



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2020
12
VOL.25

ISSN1006-8961
CN11-3758/TB



数字浮雕建模 P2647



第25卷第12期 (总第296期)
2020年12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 中国科学院空天信息创新研究院

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 吴一戎

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 Aerospace Information Research Institute, CAS

China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong

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

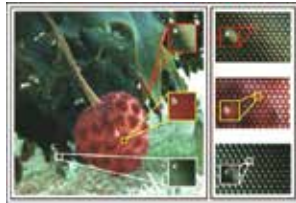
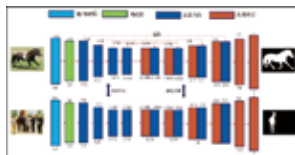
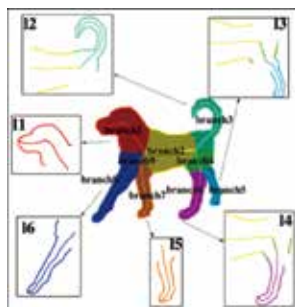
ISSN 1006-8961

CODEN ZTTXFZ

国外发行代号 M1406

国内邮发代号 82-831

国内定价 60.00元

光场显著性检测研究综述
(第2465页)融合注意力机制与知识蒸馏
的孪生网络压缩(第2563页)面向可解释性的物体拓扑结构
骨架表征方法(第2587页)

综述

光场显著性检测研究综述

刘亚美, 张骏, 张旭东, 孙锐, 高隽 2465

图像处理和编码

自适应卷积的残差修正单幅图像去雨

王美华, 何海君, 李超 2484

局部特定空间关系统计特征的RVIN噪声检测器

于海雯, 易昕炜, 徐少平, 林珍玉, 刘蕊蕊 2494

全局与局部属性一致的图像修复模型

孙劲光, 杨忠伟, 黄胜 2505

图像分析和识别

关键语义区域链提取的视频人体行为识别

马淼, 李贻斌, 武宪青, 高金凤, 潘海鹏 2517

针对形变与遮挡问题的行人再识别

史维东, 张云洲, 刘双伟, 朱尚栋, 暴吉宁 2530

多特征融合的行为识别模型

谭等泰, 李世超, 常文文, 李登楼 2541

结构特征下的可撤销人脸识别

孙浩浩, 邵珠宏, 尚媛园, 陈滨, 赵晓旭 2553

融合注意力机制与知识蒸馏的孪生网络压缩

耿增民, 余梦巧, 刘峡壁, 吕超 2563

联合聚焦度和传播机制的光场图像显著性检测

李爽, 邓慧萍, 朱磊, 张龙 2578

图像理解和计算机视觉

面向可解释性的物体拓扑结构骨架表征方法

危辉, 余莉萍 2587

自监督深度残差函数映射网络的3维模型对应关系计算

杨军, 马中昌 2603

融合自编码器和one-class SVM的异常事件检测

胡海洋, 张力, 李忠金 2614

抗高光的光场深度估计方法

王程, 张骏, 高隽 2630

计算机图形学

不同类型高度图控制浅浮雕模型生成方法

尚菁, 余文杰, 王美丽 2647

遥感图像处理

残差密集空间金字塔网络的城市遥感图像分割

韩彬彬, 张月婷, 潘宗序, 台宪青, 李芳芳 2656

结合判别相关分析与特征融合的遥感图像检索

葛芸, 马琳, 储珺 2665

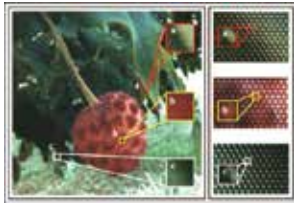
改进卷积网络的高分遥感图像城镇建成区提取

侯博文, 闫冬梅, 郝伟, 黄青青, 苏秀琴, 李青雯 2677

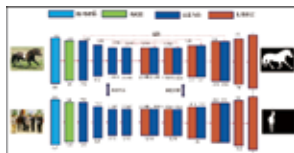
总目次 I

CONTENTS

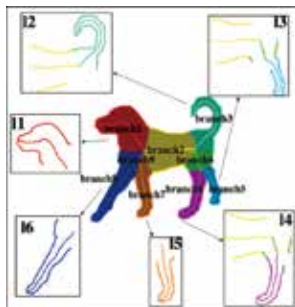
JOURNAL OF IMAGE AND GRAPHICS



Review of saliency detection on light fields(P2465)



Combining attention mechanism and knowledge distillation for Siamese network compression (P2563)



Skeleton characterization of object topology toward explainability (P2587)

Review

Review of saliency detection on light fields

Liu Yamei, Zhang Jun, Zhang Xudong, Sun Rui, Gao Jun 2465

Image Processing and Coding

Single image rain removal based on selective kernel convolution using a residual refine factor

Wang Meihua, He Haijun, Li Chao 2484

Random-valued impulse noise detector using local spatial structure statistics

Yu Haiwen, Yi Xinwei, Xu Shaoping, Lin Zhenyu, Liu Ruirui 2494

Image inpainting model with consistent global and local attributes

Sun Jinguang, Yang Zhongwei, Huang Sheng 2505

Image Analysis and Recognition

Human action recognition in videos utilizing key semantic region extraction and concatenation

Ma Miao, Li Yibin, Wu Xianqing, Gao Jinfeng, Pan Haipeng 2517

Person re-identification based on deformation and occlusion mechanisms

Shi Weidong, Zhang Yunzhou, Liu Shuangwei, Zhu Shangdong, Bao Jining 2530

Multi-feature fusion behavior recognition model

Tan Dengtai, Li Shichao, Chang Wenwen, Li Denglou 2541

Cancelable face recognition with fusion of structural features

Sun Haohao, Shao Zhuhong, Shang Yuanyuan, Chen Bin, Zhao Xiaoxu 2553

Combining attention mechanism and knowledge distillation for Siamese network compression

Geng Zengmin, Yu Mengqiao, Liu Xiabi, Lyu Chao 2563

Saliency detection on a light field via the focusness and propagation mechanism

Li Shuang, Deng Huiping, Zhu Lei, Zhang Long 2578

Image Understanding and Computer Vision

Skeleton characterization of object topology toward explainability

Wei Hui, Yu Liping 2587

Correspondence Calculation of 3D Models by a Self-Supervised Deep Residual Functional Maps Network

Yang Jun, Ma Zhongchang 2603

Anomaly detection with autoencoder and one-class SVM

Hu Haiyang, Zhang Li, Li Zhongjin 2614

Anti-specular light-field depth estimation algorithm

Wang Cheng, Zhang Jun, Gao Jun 2630

Computer Graphics

Bas-relief generation controlled by different height maps

Shang Jing, Yu Wenjie, Wang Meili 2647

Remote Sensing Image Processing

Residual dense spatial pyramid network for urbanremote sensing image segmentation

Han Binbin, Zhang Yueting, Pan Zongxu, Tai Xianqing, Li Fangfang 2656

Remote sensing image retrieval combining discriminant correlation analysis and feature fusion

Ge Yun, Ma Lin, Chu Jun 2665

Urban built-up area extraction using high-resolution remote sensing images with an improved convolutional neural network

Hou Bowen, Yan Dongmei, Hao Wei, Huang Qingqing, Su Xiuqin, Li Qingwen 2677

中图法分类号: TP751.1 文献标识码: A 文章编号: 1006-8961(2020)12-2656-09

论文引用格式: Han B B, Zhang Y T, Pan Z X, Tai X Q and Li F F. 2020. Residual dense spatial pyramid network for urban remote sensing image segmentation. Journal of Image and Graphics, 25(12): 2656-2664 (韩彬彬, 张月婷, 潘宗序, 台宪青, 李芳芳. 2020. 残差密集空间金字塔网络的城市遥感图像分割. 中国图象图形学报, 25(12): 2656-2664) [DOI: 10.11834/jig.190557]

残差密集空间金字塔网络的城市遥感图像分割

韩彬彬^{1,2,3}, 张月婷^{1,2}, 潘宗序^{1,2}, 台宪青^{1,2}, 李芳芳^{1,2}

1. 中国科学院空天信息创新研究院, 北京 100190; 2. 空间信息处理与应用系统技术重点实验室, 北京 100190;

3. 中国科学院大学, 北京 100049

摘要: 目的 遥感图像语义分割是根据土地覆盖类型对图像中每个像素进行分类, 是遥感图像处理领域的一个重要研究方向。由于遥感图像包含的地物尺度差别大、地物边界复杂等原因, 准确提取遥感图像特征具有一定难度, 使得精确分割遥感图像比较困难。卷积神经网络因其自主分层提取图像特征的特点逐步成为图像处理领域的主流算法, 本文将基于残差密集空间金字塔的卷积神经网络应用于城市地区遥感图像分割, 以提升高分辨率城市地区遥感影像语义分割的精度。**方法** 模型将带孔卷积引入残差网络, 代替网络中的下采样操作, 在扩大特征图感受野的同时能够保持特征图尺寸不变; 模型基于密集连接机制级联空间金字塔结构各分支, 每个分支的输出都有更加密集的感受野信息; 模型利用跳线连接跨层融合网络特征, 结合网络中的高层语义特征和低层纹理特征恢复空间信息。**结果** 基于 ISPRS (International Society for Photogrammetry and Remote Sensing) Vaihingen 地区遥感数据集展开充分的实验研究, 实验结果表明, 本文模型在 6 种不同的地物分类上的平均交并比和平均 F_1 值分别达到 69.88% 和 81.39%, 性能在数学指标和视觉效果上均优于 SegNet、pix2pix、Res-shuffling-Net 以及 SDFCN (symmetrical dense-shortcut fully convolutional network) 算法。**结论** 将密集连接改进空间金字塔池化网络应用于高分辨率遥感图像语义分割, 该模型利用了遥感图像不同尺度下的特征、高层语义信息和低层纹理信息, 有效提升了城市地区遥感图像分割精度。

关键词: 语义分割; 遥感影像; 多尺度; 残差卷积网络; 密集连接

Residual dense spatial pyramid network for urban remote sensing image segmentation

Han Binbin^{1,2,3}, Zhang Yueting^{1,2}, Pan Zongxu^{1,2}, Tai Xianqing^{1,2}, Li Fangfang^{1,2}

1. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China;

2. Key Laboratory of Technology in Geo-Spatial Information Processing and Application Systems, Beijing 100190, China;

3. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: **Objective** Remote sensing image semantic segmentation, in which each pixel in an image is classified according to the land cover type, presents an important research direction in the field of remote sensing image processing. However, accurately segmenting and extracting features from remote sensing images is difficult due to the wide coverage of these images and the large-scale difference and complex boundaries among these features. Meanwhile, the traditional remote sensing image processing methods are inefficient, inaccurate, and require much expertise. Convolutional neural networks are deep

收稿日期: 2019-10-30; 修回日期: 2020-03-13; 预印本日期: 2020-03-20

基金项目: 国家重点研发计划项目 (2016YFF0202700); 国家自然科学基金项目 (61701478)

Supported by: National Key Research and Development Program of China (2016YFF0202700); National Natural Science Foundation of China (61701478)

learning networks that are suitable for processing data with grid structures, such as 1D data with time series features (e. g., speech) and image data with 2D pixel matrix grids. Given its multi-layer structure, a convolutional neural network can automatically learn features at different levels. This network also has two features that facilitate image processing. First, a convolutional neural network uses the 2D characteristics of an image in feature extraction. Given the high correlation among adjacent pixels in an image, the neuron nodes in the network do not need to connect all pixels; only a local connection is required to extract features. Second, convolution kernel parameters are shared when the convolutional neural network performs convolution operations, and features at different positions of an image use the same convolution kernel to calculate their values, thereby greatly reducing the model parameters. In this paper, a full convolutional neural network based on a residual dense spatial pyramid is applied in urban remote sensing image segmentation to achieve an accurate semantic segmentation of high-resolution remote sensing images.

Method To improve the semantic segmentation precision of high-resolution urban remote sensing images, we first take a 101-layer residual convolutional network as our backbone in extracting remote sensing image feature maps. When extracting features by using classic convolutional neural networks, the repeated concatenation of max-pooling and striding at consecutive layers significantly reduces the spatial resolution of the feature maps, typically by a factor of 32 across each direction in general deep convolutional neural networks (DCNNs), thereby leading to spatial information loss. Semantic segmentation is a pixel-to-pixel mapping task whose class intensity reaches the pixel level. Reducing the spatial resolution of feature maps can lead to spatial information loss, which is not conducive to the semantic segmentation of remote sensing images. To avoid such loss, the proposed model introduces atrous convolution into the residual convolutional neural network. Compared with ordinary convolution, atrous convolution uses the parameter r to control the receptive field of the convolution kernel during the calculation. The convolutional neural network with atrous convolution can expand the receptive field of the feature map while keeping the feature map size unchanged, thereby significantly improving the remote sensing image semantic segmentation performance of the proposed model. Objects in remote sensing images often demonstrate large-scale variations and complex texture features, both of which challenge the accurate encoding of multi-scale advanced features. To accurately extract multi-scale features in these images, the proposed model cascades each branch of a spatial pyramid structure based on a dense connection mechanism, which allows each branch to output highly dense receptive field information. In these semantic segmentation of remote sensing images, not only the high-level semantic features extracted by the convolutional neural network are required to correctly determine the category of each pixel; low-level texture features are also required to determine the edges of the target. Low-level texture features can benefit the reconstruction of object edges during semantic segmentation. Our proposed model uses a simple encoder to effectively use high-level semantic features and low-level texture features in a network. A decoder also uses skip connection to fuse cross-layer network information and to combine high-level semantic features with the underlying texture features. After fusing high- and low-level information, we use two 3×3 convolutions to integrate the information among channels and to recover spatial information. We eventually input the extracted feature map to a softmax classifier for pixel-level classification and obtain the remote sensing image semantic segmentation results.

Result Full experiments are performed by using the ISPRS (International Society for Photogrammetry and Remote Sensing) remote sensing dataset of the Vaihingen area. We use intersection over union (IoU) and F_1 as our indicators for evaluating the segmentation performance of the proposed model. We also build and train our models based on the NVIDIA Tesla P100 platform and the Tensorflow deep learning framework. The complexity of tasks in the experiment increases at each stage. Experimental results show that the proposed model obtains mean IoU (MIoU) and F_1 values of 69.88% and 81.39% over six types of surface features, respectively, thereby demonstrating vast improvements compared with a residual convolutional network without atrous convolution. Our proposed method also outperforms SegNet, Res-shuffling-Net and SDFCN (symmetrical dense-shortcut fully convolutional network) in terms of mathematics and outperforms pix2pix in terms of visual effects, thereby cementing its validity. We then apply this model on the remote sensing image data of Potsdam area and obtain MIoU and F_1 values of 74.02% and 83.86%, respectively, thereby proving the robustness of our model.

Conclusion We build an end-to-end deep learning model for the semantic segmentation of remote sensing images of high-resolution urban areas. By applying an improved spatial pyramid pooling network based on atrous convolution and dense connections, our proposed model effectively extracts multi-scale features from remote sensing images and fuse high-level semantic information and low-level texture information of the network, which in turn can

improve the accuracy of the model in the remote sensing image segmentation of urban areas. Experimental results prove that the proposed model achieves an excellent performance in terms of mathematical and visual effects and has high application value in the semantic segmentation of high-resolution remote sensing images.

Key words: semantic segmentation; remote sensing images; multiscale; residual convolutional network; dense connection

0 引言

随着遥感技术的快速发展,遥感图像空间、时间以及光谱分辨率大大提高。遥感图像语义分割是根据图像中表达的语义信息对像素进行分组,得到具有逐像素语义注释的分割图像,如图1所示,已广泛应用于环境监测、农业、林业和城市规划等各个领域(冯丽英,2017),是遥感图像应用的重要组成部分。

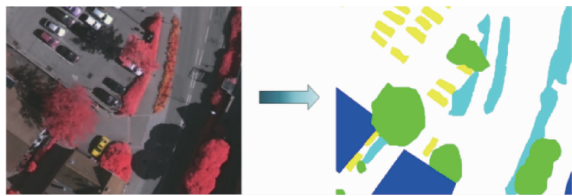


图1 语义分割实例

Fig.1 Examples for semantic segmentation

相关学者对遥感图像语义分割做了大量研究,主要的研究方法分为两种,一种是基于人工特征的传统方法,包括阈值方法,边缘检测方法和区域方法(陈天华等,2018),这种传统方法一方面效率低且不准确,另一方面需要大量的专业知识;另一种是基于卷积神经网络(convolutional neural networks, CNN)的方法。随着CNN在计算机视觉领域的成功应用,相关学者逐渐开始研究其在遥感图像语义分割中的应用。李欣等人(2019)利用带孔卷积,提出了一种基于深度残差网络的多尺度语义分割模型。Audebert等人(2016)研究了全卷积神经网络(fully convolutional network, FCN)在地球观测图像上基于像素场景标记中的应用。Liu等人(2017)使用完整的卷积神经网络作为特征提取器,结合条件随机场对高分辨率遥感影像进行语义分割。Blomley和Weinmann(2017)使用深度完全卷积神经网络处理高分辨率遥感影像的多模态数据,并将其应用于语义分割。Chen等人(2018a)研究了对称法线捷径FCN(symmetrical normal-shortcut FCN, SNFCN)和对称密线捷径FCN(symmetrical dense-shortcut FCN,

SDFCN)框架在超高分辨率遥感图像分割中的应用。Marmanis等人(2016)使用两个孪生网络分别提取遥感图像与数字地表模型(digital surface model, DSM)的特征,应用于高分辨率航空遥感图像的语义分割。Sherrah(2016)提出了一种不进行下采样的卷积神经网络,但计算效率低,且需要大量的GPU资源。

本文主要贡献是将密集连接的空间金字塔池网络应用于高分辨率城市地区遥感影像语义分割。网络融合遥感多尺度语义特征以及低层纹理特征,能够有效地提取不同尺度目标及其边缘,可以更有效地分割纹理复杂的高分辨率遥感图像。本文基于高分辨率遥感图像数据集ISPRS(International Society for Photogrammetry and Remote Sensing)(Rottensteiner等,2012)开展充分实验研究,实验结果表明,本文模型在均交并比和 F_1 值两个指标上优于Segnet、pix2pix、Res-shuffling-Net以及SDFCN算法,证明了算法的有效性。

1 方法原理

1.1 网络结构

本文模型结构包含3部分,如图2所示。1)基于带孔卷积改进的残差网络(图2中深度卷积神经网络(deep convolutional neural network, DCNN))部分用于提取特征;2)基于密集连接改进的空洞空间卷积池化金字塔(atrous spatial pyramid pooling, AS-

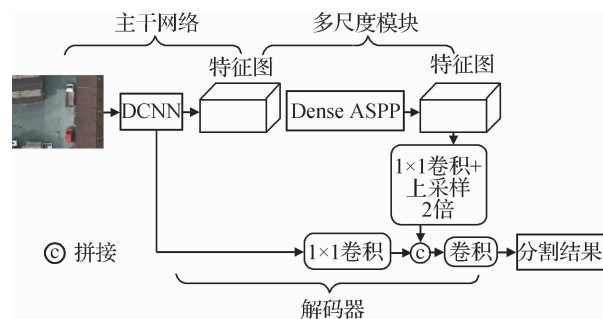


图2 网络结构图

Fig.2 Network architecture diagram

PP)模块,用于提取和融合多尺度特征;3)解码器,采用跳连接融合高低信息进行简单解码,输出语义分割图。

1.2 空洞卷积残差网络

本文使用101层残差网络(ResNet-101)(He等,2016)作为主干网络提取特征。CNN在提取特征过程中存在下采样操作扩大感受野,会造成空间信息的损失,不利于遥感图像语义分割。本文使用带孔卷积替代普通卷积。带孔卷积具体计算为

$$Y[i,j] = \sum_m \sum_n (X[i+r \cdot m, j+r \cdot n] * W[m,n]) \quad (1)$$

式中, X 表示输入特征图, W 表示卷积核,参数 r 为带孔卷积的采样率,用于控制卷积核感受野。由式(1)可知,带孔卷积在扩大特征图感受野的同时能够保持特征图尺寸不变,可有效避免空间信息的损失(Yu和Koltun,2015;Chen等,2017)。

基于带孔卷积的残差卷积网络(residual convolutional network with atrous convolution, RNA)结构如表1所示。根据输出步长大小设置网络中的conv2_x—conv4_x阶段是否采用带孔卷积,详细情况后续实验将具体讨论,conv5_x阶段3个残差瓶颈单元中的 3×3 卷积分别设置为采样率为1、2、4的带孔卷积。

表1 特征提取器网络结构

Table 1 Network structure of feature extractors

层	ResNet-101
conv1	$7 \times 7, 64$ 3×3 , max pool
conv2_x	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
conv4_x	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$
conv5_x	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$

1.3 空洞空间卷积池化金字塔(ASPP)

遥感图像中存在不同尺度的目标,本文首先采

用ASPP(Chen等,2017)结构,使用一组具有不同采样率的带孔卷积并行提取遥感图像中的多尺度特征。ASPP结构如图3所示,具体计算为

$$Y = \text{Concat}(\mathbf{I}_{\text{pooling}}(X), H_{1,3}(X), H_{6,3}(X), H_{12,3}(X), H_{18,3}(X)) \quad (2)$$

式中, $\text{Concat}(\cdot)$ 表示对特征图做第1维度上的拼接操作, $H_{r,n}(X)$ 为采样率为 r 、卷积核大小为 n 的带孔卷积, $\mathbf{I}_{\text{pooling}}$ 指图3中image pooling分支的图像级特征,即输入特征图的平均池化特征。

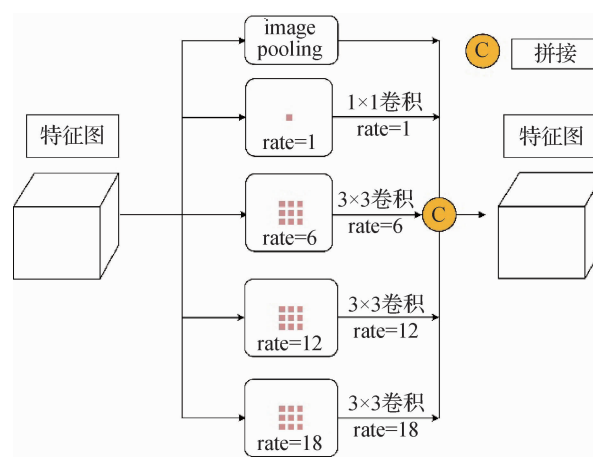


图3 ASPP结构图

Fig. 3 Structure of ASPP

1.4 解码器

由1.2节可知,主干网络用于提取特征。由于提取特征会造成特征图尺寸下降,本文使用简单的解码网络(Chen等,2018c)恢复特征图尺寸。在解码网络中,首先通过线性插值法对特征图做2倍的上采样,然后将得到的特征图与主干网络中对应尺寸的低级特征(取conv2_x中的第2个卷积组中conv3的输出)融合,最后使用 3×3 的卷积整合通道间信息,输出语义分割图。

1.5 密集空洞空间卷积池化金字塔(Dense ASPP)

高分辨率遥感图像中的目标具有纹理特征复杂和尺度变化范围大的特点,ASPP并行提取特征,在一定程度上有利于解决多尺度问题,但其在尺度轴维度上的分辨率不足以精确提取遥感图像中的目标特征。本文采用Yang等人(2018)基于密集连接提出的ASPP(Dense ASPP)代替ASPP, Dense ASPP中各分支以密集连接方式级联,且带孔卷积采样率逐渐增加,每个分支的输入是之前分支输出的拼接,每个分支的输出都有更加密集的感受野信息,具

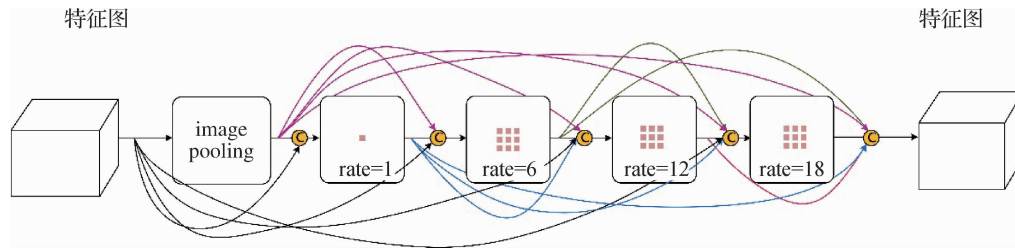


图4 Dense ASPP 结构图

Fig. 4 Structure of Dense ASPP

体结构如图4所示。图4中每一层的输出表达式为

$$Y_l = \begin{cases} I_{\text{pooling}}(X) & l=0 \\ H_{r,n}(\text{Concat}(Y_0, Y_1, \dots, Y_{l-1})) & l \neq 0 \end{cases} \quad (3)$$

式中, Y_l 表示第 l 层的输出。

2 实验

2.1 实验数据集及模型评价指标

本文使用 ISPRS (Rottensteiner 等, 2012) Vaihingen 区域的高分辨率遥感数据进行实验。数据集包含 33 幅超高分辨率遥感图像, 其中 16 幅用

于训练, 17 幅用于测试, 图像的尺寸在 1 281 ~ 3 816 像素之间, 空间分辨率为 0.09 m。数据集真值已知, 划分为 6 种常见的土地覆盖类别, 包括不透水面、建筑物、低矮的植被、树木、汽车和杂波/背景, 数据集示例如图 5 所示。实验中首先对数据进行预处理, 包括数据切割和数据增强: 1) 将数据集中的图像均匀地切割成尺寸为 400 × 500 像素的图像块; 2) 对得到的训练图像进行数据增强处理, 即对训练集中的每幅图像进行上下左右翻转, 旋转 90°、180° 和 270°。最终, 训练集中包含 2 718 幅图像, 测试集中包含 517 幅图像。

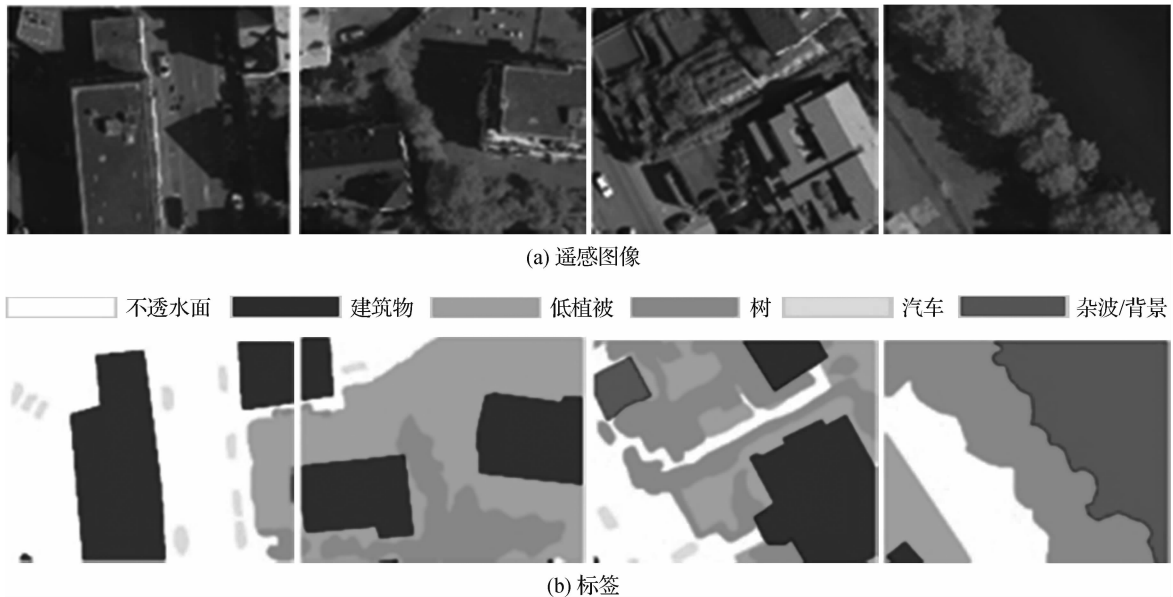


图5 ISPRS 数据集样例

Fig. 5 Samples of ISPRS dataset((a) remote sensing images; (b) labels)

本文使用交并比(intersection over union, IoU)和 F_1 值作为模型评价指标。IoU 的定义是预测和标签中都标记为某一类的像素数与在预测或标签中被标记为该像素数的比值; F_1 值的定义为精确率和召回率的调和平均值。分别计算为

$$IoU = \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (4)$$

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} \quad (5)$$

式中, k 表示类别数, p_{ij} 真实像素类别为 i 的像素被预测为类别 j 的数量, P 和 R 分别表示精确率和召回率。

2.2 实验环境及实验超参数

实验环境为 Windows Server 系统, 深度学习框架为 Tensorflow 1.9.0, 硬件平台为 NVIDIA Tesla P100。定义 S_{ratio} 为输入图像与 ResNet-101 提取特征图大小的比值。学习率策略为多项式学习率策略, 更新公式为 $\eta_0 \times (1 - s/s_{max})^p$, 初始学习率 η_0 为 0.001, s 指当前迭代次数, s_{max} 指最大迭代次数, p 为学习率策略指数, 实验中设置为 0.9; 训练批尺寸为 4, 迭代次数均为 25 000。

2.3 模型参数分析

2.3.1 输出步长设置

为了探究输出步长对模型性能的影响, 以 RNA 为训练模型并设置 4 组实验进行对比分析。在实验过程中, RNA 的输出步长分别设置为 4, 8, 16, 32。设置方式为当特征图尺寸缩小至指定输出步长大小, 残差块 (conv2_x—conv4_x) 使用带孔卷积, 否则与 ResNet-101 相同。例如, 当输出步长为 4 时, conv2_x ~ conv4_x 阶段中所有的 3×3 卷积均设置为采样率为 2 的带孔卷积, 输出步长为 8 时, conv3_x、conv4_x 阶段 3×3 卷积均设置为带孔卷积。当 RNA 的输出步长为 4 时, 由于 GPU 内存的限制, 批尺寸设置为 1, 对应的训练迭代为 100 000。实验得到的平均交并比 (mean intersection over union, MIoU) 结果如表 2 所示。由表 2 可知, RNA 的性能与输出步长呈负相关。当输出步长减小到 4 时, 模型性能没有明显改善, 但计算量大幅增加。为了平衡模型性能和计算量, 在后续实验中将模型的输出步长统一设置为 8。

2.3.2 ASPP 参数设置

由 1.3 节可知, 为了提取遥感图像中的多尺度特征, 本文使用 ASPP 结构。为了探究 ASPP 结构中带孔卷积分支采样率对性能的影响, 本节设计了 3 组对比实验, 将 ASPP 中带孔卷积的采样率分别设置为 $R_{group} = 2, 4, 6, 8$; $R_{group} = 6, 12, 18, 24$ 和 $R_{group} = 6, 12, 18$ 。实验结果如表 3 所示, 由表 3 可知, 当采样率组设置为 (6, 12, 18) 时, 模型分割效果最佳。在后续实验中, ASPP 采样率组设置为 (6, 12, 18), 以确保模型可以表现出最佳性能。

表 2 不同输出步长 RNA 的 MIoU 结果

Table 2 MIoU of RNA with different S_{ratio}

/%	
S_{ratio}	MIoU
4	67.47
8	67.24
16	66.27
32	62.12

表 3 不同比率的 RNA + ASPP 的 MIoU 结果

Table 3 MIoU of RNA + ASPP with different rates

/%	
R_{group}	MIoU
(6, 12, 18, 24)	67.53
(2, 4, 6, 8)	67.32
(6, 12, 18)	67.67

2.3.3 Dense ASPP 和解码器

在本节实验中, 将解码器和 Dense ASPP (Yang 等, 2018) 添加到 RNA 上。在解码器中, 为了平衡解码网络中高级和低级特征的权重, 首先使用通道数为 48 的 1×1 卷积来减少高级功能的通道数, 在融合高低层语义特征时, 使用两个连续的通道数为 256 的 3×3 卷积整合高级语义信息和低级纹理信息。为了验证解码器的作用, 在 2.3.2 节的基础上添加了解码器, 重新训练模型作为对比。2.3 节各阶段训练的模型在 ISPRS 测试集上的详细实验结果如表 4 和表 5 所示。由表 4 和表 5 可知, 本文模型在 F_1 和 IoU 两个指标上均有所提升, 解码器和 Dense ASPP 具有一定效果。

2.4 模型效果分析

本文消融实验结果如表 6 所示, 以 RNA 作为对比基准, 本文模型 MIoU 值提升了 2.64, 证明了模型的有效性。

为了进行横向对比, 本文基于 Vaihingen 数据集训练 SegNet (Badrinarayanan 等, 2017) 和 pix2pix (Isola 等, 2017), 并统计 Res-shuffling-Net (Chen 等, 2018b)、SDFCN (Chen 等, 2018a) 及本文模型结果, 相关数据见表 7。由表 7 可知, 本文模型的 MIoU 优于 SegNet、Res-shuffling-Net 和 SDFCN。表 7 未统计 pix2pix 的结果, 原因是 pix2pix 属于生成对抗网络 (generative adversarial network, GAN)。GAN 的机制

表 4 不同模型在 ISPRS 测试集上的 IoU 对比
Table 4 Comparison of IoU by different models on ISPRS test set

模型	土地覆盖类别						平均值
	不透水面	建筑物	低植被	树	汽车	杂波/背景	
RNA + ASPP	80.04	86.69	66.27	76.18	60.18	36.70	67.67
RNA + ASPP + decoder	80.64	87.62	66.96	76.24	62.16	36.28	68.31
本文	81.13	87.67	66.78	76.25	63.57	43.86	69.88

注:加粗字体表示各列最优结果。

表 5 不同模型在 ISPRS 测试集上的 F₁ 对比
Table 5 Comparison of F₁ by different models on ISPRS test set

模型	土地覆盖类别						平均值
	不透水面	建筑物	低植被	树	汽车	杂波/背景	
RNA + ASPP	88.91	92.87	79.71	86.48	75.14	53.69	79.47
RNA + ASPP + decoder	89.29	93.40	80.21	86.52	76.66	53.24	79.89
本文	89.58	93.42	80.08	86.52	77.73	60.97	81.39

注:加粗字体表示各列最优结果。

表 6 具有不同组件的模型的结果
Table 6 Results of models with different components

RNA	ASPP	Dense ASPP	decoder	MIoU/%
✓				67.24
✓	✓			67.67
✓	✓		✓	68.31
✓		✓	✓	69.88

注:加粗字体表示各行最优结果。

表 7 不同模型的 Vaihingen 测试集结果
Table 7 Results by different models on Vaihingen test set

方法	MIoU
SegNet	44.51
Res-shuffling-Net	60.48
SDFCN	62.38
本文	69.88

注:加粗字体表示各行最优结果。

是使生成器生成尽可能接近真实值的结果,并非原始标签值完全相同,故在数学角度上,精度非常低,但 pix2pix 模型能够生成较好的视觉效果图。在图 6 中列出了 SegNet、pix2pix 和本文模型的输出图

像。由表 7 和图 6 可知,本文模型在数学和视觉方面均表现出较好的性能。

2.5 Potsdam 地区的实验结果

为了验证本文模型的泛化能力,将本文模型应用于 Potsdam 地区的遥感图像。数据集包含 38 幅遥感影像,24 幅用于训练,14 幅用于测试,图像尺寸为 6 000 × 6 000 像素,分辨率为 0.05 m。使用 2.1 节中的方法进行数据预处理,训练集中获得 25 920 幅图像,测试集中获得 2 520 幅图像。结果如表 8 所示。可以看出,在 Potsdam 数据集上,MIoU 达到 74.02%,F₁ 均值达到 83.86%,证明了本文模型对高分辨率城市地区遥感影像语义分割的有效性。

3 结 论

为了提高城市地区遥感图像语义分割精度,本文设计了一种适用于高分辨率城市地区遥感图像语义分割的全卷积神经网络。本文模型主要有以下两个特点:

1) 采用基于带孔卷积改进的残差结构卷积网络作为主干网络提取遥感图像特征,主干网络能够通过设置超参数控制提取的特征图的尺寸,避免了

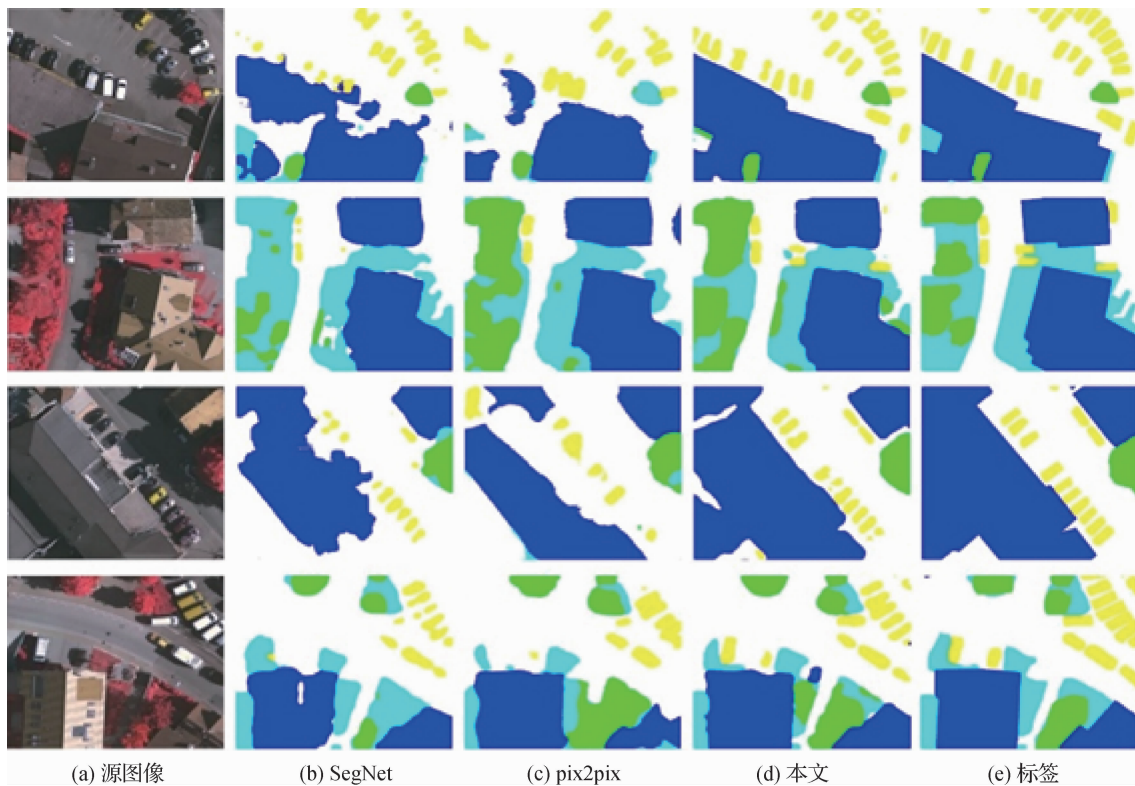


图 6 模型视觉效果图

Fig. 6 Visual effects of the models ((a) source images; (b) SegNet; (c) pix2pix; (d) ours; (e) labels)

表 8 Potsdam 地区的实验结果

Table 8 Experimental results on Potsdam area

评价指标	土地覆盖类别						平均值
	不透水面	建筑物	低植被	树	汽车	背景	
IoU	83.00	90.89	73.40	75.78	81.81	39.23	74.02
F ₁	90.71	95.23	84.66	86.22	89.99	56.35	83.86

提取特征过程中空间信息的损失,能够准确有效地提取遥感图像的高级特征。

2) 引入了基于密集连接机制构建的多尺度特征提取模块,该模块以密集连接为基础,充分利用网络中的多尺度特征,能够有效地分割遥感图像中不同尺度目标。

本文基于 Tensorflow 框架搭建网络模型,在 ISPRS 数据集上分阶段展开实验,以 RNA、RNA + ASPP、RNA + ASPP + decoder、RNA + Dense ASPP + decoder 为模型进行递进研究。同时将本文模型与 SegNet、pix2pix、Res-shuffling-Net 以及 SDFCN 的分割效果进行比较,实验结果表明,本文模型在数学指标和视觉效果方面均优于其他算法,证明了本文模型的有效性。

本文模型的不足之处在于,由于模型中多尺度模块以密集连接为基础,使得模型的计算量大幅提升,模型分割实时性较差。后期研究将进一步对模型进行优化,在不降低模型准确性的前提下,改进网络结构,研究兼具精度和速度的实时城市遥感图像语义分割模型。

参考文献 (References)

- Audebert N, Le Saux B L and Lefèvre S. 2016. Semantic segmentation of earth observation data using multimodal and multi-scale deep networks//Proceedings of the 13th Asian Conference on Computer Vision. Taipei, China; Springer: 180-196 [DOI: 10.1007/978-3-319-54181-5_12]
- Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: a deep con-

- volutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12): 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]
- Blomley R and Weinmann M. 2017. Using multi-scale features for the 3D semantic labeling of airborne laser scanning data//ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Wuhan, China: ISPRS: 43-50 [DOI: 10.5194/isprs-annals-IV-2-W4-43-2017]
- Chen G Z, Zhang X D, Wang Q, Dai F, Gong Y F and Zhu K. 2018a. Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(5): 1633-1644 [DOI: 10.1109/JSTARS.2018.2810320]
- Chen K, Weinmann M, Gao X, Yan M, Hinz S, Jutzi B and Weinmann M. 2018b. Residual shuffling convolutional neural networks for deep semantic image segmentation using multi-modal data//ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Riva del Garda, Italy: ISPRS: 65-72 [DOI: 10.5194/isprs-annals-IV-2-65-2018]
- Chen L C, Papandreou G, Schroff F and Adam H. 2017. Rethinking atrous convolution for semantic image segmentation [EB/OL]. [2019-09-30]. <https://arxiv.org/pdf/1706.05587.pdf>
- Chen L C, Zhu Y K, Papandreou G, Schroff F and Adam H. 2018c. Encoder-decoder with atrous separable convolution for semantic image segmentation [EB/OL]. [2019-09-30]. <https://arxiv.org/pdf/1802.02611v1.pdf>
- Chen T H, Zheng S Q and Yu J C. 2018. Remote sensing image segmentation based on improved DeepLab network. Measurement and Control Technology, 37(11): 34-39 (陈天华, 郑司群, 于峻川. 2018. 采用改进 DeepLab 网络的遥感图像分割. 测控技术, 37(11): 34-39) [DOI: 10.19708/j.ckjs.2018.11.008]
- Feng L Y. 2017. Research on Construction Land Information Extraction from High Resolution Images with Deep Learning Technology. Hangzhou: Zhejiang University (冯丽英. 2017. 基于深度学习技术的高分辨率遥感影像建设用地信息提取研究. 杭州: 浙江大学)
- 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 (CVPR). Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Isola P, Zhu Y J, Zhou T H and Efros A A. 2017. Image-to-image translation with conditional adversarial networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5967-5976 [DOI: 10.1109/CVPR.2017.632]
- Li X, Tang W L and Yang B. 2019. Semantic segmentation of high-resolution remote sensing image based on deep residual network. Journal of Applied Sciences—Electronics and Information Engineering, 37(2): 282-290 (李欣, 唐文莉, 杨博. 2019. 利用深度残差网络的高分遥感影像语义分割. 应用科学学报, 37(2): 282-290) [DOI: 10.3969/j.issn.0255-8297.2019.02.013]
- Liu Y S, Piramanayagam S, Monteiro S T and Saber E. Dense semantic labeling of very-high-resolution aerial imagery and LiDAR with fully-convolutional neural networks and higher-order CRFs//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, USA: IEEE: 1561-1570 [DOI: 10.1109/CVPRW.2017.200]
- Marmanis D, Wegner J D, Galliani S, Schindler K, Datcu M and Stilla U. 2016. Semantic segmentation of aerial images with an ensemble of CNNs//ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Prague, Czech Republic: ISPRS: 473-480 [DOI: 10.5194/isprs-annals-III-3-473-2016]
- Rottensteiner F, Sohn G, Jung J, Gerke M, Baillard C, Bénítez S and Breikopf U. 2012. The ISPRS benchmark on urban object classification and 3D building reconstruction//ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Melbourne, Australia: ISPRS: 293-298 [DOI: 10.5194/isprsannals-I-3-293-2012]
- Sherrah J. 2016. Fully convolutional networks for dense semantic labeling of high-resolution aerial imagery [EB/OL]. [2019-09-30]. <https://arxiv.org/pdf/1606.02585.pdf>
- Yang M K, Yu K, Zhang C, Li Z W and Yang K Y. 2018. DenseASPP for semantic segmentation in street scenes//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3684-3692 [DOI: 10.1109/CVPR.2018.00388]
- Yu F and Koltun V. 2015. Multi-scale context aggregation by dilated convolutions [EB/OL]. [2019-09-30]. <https://arxiv.org/pdf/1511.07122.pdf>

作者简介



韩彬彬, 1994 年生, 男, 硕士研究生, 主要研究方向为深度学习和遥感图像语义分割。
E-mail: hanbinbin17@mails.ucas.ac.cn



张月婷, 通信作者, 女, 副研究员, 主要研究方向为合成孔径雷达图像解释、SAR 图像模拟和散射计算。
E-mail: zhangyueying06@mails.gucas.ac.cn

潘宗序, 男, 副研究员, 主要研究方向为深度学习在遥感中的应用。E-mail: zxpan@mail.ie.ac.cn

台宪青, 男, 研究员, 主要研究方向为大数据与信息安全。

E-mail: taixianqing@ciotc.org

李芳芳, 女, 副研究员, 主要研究方向为干涉 SAR 信号处理、SAR 三维成像。E-mail: fflil@mail.ie.ac.cn