

# 基于 DeepLab-v3+ 的遥感影像分类

袁立<sup>1</sup>, 袁吉收<sup>1\*</sup>, 张德政<sup>2</sup>

<sup>1</sup>北京科技大学自动化学院, 北京 100083;

<sup>2</sup>北京科技大学计算机与通信工程学院, 北京 100083

**摘要** 遥感影像分类是模式识别技术在遥感领域的具体应用, 针对普通卷积神经网络处理遥感图像分类遇到的边缘分类不准确、分类精度低等问题, 提出了一种基于编码解码器的空洞卷积模型(DeepLab-v3+)的遥感图像分类方法。首先标注卫星图像数据; 再利用标注数据集对 DeepLab-v3+ 模型进行训练, 该模型能够提取遥感图像中具有较强稳健性的边缘特征; 最后获得遥感影像地物分类结果。在遥感数据集上进行分析可知, 所提方法比其他分类方法具有更高的分类精度, 更稳健的边缘特征, 以及更优的分类效果。

**关键词** 遥感; 编码解码器; 空洞卷积; 影像分类

中图分类号 TP751

文献标识码 A

doi: 10.3788/LOP56.152801

## Remote Sensing Image Classification Based on DeepLab-v3+

Yuan Li<sup>1</sup>, Yuan Jishou<sup>1\*</sup>, Zhang Dezheng<sup>2</sup>

<sup>1</sup> School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China;

<sup>2</sup> School of Computer and Communications Engineering, University of Science and Technology Beijing, Beijing 100083, China

**Abstract** Remote sensing image classification is a specific application of the pattern recognition technology in the remote sensing field. This study proposes an atrous convolution model based on encoder-decoder (DeepLab-v3+) for performing remote sensing image classification with respect to the inaccurate edge classification and low classification accuracy problems encountered while processing remote sensing image classification using ordinary convolutional neural networks. First, the satellite image data are marked, and the DeepLab-v3+ model is trained using a calibration dataset. This model can extract edge features exhibiting considerable robustness from the remote sensing image. Finally, the classification results of the remote sensing image is obtained. When compared with other classification methods, the proposed method achieves higher classification accuracy, more robust edge features, and better classification results when applied on a remote sensing dataset.

**Key words** remote sensing; encoder-decoder; atrous convolution; image classification

**OCIS codes** 280.4750; 100.4996; 100.1830; 100.2960

## 1 引言

遥感影像分类<sup>[1-5]</sup>是根据遥感数据自身的各种信息, 通过某种算法提取每类地物的独有特征并将其分割为互不相交的子空间, 进而将每个像素划分到对应的子空间。但是, 对于高分辨率的遥感图像, 地物的光谱特征非常丰富, “同物异谱”和“异物同谱”现象非常明显<sup>[6]</sup>。因此传统的参数化方法如最

小距离分类法(MDC)<sup>[7]</sup>、极大似然分类法(MLC)<sup>[8]</sup>等分类准确度降低。而非参数化方法如支持向量机<sup>[9]</sup>、人工神经网络(ANN)<sup>[10]</sup>、决策树(DT)<sup>[11]</sup>等在高分辨率遥感影像分类中得到了广泛的应用。然而这些算法都属于浅层学习算法<sup>[12]</sup>, 难以有效地表达复杂函数, 缺少对复杂样本的适应性, 并且分类器模型的参数无法估计或估计不准确, 导致分类效果不理想<sup>[13]</sup>。

收稿日期: 2018-12-01; 修回日期: 2018-12-14; 录用日期: 2019-03-06

\* E-mail: jishou\_yuan@163.com

2017年,加州大学伯克利分校的 Shelhamer 等<sup>[14]</sup>提出全卷积网络(FCN),使得卷积神经网络(CNN)无需全连接层即可进行密集的像素预测,卷积网络从而得到普及。使用 FCN 方法可生成任意大小的图像分割图,且该方法比图像块分类法速度快。语义分割领域几乎所有先进方法都采用了该模型。除全连接层,使用卷积神经网络进行语义分割还存在另一问题,即池化层因扩大感受野、聚合语境而造成了位置信息的丢失。2015年,Ronneberger 等<sup>[15]</sup>提出了 U-Net。U-Net 能够适应很小的训练集。与 FCN 逐点相加不同,U-Net 采用将特征在通道维度拼接在一起的方法,形成了更厚的特征,所以分割效果更好。但是,实验发现,U-Net 在划分的类别较多时收敛速度较慢,而且物体边缘分割效果不理想。2015年,谷歌发布了第一个版本的 DeepLab 模型<sup>[16]</sup>。之后,经过卷积神经网络特征提取器、目标尺度建模技术、语境信息处理、模型训练流程、深度学习硬件和软件的不断改进和优化,使得 DeepLab 模型升级到了 DeepLab-v2<sup>[17]</sup>和 DeepLab-v3<sup>[18]</sup>。谷歌通过添加一个简单又有效的解码器模块<sup>[19]</sup>以精炼分割结果(尤其是在目标边界处),将

DeepLab-v3 扩展为 DeepLab-v3+<sup>[20]</sup>,并进一步将深度可分卷积应用到空间多孔金字塔池化(ASPP)<sup>[21-23]</sup>和解码器模块上,得到了更快、更强大的语义分割编码器-解码器网络。本文为了解决传统深度神经网络在遥感影像分割时存在的精度低、边缘不准确等问题,采用 DeepLab-v3+模型来提高分割效果。

## 2 基于 DeepLab-v3+ 的遥感图像分类方法

### 2.1 系统概述

应用 DeepLab-v3+ 算法解决遥感影像分类问题。系统框图如图 1 所示。DeepLab-v3+<sup>[19]</sup> 模型把 ASPP 模块应用到编码器-解码器网络执行语义分割任务。在编码过程中,首先应用 ResNet 网络提取原图像特征;然后经过 ASPP 模块,在多个比率、多个有效视野上用滤波器通过卷积提取遥感影像的输入特征并执行池化操作来编码多尺度的上下文信息,在解码器阶段,将低层特征和经过 ASPP 后的特征串联起来再进行卷积;最后进行上采样,逐渐恢复空间信息来捕捉遥感影像更加精细的目标边界<sup>[19]</sup>。

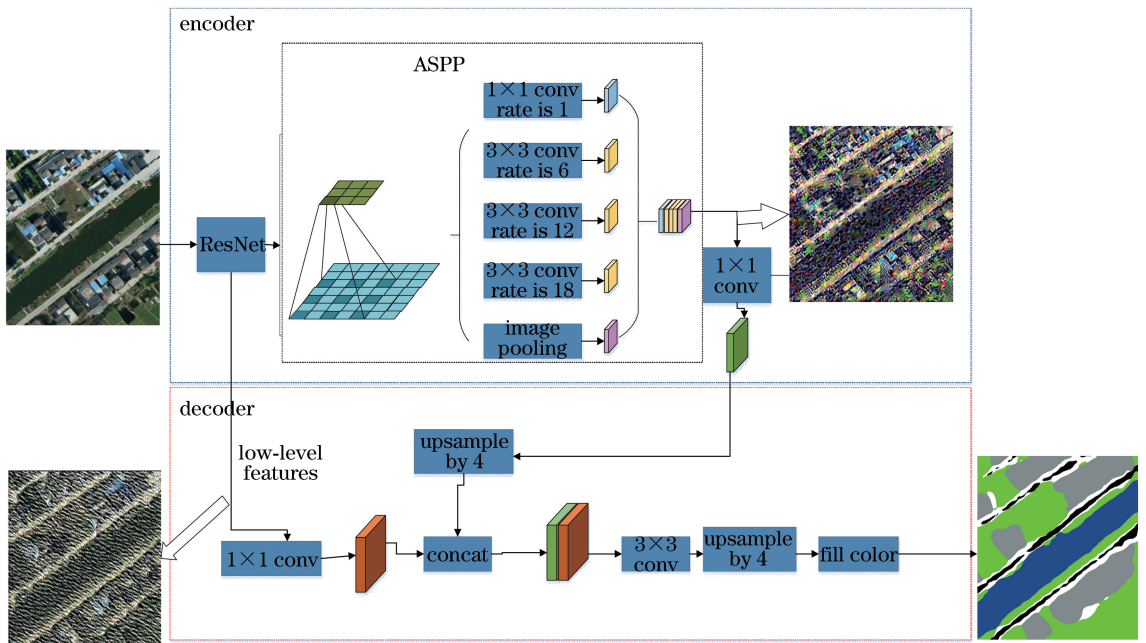


图 1 基于 DeepLab-v3+ 模型的遥感图像分类系统框图

Fig. 1 System diagram of remote sensing image classification based on DeepLab-v3+

### 2.2 ASPP

深度卷积神经网络采用全卷积方式时,会明显降低特征图的空间分辨率。为了解决该问题,有效地生成更稠密的特征图,在深度卷积神经网络

的最大池化层中避免降采样操作,将上采样滤波器加入接下来的卷积层中,在更高的采样率上计算特征图。滤波上采样就是在非零滤波器之间插入空洞,以多孔卷积作为上采样滤波卷积运算。采用多

孔卷积恢复全分辨率的特征地图,并计算更稠密的特征地图,接着在原图像和特征响应之间做简单双线性插值。这种算法为稠密预测任务中使用反卷积层提供了简单有效的方法。与常用的大滤波器卷积

相比,多孔卷积可以有效增加滤波器视野而不增加参数数量或计算量。多孔卷积示意图如图 2 所示。其中:kernel 表示卷积模板大小;stride 表示步长;pad 表示填充大小;rate 表示扩充率参数。

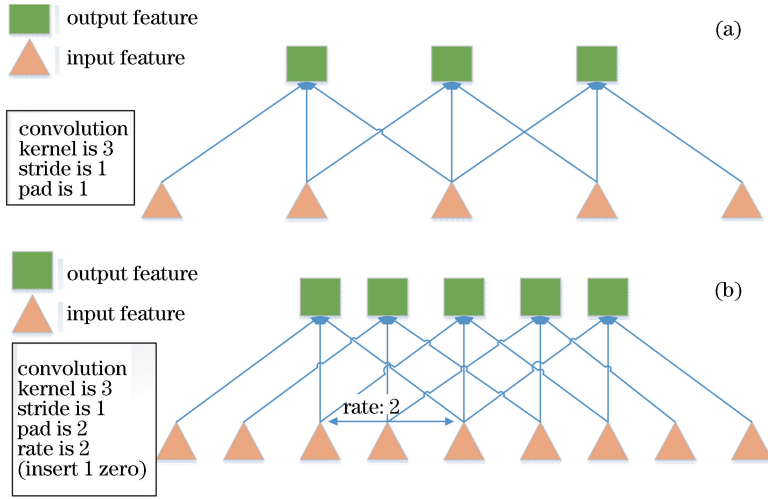


图 2 多孔卷积示意图<sup>[11]</sup>。(a)稀疏的特征提取;(b)稠密的特征提取

Fig. 2 Diagrams of atrous convolution<sup>[11]</sup>. (a) Sparse feature extraction; (b) dense feature extraction

遥感影像在多尺度图像中的状态导致其对多尺度图像的识别精度不高。受到 ASPP 的启发,采用计算效率更好的算法,在多个采样率上重采样特定的特征层来进行卷积。因此,需要用互补有效视野的多个滤波器来检测遥感影像,可以在多个尺度上

对有用图像捕捉地物特征。与真正对特征重采样不同,利用不同采样率的多个并行多孔卷积层做映射的技术被称为“ASPP”。ASPP 示意图如图 3 所示,其中:input 为 ResNet 的输出,输出结果最后送入解码模块。

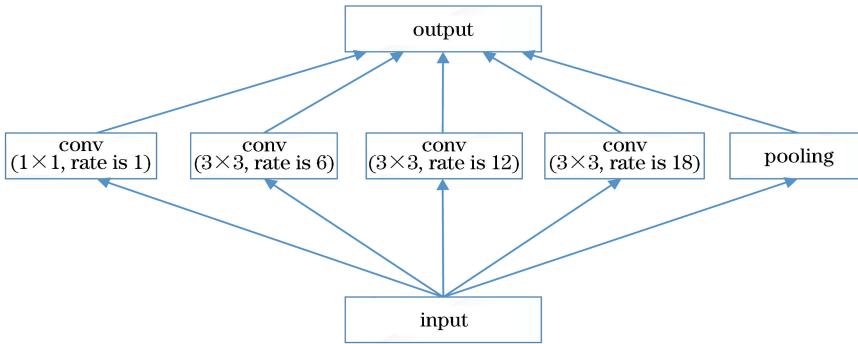


图 3 多孔空间金字塔池化结构

Fig. 3 Structure of atrous spatial pyramid pooling

### 2.3 编码解码模块

由于 ASPP 方法参考了不同尺度的特征图,并且空洞卷积的使用加强了提取稠密特征的能力,因此该结构具有可以提取遥感影像中较稠密特征的优点。该方法中存在池化和有步长的卷积,使得遥感影像的边界信息严重丢失。编码-解码(Encoder-Decoder)方法中的 Decoder 就可以起到修复遥感影像尖锐边界的作用。

具体步骤如下:Encoder 提取出的特征首先被 4 倍上采样,称之为特征图 F1;Encoder 中提取出来的与 F1 同尺度的特征 F2 先进行  $1 \times 1$  卷积,降低通道数得到特征图 F2,再进行 F1 和 F2 的串联,得到 F3;对 F3 进行常规的  $3 \times 3$  卷积微调特征;最后直接 4 倍上采样得到分割结果。示意图如图 4 所示。

## 3 实验与分析

### 3.1 实验环境

采用 Pytorch0.4.0 开源框架,搭建环境为个人

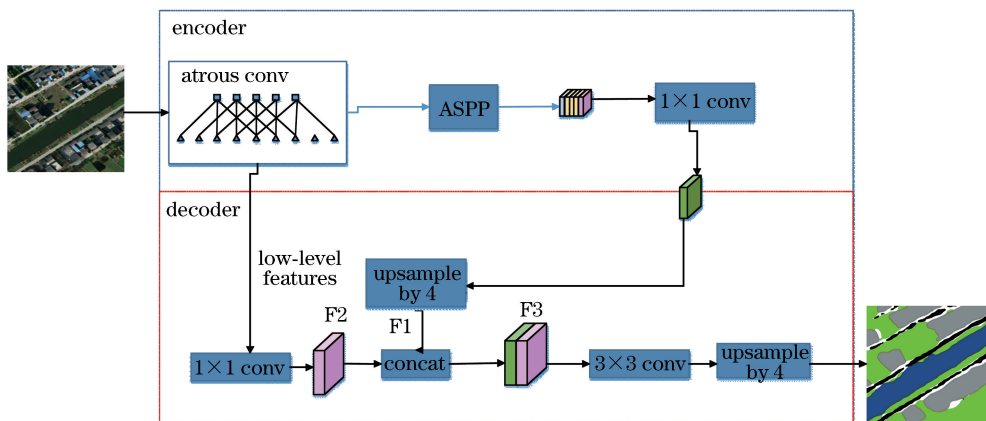


图 4 编码解码模块结构

Fig. 4 Structure of encoder-decoder module

PC,操作系统为 Ubuntu14.04,Python 版本为 3.5,实验采用型号为 Geforce GTX TITAN X 的 GPU 进行训练。

3.2 实验数据及数据增强

所使用的数据集是英国国防科学与技术实验室 (DSTL)于 2017 年 2 月在 kaggle 上发布的数据集。根据该地区实际地物分布情况,分为绿地、道路、水域、居民区以及裸地 5 类,图 5 为数据集示例。将原图切成尺度为 512 pixel × 512 pixel 的图像块,共选择 700 个图像块作为数据集,随机选择 1/10 的数据样本作为验证集,剩余数据样本作为训练集。

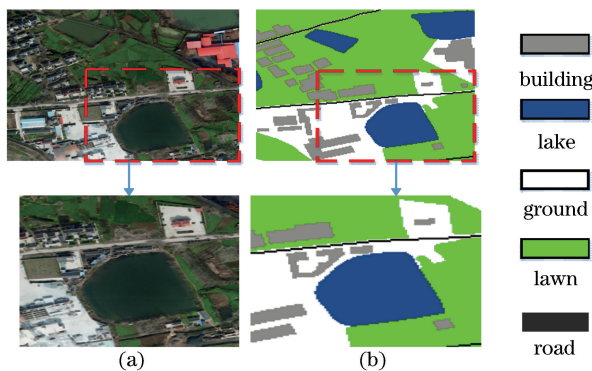


图 5 数据集的示例。(a)原图;(b)标签

Fig. 5 Example of dataset. (a) Original images; (b) labels

由于训练集图片较少,并且为了增加模型的稳健性,对遥感影像数据进行缩放、翻转、裁剪等数据增强方法。本研究图像缩放后的图像分辨率为 500 pixel×500 pixel,然后使用 480 pixel × 480 pixel的滑窗随机裁剪图像,最后以 0.5 的概率翻转图像。

3.3 训练和测试结果分析

使用上述数据集对 DeepLab-v3+模型进行训

练,其训练参数如表 1 所示。其中:“base\_lr”表示基础学习率;“lr\_decay”表示学习衰减率;“batch\_size”表示一个批次学习图像数量;“weight\_decay”表示权重衰减。为防止过拟合,实验采用 Adam 优化方法。最大迭代次数“max\_iter”为 100。

表 1 训练参数

Table 1 Training parameters

Parameter	Value
base_lr	0.0001
lr_decay	5
batch_size	10
weight_decay	0.0001
max_iter	100

每迭代一次大约耗时 45 s,在训练达到 25 次以后就基本完成了收敛,由损失曲线图可以看出,在训练收敛时,选取损失最低的模型作为测试模型,损失曲线如图 6 所示。

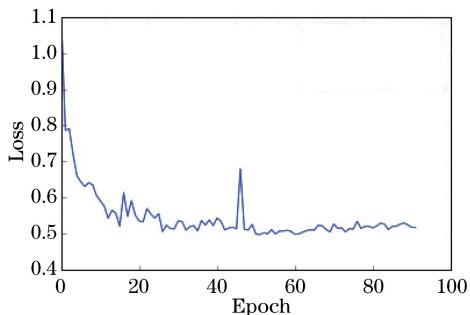


图 6 损失曲线

Fig. 6 Loss curve

提取第一层卷积层卷积后的 64 张特征图,并把得到的特征图可视化,如图 7 所示。可以看出卷积层确实学习到了遥感影像的边缘、纹理、颜色等特征。

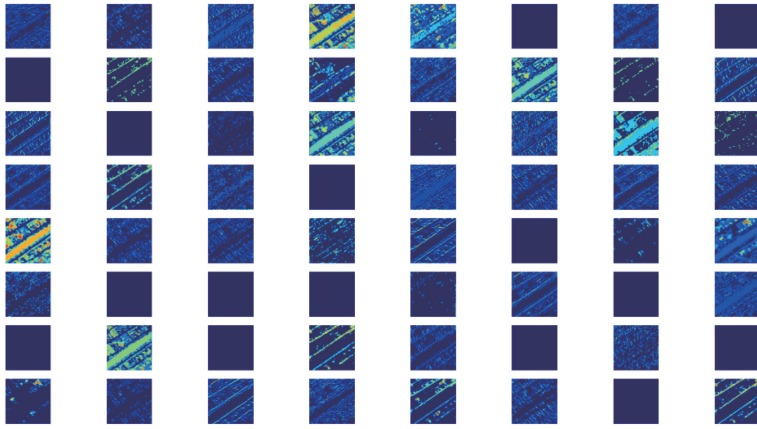


图7 第一层卷积得到的64张特征图

Fig. 7 64 feature maps obtained from the first convolution layer

从测试图像中选取具有代表性的图像,经DeepLab-v3+模型进行分割后,再对其结果进行颜色填充。图8所示为遥感影像分类结果。

由图8可以看出,DeepLab-v3+对卫星影像分割有一定的效果,特别是对于地物的边缘有较好的分割效果。

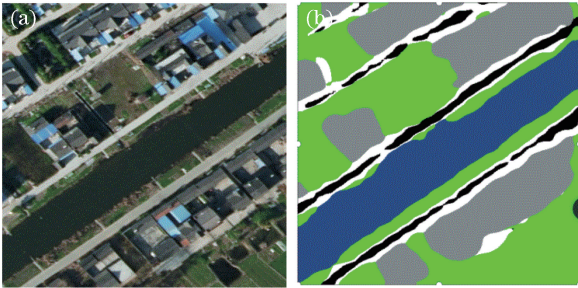


图8 DeepLab-v3+分割效果示意图。

(a)原图;(b) DeepLab-v3+分割结果

Fig. 8 Segmentation effect by DeepLab-v3+; (a) Original image; (b) segmentation result by DeepLab-v3+

### 3.4 DeepLab-v3+与其他模型分割结果对比

为证明所提方法的有效性和优越性,从测试图像中选取了具有代表性的图像,分别使用DeepLab-v3+和FCN、U-Net对遥感图像进行实验,并进行结果对比。

模型复杂度通常使用FLOPs(floating point operations)衡量。本文提出的DeepLab-v3+模型的计算消耗主要在于ResNet,其次是ASPP模块,所以模型的计算复杂度可视二者复杂度之和。本研究的FLOPs为113亿。

图像分割评估指标有4种:像素精度(PA)、均像素精度(MPA)、均交并比(MIoU)、频率加权交并比(FWIoU)。像素精度为标记正确的像素占总像

素的比例。均像素精度是PA的一种简单提升,计算每个类内被正确分类像素数的比例,之后求所有类的平均。均交并比为语义分割的标准度量,计算真实值和预测值的交集和并集之比,该比例可以变形为真正数比上真正、假负、假正(并集)之和,在每个类上计算交并比,之后取平均。频率加权交并比为MIoU的一种提升,这种方法根据每个类出现的频率为其设置权重。MIoU由于简洁、代表性强而成为最常用的度量标准。令 $n_{ij}$ 为属于类 $i$ 预测为类 $j$ 的像素数,其中存在 $n_{cl}$ 个不同的类,因此 $n_{ii}$ 是正确预测的类 $i$ 的像素数。4个指标的表达式为

$$\left\{ \begin{aligned} f_{PA} &= \sum_{i=0}^{n_{cl}} n_{ii} / \sum_{i=0}^{n_{cl}} \sum_{j=0}^{n_{cl}} n_{ij} \\ f_{MPA} &= (1/n_{cl}) \sum_{i=0}^{n_{cl}} (n_{ii} / \sum_{j=0}^{n_{cl}} n_{ij}) \\ f_{MIoU} &= (1/n_{cl}) \sum_{i=0}^{n_{cl}} \cdot \\ & \left[ n_{ii} / \left( \sum_{j=0}^{n_{cl}} n_{ij} + \sum_{j=0}^{n_{cl}} n_{ji} - n_{ii} \right) \right] \\ f_{FWIoU} &= \left( \sum_{i=0}^{n_{cl}} \sum_{j=0}^{n_{cl}} n_{ij} \right)^{-1} \cdot \\ & \left[ \sum_{i=0}^{n_{cl}} \left[ n_{ii} / \left( \sum_{j=0}^{n_{cl}} n_{ij} + \sum_{j=0}^{n_{cl}} n_{ji} - n_{ii} \right) \right] \right] \end{aligned} \right. \quad (1)$$

图9为DeepLab-v3+模型和其他模型在验证集上的损失曲线对比。由图9以及表2可以看出,DeepLab-v3+模型收敛速度比其他模型较快,而且可以达到更低的损失以及更高的精度。

图10所示为DeepLab-v3+、FCN和U-Net模型对遥感影像分类分割并进行颜色填充后的结果。

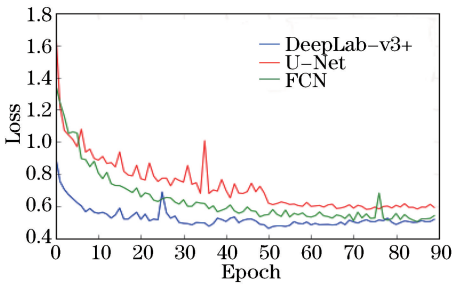


图 9 不同模型的损失对比

Fig. 9 Comparison of loss of different models

结果表明,DeepLab-v3+在分类时对于类别的整体性效果较好,分割结果也比较平滑。DeepLab-v3+

表 2 DeepLab-v3+和其他模型的精度和运行时间  
Table 2 Precision and running time of DeepLab-v3+ and other models

Model	MIoU	Time /h
FCN	0.5068	1.9
U-Net	0.5074	3.1
DeepLab-v3+	0.5743	1.2

使用了ASPP模型,因此结构上可以提取遥感影像中比较稠密的特征,参考不同尺度的特征图,同时使用空洞卷积加强了提取稠密特征的能力,并且编码-解码方法中的解码具有修复遥感影像尖锐边界的作用。

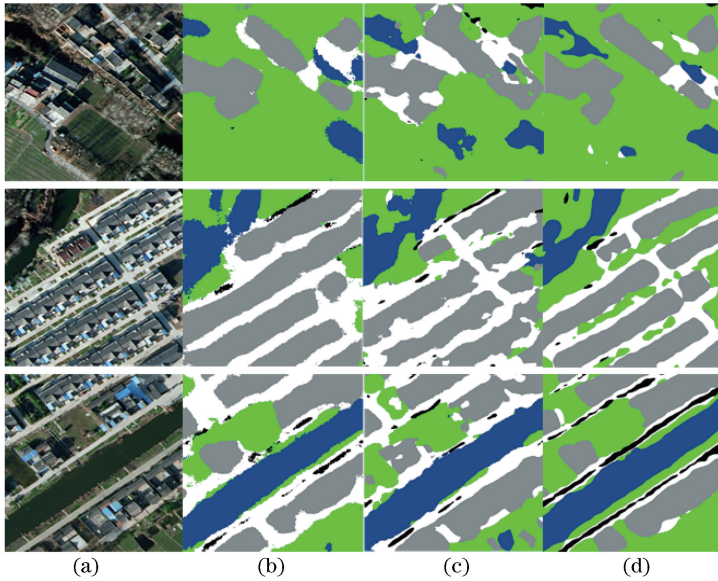


图 10 DeepLab-v3+和其他模型分割结果对比。(a)原图;(b)FCN分割结果;(c)U-Net分割结果;(d)DeepLab-v3+分割结果

Fig. 10 Comparison of segmentation results of DeepLab-v3+ and other models. (a) Original images; (b) segmentation results of FCN; (c) segmentation results of U-Net; (d) segmentation results of DeepLab-v3+

### 3.5 在公开的数据集实验

对本研究提出的带有空洞卷积的编码器-解码器方法在公开的的遥感影像数据集 GID 上进行评估<sup>[24]</sup>。从高分-2(GF-2)卫星获取的数据集的图像显示出我国人口分布与地貌的特点,人口主要分布在东部沿海地区,且以水域为主线集中分布,大量的农田集中分布在建筑物与水域周围;我国地貌呈块状的梯形分布,为地势西高东低的阶梯状分布结构。GF-2 卫星获取的数据集包含了 6 种最常见的土地覆盖类别(水域、建筑物、草地、农田、森林、杂波/背景),并且已被手动精确分类。GID 数据集中包含 150 张分辨率为 7200 pixel × 6800 pixel 的 GF-2 卫星影像。选取 5 张卫星影像,进行数据增强,然后训练。训练参数采用表 1 中的参数。表 3 为所提模型以及其他模型在 GID 数据集上的表现。图 11 为所

提方法和其他模型的分割结果,从图 11 可以看出,所提方法在该数据集上表现更好。

表 3 DeepLab-v3+和其他模型在 GID 数据集的精度  
Table 3 Precision of DeepLab-v3+ and other models on GID dataset

Model	MIoU
FCN	0.5606
U-Net	0.5782
DeepLab-v3+	0.6426

## 4 结 论

提出了一种基于编码解码器的空洞卷积的遥感图像分类方法。该模型能够提取图像中具有较强稳健性的边缘特征,并且错分情况较少。通过与其他神经网络模型相比,该方法能够得到更高的分类精度,提取特征速度更快,可获得更优的分类效果。

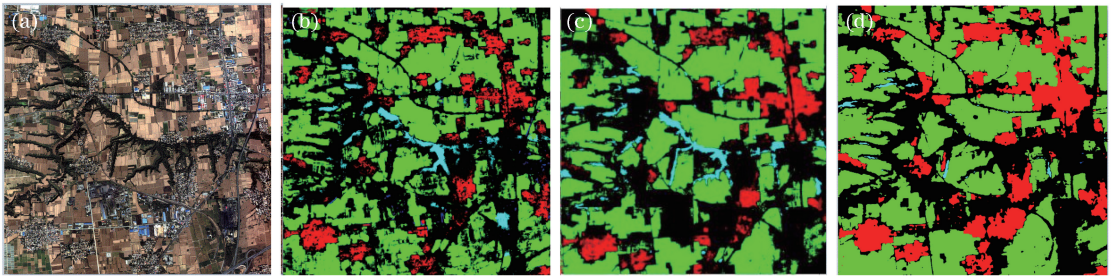


图 11 DeepLab-v3+和其他模型在 GID 数据集分割结果。(a)原图;(b) FCN 分割结果;(c) U-Net 分割结果;  
(d) DeepLab-v3+分割结果

Fig. 11 Segmentation results of DeepLab-v3+ and other models on GID dataset. (a) Original image; (b) segmentation result of FCN; (c) segmentation result of U-Net; (d) segmentation result of DeepLab-v3+

## 参 考 文 献

- [1] Li S H, Wang J L, Bi Y, *et al.* A review of methods for classification of remote sensing images [J]. Remote Sensing for Land & Resources, 2005, 17(2): 1-6.  
李石华, 王金亮, 毕艳, 等. 遥感图像分类方法研究综述[J]. 国土资源遥感, 2005, 17(2): 1-6.
- [2] Shan B H, Huo X Y, Liu Y. A stereovision measurement method using epipolar constraint to correct digital image correlation matching [J]. Chinese Journal of Lasers, 2017, 44(8): 0804003.  
单宝华, 霍晓洋, 刘洋. 一种极线约束修正数字图像相关匹配的立体视觉测量方法[J]. 中国激光, 2017, 44(8): 0804003.
- [3] Zhao F Z, Liang H Y, Wu X L, *et al.* Active contour segmentation model based on local and global Gaussian fitting [J]. Laser & Optoelectronics Progress, 2017, 54(5): 051006.  
赵方珍, 梁海英, 巫湘林, 等. 基于局部和全局高斯拟合的主动轮廓分割模型[J]. 激光与光电子学进展, 2017, 54(5): 051006.
- [4] Song Y, Wu Y Q, Bi S B. Satellite remote sensing cloud image segmentation using edge corrected CV model [J]. Acta Optica Sinica, 2014, 34(9): 0901004.  
宋昱, 吴一全, 毕硕本. 边缘修正 CV 模型的卫星遥感云图分割方法[J]. 光学学报, 2014, 34(9): 0901004.
- [5] Cheriyyadath A M. Unsupervised feature learning for aerial scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(1): 439-451.
- [6] Liu D W, Han L, Han X Y. High spatial resolution remote sensing image classification based on deep learning [J]. Acta Optica Sinica, 2016, 36(4): 0428001.  
刘大伟, 韩玲, 韩晓勇. 基于深度学习的高分辨率遥感影像分类研究[J]. 光学学报, 2016, 36(4): 0428001.
- [7] Alberga V. A study of land cover classification using polarimetric SAR parameters [J]. International Journal of Remote Sensing, 2007, 28(17): 3851-3870.
- [8] Hagner O, Reese H. A method for calibrated maximum likelihood classification of forest types[J]. Remote Sensing of Environment, 2007, 110(4): 438-444.
- [9] Niu X, Ban Y F. Multi-temporal RADARSAT-2 polarimetric SAR data for urban land-cover classification using an object-based support vector machine and a rule-based approach[J]. International Journal of Remote Sensing, 2013, 34(1): 1-26.
- [10] Heermann P D, Khazenie N. Classification of multispectral remote sensing data using a back-propagation neural network[J]. IEEE Transactions on Geoscience and Remote Sensing, 1992, 30(1): 81-88.
- [11] Pal M, Mather P M. An assessment of the effectiveness of decision tree methods for land cover classification[J]. Remote Sensing of Environment, 2003, 86(4): 554-565.
- [12] Bengio Y. Learning deep architectures for AI[M]. Foundations and Trends® in Machine Learning, 2009, 2(1): 1-127.
- [13] Du P J, Xia J S, Xue Z H, *et al.* Review of hyperspectral remote sensing image classification[J]. Journal of Remote Sensing, 2016, 20(2): 236-256.  
杜培军, 夏俊士, 薛朝辉, 等. 高光谱遥感影像分类研究进展[J]. 遥感学报, 2016, 20(2): 236-256.

- [14] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640-651.
- [15] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[M] // Navab N, Hornegger J, Wells W, *et al.* *Medical image computing and computer-assisted intervention. Lecture notes in computer science*. Cham: Springer, 2015, 9351: 234-241.
- [16] Chen L C, George P, Iasonas K, *et al.* Semantic image segmentation with deep convolutional nets and fully connected CRFs [J/OL]. (2016-06-07) [2018-12-01]. <https://arxiv.org/abs/1412.7062>.
- [17] Chen L C, Papandreou G, Kokkinos I, *et al.* DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40 (4): 834-848.
- [18] Chen L C, Papandreou G, Schroff F, *et al.* Rethinking atrous convolution for semantic image segmentation [J/OL]. (2017-12-05) [2018-12-01]. <https://arxiv.org/abs/1706.05587>.
- [19] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39 (12): 2481-2495.
- [20] Chen L C, Zhu Y K, Papandreou G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation [M] // Ferrari V, Hebert M, Sminchisescu C, *et al.* *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11211: 833-851.
- [21] Grauman K, Darrell T. The Pyramid match kernel: discriminative classification with sets of image features [C] // Tenth IEEE International Conference on Computer Vision (ICCV'05), October 17-21, 2005, Beijing, China. New York: IEEE, 2005: 8824338.
- [22] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories [C] // 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), June 17-22, 2006, New York, NY, USA. New York: IEEE, 2006.
- [23] He K M, Zhang X Y, Ren S Q, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition [M] // Fleet D, Pajdla T, Schiele B, *et al.* *Computer vision - ECCV 2014. Lecture notes in computer science*. Cham: Springer, 2014, 8691: 346-361.
- [24] Ge W Y, Liu G Y. Unsupervised classification of high-resolution remote-sensing images under edge constraints [J]. *Proceedings of SPIE*, 2017, 10609: 106091C.