



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象学报 图形学报

2020
05
VOL.25

ISSN1006-8961
CN11-3758/TB



智能疫
“小珈”
助力



第25卷第5期（总第289期）
2020年5月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 顾行发
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@radi.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Institute of Remote Sensing and Digital Earth, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Gu Xingfa
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



多相图像分割Vese-Chan模型连续最大流方法(第0926页)



保持区域平滑的3维动画形状高效编辑(第1019页)



树形结构卷积神经网络优化的城区遥感图像语义分割(第1043页)

学者观点

多源空一谱遥感图像融合方法进展与挑战

肖亮, 刘鹏飞, 李恒 0851

综述

中国图像工程: 2019

章毓晋 0864

图像处理和编码

相同编码参数HEVC视频重压缩检测

潘鹏飞, 姚晔, 王慧 0879

密集连接卷积网络图像去模糊

吴迪, 赵洪田, 郑世宝 0890

密集网络图像哈希检索

王亚鸽, 康晓东, 郭军, 李博, 张华丽, 刘汉卿 0900

图像分析和识别

域自适应城市场景语义分割

张桂梅, 潘国峰, 刘建新 0913

多相图像分割Vese-Chan模型连续最大流方法

王洁, 潘振宽, 魏伟波, 徐子森 0926

结合注意力机制和多属性分类的行人再识别

郑鑫, 林兰, 叶茂, 王丽, 贺春林 0936

多分辨率特征注意力融合行人再识别

沈庆, 田畅, 王家宝, 焦珊珊, 杜麟 0946

全卷积网络电线识别方法

刘嘉玮, 李元祥, 龚政, 刘心刚, 周拥军 0956

双核压缩激活神经网络艺术图像分类

杨秀芹, 张华熊 0967

图像理解和计算机视觉

高效检测复杂场景的快速金字塔网络SPNet

李鑫泽, 张轩雄, 陈胜 0977

结合自底向上注意力机制和记忆网络的视觉问答模型

闫茹玉, 刘学亮 0993

统筹图像变换与缝合线生成的无参数影像拼接

高炯笠, 吴军, 刘祺昌, 徐刚 1007

计算机图形学

保持区域平滑的3维动画形状高效编辑

冼楚华, 杨煜, 黄俊贤, 李桂清 1019

医学图像处理

肌骨超声图像特征检测及拼接

颜焕欢, 张培镇, 王伊依, 李璇, 阳维, 王青 1032

遥感图像处理

树形结构卷积神经网络优化的城区遥感图像语义分割

胡伟, 高博川, 黄振航, 李瑞瑞 1043

融合累积变异比和集成超限学习机的高光谱图像分类

尹玉萍, 魏林, 刘万军 1053

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Continuous max-flow method for multiphase segmentation Vese-Chan model(P0926)



Efficient 3D animation shape manipulation with region smoothness preservation(P1019)



Semantic segmentation of urban remote sensing image based on optimized tree structure convolutional neural network (P1043)

Scholar View

Progress and challenges in the fusion of multisource spatial-spectral remote sensing images

Xiao Liang, Liu Pengfei, Li Heng 0851

Review

Image engineering in China: 2019

Zhang Yujin 0864

Image Processing and Coding

Detection of double compression for HEVC videos with the same coding parameters

Pan Pengfei, Yao Ye, Wang Hui 0879

Motion deblurring method based on DenseNets

Wu Di, Zhao Hongtian, Zheng Shibao 0890

Image Hash retrieval with DenseNet

Wang Yage, Kang Xiaodong, Guo Jun, Li bo, Zhang Huali, Liu Hanqing 0900

Image Analysis and Recognition

Domain adaptation for semantic segmentation based on adaption learning rate

Zhang Guimei, Pan Guofeng, Liu Jianxin 0913

Continuous max-flow method for multiphase segmentation Vese-Chan model

Wang Jie, Pan Zhenkuan, Wei Weibo, Xu Zisen 0926

Improving person re-identification by attention and multi-attributes

Zheng Xin, Lin Lan, Ye Mao, Wang Li, He Chunlin 0936

Multi-resolution feature attention fusion method for person re-identification

Shen Qing, Tian Chang, Wang Jiabao, Jiao Shanshan, Du Lin 0946

Power line recognition method via fully convolutional network

Liu Jiawei, Li Yuanxiang, Gong Zheng, Liu Xingang, Zhou Yongjun 0956

Art image classification with double kernel squeeze-and-excitation neural network

Yang Xiuqin, Zhang Huaxiong 0967

Image Understanding and Computer Vision

Snap Pyramid Network: Real-time complex scene detecting system

Li Xinze, Zhang Xuanxiong, Chen Sheng 0977

Visual question answering model based on bottom-up attention and memory network

Yan Ruyu, Liu Xueliang 0993

Stitching of parametric-free images based on coordinated image transformation and seam-line generation

Gao Jiongli, Wu Jun, Liu Qichang, Xu Gang 1007

Computer Graphics

Efficient 3D animation shape manipulation with region smoothness preservation

Xian Chuhua, Yang Yu, Huang Junxian, Li Guiqing 1019

Medical Image Processing

Feature detection algorithm of musculoskeletal ultrasound image and its application of image stitching

Yan Huanhuan, Zhang Peizhen, Wang Yinong, Li Xuan, Yang Wei, Wang Qing 1032

Remote Sensing Image Processing

Semantic segmentation of urban remote sensing image based on optimized tree structure convolutional neural network

Hu Wei, Gao Bochuan, Huang Zhenhang, Li Ruirui 1043

Ensemble extreme learning machine with cumulative variation quotient for hyperspectral image classification

Yin Yuping, Wei Lin, Liu Wanjuan 1053

中图法分类号: TP751 文献标识码: A 文章编号: 1006-8961(2020)05-1043-10

论文引用格式: Hu W, Gao B C, Huang Z H and Li R R. 2020. Semantic segmentation of urban remote sensing image based on optimized tree structure convolutional neural network. Journal of Image and Graphics, 25(05): 1043-1052 (胡伟, 高博川, 黄振航, 李瑞瑞. 2020. 树形结构卷积神经网络优化的城区遥感图像语义分割. 中国图象图形学报, 25(05): 1043-1052) [DOI: 10.11834/jig.190324]

树形结构卷积神经网络优化的城区遥感图像语义分割

胡伟, 高博川, 黄振航, 李瑞瑞

北京化工大学信息科学与技术学院, 北京 100029

摘要: **目的** 高分辨率遥感图像通常包含复杂的语义信息与易混淆的目标, 对其语义分割是一项重要且具有挑战性的任务。基于 DeepLab V3+ 网络结构, 结合树形神经网络结构模块, 设计出一种针对高分辨率遥感图像的语义分割网络。**方法** 提出的网络结构不仅对 DeepLab V3+ 做出了修改, 使其适用于多尺度、多模态的数据, 而且在其后添加连接树形神经网络结构模块。树形结构通过建立混淆矩阵、提取混淆图、构建图分割, 能够对易混淆的像素更好地区分, 得到更准确的分割结果。**结果** 在国际摄影测量及遥感探测学会 (International Society for Photogrammetry and Remote Sensing, ISPRS) 提供的两个不同城市的遥感影像集上分别进行了实验, 模型在整体准确率 (overall accuracy, OA) 这一项表现最好, 在 Vaihingen 和 Potsdam 数据集上分别达到了 90.4% 和 90.7%, 其整体分割准确率较其基准结果有 10.3% 和 17.4% 的提升, 对比 ISPRS 官方网站上的 3 种先进方法也有显著提升。**结论** 提出结合 DeepLab V3+ 和树形结构的卷积神经网络, 有效提升了高分辨率遥感图像语义分割整体精度, 其中易混淆类别数据的分割准确率显著提高。在包含复杂语义信息的高分辨率遥感图像中, 由于易混淆类别之间的像素分割错误减少, 使用了树形结构的网络模型的整体分割准确率也有较大提升。

关键词: 卷积神经网络; 遥感图像; 语义分割; 树形结构; DeepLab V3+

Semantic segmentation of urban remote sensing image based on optimized tree structure convolutional neural network

Hu Wei, Gao Bochuan, Huang Zhenhang, Li Ruirui

College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

Abstract: **Objective** High-resolution remote sensing image segmentation refers to the task of assigning a semantic label to each pixel in an image. Recently, with the rapid development of remote sensing technology, we have been able to easily obtain very-high resolution remote sensing images with a ground sampling distance of 5 cm to 10 cm. However, the very heterogeneous appearance of objects, such as buildings, streets, trees, and cars, in very-high-resolution data makes this task challenging, leading to high intraclass variance while the inter-class variance is low. A research hotspot is on detailed 2D semantic segmentation that assigns labels to multiple object categories. Traditional image processing methods depend on the extraction technique of the vectorization model, for example, based on region segmentation, line analysis, and shadow analysis. Another mainstream study relies on supervised classifiers with manually designed features. These models were not generalized when dealing with high-resolution remote sensing images. Recently, deep learning-based technology has helped explore the high-level semantic information in imaged and provide an end-to-end approach for semantic segmentation. **Method** Based on DeepLab V3+, we proposed an adaptive constructed neural network, which contains two connected modules, namely, the segmentation module and the tree module. When segmenting remote-sensing images, which contain

收稿日期: 2019-07-09; 修回日期: 2019-09-16; 预印本日期: 2019-09-23

multiscale objects, understanding the context is important. To handle the problem of segmenting objects at multiple scales, DeepLab V3+ employs atrous convolution in cascade or in parallel captures multiscale context by adopting multiple atrous rates. We adopted a similar idea in designing the segmentation module. This module uses an encoder-decoder architecture. The encoder is composed of four structures: EntryFlow, MiddleFlow, ExitFlow, and atrous spatial pyramid pooling (ASPP). In addition, the decoder is composed of two layers of SeparableConv blocks. The middle flow has two Xception blocks, which are linear stacks of depth-separable convolutional layers with residual connections. The segmentation module could capture well the multiscale features in the context. However, these features pay less attention to the easily confused classes. The other core contribution of the proposed method is the tree module. This module is constructed adaptively during the training. In each round, the method computes the confusion matrix on the evaluation data and calculates the confusion degrees between every two classes. A graph could be constructed according to the confusion matrix, and we can obtain a certain tree structure through the minimum cut algorithm. According to the tree structure, we build the tree module, in which each node is a ResNeXt unit. These nodes are connected by the concatenated connections. The tree module helped distinguish the pixels between easily confused classes by adding several neural layers to process their features. To implement the proposed method, the segmentation model is based on the MXNet framework and uses two Nvidia GeForce GTX1080 Ti graphic cards for accelerated training. The input size of the image block is 640×640 pixels due to memory limitation. We set the momentum (momentum) to 0.9, the initial learning rate to 0.01, adjust the learning rate to 0.001 when the training reaches half, and adjust the learning rate to 0.0001 when the training reaches 3/4. We perform data augmentation before training due to the small amount of data in the ISPRS (International Society for Photogrammetry and Remote Sensing) remote sensing dataset. For each piece of raw data, we rotate the image center by 10° each time and cut out the largest square tile. In this way, each training image can obtain 36 sets of pictures after rotation. In addition, because the original training image is very large in size, to directly place the entire image into the network for training is not possible. Thus, it needs to be cropped into image blocks of 640×640 pixels. We apply an overlap-tile strategy to ensure no obvious cracks in the segmentation result map after splicing.

Result The model in this study performed the best in terms of overall accuracy (OA), which reached 90.4% and 90.7% on the Vaihingen and Potsdam datasets, respectively. This result indicates that the model can achieve high segmentation accuracy. In addition, in the performance of easily confused categories, for example, low shrub vegetation (low_veg) and trees, F_1 values have greatly improved. In experiments of the Vaihingen dataset, the F_1 values of the low shrub vegetation and tree species reached 83.6% and 89.6%, respectively. In experiments of the Potsdam dataset, the F_1 values of the low shrub vegetation and tree species reached 86.8% and 87.1%. As for the average F_1 value, the scores of the model on the Vaihingen and Potsdam datasets are 89.3% and 92.0%, respectively. This number is much higher than the other latest methods, indicating that the model in this study is the best for both the segmentation of remote sensing images and the average performance of each category. Additionally, compared with the model without the tree module, the proposed method has higher segmentation accuracy for each category. The OA by using tree module increased by 1.1%, and the average F_1 value increased by 0.6% in the Vaihingen dataset. The OA and the average F_1 value increased by 1.3% and 0.9% in the Potsdam dataset. The result shows that the tree module does not only target a certain category but also improves the overall segmentation accuracy.

Conclusion The proposed network can effectively improve the overall semantic segmentation accuracy of high-resolution remote sensing images. The experimental results show that the proposed segmentation module with the tree module is greatly improved due to the reduction of the error on easily confused pixels. The proposed method in this study is universal and suitable for a wide range of application scenarios.

Key words: convolutional neural networks (CNN); remote sensing images; semantic segmentation; tree like structure; DeepLab V3+

0 引言

高分辨率遥感图像语义分割任务是指为遥感图

像中的每一个像素分配语义标签的过程。近年来,随着遥感测绘技术的高速发展,已经可以轻而易举地获得地面采样间隔(ground sample distance, GSD)为 $5 \sim 10$ cm 的超高分辨率的光学遥感图像

(Audebert 等, 2018)。城区高分辨率遥感图像以人工地物建筑为主体部分, 辅以一些自然植被用地, 人工地物建筑主要包括房屋、机场、道路、桥梁等。例如: 对房屋目标的准确分割, 可以迅速获得住宅房屋密度等城市用地指标, 为城市进一步规划提供依据。因此, 如何准确地理解上下文语义, 对这些图像像素进行标注成为遥感图像分割领域的研究热点。

高分辨率遥感图像包含丰富的语义信息, 大多数传统方法无法有效地表征, 分割效果不理想。早期, 人工地物建筑的分割依赖矢量化的模型提取技术(孙金彦 等, 2017; 王海 等, 2014), 例如: 基于区域分割的方法、直线分析、阴影分析等。其他大量的研究依赖人工设计的特征, 通过有监督的分类器实现分割操作。人工设计的特征往往在表示高级语义信息时, 泛化性较差。深度学习技术能够自动提取特征(张康 等, 2018; 王璐 等, 2015), 可以充分挖掘图像中的高级语义信息特征。

深度学习技术已经在计算机视觉领域取得了巨大成功, 例如图像分类(Krizhevsky 等, 2012; Hu 等, 2015)、目标检测(李旭冬 等, 2017)、语义分割(Long 等, 2015)等。深度卷积神经网络接收原始图像数据输入, 以端到端的结构进行学习, 根据特定任务得到最终分割结果。在遥感图像解译领域, 深度卷积神经网络开展了广泛的研究。城区遥感图像包含的语义信息复杂(Cordts 等, 2016; 谭琨 等, 2019), 具有人工地物目标种类多样、小型目标众多(Kampffmeyer 等, 2016)的特征, 其易混淆数据往往在空间分布上是相邻或交错的, 分割难度较大。全卷积神经网络(fully convolutional network, FCN)在2016年(Maggiori 等, 2016)首次应用于遥感图像分割任务, 可以接受任意大小的图像输入和测试, 并且避免了使用像素块带来的重复存储和计算卷积的问题, 相比于传统的带有全连接的卷积神经网络(convolutional neural networks, CNN)更加高效, 但是其分割结果不够精细, 对图像的细节保留不够完整。在此基础上, “沙漏状网络”包括反卷积网络(DeconvNet)(Noh 等, 2015)、SegNet(Badrinarayanan 等, 2017)、U-Net(Ronneberger 等, 2015)和DeepUNet(Li 等, 2018)等方法陆续被提出, 应用于遥感图像分割。这些网络在其解码器结构上进行了不同调整,

并且取得了更高的分割准确率。

基于空洞卷积, DeepLab(Chen 等, 2017)提出了空间金字塔模块, 其在输入特征图上应用多采样率的空洞卷积、多感受野卷积或池化操作, 以探索多尺度上下文信息。DeepLab V3+(Chen 等, 2018)不仅采用了可分离卷积, 同时为了融合多尺度信息, 引入语义分割常用的编码器—解码器结构, 通过逐渐恢复空间信息来捕捉清晰的分割目标边界, 细化分割结果, 是目前图像语义分割领域表现较好的网络框架模型。

本文提出一种改进的 DeepLab V3+ 的网络模型, 在 DeepLab V3+ 网络后面添加一种新颖的树形优化模块, 用于提高对易混淆类别数据的分割能力。该方法采用端到端的网络训练, 在国际摄影测量及探测学会(International Society for Photogrammetry and Remote Sensing, ISPRS)两个遥感数据集上, 与最新方法相比, 得到了更高的分割精度。

1 语义分割网络结构

对于多类别的遥感图像分割任务, 由于易混淆的类别数据在空间分布上通常是相邻或交错分布, 此外一般的网络模型很难学习到有效的特征表示, 因此分割的准确率并不高。本文为了改善易混淆类别数据分割准确率低的问题, 设计了一种基于 DeepLab V3+ 结构改进的树形网络模型。网络的主体结构如图1所示, 该网络结构分为两个部分: 分割模块和树形模块。

1.1 分割模块

考虑到城区遥感图像的特点, 在提取密集特征的基础上还要兼顾分割边缘的准确性, 因此本文选用 DeepLab V3+ 网络作为分割模型, 如图2所示。DeepLab V3+ 的网络模型主要由两部分组成: 编码器部分和解码器部分。编码器模块由进入流、中间流、退出流和空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP) 4个结构组成, 解码器由两层 SeparableConv 块构成。其中多个 Xception 块是带有残差连接的深度可分卷积层的线性堆叠, SeparableConv 即可分卷积层。相比于原来的 DeepLab V3+ 结构, 训练所用的图像数据和计算资源有限, 因此本文将中间流结构中的 Xception 单元由 16 个降为 2 个。

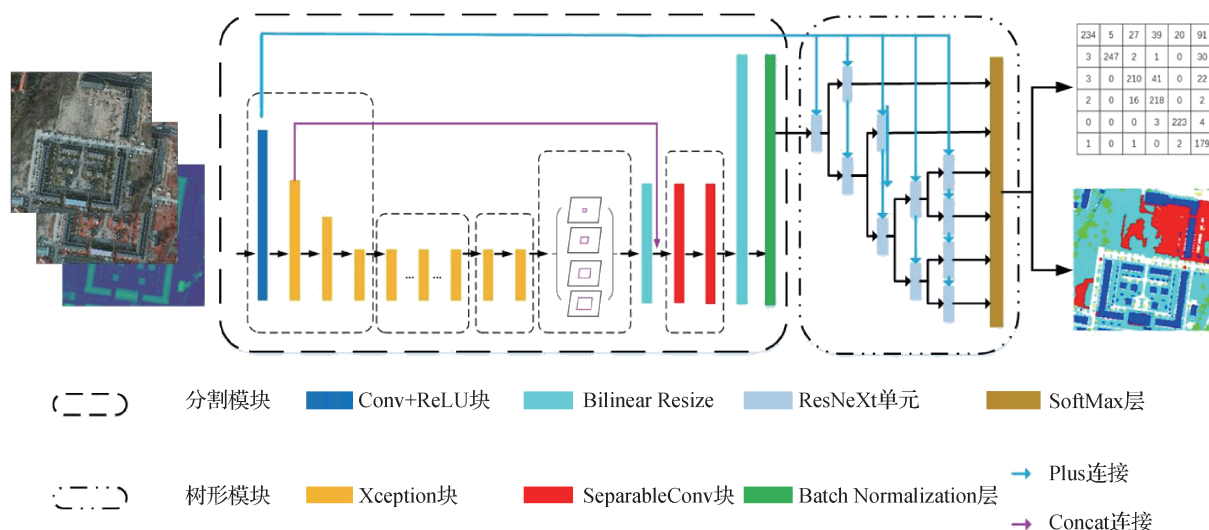


图1 网络模型整体示意图

Fig. 1 Overall schematic diagram of the network

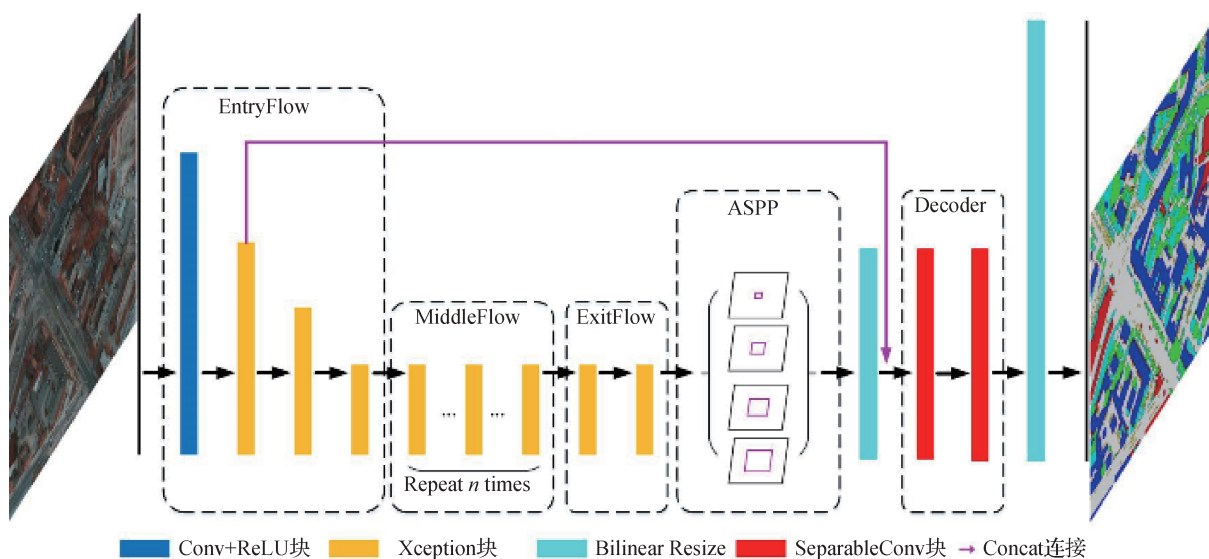


图2 DeepLab V3+:具有空洞卷积和ASPP的编码器—解码器结构

Fig. 2 DeepLab V3+: encoder-decoder with atrous convolution and ASPP

1.2 树形模块

本文网络模型的核心是树形处理结构。虽然DeepLab V3+网络具有优秀的性能,能够提取合理的上下文语义特征,也能够通过上采样最大限度保留边界信息,但是仍然解决不了在相邻区域中存在的易混淆类别像素的准确预测问题。树形网络模块是连接在基础网络模块后面的一个增强模块,它和基础网络模块一起构成了一个端到端的分割网络。树形网络通过计算混淆矩阵自适应构建,其作用是让易混淆类别的像素通过更深的树状层次,让易区分类别的像素通过较浅的树状

层次,树的节点通过残差链接有效避免了梯度消失,延迟过拟合,提升网络精度和泛化性。ISPRS遥感数据集共包含6类数据,因此将树形结构设计为一个具有6个节点的二叉树状模型。树状结构中的每一个节点均为一个ResNeXt单元,结构如图3所示,它需要经过32组相互独立且具有相同结构的整流变换流程,最后将结果进行融合。ResNeXt单元(Xie等,2017)是ResNet(He等,2016)结构的改进版,它能够在不增加参数复杂度的情况下提高准确率,每一个ResNeXt单元通过Plus连接第1层的卷积层,保证能够获得足够

的特征信息量。除此之外, ResNeXt 单元不仅可以避免在训练过程中出现梯度消失的现象, 而且

还可以通过减少超参数的数量来减小显存的训练开销。

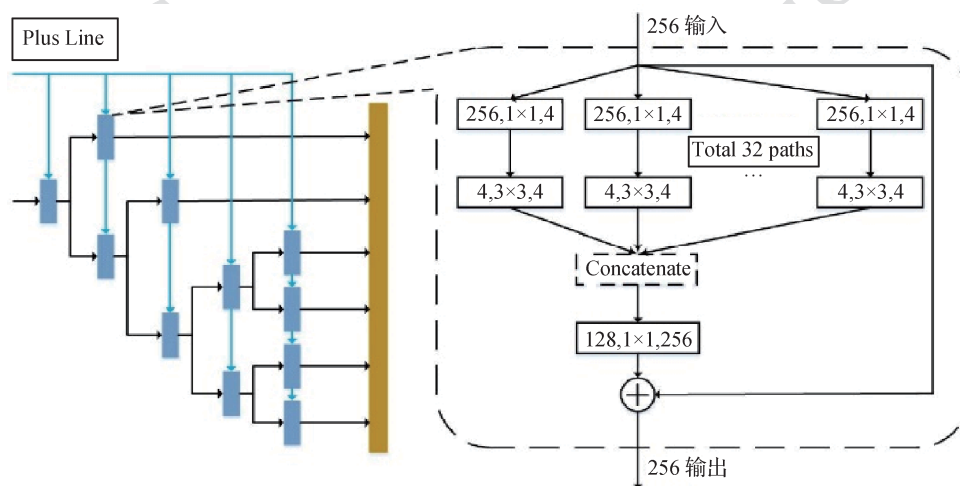


图3 树形模块示意图

Fig. 3 Illustration of tree-like block

通过先验的分割结果, 可以计算得到混淆矩阵 A 。因为数据集中一共包含 6 个类别, 所以混淆矩阵 A 是一个 6×6 的矩阵。将相对应的行列的元素 a_{ij} 、 a_{ji} 相加, 得到对应的下三角矩阵 B 。矩阵 B 的元素 b_{ij} 为

$$b_{ij} = \begin{cases} a_{ij} + a_{ji} & i > j \\ 0 & \text{其他} \end{cases} \quad (1)$$

三角矩阵 B 作为邻接矩阵, 可以得到一个无向图 G 。该无向图中共有 6 个顶点, 分别代表 ISPRS 数据集中的 6 个类别。利用最小割的方法 (Boykov 和 Kolmogorov, 2004), 依次割去图中的顶点, 可以逐步得到子类别。以此先后顺序得到一个 6 个节点的二叉树结构, 此结构即为本文网络结构中的树形模块。由于网络连接了树形模块, 能够获得足够的信息量, 在小型物体和易混淆物体的分割任务中能够得到更加精确的分割结果。

本文的网络模型是一个全卷积神经网络, 并不存在全连接层, 通过树形模块之后, 所有的特征图将会被送入一个 1×1 的卷积层, 并通过 Softmax 函数进行输出。

2 实验数据集

本文在 ISPRS Vaihingen 和 Potsdam 两个遥感数据集 (Gerke, 2014) 上进行了实验, 并且展示了实验

结果, 数据集如图 4 所示。ISPRS 遥感数据集是一个在网上公开的遥感高分辨率图像数据集, 主要包含了城市中心和周围的环境情况, 其中包含遥感图像中最常见的土地覆盖类别: 非渗透表面 (impervious surfaces)、建筑 (building)、低灌木植被 (low veg)、树木 (tree)、车辆 (car) 和杂波层 (clutter)。

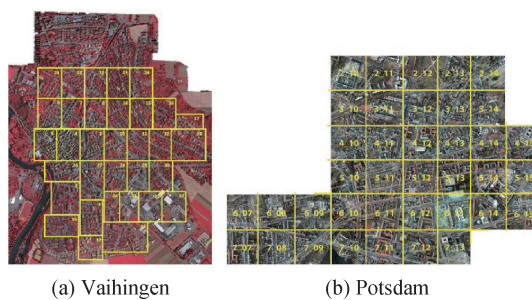


图4 ISPRS 数据集示意图

Fig. 4 Overview of the ISPRS dataset

((a) Vaihingen; (b) Potsdam)

2.1 Vaihingen 数据集

该数据集共包含 33 幅由无人机摄影机在德国 Vaihingen 镇上空拍摄的高分辨率图像, 这组数据包含红外、红和绿 (IR-Red-Green, IRRG) 三通道的格式, 数字地表模型 (digital surface model, DSM) (吴军等, 2015) 格式和数据标注。每幅图像的平均尺寸为 $2\,494 \times 2\,046$ 像素, 空间分辨率为 9 cm。本组实验将随机选取 11 幅图像作为训练集, 5 幅作为验证

集以及 17 幅作为测试集。

2.2 Potsdam 数据集

Potsdam 数据集共包含 38 幅俯拍高分辨率遥感图像,图像尺寸均为 6 000 × 6 000 像素,空间分辨率为 5 cm。这组数据有 RGB 格式、DSM 格式、IRRG 格式和数据标注。由于数据集中有一幅图像存在大量错误标注(编号 7_10),因此本组实验将随机选取 17 幅图像作为训练集、5 幅图像作为验证集和 15 幅图像作为测试集。

3 实验评价标准

分割实验以整体准确率(OA)作为评价标准。对每一类数据的分割表现,使用 F_1 值进行评价,它是由精确率(P)和召回率(R)计算得来。使用类别间的平均 F_1 值来评价整体类别分割表现,是由于整体准确率对于不平衡分布的数据的评价并不敏感。整体准确率和 F_1 值计算为

$$OA = \frac{tp + tn}{tp + tn + fp + fn} \quad (2)$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \quad (3)$$

$$P = \frac{tp}{tp + fp} \quad (4)$$

$$R = \frac{tp}{tp + fn} \quad (5)$$

式中, tp 表示将正类预测为正类数即真正率, tn 表示将负类预测为负类数即真负率, fp 表示将负类预测为正类数即假正率, fn 表示将正类预测为负类数即假负率。

4 实验环境及数据预处理

本文分割模型基于 MXNet 框架实现,使用 2 块 Nvidia GeForce GTX1080 Ti 显卡进行加速训练。由于显存的限制,训练中输入的图像块大小为 640 × 640 像素,批量大小为 4。训练 80 个回合(epoch),设置动量(momentum)为 0.9,初始学习率设置为 0.01,当训练到达一半时将学习率调整为 0.001,当训练到达 3/4 时再将学习率调整为 0.000 1。

ISPRS 遥感数据集中数据量较小,需要进行数据增广来扩充训练数据:对于每幅原始数据,以图像中心每次旋转 10°,并截取出最大的正方形图块。

这样,每幅训练图像可以在旋转后得到 36 幅图像。

另外,原始训练图像的尺寸非常大,不能直接将整幅图像放入网络中进行训练,需要裁剪为 640 × 640 像素大小的图像块。为了保证在拼接后的分割结果图中不会出现明显的裂痕,需要使用影像重叠策略(overlap-tile strategy)。影像重叠策略是在原始影像内部切割的时候进行重叠,在原始影像边缘部分进行镜像反射外推,如图 5 所示,白色实线为训练图像的边界,红色区域为切割图像时的大小,而黄色区域为实际分割时有效区域的大小。



图 5 影像重叠策略示意图

Fig. 5 Schematic diagram of overlapping tiles

5 实验结果及分析

分别在 Vaihingen 和 Potsdam 两个数据集上进行了实验,实验结果见表 1 和表 2。其中楼梯视觉库(stair vision library, SVL) (Gerke, 2014)的方法为 ISPRS 主办方提供的基准结果,基于 Dempster-Shafer 理论(Dempster-Shafer theory, DST)的方法(Sherrah, 2016)为 FCN 加入条件随机场(conditional random field, CRF)单元的结果,苏黎世大学的方法(University of Zurich, UZ) (Volpi 和 Tuia, 2017)则是反卷积网络的分割结果,以上结果均引自 ISPRS 官方网站。

5.1 实验整体结果分析

从表 1 和表 2 可以看出,本文模型均在整体准确率这一项中表现最好,在 Vaihingen 和 Potsdam 数据集上分别达到了 90.4% 和 90.7%,说明该模型具有较高的分割精度。另外,在低灌木植被(low_veg)和树木(tree)这两个相互易混淆的类别的表现上,相比较于其他网络模型, F_1 值均有提升。在 Vaihingen 数据集的实验中,本文模型的低灌木植被和树

表1 不同模型在 Vaihingen 测试数据集上的分割结果对比

Table 1 Quantitative comparisons among different models on the ISPRS Vaihingen test set

方法	imp_surf	building	low_veg	tree	car	OA	平均 F_1
SVL(Gerke, 2014)	86.6	91.0	77.0	85.0	55.6	84.8	79.0
DST(Sherrah, 2016)	90.5	93.7	83.4	89.2	72.6	89.1	85.9
UZ(Volpi 和 Tuia, 2017)	89.2	92.5	81.6	86.9	57.3	87.3	81.5
本文	92.5	94.9	83.6	89.6	85.9	90.4	89.3

注:加粗字体为每列最优结果。

表2 不同模型在 Potsdam 测试数据集上的分割结果对比

Table 2 Quantitative comparisons among different models on the ISPRS Potsdam test set

方法	imp_surf	building	low_veg	tree	car	OA	平均 F_1
SVL(Gerke, 2014)	83.5	91.7	72.2	63.2	62.2	77.8	74.6
DST(Sherrah, 2016)	92.5	96.4	86.7	88.0	94.7	90.3	91.7
UZ(Volpi 和 Tuia, 2017)	89.3	95.4	81.8	80.5	86.5	85.8	86.7
本文	93.1	97.3	86.8	87.1	95.8	90.7	92.0

注:加粗字体为每列最优结果。

木类别的 F_1 值分别为 83.6% 和 89.6%; 在 Potsdam 数据集的实验中, 本文模型的低灌木植被和树木类别的 F_1 值分别为 86.8% 和 87.1%。对于平均 F_1 值这一指标, 本文模型在 Vaihingen 和 Potsdam 数据集上的分数分别是 89.3% 和 92.0%, 远远高于其他 3 种方法, 说明本文模型对于遥感图像的整体分割表现优秀, 并且在树木之外的其他类别上, 表现都是最好的。

5.2 树状模块性能分析

本文模型的一个创新点在于提出了一种树形的处理结构模块。在本节中, 对网络是否带有树形模块进行了对照实验, 实验结果如表 3 所示, 同时也使用了经典的直线型结构(Ss)和树状结构(Ts)进行

对照实验, 实验结果如表 4 所示。

具有树形结构的网络模型相比于没有该结构的模型, 对于每个类别的分割准确率均得到不同程度的提升。在 Vaihingen 数据集上, 包含树形结构的整体准确率提升了 1.1%, 平均 F_1 值提升了 0.6%; 而在 Potsdam 数据集上, 整体准确率和平均 F_1 值分别提升了 1.3% 和 0.9%。由此说明树形结构对于网络的提升并不针对于某一个类别, 而是分割精度的全面提升。精细比较每个类别, 在分割低灌木植被(low_veg)和树木(tree)这一对易混淆的类别时, 其精确度有明显提升。此外, 在杂波层(clutter)和非渗透表面(imp_surf)的边缘细节分割上也有了显著的精度提升。

表3 是否加入树形结构的对照实验结果

Table 3 Controlled experimental result with and without the tree-like block

数据集	树形结构	imp_surf	building	low_veg	tree	car	clutter	OA	平均 F_1
Vaihingen	×	91.5	94.1	82.7	89.3	85.5	47.1	89.3	88.7
	√	92.5	94.9	83.6	89.6	85.9	52.0	90.4	89.3
Potsdam	×	92.8	97.4	84.5	85.5	95.1	44.7	89.4	91.1
	√	93.1	97.3	86.8	87.1	95.8	53.0	90.7	92.0

注:加粗字体为每列最优结果, 其中 × 代表未加入树形结构, √ 代表加入了树形结构。

表4 不同附加模块的对照实验结果

Table 4 Control experiment results with different additional modules

数据集	附加模块	imp_surf	building	low_veg	tree	car	clutter	OA	平均 F_1
Vaihingen	Ss	91.1	91.4	78.9	85.8	80.4	47.1	84.3	85.5
	Ts	92.5	94.9	83.6	89.6	85.9	52.0	90.4	89.3
Potsdam	Ss	89.1	91.7	80.8	80.0	93.2	39.1	85.2	87.0
	Ts	93.1	97.3	86.8	87.1	95.8	53.0	90.7	92.0

注:加粗字体为每列最优结果。

为了保证对照实验的一致性,使用相同结构的ResNeXt单元构建模型。因为数据集分为6个类别,因此构建了一个6层的直线型附加模块作为对照实验。

树形结构模块相较于经典的直线型结构模块,在每个类别的分割准确率都有非常明显的提升,在Vaihingen数据集上,树形结构较直线型结构的整体准确率提升了6.1%,平均 F_1 值提升了3.8%;在Potsdam数据集上,整体准确率和平均 F_1 值分别提

升了5.5%和5%。加入直线型结构的模型在整体的分割准确率上甚至低于没有添加附加结构的原始模型。说明本文设计的树形结构相较于经典的直线型结构更适合存在易混淆类别的遥感图像分割任务。

是否加入树状结构的对照实验结果示例如图6所示。从图中可以看出,本文模型加入了树形结构后,分割精度明显提升,并且分割的边缘更加顺滑,没有产生气泡化现象。

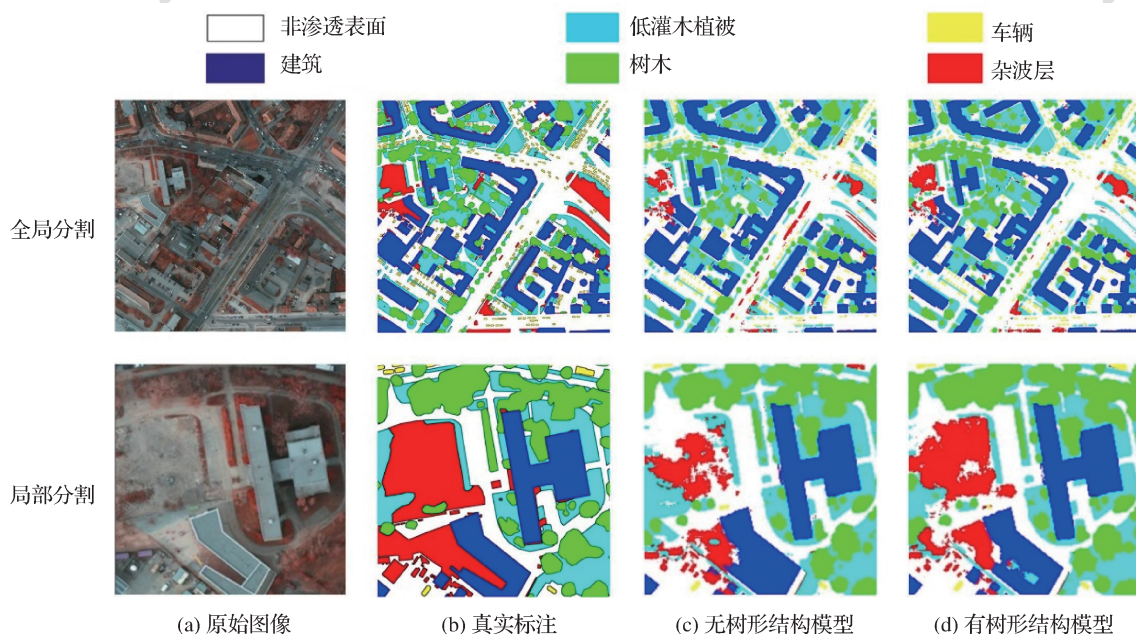


图6 测试集图像的全局和局部分割结果

Fig. 6 Global and local detailed segmentation results of test dataset

((a) original images; (b) ground truth; (c) network without tree-like block; (d) network with tree-like block)

6 结论

以DeepLab V3+作为基础结构,提出了一种基

于树形结构优化的卷积神经网络模型。该模型可以有效地改善高分辨率下的城区遥感图像语义分割问题。在ISPRS的高分辨率遥感数据集上,本模型得到的整体分割准确率较其基准结果有10.3%和

17.4%的提升,对比SVL、DST和UZ等多种先进方法也有显著提升。针对大多数类别,本模型也得到了最优的分割准确率,特别是对于易混淆类别数据,使用树形结构可以得到更顺滑更精确的分割结果。

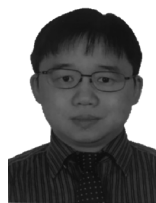
虽然树形结构对于遥感图像中易混淆类别的数据分割精度有一定程度的提升,但是未来仍有需要解决的难点:1)树形结构是否适用于所有的分割网络模型,例如基于现有的所有分割网络模型,加入该结构是否均可以在原有基础上得到一定程度的提升,仍然需要大量实验的验证;2)尽管使用了数据增广的手段,但是网络模型训练的数据量仍较小,是否可以通过生成对抗网络(generative adversarial network, GAN)的思想增大数据量,进一步提升优化模型结构,以获得更高的分割精确度,这也是未来的重点工作方向。

参考文献(References)

- Audebert N, Le Saux B and Lefèvre S. 2018. Beyond RGB: very high resolution urban remote sensing with multimodal deep networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140: 20-32 [DOI: 10.1016/j.isprsjprs.2017.11.011]
- Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]
- Boykov Y and Kolmogorov V. 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9): 1124-1137 [DOI: 10.1109/TPAMI.2004.60]
- Chen L C, Papandreou G, Schroff F and Adam H. 2017. Rethinking Atrous Convolution for Semantic Image Segmentation [EB/OL]. [2019-07-01]. <https://arxiv.org/pdf/1706.05587.pdf>
- Chen L C, Zhu Y K, Papandreou G, Schroff F and Adam H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer: 833-851 [DOI: 10.1007/978-3-030-01234-2_49]
- Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S and Schiele. 2016. The cityscapes dataset for semantic urban scene understanding//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE: 3213-3223 [DOI: 10.1109/CVPR.2016.350]
- Gerke M. 2014. Use of the Stair Vision Library within the ISPRS 2D Semantic Labeling Benchmark (Vaihingen). Holland: University of Twente
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hu W, Huang Y Y, Wei L, Zhang F and Li H C. 2015. Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, 2015: #258619 [DOI: 10.1155/2015/258619]
- Kampffmeyer M, Salberg A B and Jenssen R. 2016. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Las Vegas, NV, USA: IEEE: 680-688 [DOI: 10.1109/CVPRW.2016.90]
- Krizhevsky A, Sutskever I and Hinton G E. 2012. ImageNet classification with deep convolutional neural networks//*Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe, Nevada, USA Curran Associates Inc: 1097-1105
- Li R R, Liu W J, Yang L, Sun S H, Hu W, Zhang F and Li W. 2018. Deepunet: a deep fully convolutional network for pixel-level sea-land segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(11): 3954-3962 [DOI: 10.1109/JSTARS.2018.2833382]
- Li X D, Ye M and Li T. 2017. Review of object detection based on convolutional neural networks. *Application Research of Computers*, 34(10): 2881-2886, 2891 (李旭冬, 叶茂, 李涛. 2017. 基于卷积神经网络的目标检测研究综述. *计算机应用研究*, 34(10): 2881-2886, 2891) [DOI: 10.3969/j.issn.1001-3695.2017.10.001]
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA, USA: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Maggiori E, Tarabalka Y, Charpiat G and Alliez. 2016. Fully convolutional neural networks for remote sensing image classification//*Proceedings of 2016 IEEE International Geoscience and Remote Sensing Symposium*. Beijing, China: IEEE: 5071-5074 [DOI: 10.1109/IGARSS.2016.7730322]
- Noh H, Hong S and Han B. 2015. Learning deconvolution network for semantic segmentation//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago, Chile: IEEE: 1520-1528 [DOI: 10.1109/ICCV.2015.178]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//*Proceedings of the 18th International Conference on Medical image computing and computer-assisted intervention*. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Sherrah J. 2016. Fully convolutional networks for dense semantic labeling of high-resolution aerial imagery [EB/OL]. [2019-07-01]. <https://arxiv.org/pdf/1606.02585.pdf>

- Sun J Y, Huang Z J, Zhou S G, Xu N, Qian H M and Wang C L. 2017. Building outline vectorization from high spatial resolution imagery. *Journal of Remote Sensing*, 21(3): 396-405 (孙金彦, 黄祚继, 周绍光, 徐南, 钱海明, 王春林. 2017. 高分辨率遥感影像中建筑物轮廓信息矢量化. *遥感学报*, 21(3): 396-405) [DOI: 10.11834/jrs.20176127]
- Tan K, Wang X and Du P J. 2019. Research progress of the remote sensing classification combining deep learning and semi-supervised learning. *Journal of Image and Graphics*, 24(11): 1823-1841 (谭琨, 王雪, 杜培军. 2019. 结合深度学习和半监督学习的遥感影像分类进展. *中国图象图形学报*, 24(11), 1823-1841) [DOI: 10.11834/jig.190348]
- Volpi M and Tuia D. 2017. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2): 881-893 [DOI: 10.1109/TGRS.2016.2616585]
- Wang H, Tong H J, Zuo B X and Tang W R. 2014. Integration of multi-scale segmentation algorithm and vectorization algorithm for remote sensing image. *Computer Engineering*, 40(6): 256-261 (王海, 童恒建, 左博新, 汤文瑞. 2014. 遥感图像多尺度分割算法与矢量化算法的集成. *计算机工程*, 40(6): 256-261) [DOI: 10.3969/j.issn.1000-3428.2014.06.055]
- Wang L, Zhang F, Li W, Xie X M and Hu W. 2015. A method of SAR target recognition based on Gabor filter and local texture feature extraction. *Journal of Radars*, 4(6): 658-665 (王璐, 张帆, 李伟, 谢晓明, 胡伟. 2015. 基于 Gabor 滤波器和局部纹理特征提取的 SAR 目标识别算法. *雷达学报*, 4(6): 658-665) [DOI: 10.12000/JR15076]
- Wu J, Cheng M M, Yao Z X, Peng Z Y, Li J and Ma J. 2015. Automatic generation of high-quality urban DSM with airborne oblique images. *Journal of Image and Graphics*, 20(06): 117-128 (吴军, 程门门, 姚泽鑫, 彭智勇, 李俊, 马峻. 2015. 倾斜航空影像的城区 DSM 生成. *中国图象图形学报*, 20(06), 117-128) [DOI: 10.11834/jig.20150615]
- Xie S N, Girshick R, Dollár P, Tu Z W and He K M. 2017. Aggregated residual transformations for deep neural networks//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE: 5987-5995 [DOI: 10.1109/CVPR.2017.634]
- Zhang K, Hei B Q, Zhou Z and Li S Y. 2018. CNN with coefficient of variation-based dimensionality reduction for hyperspectral remote sensing images classification. *Journal of Remote Sensing*, 22(1): 87-96 (张康, 黑保琴, 周壮, 李盛阳. 2018. 变异系数降维的 CNN 高光谱遥感图像分类. *遥感学报*, 22(1): 87-96) [DOI: 10.11834/jrs.20187075]

作者简介



胡伟, 1979 年生, 男, 副教授, 硕士生导师, 主要研究方向为图像/视频智能处理、深度学习、海量数据可视化。

E-mail: huwei@mail.buct.edu.cn



李瑞, 通信作者, 女, 讲师, 主要研究方向为遥感图像语义分割、全景分割。

E-mail: ilydouble@gmail.com

高博川, 男, 硕士研究生, 主要研究方向为深度学习处理遥感图像的分割问题。E-mail: 2018210469@mail.buct.edu.cn

黄振航, 男, 硕士研究生, 主要研究方向为遥感图像语义分割。E-mail: pre_hzh@163.com