



多尺度特征融合空洞卷积ResNet遥感图像建筑物分割

徐胜军, 欧阳朴衍, 郭学源, Taha Muthar Khan, 段中兴

引用本文:

徐胜军, 欧阳朴衍, 郭学源, 等. 多尺度特征融合空洞卷积ResNet遥感图像建筑物分割[J]. *光学精密工程*, 2020, 28(7): 1588–1599.

XU Sheng-jun, OUYANG Pu-yan, GUO Xue-yuan, et al. Building segmentation in remote sensing image based on multiscale–feature fusion dilated convolution resnet[J]. *Optics and Precision Engineering*, 2020, 28(7): 1588–1599.

在线阅读 View online: <https://doi.org/10.37188/OPE.20202807.1588>

您可能感兴趣的其他文章

Articles you may be interested in

融合高阶信息的遥感影像建筑物自动提取

High–order statistics integration method for automatic building extraction of remote sensing images

光学精密工程. 2019, 27(11): 2474–2483 <https://doi.org/10.3788/OPE.20192711.2474>

基于多尺度特征融合的遥感图像场景分类

Scene classification of remote sensing images based on multiscale features fusion

光学精密工程. 2018, 26(12): 3099–3107 <https://doi.org/10.3788/OPE.20182612.3099>

基于卷积神经网络的光学遥感图像检索

Optical remote sensing image retrieval based on convolutional neural networks

光学精密工程. 2018, 26(1): 200–207 <https://doi.org/10.3788/OPE.20182601.0200>

基于深度语义分割的多源遥感图像海面溢油监测

Research on oil spill monitoring of multi–source remote sensing image based on deep semantic segmentation

光学精密工程. 2020, 28(5): 1165–1176 <https://doi.org/10.3788/OPE.20202805.1165>

循环神经网络多标签航空图像分类

Recurrent neural network multi–label aerial images classification

光学精密工程. 2020, 28(6): 1404–1413 <https://doi.org/10.3788/OPE.20202806.1404>

文章编号 1004-924X(2020)07-1588-12

多尺度特征融合空洞卷积 ResNet 遥感图像建筑物分割

徐胜军¹, 欧阳朴衍^{1*}, 郭学源², Taha Muthar Khan¹, 段中兴¹

(1. 西安建筑科技大学 信息与控制工程学院, 陕西 西安 710055;

2. 清华大学建筑设计院有限公司, 北京 100084)

摘要: 针对传统建筑物提取方法难以有效描述遥感图像细节特征, 导致复杂场景下道路、树木及建筑物之间分割边界不清晰等问题, 提出了一种基于多尺度特征融合空洞卷积 ResNet(Multiscale-feature Fusion Dilated Convolution ResNet, MFDC-ResNet)模型。首先, 为了获取遥感图像建筑物更大范围的特征信息, 在深度残差网络中引入空洞卷积增大特征提取的感受野, 以捕捉更丰富的多尺度细节特征; 其次, 为了增强空洞卷积中心点对图像局部区域特征的表达, 利用 3×3 卷积核提取遥感图像的中心点区域特征, 引入更多的中心点空间先验信息; 最后, 利用空间金字塔池化模型对不同尺度空洞卷积特征进行融合, 获取不同尺度的遥感图像建筑物的上下文信息。在 WHU 遥感图像数据集上的实验表明, 平均交并比 mIoU 达到 0.820, 召回率 Recall 达到 0.882。提出算法不仅提高了分割精度, 而且有效克服了道路、树木等因素的干扰, 得到了较清晰的建筑物边界。

关键词: 遥感图像; 建筑物分割; 残差网络; 空洞卷积; 多尺度特征融合

中图分类号: P236 文献标识码: A doi:10.37188/OPE.20202807.1588

Building segmentation in remote sensing image based on multiscale-feature fusion dilated convolution resnet

XU Sheng-jun¹, OUYANG Pu-yan^{1*}, GUO Xue-yuan², Taha Muthar Khan¹, DUAN Zhong-xing¹

(1. College of Information and Control Engineering,

Xi'an University of Architecture and Technology, Xi'an 710055, China;

2. Architectural Design and Research Institute of Tsinghua University Company,

Limited, Beijing 100084, China)

* Corresponding author, E-mail: jokerman1993@foxmail.com

Abstract: To solve the problem in which a traditional ResNet101 model cannot effectively describe the detailed features of remote sensing images, leading to the unclear segmentation boundary between roads, trees, and buildings in complex scenes, a Multiscale-feature Fusion Dilated Convolution ResNet (MFDC-ResNet) was proposed. First, to obtain large-scale building feature information of remote sensing images, a dilated convolution was introduced in the deep residual network to capture

收稿日期: 2020-01-02; 修订日期: 2020-03-02.

基金项目: 国家自然科学基金资助项目 (No. 51678470, No. 61803293); 陕西省教育厅专项科研项目资助 (No. 18JK0477, No. 2017JM6106); 陕西省自然科学基金基础研究计划资助项目 (No. 2020JM-472, No. 2020JM-473, No. 2019JQ-760); 西安建筑科技大学基础研究基金资助项目 (No. JC1703, No. JC1706)

richer multi-scale details. Second, to enhance the expression ability of the center point of dilated convolution on the building of feature images, a 3×3 convolution kernel was proposed to extract features in the local area of remote sensing images. Finally, a spatial pyramid pool model of multi-scale feature fusion was proposed to fuse the multi-scale features, obtain the building contextual information of different scales of remote sensing images, and complete the accurate segmentation of buildings. The results of the experiments show that the mean Intersection over Union (mIoU) of building segmentation in WHU is 0.820 and the recall rate is 0.882. The developed method can effectively overcome the influence of roads, trees, and other factors. Moreover, the building boundary can be extracted clearly and smoothly from the remote sensing images and the segmentation accuracy is improved.

Key words: remote sensing image; building segmentation; resnet; dilated convolution; multi-scale feature fusion

1 引 言

建筑物作为城市区域的重要特征,是占主体地位的地物目标,遥感图像建筑物的分割在重要目标的识别及定位、国防安全和掌握建筑物实时位置等方面都有着重要研究意义和应用价值。但是由于不同时相遥感图像受角度、光照、季节等因素的影响,同时也会受到建筑物周围背景,如道路、树木及其他地物阴影等的干扰,使得遥感图像建筑物的分割成为高分辨率遥感图像分割中最具挑战性的任务之一。

在遥感图像建筑物语义分割中,传统方法通过人工或机器学习方法^[1-4]获取梯度、灰度值、纹理信息等图像特征实现对特定目标的分割,主要包括基于阈值\基于边缘和基于区域的分割方法。ANDRES 等^[1]提出一种基于 K-means 算法的阈值确定方法对遥感图像进行分割,利用 K-means 分类来指导遥感图像的分割过程。Wang 等^[2]在遥感图像阈值分割的基础上,融合边缘提取的方法得到遥感图像分割物的边缘信息,最终对目标区域实现分割。Deng 等^[3]引入区域合并的方法来解决遥感图像过分割问题,运用基于区域面积加权的区域相似度准则解决区域合并中过分割的问题实现对遥感图像的分割。Zhao 等^[4]提出了一种基于灰度共生矩阵丰富边缘区域提取的图像去模糊方法,将区域特征信息与图像边缘信息结合,从模糊图像中提取高频信从而提升图像分割精度。这些传统方法在提取遥感图像建筑物特征

信息时受到空间相关性弱、多阈值选择限制和边缘信息不连续等因素影响较大,分割结果欠佳^[5]。

近年来,利用深度学习^[6-20]实现遥感图像建筑物语义分割得到了广泛关注。相比传统方法,基于深度学习的方法可以针对具体的遥感图像建筑物分割问题自动学习合适的特征表示,其分割的实质是建立图像到分割目标之间的复杂映射模型。Krizhevsky 等^[6]提出 AlexNet 模型,利用深度神经网络对图像进行分类。Simonyan 等^[7]基于 VGG 模型,通过反复堆叠 3×3 的卷积核和 2×2 最大池化层,不断加深网络结构来提升性能。但是随着网络的加深,出现训练集准确率下降,错误率上升的现象。He 等^[8]提出深度残差网络 ResNet(Residual Neural Network, ResNet),解决了网络模型的退化问题,利用更深层的网络提取更复杂的特征。在卷积神经网络中,很多经典网络被应用于遥感图像的语义分割。YE 等^[9]使用一种基于卷积神经网络的融合算法网络^[10]对传统融合方法中人工制定融合规则的缺点加以改进,利用网络模型将出入图像输出具有端到端特性的融合图像,通过训练数据自适应学习强鲁棒融合函数从而加强分割效果。ZHU 等^[11]利用 SegNet^[12]与图像形态学结合来消除遥感图像噪声,减少误判从而提升遥感图像的分割精度。Yang 等^[13]在 SegNet 基础上利用解码器模块将低分辨率映射特征传递到高分辨率映射特征,对高维特征进行低维提取,达到对遥感图像提取和分割的目的。Piramanayagam^[14]等采用完全卷积神经网络(Fully Convolutional Networks, FCN)

框架对多传感器图像进行像素级分类,将提取的特征经过全卷积后进行融合实现遥感图像的分类。Saito 等^[15]结合卷积层与全连接层组合成的深度神经网络模型实现遥感图像建筑物与道路的分割。Bittner 等^[16]利用 VGG 深度神经网络训练建筑物边缘距离特征信息,建立能量损失函数模型来分割遥感图像建筑物。Jiao 等^[17]提出一种基于全连接条件随机场深度多尺度模型来实现高光谱遥感图像的分割。Xu 等^[18]提出一种全卷积网络模型 FCN,去掉下采样层,将多元信息在输入端进行融合来实现遥感图像小目标的分割。但是上述方法在遥感图像的建筑物分割中,提取的固定大小的建筑物特征信息对一些较小尺寸的建筑物经常无法实现精细化分割,甚至出现漏检的问题。另外遥感图像建筑物的分割常受到复杂场景中道路、树木及建筑物之间距离等因素的影响,建筑物的分割边界容易与其他物体边界融合使得难以提取并精确分割。

针对上述问题,提出了一种基于多尺度特征融合的空洞卷积 ResNet (Multiscale-feature fusion Dilated Convolution ResNet, MFDC-ResNet) 模型,提出模型选用 ResNet 模型作为基础网络,引入空洞卷积方法提取更大范围的建筑物特征,然后利用 4 邻域加权特征算子增强空洞卷积中心点对特征的提取能力,最终将不同尺度的多个遥感图像建筑物特征融合来提升及分割精度。

2 Resnet 深度残差网络

2.1 ResNet 基本结构

令 $\mathbf{X} = \{x_1, x_2, \dots, x_N\}$ 表示遥感图像,其中 x_i 表示任意像素, N 为像素总数; $\mathbf{Z} = \{z_1, z_2, \dots, z_N\}$ 表示遥感图像建筑物分割结果。基于深度学习对遥感建筑物图像分割的实质是建立遥感图像与分割结果之间的映射关系 $f(\cdot)$:

$$\mathbf{Z} = f(\mathbf{X}, \omega), \quad (1)$$

式中 ω 为神经网络中的权重。

遥感图像具有高维性、强背景干扰以及建筑物复杂多变等特征,浅层网络难以有效建立起遥感图像与分割结果之间的映射关系 $f(\cdot)$ 。深度残差神经网络 ResNet 在深度神经网络中增加残差网络,通过学习多个网络层输入、输出之间的残

差,简化学习目标同时保护信息的完整性,解决了网络层数增多时由于误差增高、梯度消失而导致训练结果变差的现象。ResNet101 基础模型分别用 $1 \times 1, 3 \times 3, 1 \times 1$ 的卷积堆叠,其网络整体模型如图 1 所示。

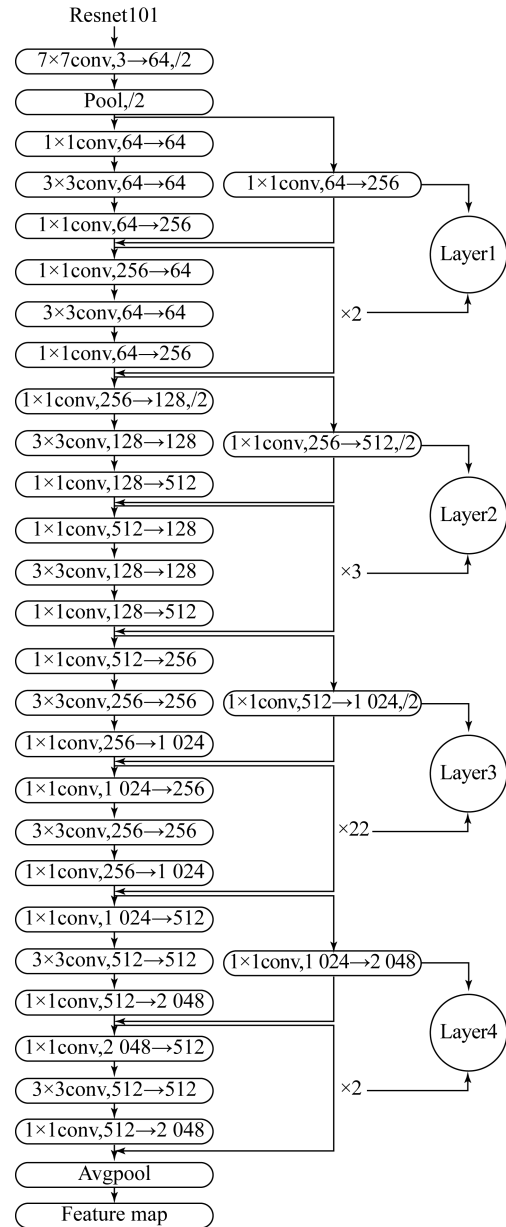


图 1 ResNet101 网络结构

Fig. 1 ResNet101 network structure

2.2 ResNet 基础残差模块

ResNet 基础残差模块主要有 2 种,图 2(a)是常规模块,由两个 64×64 通道的 3×3 卷积组成,输入和输出均为 64 通道。图 2(b)是瓶颈模块,通过一个 1×1 卷积将 256 维通道降维到 64 维通

道,最后再通过一个 1×1 卷积恢复到 256 维通道。残差模块输入输出映射关系 $f_c(\cdot)$ 为:

$$f_c(\mathbf{X}, \omega) = H(\mathbf{X}) - \mathbf{X}, \quad (2)$$

式中: $H(\mathbf{X})$ 表示网络输入 \mathbf{X} 的期望输出; $H(\mathbf{X}) - \mathbf{X}$ 表示网络输出和输入的残差。

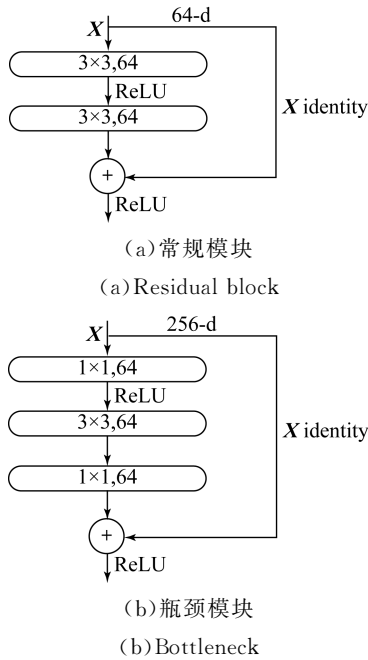


图 2 基础残差模块
Fig. 2 Basic residual module

由残差模块框架可知,在训练模型输出端引入样本的前馈通道构成闭环,把网络的恒等映射函数 $H(\mathbf{X}) = \mathbf{X}$ 训练转换为一个残差函数 $f_c(\mathbf{X}, \omega) = H(\mathbf{X}) - \mathbf{X}$ 的训练,从而使得深度神经网络的权值收敛更加有效。

3 多尺度特征融合的空洞卷积 Resnet 模型

由于遥感图像中建筑物尺寸大小不一,利用固定大小的建筑物特征常无法对一些较小尺寸的建筑物实现精细化分割,甚至出现漏检的问题,另外遥感图像建筑物的分割常受到复杂场景中道路、树木及建筑物之间距离等因素的影响,导致建筑物与其他物体边界相互混淆。针对此问题,本文提出了一种基于多尺度特征融合的空洞卷积 Resnet 模型 (MFDC-ResNet)。提出的 MFDC-ResNet 模型由空洞卷积特征提取、多尺度特征融

合和特征解码 3 个模块组成。基础网络为 ResNet101,分别运用不同扩张率的空洞卷积自动提取遥感图像中建筑物 \mathbf{X} 的特征信息,在不损失所提取图像特征信息的基础上增加特征提取的感受野,获取比 ResNet 基础网络更大范围的建筑物特征信息;其次将提取的不同层次特征信息 \mathbf{Y} 送入多尺度特征提取模块输出,将多层次的特征信息 $\mathbf{Y}^{(m)}$ 和对应的多个上下文信息 $\mathbf{C}^{(m)}$ 由深到浅从多个层次融合并逐层输出特征 $\mathbf{A}^{(m)}$,融合时保持输出步幅下采样的倍数不变;最终由 ResNet101 模块的特征解码恢复图像原始分辨率,输出遥感图像建筑物分割结果 \mathbf{Z} 。提出 MFDC-ResNet 模型框架如图 3 所示。

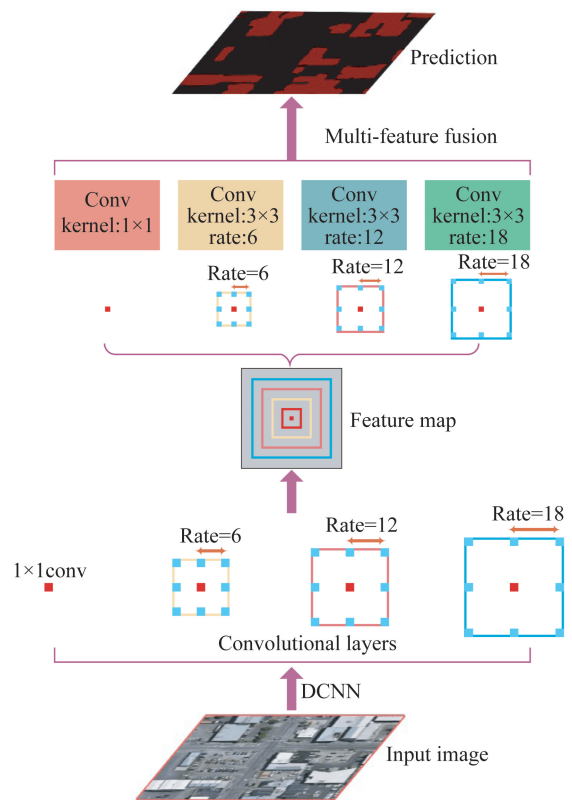


图 3 MFDC-ResNet 模型框架
Fig. 3 MFDC-ResNet model frame

3.1 改进的 ResNet 基础模块

为了避免卷积输出可能存在的负值对 ReLU 激活层输出的弱化,增强深层网络训练过程中的图像特征信息,在 ResNet101 基础残差模块的基础上,将 BN (Batch Normalization) 层和 ReLU (Rectified Linear Unit) 激活层放置于卷积层之前,改进后 ResNet 基本模块结构如图 4 所示。

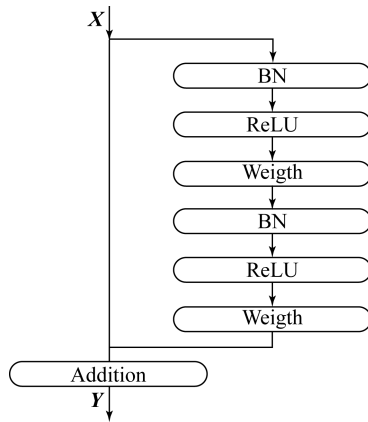


图 4 改进的 ResNet 基本模块

Fig. 4 Improved ResNet basic module

BN 层通过规范化的方式,使激活函数输入在可以激活的区域,输入的变化可以反应损失函数的变化,产生较大梯度以避免梯度消失,从而加快训练中网络收敛速度和学习速率。BN 层的实现过程如算法 1 所示:

算法 1 批量规范化(BN)

Step1: 获取容量为 m 的当前批次训练样本: $\mathbf{X} = \{x_1, x_2, \dots, x_m\}$;

Step2: 计算均值 $\mu_{\delta} : \mu_{\delta} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i$;

Step3: 计算方差 $\sigma_{\delta}^2 : \sigma_{\delta}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\delta})^2$;

Step4: 样本规范化: $\hat{x} \leftarrow \frac{x_i - \mu_{\delta}}{\sqrt{\sigma_{\delta}^2 + \epsilon}}$;

Step5: 尺度变换规范化输出: $y_i \leftarrow \gamma \hat{x}_i + \beta \equiv BN_{\gamma, \beta}(x_i)$, 其中 \hat{x}_i 为标注化参数, γ 和 β 为需要学习的参数, ϵ 为小批量数据的值。

ReLU 激活函数的作用是拟合训练数据,单侧抑制的特点可以使得 ResNet 基本模块具有稀疏激活的作用。ReLU 激活函数定义如式(3):

$$f(x) = \max(0, x), \quad (3)$$

式中 $\max(\cdot)$ 表示在输入为负值时输出为 0,神经元不会被激活。

3.2 空洞卷积结构

在将遥感图像输入深度学习过程中,对图像进行连续卷积和池化或者其他下采样操作整合多尺度上下文信息,这样在降低图像尺寸的同

时增大感受野,会损失一些分辨率,使得一些图像细节信息无法重建。针对以上问题,在 ResNet 中利用空洞卷积在不做池化操作损失特征信息的情况下增大感受野,每个卷积输出包含比普通卷积更大范围的特征信息,利于获取遥感图像中建筑物特征的全局信息。

空洞卷积向卷积层引入了一个“扩张率(Dilation Rate, DR)”的新参数,该参数定义了卷积核处理数据时各值的间距。设置不同扩张率时,感受野会不同,从而能够获取遥感图像多尺度特征信息。实验中,根据遥感图像的原始尺寸来相应设定深度神经网络中空洞卷积所提取的特征图的大小,进而设定空洞卷积扩张率。本文设定扩张率参数分别为 $\text{Rate} = 6$, $\text{Rate} = 12$, $\text{Rate} = 18$,通过不同扩张率的卷积获取多个尺度的建筑物特征信息,卷积内剩余点的权重均为 0。空洞卷积结构如图 5 所示。

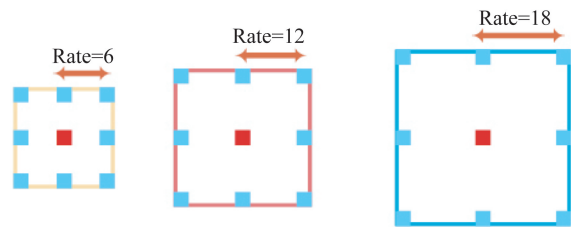


图 5 空洞卷积结构

Fig. 5 Dilated convolution structure

在遥感图像中,建筑物往往随机分布在图像任意位置,且其形状变化大,利用空洞卷积提取特征时扩张率逐渐增大,而卷积中心点的特征权重会逐渐减弱。因此在利用空洞卷积提取建筑物特征时,在增强感受野的同时需要增强卷积核中心点的对图像特征信息的描述能力。为解决这一问题,本文提出一种融合局部空间特征信息卷积核中心点提取方法来加强扩张率 $\text{Rate} = 18$ 时空洞卷积中心点对特征信息的描述能力。提出方法是在提取特征时将卷积核中心点及其 4 邻域点作为有效点一同提取特征。对于一个 3×3 卷积核,输出特征时将卷积核 4 个顶角的特征参数用 0 取代,将卷积输出作为中心点的特征信息。提出的卷积核中心点提取方法如图 6 所示。

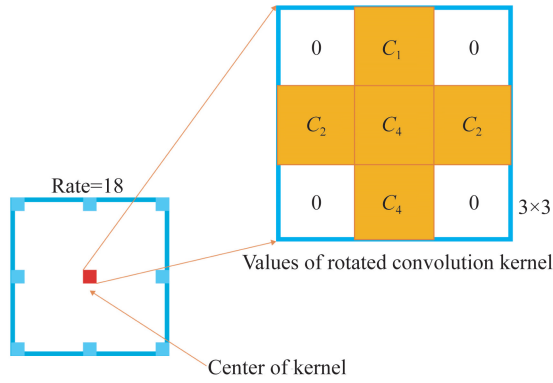


图 6 卷积核中心点改进后提取方式

Fig. 6 Improved extraction method of convolution kernel center point

在中心点 C_0 及其 4 邻域点 $C_n (n=1, 2, 3, 4)$ 构成的 3×3 卷积中,利用局部区域空间特征信息提取新的中心点权重特征 Y 。中心点及邻域的卷积过程如图 7 所示。

其中输入图像向量为 $X = \{x_1, x_2, \dots, x_N\}$, 输出特征 $Y = \{y_1, y_2, \dots, y_9\}$, 卷积过程表示为:

$$X * C = Y, \quad (4)$$

其中卷积算子 $C = \begin{bmatrix} 0 & \omega_2 & 0 \\ \omega_3 & \omega_1 & \omega_4 \\ 0 & \omega_5 & 0 \end{bmatrix}$ 。

$$X * \begin{bmatrix} 0 & \omega_2 & 0 \\ \omega_3 & \omega_1 & \omega_4 \\ 0 & \omega_5 & 0 \end{bmatrix} = Y$$

图 7 中心点及邻域卷积过程

Fig. 7 Center point and its neighborhood convolution process

3.3 多尺度特征融合结构

利用深度神经网络分割建筑物时,池化和其下采样会导致建筑物边缘信息丢失现象。空间金字塔池化对不同分辨率的特征图进行池化,而不同尺度的上下文信息有助于获得完整的建筑物边界信息。因此基于空间金字塔池化在神经网络模型中建立多尺度特征融合结构,去掉传统神经网络中的池化或者其他下采样步骤,在不丢失分辨率的条件下更容易对建筑物目标定位。

表 1 多尺度特征融合输出

Tab. 1 Multiscale-feature fusion output

	卷积核	空洞卷积扩张率	多尺度融合输出	输出尺寸
Conv1	$7 \times 7, 64$ stride2 $3 \times 3, \max \text{ pool}, \text{ stride}2$	—	—	1/2
Conv2	$1 \times 1, 64$ $3 \times 3, 64 \times 3$ $1 \times 1, 256$	—	$A^{(2)}$	1/4
Conv3	$1 \times 1, 64$ $3 \times 3, 64 \times 4$ $1 \times 1, 256$	—	$A^{(3)}$	1/8
Conv4	$1 \times 1, 64$ $3 \times 3, 64 \times 23$ $1 \times 1, 256$	6, 12, 18	$A^{(4)}$	1/8
Conv5	$1 \times 1, 64$ $3 \times 3, 64 \times 3$ $1 \times 1, 256$	6, 12, 18	$A^{(5)}$	1/8

提出的多尺度特征融合结构分别采用 3 种不同采样率的 3×3 卷积核获取多尺度特征信息。但是当采样率接近映射特征时, 3×3 的卷积核不能有效

捕捉局部细节信息,因而采用 1×1 的卷积滤波器,利用滤波器中心点来提取遥感图像建筑物边缘较小的细节信息。多尺度特征融合结构如图 8 所示。

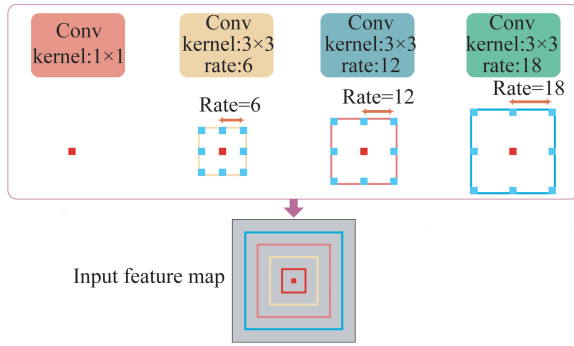


图 8 多尺度特征融合结构

Fig. 8 Multiscale feature fusion structure

空间金字塔结构包含 1×1 和采样率分别为 6, 12, 18 的空洞卷积, 输出通道的数量为 256, 最后包含一个 BN 层。在遥感图像建筑物的多尺度信息融合时, 采用不同采样率的空洞卷积保持输出步幅下采样的倍数为 8, 将空洞卷积提取的多层次特征信息经过空间金字塔模块处理的信息进行融合。提取特征时第 m 层的输出为:

$$\mathbf{Y}^{(m)} = \mathbf{W}^{(m)} \times \mathbf{X}^{(m)} + \mathbf{b}^{(m)}, \quad (5)$$

式中: 第 m 层的卷积输入为 $\mathbf{X}^{(m)}$, 输出为 $\mathbf{Y}^{(m)}$, 权重参数为 $\mathbf{b}^{(m)}$ 。空洞卷积输出的各级特征 $\mathbf{Y}^{(m)}$ 和对应的上下文特征 $\mathbf{C}^{(m)}$ 经空间金字塔模块由深到浅从多个层次融合各级特征 $\mathbf{A}^{(m)}$, 最终运用 ResNet101 中特征解码模块恢复图像原有分辨率, 输出分割结果 \mathbf{Z} 。多尺度特征融合的输出如表 1 所示。

4 实验结果及分析

4.1 实验数据集及实验平台参数设置

实验平台搭载 Inter Xeon E5 2650 处理器, 32 G 内存, 2 个 NVIDIA TitanV Volta 12 G 显卡 (GPU 处理单元); 深度学习框架采用 pytorch-1.0.2, 以及 Nvidia 公司 CUDA8.0 的 GPU 运算平台以及 cuDNN6.0 深度学习 GPU 加速库。实验数据集采用 WHU Building change detection dataset 数据集。数据集样本来自新西兰土地信息服务网站, 数据集有约 22 000 个独立建筑。图像的原始分辨率为 0.075 m。为了便于训练, 将大部分的建筑物航空影像地面分辨率减低至 0.3 m, 并将原始图片裁剪成像素大小为 512×512 的 8 189 张遥感图像建筑物图片, 其中 4 736

张作为训练集, 1 036 张作为验证集, 2 416 张作为测试集。

为了验证提出 MFDC-ResNet 模型的有效性, 在实验中将 VGG^[7], ResNet^[8], ResNetCRF^[21] 与提出模型进行定性与定量对比评价。对比模型 VGG, ResNet, ResNetCRF 和 MFDC-ResNet 等网络模型的初始化权值均来自在 ImageNet 数据集上的预训练模型。其他参数初始值为 0。MFDC-ResNet 模型通过随机梯度下降法对模型训练, 基准学习率为 0.002 0, 总迭代次数为 250 000, 每迭代 5 000 次学习率减少 0.1 倍, 训练过程中每次训练包括 1 次前向传播和 1 次反向传播, 前向传播过程预测结果, 反向传播过程更新权重参数。

4.2 分割性能指标

为了量化分析分割结果, 采用平均交并比 (mean Intersection over Union, mIoU) 和召回率 (Recall Rate, Recall) 对结果进行评价分割。本实验中在建筑物类别上计算 mIoU, 可以准确地评价建筑物的分割精度, 其计算方法如式 (6) 表示:

$$mIoU = \frac{1}{k} \times \sum_i \frac{n_{ii}}{t_i + p_i - n_{ii}}, \quad (6)$$

式中: k 表示分割类别数目, n_{ii} 表示实际类别为 i 而预测结果为 j 的像素点数, t_i 表示第 i 类的像素点总数, p_i 表示第 i 类预测结果的总像素数。

将 Recall 作为评价指标可以评价分割为建筑物的像素点与真实标注的建筑物像素点的比率, 表征在不考虑遥感图像背景的前提下建筑物的分割精度, 其计算方法如下表示:

$$Recall = \frac{B_{seg}}{B_{seg} + I_{unseg}}, \quad (7)$$

式中: B_{seg} 表示遥感图像分割中建筑物与标注结果相比分割正确的像素点数, I_{unseg} 表示遥感图像分割中建筑物与标注结果相比未正确分割的像素点数。

为了对比提出模型的有效性, 分别采用基于 VGG^[6], ResNet^[7], ResNetCRF^[21] 等方法对遥感图像建筑物进行分割实验。不同模型 mIoU 训练过程和 Recall 训对比分别如图 9 和图 10 (下图横坐标为迭代次数, 单位: 千次; 纵坐标分别为评价指标 mIoU 和 Recall 精度)。由 4 种对比模型训练过程可以看出, 提出 MFDC-ResNet 模型的平均交并比 (mIoU) 和召回率 (Recall) 曲线均高于

VGG, ResNet, ResNetCRF 的 mIoU, Recall 曲线。提出模型的 mIoU, Recall 随着训练数据集的增加精度逐渐增高,分别收敛在 0.820 和 0.882 附近,不同模型分割性能评价指标对比如

表 2 所示。由表 2 可知,和 VGG, ResNet 及 ResNetCRF 3 种模型相比,提出的 MFDC-ResNet 模型的 mIoU、Recall 性能评价指标有明显提升,可以有效提升遥感图像建筑物的分割精度。

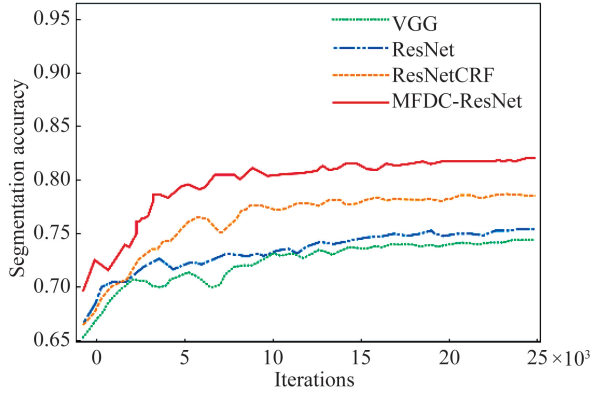


图 9 mIoU 训练过程
Fig. 9 mIoU training process

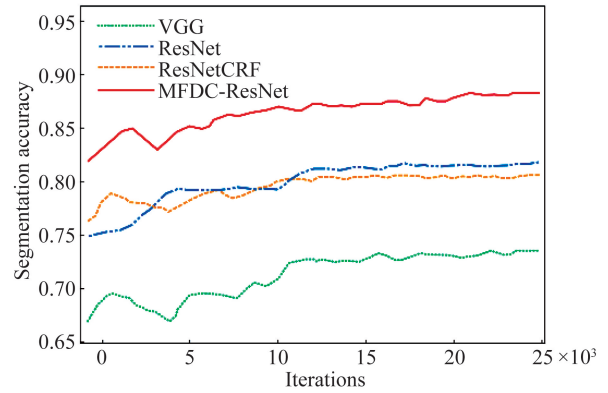


图 10 Recall 训练过程
Fig. 10 Recall rate training process

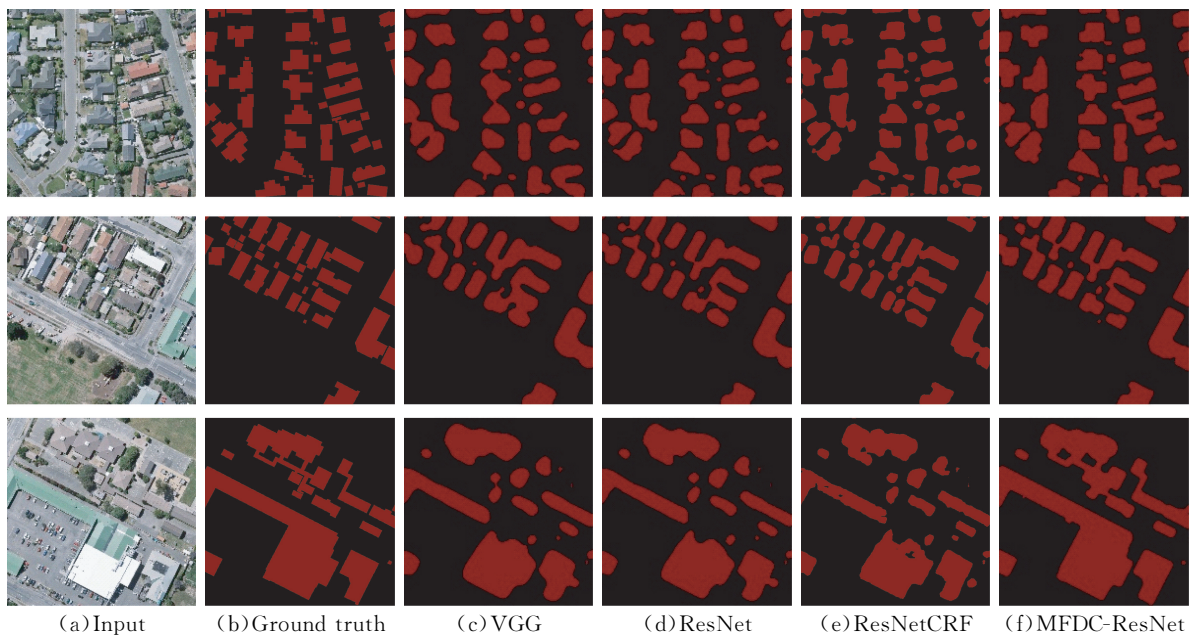
表 2 分割性能评价指标

Tab. 2 Segmentation performance evaluation index

模型	mIoU	Recall
VGG	0.743	0.736
ResNet	0.754	0.819
ResNetCRF	0.786	0.806
MFDC-ResNet	0.82	0.882

4.3 不同模型分割结果分析

不同模型的分割结果对比如图 11 所示,其中 (a)列为分辨率 512×512 的遥感图像,(b)列为建筑物标记真实值图像,(c)列为 VGG 模型分割结果,(d)列为 ResNet 模型分割结果。(e)列为 ResNetCRF 模型分割结果,(f)列为 MFDC-ResNet 模型分割结果。



(a)Input (b)Ground truth (c)VGG (d)ResNet (e)ResNetCRF (f)MFDC-ResNet

图 11 不同方法分割结果对比图

Fig. 11 Comparison of segmentation results by different methods

由实验结果可以发现,基于 VGG 模型的分割方法存在明显误分割现象,特别是对于相邻的建筑物出现许多建筑物粘连的情况,基于 ResNet 模型的分割结果与 VGG 模型相比有了一定提升,粘连现象不明显,但是分割结果仍然欠佳;基于 ResNetCRF 模型,由于 CRF 层有效利用了多特征和上下文信息提升了建筑物分割结果,但在建筑物边缘信息上分割准确性不高;提出的 MFDC-ResNet 与其它几种对比模型相比建筑物边缘清晰平滑,明显减少对树木道路等干扰物误分割的情况,对于不规则建筑物也有较好的分割效果,有效减少因建筑物距离较近而导致分割边界模糊的情况,分割结果明显优于 VGG, ResNet, ResNetCRF 模型。

4.4 干扰因素对比实验结果分析

对比实验所用的遥感图像数据集中包含了不同尺寸大小的建筑物和复杂场景下含有道路、树木等干扰的建筑物以及距离相距较近的建筑物等,为了验证提出 MFDC-ResNet 模型对不同干扰因素的抗干扰能力,分别对不同场景下遥感图像进行分割实验,并对实验结果进行对比分析。

4.4.1 复杂道路条件下分割结果分析

图 12 为包含复杂道路信息的遥感图像分割结果对比。由 4 种深度神经网络模型分割结果可知,4 种模型均可以实现建筑物的分割,其中 VGG 模型受道路影响较大,存在一些道路误分割情况,ResNet 可以确定建筑物的大致位置,但建筑物边缘信息不够准确,在遥感图像中对较窄道路有误分割情况,与 ResNet 相比,ResNetCRF 对建筑物边界提取效果有所提升,但仍存在对道路误分割情况;提出 MFDC-ResNet 模型受到遥感图像中道路的干扰较小,能够较好提取建筑物的完整边缘信息。由表 3 定量分析对比可知,提出 MFDC-ResNet 模型分割精度高于其他 3 种模型。

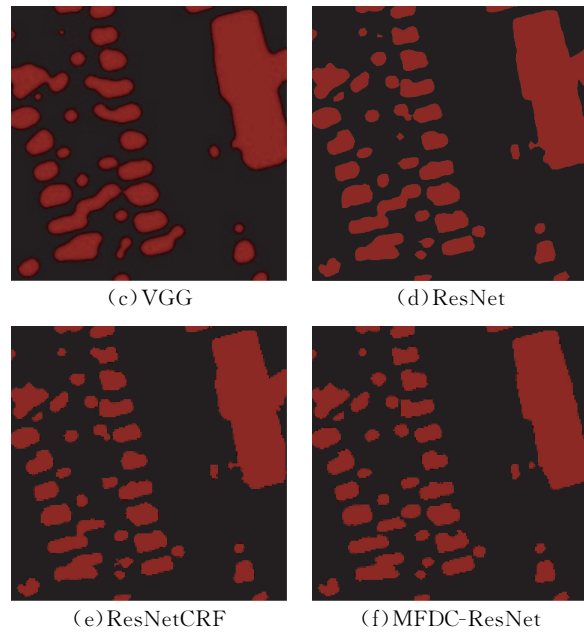
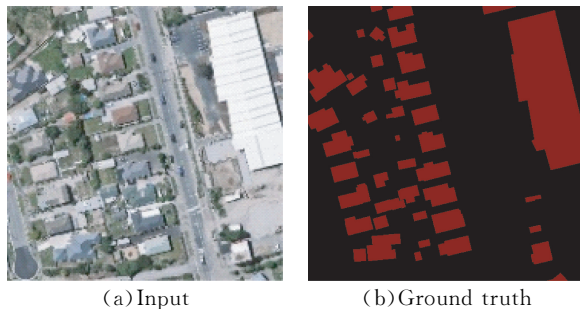


图 12 复杂道路条件下建筑物分割

Fig. 12 Building segmentation under complex road conditions

表 3 复杂道路条件下分割性能对比

Tab. 3 Comparison of segmentation performance under complex road conditions

模型	mIoU	Recall
VGG	0.738	0.738
ResNet	0.757	0.821
ResNetCRF	0.788	0.809
MFDC-ResNet	0.823	0.885

4.4.2 建筑物距离较近条件下分割结果分析

图 13 为包含着多个距离较近建筑物的遥感图像分割对比。由分割结果可以看出,VGG 模型受建筑物之间距离相近的干扰较大,分割结果中许多距离较近的建筑物未能得到清晰分割,ResNet 和 ResNetCRF 分割结果有一定提升,但是仍存在多处相邻建筑物相粘连的情况,无法从中完整提取单个的建筑物边界信息。提出 MFDC-ResNet 模型分割精度高,可以有效克服建筑物之间距离相近的干扰,实现对建筑物边缘的完整提取。由表 4 定量分析对比可知,提出 MFDC-ResNet 模型分割精度高于其他 3 种模型。

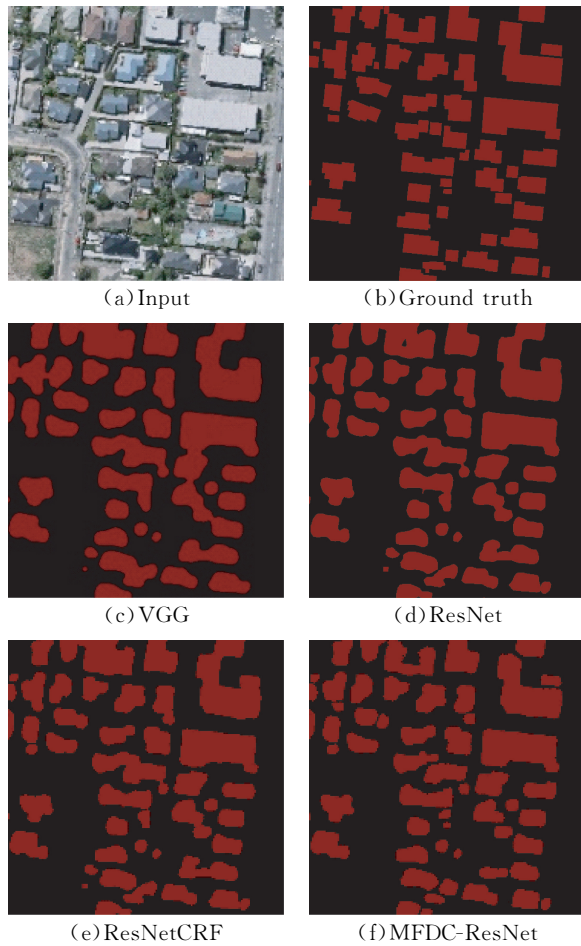


图 13 建筑物距离相隔较近条件下建筑物分割
Fig. 13 Building segmentation under the condition of close distance between buildings

表 4 建筑物距离相隔较近条件下分割性能对比

Tab. 4 Comparison of segmentation performance under the condition of close distance between buildings

模型	mIoU	Recall
VGG	0.737	0.732
ResNet	0.750	0.814
ResNetCRF	0.781	0.801
MFDC-ResNet	0.814	0.879

4.4.3 树木及建筑物阴影较多条件下分割结果分析

图 14 为包含树木和建筑物的阴影干扰的遥感图像建筑物分割对比。由对比分割结果可知, VGG 模型在提取建筑物边缘信息时与真值差别较大, ResNet 和 ResNetCRF 在对建筑物的边缘特征提取时受阴影的影响较大, 所提取的建筑物边缘与所标记的建筑物相比对阴影部分存在明显的误分割现象, 且存在将遥感图像中的树木误识为

建筑物进行分割的现象。提出 MFDC-ResNet 模型可以克服建筑物阴影的干扰对建筑物边缘提取, 且有效克服树木和建筑物阴影等干扰源的影响, 能有效提取遥感图像中的建筑物边缘信息, 特别是在对遥感图像中不规则建筑物的边缘特征的提取结果明显优于 ResNet 和 ResNetCRF 的分割结果。由表 5 定量分析对比可知, 提出 MFDC-ResNet 模型分割精度高于其他 3 种模型。

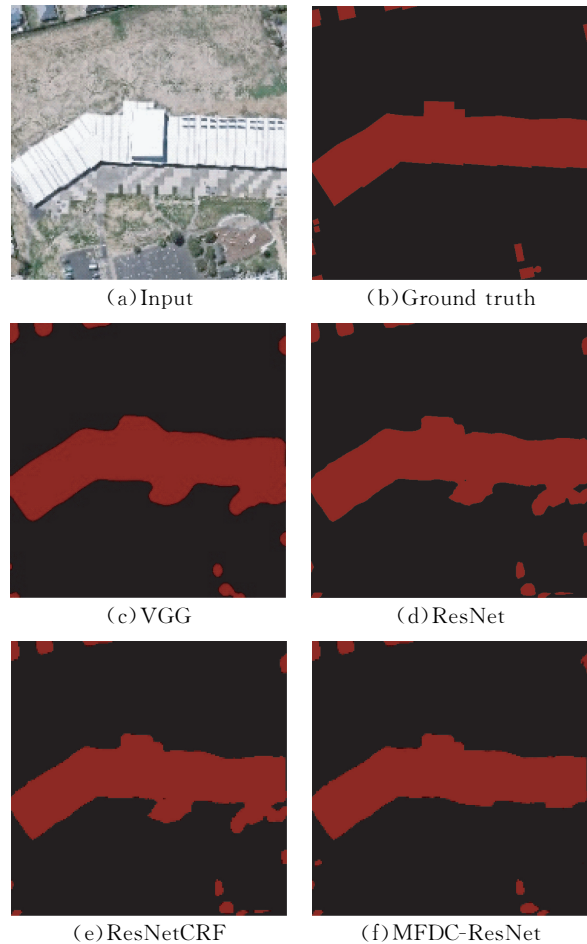


图 14 树木及建筑物阴影较多条件下建筑物分割
Fig. 14 Building segmentation under the condition of more trees and building shadows

表 5 树木及建筑物阴影较多条件下分割性能对比

Tab. 5 Comparison of segmentation performance of trees and buildings with more shadows

模型	mIoU	Recall
VGG	0.746	0.751
ResNet	0.759	0.823
ResNetCRF	0.789	0.813
MFDC-ResNet	0.825	0.886

5 结 论

针对传统建筑物提取方法难以有效描述遥感图像细节特征而导致复杂场景下道路、树木及建筑物之间分割边界不清晰等问题,提出一种基于多尺度特征融合空洞卷积 ResNet (MFDC-ResNet) 遥感图像建筑物分割模型。提出的 MFDC-ResNet 模型包括空洞卷积特征提取、多尺度特征融合和特征解码 3 个模块。空洞卷积特征提取模块改进了空洞卷积中心点特征的提取方式,增强感受野的同时增强了卷积核中心点的对

遥感图像特征信息的描述能力。多尺度特征融合模块在多个层次的特征图上获取的更大范围的遥感图像建筑物上下文信息,可以有效克服遥感图像中道路、树木、建筑物阴影,建筑物不规则及建筑物距离相隔较近等因素的干扰,提升遥感图像中建筑物的分割精度。在 WHU Building change detection dataset 遥感图像建筑物数据集的分割实验中,提出 MFDC-ResNet 模型的平均交并比 mIoU 达到 0.820,召回率 Recall 达到 0.882,与 ResNet 相比,分别提升了 0.066 和 0.063,验证了提出模型的有效性。

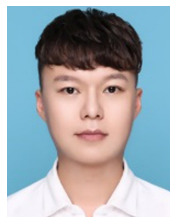
参考文献:

- [1] ANDRES T G, GANCARSKI P, BERTI-EQUILLE L. Remote sensing image analysis by aggregation of segmentation-classification collaborative agents [J]. *Pattern Recognition*, 2018, 73: 259-274.
- [2] WANG M, CUI Q, WANG J, *et al.*. Raftcultivation area extraction from high resolution remote sensing imagery by fusing multi-scale region-line primitive association features [J]. *Isprs Journal of Photogrammetry & Remote Sensing*, 2017, 123(9):104-113.
- [3] DENG J, BAI Y H, LYU F J, *et al.*. Remote sensing alteration information extraction and threshold determination by band ratio method [J]. *Journal of Geology*, 2017, 41(3):504-510.
- [4] ZHAO M, ZHANG X, SHI Z, *et al.*. Restoration of motion blurred images based on rich edge region extraction using a gray-level co-occurrence matrix [J]. *IEEE Access*, 2018, 6:15532-15540.
- [5] 闫敬文, 陈宏达, 刘蕾. 高光谱图像分类的研究进展 [J]. *光学精密工程*, 2019, 27(3):175-188.
YAN J W, CHEN H D, LIU L. Overview of hyperspectral image classification [J]. *Opt. Precision Eng.*, 2019, 27(3):175-188. (in Chinese)
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[C]. *NIPS. Curran Associates Inc*, 2012.
- [7] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. *arXiv:1409.1556*, 2014.
- [8] HE K, ZHANG X, REN S, *et al.*. Deep residual learning for image recognition[C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *IEEE Computer Society*, 2016.
- [9] YE F, LI X, ZHANG X. FusionCNN: a remote sensing image fusion algorithm based on deep convolutional neural networks [J]. *Multimedia Tools and Applications*, 2019, 78(11):14683-14703.
- [10] LIUH, ZENG D, TIAN Q. Super-pixel cloud detection using hierarchical fusion CNN[C]. 2018 *IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, *IEEE*, 2018.
- [11] ZHU B, GAO H, WANG X, *et al.*. Change detection based on the combination of improved SegNet neural network and morphology [C]. 2018 *IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, *IEEE*, 2018.
- [12] BADRINARAYANANV, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481-2495.
- [13] YANG J Y, ZHOU Z X, DU Z R, *et al.*. Rural construction land extraction from high spatial resolution remote sensing image based on SegNet semantic segmentation model [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2019, 35(5):251-258.
- [14] PIRAMANAYAGAM S, SCHWARTZKOPF W, KOEHLER F W, *et al.*. Classification of remote sensed images using random forests and deep learning framework [C]. *SPIE Remote Sensing. Proceedings of the SPIE*, 2016, 10004(10):1-8.
- [15] SAITO S, AOKI Y. Building and road detection from large aerial imagery [C]. *Image Processing*:

- Machine Vision Applications VIII*, 2015: 1814-1821.
- [16] BITTNER K, CUI S Y, REINARLZ P. Building extraction from remote sensing data using fully convolutional network [C]. *ISPRS Hannover Workshop:Hrigr*, 2017:481-486.
- [17] JIAO L, LIANG M, CHEN H, *et al.*. Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification [J]. *IEEE Transaction on Geoscience and Remote Sensing*, 2017, 55 (10):5585-5599.
- [18] XU Y L, ZHU M M, XIN P, *et al.*. Rapid airplane detection in remote sensing images based on multilayer feature fusion in fully convolutional neural networks [J]. *Sensors*, 2018, 18(7):2335-2344.
- [19] 王舒洋, 慕晓冬, 杨东方, 等. 融合高阶信息的遥感影像建筑物自动提取 [J]. *光学精密工程*, 2019, 27(11):2474-2483.
- WANG SH Y, MU X D, YANG D F, *et al.*. High-order statistics integration method for automatic building extraction of remote sensing images [J]. *Opt. Precision Eng.*, 2019, 27 (11): 2474-2483. (in Chinese)
- [20] 朱福珍, 刘越, 黄鑫, 等. 改进的稀疏表示遥感图像超分辨重建 [J]. *光学精密工程*, 2019, 27(3): 213-220.
- ZHU F Z, LIU Y, HUANG X, *et al.*. Remote sensing image super-resolution based on improved sparse representation [J]. *Opt. Precision Eng.*, 2019, 27(3):213-220. (in Chinese)
- [21] SHI Y, CHARLES OTTO, ANIL K J. Face clustering: representation and pairwise constraints [J]. *IEEE Transactions on Information Forensics & Security*, 2017, 13(7):1626-1640.

作者简介:

徐胜军(1976—),男,陕西西安人,工学博士,副教授,硕士生导师,主要从事图像处理、模式识别领域的研究。E-mail: duplin@sina.com

通讯作者:

欧阳朴衍(1993—),男,山西长治人,硕士研究生,主要从事图像处理、遥感图像等方面的研究。E-mail: jokerman1993@foxmail.com