityscapes 数据集, Foggy Cityscapes 中不同衰减系数 β 值对应不同能见度版本的数据集,实验选用 $\beta=0.01$ 的 Foggy Cityscapes 数据集。实验过程中,用训练集图片训练神经网络,用验证集图片验证改进网络的效果。评价指标为 mIoU 和模型参数量。mIoU 计算所有类别真实值和预测值的交集和并集之比的平均值,用来评价模型的分割精度;模型参数量计算模型总的可训练参数量,用来评价模型的空间复杂度。实验所用的操作系统为 64 位 Windows 10, CPU 为 Intel(R) Core(TM) i7-8700K CPU @ 3.70 GHz, RAM 为 32 GB, 编程软件平台为 python 3.7 和 PaddlePaddle 2.0.1, GPU 为英伟达 Tesla v100, cuDNN 为 7.6.2, 具体实验参数如表 1 所示。

表 1 实验参数

Table 1 Experimental parameters

Parameter name	Parameter selection
Enter picture size	769×769
Loss function	Cross entropy
Optimizer	SGD
Batch size	16
Iteration	18500

3.2 验证实验

为了验证不同特征提取骨干网络在 DeepLabv3+ 模型上的适配性,对配置不同骨干网络的 DeepLabv3+ 在 Cityscapes 验证集上进行了对比实验,评价指标为 mIoU、验证时间和模型参数量,结果如表 2 所示,表 2 中验证时间为平均验证一张图片的时间。

从表中可看出,将 DeepLabv3+ 模型的骨干网络换成 MobileNetv3 后, mIoU 为 72.94%, 验证时间为 15 ms, 模型参数量为 2.18 MB, 平均交并比变化不大,但验证时间和模型参数量有着数量级的减少,这对于部署到移动端极有优势。

表 2 不同骨干网络的 DeepLabv3+ 模型性能对比

Table 2 Comparison of Deeplabv3+ model performance of different backbone networks

Evaluation index	Xception	ResNet-50	MobileNetv3
mIoU / %	78.18	79.20	72.94
Validating time / ms	113	234	15
Parameter quantity / MB	78.53	38.72	2.18

3.3 对比实验

为了验证深度可分离卷积、SE 注意力机制模块和 GN 归一化方法的有效性,设置了相应的对比实验,均在实验环境相同的情况下进行。其中不同卷积方法对比在 Cityscapes 验证集上进行验证,结果如表 3 所示;不同模块性能对比分别在 Cityscapes 和 Foggy Cityscapes 验证集上进行验证,结果如表 4 和表 5 所示。

表 3 不同卷积方法对比

Table 3 Comparison of different convolution methods

Convolution method in ASPP	Training time / h	mIoU / %
Standard convolution	3.39	68.50
Depthwise separable convolution in ASPP	2.56	66.78

表 4 Cityscapes 上不同模块性能结果

Table 4 Performance results for different modules on Cityscape

Depthwise separable convolution in ASPP	SE	GN	mIoU / %
			68.50
✓			66.78
	✓		71.23
✓	✓		70.17
✓	✓	✓	72.94

表 5 Foggy Cityscapes 上不同模块性能结果

Table 5 Performance results for different modules on Foggy Cityscape

Depthwise separable convolution in ASPP	SE	GN	mIoU / %
			54.06
✓			53.42
	✓		57.71
✓	✓		56.31
✓	✓	✓	58.76

从表 3 中可看出,只替换骨干网络的 DeepLabv3+ 模型得到的 mIoU 为 68.50%。ASPP 模块中引入深度可分离卷积后,模型的训练时间降

低了 0.83 h,在对分割精度影响不大的前提下提高了模型的训练效率。

从表 4 中可看出,与原模型相比,仅在 ASPP 模块中引入 SE 注意力机制模块,mIoU 增加了 2.73 个百分点;在 ASPP 模块中同时引入深度可分离卷积和 SE 注意力机制模块后,mIoU 增加了 1.67 个百分点;将模型结构中的 BN 用 GN 代替,mIoU 又增加了 2.77 个百分点。从表 5 中可看出,在 Foggy Cityscapes 数据集上得到的 mIoU 较低,是因为实验选取了 $\beta=0.01$ 的 Foggy Cityscapes。 $\beta=0.01$ 时,相当于能见度为 300 m 的场景,此时对网络的分割效果有一定影响。此对比实验表明所提改进模块有一定的适用性。

在 Cityscapes 验证集上,将所提算法和目前比较流行的一些图像语义分割算法进行了对比,实验结果如表 6 所示。

从表 6 中可看出,对比 SegNet,所提算法在 Cityscapes 验证集上的 mIoU 为 72.94%,mIoU 比 SegNet 提高 14.99 个百分点,对比基于 Xception 和 ResNet-50 的 DeepLabv3+ 模型有一定的差距,但所提算法的模型参数量仅有 2.18 MB,比其他算法少非常多,明显具有更大的优势。因此,所提算法在分割精度和模型参数量之间实现了较好的平衡。

在 Cityscapes 验证集上,将所提算法和其他轻量级图像语义分割算法进行了对比,实验结果如表 7 所示。

表 6 不同算法性能对比

Table 6 Comparison of performance of different algorithms

Algorithm	Backbone network	mIoU / %	Parameter quantity /MB
PSPNet ^[14]	Paper source	78.42	84.75
SegNet ^[8]	Paper source	57.95	29.46
DeepLabv3+	Xception	78.18	78.53
DeepLabv3+	ResNet-50	79.20	38.72
Proposed algorithm	MobileNetv3	72.94	2.18

表 7 轻量级图像语义分割算法对比

Table 7 Comparison of lightweight image semantic segmentation algorithms

Algorithm	mIoU / %	Parameter quantity /MB
FSSNet ^[23]	62.32	0.20
Fast-SCNN ^[24]	69.25	2.33
ENet ^[25]	60.43	0.37
Proposed algorithm	72.94	2.18

从表 7 中可看出,对比其他轻量级图像语义分割算法,所提算法的模型参数量不是最低,但分割精度最高,所提算法参数量比 Fast-SCNN 低 0.15 MB,但 mIoU 比 Fast-SCNN 高 3.69 个百分点,实现了分割精度和模型参数量之间较好的平衡。

3.4 算法分割结果

所提算法在 Cityscapes 验证集上的分割效果如图 5 所示。

从图 5 中的分割效果图可以看出:所提算法对图

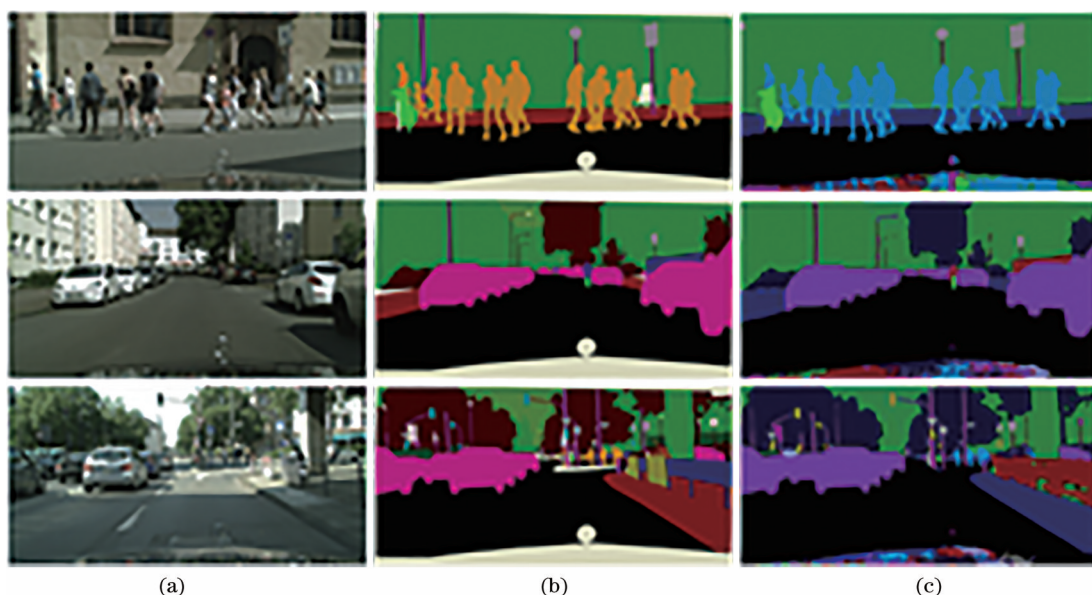


图 5 所提算法分割效果。(a)输入图片;(b)真实标签;(c)分割效果

Fig. 5 Segmentation results of proposed algorithm. (a) Input images; (b) ground truth; (c) segmentation results

片中各类场景能够进行有效的分割,分割结果接近真实场景分割;对于一些远距离和复杂场景,所提算法不能很好地对细小物体或者物体边缘进行分割。

4 结 论

提出了一种基于改进 DeepLabv3+ 网络轻量级语义分割算法,采用 MobileNetv3 代替原 DeepLabv3+ 模型特征提取骨干网络,将 ASPP 模块中的标准卷积替换为深度可分离卷积、引入注意力机制模块和组归一化方法,在极大降低模型参数量和运行时间的同时保证了算法性能,得到较好的分割效果。所提算法在 Cityscapes 验证集上得到了 72.94% 的 mIoU,验证时间为 15 ms,模型参数量仅为 2.18 MB。下一步将研究如何在降低模型参数量的同时进一步提高算法的分割精度,提出更高精度的轻量级语义分割算法。

参 考 文 献

- [1] Wang Z Y, Ni X Y, Shang Z D. Autonomous driving semantic segmentation with convolution neural networks[J]. Optics and Precision Engineering, 2019, 27(11): 2429-2438.
王中宇,倪显扬,尚振东.利用卷积神经网络的自动驾驶场景语义分割[J].光学精密工程,2019,27(11):2429-2438.
- [2] Wang L F, Yan C M. Review on semantic segmentation of road scenes[J]. Laser & Optoelectronics Progress, 2021, 58(12): 120003.
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):120003.
- [3] Huang G, Liu X L. Automatic extraction and classification of road markings based on deep learning [J]. Chinese Journal of Lasers, 2019, 46(8): 0804002.
黄刚,刘先林.基于深度学习的道路标线自动提取与分类方法[J].中国激光,2019,46(8):0804002.
- [4] Yuan L, Yuan J S, Zhang D Z. Remote sensing image classification based on DeepLab-v3+[J]. Laser & Optoelectronics Progress, 2019, 56(15): 152801.
袁立,袁吉收,张德政.基于DeepLab-v3+的遥感影像分类[J].激光与光电子学进展,2019,56(15):152801.
- [5] Wu Y C, Lin L, Wang J X, et al. Application of semantic segmentation based on convolutional neural network in medical images[J]. Journal of Biomedical Engineering, 2020, 37(3): 533-540.
吴玉超,林岚,王婧璇,等.基于卷积神经网络的语义分割在医学图像中的应用[J].生物医学工程学杂志,2020,37(3):533-540.
- [6] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3431-3440.
- [7] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [8] 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.
- [9] 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]. <https://arxiv.org/abs/1412.7062>.
- [10] 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.
- [11] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-06-17) [2021-02-10]. <https://arxiv.org/abs/1706.05587>.
- [12] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 833-851.
- [13] Lin G S, Milan A, Shen C H, et al. RefineNet: multi-path refinement networks for high-resolution semantic segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5168-5177.
- [14] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer

- Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6230-6239.
- [15] Howard A, Sandler M, Chen B, et al. Searching for MobileNetv3[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 1314-1324.
- [16] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [EB/OL]. (2015-05-02) [2019-04-03]. <http://arxiv.org/abs/1502.03167>.
- [17] Wu Y X, He K M. Group normalization[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11217: 3-19.
- [18] Kolesnikov A, Beyer L, Zhai X H, et al. Big transfer (BiT): general visual representation learning [EB/OL]. (2019-12-24) [2021-02-10]. <https://arxiv.org/abs/1912.11370>.
- [19] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1800-1807.
- [20] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [21] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 3213-3223.
- [22] Sakaridis C, Dai D X, Van Gool L. Semantic foggy scene understanding with synthetic data[J]. International Journal of Computer Vision, 2018, 126(9): 973-992.
- [23] Zhang X T, Chen Z X, Wu Q M J, et al. Fast semantic segmentation for scene perception[J]. IEEE Transactions on Industrial Informatics, 2019, 15(2): 1183-1192.
- [24] Poudel R P K, Liwicki S, Cipolla R. Fast-SCNN: fast semantic segmentation network[EB/OL]. (2019-02-12)[2021-02-10]. <https://arxiv.org/abs/1902.04502>.
- [25] Paszke A, Chaurasia A, Kim S, et al. ENet: a deep neural network architecture for real-time semantic segmentation[EB/OL]. (2016-06-07) [2021-02-10]. <https://arxiv.org/abs/1606.02147>.