第 58 卷 第 16 期/2021 年 8 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于改进 DeepLabv3＋网络的图像语义分割方法

徐聪*，王丽

天津城建大学计算机与信息工程学院，天津 300384

摘要 针对 DeepLab 网络不能充分利用多尺度特征信息，忽略了高分辨率浅层特征以及直接上采样倍数过大导致的重要像素信息丢失问题，提出了一种基于改进 DeepLabv3＋网络的图像语义分割方法。首先，充分利用了网络产生的多尺度特征信息，并采用特征金字塔网络有效融合了高分辨率的浅层特征；然后，使用逐层上采样增强图像中像素信息的连续性；最后，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，提高了网络模型的训练效率。在语义分割标准数据集 PASCAL VOC2012 验证集上的实验结果表明，本方法的平均交并比可达到 79.97%。相比 DeepLab 网络，可预测出更精细的语义分割结果。

关键词 图像处理；改进的 DeepLabv3＋网络；特征金字塔网络；空洞空间金字塔池化模块

中图分类号 TP751 **文献标志码** A **doi**：10.3788/LOP202158.1610008

Image Semantic Segmentation Method Based on Improved DeepLabv3＋ Network

Xu Cong*, Wang Li

School of Computer and Information Engineering, Tianjin Chengjian University, Tianjin 300384, *China*

Abstract This paper proposes an image semantic segmentation method based on an improved DeepLabv3＋ network to address the DeepLab network's inability to fully utilize multiscale feature information while ignoring the problem of high-resolution shallow features and the loss of important pixel information due to excessive direct upsampling multiples. First, the multiscale feature information generated by the network is fully utilized and the feature pyramid network is used to effectively fuse high-resolution shallow features. Then, layer-by-layer upsampling is used to improve the image's pixel information continuity. Finally, in the atrous spatial pyramid pooling module, the standard convolution is replaced with depthwise separable convolution, enhancing the network model's training efficiency. The experimental results on the semantic segmentation standard data set PASCAL VOC2012 verification set show that, the mean intersection over union of the method can reach 79.97%. It can predict more refined semantic segmentation results compared with the DeepLab network.

Key words image processing; improved DeepLabv3＋ network; feature pyramid network; atrous spatial pyramid pooling module

OCIS codes 100.4996; 100.6890; 150.1135

1 引 言

图像语义分割是计算机视觉领域的重要组成部分，在众多工业领域中起到了至关重要的作用[1]。图像语义分割的主要任务是为图像中的每一个像素分配一个预先定义好的语义类别标签[2]。DeepLab 系列网络[3-6]是图像语义分割领域中主流的网络模型，其中，比较成熟的 DeepLabv3＋网络[6]在实际生

收稿日期：2020-10-24；**修回日期**：2020-11-12；**录用日期**：2020-12-08

基金项目：国家重点研发计划(2019YFD1100402)、天津市教委科研计划(2016CJ12)

通信作者：*769458796@qq.com

活中也得到了广泛应用[7]。

目前，大多数语义分割网络模型都是以编码-解码模块[8]为基础。其中，编码层主要利用卷积[9]和池化操提取特征图信息，解码层主要负责还原高分辨率语义特征，而一个有效的解码层对于语义分割任务至关重要。在 DeepLab 系列网络中，除了 DeepLabv3＋网络之外，其他网络几乎都没有加入一个有效的解码模块，且存在由于上采样倍数过大导致的重要像素信息丢失问题。DeepLabv3＋网络将 DeepLabv3 网络[5]作为编码模块，设计了一个简单有效的解码模块，但其只利用了骨干网络中产生的 1 个高分辨率低级特征图，没有充分利用骨干网络产生的多阶段高分辨率特征图，导致预测结果中目标物体不连续。此外，随着网络层数的加深，特征图的分辨率也逐渐减小，而空洞空间金字塔池化(ASPP)模块[4]中空洞率较大的空洞卷积并不利于分割低分辨率特征图；且在其解码层中，将直接上采样 4 倍的特征图与高分辨率低级特征图进行拼接融合，导致特征图中重要像素信息的丢失，语义分割的预测结果不够精细。

针对上述方法存在的不足，本文提出了一种基于改进 DeepLabv3＋网络的语义分割方法。首先，充分利用了骨干网络产生的多阶段高分辨率低级特征图，并利用特征金字塔网络(FPN)将其与网络后半部分的低分辨率高级特征图进行拼接融合；然后，在骨干网络中得到多个不同分辨率的特征图，并在解码过程中逐层将不同分辨率的特征图拼接后进行卷积和上采样操作，以充分有效地利用低级特征图的语义信息，从而改善上采样幅度过大导致的重要像素信息丢失问题。此外，还提出了一种新的 ASPP 模块，将 ASPP 模块中的标准卷积替换成深度可分离卷积，提高了网络模型的训练效率。

2　网络结构

2.1　DeepLabv3＋网络

DeepLabv3＋网络的结构如图 1 所示，主要分为编码层和解码层，其骨干网络为残差网络(ResNet-101)[10]。其中，Conv 为卷积操作，rate 为空洞率。在编码层，原始图像先经过骨干网络提取特征信息。可以发现，图像依次变为原始图像大小的 1/4、1/8、1/16。在 Block_4 中，有空洞率分别为 2、4、8 的空洞卷积。由空洞卷积[11-12]的特性可知，网络可在不损失图像分辨率及不增加参数量的前提下获得更大的感受野。由 Block_4 得到的 1/16 大小特征图进入 ASPP 模块，ASPP 模块包括 1×1 卷积，空洞率分别为 6、12、18 的空洞卷积及全局平均池化[13]，之后将得到的特征图在通道维度上进行拼接融合，并通过 1×1 卷积降低特征图的通道数。在解码层，DeepLabv3＋网络将编码层得到的特征图进行 4 倍上采样，并将上采样后的特征图与骨干网络中 Block_1 经 1×1 卷积得到的特征图进行拼接融合，再将特征图经 3×3 卷积后进行 4 倍上采样，从而得到 DeepLabv3＋网络预测的语义分割结果。

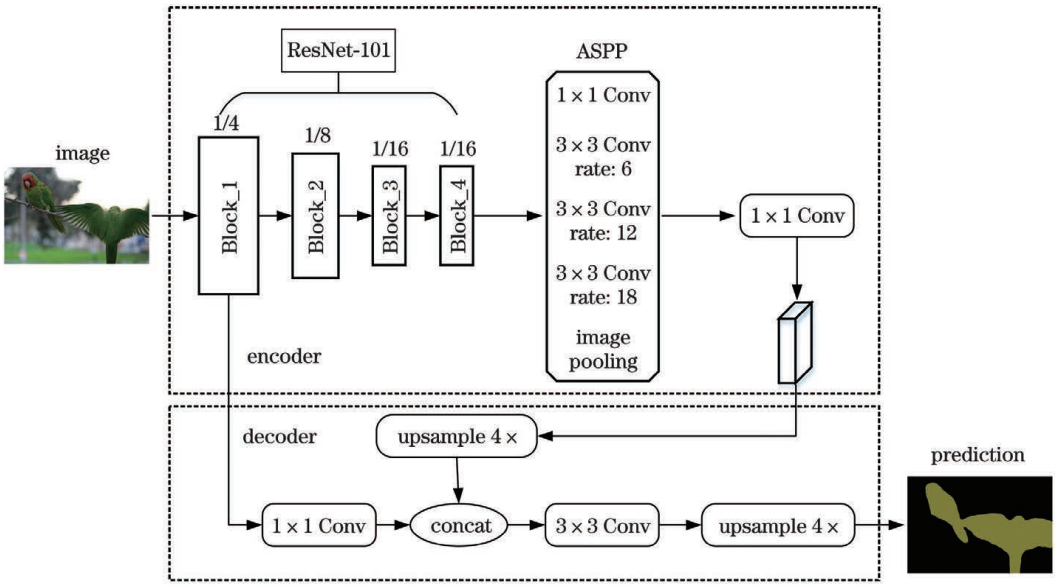

图 1　DeepLabv3＋网络的结构

Fig. 1　Structure of the DeepLabv3＋ network

2.2　改进的 DeepLabv3＋网络

改进的 DeepLabv3＋网络结构如图 2 所示，同样分为编码层和解码层，其骨干网络为 ResNet-101，具体改进如下。

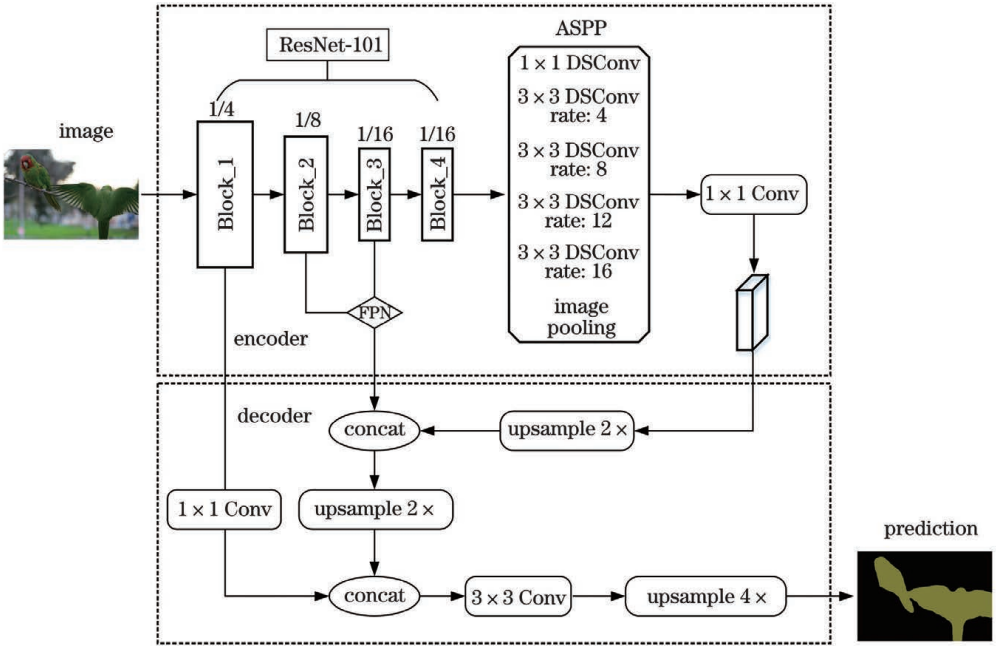

图 2　改进 DeepLabv3＋网络的结构

Fig. 2　Structure of the improved DeepLabv3＋ network

1）在编码层，将 DeepLabv3＋网络 ASPP 模块中原有空洞率分别为 6、12、18 的空洞卷积替换为空洞率分别为 4、8、12、16 的空洞卷积。随着骨干网络对图像特征的提取，特征图的分辨率也逐渐变小，设计空洞率较小的空洞卷积能更好地提取低分辨率特征图的信息。同时，将 ASPP 模块中原有的标准卷积替换成深度可分离卷积（DSConv），即将 1×1 卷积变为 1×1 深度可分离卷积，将所有 3×3 空洞卷积变为 3×3 空洞深度可分离卷积。相比标准卷积，深度可分离卷积可以大大减少训练过程中的参数量，且能在对预测精度影响不大的前提下提高网络模型的训练效率。

2）在解码层，改进的 DeepLabv3＋网络增加了一个特征融合[14]分支。由于骨干网络各个阶段产生的特征图对于最后的分割结果都是至关重要的，而 DeepLabv3＋网络只利用了骨干网络产生的 1/4 大小特征图。改进后的网络利用 FPN 融合了骨干网络产生的 1/8 和 1/16 大小特征图，并将其作为解码层中特征融合的一个重要分支。骨干网络得到 1/8 和 1/16 大小特征图的通道数分别为 512 和 1024，在 FPN 中，利用 1×1 卷积分别对其进行降维，使 2 个特征图的通道数都变为 256，之后将 1/16 大小、通道数为 256 的特征图进行 2 倍上采样，并与 1/8 大小、通道数为 256 的特征图以相加的方式融合，得到 FPN 最终产生的特征图。DeepLabv3＋网络在解码层中，将 ASPP 模块得到的特征图经 1×1 卷积后直接进行 4 倍上采样。而特征图中的目标类别由大量的像素矩阵构成，每个像素之间都存在密切的联系，将得到的特征图直接进行 4 倍上采样会导致图像中的像素不连续，丢失某些重要的像素信息，使网络的预测结果不精确。为了解决该问题，将 ASPP 模块得到的特征图经 1×1 卷积后进行 2 倍上采样，与 FPN 得到的特征图拼接融合后再进行 2 倍上采样，相当于将原始 DeepLabv3＋网络中直接的 4 倍上采样替换为 2 次 2 倍上采样，从而增强图像中像素的连续性，还原出的像素值更逼近原始图像，使网络的预测结果更精确。具体步骤：将 ASPP 模块得到的特征图进行 2 倍上采样后与 FPN 得到的特征图进行拼接融合，得到 1/8 大小、通道数为 512 的特征图，利用 1×1 卷积将其通道数变为 256 后继续进行 2 倍上采样，并将其与骨干网络 Block_1 得到的特征图进行拼接融合。综上所述，改进的 DeepLabv3＋网络在解码层中仅添加了 3 次 1×1 卷积操作，相比 DeepLabv3＋网络，增加的参数量极少。

2.3　空洞空间金字塔池化模块

改进 DeepLabv3＋网络中的 ASPP 模块如图 3 所示，ASPP 模块在提取图像多尺度语义特征方面

具有独特的优越性，被广泛应用于图像语义分割任务中。ASPP 模块由多个卷积操作及全局平均池化操作并行组成，除了 1×1 卷积外，每个卷积核都有不同的空洞率。空洞率较大的卷积核，有利于分割大目标；空洞率较小的卷积核，有利于分割小目标。因此，采用 ASPP 模块使网络具有多尺度卷积核，进而增加模型分割不同大小目标的能力。图 3 中的特征图为经骨干网络 ResNet-101 得到的 1/16 大小特征图，通道数为 2048。在 ASPP 模块中，特征图分别进行 1×1 的深度可分离卷积，空洞率依次为 4、8、12、16 的空洞深度可分离卷积以及全局平均池化，得到 6 个 1/16 大小、通道数为 256 的特征图。将 6 个特征图在通道维度上进行拼接融合，就能得到 ASPP 模块产生的特征图。

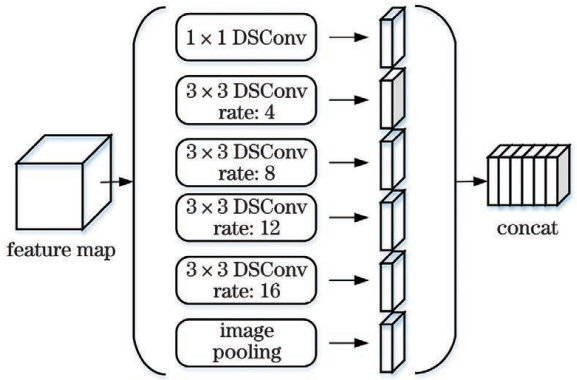

图 3 ASPP 模块的结构

Fig. 3 Structure of the ASPP module

2.4 深度可分离卷积

深度可分离卷积[15]的操作流程如图 4 所示。首先，将输入特征在所有通道依次进行3×3卷积，使其通道分离。然后，对特征图进行拼接，并进行 1×1 卷积，得到输出特征。为了更形象地表示通道分离的过程，图 4 中随机画出 5 个 3×3 卷积操作。深度可分离卷积主要体现在图 2 中的 ASPP 模块中，如 ResNet、VGG（Visual Geometry Group）[16]等骨干网络通过堆叠大量的卷积层和池化层使网络同时学习特征图的空间相关性和通道相关性。而深度可分离卷积先逐个通道学习空间相关性，再用标准卷积学习通道相关性。相比标准卷积，深度可分离卷积可以用更少的参数进行特征学习，计算速度更快。因此，实验在 ASPP 模块中将标准卷积替换为深度可分离卷积，在对预测精度影响不大的前提下，提升了网络模型的训练效率。

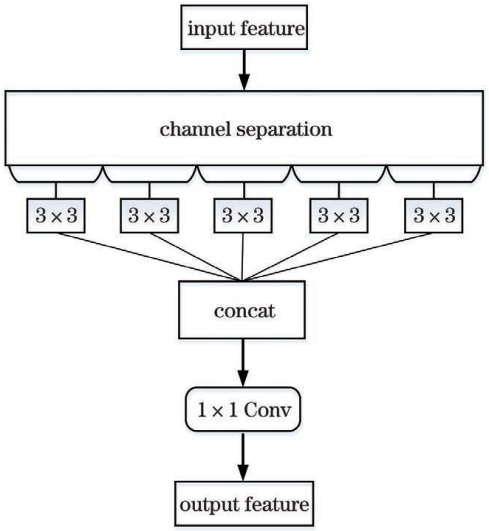

图 4 深度可分离卷积的流程图

Fig. 4 Flow chart of the depth separable convolution

2.5 特征金字塔网络

FPN[17]是一种融合不同层级特征图的方式，其结构如图 5 所示。其中，左上方为 3 个不同层级的特征图，其分辨率自下而上依次变小，特征图经 FPN 后得到右上方的融合特征。由于网络进行特征学习时图像的分辨率会逐渐减小，一般高分辨率浅层特征靠近输入端，而低分辨率深层特征靠近输出端。因此，先将骨干网络中 Block_3 产生的 1/16 大小、通道数为 1024 的特征图进行 1×1 卷积降维，使其通道数变为 256；然后对其进行 2 倍上采样，得到 FPN 中的一个分支；其次，将骨干网络中 Block_2 产生的 1/8 大小、通道数为 512 的特征图进行 1×1 卷积降维，使其通道数变为 256，得到 FPN 中的另一个分支；最后，将这两个分支的特征图以相加的方式融合。采用 FPN 可以得到更丰富的语义信息和空间信息，有效提高网络模型的预测精度。

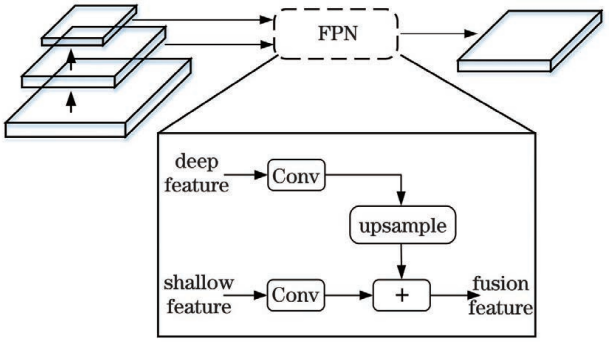

图 5 FPN 的结构

Fig. 5 Structure of the FPN

3 实验结果及分析

3.1 数据集介绍

实验采用的是 PASCAL VOC2012 增强版数据

集[18]，由语义分割标准数据集 PASCAL VOC2012[19]和 SBD 数据集合并而成，共有 21 个语义分割类别，包括 20 个前景类别和 1 个背景类别。增强版数据集中用额外标注的 10582 张图像进行训练，用 1449 张图像进行验证，用 1456 张图像进行测试。PASCAL VOC2012 数据集是目前语义分割领域中最常用的数据集，其增强版数据集的数据量更大，可使训练得到的语义分割模型表现出更强的泛化能力，逐渐得到了人们的广泛应用。实验主要采用网络模型在语义分割标准数据集 PASCAL VOC2012 验证集上的测试效果作为评价准则。

3.2 实验环境

实验程序采用主流深度学习框架 Pytorch 实现，机器的软硬件配置如表 1 所示。

表 1 机器的软硬件配置

Table 1 Hardware and software configurations of the machine

Project	Detail
CPU	Inter i7-4770k
RAM	32 G
Graphics card	TITAN V
Operating system	64-bit Ubuntu 18.04.3
CUDA	Cuda 10.2
Data processing	Pytorch 3.6

实验中采用一种 poly 学习策略[20]，学习率可表示为

$$l = l_{\text{base}}\left(1-\frac{i_{\text{c}}}{i_{\max}}\right)^{p}, \tag{1}$$

式中，i_{c} 为当前训练的迭代次数，$i_{\max}$ 为最大迭代次数，实验设置为 13 万。实验中设置的初始学习率 l_{base} 为 0.003，参数 p 为 0.9。骨干网络采用 ResNet-101，使用随机梯度下降(SGD)网络模型优化器，动量为 0.9，为了防止过拟合，将权重衰减率设为 0.0005，输入图像的大小为 513×513，每次输入网络中图像的批次大小(Batch size)为 8，1 个 Epoch 指所有数据送入网络中完成一次前向计算及反向传播的过程，总 Epoch 数设为 100。

在图像语义分割领域中，常用的性能评价指标主要包括像素准确率(PA)、平均准确率(MA)、平均交并比(mIoU)和频率加权交并比(FWIoU)。实验采用的主要评价指标是 mIoU，可表示为

$$X_{\text{mIoU}} = \left[\sum_{i=1}^{N}\frac{X_{ii}}{T_i+\sum_{j=1}^{N}(X_{ji}-X_{ii})}\right]/N, \tag{2}$$

式中，N 为图像像素的类别数量，T_i 为第 i 类的像素总数，X_{ii} 为实际类别为 i、预测类别为 i 的像素总数，X_{ji} 为实际类别为 i、预测类别为 j 的像素总数。mIoU 可表示分割结果与其真值的重合度，是目前图像语义分割领域最常用的评价指标。

3.3 实验结果对比

3.3.1 ASPP 模块中不同空洞率组合的选择

空洞卷积可以在不增加参数量的同时扩大特征图的感受野，因此，选择合适的空洞率对于网络模型来说是至关重要的。表 2 为改进 DeepLabv3+网络中 ASPP 模块在不同空洞率组合下的测试结果，其中，ASPP 模块中的卷积方式为标准卷积。由于 DeepLab 网络中 ASPP 模块的空洞率组合为 6、12、18 和 6、12、18、24，因此，将其加入对比实验。可以发现，当 ASPP 模块中的空洞率组合为 6、12、18 时，网络的 mIoU 为 79.73%；当 ASPP 模块中的空洞率组合为 6、12、18、24 时，网络的 mIoU 为 79.80%；经过骨干网络后，特征图的分辨率不断变小，仅为原始图像的 1/16，空洞率较小的空洞卷积能有效提取较小分辨率的特征图信息。当改进 DeepLabv3+网络中 ASPP 模块的空洞率组合为 4、8、12、16 时，网络的 mIoU 为 79.97%，优于其他空洞率组合下的测试结果。因此，改进 DeepLabv3+网络中 ASPP 模块选择的空洞率组合为 4、8、12、16。

表 2 ASPP 模块中不同空洞率组合下的测试结果

Table 2 Test results under different combinations of dilation rate in ASPP module unit: %

Dilation rate	mIoU
(6,12,18)	79.73
(6,12,18,24)	79.80
(4,8,12,16)	79.97

3.3.2 深度可分离卷积 ASPP 模块的训练时间和精度

在 DeepLabv3+网络中，ASPP 模块由 1×1 卷积、空洞率为 6、12、18 的空洞卷积以及全局平均池化组成。本方法将原始空洞率为 6、12、18 的空洞卷积替换为空洞率为 4、8、12、16 的空洞卷积，在不损失图像分辨率且不增加参数量的前提下增大了特征图的感受野，但多加入了一次空洞卷积也会消耗更长的训练时间。因此，将原始 ASPP 模块中的标准卷积替换为深度可分离卷积，以减少训练过程中的参数量，提高网络模型的训练效率。

表 3 为不同网络的训练时间和在 PASCAL VOC2012 验证集上的测试精度，可以发现，当 ASPP 中的卷积方式为标准卷积时，ASPP 模块由 1×1 卷积、空洞率分别为 4、8、12、16 的空洞卷积以及全局平均池化操作并联而成，网络的训练时间为 23.2 h，mIoU 为 79.97%；将 ASPP 模块中的标准卷积替换为深度可分离卷积后，ASPP 模块由 1×1 深度可分离卷积、空洞率分别为 4、8、12、16 的空洞深度可分离卷积以及全局平局池化操作并联而成，网络的训练时间变为 17.3 h，mIoU 为 79.46%。这表明深度可分离卷积在对预测结果影响不大的前提下可提高网络模型的训练效率。

表 3 不同网络的训练时间和 mIoU

Table 3 Training time and mIoU of different networks

Convolution method in ASPP	Time/h	mIoU/%
Standard convolution	23.2	79.97
Depthwise separable convolution	17.3	79.46

3.3.3 不同网络的性能对比

随着网络层数的不断加深，网络模型的参数量会逐渐增加，相应的模型也会变得越复杂。表 4 为用本方法改进的 DeepLabv3＋与 DeepLab 系列网络的参数量、模型复杂度及训练速度。其中，方法 1(Ours 1)改进的 DeepLabv3＋网络 ASPP 模块中的卷积方式为标准卷积，方法 2(Ours 2)ASPP 模块中的卷积方式为深度可分离卷积，FLOPs 为浮点运算数，GMAC 为一次乘加操作。可以发现，相比 DeepLabv3＋网络，方法 1 仅增加了 8.9%的参数量及 7.1%的模型复杂度，训练速度略低；但方法 2 却减少了 19.2%的参数量及 12.5%的模型复杂度，训练速度为 16.96 frame/s。相比 DeepLab 系列网络，本方法的模型参数量更少，训练效率更高，且本方法均能获得较好的预测结果。

表 4 不同网络的性能

Table 4 Performances of different networks

Network	Parameter /M	FLOPs / GMAC	Speed / (frame·s^{-1})
DeepLabv2	61.41	75.40	12.55
DeepLabv3	58.04	71.16	13.07
DeepLabv3＋	59.34	92.93	13.45
Ours 1	**64.65**	**99.53**	**12.66**
Ours 2	**47.97**	**81.34**	**16.96**

3.3.4 不同网络在 PASCAL VOC2012 验证集上的测试结果

DeepLabv1 网络的骨干网络为 VGG-16，且其用空洞率较大的空洞卷积获取更大的感受野，引入了全连接条件随机场作为后处理。相比 ResNet-101，VGG-16 网络中的卷积层更少，提取特征的能力较差。表 5 为不同方法在 PASCAL VOC2012 验证集上的测试结果，可以发现，DeepLabv1 在 PASCAL VOC2012 验证集上的 mIoU 为 68.70%；在 DeepLabv1 基础上改进的 DeepLabv2 网络将骨干网络 VGG-16 替换成了卷积层数更多、特征提取能力更强的 ResNet-101，且提出了由空洞率为 6、12、18、24 的空洞卷积组成的 ASPP 模块，以提取图像的多尺度特征。在不添加全连接条件随机场后处理的情况下，在 PASCAL VOC2012 验证集上的 mIoU 为 76.35%，比 DeepLabv1 网络提高了 7.65 个百分点。

表 5 不同网络在 PASCAL VOC2012 验证集上的 mIoU

Table 5 mIoU of different networks on the PASCAL VOC2012 verification set

Network	mIoU/%
DeepLabv1	68.70
DeepLabv2	76.35
DeepLabv3	77.21
DeepLabv3＋	78.85
Ours 2	**79.46**
Ours 1	**79.97**

DeepLabv3 网络在 DeepLabv2 网络的基础上改进了 ASPP 模块，将原始空洞率为 24 的空洞卷积操作替换成了 1×1 卷积和全局平均池化，改进 ASPP 模块后的 DeepLabv3 网络在 PASCAL VOC2012 验证集上 mIoU 为 77.21%；DeepLabv3＋网络在 DeepLabv3 网络的基础上，添加了一个简单且有效的解码层，利用 DeepLabv3 网络作为编码模块，并设计了一个简单有效的解码模块。在解码过程中，将 ASPP 输出的特征图卷积上采样后与骨干网络中卷积后的高分辨率低级特征图进行拼接融合。将融合后的特征图经卷积后上采样到原始图像的大小，得到预测结果。DeepLabv3＋网络在 PASCAL VOC2012 验证集上的 mIoU 为 78.85%，比 DeepLabv3 网络提高了 1.64 个百分点。对 DeepLab 系列网络进行深入研究，提出了更

有效的网络结构，将 DeepLabv3＋网络 ASPP 模块中原始空洞率为 6、12、18 的空洞卷积替换成空洞率为 4、8、12、16 的空洞卷积，提取了更丰富的多尺度特征语义信息；为了提高网络的训练效率，将 ASPP 模块中的标准卷积替换为深度可分离卷积，并设计了更有效的解码模块，将骨干网络 Block_2 和 Block_3 得到的特征图经 FPN 后作为解码层多尺度特征融合的一个重要分支，并采用逐层上采样方式进行拼接融合。可以发现，方法 2 在 PASCAL VOC2012 验证集上 mIoU 为 79.46%，比 DeepLabv3＋网络的 mIoU 提高了 0.61 个百分点；而方法 1 在 PASCAL VOC2012 验证集上的 mIoU 为 79.97%，比 DeepLabv3＋网络提高了 1.12 个百分点，这验证了本网络模型的有效性。

3.3.5　不同方法的分割结果

本方法与 DeepLabv3＋网络在 PASCAL VOC2012 验证集上的分割结果如图 6 所示，其中，图 6(a)为输入图像，图 6(b)为输入图像对应的标签图，图 6(c)为输入图像在 DeepLabv3＋网络上的分割结果，图 6(d)为方法 1 对输入图像的分割结果。

改进方法的精度与网络分割结果的好坏正相关，而网络分割结果的好坏取决于网络模型得到的分割图与其对应标签的相似程度，这也是评价网络模型好坏的基本准则。网络分割图与其对应标签的相似程度越高，表明网络的分割精度就越高。从图 6(a1)的预测结果可以发现，DeepLabv3＋网络对图像中人手握的酒瓶轮廓预测结果比较粗糙，对人头上的鸭舌帽边缘预测不够突出，且对人面部特征预测的不够具体，而本方法可以很好地解决这些问题；从图 6(a2)的预测结果可以发现，相比 DeepLabv3＋网络，本方法预测的图像中桌子和椅子的形状信息更全面；从图 6(a3)的预测结果可以发现，本方法预测的酒瓶边缘更平滑，且能成功预测图中最右侧站着的人；从图 6(a4)的预测结果可以发现，本方法可预测出边界比较清晰的摩托车把手以及比较形象的车身轮廓；从图 6(a5)的预测结果可以发现，本方法虽然将图像中的树枝预测成了鸟身体的一小部分，但其预测出了更完善的鸟尾巴轮廓以及圆滑的鸟身体边界。综上所述，改进 DeepLabv3＋网络的分割结果与对应标签的相似程度更大，即分割精度更高，具有更强的鲁棒性，同时在图像语义分割标准数据集 PASCAL VOC2012 验证集上可以预测出更优良的分割结果。

图 6　不同方法的分割结果。(a)输入图像；(b)带标签的图像；(c) DeepLabv3＋网络；(d) Ours 1

Fig. 6　Segmentation results of different methods. (a) Input image; (b) labeled image; (c) DeepLabv3+ network; (d) Ours 1

4　结　　论

提出了一种基于改进 DeepLabv3＋网络的语义分割方法，首先，通过 FPN 充分利用了骨干网络产生的高分辨率低级特征；然后，将 DeepLabv3＋网络 ASPP 模块中原始空洞率为 6、12、18 的空洞卷积替换为空洞率为 4、8、12、16 的空洞卷积，有效提取了图像中的多尺度特征信息；其次，将 ASPP 模块中的标准卷积替换为深度可分离卷积，提高了网络模型的训练效率；最后，采用逐层上采样方式有效改善了由上采样倍数过大导致的重要像素信息丢失问题。实验结果表明，改进的方法 1 在语义分割标准数据集 PASCAL VOC2012 验证集上的 mIoU 达到了 79.97%。在后续研究中，还需考虑利用深度监督损失优化网络模型，并引入注意力机制模块捕获图像的上下文依赖信息，进而更有效地提升网络模型的预测精度。

参 考 文 献

[1] Tian X, Wang L, Ding Q. Review of image semantic segmentation based on deep learning[J]. Journal of Software, 2019, 30(2): 440-468.
田萱, 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.

[2] Csurka G, Perronnin F. An efficient approach to semantic segmentation[J]. International Journal of Computer Vision, 2011, 95(2): 198-212.

[3] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[EB/OL]. (2016-06-07) [2020-10-15]. https://arxiv.org/abs/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, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-12-05) [2020-10-15]. https://arxiv.org/abs/1706.05587.

[6] 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.

[7] 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.

[8] Zhang Z H, Fang W, Du L L, et al. Semantic segmentation of remote sensing image based on encoder-decoder convolutional neural network[J]. Acta Optica Sinica, 2020, 40(3): 0310001.
张哲晗, 方薇, 杜丽丽, 等. 基于编码-解码卷积神经网络的遥感图像语义分割[J]. 光学学报, 2020, 40(3): 0310001.

[9] Le Cun Y, Bengio Y. Convolutional networks for images, speech, and time series[J]. The Handbook of Brain Theory and Neural Networks, 1995, 3361(10): 255-258.

[10] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[11] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2016-04-30) [2020-10-15]. https://arxiv.org/abs/1511.07122.

[12] Cheng X Y, Zhao L Z, Hu Q, et al. Real-time semantic segmentation based on dilated convolution smoothing and lightweight up-sampling[J]. Laser & Optoelectronics Progress, 2020, 57(2): 021017.
程晓悦, 赵龙章, 胡穹, 等. 基于膨胀卷积平滑及轻型上采样的实时语义分割[J]. 激光与光电子学进展, 2020, 57(2): 021017.

[13] Lin M, Chen Q, Yan S. Network in network[EB/OL]. (2014-03-04) [2020-10-15]. https://arxiv.org/abs/1312.4400.

[14] Cai Y, Huang X G, Zhang Z A, et al. Real-time semantic segmentation algorithm based on feature fusion technology[J]. Laser & Optoelectronics Progress, 2020, 57(2): 021011.
蔡雨, 黄学功, 张志安, 等. 基于特征融合的实时语义分割算法[J]. 激光与光电子学进展, 2020, 57(2): 021011.

[15] 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.

[16] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10) [2020-10-15]. https://arxiv.org/abs/1409.1556.

[17] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.

[18] Hariharan B, Arbelaez P, Bourdev L, et al. Semantic contours from inverse detectors[C]//2011 IEEE International Conference on Computer Vision (ICCV), November 6-13, 2011, Barcelona, Spain. New York: IEEE Press, 2011: 991-998.

[19] Everingham M, Eslami S M A, Gool L, et al. The pascal visual object classes challenge: a retrospective[J]. International Journal of Computer Vision, 2015, 111(1): 98-136.

[20] Liu W, Rabinovich A, Berg A C. Parsenet: looking wider to see better[EB/OL]. (2015-11-19)[2020-10-15]. https://arxiv.org/abs/1506.04579.