

基于多特征图金字塔融合深度网络的遥感图像语义分割

赵 斐^{①②} 张文凯^{*①③④} 闫志远^{①③④} 于泓峰^{①③④} 刁文辉^{①③④}

^①(中国科学院大学 北京 100049)

^②(北京跟踪与通信技术研究所 北京 100094)

^③(中国科学院电子学研究所 北京 100190)

^④(中国科学院电子学研究所空间信息处理技术与应用院重点实验室 北京 100190)

摘 要: 在遥感图像语义分割中, 利用多元数据(如高程信息)进行辅助是一个研究重点。现有的基于多元数据的分割方法通常直接将多元数据作为模型的多特征输入, 未能充分利用多元数据的多层次特征, 此外, 遥感图像中目标尺寸大小不一, 对于一些中小型目标, 如车辆、房屋等, 难以做到精细化分割。针对以上问题, 提出一种多特征图金字塔融合深度网络(MFPNet), 该模型利用光学遥感图像和高程数据作为输入, 提取图像的多层次特征, 然后针对不同层次的特征, 分别引入金字塔池化结构, 提取图像的多尺度特征, 最后, 设计了一种多层次、多尺度特征融合策略, 综合利用多元数据的特征信息, 实现遥感图像的精细化分割。基于Vaihingen数据集设计了相应的对比实验, 实验结果证明了所提方法的有效性。

关键词: 语义分割; 深度卷积神经网络; 特征图融合; 金字塔池化

中图分类号: TP391.41

文献标识码: A

文章编号: 1009-5896(2019)10-2525-07

DOI: 10.11999/JEIT190047

Multi-feature Map Pyramid Fusion Deep Network for Semantic Segmentation on Remote Sensing Data

ZHAO Fei^{①②} ZHANG Wenkai^{①③④} YAN Zhiyuan^{①③④}

YU Hongfeng^{①③④} DIAO Wenhui^{①③④}

^①(University of Chinese Academy of Sciences, Beijing 100049, China)

^②(Beijing Institute of Tracking and Telecommunications Technology, Beijing 100049, China)

^③(Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China)

^④(Key Laboratory of Spatial Information Processing and Application System Technology,
Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Utilizing multiple data (elevation information) to assist remote sensing image segmentation is an important research topic in recent years. However, the existing methods usually directly use multivariate data as the input of the model, which fails to make full use of the multi-level features. In addition, the target size varies in remote sensing images, for some small targets, such as vehicles, houses, etc., it is difficult to achieve detailed segmentation. Considering these problems, a Multi-Feature map Pyramid fusion deep Network (MFPNet) is proposed, which utilizes optical remote sensing images and elevation data as input to extract multi-level features from images. Then the pyramid pooling structure is introduced to extract the multi-scale features from different levels. Finally, a multi-level and multi-scale feature fusion strategy is designed, which utilizes comprehensively the feature information of multivariate data to achieve detailed segmentation of remote sensing images. Experiment results on the Vaihingen dataset demonstrate the effectiveness of the proposed method.

Key words: Semantic segmentation; Deep Convolutional Neural Network(DCNN); Feature map fusion; Pyramid pooling

收稿日期: 2019-01-17; 改回日期: 2019-04-08; 网络出版: 2019-04-20

*通信作者: 张文凯 iecas_wenkai@yahoo.com

基金项目: 国家自然科学基金(41701508)

Foundation Item: The National Natural Science Foundation of China (41701508)

1 引言

图像语义分割是指对图像中的像素点进行分类,在场景理解和目标检测等领域中具有广泛应用。近年来,国内外学者在图像语义分割领域做了大量的研究,提出了许多分割方法。传统的图像语义分割方法通常依赖手工设计的特征(方向梯度直方图(Histogram of Oriented Gradient, HOG)^[1],尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)^[2]),然后通过设置阈值^[3]或利用机器学习分类器^[3](随机森林,支持向量机),对图像像素进行分类。这类方法较为简单,分割速度较快,但特征设计存在诸多困难,限制了这类方法的实用性。

近年来涌现出大量的图像语义分割算法,尤其是基于深度卷积神经网络(Convolutional Neural Networks, CNN)^[4]的方法,在计算机视觉领域得到了快速发展和广泛应用。随着深度学习的快速发展,卷积神经网络以其优秀的特征表达能力,在图像处理领域大放异彩,基于深度卷积神经网络的图像语义分割方法逐渐成为研究热点。Long等人^[5]在2015年提出了全卷积神经网络(Fully Convolutional Networks, FCN),首次实现了以端对端的形式进行图像语义分割,并在自然场景图像分割中取得了较好的结果。此后,许多学者以此为基础,将卷积神经网络模型应用到遥感图像语义分割中来,但是与自然场景图像相比,遥感图像往往包含更丰富的多元信息,如光学遥感图像、高程数据等,如何充分利用图像的高程信息,辅助图像分割,提升分割精度,仍是遥感图像语义分割中的难点,相关方法不断被提出。例如,文献[6-8]直接将遥感图像的多元信息作为统一的数据,输入到全卷积神经网络中,得到分割结果,相比于单独使用光学遥感图像进行分割,这类方法的效果提升并不明显。Marmanis等人^[9]提出使用包含两个分支的卷积神经网络模型,将光学遥感图像和高程数据分别作为这两个分支的输入,经过一系列卷积操作,最后在预测层融合这两个分支产生的分割结果。考虑到网络下采样会丢失图像信息,Sherrah^[10]设计了一个不带下采样的全卷积神经网络模型,并同样在最后融合光学遥感图像分支和高程数据分支的分割结果。以上方法在处理多元遥感信息时,或者直接在输入端进行融合,或者最后在预测层进行融合,都未对卷积神经网络的中间层信息进行研究,并且对遥感图像小目标的分割结果不够理想。

针对以上问题,考虑到卷积神经网络的浅层特征包含丰富的空间信息,深层特征包含丰富的语义信息,若能综合利用这些多层次特征,将能有效提

高遥感图像的分割精度,因此本文提出了一种基于多特征图金字塔融合深度网络(Multi-Feature map Pyramid fusion deep Network, MFPNet)的遥感图像语义分割模型,具体来说,本文的贡献与创新点如下:

(1) 本文提出了一种多元数据特征图融合策略,本策略考虑使用网络的中间层信息,将光学遥感图像和高程数据的各中间层特征分别对应融合,然后将融合后的深层特征逐步上采样,与浅层特征融合,综合利用浅层特征的空间信息与深层特征的语义信息,精细化分割目标边界,优化分割效果;

(2) 考虑到遥感图像中目标尺寸大小不一,对于一些中小型目标,如车辆、房屋等,往往存在漏检的问题,难以做到精细化分割。针对这个问题,本文设计引入金字塔场景解析网络(Pyramid Scene Parsing Network, PSPNet)^[11]中的金字塔池化结构,针对不同层次的特征,使用金字塔池化结构提取图像的多尺度特征,提高对不同尺度目标的分割精度。

本文的结构安排如下:第2节详细介绍本文提出的网络模型;第3节提供了相关实验与分析,最后第4节总结全文工作。

2 多特征图金字塔融合深度网络模型

本文所提的MFPNet如图1所示。本模型包含3个部分:特征编码、金字塔池化和特征解码。特征编码部分包含两个输入分支,输入分别为光学遥感图像和高程数据,基础网络采用残差网络模型(Residual Network, ResNet)^[12],分别提取光学遥感图像和高程数据的多层次特征(L_1, L_2, L_3, L_4, L_5)。然后,本文将这两个分支的不同层次特征分别对应相加融合,将融合后的特征图(C_2, C_3, C_4, C_5)分别送入金字塔池化结构,提取对应层次的多尺度特征与上下文信息。最后,在特征解码部分,我们将提取到的多层次、多尺度特征(F_5, F_4, F_3, F_2)逐步前向融合,最终送入softmax分类器,产生遥感图像的分割结果。网络的具体细节将在下面小节详细介绍。

2.1 特征编码

特征编码部分使用基础网络提取图像的多层次特征,本文选用ResNet-101^[12]模型作为基础网络。原始的ResNet-101模型是用于自然场景图像分类任务,为了将其应用到多元遥感数据的语义分割,本文对其进行了结构上的修改,修改后的网络结构如表1所示。

首先去掉了原始ResNet-101模型最后用于图像分类的全连接层,将网络变成了全卷积神经网络,以用于图像分割任务。为了充分利用遥感图像的多

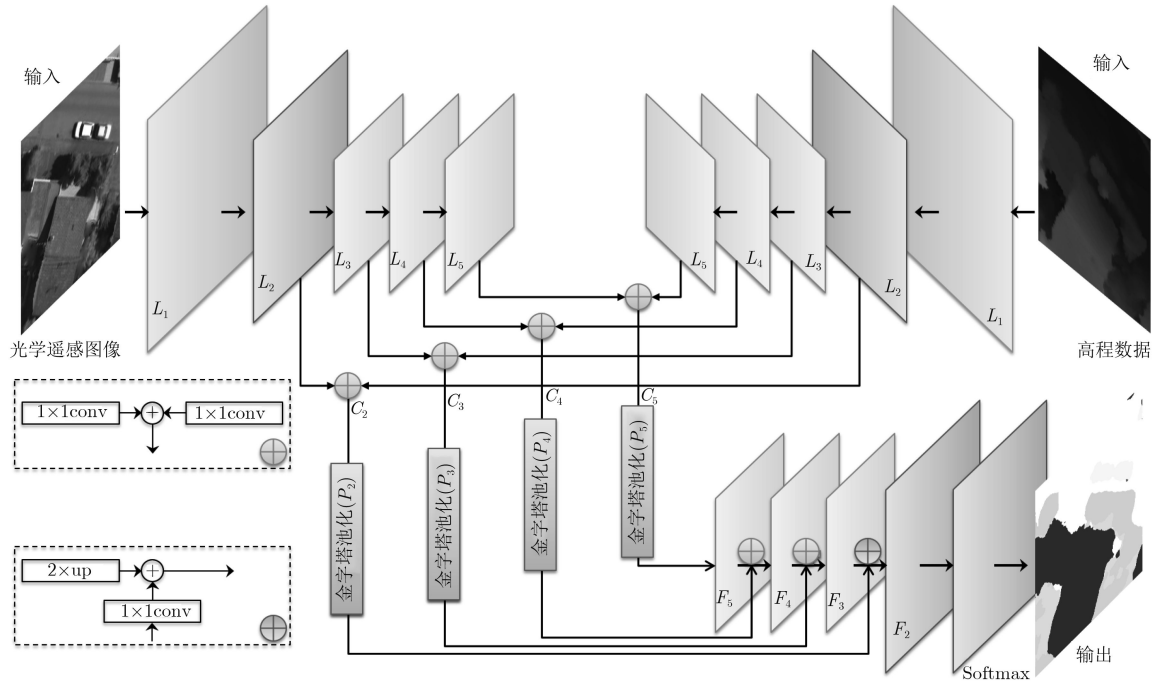


图1 多元特征图融合网络模型框架图

表1 特征编码网络结构

ResNet卷积层	光学遥感图像分支输出	高程数据分支输出	多元特征融合	融合输出	输出尺寸
7×7, 64, 步幅2	$L_{1-\text{img}}$	$L_{1-\text{ele}}$			1/2
3×3, 最大值池化, 步幅2 $\left. \begin{array}{l} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{array} \right\} \times 3$	$L_{2-\text{img}}$	$L_{2-\text{ele}}$	✓	C_2	1/4
$\left. \begin{array}{l} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{array} \right\} \times 4$	$L_{3-\text{img}}$	$L_{3-\text{ele}}$	✓	C_3	1/8
$\left. \begin{array}{l} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{array} \right\} \times 23$ (带孔卷积)	$L_{4-\text{img}}$	$L_{4-\text{ele}}$	✓	C_4	1/8
$\left. \begin{array}{l} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{array} \right\} \times 3$ (带孔卷积)	$L_{5-\text{img}}$	$L_{5-\text{ele}}$	✓	C_5	1/8

元信息，本文添加了高程数据分支，使用该分支单独提取高程数据的特征。为了避免下采样造成的高层特征图尺寸太小，丢失太多空间信息，在第 L_4 层和 L_5 层使用带孔卷积^[12]，这样在不减小特征图的情况下，有效扩大了感受野。最后，本文添加了多元数据特征融合层，将两个分支的 L_2, L_3, L_4, L_5 层特征分别对应逐像素相加，进行多元信息融合，融合后的特征标记为 C_2, C_3, C_4, C_5 ，具体融合策略如下：

假设在一个 L 层的卷积神经网络中，使用 $X^{(l)}$ 表示第 l 层的输入， $Z^{(l)}$ 表示第 l 层的输出， $W^{(l)}$ 表示第 l 层的权值参数， $b^{(l)}$ 表示第 l 层的偏置项， $f(\cdot)$ 表示激活函数，则：

第 l 层的卷积输出为

$$Z^{(l)} = W^{(l)} * X^{(l)} + b^{(l)} \quad (1)$$

经过ReLU激活后输出为

$$A^{(l)} = f(Z^{(l)}), \text{ 其中 } f(x) = \max(0, x) \quad (2)$$

本文中，使用 $Z^{(l)}(\text{img})$ 表示第 l 层光学遥感图像的卷积输出， $Z^{(l)}(\text{ele})$ 表示第 l 层高程数据的卷积输出，若在ReLU激活之前融合，其融合输出为

$$Z^{(l)} = Z^{(l)}(\text{img}) + Z^{(l)}(\text{ele}) \quad (3)$$

Hazirbas等人^[13]的研究表明，考虑到卷积输出的值可能存在负值，若在ReLU激活之前融合，会弱化ReLU层的输出。而ReLU激活后的值都是非负

值,若在ReLU激活后再进行特征融合,则可以增强图像特征信息,用式(4)可以表示为

$$f(Z^{(l)}(\text{img}) + Z^{(l)}(\text{ele})) \leq f(Z^{(l)}(\text{img})) + f(Z^{(l)}(\text{ele})) \quad (4)$$

基于以上推论,本文采用在ReLU激活层后,将对应特征图逐像素相加,进行多元特征融合。

2.2 金字塔池化

遥感图像中目标尺寸大小不一,在语义分割过程中容易漏掉一些中小型目标,如车辆、房屋等,这往往是由于卷积神经网络中连续的池化操作造成的空间信息损失,传统的池化方式只在某个固定大小的池化窗口内下采样,感受野大小固定,提取到的上下文信息有限,而提取不同尺度的上下文信息,将有利于对不同尺度目标的分割。针对以上问题,本文在特征编码后的不同层次特征上(C_2, C_3, C_4, C_5),分别引入金字塔池化结构^[11],提取图像的多尺度特征,提高对不同尺度目标的分割精度。

金字塔池化结构如图2所示,它是一个多级池化模块。金字塔顶部使用全局平均池化,提取输入的全局上下文特征;第2层、第3层分别将输入划分为不同大小的子区域,例如第2层将输入划分为4个子区域,第3层将输入划分为16个子区域,然后分别在这些子区域内使用全局平均池化,提取了不同大小的局部特征。然后使用 1×1 卷积降维,减少计算量。最后,将不同大小的池化结果上采样到输入特征图的大小,并进行通道叠加的融合,作为该金字塔池化模块最后的输出。通过这种方式,不仅可以提取全局的上下文特征,而且可以提取多尺度的局部上下文特征。本文中,金字塔分为3个级别,每个级别的子区域划分为 $1 \times 1, 2 \times 2$ 和 4×4 。

2.3 特征解码

遥感图像语义分割中,往往存在目标边界模糊的问题,这通常是由于卷积神经网络深层特征包含的空间信息较少的原因。而卷积神经网络的浅层特

征包含丰富的边缘、纹理信息,深层特征包含丰富的语义信息,若能综合利用网络的浅层空间信息与深层语义信息,则能精细化分割目标边界,优化输出结果。基于此,本文在特征解码部分,设计了一种多层次、多尺度特征融合策略,将经过金字塔池化的各级特征(F_5, F_4, F_3, F_2),由深到浅逐步融合,最后将融合后的特征送入softmax分类器,输出预测结果。

具体来说,从最深层的 F_5 开始,经过 1×1 卷积改变通道数,使 F_5 与 F_4 的通道数保持一致,由于 F_5 与 F_4 具有相同的分辨率,可以不用经过上采样,直接进行逐像素相加的融合;同理 F_4 与 F_3 ;由于 F_3 的分辨率只有 F_2 的 $1/2$,所以需要先对 F_3 进行2倍的上采样,再与 F_2 逐像素相加融合。最后将 F_2 融合后的特征送入softmax多分类器,产生最后的预测结果。

3 实验与分析

3.1 实验数据集介绍

实验采用Vaihingen遥感影像数据集^[14],该数据集包含16张宽幅图像,每幅图像含有正射影像和相应的高程数据,正射影像为IRG(Infrared-Red-Green)类型,分辨率为0.9 m,高程数据的采样距离为0.9 m。实验随机选取12幅图像作为训练集,剩余4幅作为测试集。为了便于训练,本文对大图进行裁剪,裁剪尺寸为 256×256 ,同时对数据进行增强,将裁剪后的图片,每隔 90° 旋转1次,并进行左右镜像处理。最终得到训练集共15780张切片,测试集2016张切片。

3.2 实验参数设置

本文实验基于深度学习框架tensorflow^[15]。在模型的输入端,光学遥感图像的输入大小为 $256 \times 256 \times 3$,高程数据的输入大小为 $256 \times 256 \times 1$ 。本文使用在ImageNet数据集上预训练的ResNet-101网络权值初始化MFPNet的基础网络,MFPNet的其

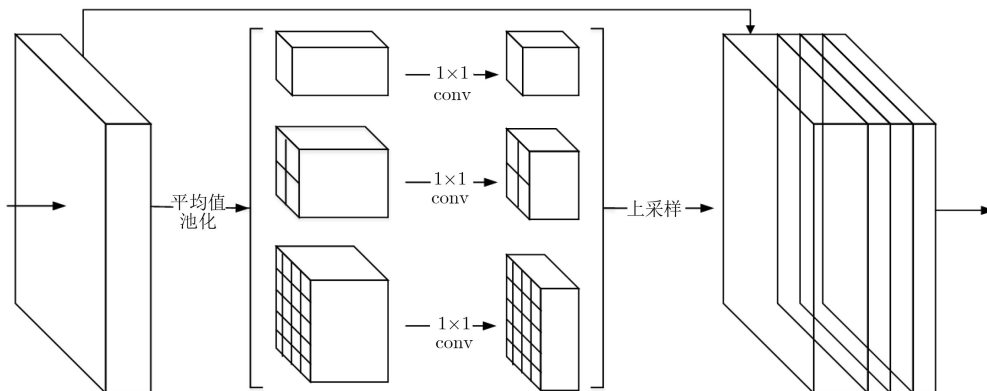


图2 金字塔池化结构

它层参数使用均值为0，方差为0.2的高斯分布随机初始化。训练迭代总次数为60000，使用随机梯度下降法训练模型，采用步进策略加速网络的收敛，初始的基准学习率为0.0025，每迭代20000次将学习率减少0.1倍，遗忘因子 μ 设置为0.9。每次训练包含1个前向传播和反向传播过程，前向传播推理预测结果，并和真值标注比较生成交叉熵损失，反向传播更新模型权值参数。

3.3 评价标准

为了评价MFPNet在Vaihingen数据集上的分割效果，本文使用了3个常用的评价指标，分别是平均交并比(mean Intersection Over Union, mIOU)、总体分类准确率(Overall Accuracy, OA)和F1值(F1 Score)。假设有 k 个不同的类别，令 n_{ij} 表示实际类别为 i 但是预测结果为 j 的像素个数， $t_i = \sum_j n_{ij}$ 表示类别 i 的总像素个数， $p_i = \sum_j n_{ji}$ 表示预测结果为类别 i 的总像素个数。它们的计算公式分别为

$$\text{mIOU} = \frac{1}{k} \times \sum_i \frac{n_{ii}}{t_i + p_i - n_{ii}} \quad (5)$$

$$\text{OA} = \frac{\sum_i n_{ii}}{\sum_i t_i} \quad (6)$$

$$\text{F} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

其中，精确率 $\text{Precision} = \frac{n_{ii}}{p_i}$ ，召回率 $\text{Recall} = \frac{n_{ii}}{t_i}$ 。

3.4 消融实验结果与分析

为了验证本文方法的有效性，本文设计了模型消融实验，实验结果如表2所示。其中，Color-E模型

是指在输入端同时输入光学遥感图像和高程数据，不包含多元数据融合结构与金字塔池化结构，只利用 L_5 分支进行结果预测；MFFNet模型是指在Color-E模型的基础上，添加多元数据融合结构，将光学遥感图像分支和高程数据分支的不同层次特征分别进行融合，然后使用融合后的特征进行预测输出；本文提出的MFPNet模型则是在MFFNet模型的基础上，添加金字塔池化结构，也就是如图1所示的整体模型。

根据表2实验结果，MFFNet模型的各项指标都高于Color-E模型，其中mIOU提高了6.85，OA提高了2.98，不同类别的F1也都有相应提高，这充分说明了多元数据融合结构的有效性。而本文提出的MFPNet模型相比于MFFNet模型，各项指标又有了进一步提高，尤其是“车辆”的F1提高了0.14，“建筑物”的F1提高了0.05，这说明了本文引入的金字塔池化结构对遥感图像的中小型目标分割有明显的提升作用。

3.5 与其它方法对比实验结果与分析

为了验证本文方法MFPNet的效果，与其它方法进行了对比，对比模型选择了FCN^[8]，DeepLab^[16]和PSPNet^[11]，FCN是图像语义分割领域使用全卷积神经网络的基石，DeepLab和PSPNet是当前比较流行的语义分割模型。对于这3个模型，本文在输入端将光学遥感图像和高程数据合并为4通道数据进行输入。

表3展示了本文提出的MFPNet模型与其他方法的对比结果。由表3可知，从整体上看，本文的MFPNet模型相比于其他方法，取得了较高的mIOU和OA。通过对比每个类别的F1，可以发现我们的模型对建筑物、车辆等中小型目标分割有明显

表 2 MFPNet模型消融实验结果

模型	mIOU	OA	F1					
			道路	建筑物	草地	树木	车辆	其它
Color-E	68.96	81.77	0.85	0.88	0.72	0.83	0.50	0.59
MFFNet	75.81	84.75	0.89	0.91	0.79	0.87	0.62	0.68
MFPNet	77.10	85.95	0.91	0.96	0.82	0.88	0.76	0.75

表 3 MFPNet与其他方法的对比结果

方法	mIoU	OA	F1					
			道路	建筑物	草地	树木	车辆	其它
FCN	59.65	79.67	0.82	0.86	0.69	0.81	0.56	0.59
Deeplab	70.85	82.75	0.86	0.89	0.72	0.82	0.60	0.61
PSPNet	74.96	83.92	0.90	0.93	0.74	0.81	0.65	0.63
MFPNet	77.10	85.95	0.91	0.96	0.82	0.88	0.76	0.75

的提升作用,同时对草地和树木这类不规则目标也有较好的分割效果,这都说明了MFPNet模型的有效性。

图3展示了不同方法在Vaihingen数据集上的分割结果对比图。通过对比图3(a)可以发现,如果直接将高程数据和光学遥感图像放在输入端,难以处

理建筑物高低的问题,而本文提出的多元特征图融合网络模型能够较好地利用两种数据的互补性,较好的解决了这个问题,提升了遥感图像语义分割的效果。通过对比图3(b)可以发现,本文提出的MFPNet模型对不同尺度的目标都有较好的分割效果,同时在分割边界的处理上也明显优于其他方法。

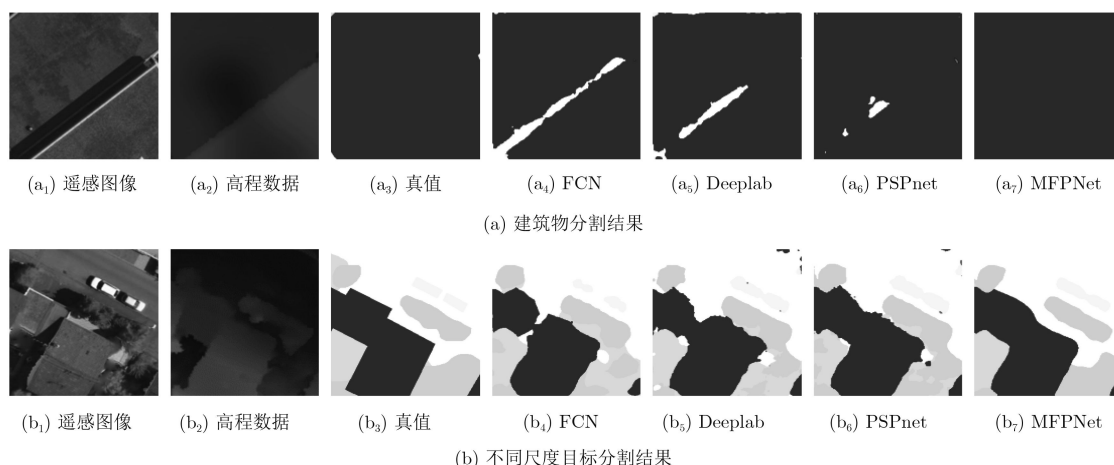


图 3 不同方法分割结果对比图

4 总结

本文针对遥感图像语义分割中多元数据的处理问题与小目标分割问题,提出了多特征图金字塔融合深度网络(MFPNet)。该模型由特征编码,金字塔池化和特征解码3部分组成。特征编码从多元数据中提取多层次特征,金字塔池化在多层次特征的基础上进一步提取多尺度特征,在特征解码阶段对多层次、多尺度特征进行融合,提升遥感图像的分割效果。本文在Vaihingen数据集上进行了实验,实验证明了提出模型的有效性。

参 考 文 献

- [1] DALAL N and TRIGGS B. Histograms of oriented gradients for human detection[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005: 886–893.
- [2] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91–110. doi: [10.1023/B:VISI.0000029664.99615.94](https://doi.org/10.1023/B:VISI.0000029664.99615.94).
- [3] SHOTTON J, JOHNSON M, and CIPOLLA R. Semantic texton forests for image categorization and segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, USA, 2008: 1–8.
- [4] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. ImageNet classification with deep convolutional neural networks[C]. The 25th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, 2012: 1097–1105.
- [5] LONG J, SHELHAMER E, and DARRELL T. Fully convolutional networks for semantic segmentation[C]. The IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 3431–3440.
- [6] KAMPPFFMEYER M, SALBERG A B, and JENSSEN R. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks[C]. The IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, USA, 2016: 1–9.
- [7] MAGGIORI E, TARABALKA Y, CHARPIAT G, *et al.* Convolutional neural networks for large-scale remote-sensing image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(2): 645–657. doi: [10.1109/TGRS.2016.2612821](https://doi.org/10.1109/TGRS.2016.2612821).
- [8] SHELHAMER E, LONG J, and DARRELL T. Fully convolutional networks for semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640–651. doi: [10.1109/TPAMI.2016.2572683](https://doi.org/10.1109/TPAMI.2016.2572683).
- [9] MARMANIS D, WEGNER J D, GALLIANI S, *et al.* Semantic Segmentation of Aerial Images with an Ensemble of CNNs[J]. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016, III-3: 473–480. doi: [10.5194/isprsannals-III-3-473-2016](https://doi.org/10.5194/isprsannals-III-3-473-2016).
- [10] SHERRAH J. Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery[J].

- arXiv: 1606.02585, 2016.
- [11] ZHAO Hengshuang, SHI Jianping, QI Xiaojuan, *et al.* Pyramid scene parsing network[C]. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2016: 6230–6239.
- [12] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778.
- [13] HAZIRBAS C, MA L N, DOMOKOS C, *et al.* FuseNet: Incorporating depth into semantic segmentation via fusion-based CNN architecture[C]. The 13th Asian Conference on Computer Vision, Taipei, China, 2016.
- [14] ISPRS 2D semantic labeling contest[EB/OL]. <http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html>, 2019.
- [15] ABADI M, BARHAM P, CHEN Jianmin, *et al.* TensorFlow: A system for large-scale machine learning[C]. The 12th USENIX Conference on Operating Systems Design and Implementation, Savannah, USA, 2016.
- [16] 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. doi: [10.1109/TPAMI.2017.2699184](https://doi.org/10.1109/TPAMI.2017.2699184).
- 赵 斐：男，1974年生，高级工程师，研究方向为遥感图像目标检测。
- 张文凯：男，1990年生，助理研究员，研究方向为图像集视觉总结，遥感图像分类。
- 闫志远：女，1994年生，硕士，研究方向为遥感图像语义分割。
- 于泓峰：男，1991年生，助理研究员，研究方向为遥感图像智能解译。
- 刁文辉：男，1988年生，助理研究员，研究方向为遥感图像目标检测。