DOI:10.13203/j.whugis20200305 文章编号:1671-8860(2022)07-1071-10

结合空洞卷积的FuseNet变体网络高分辨率遥感影像语义分割

杨 军[1,2,3,4] 于茜子[2,3,4]

1 兰州交通大学电子与信息工程学院,甘肃 兰州,730070
2 兰州交通大学测绘与地理信息学院,甘肃 兰州,730070
3 地理国情监测技术应用国家地方联合工程研究中心,甘肃 兰州,730070
4 甘肃省地理国情监测工程实验室,甘肃 兰州,730070

摘 要:针对多模态、多尺度的高分辨率遥感影像分割问题,提出了结合空洞卷积的FuseNet变体网络架构对常见的土地覆盖对象类别进行语义分割。首先,采用FuseNet变体网络将数字地表模型(digital surface model,DSM)图像中包含的高程信息与红绿蓝(red green blue,RGB)图像的颜色信息融合;其次,在编码器和解码器中分别使用空洞卷积来增大卷积核感受野;最后,对遥感影像逐像素分类,输出遥感影像语义分割结果。实验结果表明,所提算法在国际摄影测量与遥感学会(International Society for Photogrammetry and Remote Sensing,ISPRS)提供的Potsdam、Vaihingen数据集上的m_{F1}得分分别达到了91.6%和90.4%,优于已有的主流算法。

关键词:高分辨率遥感影像;深度卷积神经网络;空洞卷积;语义分割;FuseNet

中图分类号:P237 **文献标志码**:A

遥感影像语义分割是遥感影像信息获取的关键环节,也是近年来的研究热点,相关研究成果已广泛应用于土地利用变化检测、交通监测和灾害预警评估等方面[1]。高分辨率遥感影像能够表现丰富的地物信息,有利于提取地物的复杂特征以识别复杂的人造目标。

传统的遥感图像语义分割主要是通过提取图像的低级特征进行分割,分割结果缺乏语义标注。文献[2]通过随机森林分类器提取语义特征进行语义分割。文献[3]利用Logistic回归分类器提取颜色、纹理特征,通过条件随机场(conditional random field,CRF)模型训练实现语义分割。然而,传统的遥感图像语义分割方法对特征的提取和表达,需要依靠先验知识进行人工选择和设计,并且在建立相应语义分割模型的过程中,人工设计的特征和高层语义特征之间存在差距,因此建立的语义分割模型泛化能力较差。

随着深度学习理论的发展与普及,深度神经网络模型已广泛应用于不同行业[4]。研究者在遥感影像分析处理中应用深度学习方法,取得了较为理想的效果[5-6]。全卷积网络[7](fully convolutional networks,FCN)和SegNet网络[8]在高分辨率遥感影像语义分割中展现出了较为优异的性能与分割效果,但FCN对像素进行分类时没有考虑到像素之间的关系,忽略了基于像素分类的空间规划步骤,缺乏空间一致性。SegNet的基本网络结构为编码器-解码器,编码器对图像进行高维特征提取和下采样,解码器对提取的特征图进行上采样操作,因此编码器-解码器结构可以以1:1的分辨率进行像素预测,但上采样的过程中易丢失细节信息,使得小目标地物的分割效果较差。文献[9]分别提取红绿蓝(red green blue,RGB)信息和数字地表模型(digital surface model,DSM)信息,并将它们融合集成到SegNet结构中进行语义分割,获得高分辨率的多模态预测RGB-DSM数据用于异构数据源的联合学习。然而该融合策略无法平衡高程信息和颜色信息,导致图像分割不准确。因此,本文针对高分辨率遥感影像中多模态数据融合效果不佳、边缘分割效果不理想、类边界模糊和易产生误分割现象等问题,受

收稿日期:2020-09-24

项目资助:国家自然科学基金(61862039);甘肃省科技计划(20JR5RA429);2021年度中央引导地方科技发展资金(2021-51);兰州交通大学优秀平台支持项目(201806)。

第一作者:杨军,博士,教授,博士生导师,主要从事计算机图形学、数字图像处理和地理信息系统等方面的研究。 yangj@mail.lzjtu.cn

编码器-解码器和文献[6]中FuseNet网络结构的启发，对FuseNet网络结构进行改进，提出了一种结合空洞卷积的FuseNet变体网络（improved FuseNet with atrous convolution-convolutional neural network，IFA-CNN）模型。在编码器部分，提出虚拟融合单元来提高遥感影像语义分割效果；针对遥感影像提取特征部分，引入空洞卷积调整感受野捕获遥感影像多尺度信息，提高目标分割效果；在解码器部分，链接编码器并提取融合特征，以提高网络鲁棒性。

1 高分辨率遥感影像语义分割模型

1.1 多模态遥感数据融合

文献[10]中，FuseNet采用了编码器-解码器结构将二维图像数据融合。FuseNet架构如图1所示，其中，pooling为池化操作，conv为卷积操作，unpooling为反池化操作，IRRG为近红外、红外和绿波段，Ⓒ为融合操作。图1中使用了两个编码器对RGB和DSM进行联合编码，首先将编码后的特征图输入到解码器中进行上采样，然后由分类器进行弱分类，通过softmax得到最终分割结果。同时，FuseNet选择深度信息作为辅助特征进行多模态数据融合，如图2(a)所示，其中，aux为辅助分支，main为主分支，mix为虚拟融合操作。但FuseNet在进行多模态数据融合时，DSM分支与RGB分支存在不对称，使得DSM分支仅提取深度特征，RGB分支需要提取DSM与RGB数据的融合。此外，这种不对称的融合方案导致在解码过程中只使用主分支编码时的索引进行上采样，在一定程度上会影响遥感影像的分割效果。

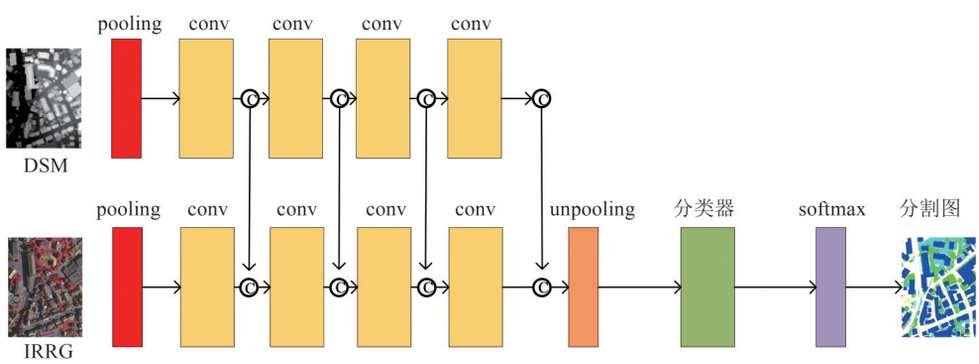
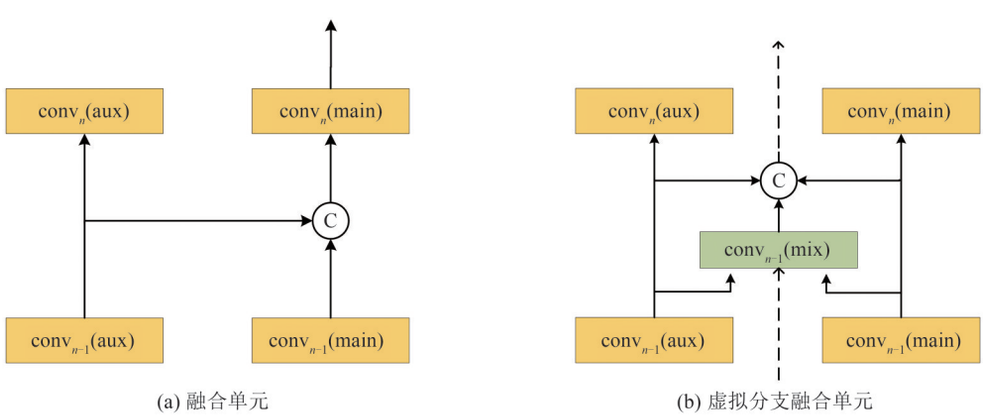


图1 用于遥感数据融合的FuseNet架构[9]

Fig.1 FuseNet Architecture for Fusion of Remote Sensing Data [9]



(a) 融合单元 (b) 虚拟分支融合单元

图2 多模态数据融合策略

Fig.2 Multimodal Data Fusion Strategy

为了更好地提取RGB-DSM图像的特征，解决主数据源及辅助数据源数据分配不均的问题，本文提出了一种虚拟分支融合单元，对主数据源和辅助数据源进行一次卷积运算，从而产生一种虚拟模态。将该虚拟模态作为融合数据源之一，将DSM分支提取的特征和RGB分支提取的特征进行融合。如图2(b)所示，通过这种方法调整FuseNet结构，使其在一定程度上可以解决对主数据源和辅助数据源进行选择的问题，以解决数据处理不均衡的问题。另外，为解决解码过程中只使用主分支编码时产生的索引进行上采样的问题，本文将虚拟分支融合单元中最大池化操作

产生的索引应用于解码阶段的上采样,从而提高语义分割的精度。

1.2　多尺度空洞卷积

空洞卷积[11]是在不减少图像尺寸的同时获得比较大的感受野,所以其主要优势在于允许灵活地调整感受野的大小来捕获多尺度信息,提高多目标分类和分割任务的性能[12]。二维空洞卷积算子定义为:

$$g_{i,j}(x_\ell)=\sum_{c=0}^{C_\ell}\theta_{k,r}^{i,j}*x_\ell^c \tag{1}$$

式中,$g_{i,j}$是对输入特征图的卷积操作$\mathbf{R}^{H_\ell\times W_\ell\times C_\ell}\rightarrow\mathbf{R}^{H_{\ell+1}\times W_{\ell+1}}$;$*$表示卷积算子;$x_\ell\in\mathbf{R}^{H_\ell\times W_\ell\times C_\ell}$为在第$i$行和第$j$列中属于通道$c\in\{0,1,2\cdots C_\ell\}$的特征图;$\theta_{k,r}$为卷积核大小为$k$和扩张率为$r\in\mathbf{Z}^+$的空洞卷积。在空洞卷积中,卷积核大小$k$增加为$k+(k-1)(r-1)$,当$r=1$时,空洞卷积相当于标准卷积。标准卷积的卷积层感受野与之前所有层卷积核的大小和步长有关,感受野呈线性增长,而空洞卷积感受野为$(2^{r+1}-1)\times(2^{r+1}-1)$,因此空洞卷积的级联可以实现感受野呈指数增长,使得每个卷积输出都包含较多的信息。

2　结合空洞卷积的FuseNet变体网络

本文使用编码器-解码器作为基本网络结构,如图3所示。编码器-解码器是一种输出近似于输入的网络结构。因此,在影像分割阶段,原始图像分辨率与分割图像分辨率保持一致。解码器能够使用反池化操作对特征图进行上采样,因此可使输出图像分辨率逼近输入图像分辨率。编码器部分采用VGG-16架构,包含5个卷积模块,每个卷积模块分别包含2个或者3个卷积核为3×3的卷积层,然后利用池化核为2×2的最大池化层对每个卷积模块提取的特征进行特征降维。每个卷积层中均使用修正线性单元(rectified linear unit, ReLU)作为激活函数,并利用批归一化(batch normalization, BN)使数据服从正态分布。

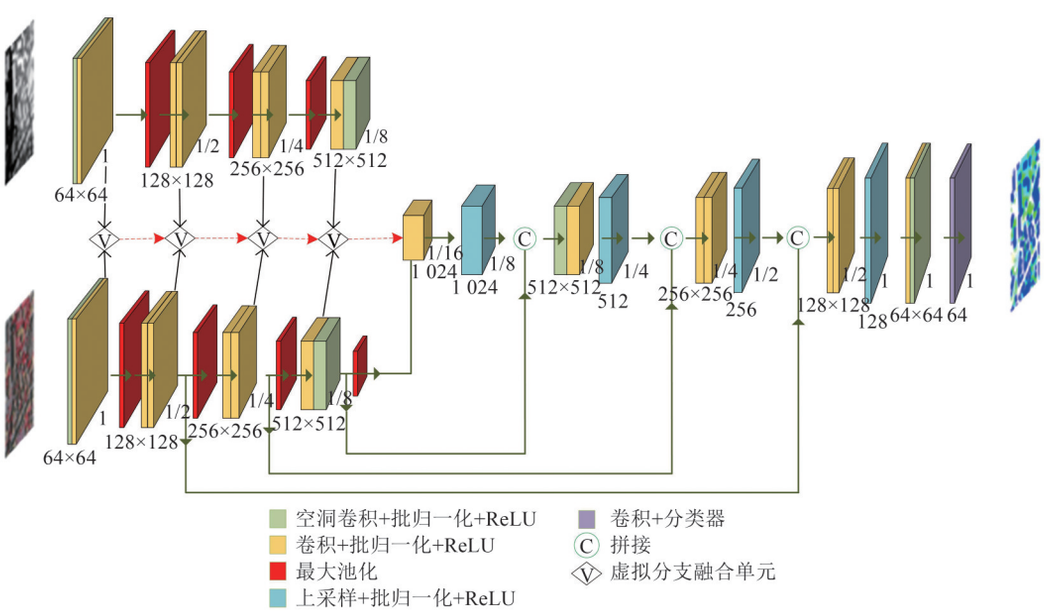


图3　结合空洞卷积的FuseNet变体网络结构

Fig.3　Structure of Improved Network Based on FuseNet and Atrous Convolution

解码器则是执行上采样和分类的过程。上采样是将编码后的特征图恢复到原始空间分辨率,在解码过程中池化层被反池化层替换,反池化是根据最大池化过程中的索引从较小的特征图映射到一个零填充的上采样特征图。如图4所示,给定一个特征图,定义其大小为4×4,步长为2,通过最大池化操作得到特征图以及特征图中各值在原特征图中的索引。反池化操作是根据索引和特征图进行补0,这种反池化操作将抽象特征转换为几何特征。在反池化操作后,卷积块增加稀疏特征图的密度。重复此过程,直到特征图与输入分辨率一致。相比于其他网络结构,降采样操作会丢失细节信息,虽然底层特征具有丰富细节,但判别能力较弱,使得网络对小目标地物的分割性能较差。编码器-解码器结构中通过将上采样操作与跳跃连接相结合,利用反池化操

作把浅层信息和高层信息融合，一定程度上缓解了细节丢失问题，使得该基本结构对于分割小目标地物效果也较好。

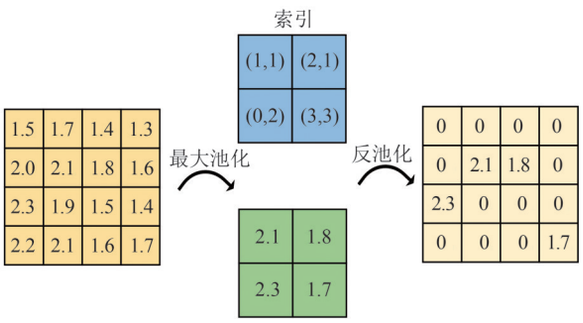


图4 最大池化和反池化操作对4×4特征图的影响

Fig.4 Influence of Max Pooling and Unpooling Operation on 4×4 Characteristic Pattern

在编码器-解码器的特征图处理过程中，如果空间分辨率一致，则可以直接通过跳跃连接进行特征融合；如果空间分辨率不一致，则将输入特征图通过1×1的卷积核投影成与输出特征图相同的维度。为了保持空间分辨率不变，本文提出的网络保留了初始2×2的最大池化，但需将所有卷积的步长减小为1。为了将特征图恢复到原始分辨率，反池化操作后进行标准卷积操作。最后计算损失函数L，并在像素块上取均值。L的计算公式为：

$$L=\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{k}y_j^i\lg\frac{\exp(z_j^i)}{\sum_{l=1}^{k}\exp(z_l^i)} \tag{2}$$

式中，N为输入图像的像素个数；k为类的个数；对于特定像素i，y_j^i表示像素i属于第j类标签；z_j^i表示像素损失预测值。本文在不进行任何空间正则化的情况下，将平均逐像素分类损失降到最低。此外，本文算法不使用任何影像后处理过程，提高了计算速度。

3 IFA-CNN实验结果与分析

3.1 数据集

本文在国际摄影测量与遥感学会(International Society for Photogrammetry and Remote Sensing，ISPRS)[13]航空影像Vaihingen数据集和Potsdam数据集上验证所提出算法的可行性。分别对建筑、不透水域表面(如道路)、低矮植被、树木、汽车和杂波等6个类别的地物进行语义分割。实验中，杂波的像素面积仅占总影像像素的0.88%。

Vaihingen数据集是由33张航拍影像组成的，采集于德国Vaihingen市1.38 km^{2}的区域内。每幅影像的平均大小为2 494×2 064像素，空间分辨率为9 cm，含3个波段：近红外(near infrared，NIR)、红(red，R)、绿(green，G)波段。影像中提供物体表面高度的DSM作为补充数据。本文选择29幅影像进行训练，4幅影像进行测试。

Potsdam数据集由38幅高分辨率航拍影像组成，其中24幅影像包含真实标签，覆盖面积3.42 km^2，每幅航拍影像由4个波段组成，分别为NIR、R、G、蓝(blue，B)，本文使用NIR、R、G波段。影像的大小为6 000×6 000像素，以6个类别的像素级标签作为标注，空间分辨率为5 cm，同样有DSM补充数据。实验中选择20幅影像进行训练，4幅影像进行测试。

3.2 多尺度数据增强及标准化

数据增强的目的是生成新的样本实例。当训练样本较少时，数据增强对提高网络的泛化能力起到关键性的作用。在Potsdam数据集中对高分辨率遥感影像随机裁剪，得到5 000个大小为256×256像素的图像块，并通过旋转、缩放等操作扩充数据集的规模，用于IFA-CNN网络的训练，从而增强网络的泛化能力。本文使用的高分辨率遥感影像的所有波段(NIR、R、G)都被标准化在[0，1]区间内。

神经网络的参数和激活函数通常初始化为[0，1]之间的随机数，需要采用标准化方法避免梯度爆炸、梯度弥散情况的出现。Z分数标准化方法[14]将输入图像的像素值逼近于正态分布，有利于提高网络收敛速度。标准化公式为：

$$X_{\text{out}}=\frac{X/\max(X)-\lambda}{\sigma} \tag{3}$$

式中，X_{out}为输出值；X为输入值；$\max(X)$为输入最大值；λ、σ分别为$X/\max(X)$的均值和标准差。

3.3 网络训练

由于本文使用的数据集是高分辨率遥感影像数据集，无法在深层网络中直接处理，因此使用滑动窗口的方法来提取256×256像素的小块。滑动窗口的步长也定义了两个连续小块之间重叠区域的大小。在训练时，较小的步长可以提取更多的训练样本，起到数据扩充的作用，所以将Vaihingen数据集和Potsdam数据集的步长分别设定为64像素和32像素。在测试时，较小的步长允许对重叠区域进行平均预测，以提高整体精度，本文分别使用32像素步长和16像素步长滑动窗口对Vaihingen数据集和Potsdam数据集中

的测试图像提取256×256像素的小块。

本文设置初始学习率为0.01,每隔5个迭代次数将学习率除以10直至0.000 01;动量参数为0.9,权重衰减为0.000 5,批归一化大小为10。对于编码器-解码器结构,采用迁移学习的方法利用ImageNet数据集上训练好的VGG-16的权值作为本文初始化编码器的权值,并随机初始化解码器的权值,有效缩短了模型的训练时间。将初始化后权值的学习率设定为新权值学习率的一半,并在每个数据集上对结果进行交叉验证。本文提出的深度学习网络的损失值曲线如图5所示,图5(a)为Vaihingen数据集在网络训练过程中的损失值曲线,在25 000次迭代后基本处于收敛状态,但当损失值第一次收敛趋近0.25时,损失曲线突然上升,其原因为后期训练中学习率相对过大。图5(b)为Potsdam数据集在网络训练过程中的损失值曲线。

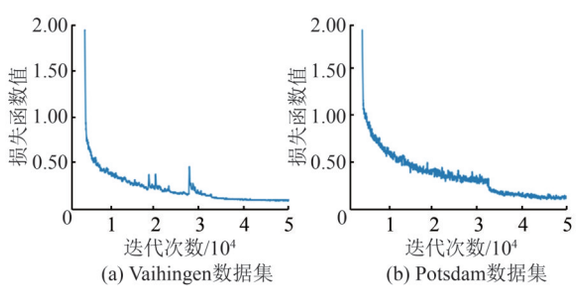


图5　损失值曲线

Fig.5　Loss Curve

3.4　结果与分析

3.4.1　评价标准

本文使用$F1$得分评估深度学习网络的性能,其计算公式为:

$$P=T_P/\left(T_P+F_P\right) \tag{4}$$

$$R=T_P/\left(T_P+F_N\right) \tag{5}$$

$$m_{F1}=2PR/\left(P+R\right) \tag{6}$$

式中,T_P为真正例,表示预测值为1,真实值为1;F_P为假正例,表示预测值为1,真实值为0;F_N为假反例,表示预测值为0,真实值为1;P为预测正确的正例数占预测为正例总量的比率,即查准率;R为预测正确的正例数占真正的正例数的比率,即查全率。本文实验中,通过计算$F1$得分的平均值m_{F1}评估网络的分割准确率,m_{F1}的值越大,表示网络性能越好,且分割准确率越高。

此外,本文还利用总体精度(overall accuracy,OA)评估算法的分割准确率。OA的计算公式为:

$$O_A=\frac{T_P+T_N}{T_P+F_P+T_N+F_N} \tag{7}$$

式中,O_A为OA值;T_N为真反例,表示预测值为0,真实值为0。

3.4.2　实验结果与分析

本文算法(IFA-CNN)得到的部分实验数据结果与真实标签之间的对比如图6所示。可以看出,IFA-CNN在整体上得到了比较理想的分割结果,尤其是对较大目标地物的分类效果很好,但在图像中也存在一些分割错误的区域。对比Vaihingen分割图像与真实标签可以看出,分割错误的区域较少,分割效果较好;但在Potsdam数据集的分割图像中出现了小块区域分割效果不佳的情况,主要原因为Potsdam数据集地物分布较复杂,且模糊区域较多,而Vaihingen数据集地物分布较均匀,分割难度较低。

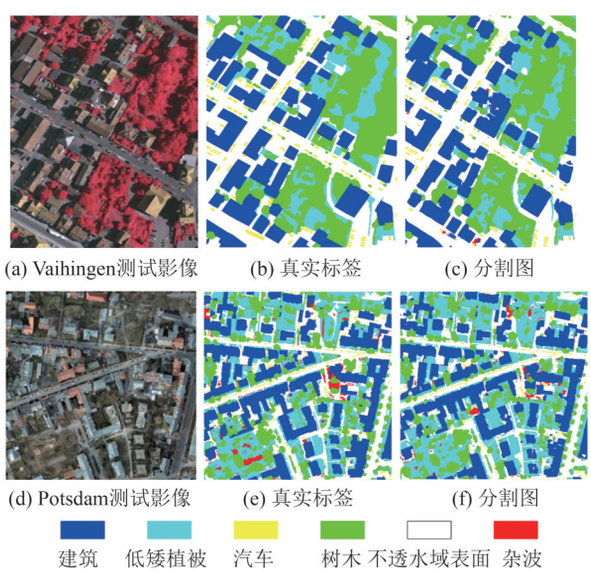


图6　ISPRS Vaihingen和Potsdam数据集的分割结果

Fig.6　Segmentation Results of ISPRS Vaihingen and Potsdam Dataset

表1为本文算法在ISPRS Vaihingen和Potsdam测试集上的分割准确率计算结果。可以看出,本文算法取得了不错的分割结果。本文提出的网络与文献[8]中的FuseNet网络在Vaihingen数据集和Potsdam数据集上分别进行实验对比,实验结果如表2、表3所示。在实验中,除融合单元部分不同外,其他网络结构部分一致。可以看出,本文多模态数据融合策略中使用的虚拟分支融合(virtual fusion,V-Fusion)单元对各类别地物的分割准确率均高于FuseNet网络的融合单元,进一步证明了虚拟分支融合单元解决了数据分配不均的问题,它将DSM分支提取的特征与RGB分支提取的特征在此单元进行融合,更好地

提取RGB-DSM图像的特征，因此添加虚拟分支融合单元的FuseNet网络分割准确率更高。

表1 ISPRS Vaihingen和Potsdam数据集的分割准确率

Tab.1 Segmentation Accuracy Results on ISPRS Vaihingen and Potsdam Dataset

数据集	*F*1得分					OA	平均*F*1得分
	建筑	树木	低矮植被	不透水域表面	汽车		
Vaihingen	0.955	0.921	0.836	0.937	0.871	0.915	0.904
Potsdam	0.956	0.864	0.906	0.917	0.939	0.909	0.916

表2 V-Fusion单元与Fusion单元在Vaihingen数据集上的分割准确率比较

Tab.2 Comparison of Segmentation Accuracy Between V-Fusion Unit and Fusion Unit on Vaihingen Dataset

FuseNet网络	*F*1得分					OA	平均*F*1得分
	建筑	树木	低矮植被	不透水域表面	汽车		
Fusion单元	0.939	0.846	0.833	0.911	0.853	0.898	0.876
V-Fusion单元	0.955	0.921	0.836	0.937	0.871	0.915	0.904

表3 V-Fusion单元与Fusion单元在Potsdam数据集上的分割准确率比较

Tab.3 Comparison of Segmentation Accuracy Between V-Fusion Unit and Fusion Unit on Potsdam Dataset

FuseNet网络	*F*1得分					OA	平均*F*1得分
	建筑	树木	低矮植被	不透水域表面	汽车		
Fusion单元	0.942	0.863	0.828	0.909	0.930	0.882	0.894
V-Fusion单元	0.956	0.864	0.906	0.917	0.939	0.909	0.916

为探索使用编码器-解码器结构对小目标地物分割准确率的影响，采用本文提出的网络与文献[15-17]中的网络在Vaihingen数据集上进行实验对比，与文献[7,18-19]中的网络在Potsdam数据集上进行实验对比，实验结果如表4、表5所示。可以看出，IFA-CNN采用的编码器-解码器结构对于汽车及低矮植被这两类小目标地物的分割准确率均高于非编码器-解码器结构网络的。由于小目标地物的细节信息较少，相比于其他网络结构，编码器-解码器结构在编码过程中能够较好地提取高分辨率遥感影像的语义特征，并在解码过程中通过反卷积将特征有效恢复为语义分割预测图，还原小目标地物的语义特征，减少细节信息的丢失。

表4 Vaihingen数据集上IFA-CNN与非编码器-解码器结构网络对小目标地物分割准确率比较

Tab.4 Comparison of Segmentation Accuracy for Small Objects Between IFA-CNN and Non-Encoder-Decoder Network on Vaihingen Dataset

类别	*F*1得分			
	UOA[15]	ADL_3[16]	DST_2[17]	IFA-CNN
汽车	0.820	0.633	0.726	0.871
低矮植被	0.804	0.823	0.834	0.836

表5 Potsdam数据集上IFA-CNN与非编码器-解码器结构网络对小目标地物分割的准确率比较

Tab.5 Comparison of the Segmentation Accuracy for Small Objects Between IFA-CNN and Non-Encoder-Decoder Network on Potsdam Dataset

类别	*F*1得分			
	FCN[7]	SCNN[18]	RGB+Iensembel[19]	IFA-CNN
汽车	0.893	0.912	0.892	0.939
低矮植被	0.800	0.837	0.822	0.906

为验证本文算法的有效性，对IFA-CNN与其他方法在ISPRS数据集上进行实验对比，结果如表6、表7所示。表6为在Vaihingen数据集上IFA-CNN与文献[16,20-21]算法的分割准确率对比，IFA-CNN无论从平均*F*1得分还是OA都取得了比较理想的结果。特别是树木类别的*F*1得分比文献[20-21]提高了0.22%，汽车类别的*F*1得分比文献[21]提高了0.47%。表7为在Potsdam数据集上IFA-CNN与文献[19,22-24]算法的分割准确率对比，IFA-CNN除汽车的分割准确率略低于文献[24]，不透水域表面的分割准确率与文献[22]算法持平，其余类别地物的分割准确率均高于其他算法。另外，IFA-CNN的OA和平均*F*1得分均高于其他算法，证明了

IFA-CNN的有效性。

由表6、表7可知,IFA-CNN在Vaihingen数据集和Potsdam数据集上,对各类别地物的分割效果都有着较好的表现。相较于其他算法,IFA-CNN的优点在于多个模式之间的互补性得到了更有效的利用,联合特征明显增强,更适用于将较弱的辅助数据(如DSM数据)集成到主学习网络中,并且虚拟分支融合单元很好地解决了特征融合效果不佳的问题。此外,由于IFA-CNN使用了多模态数据融合方案,同时空洞卷积通过扩大感受野的大小来捕获多尺度信息,提高了多目标分割任务的性能,所以IFA-CNN网络更好地提高了各类别地物的分割准确率。

表6　本文方法与其他方法在Vaihingen上的分割准确率对比

Tab.6　Comparison of the Accuracy of the Proposed Method with Other Methods on Vaihingen Dataset

模型	*F*1得分					OA	平均*F*1得分
	建筑	树木	低矮植被	不透水域表面	汽车		
ADL_3[16]	0.932	0.882	0.823	0.895	0.633	0.880	0.833
ONE_7[20]	0.945	0.899	0.844	0.910	0.778	0.898	0.875
GSN[21]	0.951	0.899	0.837	0.922	0.824	0.903	0.887
IFA-CNN	0.955	0.921	0.836	0.937	0.871	0.915	0.904

表7　本文方法与其他方法在Potsdam上的分割准确率对比

Tab.7　Comparison of the Accuracy of the Proposed Method with Other Methods on Potsdam Dataset

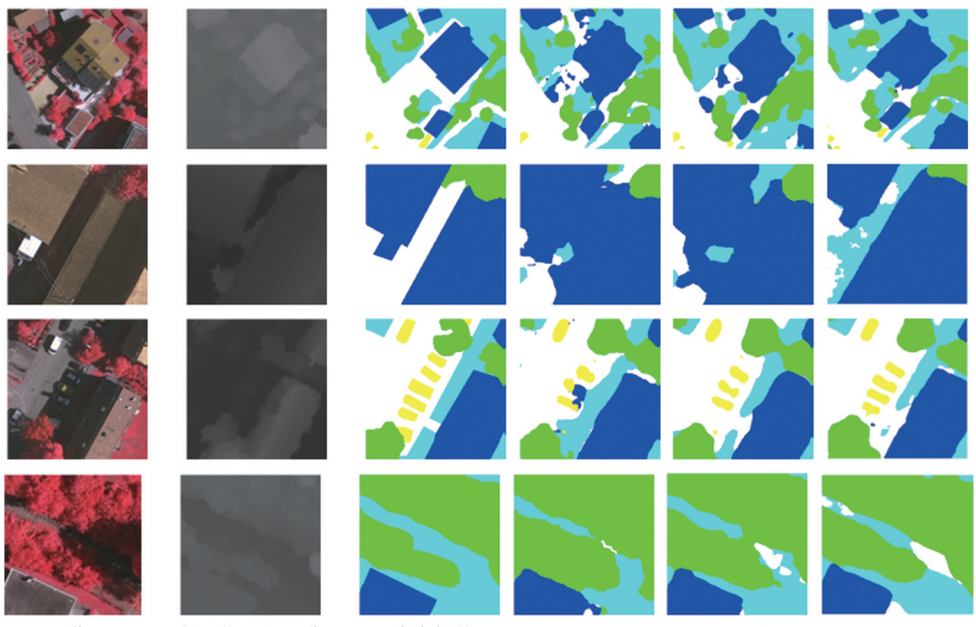
模型	*F*1得分					OA	平均*F*1得分
	建筑	树木	低矮植被	不透水域表面	汽车		
RiFCN[22]	0.930	0.819	0.837	0.917	0.937	0.883	0.861
RGB+I-ensemble[19]	0.936	0.845	0.822	0.870	0.892	0.900	0.873
Hallucination[23]	0.938	0.848	0.821	0.873	0.882	0.901	0.872
S-RA-FCN[24]	0.947	0.835	0.868	0.913	0.945	0.886	0.880
IFA-CNN	0.956	0.864	0.906	0.917	0.939	0.909	0.916

为使实验更具科学性,将IFA-CNN与网络结构为编码器-解码器且使用DSM数据的文献进行影像分割细节对比,实验结果如图7所示。图7中,第1列为输入遥感影像的局部细节,第2列为局部DSM细节,第3列为局部真实标签。图7(a)中,IFA-CNN与ONE_7[20]和GSN[21]的分割细节实例相比,IFA-CNN有效改善了分割影像中的边缘毛刺、细化类的边界,使得目标边缘更加接近场景的真实边缘。在图7(b)中,IFA-CNN与RiFCN[22]和S-RA-FCN[24]相比,对建筑、树木等较大目标地物的分割更加准确,有效地减少了误分割现象,阴影覆盖区域部分分割效果也较为理想。

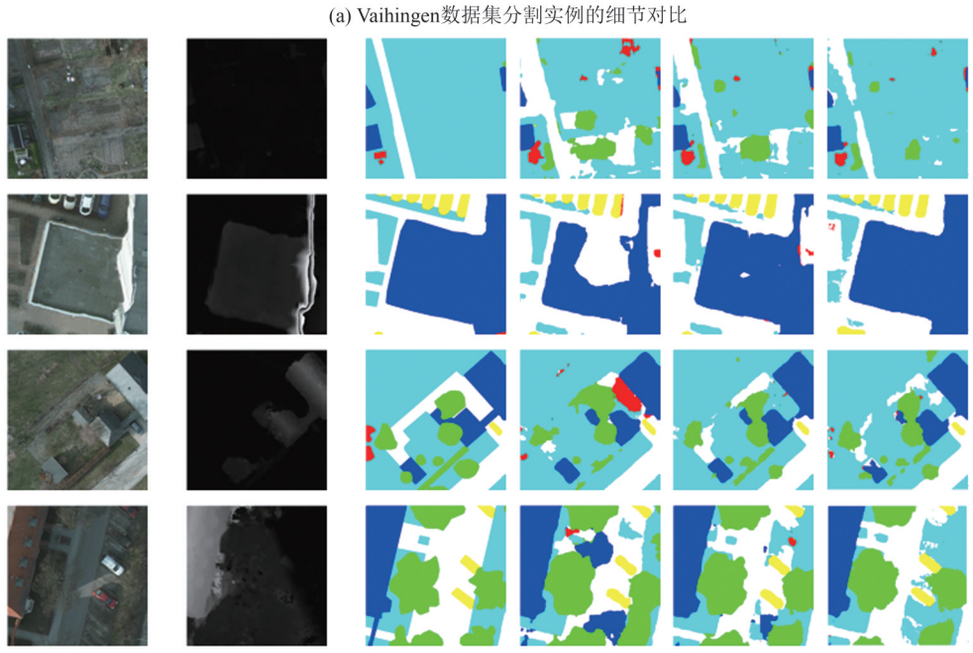
4　结　语

本文提出了一种结合空洞卷积的FuseNet变体深度学习网络架构,实现了高分辨率遥感影像语义分割。FuseNet变体的多模态数据融合可以使网络学习到更强的特征并有效地利用异构数据的互补性,将高分辨率遥感影像的DSM信息与RGB信息融合。在编码器-解码器架构中使用了跳跃连接,将高级特征与低级特征结合,使网络的整体分割精度提高。感受野在编码器-解码器部分均使用了空洞卷积,采用大滤波器的转置卷积进行上采样,获得了更大的接受域。

在公开数据集ISPRS Vaihingen和Potsdam上进行了实验,并与相关文献方法进行对比,实验结果表明本文所提出的IFA-CNN取得了较好的分割准确率。然而,本文方法仍存在改进的空间。一方面,对于高分辨率遥感影像的边缘还存在分割不准确的情况;另一方面,尝试在保证分割准确率的情况下减少网络层数,提高网络运算效率。



(a) Vaihingen数据集分割实例的细节对比



(b) Potsdam数据集分割实例的细节对比

图7 ISPRS Vaihingen和Potsdam数据集分割实例细节对比

Fig.7 Comparison of Detailed Segmentation Results on ISPRS Vaihingen and Potsdam Dataset

参 考 文 献

[1] Kampffmeyer M, Salberg A B, Jenssen R. Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks [C]//IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 2016

[2] Wang H, Wang Y, Zhang Q, et al. Gated Convolutional Neural Network for Semantic Segmentation in High-Resolution Images[J]. *Remote Sensing*, 2017, 9(5): 1-15

[3] Mou Lichao, Hua Yuansheng, Zhu Xiaoxiang. A Relation-Augmented Fully Convolutional Network for Semantic Segmentation in Aerial Scenes[C]//IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA, 2019

[4] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions[C]//IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA,

USA, 2015

[5] Hoffman J, Gupta S, Darrell T. Learning with Side Information Through Modality Hallucination[C]// IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016

[6] Hazirbas C, Ma L, Domokos C, et al. FuseNet: Incorporating Depth into Semantic Segmentation via Fusion-Based CNN Architecture[C]//Asian Conference on Computer Vision, Taipei, China, 2016

[7] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 39(4): 640-651

[8] Badrinarayanan V, Kendall A, Segnet R C. A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495

[9] Sherrah J. Fully Convolutional Networks for Dense Semantic Labelling of High-Resolution Aerial Imagery [EB/OL]. (2016-06-08) [2020-06-22]. https://www.doc88.com/p-0704858988942.html

[10] Nogueira K, Penatti O A B, Santos J A D. Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification[J]. *Pattern Recognition*, 2017, 61: 539-556

[11] Zhang Kang, Hei Baoqin, Zhou Zhuang, et al. CNN with Coefficient of Variation-Based Dimensionality Reduction for Hyperspectral Remote Sensing Images Classification [J]. *Journal of Remote Sensing*, 2018, 22(1): 91-100 (张康, 黑保琴, 周壮, 等. 变异系数降维的CNN高光谱遥感图像分类[J]. 遥感学报, 2018, 22(1): 91-100)

[12] Everingham M, Eslami S M A, van Gool L, et al. The Pascal Visual Object Classes Challenge: A Retrospective [J]. *International Journal of Computer Vision*, 2015, 111(1): 98-136

[13] Gerke M, Rottensteiner F, Wegner J D, et al. ISPRS Semantic Labeling Contest [J]. *Remote Sensing*, 2020, 12(3): 417-446

[14] Ngiam J, Khosla A, Kim, et al. Multimodal Deep Learning[C]// The 28th International Conference on Machine Learning, Washington DC, USA, 2011

[15] 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

[16] Luo W, Li Y, Urtasun R, et al. Understanding the Effective Receptive Field in Deep Convolutional Neural Networks[C]// The 30th Conference on Advances in Neural Information Processing Systems, Barcelona, Spain, 2016

[17] Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions[C]//International Conference on Learning Representations, San Juan, Puerto Rico, 2016

[18] Liu Y, Piramanayagam S, Monteiro S T, et al. Dense Semantic Labeling of Very-High-Resolution Aerial Imagery and LiDAR with Fully-Convolutional Neural Networks and Higher-Order CRFs[C]// IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawii, USA, 2017

[19] Zhao Jun, Guo Feixiao, Li Qi. Fisher-Score Algorithm of WTLS Estimation for PEIV Model[J]. *Geomatics and Information Science of Wuhan University*, 2019, 44(2): 214-220 (赵俊, 郭飞霄, 李琦. PEIV模型WTLS估计的Fisher-Score算法[J]. 武汉大学学报·信息科学版, 2019, 44(2): 214-220)

[20] Chen G, Zhang X, Wang Q, et al. Symmetrical Dense-Shortcut Deep Fully Convolutional Networks for Semantic Segmentation of Very-High-Resolution Remote Sensing Images[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, 11(5): 1633-1644

[21] Wei Y, Xiao H, Shi H, et al. Revisiting Dilated Convolution: A Simple Approach for Weakly-and Semi-supervised Semantic Segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018

[22] Lin G S, Shen C H, van den Hengel A, et al. Efficient Piecewise Training of Deep Structured Models for Semantic Segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016

[23] Paisitkriangkrai S, Sherrah J, Janney P, et al. Effective Semantic Pixel Labelling with Convolutional Networks and Conditional Random Fields[C]// IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 2015

[24] Audebert N, Saux B L, Lefèvre S. Semantic Segmentation of Earth Observation Data Using Multimodal and Multi-Scale Deep Networks[C]//The 13th Asian Conference on Computer Vision, Taipei, China, 2016

Semantic Segmentation of High-Resolution Remote Sensing Images Based on Improved FuseNet Combined with Atrous Convolution

YANG Jun[1,2,3,4] *YU Xizi*[2,3,4]

1 School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China
2 Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China
3 National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou 730070, China
4 Gansu Provincial Engineering Laboratory for National Geographic State Monitoring, Lanzhou 730070, China

Abstract: **Objectives**: With the development and popularization of deep learning theory, deep neural networks are widely used in image analysis and interpretation. The high-resolution remote sensing images have the characteristics of a large amount of information, complex data, and rich feature information, and most of the current semantic segmentation neural networks of the natural image are not completely designed for the characteristics of high-resolution remote sensing images, so it cannot effectively extract the detailed features of the ground objects in remote sensing images, and the segmentation accuracy needs to be improved. **Methods**: We propose the process of improved FuseNet with the atrous convolution-convolutional neural network(IFA-CNN). Firstly, we use the improved FuseNet to fuse the elevation information of DSM(digital surface model) images with the color information of RGB(red green blue) images. At the same time, we propose a multimodal data fusion scheme to solve the problem of poor fusion of the RGB branch and DSM branch. Secondly, multiscale features are captured through flexibly adjusting the receptive field by the atrous convolution. Through deconvolution and upsampling, a decoder that increases the feature maps is formed. Finally, the Softmax classifier is used to procure the semantic segmentation results. **Results**: Compared with relevant algorithms, IFA-CNN effectively improves the edge burr and thinning boundaries in segmented images, and is more accurate for segmentation of larger objects such as buildings and trees, it also reduces the miss segmentation condition with effect, the segmentation of the shadow covered areas is close to being perfect. The m_{F1} score achieved when our model is applied to the open ISPRS(International Society for Photogrammetry and Remote Sensing) Potsdam and Vaihingen dataset are 91.6% and 90.4% respectively, exceeding by a considerable margin of relevant algorithms. **Conclusions**: (1) The virtual fusion (V-Fusion) unit used for segmentation by the multimodal data fusion strategy is more accurate than the one used by the FuseNet network. (2) The encoder-decoder structure is arranged in such a way that the effective improvement of the segmentation accuracy of small target features is guaranteed. So, the loss of detailed information can be decreased. (3) While the multimodal data fusion is being carried out by IFA-CNN, the atrous convolution expands the receptive field accordingly to extract the multiscale information.

Key words: high-resolution remote sensing image; deep convolutional neural network; atrous convolution; semantic segmentation; FuseNet

First author: YANG Jun, PhD, professor, specializes in computer graphics, image processing, and geographic information system. E-mail: yangj@mail.lzjtu.cn

Foundation support: The National Natural Science Foundation of China(61862039); Science and Technology Program of Gansu Province (20JR5RA429); 2021 Central Government Funds for Guiding Local Science and Technology Development(2021-51); Excellent Platform Support Project of Lanzhou Jiaotong University(201806).

引文格式: YANG Jun, YU Xizi. Semantic Segmentation of High-Resolution Remote Sensing Images Based on Improved FuseNet Combined with Atrous Convolution[J]. Geomatics and Information Science of Wuhan University, 2022, 47(7): 1071-1080. DOI: 10.13203/j.whugis20200305(杨军, 于茜子. 结合空洞卷积的FuseNet变体网络高分辨率遥感影像语义分割[J]. 武汉大学学报·信息科学版, 2022, 47(7): 1071-1080. DOI:10.13203/j.whugis20200305)