

DOI:10.16452/j.cnki.sdkjzk.2021.06.010

文章编号:1672-3767(2021)06-0076-09

# 一种基于空洞卷积组合的轻量级语义分割方法

张晓庆<sup>1</sup>, 刘伟科<sup>2</sup>

(1. 山东科技大学 计算机科学与工程学院, 山东 青岛 266590;

2. 山东科技大学 网络安全与信息化办公室, 山东 青岛 266590)

**摘要:**编-解码结构的卷积神经网络是近年来出现的一类高准确率的图像语义分割方法,但是参数量大、对算力要求高的特点,束缚了其在无人驾驶、道路监控、遥感分类以及移动端物体检测等算力有限、实时性强的领域中的应用。针对以上问题,首先设计空洞卷积组合模块—NG-APC模块,通过规范空洞率,在扩大感受野的同时解决空洞卷积中的栅格问题;再利用 NG-APC 模块结合深度可分离卷积搭建编码-解码结构的 NA-U-Net。最后利用该网络,提出一种基于空洞卷积组合的轻量级语义分割方法,在保持较高的分割准确率的同时大幅降低卷积模型的参数量和计算量。通过在公开数据集 Cityscapes 上进行实验,并与经典的 FCN-8s、U-Net 以及轻量级的 ESP-Netv2、Refine Net-LW、LiteSeg 进行对比,验证本方法的有效性。

**关键词:**语义分割;空洞卷积组合;感受野;栅格问题;深度可分离卷积

**中图分类号:**TP30

**文献标志码:**A

## A lightweight semantic segmentation method based on combination of dilated convolution

ZHANG Xiaoqing<sup>1</sup>, LIU Weike<sup>2</sup>

(1. College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao,

Shandong 266590, China; 2. Network Security and Information Office,

Shandong University of Science and Technology, Qingdao, Shandong 266590, China)

**Abstract:** Convolutional neural network with encoding-decoding structure is a kind of high accuracy image semantic segmentation method that has just emerged in recent years. However, the large number of parameters and the high requirements for computing power restrict its application in fields with limited computing power and strong real-time performance, such as unmanned driving, road monitoring, remote sensing classification, mobile object detection and so on. In view of the above problems, this paper proposed a lightweight semantic segmentation method based on dilated convolution combination. Firstly, the dilated convolution combination module NG-APC module was designed by regulating the dilated rate to expand the receptive field and to solve the gridding problem in dilated convolution. Then, the NA-U-Net of encoding-decoding structure was built by using NG-APC module combined with depthwise separable convolution. With the NA-U-Net, a lightweight semantic segmentation method based on the combination of dilated convolution was put forward to maintain high segmentation accuracy and drastically reduce the amount of parameters and computation of convolution model. Extensive experiments were conducted on the Cityscapes public datasets and comparisons were made with classic FCN-8s and U-Net as well as lightweight ESP-Netv2, Refine Net-LW and LiteSeg. The results verified the effectiveness of the proposed method.

**Key words:** semantic segmentation; combination of dilated convolution; receptive field; gridding problem; depth-wise separable convolution

收稿日期:2020-09-26

作者简介:张晓庆(1980—),女,山东泰安人,讲师,博士,研究方向为深度学习与图像识别。

刘伟科(1980—),男,山东临沂人,高级工程师,博士,研究方向为深度学习与大数据处理技术,本文通信作者。

E-mail:lwk@sdust.edu.cn

图像语义分割对自动驾驶<sup>[1]</sup>、机器人传感等技术具有潜在应用价值,通过对每个像素赋予具体意义的语义标签并根据语义标签信息分割图像,是机器视觉中场景理解与分析的一个重要研究课题。近年来,随着计算机软硬件技术的快速发展,以卷积网络为代表的深度学习方法取得了较大的突破。2014年,Long等<sup>[2]</sup>提出利用深度学习方法进行图像语义分割的网络—全卷积神经网络(fully convolutional networks, FCN),随后 SegNet<sup>[3]</sup>、U-Net<sup>[4]</sup>、DeepLab<sup>[5]</sup>系列、ESP-Net<sup>[6]</sup>系列等先进的语义分割网络<sup>[7]</sup>相继被提出。2015年,以U-Net为代表的编-解码结构<sup>[8]</sup>的网络,因为分割准确率高而受到广大学者的重视,但是这类网络体量大、参数多,对算力的要求高。为降低算力,提高网络训练速度,多种轻量级的语义分割网络被提出,如 Refine Net-LW<sup>[9]</sup>、LiteSeg<sup>[10]</sup>、PSPNet<sup>[11]</sup>、ESPNet 系列<sup>[12-13]</sup>,但这类网络的准确率相比非轻量级的网络较低。针对这一问题,本研究在保证准确率的前提下,以提高卷积效率、降低网络参数数量为目标,提出一个级联多分枝的空洞卷积组合 NG-APC 模块,通过规范空洞卷积组合中的空洞率,解决空洞卷积中出现的栅格问题,并结合深度可分离卷积降低网络的参数量,搭建编-解码结构的 NA-U-Net,得到一种高效的轻量级语义分割方法。

## 1 相关工作

### 1.1 XU-Net

以 U-Net 为代表的编-解码的卷积神经网络在分割精确度方面表现出色,其中 U-Net 中的长连接(图 1 中的 Concat 部分)将对应层的特征直接连接,保留了分割结果的边界细节,提高了较难分割的边缘像素的分割准确率。U-Net++<sup>[14]</sup>、Refine-Net、SegNet、Deeplab-v3+<sup>[15]</sup>等语义分割新方法中均同样使用了编-解码结构。然而,编-解码结构网络的缺点是卷积层数多、参数多、训练时间长、优化难,不能满足移动端的轻量级需求。利用深度可分离卷积将卷积层通道间相关性和空间相关性的映射去耦合后分开映射,能达到同样的卷积效果,同时参数量可以大幅降低。Xception<sup>[16]</sup>、MobileNet<sup>[17]</sup>、Shuffle-Net<sup>[18]</sup>、Squeeze-Net<sup>[19]</sup>等轻量级语义分割方法中均使用了深度可分离卷积。据此,本研究将深度可分离卷积融合到 U-Net,提出 XU-Net<sup>[20]</sup>,其网络结构如图 1 所示。

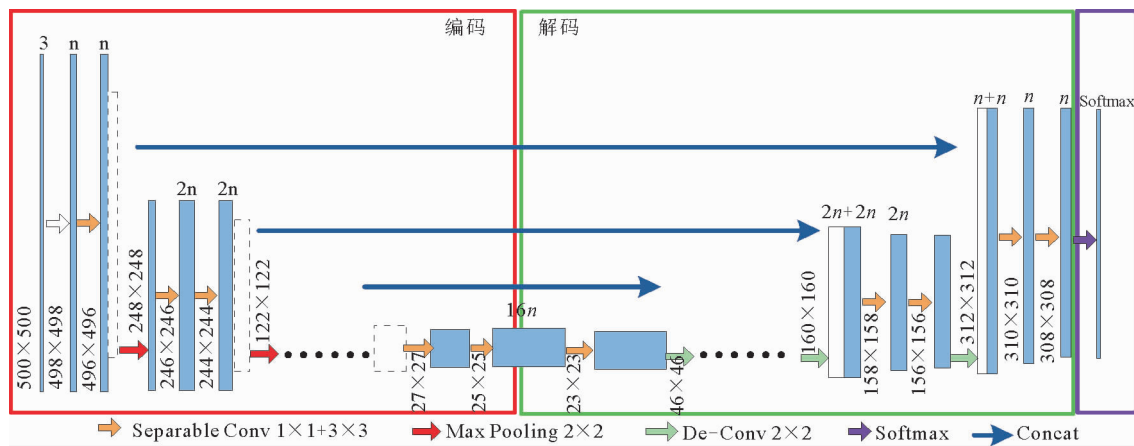


图 1 XU-Net 结构图

Fig. 1 XU-Net structure chart

具体地,将 U-Net 中的标准卷积(如图 2(a))(输入层除外)改进为可分离的两步卷积结构(如图 2(b)),首先对每一个输入通道的特征图做空间卷积,通过卷积核参数值的训练获取一个输入通道的特征组合,然后以通道维为单位利用  $1 \times 1$  的小卷积核做混淆,获取通道维的组合特征。在图像的卷积过程中,输入和输出通道数通常数值较大(通常为  $64 \sim 1024$ ),因此在标准卷积中做全混淆参数量巨大,通过深度可分离卷积,在空间维(输入通道数)和通道维(输出通道数)分别做混淆操作,总参数量比标准卷积大幅降低,参数量降低幅

度的比值为:

$$\frac{M \times n^2 + M \times N \times 1^2}{M \times N \times n^2} = \frac{1}{N} + \frac{1^2}{n^2} \quad (1)$$

其中:  $M$  为输入通道数,  $N$  为输出通道数,  $n^2$  为卷积核尺寸。式(1)中左侧分子表示可分离卷积的参数量, 分母为标准卷积参数量。

## 1.2 空洞卷积组合与栅格问题

空洞卷积<sup>[21]</sup> (atrous convolution) 也称为膨胀卷积(dilated convolution), 是在标准卷积核中注入空洞, 将小的卷积核尺寸变大同时保持卷积的参数量不变, 扩大的幅度称为空洞率(dilation rate)。其优点是在不使用大卷积核、不增加参数量、不增加卷积深度的基础上扩大感受野, 获取更大范围内独立的特征信息, 可以提高对大目标的分割准确率; 缺点是由于卷积核内有空洞会造成采样特征不连续, 形成特征漏采, 即栅格问题(griding problem)<sup>[22]</sup>, 且级联的空洞卷积可能会因为感受野过大而造成不相关目标的特征组合。

为解决栅格问题, 可通过空洞卷积的组合方式: 一是不同空洞率的空洞卷积做级联; 二是在多个分枝中使用不同空洞率的空洞卷积做并联。DeepLabv3<sup>[23]</sup> 中使用多分枝并联的空洞卷积组合, 称为空间金字塔空洞卷积池化层(atrous spatial pyramid pooling, ASPP), 空洞率分别为 6、12、18、24, 同时提出级联的空洞卷积组合, 空洞率分别为 2、4、8、16; 在 ESPNet 中使用多分枝并联的空洞卷积组合, 空洞率分别为 1、2、4、8、16。后来, 发现使用空洞率大的空洞卷积虽然能大幅扩大感受野, 但并不是感受野越大越好, 因此在 ESPNetv2 中将空洞率更新为 1、2、3、4 的多分支并联; HDC 网络<sup>[22]</sup> 为解决栅格问题使用了空洞率为 1、2、3 的空洞卷积级联。

虽然通过级联的空洞卷积组合可解决栅格问题, 但对如何通过最优的空洞卷积组合在解决栅格问题的同时又体现空洞卷积扩大感受野的优势没有深入研究。针对此问题, 本研究提出一种并联和级联结合的空洞卷积组合方式, 并规范了空洞率的组合, 在解决栅格问题的前提下, 达到感受野的最大化。

## 2 NA-U-Net

本研究提出的编-解码结构的神经网络结构为: 在编码部分针对空洞卷积组合结构中的栅格问题, 通过规范空洞率的组合, 得到感受野最大化的级联多分支的空洞卷积组合模块—NG-APC 模块, 通过模块级联构成编码部分; 在解码部分使用深度可分离卷积, 大幅度降低了网络的参数量, 搭建一个轻量级的编码-解码结构的卷积神经网络—NA-U-Net, 并以该网络为基础, 提出一种图像语义分割方法。

### 2.1 NG-APC 模块

空洞卷积组合的级联与多分支并联两种方式在感受野、参数量与栅格问题的解决 3 个方面各有所长:

1) 感受野方面。级联的空洞卷积组合的感受野是根据级联的层数倍数递增, 递增速度快; 多分支并联的空洞卷积组合的感受野是多分支中感受野最大的一个分支。

2) 参数量方面。级联的空洞卷积组合的参数量与级联的标准卷积参数量相同, 是根据级联的层数倍数递增; 由于卷积核总数量不变, 只是分成多个分支, 因此多分支并联的空洞卷积组合的参数量没有增加。

3) 解决栅格问题方面。级联的空洞卷积组合解决栅格问题是通过上一层的特征组合与本层卷积核的参数相连, 层层扩张, 减少特征漏采; 多分支并联的空洞卷积组合解决栅格问题是通过小空洞率的特征补充大空洞率的特征漏采部分, 但若达到完全解决栅格问题, 需要使用连续的空洞率, 即退化成一个标准卷积, 失去了空洞卷积的优势。

基于以上分析, 在解决栅格问题的前提下, 为了达到感受野最大, 同时兼顾参数量, 本研究设计了三层的

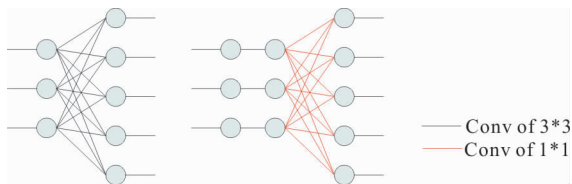


图2 可分离卷积示意图

Fig. 2 Separable convolution diagram

级联与多分支并联结合的空洞卷积组合,模块结构如图 3 所示。

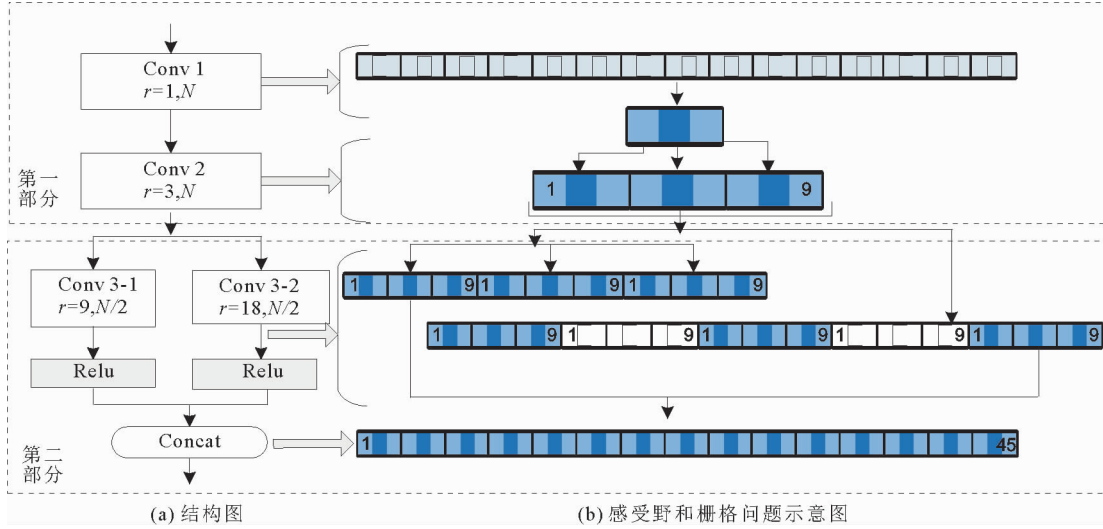


图 3 NG-APC 模块图

Fig. 3 NG-APC module diagram

第一部分为级联部分,为标准卷积后级联空洞率为 3 的空洞卷积,第一层设计为标准卷积,因此如将第一层选用空洞卷积,会造成采样中心点周围特征被漏采,通过级联后,这种漏采现象会加剧,无法做到完全去栅格;第二层设计为空洞率为 3 的空洞卷积,其漏采部分正好由上一层的特征组合补齐,获得感受野为 9 的特征组合。

第二部分的多分支并联是在第一部分的基础上进行分组卷积,每一个采样点可以采样第一部分的感受野为 9 的特征组合,因此设计第一分支的空洞率为 9 可以做到无特征漏采,感受野为 27。为了进一步扩大感受野,第二分支设计为空洞率为 18 的空洞卷积,其漏采部分的特征由第一分支补全,由此可以做到解决栅格问题的最大化感受野的空洞卷积组合,其感受野为 45,感受野和栅格问题如图 3(b)所示。第二部分紧接激活函数和特征拼接,形成一个空洞卷积模块。模块为无栅格的金字塔空洞卷积组合(non-gridding atrous pyramid convolution),称为 NG-APC 模块。

NG-APC 模块在达到感受野最大化的同时,采样点是无重叠的,但对不同的应用场景,并不意味着感受野越大越好,空洞率减小的同时可以提升采样的冗余,采样保持一定的冗余度有助于特征提取。因此模块的层数、分支数、空洞率可以调整,但为了保证去栅格问题的空洞卷积组合,空洞率的取值应不大于式(4)、(5)中的  $r_{\max}$ ,其中 NG-APC 模块中使用的是  $3 \times 3$  的卷积核,卷积步长为 1,感受野的计算公式为:

$$F = (k - 1) \times r + 1, \quad (2)$$

其中:  $F$  为感受野,  $k$  为卷积核尺寸,  $r$  为空洞率。

除了三层二分支的 NG-APC 模块之外,本研究设计了普适性的 NG-APC 模块:

第一部分为空洞卷积级联。第一层为标准卷积,其中  $r_1 = 1$ ,  $F_1 = 3$ ;第二层为空洞卷积,  $r_{2\max} = 3$ 。若级联为  $d$  层,空洞率和感受野的取值分别由公式(3)、(4)得到,即由上一层的感受野确定下一层的空洞率。

$$r_{d\max} = F_{d-1}, \quad (3)$$

$$F_d = F_{d-1} + 2r_d. \quad (4)$$

第二部分为多分支并联的空洞卷积。若为二分支,分支 1 的  $r_{3.1} = 9$ ,  $F_{3.1} = 27$ ;分支 2 的  $r_{3.2} = 18$ ,  $F_{3.2} = 45$ 。若为多个分支并联,分支数为  $b$ ,输出通道数为  $N$ ,则每个分支的卷积核数量为  $N/b$ ,每个分支使用不同的空洞率进行卷积,空洞率如式(5)所示,每个分支卷积后使用 Relu 激活函数,再将多个分支进行特征拼接后输出。

$$\begin{cases} r_{d_{\max}}^i = iF_{d-1}, \\ F_{d_{\max}}^i = F_{d-1} + 2r_{d_{\max}}^i, \end{cases} \quad i=1,2,\dots,b. \quad (5)$$

为了在提高网络获取图像特征能力的同时保证较小的参数量与计算量,根据 2.1 节,将 NG-APC 模块中的标准卷积用深度可分离卷积代替,在保证网络分割能力的同时可以将 NA-U-Net 的参数量由 16.9 M 降至 2.35 M。

### 3.2 NA-U-Net

利用 NG-APC 模块与深度可分离卷积搭建的编-解码结构的 NA-U-Net 的结构如图 4 所示。以该网络为基础,提出一种轻量级的图像语义分割方法。本方法属于利用深度学习方法完成图像语义分割,关键在于利用 NA-U-Net 训练自动获取图像语义特征。在网络训练过程中,将图像送入网络,与标签对比后通过损失函数计算与真值的差距,通过反馈逐层调整网络参数,直至准确率稳定。得到精确到像素的图像语义分割结果。

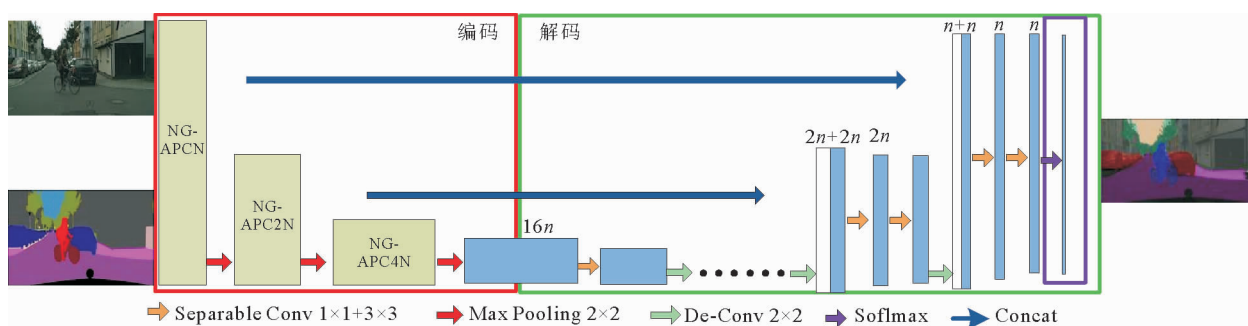


图 4 NA-U-Net 结构图

Fig. 4 Structure diagram of NA-U-Net

实现语义分割的 NA-U-Net 编码部分由多个 NG-APC 模块级联,经过多次实验,使用三层二分支的 NG-APC 模块,根据编码部分下采样过程中卷积核数量倍增的特点,修改了 NG-APC 模块第三层的卷积核数量,由  $N/2$  改为  $N$ ,达到卷积核倍增的结果。经过最大池化下采样后,恰好与编码部分特征通道数量的要求相符。解码部分将逐层编码的结果与上采样的结果连接后,经过逐层解码,使用 XU-Net 模型的解码部分,以降低参数量与计算量,最后经过一层 Softmax 分类器,完成网络的设计与搭建。

## 3 实验验证与分析

为了验证本方法的有效性,使用 Pytorch 开源深度学习库编写代码,利用公开的大型街景数据集—Cityscapes<sup>[24]</sup>进行训练及测试。

### 3.1 实验数据集介绍

Cityscapes 数据集采集了 18 个城市的 5 000 张  $2\,048 \times 1\,024$  像素的精细标注的街景图像,其中 2 975 张作为训练图,500 张为验证图,1 525 张为测试图。数据集中的图像是以汽车驾驶员视角拍摄的道路街景图像,包含行人、道路、汽车等 19 个语义类别,加上背景,共 20 个类别。图 5 为 Cityscapes 数据集中的一幅典型图像,图 5(a)为示例原图,图 5(b)为示例图的语义标签,标签分类如图 5(c)所示。

### 3.2 实验过程及参数设置

首先使用小分辨率的图像( $512 \times 256$  像素)进行训练得到粗略的训练结果,然后提高训练图像的分辨率( $1\,024 \times 512$  像素),训练得到最终分类结果。初始学习率初值为 0.1,逐次递减 5%;激活函数为 Relu;Batch=8;Epoch=300;损失函数使用交叉熵;优化策略使用随机梯度下降法。使用的 GPU 为 NVIDIA GTX 1080。评价标准使用语义分割领域的通用评价指标 mIoU,即图像内各类标签 IoU 的平均值。



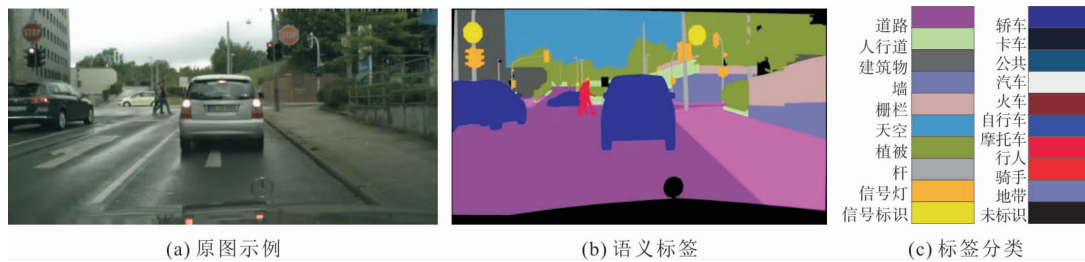


图 5 Cityscapes 数据集示例及分类标签

Fig. 5 Example and classification label of Cityscapes dataset

### 3.3 实验结果及分析

NA-U-Net 在 Cityscapes 数据集上经过 300 次迭代循环,与经典的 FCN-8s<sup>[2]</sup>、改进前的 U-Net<sup>[4]</sup>、先进且轻量级的 ESP-Netv2<sup>[13]</sup> 以及轻量级 LiteSeg<sup>[10]</sup>、RefineNet-LW<sup>[9]</sup> 进行准确率与 mIoU 值对比,如表 1 所示。

表 1 本文方法与其他方法在 Cityscapes 测试集上的准确率与 mIoU 值对比

Tab. 1 Comparison of accuracy and mIoU value cityscapes dataset between the proposed method and other methods

类别	FCN-8s <sup>[2]</sup>	U-Net <sup>[4]</sup>	ESP-Netv2 <sup>[13]</sup>	LiteSeg <sup>[10]</sup>	Refine Net-LW <sup>[9]</sup>	NA-U-Net
道路	0.705	0.832	0.833	0.859	0.839	0.895
人行道	0.304	0.351	0.657	0.891	0.897	0.883
建筑物	0.657	0.699	0.702	0.892	0.832	0.748
墙	0.374	0.495	0.472	0.351	0.312	0.58
栅栏	0.322	0.236	0.405	0.454	0.517	0.511
天空	0.831	0.871	0.852	0.961	0.861	0.932
植被	0.633	0.751	0.823	0.792	0.78	0.852
杆子	0.551	0.633	0.437	0.584	0.511	0.623
信号灯	0.204	0.536	0.517	0.433	0.302	0.721
信号标识	0.518	0.629	0.672	0.713	0.703	0.539
轿车	0.728	0.693	0.732	0.761	0.781	0.792
卡车	0.747	0.815	0.817	0.712	0.821	0.817
公共汽车	0.718	0.781	0.79	0.681	0.813	0.838
火车	0.703	0.576	0.748	0.795	0.638	0.704
自行车	0.589	0.515	0.637	0.577	0.708	0.787
摩托车	0.539	0.692	0.497	0.479	0.511	0.613
行人	0.571	0.717	0.734	0.643	0.673	0.794
骑手	0.445	0.594	0.497	0.478	0.696	0.642
mIoU/%	56.33	63.42	65.7	66.98	67.75	73.73

由表 1 中不同种类物体的分割结果可以看出,各类分割方法对 Cityscapes 测试集中背景类的物体,如天空、植被、墙、道路、轿车等特征变化小或者目标大的物体识别准确率都比较高,但对特征变化大的物体(如自行车、行人等)以及目标小的物体(如信号灯、杆子等)识别准确率差别较大。本方法对特征变化大的物体区分较好,对目标小的物体仅次于 U-Net 模型,但均优于其他轻量级的模型。另外,与轻量级的 ESP-Netv2、LiteSeg 以及 Refine Net-LW 模型相比,本方法的 mIoU 值分别提高了 8.03%、6.75%以及 5.98%,验证了本方法在算法准确率方面的优势。

表2 本文方法与其他方法的参数量、GFLOPs 以及 OA 对比

Tab. 2 Comparison parameters, GFLOPs and OA between the proposed method and other methods

	FCN-8s	U-Net	ESP-Netv2	LiteSeg	Refine Net-LW	NA-U-Net
参数量/MB	133	102	6.79	20.55	46	2.35
GFLOPs/G	181	236	22.6	7.4	52	35.17
OA/%	65.8	86.9	91.6	87.9	91.4	92.5

从表2的参数量、GFLOPs、准确率(OA)对比情况可见,本方法的参数量和 GFLOPs 比改进前的 U-Net 有大幅降低,同时还小于轻量级编-解码结构的 Refine Net-LW、ESP-Netv2 和 LiteSeg 方法;从准确率可以看出,本方法准确率达 92.5%,比改进前的 U-Net 高 5.6%,高于三种编-解码结构的轻量级语义分割方法,是多种对比方法中最好的。

综上,在 Cityscapes 的测试集中,本方法不仅在参数量与计算量方面具有一定的优势,而且分割的准确率与 mIoU 也高于经典的 FCN-8s、U-Net 与轻量级编-解码结构的 Refine Net-LW、ESP-Netv2 和 LiteSeg。

为了更细致地展示分割效果,在 Cityscapes 数据集里选取两张具有代表性的图像 jena\_000110\_000019\_leftImg8bit(用原图 1 表示)和 jena\_000070\_000019\_leftImg8bit(用原图 2 表示),并分别给出 U-Net、ESP-Netv2、LiteSeg、Refine Net-LW 以及 NA-U-Net 的语义分割效果如图 6。原图 1(图 6(a))为街景中的典型图像,包含建筑物、道路、轿车、自行车、骑手、植被等。原图 2(图 6(c))为街景中较少出现的特殊场景,不仅包含典型场景中建筑物、人群、汽车,还包括特殊着装的消防员与特殊车辆消防车,有栅栏、特征复杂的广告墙以及标识牌。通过两幅图像的实验说明本方法在典型图像和非典型图像中的分割效果。

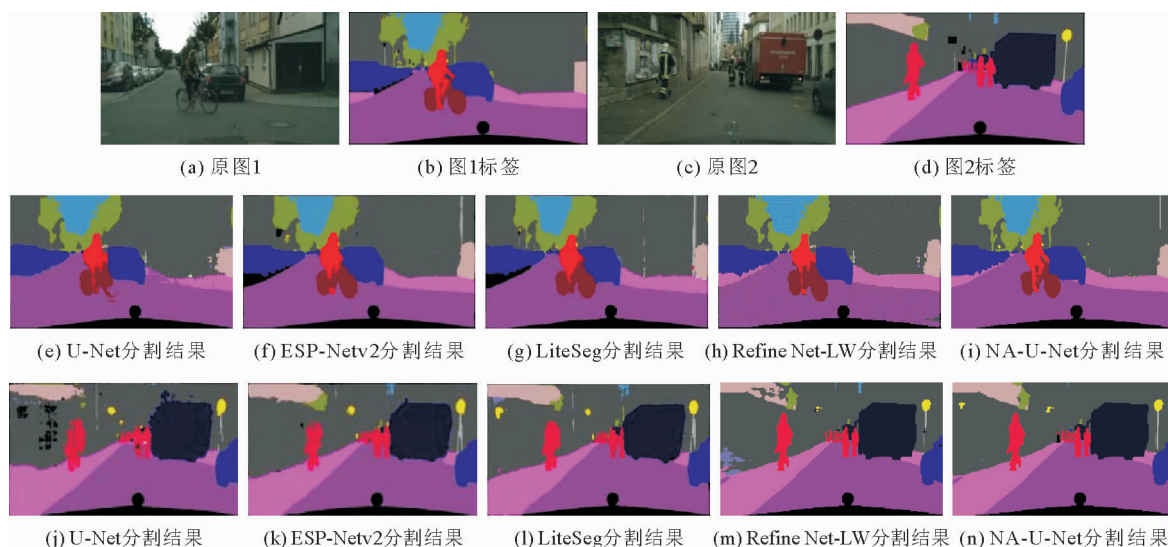


图6 U-Net、ESP-Netv2、LiteSeg、RefineNet-LW 以及 NA-U-Net 方法的分割效果图

Fig. 6 Segmentation results of U-Net、ESP-Netv2、LiteSeg、RefineNet-LW and NA-U-Net

由分割效果图 6(e)~6(n)可以看出:

- 1) 本方法对自行车、骑手、轿车的分割结果明显优于其他方法,尤其是对树木和骑手双脚的轮廓分辨较为清晰;
- 2) 各方法对特殊着装的消防员识别结果差距较大,本方法对近处消防员以及较远处的行人分割准确,对远处 4 个行人的分割精度甚至超过了标签;
- 3) 各方法对杆子、栅栏等分割准确度较低;
- 4) 所有方法均将消防车识别为货车。

产生结果 1)、2)的原因是 NG-APC 模块在扩大感受野的同时,解决了栅格问题,提高了卷积的效率,更好地获取了图像的特征;产生结果 3)的原因是扩大感受野对目标小的物体分割效果不明显;产生结果 4)的原因是标签中没有设置消防车类别,所以所有方法均将消防车识别为货车,说明标签的准确性直接影响分割结果的准确性。选取图像中不同种类的物体边界分割清晰。

表 3 为本文方法在 Cityscapes 数据集上部分数据测试结果的准确率与 mIoU 值的情况,从表 3 可见:分割方法整体的 mIoU 值分别为 73.8%和 67.7%,其中,对道路、建筑物、植被的分割效果较好,但是杆子的识别准确率较低,说明本文方法对易混淆物体的准确率仍有待提高。

表 3 本文方法在 Cityscapes 数据集上部分数据测试结果的准确率与 mIoU 值

Tab. 3 The accuracy and mIoU value of test results with part of data of Cityscaps by this thesis method

	道路	人行道	轿车	建筑物	杆子	信号标识	植被	天空	行人	栅栏	自行车	骑手	mIoU/%
原图 1	0.94	0.86	0.74	0.94	0.09	0.51	0.88	0.89	0	0.61	0.78	0.88	73.8
原图 2	0.95	0.84	0.75	0.91	0.54	0.45	0.65	0.36	0.67	0.65	0	0	67.7

## 4 结论

本研究首先通过规范空洞卷积组合中的空洞率,在解决空洞卷积组合中的栅格问题同时保证感受野最大化,并设计了一种级联多分支的空洞卷积组合模块,然后结合深度可分离卷积搭建了编-解码结构的 NA-U-Net 网络,基于该网络,提出了一种轻量级的图像语义分割方法。通过在 Cityscapes 数据集上进行实验验证,该网络 mIoU 达 73.73%,OA 达 92.5%;参数量仅为 2.35 M,小于多种轻量级卷积神经网络。结果表明本方法是一种轻量且高效的图像语义分割方法。虽然本文方法在参数量与计算量方面有一定幅度的降低,但是仍有下降的空间,将进一步进行网络轻量级方面的研究。

### 参考文献:

- [1]李云鹏,侯凌燕,王超.基于 YOLOv3 的自动驾驶中运动目标检测[J].计算机工程与设计,2019,40(4):1139-1144.  
LI Yunpeng, HOU Lingyan, WANG Chao. Moving objects detection in automatic driving based on YOLOv3[J]. Computer Engineering and Design, 2019, 40(4): 1139-1144.
- [2]LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.
- [3]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.
- [4]RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015: 234-241.
- [5]LIANG C, GEORGE P, IASONAS K, 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.
- [6]MEHTA S, RASTEGARI M, CASPI A, et al. ESPNet: Efficient spatial pyramid of dilated convolutions for semantic segmentation[C]// European Conference on Computer Vision. Springer Cham, 2018: 561-580.
- [7]田莹,王亮,丁琪.基于深度学习的图像语义分割方法综述[J].软件学报,2019,30(2):440-468.  
TIAN Xuan, WANG Liang, DING Qi. Review of image semantic segmentation based on deep learning[J]. Journal of Software, 2019, 30(2): 440-468.
- [8]韩慧慧,李帷涛,王建平,等.编码-解码结构的语义分割[J].中国图像图形学报,2020,25(2):255-266.  
HAN Huihui, LI Weitao, WANG Jianping, et al. Semantic segmentation of encoder-decoder structure[J]. Journal of Image and Graphics, 2020, 25(2): 255-266.
- [9]NEKRASOV V, SHEN C, REID I. Light-weight RefineNet for real-time semantic segmentation[C]// BMVC, England, New-



castle upon Tyne, Sep. 3-6, 2018; 1-15.

- [10] EMARA T, MUNIM HEAE, ABBAS H M. LiteSeg: A novel lightweight convnet for semantic segmentation[C]// 2019 Digital Image Computing: Techniques and Applications (DICTA), Perth, Australia, 2019; 1-7.
- [11] ZHAO H S, SHI J P, QI X, et al. Pyramid scene parsing network[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017; 6230-6239.
- [12] MEHTA S, RASTEGARI M, CASPI A, et al. ESPNet: Efficient spatial pyramid of dilated convolutions for semantic segmentation[C]// ECCV, Munich, Germany, 2018; 561-580.
- [13] MEHTA S, RASTEGARI M, SHAPIRO L, et al. ESPNetv2: A light-weight, power efficient, and general purpose convolutional neural network[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019; 9182-9192.
- [14] ZHOU Z, MAHFUZUR M, TAJBAKHS N, et al. UNet++: A nested U-Net architecture for medical image segmentation[C]// DLMIA, Granada, Spain, 2018; 3-11.
- [15] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]// ECCV 2018, Munich, Germany, 2018; 833-851.
- [16] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017; 1800-1807.
- [17] WANG Y, YAN J, SU Q, et al. A mobileNets convolutional neural network for GIS partial discharge pattern recognition in the ubiquitous power Internet of Things context: Optimization, Comparison, and Application[J]. IEEE Access, 2019, 7: 150226-150236.
- [18] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition, US, Salt Lake City, 2018; 6848-6856.
- [19] LIU Y, LI Z Y, CHEN X, et al. Improving the accuracy of SqueezeNet with negligible extra computational cost[C]// HPBD & IS, Shenzhen, China, 2020; 1-6.
- [20] ZHANG X Q, ZHENG Y G, LIU W K, et al. An improved architecture for urban building extraction based on depthwise separable convolution[J]. Journal of Intelligent and Fuzzy Systems, 2020, 38(11); 1-9.
- [21] ZHANG Q, CUI Z, NIU X, et al. Image segmentation with pyramid dilated convolution based on ResNet and U-Net[C]// International Conference on Neural Information Processing. Springer, Cham, 2017; 364-372.
- [22] WANG P, CHEN P F, YUAN Y, et al. Understanding convolution for semantic segmentation[C]// 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, 2018; 1451-1460.
- [23] Zhang X, Zhang C, Dai C, et al. High resolution remote sensing image classification based on DeepLabv3 architecture[J]. Hydrographic Surveying and Charting, 2019, 39(2); 40-44.
- [24] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding[C]// CVPR, Las Vegas, USA, 2016; 3213-3223.

(责任编辑:傅 游)