计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2021,30(10):319−324 [doi: 10.15888/j.cnki.csa.008175]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

遥感影像中建筑物的 Unet 分割改进①

黄　杰, 蒋　丰

(广东工业大学 自动化学院, 广州 510006)
通讯作者: 黄　杰, E-mail: 694141167@qq.com

摘　要: 针对经典 Unet 算法在提取遥感影像中建筑物特征时存在编码信息丢失、对多尺度建筑目标适应性差和上下文特征联系不足的问题, 本研究提出了一种多尺度融合的变形残差金字塔编解码网络. 首先, 引入深度编码网络与下采样旁路网络替换原编码结构, 共同完成对建筑物目标高阶特征信息的提取; 其次, 在编码网络次末端节点引入联合变形卷积的残差金字塔结构, 以提升网络对建筑物多尺度特征和边缘模糊特征的辨识能力; 最后, 将高阶和低阶特征逐层级联融合, 在解码网络末端获取对建筑物的分割结果. 实验结果表明, 改进后模型相比原模型在 *F*1-*score* 和 *MIOU* 指标上分别提升了 1.6% 和 2.1%.

关键词: 建筑物提取; Unet 算法; 语义分割; 金字塔结构; 变形卷积; 编解码网络

引用格式: 黄杰,蒋丰.遥感影像中建筑物的 Unet 分割改进.计算机系统应用,2021,30(10):319–324. http://www.c-s-a.org.cn/1003-3254/8175.html

Segmentation of Buildings in Remote Sensing Images by Improved Unet Algorithm

HUANG Jie, JIANG Feng

(School of Automation, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: The loss of coding information, the poor adaptability to multi-scale building targets, and the insufficient contextual feature connection can be found in the classic Unet algorithm during the extraction of building features from remote sensing images. To tackle these problems, this study proposes a deformed-residual-pyramid codec network with multi-scale fusion. First, the original coding structure is replaced by the deep coding network and the down-sampling bypass network, which jointly extract the high-level feature information of the building target. Second, the residual pyramid structure combined with deformed convolution is introduced at the penultimate node of the coding network to improve the network's ability to recognize multi-scale features and edge fuzzy features of buildings. Finally, the high- and low-level features are cascaded and merged layer by layer, and the segmentation result of the building is obtained at the end of the decoding network. The experimental results show that compared with the original model, the improved model has increased *F*1-score and MIoU by 1.6% and 2.1%, respectively.

Key words: building extraction; Unet algorithm; semantic segmentation; pyramid structure; deformed convolution; codec network

建筑物是人们在工作和学习中不可或缺的活动场所, 从遥感影像中提取建筑物相关目标对于 GIS 数据库更新、土地利用、城市规划和自然灾害探测等工程具有重要意义[1]. 因此研究人员针对建筑物的提取提出了许多基于传统或者深度学习的分割方法.

建筑物目标丰富的直线、直角和阴影等特性可被传统方法作为建模和分割的依据. 然而传统方法的构建需强烈依赖于对特定目标的先验知识, 过程费时费

① 收稿时间: 2021-01-05; 修改时间: 2021-02-03; 采用时间: 2021-03-16

力, 因此近年来, 更多基于深度学习语义分割的建筑物提取方法被研究人员所提出. Zhong 等人[2] 利用预训练参数对网络进行训练, 通过对比分析 FCN 网络中解码器的特征融合层数对模型精度的影响, 提出了改善后的网络模型, 但由于其较为简单的网络结构调整, 得到的遥感影像仍存在信息缺失问题. 尚群锋等人[3] 针对遥感影像中小物体特征在高纬度难以被提取的问题, 提出了改进的 DeconvNet 网络, 该网络通过记录编码过程的池化索引并将其应用到解码恢复过程的方式改进网络解码部分, 从而减少了图像恢复的盲目性, 并最终提高了对小物体的分割效果, 但该方法需占用较大的机器内存, 对于大物体容易出现边缘不平滑的情况. 赵斐等人[4] 提出了一种端到端的语义分割模型. 该模型秉承 Unet 算法中编解码结构的思想, 通过引入注意力机制调整金字塔中各个通道中特征的权重, 提取具有信息侧重的多尺度特征, 解决物体边缘分割模糊的问题, 同时小目标漏检情况也得到了改善. 苏健民等人[5] 专注于像素间的联系问题, 引入神经网络中常被人们忽略的后处理操作并提出了一种基于 Unet 的改进方法, 其首先采用集成学习的策略, 为建筑、道路和水体等每一类地物目标训练一个二分类模型, 随后将各预测的子图进行组合以生成最终的分割结果, 该模型性能虽获得一定的提高, 但是“分类训练+后处理”的分割策略在操作上仍稍显繁琐, 且部分空间信息仍存在丢失问题.

尽管上述方法相比传统方法能更便捷地实现对遥感影像中建筑物等目标的分割, 但他们未能综合考虑建筑物目标轮廓的多样性、网络编码过程空间和细节信息的丢失以及深层语义信息间上下文联系存在不足等问题, 导致了网络模型在面对建筑物边缘以及对应的分割完整性上仍有提升的空间. 为此本文基于经典 Unet 算法[6], 通过设计下采样旁路网络和联合变形卷积的残差金字塔网络, 提出了多尺度融合的变形残差金字塔网络方法, 有效提高了模型的分割精度.

1 模型架构与改进

1.1 多尺度融合的变形残差金字塔网络模型

本文所提多尺度融合的变形残差金字塔网络模型 (Multi-scale fusion of Deformation Residual Pyramid Network, MDRP-Net) 如图 1 所示. 其主要包含 3 个部分: 下采样旁路主干网络、联合变形卷积的残差金字塔网络结构和级联上采样解码器. 下采样旁路主干网络由 VGG16[7] 主干网络和下采样旁路网络组成, 主干网络主要用于挖掘建筑物深层次特征; 下采样旁路网络结构则把输入影像进行不同程度的下采样, 用于对 VGG16 网络获取的多层次特征图进行融合补充. 对于 VGG16 主干网络次末端的卷积层输出, 其既作为提取网络最深层特征的卷积层输入, 也作为联合变形卷积的残差金字塔网络结构的初始输入, 以并行融合方式增加深层语义的丰富程度. 网络的级联上采样解码器, 接收综合下采样旁路主干网络和联合变形卷积的残差金字塔结构两部分的多层次、多尺度特征信息图, 然后把获取的多特征融合图向前上采样逐步恢复图像尺寸与细节, 最后将其送入网络的末端判别器实现对遥感建筑物影像的预测和分割

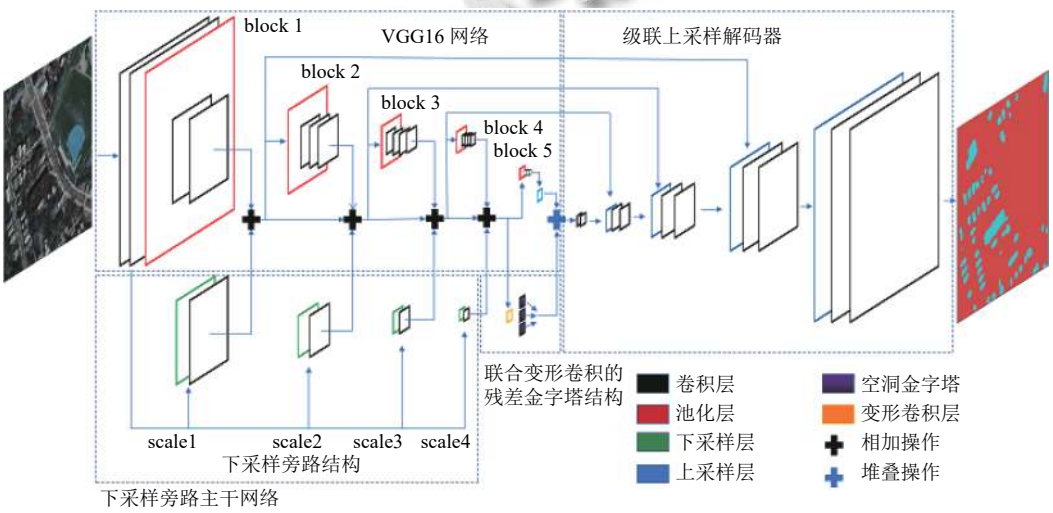

图 1 多尺度融合的变形残差金字塔网络模型

1.2 网络的改进

1.2.1 下采样旁路主干网络

经典 Unet 模型被广泛地应用医学图像分析的领域, 但由于简单的编码结构, 使其无法适应背景信息更为复杂、干扰信息更多的对象. 而 VGG 网络相比 18 个主流特征提取网络具有更优的迁移性[8], 因此本文把网络的编码结构使用 VGG16 网络将其替换并作为主干网络, 同时, 对修改后主干网络进一步添加一个下采样旁路结构作为网络补充.

在该旁路网络中, 本文使用最大池化操作将网络最初输入影像分别下采样至原大小的 1/4、1/16、1/64 和 1/256 倍, 此时能得到 4 种不同尺寸的图像, 并将其记录为 scale1–scale4. 在主干网络中, 每个包含卷积池化的 block$l$ (l=1, 2, 3, 4) 块也能得到 4 种不同分辨率的输出图像, 这些输出图像刚好与 scale$l$ 图像大小相同. 我们将 block$l$ 块的输出图像和 scale$l$ 的图像进行相加融合, 分别作为下一层网络的输入进而使下一层卷积层获得两个尺度的特征信息.

1.2.2 联合变形卷积的残差金字塔结构

根据变形卷积方法的思想[9], 其可通过训练获取卷积核偏移坐标从而指导卷积核采样点的选取. 这意味着利用该偏移坐标网络可以更针对性地对建筑物轮廓特征进行模拟与提取. 然而偏移坐标存在着偏移大小的限制, 这使得变形卷积核的感受野与传统卷积核相差不大, 导致变形卷积在面对多尺度目标时仍存在不足, 因此本文引入金字塔池化结构以扩大变形卷积对不同尺度特征的捕获能力. 同时, 在 Deeplab[10] 系列中, 作者强调空洞卷积的使用和提出 ASPP 模块来聚合不同模块和不同尺度间的上下文信息. 这些方法虽然有效, 但是他们仅简单地对尾部特征进行拼接的方式会导致上下文间仍存在语义鸿沟的问题. 综合上述问题, 本文设计一种联合变形卷积的残差金字塔模块 (Deformation Residual Spatial Pyramid, DRSP), 如图 2 所示.

与 DeepLabV3+[11] 方法使用金字塔结构的方式相比, 本文提出的 DRSP 模块是基于主干网络 block4 特征图作为输入的, 其首先经过变形卷积获取变形特征, 再进一步对变形特征提取多尺度上下文信息. 同时, 为了减少上下文语义信息的差距,不同尺度特征之间使用残差模块来逐层聚合它们. 在形式上可描述为式 (1).

$$X_{\text{raspp}}=\begin{cases} H([h_1,h_2,h_3,\cdots,h_n]) \\ h_n=f(f(f(X_1)\oplus X_2)\oplus X_3)\oplus\cdots\oplus X_n) \\ d_1<d_2<d_3<\cdots<d_n \end{cases} \tag{1}$$

其中, X_{raspp} 为 DRSP 模块的最终聚合特征, d_n 为卷积核膨胀率, $H([\cdot])$ 为通道串联操作, X_n 代表从变形特征获取的不同尺度特征, f 代表残差模块[12], $\oplus$ 表示元素求和. 在 DRSP 模块逐层聚合上下文信息的过程中, 卷积核膨胀率逐渐增大, 同时其膨胀率大小根据 Wang 等人[13] 的公式推荐以及实验的尝试, 设定为 1、2、5、9、13.

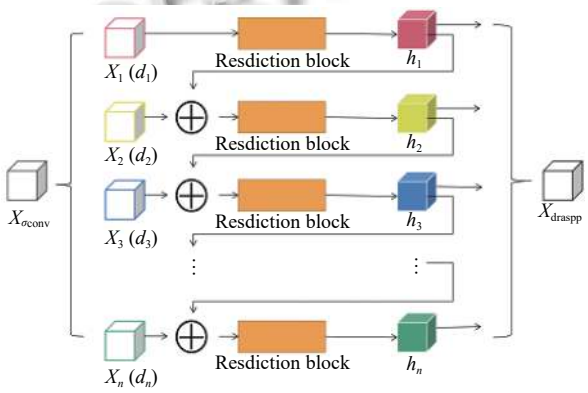

图 2 联合变形卷积的残差金字塔结构

2 实验与分析

2.1 实验数据集

本实验数据集选取遥感建筑物影像 Massachusetts Buildings[14]. 数据集中包含了 137 张训练影像数据、4 张验证影像数据、10 张测试影像数据, 每张图像尺寸为 1500×1500 像素. 为了适应硬件条件和便于训练, 本文对原图按 256×256 像素大小进行裁剪. 裁剪后按随机旋转、引入高斯噪声、随机缩放策略对训练数据进行扩增, 最终获得训练集大小为 11 664 张, 测试集大小为 360 张, 验证集大小为 144 张.

2.2 实验设计和参数设定

实验设计部分, 选用两个使用了金字塔池化结构的网络方法 PSPNet[15] 和 DeepLabV3+与本文方法进行对比, 同时, 另设计 3 组实验对比各改动方法对网络性能的影响. 实验 1: 在经典 Unet 算法基础上, 单独添加下采样旁路主干网络; 实验 2: 在经典 Unet 算法上, 单独添加 DRSP 模块; 实验 3: 在经典 Unet 算法上, 同时添加下采样旁路主干网络和 DRSP 模块.

训练样本输入大小为 256×256, batchsize 大小为 4, 训练 100 代. 网络训练过程, 不同网络模型使用超参数相同: 初始学习率为 0.01, 学习率衰减率为 $1e^{-2}$, 动量值为 0.9. 训练过程中使用监测器对测试集损失值进行监测, 当损失值连续 50 代没有下降, 则认为模型训练完毕, 训练提前停止.

对于建筑物遥感影像语义分割, 是属于二分类的任务, 网络模型在训练过程中将使用交叉熵作为损失函数, 其表达式如下:

$$Loss = -\frac{1}{n}\sum_{i=1}^{n}(y_i \cdot \ln \hat{y}_i + (1-y_i)\cdot \ln(1-\hat{y}_i)) \tag{2}$$

其中, n 表示类别数量, y_i 表示真值, $\hat{y}_i$ 表示当前像素预测的值.

2.3 实验评价指标

实验结果评价指标采用均交并比 *MIoU*[2] 和可用于衡量二分类模型精确度的指标 *F*1-*score*[16], 计算公式如下:

$$MIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{P_{ii}}{\sum_{j=0}^{k}P_{ij}+\sum_{j=0}^{k}P_{ji}-P_{ii}} \tag{3}$$

$$F1\text{-}score = 2\times\frac{Precision\times Recall}{Precision+Recall} \tag{4}$$

$$Precision = \sum_{i=0}^{k}\frac{P_{ii}}{P_{ii}+\sum_{j=0}^{k}P_{ij}} \tag{5}$$

$$Recall = \sum_{i=0}^{k}\frac{P_{ii}}{P_{ii}+\sum_{j=0}^{k}P_{ji}} \tag{6}$$

式中, P_{ii} 表示预测正确的像素, P_{ij} 表示预测为建筑物, 实际为非建筑物的像素, P_{ji} 表示预测为非建筑物, 实际为建筑物的像素, *Precision* 表示精确率, *Recall* 表示召回率.

2.4 实验结果汇总与分析

图 3 和表 1 分别是各实验模型损失值对比曲线和模型测试结果的汇集.

PSPNet 与 DeepLabV3+是语义分割网络中具有代表性的方法, 两者曾在 PASCAL VOC-2012 数据集获得过优异的成绩, 尽管在面对遥感建筑物数据集时其损失函数曲线相比 Unet 更加平滑, 然而两者在最终的评价指标以及可视化结果上的表现均不如经典的 Unet 网络.

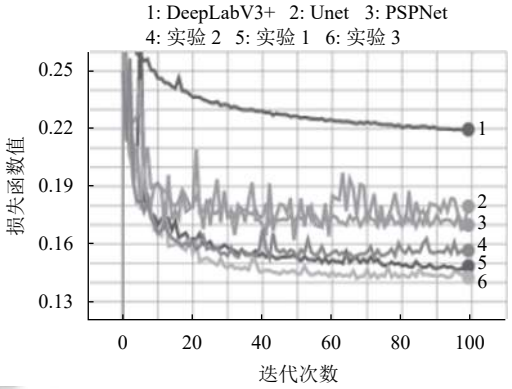

图 3 损失函数值对比曲线

表 1 模型测试结果汇集

实验名称	*Precision*	*Recall*	*F*1-*score*	*MIoU*
DeepLabV3+	0.792	0.64	0.708	0.588
PSPNet	0.802	0.751	0.775	0.671
Unet	0.841	0.854	0.847	0.754
实验1	0.878	0.840	0.858	0.768
实验2	0.864	0.843	0.853	0.762
实验3	0.879	0.848	**0.863**	**0.775**

实验 1 通过将 VGG16 主干网络与下采样旁路网络两者特征按层次融合的方式, 使得网络在挖掘更深层特征的同时具备与浅层特征信息的联系. 相比改进前的网络, 改进后网络损失函数值波动幅度明显减小, 整体损失值降低了约 0.02, 且训练迭代约 75 次时损失值再度降低并最终进入稳态. 经测试, 改进后网络最终在 *F*1-*score* 和 *MIoU* 指标分别获得了 1.1% 和 1.4% 的提升.

实验 2 将 DRSP 结构与主干网络两者的输出特征进行融合, 尽管该网络损失值函数曲线没有实验 1 平滑, 但相比改进前网络其损失函数波动浮动和损失值均有一定程度地改善, 经测试, 实验 2 网络在 *F*1-*score* 和 *MIoU* 指标获得了 0.6% 和 0.8% 的提升.

实验 3 通过把实验 1 与实验 2 改进方法共同作用于原网络, 图 3 中对应的曲线显示表明改进后的网络缓解了单独引入 DRSP 模块时存在的损失函数曲线的波动, 且训练至大约 20 代时就达到此前实验最优损失值附近, 同时在迭代约 60 代时进入稳态. 最终测试结果也比两组单独的改进实验效果更好, 最终其在 *F*1-

score 和 *MIoU* 指标上相比 Unet 算法分别提升了 1.6% 和 2.1%.

为了更直观感受模型的改进对分割性能所带来的影响, 本文把个各实验模型语义分割的部分预测图进行了可视化, 如图 4 所示. 图中展示了本文所提方法的优势, 其主要体现在建筑物与背景模糊分界的区域以及对中大型建筑物分割的完整性这两个方面. 受光线和阴影影响, 建筑物边缘与背景区域区分度低, 如图 4(a)–图 4(d) 中建筑物边缘存在绿植、阴影或者颜色相似的道路等干扰, 导致建筑物与背景出现分界模糊的情况, 但相较原 Unet 网络, 本文所提方法能更好地区分此类建筑物的边界区域, 以改善对建筑物边缘分割的准确性. 另一方面, 由于原始模型仅使用单一规则的卷积核和较简单主干网络, 致使其对不同尺寸特别是较大型建筑物特征信息捕获能力存在一定限制, 如图 4(e)–图 4(g) 中建筑物中间部分出现的漏空现象. 可以看出, 相对未改进的方法, 本文所提方法拥有更强的多尺度目标的适应能力和特征信息保留的能力, 从而在面对中大型建筑物时具有更完整的分割.

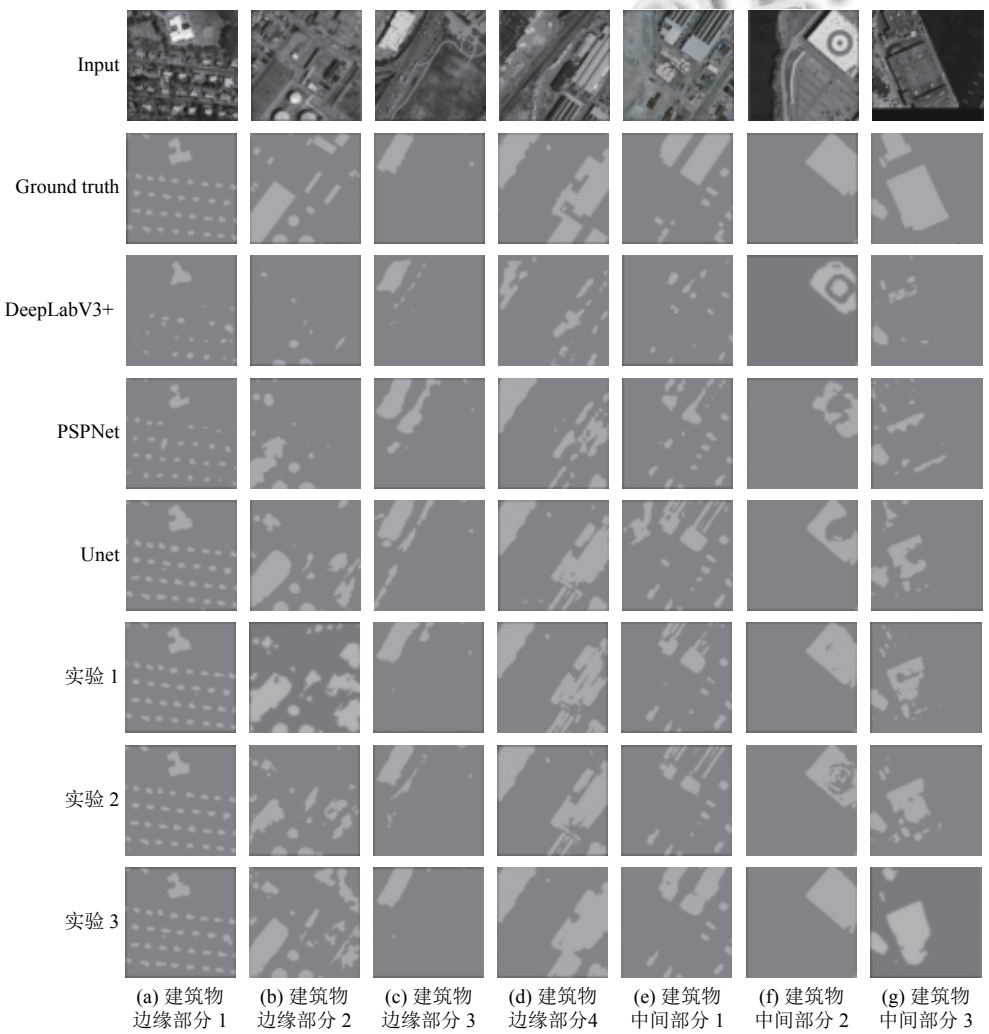

图 4 各实验模型建筑物分割效果对比

3 结语

本文研究了相关语义分割网络在遥感建筑物影像中的应用, 针对网络中传统卷积核模拟几何结构特征能力存在不足、对目标尺寸适应能力不足和编码网络中特征信息容易丢失的问题, 提出了下采样旁路主干网络和多尺度融合的变形残差金字塔卷积网络. 该网络模型融合下采样旁路主干网络、变形残差金字结构和级联上采样解码器 3 部分特征, 实现了对原模型网

络结构的优化. 最后, 本文在 Mnih 遥感建筑物数据集上进行了对照实验, 其实验指标和可视化结果均验证了本文改进措施的有效性.

参考文献

1 杨州, 慕晓冬, 王舒洋, 等. 基于多尺度特征融合的遥感图像场景分类. 光学精密工程, 2018, 26(12): 3099–3107.

2 Zhong ZL, Li J, Cui WH, *et al*. Fully convolutional networks for building and road extraction: Preliminary results. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Beijing: IEEE, 2016. 1594–1594.

3 尚群锋, 沈炜, 帅世渊. 基于深度学习高分辨率遥感影像语义分割. 计算机系统应用, 2020, 29(7): 180–185. [doi: 10.15888/j.cnki.csa.007487]

4 赵斐. 基于金字塔注意力机制的遥感图像语义分割. 国外电子测量技术, 2019, 38(8): 150–154.

5 苏健民, 杨岚心, 景维鹏. 基于 U-Net 的高分辨率遥感图像语义分割方法. 计算机工程与应用, 2019, 55(7): 207–213. [doi: 10.3778/j.issn.1002-8331.1806-0024]

6 Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. Proceedings of 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015. 234–241.

7 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations. arXiv: 1409.1556v3, 2014.

8 Su D, Zhang H, Chen HG, *et al*. Is robustness the cost of accuracy?—A comprehensive study on the robustness of 18 deep image classification models. Proceedings of 15th European Conference on Computer Vision. Cham: Springer, 2018. 644–661.

9 Dai JF, Qi HZ, Xiong YW, *et al*. Deformable convolutional networks. Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017. 764–773.

10 Chen LC, Papandreou G, Kokkinos I, *et al*. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834–848. [doi: 10.1109/TPAMI.2017.2699184]

11 Chen LC, Zhu YK, Papandreou G, *et al*. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the 15th European Conference on Computer Vision (ECCV). Cham: Springer, 2018. 833–851.

12 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 770–778.

13 Wang PQ, Chen PF, Yuan Y, *et al*. Understanding convolution for semantic segmentation. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Tahoe: IEEE, 2018. 1451–1460.

14 Mnih V. Machine learning for aerial image labeling. Toronto: University of Toronto, 2013.

15 Zhao HS, Shi JP, Qi XJ, *et al*. Pyramid scene parsing network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 6230–6239.

16 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. 光学精密工程, 2019, 27(11): 2474–2483.