

DOI:10.13203/j.whugis20220580



文章编号:1671-8860(2022)10-1546-09

面向多源数据地物提取的遥感知识感知与多尺度特征融合网络

龚健雅¹ 张展² 贾浩巍² 周桓³ 赵元昕¹ 熊汉江²

1 武汉大学遥感信息工程学院,湖北 武汉,430079

2 武汉大学测绘遥感信息工程国家重点实验室,湖北 武汉,430079

3 香港理工大学土地测量与地理资讯学系,香港,999077

摘要:遥感地物自动提取是遥感能力解译中的关键问题,对空间信息的理解和知识发现具有重要意义。近年来,使用全卷积神经网络(fully convolutional networks, FCN)从高分影像和三维激光雷达(light detection and ranging, LiDAR)数据中提取地物信息因取得了较好效果而受到广泛关注。现有FCN网络在地物提取精度和效率等方面仍存在不足,由此提出一种基于多源数据的遥感知识感知与多尺度特征融合网络(knowledge-aware and multi-scale feature fusion network, KMFNet)。在网络编码器端融入遥感知识感知模块(knowledge-aware module, KAM),高效挖掘多源遥感数据中的遥感知识信息;在网络编码器和解码器之间添加了串并联混合空洞卷积模块(series-parallel hybrid convolution module, SPHCM),提高网络对地物多尺度特征的学习能力;在解码器端使用了渐进式多层特征融合策略,细化最终的地物分类结果。基于公开的ISPRS语义分割标准数据集,在LuoJiaNET遥感能力解译开源深度学习框架上将KMFNet与当前主流方法进行了对比。实验结果表明,所提方法提取出的地物更为完整,细节更加精确。

关键词:高分影像;三维激光雷达;地物提取;全卷积神经网络;遥感知识;多尺度特征

中图分类号:P237

文献标志码:A

遥感影像的地物自动提取是遥感能力解译领域的一个重要研究课题,在国土资源规划、自然环境监测、智慧城市、国防等领域发挥了重要作用^[1-2]。近年来,对地观测技术的快速发展极大地提高了遥感数据的获取质量和更新速度,使得多源遥感数据变得广泛可用,如高分影像和三维激光雷达(light detection and ranging, LiDAR)等数据为高精度地物信息提取提供了巨大机遇^[3]。然而,细节信息复杂、信息量冗余的多源遥感数据也同时给地物提取任务带来了一系列问题与挑战^[4]。

传统基于单一高分影像的地物自动提取方法主要包括基于像元方法与面向对象方法两大类。前者主要聚焦于像素的光谱、纹理等特征信息,使用基于阈值或基于特征空间聚类的方法来进行地物提取^[5];后者以地物斑块作为最小分析对象,利用地物斑块的光谱特征、几何特

征和多个特征组合来实现地物提取,其主要包括基于几何边界特征的方法、基于区域分割的方法和基于图模型的方法^[6]。随着LiDAR技术的快速发展,其获取的三维点云数据的处理理论得到了广泛且深入的研究。传统基于LiDAR数据的地物自动提取方法有以支持向量机、随机森林等算法为代表的监督学习方法和以随机抽样一致性和霍夫变换等算法为代表的非监督学习方法^[7]。在地物提取任务中,高分影像可提供丰富的光谱、纹理以及几何等信息,LiDAR数据则可通过其得到的数字地表模型(digital surface model, DSM)来提供地物的空间几何信息。基于数据融合的地物提取方法可以充分发挥这两者的数据优势,有效提高地物提取结果的精度和可靠性^[8]。然而,传统基于多源遥感数据的地物提取方法仍存在精度有限、算法鲁棒性低等局限。

收稿日期:2022-09-08

项目资助:国家自然科学基金(42090011,41971402)。

第一作者:龚健雅,博士,教授,中国科学院院士,长期从事地理信息理论和摄影测量与遥感基础研究。gongjy@whu.edu.cn

通讯作者:张展,博士生。zhangzhanstep@whu.edu.cn

近年来,基于深度学习的全卷积神经网络(fully convolutional networks, FCN)因具有端到端的强大特征表达和像素级分类能力,在遥感地物提取中取得了很好的效果^[9]。目前,使用基于编码器-解码器架构的FCN模型来学习多源遥感数据中的地物特征能够有效提高地物提取的精度和可靠性^[10]。其中,常用的处理方式是将基于LiDAR数据得到的DSM和基于影像得到的归一化植被指数(normalized difference vegetation index, NDVI)等作为网络辅助或额外的图像特征,或对不同数据源使用多个网络分别进行特征提取后,再进行特征融合来实现地物提取^[11-13]。然而,这种处理方式不仅缺乏对多源遥感数据像元内在知识信息的充分挖掘,也增加了模型大小和计算量,提升了模型学习难度。此外,在不同遥感场景下地物复杂多变的空间尺寸大小也给模型的特征提取和学习带来了巨大挑战。尽管目前不少学者基于图像特征金字塔思想^[14],通过设计

出的多尺度网络结构极大提升了地物提取精度^[15-16],但地物的多尺度特征信息仍有进一步挖掘的潜力。

为解决上述问题,本文提出了遥感知识感知模块(knowledge-aware module, KAM)与串并联混合空洞卷积模块(series-parallel hybrid convolution module, SPHCM),设计了一种遥感知识感知与多尺度特征融合网络(knowledge-aware and multi-scale feature fusion network, KMFNet),提升了多源遥感数据的地物提取精度。

1 研究方法

1.1 网络整体架构

KMFNet 总体架构如图 1 所示。网络输入为 512×512 大小的多波段高分影像(近红外、红光、绿光波段)和 LiDAR 生成的归一化数字表面模型(normalized digital surface model, nDSM),经过网络模型后输出得到相同大小的地物分类图。

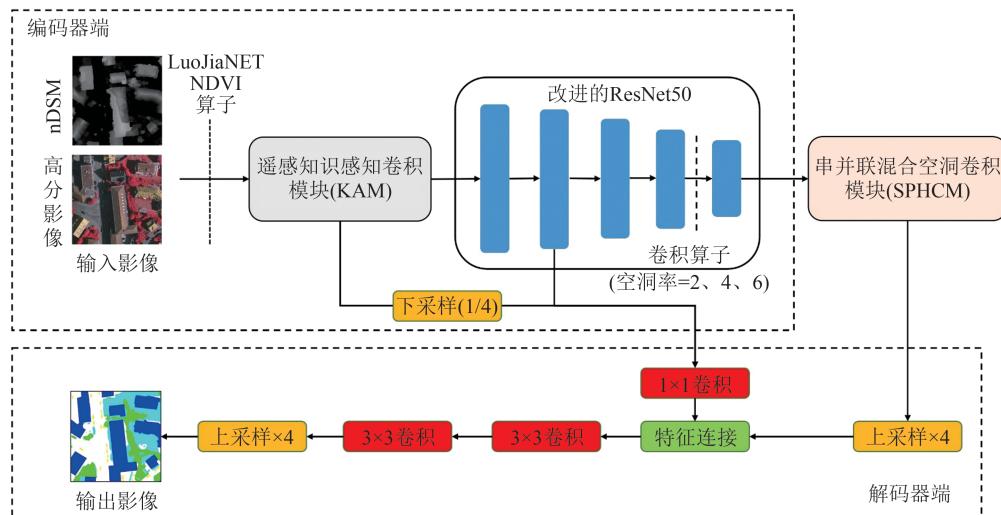


图 1 遥感知识感知与多尺度特征融合网络

Fig.1 Structure of KMFNet

KMFNet 网络在编码器端采用深度残差网络 ResNet50^[17]作为其基础架构,同时进行了相应改进:(1)前端添加了遥感知识感知卷积模块(详见§1.2),该模块能够灵活、有效地将提取得到的遥感知识显式地融入网络中;(2)在第 5 个网络层,使用了空洞率分别为 2、4、6 的空洞卷积^[18]来增大网络的感受野大小,提高网络对地物上下文语义信息的获取能力;(3)在网络编码器端和解码器端之间添加了针对高层次语义特征进行处理的串并联混合空洞卷积模块(详见§1.3),有效提高了网络对于地物多尺度上下文

语义特征的学习能力。

在网络解码器端,KMFNet 采用了一种渐进式多层特征融合策略来最大程度恢复影像分类结果中的细节信息。该策略共汇总了三支来自编码器端网络的特征信息,包括编码器端的 KAM 模块的遥感知识特征、改进的 ResNet50 模块第 2 个网络层的浅层特征以及 SPHCM 模块的深层次特征。为实现不同类型特征信息的融合,在编码器端将遥感知识特征影像空间大小下采样至 1/4 尺寸,与改进的 ResNet50 提供的浅层特征信息一同输入到解

码器端。为了减少特征波段数量、提高网络学习效率,在解码器端采用 1×1 卷积和上采样操作来实现浅层特征和深层特征的融合,随后经过2层 3×3 卷积和上采样操作,得到最终细化的地物分类结果图。

1.2 遥感知识感知模块

本文在网络前端设计的KAM由3条不同网络分支并联组成,其结构如图2所示。第1条网络分支接受3波段影像的输入,由 7×7 标准卷积、批量归一化层和激活层组成,第2条和第3条网络分支则分别接受nDSM和NDVI影像的输入,它们经 7×7 遥感知识感知卷积层处理得到对应的遥感知识特征图,最后将所有分支的特征进行融合。该模块能够有效挖掘多源遥感数据中像元内在知识信息,将其无缝融入网络模型中。

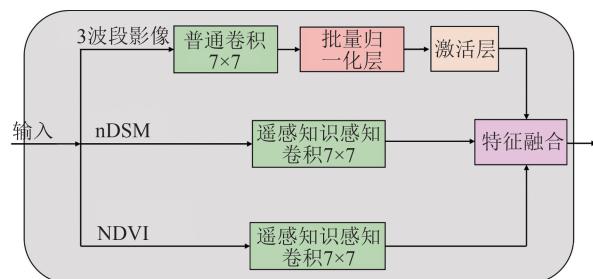


图2 遥感知识感知模块

Fig.2 Knowledge-Aware Module

KAM中所用到的遥感知识感知卷积,其主要设计思路是使用了一种以高斯函数模型为计算函数的卷积核,来定量化描述像元间的相似信息,并将卷积核内所有像元对应的函数返回值的均值作为卷积窗口中间像元的输出结果。计算函数和求均值的计算式分别为:

$$F(v_{ij}) = e^{-\frac{(v_{ij} - v_0)^2}{2\delta^2}} \quad (1)$$

$$K(P_k) = \frac{\sum_{i=1}^{n^2} F(v_{ij})}{n^2} \quad (2)$$

式中, P_k 代表特征图上的像元; v_{ij} 代表卷积核大小内的所有像元, i,j 分别代表卷积核内像元位置的宽、高索引; v_0 代表卷积核中心的像元值; δ 代表高斯函数模型中的标准差项,取值为8.1; n 代表卷积核窗口大小,取值为7; $F(v_{ij})$ 代表卷积核内所有像元经过高斯函数运算后的返回值; $K(P_k)$ 代表KAM模块在像元 P_k 处的返回值。

该模型的基本假设为:在卷积窗口大小范

围内,与窗口中心像元具有相似值的像元会被认为与其有更大的类别相关性,进而被赋予更大的像元值,从而得到能够表达像元间相似信息的遥感知识特征图,其整体计算过程如图3所示。

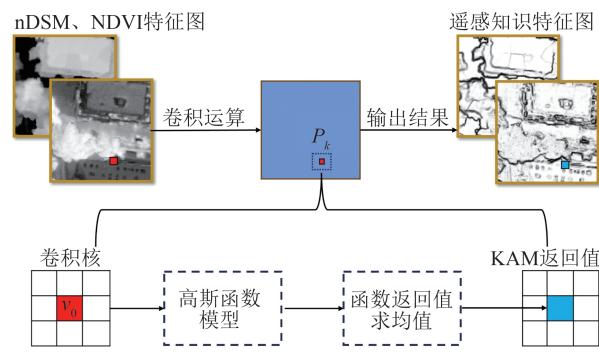


图3 遥感知识感知卷积运算示意图
Fig.3 Calculation Process of Knowledge-Aware Convolution

1.3 串并联混合空洞卷积模块

为加强网络对各类地物的多尺度及上下文信息的学习能力,本文在KMFNet的编码器和解码器端之间插入了串并联混合空洞卷积模块(series-parallel hybrid convolution module, SPHCM)。如图4所示,在SPHCM模块中,网络编码器端输出的高层次特征图在该模块中经过了三条主分支,中间分支通过 1×1 卷积层降低其通道数,并结合了不同空洞率($d=4, 8, 12, 16$)的空洞卷积层、 1×1 卷积层、归一化层和激活层,这些不同类型的网络层通过串并联的方式结合在一起,在多个网络感受野大小下得到4类不同的输出特征图。其他两条分支得到的输出特征图则作为额外添加的输入特征,输入到SPHCM模块后端特征融合层内。

在SPHCM模块后端,将6条不同路径分支输出特征图在相同空间大小基础上进行融合,随后经过 1×1 卷积运算层、归一化层和激活层,最终输出到网络模型的解码器端。本文提出的SPHCM模块能够通过不断叠加和组合空洞卷积层,在不同感受野大小下捕捉地物的多尺度特征,有效提高了其对于各类地物的全局上下文特征信息的挖掘能力。

2 实验与结果分析

2.1 数据集介绍与预处理

本文选用ISPRS语义分割标准数据集对各类地物提取方法进行了对比与测试(<http://>

www.isprs.org/commissions/comm3/wg4/semantic-labeling.html)。该数据集具有多个由不同空间尺寸大小地物(不透水面、建筑、低矮植被、树木、车辆)组成的复杂城市场景,能够很好地验证不同地物提取方法的精度和泛化性。数据集主要包括:(1)高分正射遥感影像;(2)与遥感影像对应的nDSM;(3)基于遥感影像人工标注得到的地物真实类型标签。其中,斯图加特(Vaihingen)数据集原始影像空间分辨率为0.09 m,波茨坦(Potsdam)数据集原始影像空间分辨率为0.05 m。Potsdam数据集影像含有近

红外波段、红光波段、绿光波段和蓝光波段,而Vaihingen数据集影像只含有3个波段(近红外、红光、绿光)。为保证与Vaihingen数据集的可比性,本文只保留了Potsdam数据集影像的近红外波段、红光波段和绿光波段作为网络模型的输入。考虑到当前图形处理器(graphics processing unit, GPU)性能,本文分别将这两个数据集随机裁减成4 200个和12 800个 512×512 像素的图块,从中选取部分影像进行实验,其基本参数以及训练集、验证集与测试集分配如表1所示。

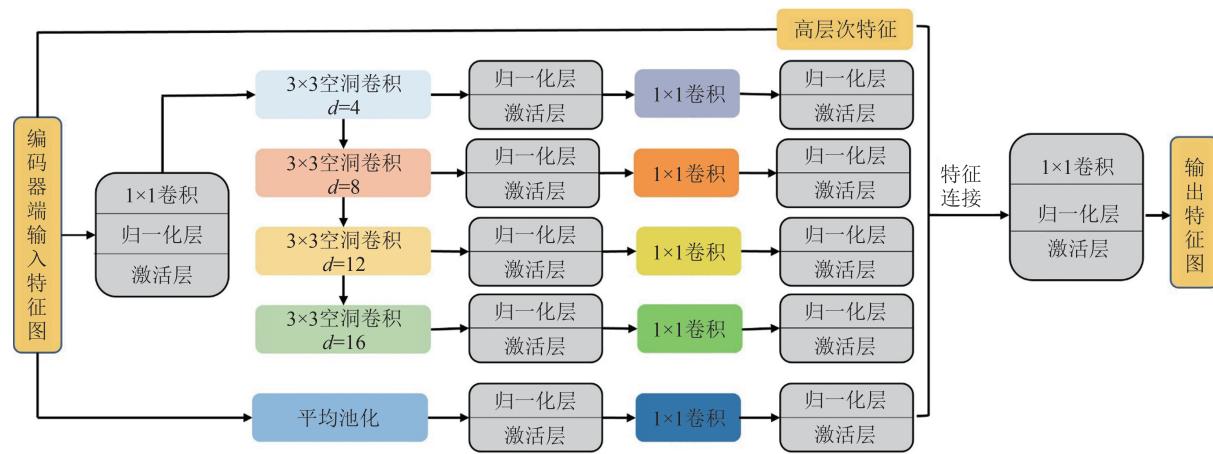


图4 串并联混合空洞卷积模块
Fig.4 Series-Parallel Hybrid Convolution Module

表1 数据集属性与训练分配

Tab. 1 Basic Attribute and Training Assignment
of Datasets

数据集	分辨率 /m	图块尺寸 /像素	训练集 /张	验证集 /张	测试集 /张
Vaihingen	0.09	512×512	1 200	150	350
Potsdam	0.05	512×512	6 000	800	2 000

2.2 实验环境与参数设置

FCN语义分割网络的特征学习过程通常需要大量的训练数据,使用一定的数据增强方法能够有效改善网络的学习效率,同时避免网络的过拟合问题来增强模型的泛化能力^[19]。在网络的每一个训练批次中,网络对于每一个影像块随机使用水平翻转、垂直翻转、镜面翻转、转置等数据增强操作。本文训练所使用的目标函数的表达式为:

$$L = - \sum_{i=1}^K y_i \log(P_i) \quad (3)$$

式中,K代表类别数量; y_i 代表符号函数,如果样本类别是*i*,则 $y_i=1$,否则为0; P_i 是网络的输出,也是观测样本类别*i*的预测概率。

为检验KMFNet的地物提取效果,引入了目前性能较好、应用广泛的GRRNet^[20]、V-FuseNet^[11]、DLR^[12]、Res-U-Net^[21]作为对比方法。这5种网络都在ISPRS语义分割标准数据集上进行测试,使用相同的训练集、验证集与测试集。本文实验统一在单块NVIDIA RTX Titan 2080上使用LuoJiaNET遥感能解译开源深度学习框架进行了实现(<https://github.com/WHULuoJiaTeam/luojianet>)。

在网络的每一个训练批次中,网络使用梯度自动下降(stochastic gradient descent, SGD)策略来进行自动学习与训练。其中,网络设置的初始学习率为0.000 1,学习率的权重衰减值为0.000 5,动量值为0.9,批尺寸大小为4,训练轮数为300,网络模型中总的训练循环次数为40 000,当循环次数在15 000次与25 000次之间时,网络学习率每隔5 000次会减少为当前学习率的1/10。网络中的初始化参数均使用了He初始值方法来进行设置。

KMFNet的输入数据包括多波段的高分影像及其对应的nDSM,基于遥感影像人工标注

得到的地物真实类型标签则作为网络监督训练过程中的真实参考数据。另外,对于遥感知识感知模块所输入的NDVI影像,本文基于网络输入的高分影像,使用LuoJiaNET框架自带的NDVI遥感先验知识提取算子来计算获得。

2.3 精度评价指标

本文使用了语义分割任务中最常用的总体精度(overall accuracy, OA)和交并比分数(intersection over union, IoU)这两类评价指标来评价各方法的提取效果,其计算式分别为:

$$OA = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (4)$$

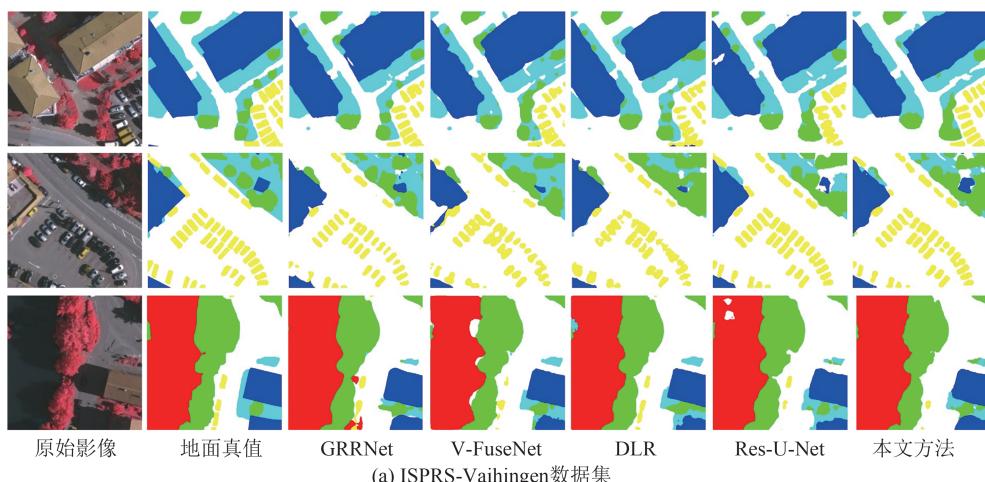
$$IoU = \frac{TP}{TP + FN + FP} \times 100\% \quad (5)$$

式中,TP(true positive)表示真实为正类且模型预测为正类的样本数;FP(false positive)表示真实为负类但模型预测为正类的样本数;FN(false negative)表示真实为正类但模型预测为负类的样本数;TN(true negative)表示真实为

负类且模型预测为负类的样本数。本文在§2.4中使用的平均交并比分数(mean intersection over union, mIoU)可看作每个地物类别IoU的平均值。

2.4 实验结果分析

将GRRNet、V-FuseNet、DLR、Res-U-Net与KMFNet在ISPRS语义分割标准数据集上进行测试,选取了不同遥感场景来进行对比,结果如图5所示。从图5的视觉表现来看,KMFNet在不同场景下的提取结果与地面真值的相似度最高,所提取地物的整体和边界细节最为完整,效果要明显优于其他4种对比方法。例如KMFNet可以保留建筑物精确的位置信息,其内部出现的空洞较少,提取出的建筑物具有丰富的边界信息,而其他方法则出现了不同程度的建筑物粘连或边界细节缺失的问题;在密集的车辆停放区域,由于车辆之间间距小、密度大,导致其难以被区分,KMFNet能够很好地区分出不同的车辆。



(a) ISPRS-Vaihingen数据集

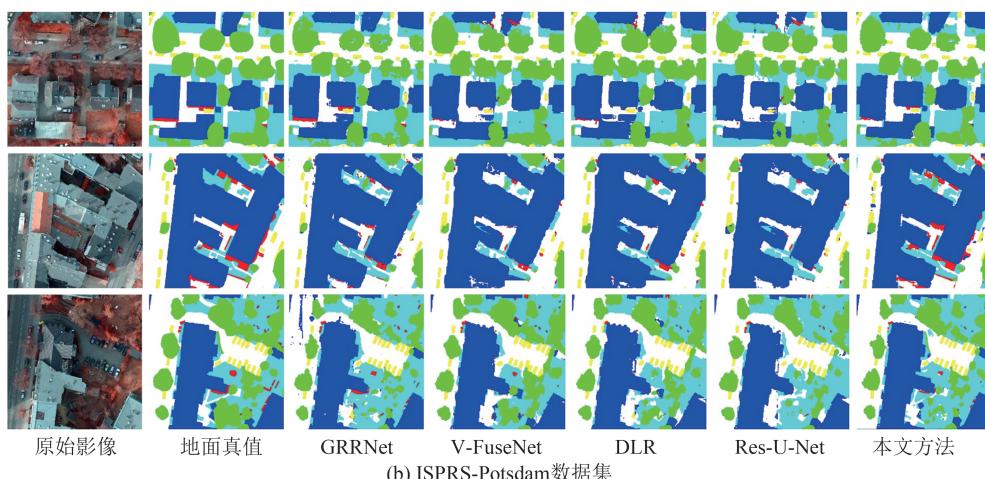


图5 不同方法地物提取结果

Fig.5 Results of Ground Object Extraction Using Different Methods

表 2 为 KMFNet 在 ISPRS 两个不同数据集上的地物提取精度。为了验证本文提出的 KAM 和 SPHCM 模块的效果, 将基于 ResNet50 改进的编码器-解码器端语义分割框架作为基准网络 (Baseline), 在 ISPRS 数据集上进行了一系列的消融实验, 结果见表 3。从表 3 可以看出, 在 Baseline 上添加 KAM 模块后, 网络模型在 Vaihingen 数据集上的 OA 和 mIoU 分别提高了

2.76% 和 3.11%; 在 Potsdam 数据集上的 OA 和 mIoU 分别提高了 1.01% 和 2.12%。而在 Baseline 上同时添加 KAM 模块和 SPHCM 模块后, 网络模型在 Vaihingen 数据集上的 OA 能达到 85.69%, mIoU 能达到 72.84%; 在 Potsdam 数据集上的 OA 能达到 86.67%, mIoU 能达到 73.69%。其精度结果相较于 Baseline 有明显提高, 可见本文提出模块的有效性。

表 2 KMFNet 在不同数据集上的分类精度 /%

Tab.2 Classification Accuracy of KMFNet in Different Datasets /%

数据集	不透水面		建筑		低矮植被		树木		车辆		平均	
	IoU	OA										
Vaihingen	79.21	86.21	85.32	90.27	67.35	86.36	76.54	88.44	55.78	77.18	72.84	85.69
Potsdam	78.05	85.32	86.76	91.71	68.12	86.44	74.92	87.26	60.02	82.63	73.69	86.67

表 3 本文所提模块在不同数据集上的消融实验结果 /%

Tab.3 Ablation Study of the Proposed Modules in Different Datasets /%

模型	Vaihingen 数据集		Potsdam 数据集	
	mIoU	OA	mIoU	OA
Baseline	67.32	80.11	67.45	83.21
Baseline + KAM	70.43	82.87	69.57	84.22
Baseline + KAM + SPHCM	72.84	85.69	73.69	86.67

为了比较 KMFNet 与其他 4 种方法的提取精度, 本文在 ISPRS 语义分割标准数据集上进行地物提取, 其统计结果见表 4。由表 4 可以看出,

KMFNet 在总体精度上具有更好的表现。在 Vaihingen 数据集上, 相较于 GRRNet、V-FuseNet、DLR、Res-U-Net, KMFNet 的总体精度分别提升了 2.19%、4.63%、1.93%、4.02%; 相较于其他方法, 对不透水面、建筑、低矮植被、树木这 4 类地物的提取精度最高, 对车辆的提取结果略低于 DLR。在 ISPRS 的 Potsdam 数据集上, 相较于 GRRNet、V-FuseNet、DLR、Res-U-Net, KMFNet 的总体精度分别提升了 1.52%、3.46%、1.75%、4.54%; 相较于其他方法, 对建筑、低矮植被、树木、车辆这 4 类地物的提取精度最高, 对不透水面的提取结果略低于 GRRNet。

表 4 不同地物自动提取方法在不同数据集上的分类精度 /%

Tab.4 Classification Accuracy of Different Methods in Different Datasets /%

方法	Vaihingen 数据集						Potsdam 数据集					
	不透水面	建筑	低矮植被	树木	车辆	总体精度	不透水面	建筑	低矮植被	树木	车辆	总体精度
GRRNet	84.23	90.11	82.56	85.32	75.26	83.50	85.56	89.21	83.21	86.01	81.77	85.15
V-FuseNet	82.61	86.88	80.2	83.11	72.48	81.06	84.01	87.33	82.64	83.06	79.02	83.21
DLR	81.24	88.25	84.27	87.02	78.01	83.76	83.65	89.69	84.23	85.23	81.79	84.92
Res-U-Net	81.99	86.31	81.21	84.94	73.92	81.67	81.22	85.98	80.26	83.54	79.67	82.13
KMFNet	86.21	90.27	86.36	88.44	77.18	85.69	85.32	91.71	86.44	87.26	82.63	86.67

Res-U-Net 整合了两种常用网络框架, 在网络模型的输出端使用了一个导向滤波后处理模块来优化建筑物提取结果, 但是在特征提取阶段没有针对性地进行多尺度特征提取, 导致在地物对象提取完整性上表现不佳。V-FuseNet 注重对多源数据不同特征的融合方式进行挖掘, 设计了一种残差融合结构来完成特征融合, 但是在网络解码器端没有对地物特征

细节进行补充, 提取效果有待提升。DLR 在多个分支网络中使用了一种地物对象边界检测机制, 在实验中取得了较好的地物提取效果, 但在特征提取阶段缺乏对像元间遥感先验知识的显式挖掘, 故在建筑物、植被等地物的提取效果上不如 KMFNet。GRRNet 通过在 ResNet50 的基础上引入门控特征标记单元, 通过提高网络中特征的传输和学习效率, 达到了

较高的地物提取精度,但是该方法仅简单地将nDSM作为高分影像额外的特征波段一起输入到网络中,缺乏对多源遥感数据特征的深度融合,削弱了地物提取的效果。KMFNet在特征提取阶段使用了KAM模块来对像元间的相似信息进行高效、显式地挖掘,对遥感知识特征进行了深度融合;在特征处理阶段引入了SPHCM模块,增强对不同尺度地物特征的学习能力,故具有最好的提取效果。另外,KMFNet在网络解码器端使用了一种渐进式多层次特征融合策略,极大补充了地物提取结果的细节特征,在保证精度的同时具有较好的计算效率。

在性能测试上,当训练影像为1 200张、影像大小为 512×512 像素、批尺寸为4时,KMFNet、GRRNet、V-FuseNet、DLR、Res-U-Net在LuoJiaNET框架中完成一轮训练的平均耗时分别为840.24 s、893.23 s、1 284.33 s、1 547.34 s、1 165.79 s,可见KMFNet在性能上具有一定优势。

3 结语

针对目前FCN语义分割网络对于多源遥感数据学习效率不高、地物提取结果缺乏细节信息等问题,本文在改进的深度残差网络ResNet50结构基础上,融入提出的遥感知识感知模块和串并联混合空洞卷积模块,设计了一种遥感知识感知与多层次特征融合网络KMFNet,同时在网络解码器端采用了渐进式多层次特征融合策略,得到了高精度地物提取结果。遥感知识感知模块使用了一种遥感知识感知卷积核来高效、显式地挖掘多源遥感数据中的遥感知识特征。串并联混合空洞卷积模块在6种不同路径的感受野尺寸下捕捉了地物的多层次特征,加强了网络对于上下文全局信息的学习能力。本文使用LuoJiaNET遥感能力解译开源深度学习框架,在公开的ISPRS语义分割标准数据集上对多种主流的地物提取方法进行了大量实验和评估。实验结果表明,与GRRNet、V-FuseNet、DLR、Res-U-Net这4种方法相比,KMFNet能够在多源遥感数据上取得最佳地物提取效果。为进一步提高地物提取的准确率,可尝试在网络中融入更多遥感先验知识。

参考文献

- [1] Gong Jianya. Chances and Challenges for Development of Surveying and Remote Sensing in the Age of Artificial Intelligence [J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(12): 1788-1796 (龚健雅. 人工智能时代测绘遥感技术的发展机遇与挑战[J]. 武汉大学学报·信息科学版, 2018, 43(12): 1788-1796)
- [2] Li Yansheng, Zhang Yongjun. A New Paradigm of Remote Sensing Image Interpretation by Coupling Knowledge Graph and Deep Learning [J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(8): 1176-1190 (李彦胜, 张永军. 植物识别图谱和深度学习的新一代遥感影像解译范式[J]. 武汉大学学报·信息科学版, 2022, 47(8): 1176-1190)
- [3] Xu Enen, Guo Ying, Chen Erxue, et al. An Estimation Model for Regional Forest Canopy Closure Combined with UAV LiDAR and High Spatial Resolution Satellite Remote Sensing Data [J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(8): 1298-1308 (徐恩恩, 郭颖, 陈尔学, 等. 基于无人机LiDAR和高空间分辨率卫星遥感数据的区域森林郁闭度估测模型[J]. 武汉大学学报·信息科学版, 2022, 47(8): 1298-1308)
- [4] Zhang Chenglong, Li Zhenhong, Zhang Shuangcheng, et al. Surface Ruptures of the 2022 Mw 6.7 Menyuan Earthquake Revealed by Integrated Remote Sensing [J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(8): 1257-1270 (张成龙, 李振洪, 张双成, 等. 综合遥感解译2022年Mw 6.7青海门源地震地表破裂带[J]. 武汉大学学报·信息科学版, 2022, 47(8): 1257-1270)
- [5] Su T F, Li H Y, Zhang S W, et al. Image Segmentation Using Mean Shift for Extracting Croplands from High-Resolution Remote Sensing Imagery [J]. *Remote Sensing Letters*, 2015, 6(12): 952-961
- [6] Kotaridis I, Lazaridou M. Remote Sensing Image Segmentation Advances: A Meta-Analysis [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 173: 309-322
- [7] Du S J, Zhang Y S, Zou Z R, et al. Automatic Building Extraction from LiDAR Data Fusion of Point and Grid-Based Features [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2017, 130: 294-307
- [8] Zhang J X. Multi-Source Remote Sensing Data Fusion: Status and Trends [J]. *International Journal*

- of Image and Data Fusion*, 2010, 1(1): 5–24
- [9] Fu G, Liu C J, Zhou R, et al. Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network [J]. *Remote Sensing*, 2017, 9(5): 498
- [10] Sun W W, Wang R S. Fully Convolutional Networks for Semantic Segmentation of very High Resolution Remotely Sensed Images Combined with DSM [J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(3): 474–478
- [11] Audebert N, Le Saux B, Lefèvre S. Beyond RGB: Very High Resolution Urban Remote Sensing with Multimodal Deep Networks [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, 140: 20–32
- [12] Marmanis D, Schindler K, Wegner J D, et al. Classification with an Edge: Improving Semantic Image Segmentation with Boundary Detection [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, 135: 158–172
- [13] Pan X R, Gao L R, Marinoni A, et al. Semantic Labeling of High Resolution Aerial Imagery and LiDAR Data with Fine Segmentation Network [J]. *Remote Sensing*, 2018, 10(5): 743
- [14] Zhao H S, Shi J P, Qi X J, et al. Pyramid Scene Parsing Network [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017
- [15] Liu Y C, Fan B, Wang L F, et al. Semantic Labeling in very High Resolution Images via a Self-Cascaded Convolutional Neural Network [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, 145: 78–95
- [16] Li Y H, Chen Y T, Wang N Y, et al. Scale-Aware Trident Networks for Object Detection [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, South Korea, 2019
- [17] 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, Las Vegas, NV, USA, 2016
- [18] Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions [EB/OL]. [2015-04-12] <https://arxiv.org/abs/1511.07122>
- [19] Shorten C, Khoshgoftaar T M. A Survey on Image Data Augmentation for Deep Learning [J]. *Journal of Big Data*, 2019, 6(1): 1–48
- [20] Huang J F, Zhang X C, Xin Q C, et al. Automatic Building Extraction from High-Resolution Aerial Images and LiDAR Data Using Gated Residual Refinement Network [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019, 151: 91–105
- [21] Xu Y Y, Wu L, Xie Z, et al. Building Extraction in very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters [J]. *Remote Sensing*, 2018, 10: 144

Multi-source Data Ground Object Extraction Based on Knowledge-Aware and Multi-scale Feature Fusion Network

GONG Jianya¹ ZHANG Zhan² JIA Haowei² ZHOU Huan³
ZHAO Yuanxin¹ XIONG Hanjiang²

1 School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

2 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan 430079, China

3 Department of Land Surveying and Geo-Informatics, Hong Kong Polytechnic University, Hong Kong 999077, China

Abstract: **Objectives:** In recent years, the automatic ground object extraction from remote sensing images has been dramatically advanced by the fully convolutional networks (FCNs). It is an effective method to fuse high-resolution images and light detection and ranging (LiDAR) data in FCNs to improve the extraction accuracy and the robustness. However, the existing FCN-based fusion networks still face challenges in efficiency and accuracy. **Methods:** We propose a knowledge-aware and multi-scale fusion network (KMFNet) for robust and accurate ground object extraction. The proposed network incorporates a knowledge-aware module in the network encoder for better exploiting remote sensing knowledge between pixels. A series-parallel hybrid convolution module is developed to enhance multi-scale representative features from ground objects. Moreover, the network decoder uses a gradual bilinear interpolation strategy to obtain fine-grained extraction results. **Results:** We evaluate KMFNet in the LuoJiaNET with four current mainstream

ground object extraction methods (GRRNet, V-FuseNet, DLR and Res-U-Net) on ISPRS 2D semantic segmentation dataset. The comparative evaluation results show that KMFNet can obtain the best overall accuracy. Compared with the other four methods, KMFNet achieves a better effect by improving the overall accuracy by 3.20% and 2.82% on average in ISPRS-Vaihingen dataset and ISPRS-Potsdam dataset, respectively. **Conclusions:** KMFNet achieves the best extraction results by capturing the intrinsic pixel relationships and strengths the multi-scale representative and detailed features of ground objects. It shows great potential in high-precision remote sensing mapping applications.

Key words: high-resolution image; light detection and ranging (LiDAR); ground object extraction; fully convolutional network (FCN); remote sensing knowledge; multi-scale features

First author: GONG Jianya, PhD, professor, Academician of Chinese Academy of Sciences, specializes in geo-informatics and photogrammetry. E-mail: gongjy@whu.edu.cn

Corresponding author: ZHANG Zhan, PhD candidate. E-mail: zhangzhanstep@whu.edu.cn

Foundation support: The National Natural Science Foundation of China (42090011, 41971402).

引文格式: GONG Jianya, ZHANG Zhan, JIA Haowei, et al. Multi-source Data Ground Object Extraction Based on Knowledge-Aware and Multi-scale Feature Fusion Network[J]. Geomatics and Information Science of Wuhan University, 2022, 47(10):1546-1554. DOI:10.13203/j.whugis20220580(龚健雅,张展,贾浩巍,等.面向多源数据地物提取的遥感知识感知与多尺度特征融合网络[J].武汉大学学报·信息科学版,2022,47(10):1546-1554. DOI:10.13203/j.whugis20220580)