

基于改进Encoder-Decoder网络的遥感影像道路提取

尹耀, 张春亢, 吉雨田, 邵小美, 韦永昱

贵州大学矿业学院, 贵州 贵阳
Email: 1538912058@qq.com

收稿日期: 2020年12月25日; 录用日期: 2021年1月19日; 发布日期: 2021年1月28日

摘要

本文针对高分遥感影像中蕴含的丰富信息给道路提取结果带来的干扰问题, 提出了一种基于改进Encoder-Decoder网络的高分遥感影像道路提取方法。首先在编码区引入残差模块提取图像特征信息, 然后在网络的中心区域引入空洞卷积模块, 进一步拓展了识别道路特征像素信息的感受野, 并保障了特征图分辨率以及像素点空间信息保持不变, 增强了网络的细节提取能力, 最后使用Sigmoid函数对特征图进行分类。本次实验采用马萨诸塞州道路数据集作为训练数据, 实验结果表明, 本文所提出的基于改进编码-解码网络道路提取方法将整体的精度、召回率以及F1-score指标分别提升至91%, 58%和71%, 与U-Net模型相较有着明显的提升。

关键词

高分遥感影像, 道路提取, 深度学习, 编解码网络, 膨胀卷积

A Remote Sensing Image Road Extraction Method Based on Improved Encoder-Decoder Network

Yao Yin, Chunkang Zhang, Yutian Ji, Xiaomei Shao, Yongyu Wei

The Mining College of Guizhou University, Guiyang Guizhou
Email: 1538912058@qq.com

Received: Dec. 25th, 2020; accepted: Jan. 19th, 2021; published: Jan. 28th, 2021

文章引用: 尹耀, 张春亢, 吉雨田, 邵小美, 韦永昱. 基于改进 Encoder-Decoder 网络的遥感影像道路提取[J]. 应用数学进展, 2021, 10(1): 274-281. DOI: 10.12677/aam.2021.101031

Abstract

Aiming at the rich details and simple semantics of road information in high-resolution remote sensing images, a high-resolution remote sensing image road extraction model based on improved Encoder-Decoder network is designed. The residual module is introduced in the encoding area to extract image feature information. The introduction of dilated convolution module in the central area of the network further expands the receptive field of identifying the characteristic pixel information of the road, and ensures that the resolution of the characteristic map and the spatial information of the pixel remain unchanged, which enhances the detail extracting ability of the network. Finally, the Sigmoid function is used to classify the feature map. Through the comparison test of the experimental verification set, the overall accuracy, recall rate and F1-score index of the road extraction method based on the improved encoding-decoding network proposed in this paper reached 91%, 58%, and 71% respectively. Comparing to the U-Net model, there is a significant improvement.

Keywords

VHR Image, Road Extraction, Deep Learning, Encoder-Decoder Network, Dilated Convolution

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

交通道路网作为地图以及城市地理信息系统中一个必不可少的元素,对城市规划、地图测绘、地理信息系统更新[1]都具有非常重要的意义。从遥感影像中提取道路信息一直都是遥感应用领域的热门研究方向,随着高分辨率遥感影像在遥感研究领域的普及和广泛应用,如何利用高分辨率遥感影像实现交通道路信息的自动提取已然成为诸多学者研究的热点问题之一。基于不同的研究思路,遥感影像的道路提取方法,可分为基于像元[2] [3] [4]、基于面向对象[5] [6] [7]、基于深度学习[8] [9]三大类[10]。

随着计算机软硬件的快速更新迭代,深度学习作为人工智能领域中的一个热门研究方向,得到了迅速的发展,开始从众多的传统机器学习方法中脱颖而出,在图像分类[11] [12]、图像分割以及目标检测[13] [14]等计算机视觉领域表现出巨大的潜力与价值。深度学习模型能够从原始图像数据的像素级层面,逐层提取抽象的语义概念,这使得它在图像的特征提取方面具有显著优势。Mnih V [15]首次将深度学习引入到遥感影像道路提取的研究,提出了一种基于受限波尔兹曼机的高分辨率航拍影像的道路提取方法。刘笑等[8]结合图像的高维特征和深度神经网络隐藏层信息,提出一种基于全卷积神经网络的道路提取方法,较好的实现了道路的初提取,但在反向传播过程中数据降维的需要导致提取结果粗糙。郭正胜等[9]利用在分割医学图像方面效果显著的 U-net 模型,提高了道路的初提取精度,但是提取结果会出现粘连现象。上述基于深度学习语义分割模型的道路提取方法一定程度上提高的道路提取的精度和效果,但是在泛化能力上仍需要进一步的增强。

本文提出基于改进 Encoder-Decoder 网络模型,可以实现高效、智能化遥感影像的道路提取,弥补传统方法场景适用性方面的不足。在 Encoder-Decoder 网络构架的基础上,将由多组残差模块堆叠而成的深度残差网络引入编码器中,并在网络的中心区域加入膨胀卷积层,然后以瓶颈连接结构组成反卷积层,

完成本文编码解码网络结构的搭建。本文所提出方法有效的结合以上各模块的提取优势，可以在减小特征图分辨率损失的同时更细致的挖掘图像深层次特征，获得良好的道路提取效果。

2. 算法原理

深度学习应用到遥感图像提取时，传统卷积神经网络中的全连接层将被舍弃，并以可以增大图像尺寸的反卷积层替代，由此保留的卷积神经网络主要由输入层、隐藏层、输出层构成，其中隐藏层又包括卷积层、池化层和激活函数。此类隐藏层仅由卷积层、池化层、激活函数组成的网络结构通常被称作全卷积神经网络，它能对输入图像的每个像素都产生预测，输出跟任意尺寸的输入图像保持一致大小的结果。

在前馈神经网络的某一层中引入卷积运算代替一般矩阵乘法，组成卷积层，其作用是提取输入图像的目标特征，通过增加卷积层的层数可以提取深层次的语义特征。卷积运算是指通过数据以及卷积核的内积得到特征图的过程。其公式可以表示为：

$$X_j^l = f \left[\sum_{i \in M_j} X_i^{l-1} \cdot w_{ij}^l + b_j^l \right] \quad (1)$$

式中 i 为卷积行数， j 为卷积列数， l 为对应卷积层的层数， M_j 为输入的特征图， w 为该层卷积核权重， b 为该层的偏置项， f 为对应的激活函数，可以对上层卷积层提取得到的简单特征输出做非线性映射以增强模型对异 XOR 等非线性函数的学习能力。

卷积层输出的特征图会成为池化层的输入，通过对输入的特征图的池化可帮助实现其进一步的降维压缩，这一步骤又被称作特征图的下采样过程。池化对于整个网络的意义在于它可以避免过拟合现象的同时扩大网络模型的感受野，对输入特征图中的目标特征进行突出以保持数据的内在联系。池化方式通常分为最大池化和平均池化两种，本文选取最大池化作为主要的池化方法，通过将图像分为相同大小的子域并把子域中最大值作为输出值，可以加快网络收敛速度，其公式可以表示为：

$$X_{i,j}^l = \max_{m>0, s>n} \{ X_{i-s+m, j-s+n}^{l-1} \} \quad (2)$$

式中： s 表示池化模板的尺寸大小； m 和 n 表示相应方向的步长。

仅由卷积层和池化层构成的网络模型由于不具有非线性特性而无法训练非线性模型。为解决这一问题，需要在隐藏层中拓展一个非线性函数，通常称为激活函数。常用的激活函数包括 Sigmoid 函数、Tanh 函数、ReLU 函数和 ELU 函数。本文选取 ReLU 函数作为网络的激活函数，其公式可以表示为：

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

本文采用将二元交叉熵(binary cross entropy, BCE)同 DICE (dice coefficient)结合的损失函数其公式可以表示为：

$$\text{BCELoss}(P, GT) = - \sum_{i=1}^W \sum_{j=1}^H [gt_{ij} \cdot \log_2 p_{ij} + (1 - gt_{ij}) \cdot \log_2 (1 - p_{ij})] \quad (4)$$

$$\text{Dice}(P, GT) = (1 + w) \frac{\sum_{i=1}^N |PIGT|}{\sum_{i=1}^N |P| + |GT|} \quad (5)$$

$$\text{Loss} = \text{Dice}(P_i, GT_i) + \sum_{i=1}^N \text{BCELoss}(P_i, GT_i) \quad (6)$$

式中： P 表示模型的预测图像； GT 表示真值图像； N 表示批量大小； W 和 H 分别表示图像的宽度和高度； gt 表示 GT 真值图像中的一个像素值； p 表示预测图像 P 中的像素值； ω 表示权重，取值范围是 0.5 到 1。

2.1. 残差模块

Encoder-Decoder 网络中的编码过程实际是编码器对输入影像进行多次的卷积和池化操作的过程，经过下采样的高维特征图的道路特征会以降维的方式往深处传递。随着网络下采样深度的加深，会提取出到不同层次的特征信息。然而深度网络训练由于分辨率不断下降会产生梯度消失或是梯度爆炸的问题，深层网络冗余的权重参数会妨碍模型的收敛。何恺明等[16]针对堆叠更深的网络所导致的网络性能退化问题，构建了残差网络 ResNet (Residual neural network),很好的改进分割网络模型的寻优过程，并缩减模型收敛所要消耗的时间。残差块结构图如图 1 所示。一个残差块有两条前向传播路径，即 $F(x)$ 和 x ， $F(x)$ 路径为拟合路径， x 路径为恒等映射分支，即“捷径” (short connection)。梯度值在反向传播阶段的从残差块的一段输入，会按照 x 的路径返还，避开了由路径 $F(x)$ 返还而造成的 2 次更新计算导致的梯度爆炸的风险。

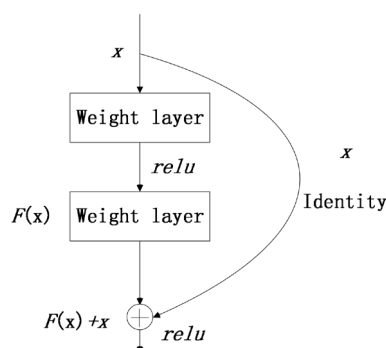


Figure 1. Residual block structure
图 1. 残差块结构

2.2. 网络框架图

本文所提出的道路提取方法主要框架如图 2 所示，首先本文神经网络模型主体采用了 Encoder-Decoder 的骨架结构。基于编码解码网络的一贯特点，本文网络同样可分为下采样编码和上采样解码两部分，在编码区接入由 4 个残差模块组成的深度残差网络 ResNet-34 作为基础编码器提取图像特征信息，而在解码区通过跳层连接，底层的局部特征信息和通过反卷积过后的高层语义信息得到融合，在此基础之上，本文在编码解码网络的中央区域加入膨胀卷积模块[17]，该模块的主要作用是保持参数个数不变的情况下增大卷积核的感受野，防止特征图在卷积过程中分辨率上的损失。

3. 实验与分析

3.1. 实验数据及环境配置

本文实验数据为马萨诸塞州道路数据集(Massachusetts Roads Dataset)。该数据集一共涵盖了 1171 张马萨诸塞州的卫星遥感影像且每张遥感影像都有对应的标签数据。数据合集集中的遥感影像空间分辨率为 1 m，尺寸大小为 1500 像素 × 1500 像素。标签数据为二值化影像，被标记的道路像素值为 1，背景像素值为 0。由于 GPU 的运算能力有限，需要将 1500 × 1500 的图像压缩至 512 × 512 大小。为保证训练后的效果，需要对原始的数据集进行数据增强，本文实验涉及到的数据增强方式包括水平及垂直镜像、旋转

等几何变换。最后得到一个由 3268 张遥感图像及所对应的道路标签图像组成的增强道路数据集，并将其按 2:8 的比例分割，选取 2652 张遥感图像及其标签数据组成训练集，选取剩下的 616 张遥感图像及对应标签组成验证集，为训练后的得到权值进行测试，数据增强结果如图 3 所示。

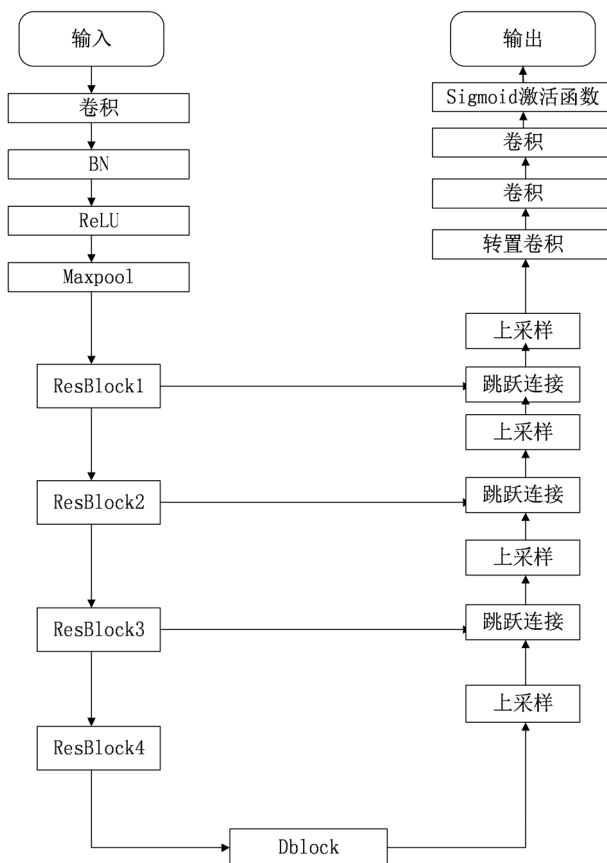


Figure 2. Improved Encoder-Decoder structure
图 2. 改进 Encoder-Decoder 结构框架图

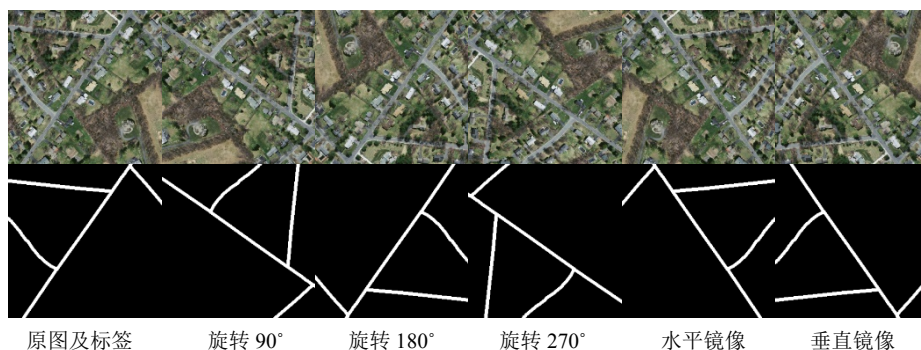


Figure 3. Augment training dataset
图 3. 增强训练数据集

本文实验环境为 Windows 操作系统，显卡型号为 NVIDIA RTX 2080Ti，显存容量 8 GB，采用 Pytorch 1.6.0 作为深度学习框架，Python 3.7 作为编辑环境，实验中超参数的设置为：基本学习率 base_lr 设置为 0.0001，迭代轮数 epoch 设置为 20，训练集和验证集的批大小均设置为 4。

3.2. 实验结果及分析

为有效评估本文提出的道路提取方法的有效性和准确性,并为该方法的量化对比分析提供可靠数据,本文采用二分类问题常用的混淆矩阵式的评估指标如精确率(precision)、召回率(recall)、F1-score 三项评价指标对提取结果进行统计。其定义分别为:

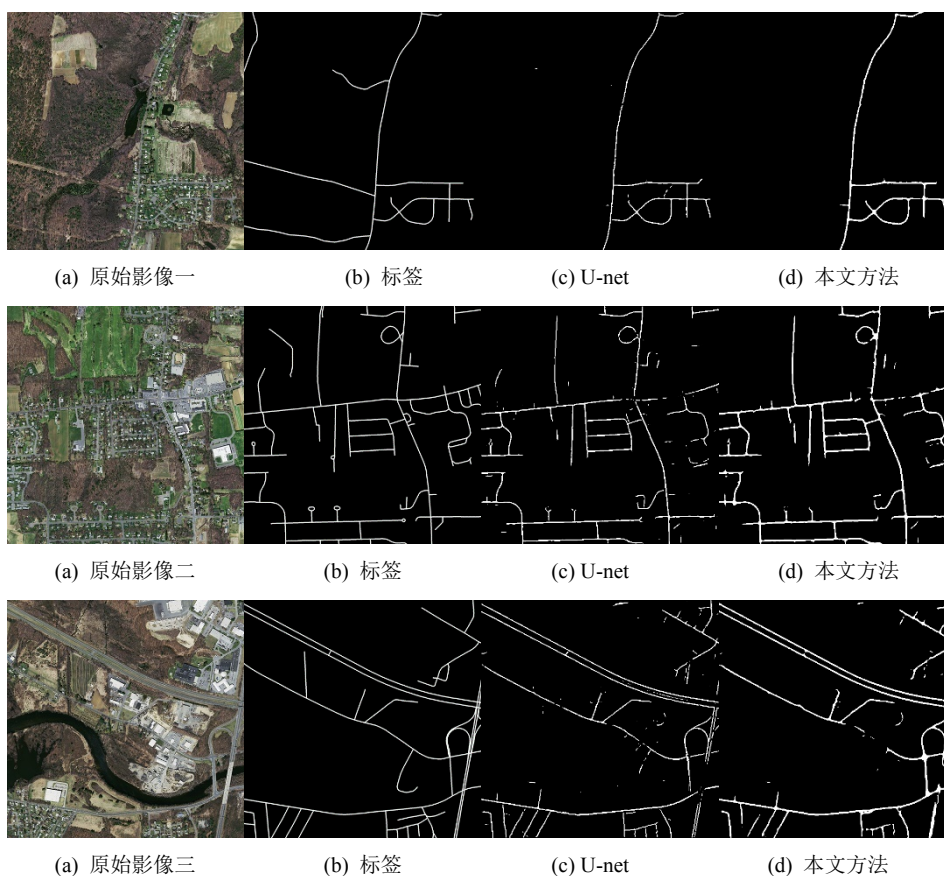
$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{F1} = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \quad (9)$$

式中: TP 表示预测为 1, 实际为 1, 即预测为道路且实际为道路; FP 表示预测为 1, 实际为 0, 即预测为道路但实际为背景; FN 表示预测为 0, 实际为 1, 即预测为背景但实际为道路。

本文选取语义分割领域中应用广泛的 U-net 模型作为参照,设计一组对比实验,并采用上文提到的评价指标进行量化比较。本文对比实验将在相同数据集,相同迭代次数等条件下进行验证测试。为了充分展现两种方法在完整性以及细节上的提取效果,特地选取背景各异的道路图像作为测试数据,其中包含了乡村道路、城区道路、环城高速路等多种道路场景。图 4 直观的展示了两种道路提取方法在不同场景下的道路提取效果。



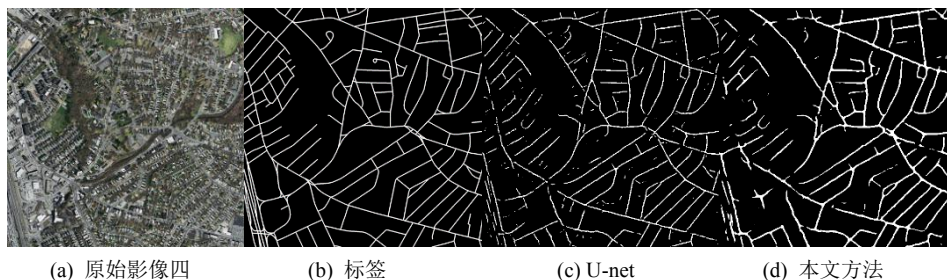


Figure 4. Comparison of road extraction results
图 4. 道路提取结果对比图

从图 4 可以看出, U-Net 网络模型虽然较完整的提取出了道路网络, 但提取结果仍存在细节上的不足, 如在背景复杂、房屋较为密集的城市区域容易出现斑点噪声, 提取的道路在某些路段上存在断裂的现象。相比 U-Net, 本文方法在道路网的整体提取上, 过滤掉了许多不必要的提取噪点, 并且在道路边缘信息的提取上也更加细致。

表 1 为两种训练后的模型对测试集进行预测后得到的定量评估数据。可以看到, 本文方法相比 U-net 在精确度上的改善虽然不是很大, 但是在召回率和 F1 值上还是取得了较为明显的提升, 召回率达到了 58%, F1 值达到了 71%。两项指标跟 U-net 相比都取得了 10% 的增长, 提升较为显著。实验结果表明, 本文方法在马萨诸塞州道路数据集上的表现要优于原始的 U-net 语义分割网络。

Table 1. Road extraction results of two different methods
表 1. 两种模型的道路提取结果对比

模型	Precision	Recall	F1-score
U-net	0.89	0.48	0.62
本文	0.91	0.58	0.71

4. 结束语

本文针对现有语义分割算法在进行高分辨率遥感影像道路信息提取时存在的完整性不足、漏提、错提、断裂等问题, 提出了一种改进的 Encoder-Decoder 网络模型。本方方法将残差网络模块、膨胀卷积模块引入到编解码结构的全卷积网络中, 从而提升道路提取的效果。残差网络模块可以防止加深网络层数情况下可能会出现的过拟合现象, 而膨胀卷积网络的加入可以在扩大特征点感受野的同时, 有效的防止特征图分辨率的损失并保护图像的边缘信息。实验结果表明, 本文所提出方法相较原始 U-Net 网络, 在遥感影像道路提取方面的效果得到一定的提升, 在背景复杂的城市场景以及道路分支结构复杂的场景下都表现出了较强的鲁棒性, 更加精确且完整的分割出道路与非道路。虽然本文方法一定程度上提升了 U-Net 网络在复杂场景下的道路提取效果, 但从提取结果以及召回率等指标来看, 本文方法在道路像素点的识别方面也依然存在优化的空间。

基金项目

贵州大学培育项目(贵大培育[2019] 26 号)。

参考文献

- [1] 吴亮, 胡云安. 遥感图像自动道路提取方法综述[J]. 自动化学报, 2010, 36(7): 912-922.

- <http://dx.chinadoi.cn/10.3724/SP.J.1004.2010.00912>
- [2] 周家香, 周安发, 陶超, 高陈强, 李静. 一种高分辨率遥感影像城区道路网提取方法[J]. 中南大学学报(自然科学版), 2013, 44(6): 2385-2391.
- [3] 张霞, 张春亢, 李红梅, 罗竹, 林健云. 结合笔画宽度变换与均值漂移的遥感影像道路提取[J]. 测绘科学技术学报, 2019, 36(3): 287-292.
- [4] 谭媛, 黄辉先, 徐建闽, 陈任. 基于改进 Sobel 算子的遥感图像道路边缘检测方法[J]. 国土资源遥感, 2016, 28(3): 7-11. <http://dx.chinadoi.cn/10.6046/gtzyyg.2016.03.02>
- [5] 张宗军, 杨风暴. 基于改进最大期望聚类的遥感影像道路提取算法[J]. 激光与光电子学进展, 2020, 57(6): 96-102. <http://dx.chinadoi.cn/10.3788/LOP57.061005>
- [6] Li, M., Stein, A., Bijker, W. and Zhan, Q.M. (2016) Region-Based Urban Road Extraction from VHR Satellite Images Using Binary Partition Tree. *International Journal of Applied Earth Observation and Geoinformation*, **44**, 217-225. <https://doi.org/10.1016/j.jag.2015.09.005>
- [7] Zang, Y., Wang, C., Yu, Y., Luo, L., Yang, K. and Li, J. (2017) Joint Enhancing Filtering for Road Network Extraction. *IEEE Transactions on Geoenvironment & Remote Sensing*, **55**, 1511-1525. <https://doi.org/10.1109/TGRS.2016.2626378>
- [8] 刘笑, 王光辉, 杨化超, 刘宇, 王耀. 全卷积神经网络遥感影像道路提取方法[J]. 遥感信息, 2018, 33(1): 69-75. <http://dx.chinadoi.cn/10.3969/j.issn.1000-3177.2018.01.011>
- [9] 郭正胜, 李参海, 王智敏. U 型卷积神经网络的 ZY-3 影像道路提取方法[J]. 测绘科学, 2020, 45(4): 51-57. <http://dx.chinadoi.cn/10.16251/j.cnki.1009-2307.2020.04.009>
- [10] 张永宏, 何静, 阚希, 夏广浩, 朱灵龙, 葛涛涛. 遥感图像道路提取方法综述[J]. 计算机工程与应用, 2018, 54(13): 1-10+51. <http://dx.chinadoi.cn/10.3778/j.issn.1002-8331.1804-0271>
- [11] Xu, R.D., Tao, Y.T., Lu, Z.Y. and Zhong Y.F. (2018) Attention-Mechanism-Containing Neural Networks for High-Resolution Remote Sensing Image Classification. *Remote Sensing*, **10**, 1602. <https://doi.org/10.3390/rs10101602>
- [12] Ma, W.P., Yang, Q.F., Wu, Y., Zhao, W. and Zhang, X.R. (2019) Double-Branch Multi-Attention Mechanism Network for Hyperspectral Image Classification. *Remote Sensing*, **11**, 1307. <https://doi.org/10.3390/rs11111307>
- [13] Zaimbashi, A. and Li, J. (2020) Tunable Adaptive Target Detection with Kernels in Colocated MIMO Radar. *IEEE Transactions on Signal Processing*, **68**, 1500-1514. <https://doi.org/10.1109/TSP.2020.2975371>
- [14] Corona, E., Alenya, G., Gabas, A. and Torras, C. (2018) Active Garment Recognition and Target Grasping Point Detection Using deep Learning. *Pattern Recognition*, **74**, 629-641. <https://doi.org/10.1016/j.patcog.2017.09.042>
- [15] Mnih, V. and Hinton, G.E. (2010) Learning to Detect Roads in High-Resolution Aerial Images. *European Conference on Computer Vision*, Heraklion, 5-11 September 2011, 210-223. https://doi.org/10.1007/978-3-642-15567-3_16
- [16] He, K.M., Zhang, X.Y., Ren, S.Q. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [17] Yu, F. and Koltun, V. (2015) Multi-Scale Context Aggregation by Dilated Convolutions. arXiv Preprint arXiv: 1511.07122. <https://arxiv.org/abs/1511.07122>