基于多级特征图联合上采样的实时语义分割

宋 宇,王小瑀,梁 超,程 超
(长春工业大学 计算机科学与工程学院,吉林 长春 130012)

摘 要:视觉感知是无人驾驶技术中的重要一环,而语义分割技术又是实现视觉感知的主要技术手段之一。现在的语义分割技术多采用计算量大、内存占用高的空洞卷积来提取高分辨率特征图,从而导致现在主流的语义分割网络分割速度不足,无法有效应用于无人驾驶的场景中。针对这一问题,提出了一种实时性更好的语义分割网络。首先,采用了一种轻量级的卷积神经网络作为编码器,并且使用跨步卷积和常规卷积替换了耗时、耗内存的空洞卷积。然后,为了得到与 DeepLabv v3+相似的特征图,提出了一种新的联合上采样模块:多级特征图联合上采样模块(multi-scale feature map joint pyramid upsamping,MJPU),通过融合编码器的多个特征图,生成了语义信息更加丰富的高分辨率特征图。通过 Cityscapes 数据集上的实验表明,相比于主流语义分割网络 Deeplabv3+,该网络在不损失大量性能的前提下,可以将分割速度提高 2.25 倍,达到 32.3 FPS/s。从而使网络具有更好的实时性,更加适合应用于无人驾驶场景。

关键词:无人驾驶;语义分割;卷积神经网络;深度学习;空洞卷积
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2022)02-0082-06
doi:10.3969/j.issn.1673-629X.2022.02.013

Real-time Semantic Segmentation Based on Multi-scale Feature Map Joint Pyramid Upsamping

SONG Yu,WANG Xiao-yu,LIANG Chao,CHENG Chao
(School of Computer Science and Engineering,Changchun University of Technology,Changchun 130012,China)

Abstract:Vision-based perception is an import link in driverless technology,and semantic segmentation is one of the main technique to realize visual perception in driverless technology. The current semantic segmentation technology mostly uses atrous convolution with a large amount of computation and high memory consumption to extract high-resolution feature maps. As a result,the current mainstream semantic segmentation network lacks the segmentation speed and cannot be effectively applied in driverless technology. To solve this problem,a semantic segmentation network with better real-time performance is proposed. Firstly,a lightweight convolutional neural network is used as the encoder,and stride convolution and regular convolution are used to replace the time-consuming and memory-consuming atrous convolution. Secondly,in order to obtain the feature map similar to Deeplabv V3 +,a new joint upsampling module,multi-scale feature map joint pyramid upsamping,is proposed. By fusing multiple feature maps in the encoder,a high resolution feature map with richer semantic information is generated. Experiments on the Cityscapes dataset show that compared with the popular semantic segmentation network Deeplab V3+,the proposed network can improve the segmentation speed by 2.25 times to 32.3 FPS/s without losing a lot of performance. Therefore,the network proposed has better real-time performance and is more suitable for driverless scenes.

Key words:driverless;semantic segmentation;convolution neural network;deep learning;atrous convolution

1 概 述

近年来,对于无人驾驶视觉感知系统,大多使用语义分割技术来处理感知到的物体[1-4]。因此语义分割在无人驾驶领域有着极其重要的作用。由于无人驾驶的特殊性,使其不仅对语义分割网络的准确度有要求,对实时性的需求也非常迫切。

Long 等人[5]提出了最原始的语义分割网络 FCN。继 FCN 之后应用于无人驾驶的语义分割算法总体来说可以分为两大类:第一类是基于编码器-解码器结构的网络,例如 Ronneberger 等人[6]提出的 Unet 网络,

收稿日期:2021-03-16 **修回日期**:2021-07-16
基金项目:吉林省科技发展计划技术攻关项目(20200401127GX);吉林省科技发展计划重点研发项目(20200403037SF);吉林省发改委项目(2019C040-3)
作者简介:宋 宇(1969-),男,硕导,教授,研究方向为人工智能、嵌入式系统;通讯作者:梁 超(1980-),男,硕士,讲师,研究方向为图像处理与视频分析。

在进行少类别分割任务时速度快、精确高。但是当分割类别增多,网络分割速度将大幅度降低;剑桥大学提出的 SegNet[7],网络中采用最大池化索引指导上采样存在特征图稀疏问题,进而导致算法虽然达到了实时分割速度,但分割精度低;Paszke 等人[8]提出的 Enet 网络通过减少神经元权重数量以及网络体积使得网络可以达到实时分割的要求,但是算法的拟合能力弱导致分割精度较低。第二类是基于上下文信息的网络,例如 Zhao H 等人[9]提出的 PSPNet 网络,通过引入更多上下文信息提高了网络的场景解析能力,但是由于多次下采样操作导致特征图丢失大部分空间信息。针对这个问题,Zhang H 等人[10]提出了 EncNet,但是由于采用 Resnet101[11]网络作为主干,参数量庞大,网络实时性较差;Chen 等人[12]提出的 DeepLab 网络,引入了空洞卷积来保持感受野不变。并且后续又提出了 DeepLab v2[13] 以及 DeepLab v3 +[14] 网络,其中 DeepLab v3+结合了两大类网络的优点,使用 DeepLab v3 作为编码器并且在最终特征图的顶部采用了空洞金字塔池化模块(atrous convolution spatial pyramid pooling,ASPP),在避免下采样操作的同时获取了多感受野信息。但是网络在分割速度方面存在不足,主要是因为引入空洞卷积带来了大量的计算复杂度和内存占用。以 Restnet-101 为例,空洞卷积的引用使得其中 23 个残差模块,需要占用 4 倍的计算资源和内存,最后 3 个残差模块需要占用 16 倍以上的资源。

针对以上各种网络无法同时兼顾准确度以及实时性的问题,该文提出了一种实时语义分割网络,采用参数量较少的轻量卷积网络 FCN8s 代替 DeepLab v3+中的 ResNet101 作为网络的主干。文中算法主干与 DeepLab v3+的区别在于最后两个卷积阶段。以第四个卷积阶段(Conv4)为例。在 DeepLab v3+中首先对输入图片进行卷积处理,然后再进行一系列的空洞卷积处理。不同的是,文中方法首先使用跨步卷积来处理输入的特征图,然后使用几个常规卷积来生成输出特征图。并且使用多级特征图联合上采样模块(multi-scale feature map joint pyramid upsamping,MJPU)来代替 DeepLab v3+中耗时、耗内存的空洞卷积,大大减少了整个分割框架的计算时间和内存占用。最重要的是,MJPU 在大幅度减少运算量的同时,不会造成性能上的损失,让算法应用在无人驾驶实时语义分割场景中变得可行。

2　文中算法

为了获得高分辨率的最终特征图,DeepLab 网络中的方法是将 FCN 最后两个下采样操作删除,这两个操作由于扩大了特征图的感受野而带来了大量的计算复杂度以及内存占用量。该文的目标是寻找一种替代方法来近似最终的特征图。

为了实现这一目标,首先将所有被 DeepLab v3+删除的跨步卷积全部复原,之后用普通的卷积层替换掉空洞卷积。如图 1 所示,文中方法主干与原始的 FCN 相同,其中五个特征图(Conv1-Conv5)的空间分辨率逐渐降低 2 倍。

为了得到与 DeepLab v3+相似的特征图,提出了一个新的模块,叫做多级特征图联合上采样模块(MJPU),它以编码器最后三个特征图(Conv3-Conv5)为输入,然后使用一个改进的多尺度上下文模块(ASPP)来产生最终的预测结果。在算法执行过程中,输入图片的格式为 $H \times W \times 3$,经过文中设计的编码器网络,编码器网络由轻量级网络构成,可以减少算法在编码阶段的计算时间;其次使用 MJPU 模块来生成一个特征图,该特征图的作用类似于 Deeplab v3+主干网络中最后一个特征图的激活作用。MJPU 的使用避免了 DeepLab v3+中参数量庞大的空洞金字塔池化网络与高分辨率的最终特征图做卷积运算而大幅度降低了分割速度。MJPU 模块是该文可以大幅度增加实时性的重要因素。最后网络经过 ASPP 获取不同大小的感受野信息,增加网络对不同尺度物体的分割能力,提升算法的分割精度。下面将详细介绍替换空洞卷积的方法以及多级特征图联合上采样模块的结构。

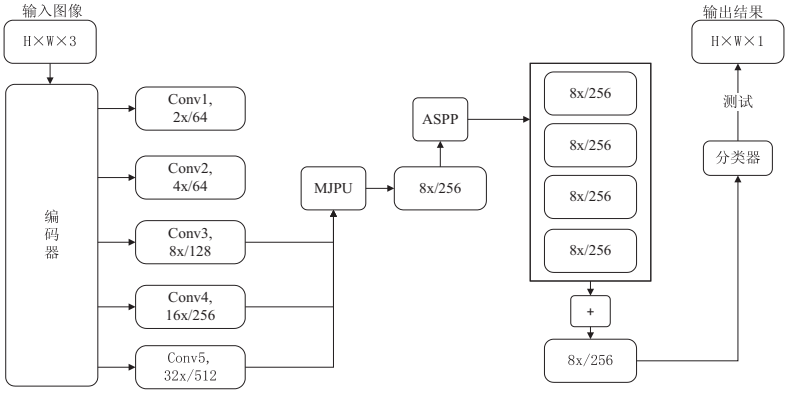
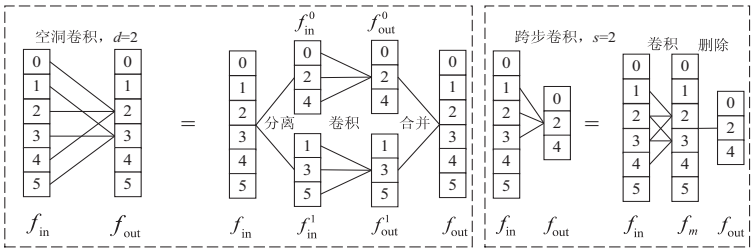


图 1　网络执行流程

2.1 空洞卷积的替代

在 DeepLab 中引入了空洞卷积(atrous convolution),在保持感受野的同时获得了高分辨率的特征图。图 2(a)给出了一维时的空洞卷积(atrous rate=2),具体可以分为以下三个步骤:(1)根据索引的奇偶性,将输入特征 f_{in} 分成 f_{in}^0 和 f_{in}^1 两组;(2)使用相同的卷积层对每组特征进行处理,得到 f_{out}^0 和 f_{out}^1;(3)将生成的两组特征进行交叉融合,得到输出特征 f_{out}。

跨步卷积被用来将输入特征转化为空间分辨率更低的输出特征[15],这相当于图 2(b)所示的两个步骤:(1)对输入特征 f_{in} 做普通卷积,得到中间特征 f_m;(2)删除索引为奇数的元素,得到 f_{out}。



(a)空洞卷积(d=2) (b)跨步卷积(s=2)

图 2 一维空洞卷积与跨步卷积示意图

形式上,给定输入特征图的 x,DeepLab v3+网络中得到输出特征图的 y_d 如下:

$$\begin{aligned} y_d = x \to C_r \to \underbrace{C_d \to \cdots\cdots \to C_d}_{n} = \\ x \to C_r \to \underbrace{SC_rM \to \cdots\cdots \to SC_rM}_{n} = \\ x \to C_r \to S \to \underbrace{C_r \to \cdots\cdots \to C_r}_{n} \to M = \\ y_m \to S \to C_r^n \to M = \\ \{y_m^0, y_m^1\} \to C_r^n \to M \end{aligned} \tag{1}$$

而在文中方法中,生成的输出特征图 y_s 如下:

$$\begin{aligned} y_s = x \to C_s \to \underbrace{C_r \to \cdots\cdots \to C_r}_{n} = \\ x \to C_r \to R \to \underbrace{C_r \to \cdots\cdots \to C_r}_{n} = \\ y_m \to R \to C_r^n = y_m^0 \to C_r^n \end{aligned} \tag{2}$$

其中,C_r 代表普通卷积,C_d 代表空洞卷积,C_s 代表跨步卷积。S、M、R 分别代表图 2 中的分离、合并、删除操作。相邻的 S 和 M 操作是可以相互抵消的。为了简单起见,上述两个方程中的卷积是一维的,对于二维卷积可以得到类似的结果。

2.2 多级特征图联合上采样模块

在给定低分辨率目标图像和高分辨率的指导图像的情况下,联合上采样旨在通过传递来自指导图像(高分辨率)的细节和结构来生成高分辨率的目标图像[16]。通常,低分辨率目标图像 y_l 是通过对低分辨率特征图 x_l 进行 $f(\cdot)$ 变换生成的,即 $y_l = f(x_l)$。对于所给出的 x_l 和 y_l,需要获得一个计算复杂度远远低于 $f(\cdot)$ 的转换函数 $\hat{f}(\cdot)$ 来近似 $f(\cdot)$。例如,如果 $f(\cdot)$ 是一个多层感知器(MLP),那么 $\hat{f}(\cdot)$ 可以简化为一个线性变换。而高分辨率目标图像 y_h 可以通过应用 $\hat{f}(\cdot)$ 指导高分辨率特征图 x_h 图像,即 $y_h = \hat{f}(x_h)$。形式上,给出 x_l,y_l 和 x_h 的联合上采样定义如下:

$$y_h = \hat{f}(x_h),\ \hat{f}(\cdot) = \underset{h(\cdot)\in H}{\operatorname{argmin}} \| y_l - h(x_l) \| \tag{3}$$

其中,H 是所有可能的变换函数集合,$\|\cdot\|$ 是一个给定的距离度量。

根据上述几个方程可知,在输入不同的 y_m^0 和 y_m 时,y_s 和 y_d 可以通过相同的函数 C_r^n 获得,而且 y_s 还是 y_d 的下采样。因此给定 x 和 y_s,可以得到 y_d 的近似特征图 y:

$$\begin{aligned} y = \{y_m^0, y_m^1\} \to \hat{h} \to M \\ \hat{h} = \underset{h\in H}{\operatorname{argmin}} \| y_s - h(y_m^0) \| \\ y_m = x \to C_r \end{aligned} \tag{4}$$

方程 4 可以看作是一个使用梯度下降法的优化问题。因此文中使用 CNN 模块来近似这个优化过程。为了达到这个目的,首先需要生成由 x 给定的 y_m,然后需要收集来自 y_m^0 和 y_s 的特征以学习映射 $\hat{h}$,最后使用卷积模块将特征转换为最终的预测 y。

根据上述分析,设计了如图 3 所示的 MJPU 模块。

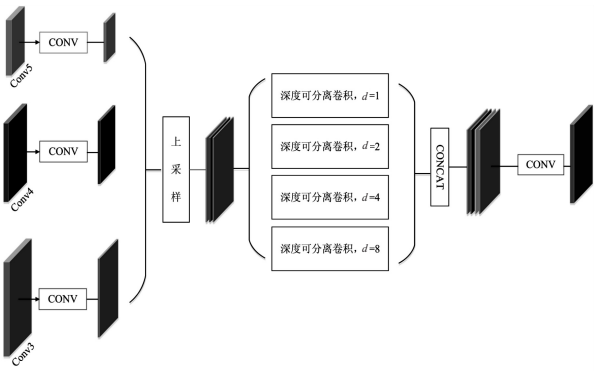


图 3 MJPU 模块

在 MJPU 模块中,每个输入的特征映射先由一个

卷积进行处理。该卷积块的作用是:(1)根据特征图 x 生成 y_m ;(2)将 f_m 放入一个更低维度的空间。这样就将所有的输入特征映射到了同一个空间中,实现了更好的特征融合,并且降低了 y_c 的计算复杂度。之后对生成的特征图进行上采样操作和拼接操作得到 y_c 。采用 4 个不同膨胀率(rate=1,2,4,8)的深度可分离卷积[17]提取的特征。不同的膨胀率具有不同的功能。膨胀率为 1 的卷积用来捕获 y_m^0 与 y_m 之间的关系。膨胀率为 2,4,8 的卷积用来学习将 y_m^0 转换成 y_s 的映射 $\hat{h}$,如图 4 所示。因此,MJPU 可以从多级特征映射中提取多尺度上下文信息,从而获得更好的性能。经过深度可分离卷积提取的特征编码 y_m^0 与 y_s 的映射以及 y_m^0 与 y_m 其他部分的关系。因此,还需要应用一个卷积层,将特征转换为最终的预测结果。

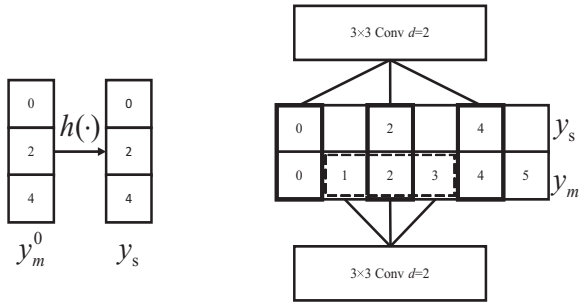


图 4　深度可分率卷积作用示意图

3　实验与分析

3.1　评价指标

实验主要是在准确率以及速度两方面对网络进行了评价。在准确率方面使用的是像素精度(PixAcc)以及平均交并比(mIoU)作为评价指标。在速度方面则使用的是每秒处理帧数(FPS)作为评价指标。

PixAcc 是语义分割中正确分割像素占全部像素的比值,而 mIoU 指的是真实分割与预测分割之间重合的比例,其计算公式如下:

$$\mathrm{mIoU}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{\mathrm{TP}}{\mathrm{FN+FP-TP}} \tag{5}$$

式中, k 表示类别数量,文中为 34 类。TP 表示正类判断为正类的数量,FP 表示负类判断为正类的数量,FN 表示正类判断为负类的数量。

FPS 作为常见的测量网络速度的评价指标,计算公式如下:

$$\mathrm{FPS}=\frac{N}{\sum_{j}^{N}T_j} \tag{6}$$

其中, N 表示处理图像数量, T_j 表示处理第 j 张图像所用的时间。

3.2　实验环境

语义分割模型使用 TensorFlow2.0 深度学习网络框架搭建,在训练和测试阶段的服务器配置均具有英特尔 Core i9-9900K 5.0 GHz 的 CPU、32 G DDR4 2 666 MHz 的内存和 RTX-2080TI(具有 11 GB 显存)的 GPU,并且基于 Window10 的操作系统,在 CUDA10.0 架构平台上进行并行计算,并调用 CuDNN7.6.5 进行加速运算。

在训练过程中网络采用 Adam 优化器,初始学习率是 0.001,学习率策略为逆时间衰减策略,权重衰减使用 L2 正则化。其中 decay_steps=74 300、decay_rate=0.5,代表每过 100 个 epoch,学习率衰减为原来的三分之二。图 5 是网络应用此种学习策略的 loss 值衰减曲线。可以清楚看出逆时间衰减策略可以使模型在较少的 epoch 次数内达到全局最优。

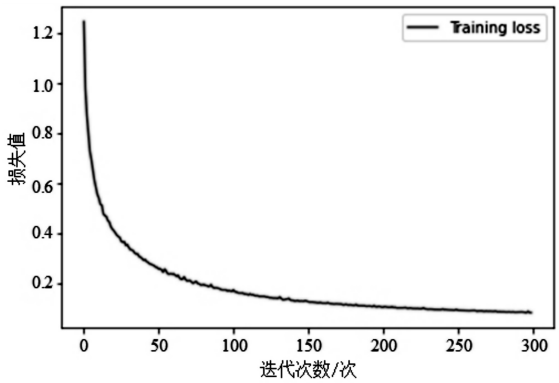


图 5　网络训练过程中 loss 值展示

3.3　数据集

文中选择的无人驾驶数据集为国际公开的由奔驰公司推动发布的数据集 Cityscapes[18]。Cityscapes 是在无人驾驶环境语义分割中使用最广泛的一个数据集。它包含了 50 个城市的不同场景、背景、季节的街景图片,具有 5 000 张精细标注的图像、20 000 张粗标注的图像。在实验过程中,文中只使用了 5 000 张精细标注的图像,将其划分为训练集、验证集和测试集。分别 2 975 张、500 张、1 525 张图像,并且使用了全部 34 类物体作为分割对象。由于原有图像分辨率为 2 048×1 024,分辨率过高导致硬件无法进行大批量训练,因此对图像进行缩放并裁剪成 512×512 大小。

3.4　实验结果分析

为了验证该网络的分割性能,选取了六种网络与文中算法做对比,选取的网络分别为 Unet、SegNet、ENet、PSPNet、EncNet、DeepLab v3+。算法性能对比见表 1。

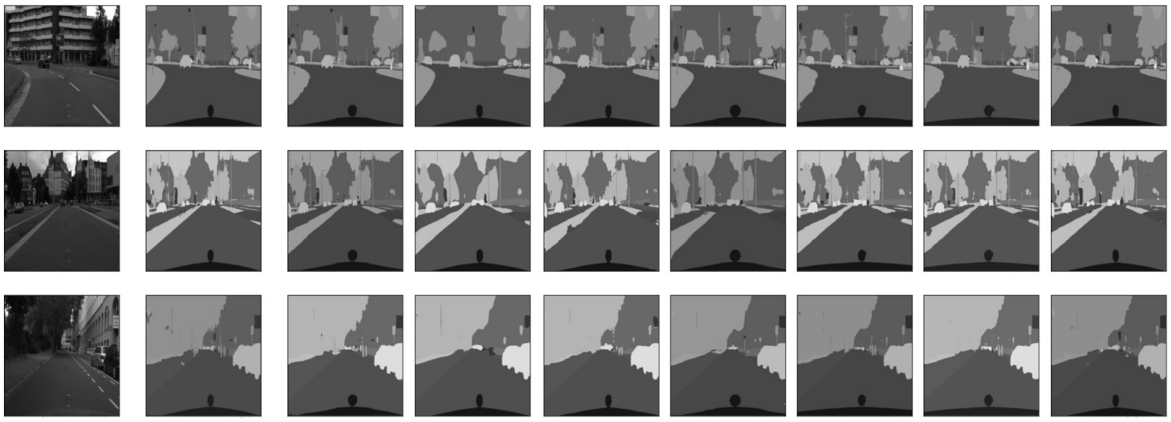
首先与表 1 中的前三种轻量级网络进行对比,文中网络的 mIoU 分别高出 6.72%、10.03% 和 13.32%,PixAcc 也有平均 5% 以上的提升。在网络实时性方面虽然略低于 Enet,但这是由于 Enet 通过减少神经元权重数量的结果,虽然 Enet 的实时性较好但是算法严重

表 1 各算法在 Cityscapes 数据集(val)上的不同评估指标对比

算法	主干网络	PixAcc/%	mIoU/%	FPS/(帧/s)
Unet	VGG16	87.07	37.06	16.6
SegNet	VGG16	85.48	33.75	24.7
Enet	Seratch	85.75	30.46	37.8
PSPNet	Resnet101	89.24	41.65	11.2
EncNet	Resnet101	92.68	45.65	13.6
Deeplab v3+	Resnet101	93.24	44.76	14.3
文中算法	FCN8s	91.85	43.78	32.3

非线性,网络分割精度较低。而对比以分割精度为主要指标的 PSPNet、EnecNet 和 DeepLab v3+,由于文中主干网络采用的是轻量级网络,导致网络精确度略微落后,但是在分割速度方面最多高出 200%。

图 6 展示了不同算法的分割结果。从图 6 中可以看出,文中算法对楼房、道路、树、汽车、天空等都具有较好的分割效果,对于这些大物体该网络均未产生分类错误区域;从图中还可以看出该网络对路灯等杆状物体以及远处车辆等小目标分割效果良好,这主要受益于所提出的 MJPU 模块结合了多层特征图的语义信息。



(a)输入图像 (b)真实标签 (c)Unet (d)Segnet (e)Enet (f)PSPNet (g)EncNet (h)DeepLab (i)文中算法

图 6 不同算法在 Cityscapes 数据集上语义分割效果

3.5 网络结构分析

为了进一步验证 MJPU 的有效性,该文将其与经典的双线性插值上采样和特征金字塔网络(FPN)[19]进行了对比实验。使用 FPS 作为评价指标,在 GPU 上以 512×512 图像作为输入进行测量。结果如表 2 和表 3 所示。对于 ResNet-50,文中方法的测试速度大约是 Encoding 结构(EncNet)的两倍。当主干更改为 ResNet101 时,文中方法的检测速度比 Encoding 结构的快三倍以上。并且可以由图看出文中方法的检测速度可以和 FPN 相媲美。但是对于 FPN 来讲,MJPU 模块可以获得更好的性能。因此对于 DeepLabv3+(ASPP)和 PSP,文中提出的模块可以在提高性能的同时,对网络进行一定程度的加速。

表 2 Resnet-50 中不同上采样方式计算复杂度对比

主干网络	预测结构	上采样方式	FPS
Resnet-50	Encoding	None	18.45
		Bilinear	43.65
		FPN	37.52
		MJPU(our)	38.15
	ASPP	None	15.24
		MJPU(our)	20.46
	PSP	None	18.47
		MJPU(our)	28.58

表 3 Resnet-101 中不同上采样方式计算复杂度对比

主干网络	预测结构	上采样方式	FPS
Resnet-101	Encoding	None	10.15
		Bilinear	35.84
		FPN	32.54
		MJPU(our)	34.08
	ASPP	None	10.42
		MJPU(our)	17.95
	PSP	None	11.68
		MJPU(our)	23.74

4 结束语

为了使语义分割网络更加满足无人驾驶实时分割任务的需求,提出了一种新的实时语义分割网络。首先,采用了一种轻量级的卷积神经网络作为编码器。并且分析了空洞卷积和跨步卷积的区别和联系,使用跨步卷积和普通卷积的组合代替了耗时、耗内存的空洞卷积。在此基础上,将高分辨率特征图的提取问题转化为一种联合上采样问题,提出了一种新的多级特征图联合上采样模块,通过该模块可以在获得近似与 DeepLab v3+相似的特征图的前提下,将网络计算复杂

度最多降低三倍以上。通过在 Cityscapes 数据集上的实验表明(mIou = 43.78%,FPS = 32.3),所提出的实时分割算法在大幅度降低计算复杂度的同时,取得了较好的分割效果。从而使该网络更加适合应用于无人驾驶场景当中。

参考文献:

[1] 马书浩,安居白,于　博. 改进 DeepLabv2 的实时图像语义分割算法[J]. 计算机工程与应用,2020,56(18):157-164.

[2] SU W,WANG Z F. Widening residual skipped network for semantic segmentation[J]. IET Image Processing,2017,11(10):880-887.

[3] 杨　鑫,于重重,王　鑫,等. 融合 ASPP-Attention 和上下文的复杂场景语义分割[J]. 计算机仿真,2020,37(9):204-208.

[4] 宋小娜,芮　挺,王新晴. 集合语义边界的道路环境语义分割方法[J]. 计算机应用,2019,39(9):2505-2510.

[5] LONG J,SHELHAMER E,DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Washington D. C. ,USA:IEEE,2015:3431-3440.

[6] 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. Munich,Germany:Springer,2015:234-241.

[7] 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.

[8] PASZKE A,CHAURASIA A,KIM A. ENet:a deep neural network architecture for real-time semantice segmentation[J]. arXiv:1606.02147,2016.

[9] ZHAO H, SHI J, QI X. Pyramid scene parsing network[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Honolulu, HI, USA: IEEE, 2017: 2881-2890.

[10] ZHANG H,DANA K,SHI J,et al. Context encoding for semantic segmentation[C]//2018 IEEE/CVF conference on computer vision and pattern recognition. Salt Lake City,UT,USA:IEEE,2018:85-97.

[11] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Las Vegas, NV, USA:IEEE,2016:770-778.

[12] CHEN L C,PAPANDREOU G,KOKKINOS I,et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[J]. Computer Science,2014(4):357-361.

[13] 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 & Machine Intelligence,2017,40(4):834-848.

[14] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Computer vision - ECCV 2018. Munich,Germany:Springer,2018:833-851.

[15] LI Y J,HUANG J B,AHUJIA N,et al. Deep joint image filtering[C]//Computer vision - ECCV 2016. Amsterdam, The Netherlands:Springer,2016:154-169.

[16] WU Huikai,ZHANG Junge,HUANG Kaiqi,et al. FastFCN:rethinking dilated convolution in the backbone for semantic segmentation[J]. arXiv:1903.11816,2019.

[17] CHOLLET F. Xception:deep learning with depth wise separable convolutions[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Honolulu,HI,USA:IEEE,2017:1251-1258.

[18] CORDTS M,OMRAN M,RAMOS S,et al. The cityscapes dataset for semantic urban scene understanding[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Las Vegas,NV,USA:IEEE,2016:3213-3223.

[19] LIN T, DOLLAR P, GIRSHICK R, et al. Feature pyramid Lonnetworks for object detection[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Honolulu,HI,USA:IEEE,2017:2117-2125.

(上接第 81 页)

computer vision and pattern recognition (CVPR). Long Beach,CA,USA:IEEE,2019:1438-1447.

[54] LI Ang,QI Jianzhong,ZHANG Rui,et al. Generative image inpainting with submanifold alignment[C]//International joint conferences on artificial intelligence. Macao,China:[s. n.],2019:811-817.

[55] SAGONG Min-Cheol, SHIN Yong-Goo, KIM Seung-Wook,et al. PEPSI:fast image inpainting with parallel decoding network[C]//2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Long Beach, CA, USA:IEEE,2019:11360-11368.

[56] LI C,HE K,LIU K,et al. Image inpainting using two-stage loss function and global and local Markovian discriminators[J]. Sensors,2020,20(21):6193.