文章编号：1005-5630(2023)02-0046-09 **DOI**: 10.3969/j.issn.1005-5630.2023.002.006

一种双分支结构的图像语义分割算法

王 兵，瑚 琦，卞亚林
（上海理工大学 光电信息与计算机工程学院，上海 200093）

摘要：图像语义分割需要精细的细节信息和丰富的语义信息，然而在特征提取阶段，连续下采样操作会导致图像中物体的空间细节信息丢失。为解决该问题，提出一种双分支结构语义分割算法，在特征提取阶段既能有效获取丰富的语义信息又能减少物体细节信息的丢失。该算法的一个分支使用浅层网络保留高分辨率细节信息有助于物体的边缘分割，另一个分支使用深层网络进行下采样获取语义信息有助于物体的类别识别，再将两种信息有效融合可以生成精确的像素预测。通过 Cityscapes 数据集和 CamVid 数据集上的实验验证，与现有语义分割算法相比，所提算法在较少的参数条件下，获得了较好的分割效果。

关键词：图像语义分割；双分支结构；细节信息；语义信息
中图分类号：TP 391 **文献标志码**：A

An image semantic segmentation algorithm with a two-branch structure

WANG Bing，HU Qi，BIAN Yalin
（School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China）

Abstract: Image semantic segmentation requires fine detail information and rich semantic information, but in the stage of feature extraction, continuous down-sampling operation will lead to the loss of spatial details of objects in the image. To solve this problem, a semantic segmentation algorithm based on double-branch structure is proposed, which can obtain rich semantic information effectively and reduce the loss of object details in feature extraction stage. One branch of the algorithm uses shallow network to retain high-resolution detail information which is helpful for object edge segmentation, and the other branch uses deep network for downsampling to obtain semantic information which is helpful for object category recognition, and then the effective fusion of the two kinds of information can generate accurate pixel prediction. Experimental results on Cityscapes and CamVid datasets show that the proposed algorithm achieves better segmentation performance under fewer parameters than existing semantic segmentation algorithms.

收稿日期：2022-04-18
基金项目：国家自然科学基金(61975125)
第一作者：王 兵 (1994—)，男，硕士研究生，研究方向为语义分割。E-mail：1812861208@qq.com
通信作者：瑚 琦 (1977—)，男，副教授，研究方向为光电检测、机器视觉。E-mail：hq_0519@163.com

Keywords: image semantic segmentation; double branch structure; detail information; semantic information

引　言

图像语义分割是计算机视觉的重要研究内容之一，其研究发展经历了 3 个阶段[1]：传统方法的研究阶段、传统方法和深度学习相结合的研究阶段、基于深度学习的研究阶段。2015 年 Long 等[2] 提出了全卷积网络(FCN)，开创性地将分类网络 VGG16[3] 中的全连接层改编成卷积层，并在浅层特征和深层特征之间采用跳跃连接，显著改善了语义分割的性能。FCN 方法的提出促使深度学习广泛应用于语义分割领域，大量基于卷积神经网络(CNN)的语义分割算法随后相继出现。

这些算法各有特点，针对性较强，其中的一个重点是改善连续下采样操作导致图像中物体细节信息丢失的问题。例如，Noh 等[4] 构建编码器–解码器网络 DeconvNet，在解码器中通过反池化和反卷积操作捕获物体更精细的细节信息，能够解决物体的详细结构丢失的问题。Badrinarayanan 等[5] 提出对称的编码器–解码器网络 SegNet，编码器中采用池化索引存储像素的位置信息，解码器中使用相应编码器的池化索引执行上采样，进而改善了物体的边缘分割。Yu 等[6] 在 Dilation10 网络中使用空洞卷积进行特征提取，在不降低特征图分辨率的同时扩大了感受野。Chen 等提出的 DeepLab[7] 和 DeepLabv2[8]网络使用全连接条件随机场(CRF)对分割结果进行后处理，提高了模型捕捉精细边缘细节的能力。Lin 等[9] 提出的 RefineNet 语义分割网络，为了充分利用下采样阶段的每一层特征，在低级特征和高级特征之间建立多个远程连接，用细粒度的低级特征细化低分辨率的语义特征。Nekrasov等[10] 在 RefineNet 的基础上用内核较小的卷积替代内核较大的卷积构建 RefineNet-LW 网络，有效降低了模型参数量，同时保持性能基本不变。Paszke 等[11] 提出的 ENet 网络和 Treml 等[12] 提出的 SQ 网络均采用轻量级的编码器–解码器结构进行实时语义分割，减少了模型的参数量却降低了模型的性能。Pohlen 等[13] 构建全分辨率残差网络 FRRN，残差流中以完整图像分辨率携带信息来实现精确的边界分割。

对图像进行特征提取时，池化层和跨步卷积有效增加了感受野，虽然有利于获取语义信息，但减小了图像分辨率，导致物体细节信息丢失。上述算法为解决物体细节信息丢失的问题，采用了编解码器网络结构、跳跃连接或 CRF 后处理等方法，但是这些算法结构仍然冗余繁杂，导致网络参数量大幅增加。

为了在特征提取阶段既能有效获取丰富的语义信息又能减少物体细节信息的丢失，同时尽可能降低网络的参数量，本文提出一种双分支网络模型，通过两个分支分别获取物体的细节信息和语义信息。其中使用浅层网络分支来保留图像中的细节信息，生成高分辨率特征，使用深层网络分支进行下采样获取语义信息。浅层网络分支能有效减少细节信息丢失，提高像素定位的准确性；深层网络分支采用轻量级主干网络下采样，既能提取语义信息又能降低模型参数量和计算量。最后，将两个分支获取的特征信息有效融合，进而提升网络分割的性能。本文主要工作为：1)提出一种双分支结构语义分割算法，建立细节分支(detail branch，DB)和上下文分支(context branch，CB)分别有效获取细节信息和语义信息；2)构建融合模块(fusion module，FM)，将得到的低级细节信息和高级语义信息进行有效融合；3)在 Cityscapes 数据集和 CamVid 数据集上验证了所提算法的有效性，分别获得 67.4%、58.5% 的均交并比。

1　背景知识

得益于大量已标注的公开数据集和不断提高的计算机性能，深度学习获得快速发展，研究学者将各种深度学习技术应用在不同的计算机视觉任务中并改善了相应任务的性能。本文提出的双分支结构语义分割算法涉及密集连接、注意力机制等重要概念，下面对其进行简述。

1.1　密集连接

密集连接网络(DenseNet)由 Huang 等[14]在 2017 年的 CVPR(Computer Vision and Pattern Recognition)会议上提出。DenseNet 主要由密集模块组成，该模块中，对于每一层，所有先前层的特征图用作其输入，该层的特征图用作所有后续层的输入，这种密集连接的方式改善了网络中各层之间的信息流动并有效实现了特征重用。密集模块的结构图如图 1 所示，该结构一经提出就在图像分类任务上展现出了非常出色的结果。鉴于密集连接能有效实现特征重用的优点，本文在细节分支中构建一个密集模块来保留物体更多的空间细节信息。

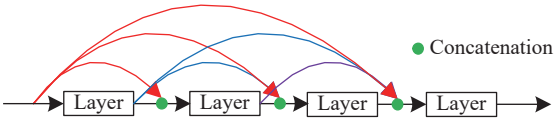


图 1　密集模块结构图

Fig. 1　Dense module structure

1.2　注意力机制

注意力机制的主要功能是对特征图进行特征细化，以增强信息丰富的特征并抑制无用的特征。注意力模块具有结构简单和轻量级的特点，可以很容易地嵌入 CNN 相关网络中，因此注意力机制被广泛应用在计算机视觉的许多任务上，如图像分类[15]、语义分割[16]等。通过注意力模块，网络在训练学习过程中能获取到特征图每个空间位置或每个通道的权重，按照权重大小完成对不同空间位置或通道的增强或抑制，从而自适应地重新校准特征图。

2　双分支网络结构设计

双分支网络整体的设计结构图如图 2 所示。图 2(a)为整体的网络结构，在输入图像之后即分成细节分支和上下文分支两部分，完成各自特征提取后通过融合模块进行特征融合，最后经过分类层(classifier)处理得到分割结果图。图 2(b)

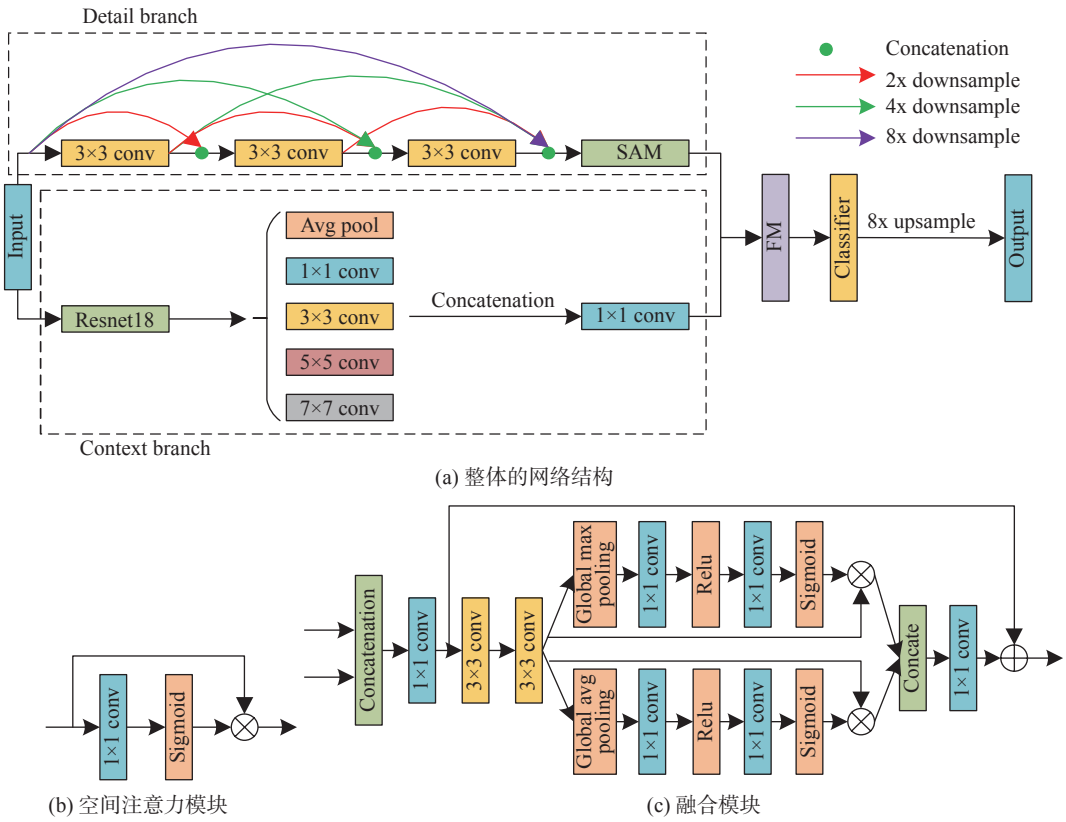


图 2　双分支网络结构图

Fig. 2　Two-branch network structure

为空间注意力模块（spatial attention module，SAM），图 2(c)为融合模块。

2.1　细节分支

该分支由一个密集模块构成，密集模块中包括 3 个 3×3 的卷积和一个空间注意力模块，这种设计既能编码物体的细节信息，以提取物体的边缘轮廓特征，又能对提取的特征进行细化。3 个卷积和空间注意力模块之间采用密集方式连接，后续层可以使用所有先前层的特征，从而有效实现特征重用并弥补细节信息的丢失，保留物体更多的空间细节特征。为尽可能减少网络参数量，以降低网络复杂度，密集模块只使用 3 个步长为 2 的 3×3 卷积对图像进行特征提取，每个卷积后面执行批标准化（BN）[17] 和 ReLU 操作。

在细节分支的末端使用空间注意力模块，能从空间维度细化所有先前层提取的特征，为特征图的每个空间位置施加一个权重以重新校准特征图，完成对不同空间位置的增强或抑制，进而得到有效的特征信息。最终，细节分支得到的特征图的分辨率为原图的 1/8，能保留图像的高分辨率细节信息。为进一步减少网络参数量，空间注意力模块只使用一个 1×1 的卷积计算空间注意力权重，其结构图如图 2(b)所示。利用空间注意力模块重新校准特征图时，对于给定的输入特征图 X，其空间注意力权重的计算和重新校准过程可表示为

$$\alpha = \sigma\left(f^{1\times1}(X)\right) \tag{1}$$

$$X_{\mathrm{SA}} = f_{\mathrm{SA}}(X,\alpha) \tag{2}$$

式中：σ 为 sigmoid 函数；$f^{1\times1}$ 表示卷积核大小为 1×1 的卷积操作；α 为空间注意力权重；$f_{\mathrm{SA}}(.)$ 表示输入特征图和相应空间注意力权重相乘；X_{SA} 为校准后的特征图。

执行卷积和细化操作过程中，当特征图尺寸不匹配时，采用平均池化操作对先前层的特征图进行下采样，如图 2(a)中彩色箭头所示。

2.2　上下文分支

该分支使用轻量级网络 ResNet18[18] 作为主干网络，将特征图下采样到原图的 1/32，从而获得丰富的语义信息。语义分割任务中，图像中往往存在多种尺度的物体，然而固定大小的感受野会导致物体分类错误。虽然 DeepLabv3[19] 网络中的空洞空间金字塔池化（ASPP）模块能有效获取多尺度上下文信息，但是该模块中使用空洞卷积导致图像中的像素不能全部用于计算且不利于小尺寸物体的分割[20]。与之不同，本文在主干网络的顶端并行使用全局平均池化和 1×1、3×3、5×5、7×7 的卷积以获取全局和局部上下文信息，一方面，使用标准卷积能充分利用所有像素用于计算，避免特征信息的丢失；另一方面，主干网络提取的特征图分辨率较低，使用较大的卷积获取多尺度上下文信息不会大幅增加网络的计算量。随后，将获得的多尺度上下文信息进行级联，再经过 1×1 的卷积对信息进行进一步融合同时缩减特征图的通道数来减少参数量。最后将特征图上采样到原图的 1/8，以匹配细节分支特征图的大小。

2.3　融合模块

细节分支得到的高分辨率空间细节信息有助于物体边缘的分割，上下文分支得到的低分辨率语义信息有助于物体类别识别，将二者获取的特征信息通过融合模块进行融合互为补充，以实现更好的分割效果。融合模块的结构图如图 2(c)所示，该模块由级联操作、一个 1×1 的卷积和残差通道注意力模块构成。

为了更有效捕获特征图通道之间的相互依赖性，在残差通道注意力模块中使用全局平均池化和全局最大池化两种池化方式，分别沿空间轴压缩特征图计算通道注意力权重来重新校准特征图。利用通道注意力模块重新校准特征图时，对于给定的输入特征图 X，使用全局平均池化聚合空间信息时，其通道注意力权重的计算和重新校准过程可表示为

$$\beta_{\mathrm{avg}} = \sigma(W_2\delta(W_1(GlobalAvgPool(X)))) \tag{3}$$

$$X_{\mathrm{avg}} = f_{\mathrm{CA}}\left(X,\beta_{\mathrm{avg}}\right) \tag{4}$$

使用全局最大池化聚合空间信息时，其通道注意力权重的计算和重新校准过程可表示为

$$\beta_{\max} = \sigma(W_2\delta(W_1(GlobalMaxPool(X)))) \tag{5}$$

$$X_{\max} = f_{CA}(X, \beta_{\max}) \tag{6}$$

式中：*GlobalAvgPool* 为全局平均池化；*GlobalMaxPool* 为全局最大池化；δ 为 ReLU 函数；σ 为 sigmoid 函数；W_1 和 W_2 为两个卷积层的权值矩阵；β_{avg} 和 $\beta_{\max}$ 为通道注意力权重；$f_{CA}(.)$ 表示输入特征图和相应通道注意力权重相乘；X_{avg} 和 $X_{\max}$ 为校准后的特征图。

使用融合模块进行特征融合时，为保留更多提取的原始特征，首先将不同级别的信息进行级联，并通过 1×1 的卷积对信息进行融合。然后，为了保证信息融合的有效性，采用残差通道注意力模块为特征图的每个通道施加一个权重以重新校准融合后的特征图，有助于网络在训练学习过程中关注信息丰富的通道。

2.4　分类层

分类层采用一个 3×3 的卷积和一个 1×1 的卷积。其中，3×3 的卷积作用是对残差通道注意力模块生成的特征图进行特征融合；1×1 的卷积作用是将特征图的通道数映射为物体类别数，并得到最终的网络预测图。

3　实　验

为评估双分支结构语义分割算法的有效性，选取最常用的 Cityscapes[21] 数据集和 CamVid[22] 数据集作为运算实验对象。这里所有运算实验均在 Ubuntu18.04 操作系统上进行，实验的软件环境为 pytorch1.2，cuda10.0，cudnn7.6.5，硬件环境采用 2 块 GTX 1080Ti GPU 加速。

3.1　数据集

Cityscapes 是城市街道场景大型数据集，拍摄于 50 个不同的城市。该数据集共有 5000 张精细标注的图片和 20000 张粗略标注的图片，本文所有实验仅使用精细标注的图片。精细标注的图片划分为训练集、验证集和测试集，分别包含 2975、500 和 1525 张图片，所有图片的分辨率为 2 048×1024。像素标注包括 30 个类别物体，其中 19 个类别用于训练和评估。

CamVid 是基于视频序列的街道场景数据集，该数据集共有 701 张图片和 11 个语义类别的像素标注。训练集、验证集和测试集分别包含 367、101 和 233 张图片，所有图片的分辨率为 480×360。

训练网络时防止出现过拟合现象，需对数据集进行增强处理，包括随机水平翻转、随机旋转和随机缩放，其中缩放尺度为{0.5，0.75，1.0，1.25，1.5，1.75，2.0}，最后将图片随机裁剪至固定大小进行训练。

3.2　实验参数设置和评价指标

实验中合理调节学习率的大小有利于网络的训练，例如，Chen 等[8] 采用 poly 学习率策略调节学习率大小，网络每训练完一个 iteration，都会对学习率进行衰减，直到网络训练完成时学习率下降为 0。学习率的迭代更新表达式为

$$lr = baselr \times \left(1 - \frac{iter}{max\_iter}\right)^{power} \tag{7}$$

式中：*iter* 为当前迭代次数；*max_iter* 为总迭代次数；*lr* 为当前迭代次数的学习率；*baselr* 为初始学习率；*power* 的值设置为 0.9。这种方法通过动态调整学习率的大小，能使网络在训练时收敛得更好，因此本文所有实验也采用此种方法调节学习率的大小。实验参数详细信息如表 1 所示。

表 1　实验参数设置

Tab. 1　Experimental parameter settings

数据集	批处理大小	初始学习率	权重衰减系数	裁剪大小	优化函数
Cityscapes	10	0.01	0.0005	1024×1024	SGD
CamVid	8	0.001	0.0001	480×360	Adam

为了定量评估所提算法的分割精度，选取均交并比（mean intersection over union，mIoU）[23] 作为评价指标。该评价指标是真实标签值和网络预测值两个集合的交集与并集之比，其计算表达式为

$$mIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij} + \sum_{j=0}^{k}p_{ji} - p_{ii}} \tag{8}$$

式中：k 为像素标签类别数；$k+1$ 为包含空类

或背景在内的总类别数；p_{ii}表示正确分类的像素数量；p_{ij}表示应为类别i但被预测为类别j的像素数量；p_{ji}表示应为类别j但被预测为类别i的像素数量。另外，在实验中还使用参数量（Parameters）来评估不同网络结构的复杂度。

3.3　消融实验及结果分析

为验证本文提出的双分支网络结构中上下文分支、细节分支和融合模块的有效性，需要进行消融实验，所有消融实验均在 Cityscapes 验证集上进行评估验证。消融实验的 epoch 均设置为 300，所得实验结果如表 2 所示，表中列出了不同情况下使用基准模型（baseline）、上下文分支、细节分支和融合模块进行实验所得分割精度和参数量。

表 2　消融实验结果

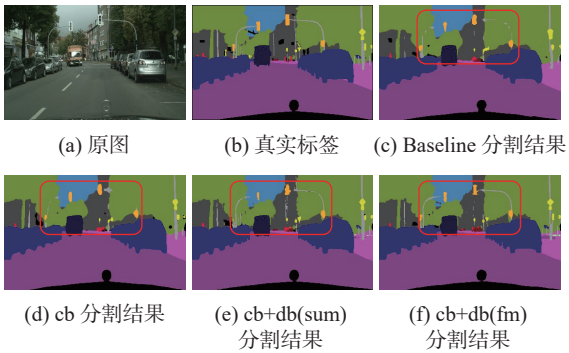
Tab. 2　Ablation experiment results

Method	mIoU/%	Parameters/10^6
baseline	57.6	11.8
CB	59.4	12.7
CB+DB（sum）	61.5	13.1
CB+DB（FM）	62.5	13.3

实验中首先选取残差网络 ResNet18 作为基准模型。使用残差网络 ResNet18 作为主干网络进行特征提取，提取的特征图经过分类层处理，将网络输出特征图进行 32 倍上采样得到原图大小。从表 2 可知，其在 Cityscapes 验证集上的精度为 57.6%。

当只使用上下文分支进行特征提取时网络的性能从 57.6% 提升至 59.4%，表明多尺度上下文信息有助于不同尺度物体的分割；当同时使用上下文分支和细节分支进行特征提取时，将两个分支获得的特征进行简单相加（sum），网络的性能从 59.4% 提升至 61.5%，从而验证细节分支保留的高分辨率细节信息有助于分割性能的提升；当使用融合模块将两个分支提取的特征进行融合时，网络的性能从 61.5% 提升至 62.5%，相比简单相加不同级别的信息，通过融合模块进行有效融合更有利于改善网络的性能。通过以上消融实验表明，同时使用细节分支、上下文分支和融合模块能达到最佳分割效果，相比基准模型，网络的参数量只有略微增加。

图 3 展示了不同结构在 Cityscapes 验证集上的部分可视化图，其中（a）为原始图像；（b）为真实标签；（c）为基准模型分割图；（d）为只使用上下文分支所得分割图；（e）为同时使用细节分支和上下文分支所得分割图，其中两个分支的特征只是简单相加；（f）为同时使用细节分支、上下文分支和融合模块所得分割图。从图 3 可以看出，随着细节分支的加入，分割效果越来越好，例如交通信号灯（图中红框部分），这也表明设计双分支网络结构可以更好地保留物体的细节信息，从而获得更好的分割效果。



(a) 原图　(b) 真实标签　(c) Baseline 分割结果

(d) cb 分割结果　(e) cb+db(sum) 分割结果　(f) cb+db(fm) 分割结果

图 3　不同结构可视化图

Fig. 3　Visualization of different structures

3.4　同其他算法的对比与分析

对于 Cityscapes 数据集，为实现更好的分割效果，将 epoch 增加至 500。由于 Cityscapes 测试集没有提供真实标签，因此需将预测标签图提交至 Cityscapes 官方网站（https://www.cityscapes-dataset.com）进行评估，才能得到测试集上的分割精度。为了更好地体现本文所提双分支结构语义分割算法的有效性，选取 SegNet[5]、ENet[11]、SQ[12]、FRRN A[13]、DeepLab[7]、FCN-8s[2]、Dilation10[6]、DeepLabv2[8]、RefineNet-LW[10]、RefineNet[9] 等算法与本文算法进行性能对比，对比结果如表 3 所示。

从表 3 可知，所提双分支结构语义分割算法相比其他大部分算法在均交并比上有所提升，表明细节分支保留的细节信息提高了像素定位的准确性，有助于目标物体边缘轮廓的分割，从而达到了更好的分割效果。DeepLabv2、RefineNet-

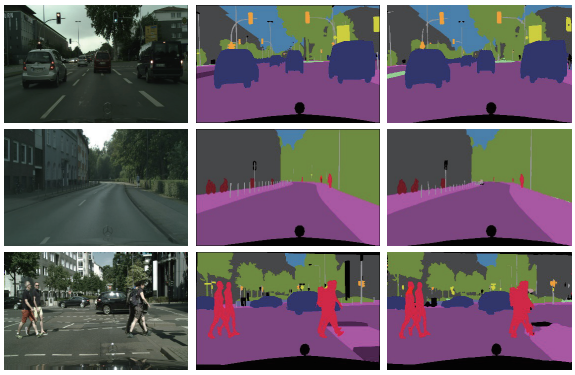
表 3　不同算法在 Cityscapes 测试集上的分割精度

Tab. 3　Segmentation accuracy of different algorithms on the Cityscapes test set

算法	预训练	主干网络	分辨率	均交并比/%	参数量/10^6
SegNet	Yes	VGG16	360×640	56.1	29.5
ENet	No	No	512×1 024	58.3	0.4
SQ	Yes	SqueezeNet[24]	1 024×2048	59.8	—
FRRN A	No	No	256×512	63.0	17.7
DeepLab	Yes	VGG16	512×1 024	63.1	37.3
FCN-8s	—	VGG16	1 024×2 048	65.3	134.5
Dilation10	Yes	VGG16	1 024×2048	67.1	—
DeepLabv2	Yes	ResNet101	1 024×2048	70.4	44.0
RefineNet-LW	Yes	ResNet101	—	72.1	46.0
RefineNet	—	ResNet101	1 024×2048	73.6	118.0
Ours	No	ResNet18	1 024×2048	67.4	13.3

LW 和 RefineNet 等算法取得了优越的均交并比，一方面是使用了结构复杂的 ResNet101 作为主干网络，其特征提取能力更强；另一方面是在网络中使用了跳跃连接或 CRF 后处理等方法，改善了下采样过程导致细节信息丢失的问题；然而这三种算法参数量较大，其中 RefineNet 的参数量是本文所提算法的 9 倍。表 3 中 ENet 参数量最少，但以牺牲精度为代价，本文所提算法与其相比，参数量有所增加，但均交并比提升了 9.1%。综合对比，本文算法以较少参数量实现了较好的分割效果。图 4 展示了所提算法在 Cityscapes 验证集上的部分可视化图。

对于 CamVid 数据集，使用训练集和验证集的图片一起训练模型，epoch 设置为 1 000，所提算法在 CamVid 测试集上的分割精度如表 4 所示。表 4 中对比了 DeconvNet[4]、ENet[11]、SegNet[5]、FCN-8s[2]、BiSeNet[25]、文献 [26] 等算法与所提双分支结构语义分割算法的分割性能。从对比结果可知，本文算法相比大部分对比算法在均交并比上有所提高，进一步验证了双分支结构的有效性，而模型参数量仅比 ENet 有所增加。虽然 BiSeNet 和文献 [26] 等算法获得了较高的均交并比，但是其网络结构复杂，参数量较大。



(a) 原图　(b) 真实标签　(c) 预测图

图 4　所提算法可视化图

Fig. 4　Visualization of the proposed algorithm

4　结　论

本文提出一种双分支结构语义分割算法，在特征提取阶段既能有效获取丰富的语义信息又能保留图像的高分辨率细节信息，将得到的细节信息和语义信息通过融合模块有效融合以充分利用不同级别的信息，进而改善语义分割的性能。在 Cityscapes 数据集和 CamVid 数据集上验证了所提算法的有效性，分别获得 67.4% 和 58.5% 的均交并比，与大部分现有算法相比，所提算法

表 4　不同算法在 CamVid 测试集上的分割精度

Tab. 4　Segmentation accuracy of different algorithms on the CamVid test set

Method	DeconvNet	ENet	SegNet	FCN-8s	BiSeNet	文献[26]	Ours
Building	—	74.7	88.8	77.8	83.0	—	75.8
Tree	—	77.8	87.3	71.0	75.8	—	66.6
Sky	—	95.1	92.4	88.7	92.0	—	89.5
Car	—	82.4	82.1	76.1	83.7	—	77.0
Sign	—	51.0	20.5	32.7	46.5	—	35.1
Road	—	95.1	97.2	91.2	94.6	—	93.2
Pedestrian	—	67.2	57.1	41.7	58.8	—	39.6
Fence	—	51.7	49.3	24.4	53.6	—	26.9
Pole	—	35.4	27.5	19.9	31.9	—	17.0
Sidewalk	—	86.7	84.4	72.7	81.4	—	78.5
Bicyclist	—	34.1	30.7	31.0	54.0	—	44.5
mIoU/%	48.9	51.3	55.6	57.0	68.7	69.1	58.5
Parameters/10^6	252	0.4	29.5	134.5	49.0	62.0	13.3

分割精度有所提高且参数量使用较少。然而，相比轻量级网络模型，所提双分支网络模型参数量还有继续减少的空间；相比均交并比较高的网络模型，本文网络模型分割精度有待提高，如何权衡网络的复杂度和分割精度将是后续的一个研究方向。

参考文献：

[1] 王嫣然, 陈清亮, 吴俊君. 面向复杂环境的图像语义分割方法综述 [J]. 计算机科学, 2019, 46(9): 36 – 46.

[2] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015: 3431 − 3440.

[3] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//Proceedings of 3rd International Conference on Learning Representations. San Diego, 2015: 1 − 14.

[4] NOH H, HONG S, HAN B. Learning deconvolution network for semantic segmentation[C]//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015: 1520 − 1528.

[5] 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.

[6] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[C]//Proceedings of the 4th International Conference on Learning Representations. San Juan: ICLR, 2016: 1 − 13.

[7] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[C]//Proceedings of the 3rd International Conference on Learning Representations. San Diego: ICLR, 2015.

[8] 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.

[9] LIN G S, MILAN A, SHEN C H, et al. RefineNet: multi-path refinement networks for high-resolution semantic segmentation[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 5168 − 5177.

[10] NEKRASOV V, SHEN C H, REID I. Light-weight RefineNet for real-time semantic segmentation[C]// British Machine Vision Conference 2018. Newcastle: BMVC, 2018.

[11] PASZKE A, CHAURASIA A, KIM S, et al. ENet: a deep neural network architecture for real-time semantic segmentation[J]. arXiv preprint arXiv:, 1606, 02147: 2016.

[12] TREML M, ARJONA-MEDINA J, UNTERTHINER T, et al. Speeding up semantic segmentation for autonomous driving[C]//Proceedings of the 29th Conference on Neural Information Processing Systems Workshop. Barcelona: MIT Press, 2016: 1 − 7.

[13] POHLEN T, HERMANS A, MATHIAS M, et al. Full-resolution residual networks for semantic segmentation in street scenes[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 3309 − 3318.

[14] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 2261 − 2269.

[15] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7132 − 7141.

[16] LI H F, QIU K J, CHEN L, et al. SCAttNet: semantic segmentation network with spatial and channel attention mechanism for high-resolution remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 18(5): 905 – 909.

[17] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//Proceedings of the 32nd International Conference on Machine Learning. Lille: JMLR. org, 2015: 448 − 456.

[18] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 770 − 778.

[19] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:, 1706, 05587: 2017.

[20] 邝辉宇, 吴俊君. 基于深度学习的图像语义分割技术研究综述 [J]. 计算机工程与应用, 2019, 55(19): 12 – 21,42.

[21] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 3213 − 3223.

[22] BROSTOW G J, FAUQUEUR J, CIPOLLA R. Semantic object classes in video: a high-definition ground truth database[J]. Pattern Recognition Letters, 2009, 30(2): 88 – 97.

[23] 田启川, 孟颖. 卷积神经网络图像语义分割技术 [J]. 小型微型计算机系统, 2020, 41(6): 1302 – 1313.

[24] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. Squeezenet: alexnet-level accuracy with 50x fewer parameters and <0.5MB model size[J]. arXiv preprint arXiv:, 1602, 07360: 2016.

[25] YU C Q, WANG J B, PENG C, et al. BiSeNet: bilateral segmentation network for real-time semantic segmentation[C]//Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018: 334 − 349.

[26] 陈劲宏, 陈玮, 陈舒曼. 单块嵌入式 GPU 下对街景图像的实时分割研究 [J]. 控制工程, 2021, 28(11): 2165 – 2173.

（编辑：张　磊）