

文章编号:1007-2780(2021)11-1497-09

结合 ASPP 与改进 HRNet 的多尺度图像 语义分割方法研究

史健锋¹, 高治明², 王阿川^{1*}

(1. 东北林业大学 信息与计算机工程学院, 黑龙江 哈尔滨 150040;

2. 东北林业大学 机电工程学院, 黑龙江 哈尔滨 150040)

摘要:针对经典语义分割算法中存在的模型庞大、训练困难以及多尺度目标分割等问题,基于空洞空间金字塔池化(ASPP)和高分辨率网络(HRNet)提出了一种高效的多尺度图像语义分割方法。首先利用深度可分离卷积结合 1×1 卷积的方式改进了 HRNet 的基础模块,减少了模型参数;其次通过在全部的卷积层之后、修正线性单元(relu)激活函数之前添加批归一化层(BN)改善 DeadRelu 问题;最后添加了使用混合扩张卷积框架重构的 ASPP,使用并行的上采样通道融合二者的优势,获得空间精准的分割结果,提出了 Re-ASPP-HRNet。在公开数据集 PASCAL VOC2012 和 CityScapes 上的实验表明,改进后的方法相比于原 HRNet 分别实现了 0.8%、0.5% 平均交并比的精度提升,且减少了 1/2 的参数数量以及 1/3 占用内存。进一步提升了网络的性能,实现了更加高效可靠、有普适性的多尺度语义分割算法。

关键词:语义分割;深度学习;神经网络;高分辨率网络

中图分类号:TP391.4 **文献标识码:**A **doi:**10.37188/CJLCD.2021-0093

Multi-scale image semantic segmentation based on ASPP and improved HRNet

SHI Jian-feng¹, GAO Zhi-ming², WANG A-chuan^{1*}

(1. College of Information and Computer Engineering, Northeast Forestry University,
Harbin 150040, China;

2. College of Mechanical and Electrical Engineering, Northeast Forestry University,
Harbin 150040, China)

Abstract: Aiming at the problems of huge model and multi-scale object segmentation in the classical semantic segmentation algorithm, an efficient multi-scale image semantic segmentation method based on ASPP and HRNet is proposed. Firstly, the basic block is improved by using deep separable convolution combined with 1×1 convolution to reduce the model parameters. Secondly, a batch normalization layer (BN) is added after all convolution layers and before the relu activation function to improve the dead relu problem. Finally, the improved ASPP module based on the hybrid dilated convolution is

收稿日期:2021-04-08; **修订日期:**2021-05-18.

基金项目:黑龙江省自然科学基金(No. F2018002)

Supported by Natural Science Foundation of Heilongjiang Province(No. F2018002)

* 通信联系人, E-mail: wangca1964@126.com

added, and the advantages of the two are fused by using the parallel upsampling channels to obtain the spatial accurate segmentation results. The RE-ASPP-HRNet is proposed. Results on Pascal voc2012 and CityScapes show that the proposed method is effective. Compared with the original HRNet, it can improve the accuracy of 0.8% or 0.5% MIoU, and reduce the number of parameters by 1/2 and memory by 1/3. We implement a more efficient and reliable multi-scale semantic segmentation algorithm.

Key words: semantic segmentation; deep learning; neural network; high resolution network

1 引言

在计算机视觉识别任务中,分类任务占据主导地位^[1],其中语义分割更是分类任务中的核心,主要工作是在语义层次上依次分类和标记图像的每个像素。不同于一般的图像级分类任务,像素分类通常更加困难^[2],需要稠密的像素级别的预测能力和大量的计算成本。随着地质检测、场景理解等问题的研究不断深入,图像语义分割问题,尤其是对于拥有复杂多尺度目标图像的理解,仍然是低级计算机视觉领域尚未解决的重要问题。因此,对于高效、可以准确分割各种尺度物体的语义分割模型的研究也愈发重要。

最近,新兴的深度学习避免了人工进行复杂和繁琐的特征设计,可以通过前向传播和反向传播来自学习更加丰富的细节特征,在多个人工智能领域获得了最先进的成果。2015年,Long等人首次提出了可以输入任意大小的图像以提取特征的完全卷积网络(Fully Convolutional Networks, FCN)^[3],模型通过反卷积方式增加特征图的分辨率以进行像素级的稠密估计,成功地将图像分类网络转化成为图像分割网络,促进了图像语义分割的研究。但其仍存在一些不足:上采样过程粗糙,细节丢失严重;未能有效地计算像素之间的联系,语义分类不准确;固定大小的常规卷积难以充分提取多尺度目标的特征。

为了解决 FCN 存在的问题,学者们在其基础上提出了一系列优化的算法:基于编码器-解码器方法,如 U-net^[4]等,这类网络采用对称的编码解码结构,逐步恢复物体细节和图像分辨率,但是连续的上采样并不能恢复多次最大值池化损失的物体细节信息,精度有限。基于优化卷积方法,如 DeepLab 系列等,通过空洞卷积等优化卷积技术,在不增加参数数量的同时成倍扩大感受野^[5],获得多尺度特征信息从而提高分割精度。基于循环

神经网络或概率图模型的方法,如 CRFasRNN 等,在卷积神经网络基础上,以像素点为节点,像素点之间的概率相关关系为边,获取图像全局信息和像素级语义信息,为语义分割过程提供丰富的上下文内容^[6],但是复杂的全连接概率图模型增加了参数数量和计算时间,分割速度缓慢。

2019年,KeSun等人共同提出了 HRNet。其以一个高分辨率子网作为第一个阶段,逐渐增加高分辨率到低分辨率的子网以形成更多的阶段,多个阶段的子网并行连接^[7]。网络通过不断地交换信息和多尺度的特征融合从而在整个过程中保持高分辨率,得到的特征表示不仅很强,在空间上也很精确,大为改善了目前已有算法存在的细节丢失、精度不够等问题。但是多个阶段之间大量重复的相互融合也会产生数倍的计算复杂度和参数量,密集的特征融合也会计算大量冗余和相同的信息。

为了高效解决多尺度目标分割问题,本文对 HRNet 进行了改进,进一步提升模型的整体性能。通过使用深度可分离卷积结合批量归一化层的方式减少参数、加速模型训练、优化 DeadRelu 问题。其次利用混合扩张卷积框架(Hybrid Dilated Convolution, HDC)改进了空洞空间金字塔池化模块(Atrous Spatial Pyramid Pooling, ASPP),将其添加进网络之中,利用并行的双通道上采样融合机制,获得最终的语义分割模型 Re-ASPP-HRNet。实验表明,本文改进的 Re-ASPP-HRNet 网络相比于 HRNet,参数数量下降了 1/2,内存占用降低了 1/3,并且平均交并比(Mean Intersection Over Union, MIoU)也有所提升。

2 改进的 Re-ASPP-HRNet

2.1 网络模型结构

目前现有的方法基本是通过下采样操作降低

分辨率,然后通过上采样恢复分辨率以获得输出。这些网络通常是线性的,例如依赖于对称的编码-解码结构,代表网络有 Seg-Net^[8] 等。编码网络之后利用跳跃链接和特征融合进行上采样的融合结构。重量级的编码结构之后采用简单的上采样方式快速恢复特征图分辨率的结构,代表网络有 FCN 等。串行网络结构如图 1 所示。

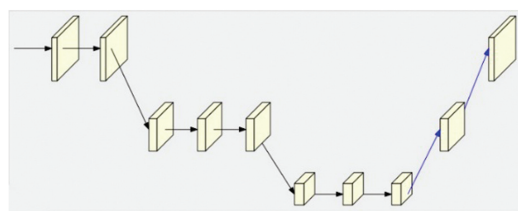


图 1 串行网络模型结构图

Fig.1 Serial network model structure

不同于目前流行的串行连接的网络,KeSun 等人共同提出的 HRNet 是一种全新的并行体系结构。其通过 4 个阶段不断地相互融合全程保持分辨率,避免了下采样带来的信息丢失。这样的并行网络有两个优点:(1)并行结构能够保持高分辨率,而不是通过从低到高的过程恢复分辨率,因此预测的热图在空间上更加精确^[9]。(2)大多数的融合方案利用的是串行模型在低级和高级上的

特征表示。相反,并行结构在相同深度和相似级别的低分辨率的特征表示的帮助下,执行重复的多尺度融合来增强高分辨率的特征表示^[10],输出的语义分割热图更加准确。HRNet 模型结构如图 2 所示。

为了高效地实现多尺度目标分割算法,本文设计了如图 3 所示的 Re-ASPP-HRNet。首先将

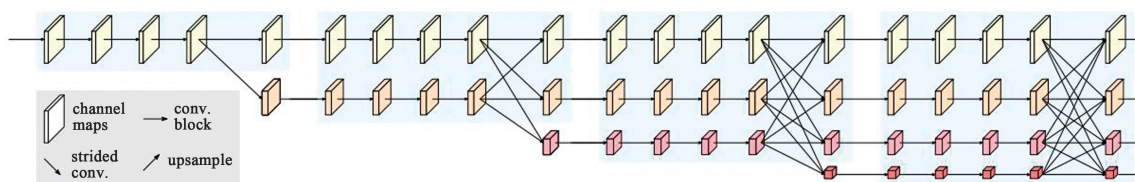


图 2 HRNet 模型结构图

Fig.2 HRNet model structure

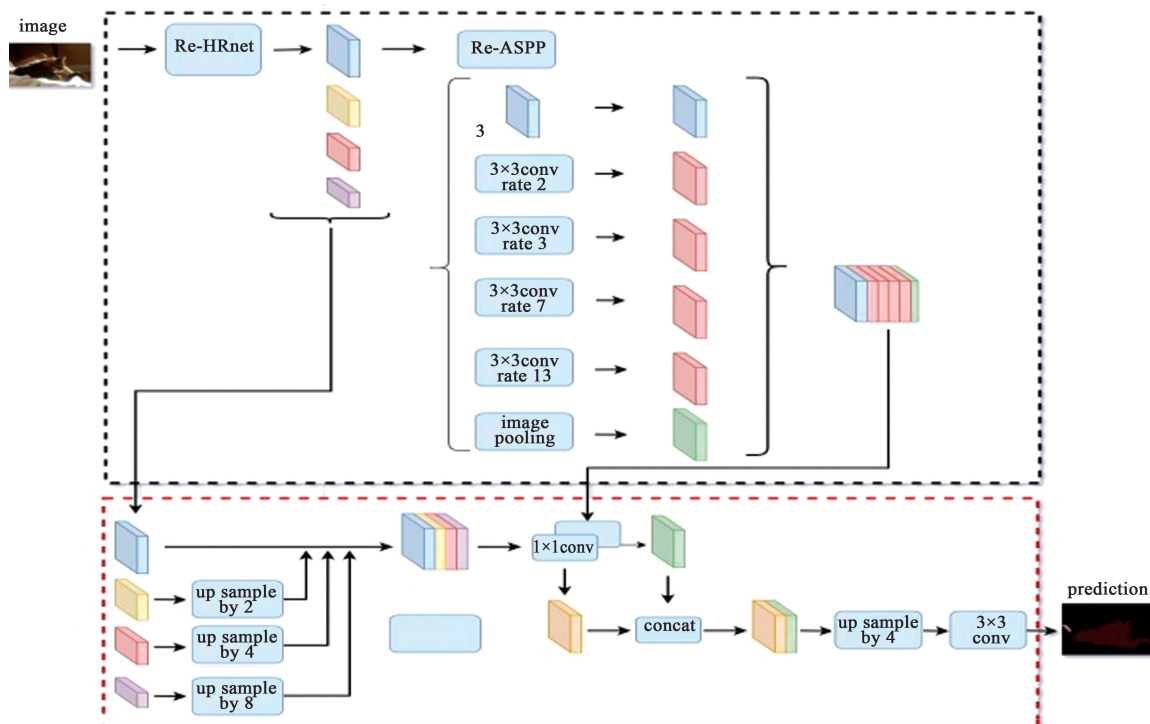


图 3 Re-ASPP-HRNet 结构图

Fig.3 Structure of Re-ASPP-HRNet

数据批次输入到改进的 HRNet 中进行特征提取,通过 4 个阶段最终获得 4 份分辨率分别为原图像 $1/4$ 、 $1/8$ 、 $1/16$ 、 $1/32$ 的特征图。由于高分辨率网反复融合多尺度特征的特性,所获得的特征图并没有因为连续下采样而损失大量细节信息。解码部分通过结合两种办法来获得预测输出:(1)将 4 种尺寸的特征图分别上采样至原图像 $1/4$ 大小后进行连接,得到一组细节丰富且在空间上更加精准的特征图表示。(2)将 4 种尺寸的特征图通过改进的 Re-ASPP 模块,得到一组多尺度特征信息丰富的特征图表示。两方面的输出分别通过 1×1 卷积进行连接,上采样恢复至原图像分辨率,最终通过 3×3 卷积和 softmax 激活函数获得空间精准、尺度丰富、细节准确的语义分割预测图。

2.2 改进的高分辨率网络

HRNet 复杂的并行子网和反复的特征融合导致了庞大的参数数量和计算复杂度,而且难免存在着大量重复计算、过拟合和占用大量内存等问题。因此本文对 HRNet 进行了以下改进。残差模块是 HRNet 的基础模块,其结构如图 4(a) 所示。

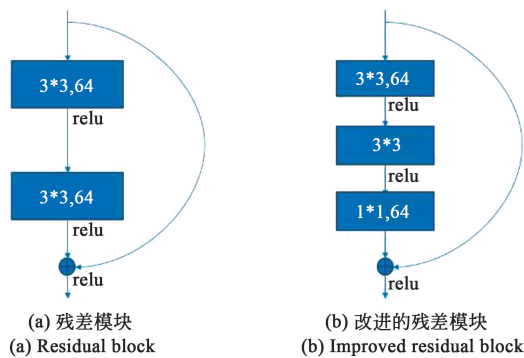


图 4 残差模块结构图

Fig.4 Structure of residual block

大量的残差模块带来了庞大的参数数量,为此本文选择使用深度可分离卷积结合 1×1 卷积的方式替换常规卷积^[11],改进了残差模块,结构如图 4(b)所示。深度可分离卷积中一个卷积核只负责一个通道,一次卷积运算之后输出相同通道数量的特征图。而这种运算对输入的每个通道独立地进行卷积运算,没有有效地利用不同通道在相同空间位置上的特征信息^[12],因此通常结合 1×1 卷积运算来改变通道数量,组合生成新的特征图。以三通道

的 7×7 像素的输入为例,想要获得 $5 \times 5 \times 128$ 的输出结果需要进行的常规卷积操作如图 5 所示。其参数数量为 $128 \times 3 \times 3 \times 3 = 3456$ 。而换成深度可分离卷积结合 1×1 卷积方式的话,卷积过程如图 6 所示。其参数数量为 $3 \times 3 \times 3 = 27$, $1 \times 1 \times 3 \times 128 = 384$ 。二者之和为 411。

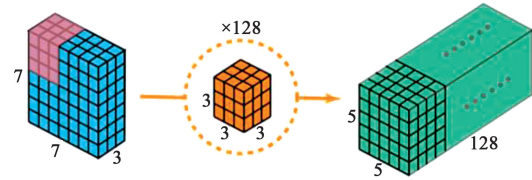


图 5 常规卷积运算示意图

Fig.5 Conventional convolution

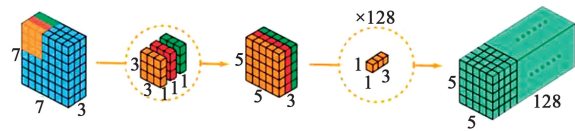


图 6 深度可分离卷积运算示意图

Fig.6 Depth separable convolution

由此可见在相同的情况下,深度可分离卷积结合 1×1 卷积的参数数量远小于常规卷积操作。当输出通道数量较大时,差距会更加明显。因此同样的参数数量使用深度可分离卷积可以搭建更深的神经网络,通常来说意味着更强的性能。但是大量减少参数数量对于一般大小的模型而言,会显著下降模型的能力,得到的结果是次优的。因此本文综合考虑以上原因,选择保留一层常规卷积,另一层以深度可分离卷积结合相同通道数量的 1×1 卷积来代替,减少了大约 $1/2$ 的模型参数,并且没有损失精度。

由于 relu 存在着致命的 DeadRelu 问题,本文将批归一化层应用于全部的 relu 激活函数之前。批归一化层可以将输入的分布拉回到均值为 0、方差为 1 的正态分布上,使梯度更新的幅度均匀,从而解决 DeadRelu 问题。SergeyIoffe 等人指出,通过将规范化作为模型体系结构的一部分并为每个训练小批量执行规范化可以获得训练优势^[13]。在深度网络中,过高的学习率可能导致梯度爆炸或消失,以及陷入局部极小值,而批次归一化不仅允许我们使用更高的学习率,还可以大幅减少训练次数并实现相同精度。因此本文基于以上分析,将批归一化层引入至 HRNet 中全部的

卷积层之后,relu 激活函数之前以优化训练,二者结合使用可以达到最佳的性能。

综上,本文通过引入深度可分离卷积结合 1×1 卷积的方式来降低网络的复杂度,减少模型参数,并通过在全部卷积层之后、relu 激活函数之前,增加批次归一化层,以加速网络训练,改善 DeadRelu 问题。

2.3 ASPP 模块的改进

虽然 HRNet 复杂的并行结构可以进行反复的特征融合,但是普通的卷积结构仍然无法充分地提取多尺度的特征。因此本文选择使用 ASPP 模块对改进 HRNet 输出的 4 种尺寸的特征图进行处理,加强模型分割不同尺度物体的能力。

Chen 等人首次在 DeepLab 系列中提出了一组并行的不同空洞率的空洞卷积组成的 Aspp 模块,在不使用下采样操作的情况下增大卷积核的感受野进而有效地提取多尺度物体的特征^[14],其结构如图 7 所示。

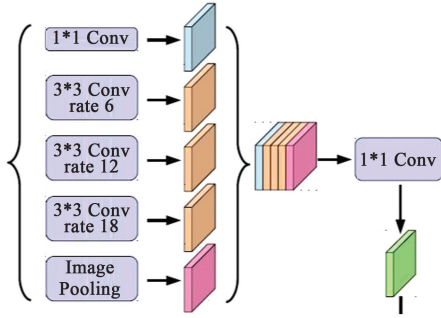


图 7 空洞空间金字塔池化模块

Fig.7 Atrous spatial pyramid pooling module

虽然空洞卷积能够有效增加感受野,但是其存在着“网格问题”。在叠加使用多个不适合空洞率的空洞卷积时网格问题更加严重。从图 8(a)可以看出,当使用一组空洞率为 2、2、2 的空洞卷积时仅有部分像素参与了运算。从图 8(b)可以看出,当空洞率为 1、2、3 时感受野覆盖较为完整。

因此 Panqu Wang 等人^[15]设计了混合扩张卷积框架以减轻叠加使用空洞卷积时产生的网格问题,扩大接收野聚合全局信息。基于以上分析,本文利用混合扩张卷积框架重新构建了 ASPP 模块,提出了 Re-ASPP 结构。

考虑到 N 个尺寸为 $K \times K$ 的扩张卷积,对应的扩张率为 $[r_1, \dots, r_i, \dots, r_n]$,HDC 的目标是让最后的接收野全覆盖整个区域(没有任何空洞

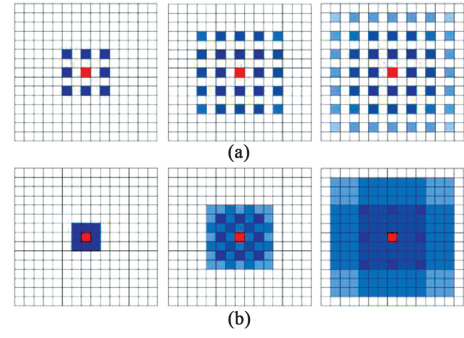


图 8 网格问题示意图

Fig.8 Gridding issue diagram

或丢失边缘),我们定义两个非零点之间最大距离为:

$$M_i = \max [M_{i+1} - 2r_i, M_{i+1} - 2(M_{i+1} - r_i), r_i], \quad (1)$$

其中 $M_n = r_n$, 设计的目标是让 $M_2 \leq K$ 。需要注意的是,一个组内的卷积不应该有一个固定的变换因子,即不要有大于 1 的公约数(例如 2、4、8 的公约数为 $2 > 1$),否则依旧无法减小网格效应。因此设计了空洞率分别为 2、3、7、13 的一组空洞卷积,显然不存在公约数,且

$$M_2 = \max [M_3 - 2r_2, -M_3 + 2r_2, r_2] = 3 \leq 3(2), \quad (2)$$

4 种尺度的感受野可以覆盖整个区域,改善网格问题的同时有效提取不同尺度的物体特征。本文定义了 6 层的 Re-ASPP 结构,如图 9 所示:首先利用原特征图作为第一层以补充空间信息;其次利用 4 种不同空洞率的空洞卷积提升模型提取不同尺度目标特征的能力形成 2 到 5 层;最后使用平均池化获得图像级别的全局特征形成第六层,并且增加了可以提高训练速度和加快网络训练和收敛的批归一化层。并行的 6 种方式有效增加了

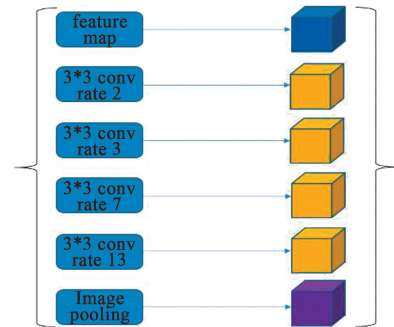


图 9 改进的 Re-ASPP 模块

Fig.9 Improved Re-ASPP module

HRNet 提取多尺度的物体特征的能力。

2.4 解码部分结构设计

HRNet 的并行网络结构输出了 4 种分别为原图像大小 1/4、1/8、1/16、1/32 的特征图, 第一张 1/4 的特征图复制备用, 将后 3 种分别上采样至原图像的 1/4 大小, 四者连接起来并以 1×1 卷积改变通道数量, 得到第一份特征输出。将之前保存的 1/4 大小的特征图通过改进后的 Re-ASPP 模块, 通过 1×1 卷积改变通道数量, 获得第二份特征输出, 并且与第一份特征输出连接起来, 上采样 4 倍之后, 通过 3×3 卷积和 softmax 激活函数获得最终的语义分割结果。

双通道并行设计的解码结构结合了两方面信息, 一方面全程保持高分辨率的输出可以更好地保留物体细节和空间定位信息, 另一方面改进的空间金字塔池化模块使用了多种尺度的空洞卷积和全局平均池化, 可以更好地提取多尺度目标的特征。二者综合起来可以结合各自的优势从而取得更好的结果。

3 实验与分析

3.1 数据集

本文使用 PASCAL VOC2012 和 CityScapes 公开数据集进行实验。前者包含了日常生活中较为常见的人、鸟、沙发等 20 个语义类别和 1 个背景类别, 分为包含 1 464 张图片的训练集, 包含 1 449 张图片的验证集以及包含 1 456 张图片的测试集。后者记录了欧洲多个城市的城市景观, 拥有 30 个类别, 分为包含 2 975 张图片的训练集, 包含 500 张图片的验证集以及包含 1 525 张图片的测试集^[16]。

3.2 评价标准

本文使用平均交并比 MIoU 来评价 Re-ASPP-HRNet 分割效果。平均交并比的计算公式如式 (3) 所示。其中 k 代表物体类别数量, $k+1$ 为物体加背景类别数, p_{ij} 表示实际类别为 i 类且预测为 j 类的像素数目。

$$\text{MIoU} = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (3)$$

3.3 模型训练

本实验基于英特尔 i5-10400F CPU, NVIDIA GeForce RTX 2070 SUPER 8 GB 显卡, 编程语言 python, 使用 tensorflow 和 keras 进行实验。受到 GPU 显存大小的限制, 训练时需要分别调整图片大小至 320×320 像素和 512×512 像素并将训练样本分批输入模型中。使用交叉熵损失函数和 adam 优化器, adam 可以自适应学习率的同时对其加以约束。批次大小分别设置为 12 和 6, 训练次数设置为 300。

3.4 实验设计

为了验证本文提出的 Re-ASPP-HRNet 的实际效果, 选择与当前主流语义分割算法 (FCN-2S, BlitzNet, PSPNet, DeepLabv3, HRNet) 进行了两组对比实验。(1) 不同模型对于 PASCAL VOC2012 中的多种物体的分割情况和平均交并比如表 1 所示。由表 1 可知, Re-ASPP-HRNet 模型在 PASCAL VOC2012 数据集上的平均交并比均高于主流模型, 验证了本文算法的优势。不同模型分割效果如图 10 所示。(2) 对于 CityScapes 数据集的实验结果如图 11 所示。可以看出改进的 HRNet 对于物体细节更加敏感,

表 1 PASCAL VOC2012 结果对比

Tab.1 Comparison of PASCAL VOC2012 result

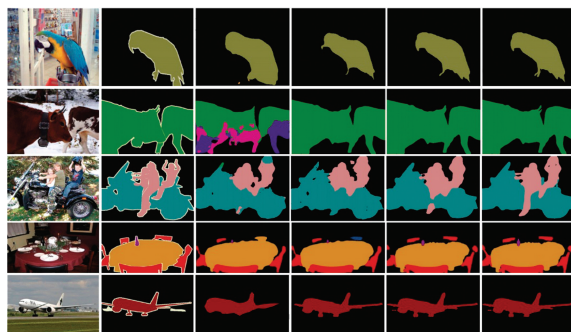
(%)

Model	FCN-2S	BlitzNet	PSPNet	DeepLabv3	HRNet	OurHRNet
Aero	81.8	90.1	95.8	96.4	96.4	96.9
Bicycle	37.0	38.7	72.7	76.6	69.6	73.3
Bird	79.5	87.5	95.0	92.7	93.6	95.5
Boat	57.2	68.6	78.9	77.8	76.1	78.8
Bottle	67.5	70.1	84.4	87.6	86.6	86.4

续 表

(%)

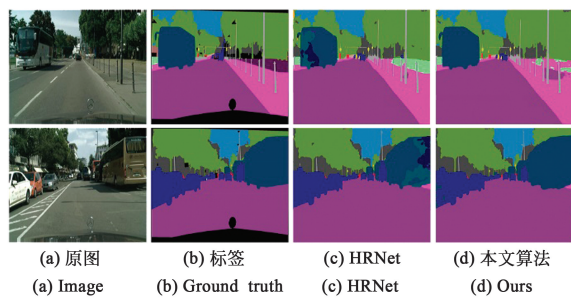
Model	FCN-2S	BlitzNet	PSPNet	DeepLabv3	HRNet	OurHRNet
Bus	83.8	93.1	94.7	96.7	96.4	96.6
Car	79.3	86.4	92.0	90.2	92.3	90.0
Cat	83.0	89.2	95.7	95.4	95.7	97.2
Chair	28.5	32.3	43.1	47.5	47.8	53.4
Cow	74.5	81.7	91.0	93.4	95.0	95.1
Table	57.5	67.9	80.3	76.3	81.5	74.4
Dog	76.0	82.2	91.3	91.4	93.1	93.9
Horse	75.9	82.9	96.3	97.2	95.9	96.8
M-bike	79.5	84.7	92.3	91.0	91.2	93.1
Person	78.6	81.5	90.1	92.1	90.7	92.5
Plant	57.0	63.3	71.5	71.3	77.0	81.0
Sheep	77.8	85.5	94.4	90.9	94.7	95.1
Sofa	45.3	55.5	66.9	68.9	68.9	67.1
Train	73.7	83.1	88.8	90.8	89.3	90.8
Tv	63.2	70.6	82.0	79.3	81.7	81.8
MIoU	69.0	75.6	85.4	85.7	86.2	87.0



(a) 原图 (b) 标签 (c) FCN-2S (d) PSPNet (e) HRNet (f) 本文算法
(a) Image (b) Ground truth (c) FCN-2S (d) PSPNet (e) HRNet (f) Ours

图 10 不同模型分割结果

Fig.10 Segmentation results of different models



(a) 原图 (b) 标签 (c) HRNet (d) 本文算法
(a) Image (b) Ground truth (c) HRNet (d) Ours

图 11 CityScapes 数据集分割效果

Fig.11 Segmentation results of CityScapes

验证了本文方法对于不同数据集的普适性。

下面将从参数数量、平均交并比两个方面来比较深度可分离卷积代替常规卷积带来的影响。表 2 中列出了不同模型参数数量和在两种数据集上的平均交并比, Ours-w100 中的 w100 表示替换了残差模块中全部的传统卷积层, 同理 w50 表示仅替代了 1/2 的传统卷积层。

表 2 不同模型参数数量和 MIoU 对比

Tab.2 Comparison of different model parameters and MIoU

Model	Param/M	MIoU(P)/%	MIoU(C)/%
Unet++	59.5	80.4	75.5
DeepLabv3	58.0	85.7	81.3
PSPNet	65.9	85.4	81.2
HRNet	65.9	86.2	81.6
Ours-w100	17.5	83.6	78.6
Ours-w50	34.8	87.0	82.1

从表 2 可以看出,原 HRNet 网络的参数量巨大,在改进其基础模块并增加本文设计的 Re-ASPP 模块后,可以在减少大量参数的基础上,获得同级别的分割精度。究其原因,HRNet 复杂的并行结构可以获得更精确的结果。然而,庞大的参数数量限制了模型进一步提高性能。因此本文通过使用深度可分离卷积结合 1×1 卷积大幅减少了参数数量,且以改进的 Re-ASPP 模块,提高网络对不同尺度物体的检测能力,最终获得的 Re-ASPP-HRNet-w50 模型,以一半于 Unet++、PSPNet、HRNet 的参数数量提升了精度,可见本文方法的有效性。

4 结 论

本文针对已有模型对于小尺度目标分割困难等问题,对高分辨率网络进行了改进,提出了 Re-ASPP-HRNet:(1)通过引入深度可分离卷积以及批归一化层,减少网络参数的同时增强泛化性;(2)引入了消除网格问题且能充分提取多尺度目标特征的改进 ASPP 模块;(3)使用并行的上采样通道融合二者的优势,获得了空间精准的分割结果。实验表明,本文设计的 Re-ASPP-HRNet 实现了 0.8% 和 0.5% MIoU 的精度提升,对比 HRNet 减少了 1/2 的参数数量以及 1/3 占用内存,而且能够准确分割小尺度目标。

参 考 文 献:

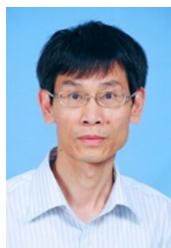
- [1] 王曦,于鸣,任洪斌.UNET 与 FPN 相结合的遥感图像语义分割[J].液晶与显示,2021,36(3):475-483.
WANG X, YU M, REN H E. Remote sensing image semantic segmentation combining UNET and FPN [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(3): 475-483. (in Chinese)
- [2] 陈晓冬,盛婧,杨晋,等.多参数 Gabor 预处理融合多尺度局部水平集的超声图像分割[J].中国光学,2020,13(5):1075-1084.
CHEN X D, SHENG J, YANG J, *et al.* Ultrasound image segmentation based on a multi-parameter Gabor filter and multiscale local level set method [J]. *Chinese Optics*, 2020, 13(5): 1075-1084. (in Chinese)
- [3] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015: 3431-3440.
- [4] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]//*18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich: Springer, 2015: 234-241.
- [5] 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.
- [6] ZHENG S, JAYASUMANA S, ROMERA-PAREDES B, *et al.* Conditional random fields as recurrent neural networks [C]//*IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015: 1529-1537.
- [7] SUN K, XIAO B, LIU D, *et al.* Deep high-resolution representation learning for human pose estimation [C]//*2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach: IEEE, 2019: 5686-5696.
- [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] WU H S, LIANG C X, LIU M S, *et al.* Optimized HRNet for image semantic segmentation [J]. *Expert Systems with Applications*, 2021, 174: 114532.
- [10] SUN K, ZHAO Y, JIANG B R, *et al.* High-resolution representations for labeling pixels and regions [J]. *arXiv*: 1904.04514, 2019: 1-4.
- [11] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C]//*2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2017: 1800-1807.
- [12] SZEGEDY C, VANHOUCKE V, IOFFE S, *et al.* Rethinking the inception architecture for computer vision

- [C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 2818-2826.
- [13] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [J]. *arXiv*: 1502.03167, 2015: 2-4.
- [14] CHEN L C, ZHU Y K, PAPANDREOU G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation [C]//*15th European Conference on Computer Vision*. Munich: Springer, 2018: 833-851.
- [15] WANG P Q, 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: IEEE, 2018: 1451-1460.
- [16] CORDTS M, OMRAN M, RAMOS S, *et al.* The cityscapes dataset for semantic urban scene understanding [C]//*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas: IEEE, 2016: 3213-3223.

作者简介:



史健锋(1997—),男,黑龙江齐齐哈尔人,硕士研究生,2019 年于黑龙江大学获得学士学位,主要从事计算机视觉研究。E-mail:2020111879@nefu.edu.cn



王阿川(1964—),男,黑龙江哈尔滨人,博士,教授,2011 年于东北林业大学获得博士学位主要从事计算机视觉,高分辨率遥感影像研究。E-mail:wangca1964@126.com