

基于深度学习的无人机遥感建筑物提取方法

吴艳兰^{1,3}, 王彪¹, 杨辉², 姚学东¹, 伍万祥¹

(1: 安徽大学资源与环境工程学院, 合肥 230601)

(2: 安徽大学物质科学与信息技术研究院, 合肥 230601)

(3: 安徽省地理信息智能技术工程研究中心, 合肥 230001)

摘要:建筑物信息提取在国土资源调查、地籍、不动产、城市规划、土地利用变化等方面具有重要意义。传统建筑物信息采集仍以人工立体采集为主, 且在实际生产过程中建筑物背景条件复杂, 需要耗费大量人力、物力, 迫切需要一种能够推进建筑物信息智能化、自动化提取的新方法、新技术。因此, 本文以“如何利用深度学习, 模拟人眼立体视觉中建筑物多特征表达形式”为核心, 综合利用飞马 D200 无人机遥感数据, 研究基于孪生网络的 DSM (数字表面模型) 与 DOM (数字正射影像) 的多特征融合方法, 设计建筑物多特征的特征抽取、分析能力, 通过密集注意力机制进一步增强建筑物的特征传递和累积整合特性, 实现顾及建筑物多层次特征的无人机遥感建筑物自动提取新方法。实验结果表明, 建筑物提取总精度达到 95.90%, F1-score 为 88.89%, 均交并比(MIOU)为 80.55%, 在提取精度和泛化性方面表现出明显优势, 有望进一步促进无人机遥感监测理论方法和实际应用价值的发展。

关键词: 无人机影像; 建筑物提取; 深度学习;

1 引言

无人机遥感监测由于其时效性强、效率高、灵活机动、成本较低等方面的优势, 目前已被广泛应用于大比例尺地形图测绘、城市空间数据基础设施建设、城市快速真三维建模等^[1], 无人机遥感建筑物提取的自动化直接影响着大规模的数据生产、广泛的行业应用、以及数据的充分应用。自无人机遥感监测技术提出以来, 国内外学者针对无人机遥感数据建筑物提取做了大量研究, 如基于几何规则加通过人工筛选提取建筑物^[4]、运用面向对象的影像分析思想^[5,6]等无监督方法, 以及支持向量机法^[7]、多核 SVM 学习法^[8]等监督方法。该类方法在建筑物提取方面取得了一定效果, 但难以顾及建筑物信息的多层次语义特征, 易造成建筑物精度和完整性损失。

随着深度学习在遥感人工智能领域的快速发展, 建筑物自动提取通过多层次特征提取^[9]、基于全卷积神经网络的建筑物提取^[10]和基于补丁的 CNN 模型^[11]等方法, 表现出强大的应用潜力。同时, U-Net^[12], Segnet^[13], RefineNet^[14]和 DeepLab^[15]等语义分析模型的出现, 进一步推动了建筑物多层次特征的深入挖掘。深度学习在自动学习有效的特征和解决同一类问题具有普适性方面有着巨大优势, 能够有效解决无人机遥感建筑物提取的问题^[16-18], 进而提升建筑物自动提取的自动化和智能化水平。

有基于此, 本文提出一种基于孪生神经网络模型, 并结合当前主流神经网络模型的混合模型来进行建筑物的提取。孪生神经网络用于解决在分类过程中出现样本数据数量不足的问题, 其与卷积神经网络相比, 最大的特点在于, 将原来使用单网络结构变为使用对称的网络结构, 并且两个对称的网络结构完全相同。本研究以孪生网络^[19]和 DenseNet^[20]为我们网络的整体结构, 并结合注意力机制, 构建混合模型, 深入挖掘无人机遥感建筑物的多层次特征和空间关系, 在分析目标语义特征的基础上, 研究无人机遥感建筑物自动提取方法。

2 孪生神经网络

2.1 密集连接网络

卷积神经网络(CNN)由一系列卷积层组成,每个卷积层之间都存在非线性转换函数 F_L 。一般非线性函数包括卷积操作、激活单元和池化操作。假设第 L 个卷积层的输入和输出分别为 X_{L-1} 和 X_L , 则相邻卷积层之间的信息转换可表示为:

$$X_L = F_L(X_{L-1}) \quad (1)$$

这种信息的简单转换会造成特征信息的丢失并降低了相邻卷积层间的信息反馈。在影像分割中, DenseNet 被广泛用来解决此问题。因其密集连接结构能够复用先前所有卷积层的特征信息,有效的减少训练中网络参数,使网络结构更易训练。在密集连接结构中,先前所有卷积层的特征映射都与后面层特征有关联,即第 L 个卷积层能够获取先前所有卷积层 ($X_0, X_1, X_2, \dots, X_{L-1}$) 的特征映射作为输入:

$$X_L = F_L([X_0, X_1, X_2, \dots, X_{L-1}]) \quad (2)$$

其中 $[X_0, X_1, X_2, \dots, X_{L-1}]$ 表示先前所有特征映射的融合。而非线性转换函数 F_L 包括批归一化层、激活函数和卷积层。此外, DenseNet 中构建了一个增长率 (growth rate) 降低密集连接结构中特征层的冗余,提高网络结构的效率。

由于密集连接结构的高效性,本文基于全卷积密集连接网络(FC-DenseNet)设计一个新的编码-解码结构(如图1所示)。为了充分利用建筑物的多特征信息,在网络的顶端本文设计一个孪生网络结构,采用两支浅网络分别抽取 DSM 和 DOM 特征,并在进入编码结构之前融合抽取的空间特征和光谱特征。编码结构中主要包括密集连接块和转换层,前者利用密集连接结构增强特征,后者通过下采样扩大特征感受野进一步提取建筑物的抽象特征。在解码结构中,反卷积的高级特征通过跳跃连接层融合编码结构中低级特征,并利用注意力机制指导低级特征获取不同类的空间细节信息。最终,网络输出建筑物分割结果。

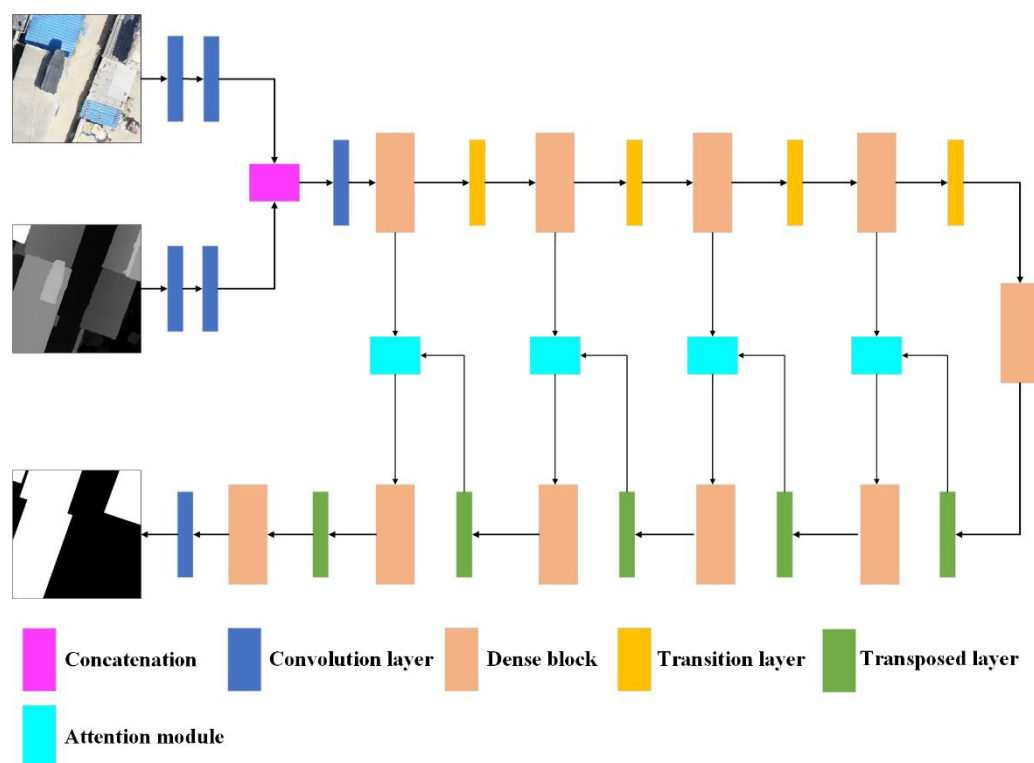


图1 本文模型结构图

2.2 浅层孪生结构

孪生神经网络是指两个或两个以上相同子网络结构的神经网络架构。孪生网络在训练中权重信息共享，需要更少的参数不容易出现过拟合。传统单支网络将不同特征信息共同输入网络，忽略了不同特征对分割目标的贡献大小，不能有效的利用特征信息。孪生网络中的每个子网络能够对不同特征单独抽取，充分利用不同特征信息，更有利于影像分割。本文设计一个浅层孪生网络结构用 DOM 和 DSM 作为网络的输入，其中一支网络提取建筑物的光谱信息同时另一支网络充分利用建筑的空间信息。如图 2 所示：两支网络具有相同的卷积层，经过两次卷积操作后，将两特征的输出结果合并，最后选用 5×5 的卷积核进行特征融合。

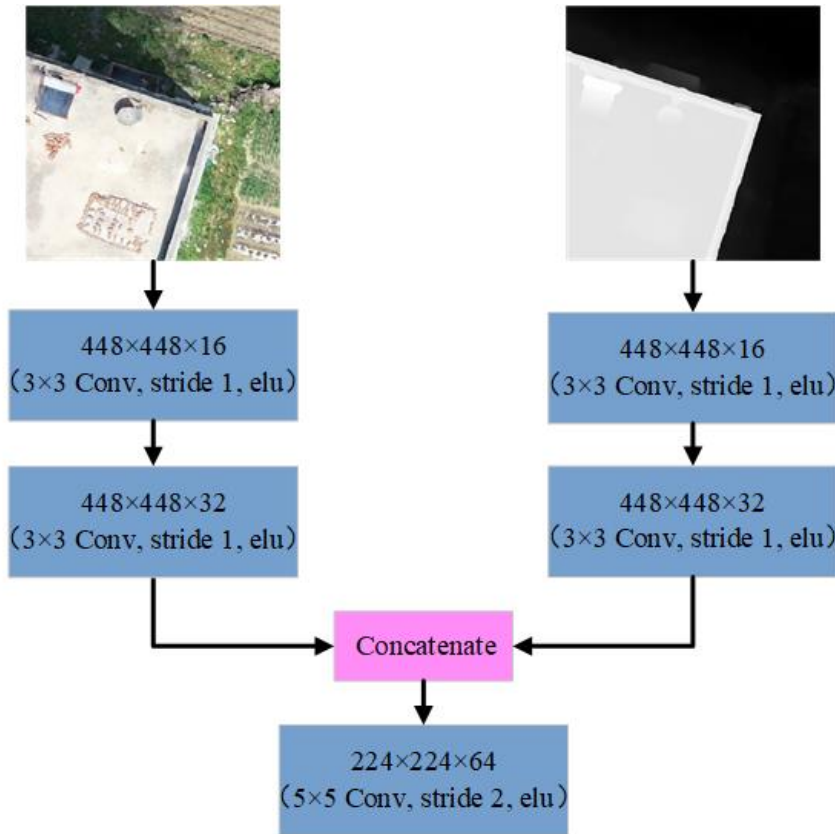


图 2 浅层孪生结构图

2.3 注意力机制

由于下采样过程中空间信息的丢失，许多卷积神经网络如 FCNs、U-Net、SegNet 等通过跳跃连接层融合下采样中的低级特征，弥补空间信息的损失。跳跃连接结构在一定程度上减少了信息的丢失，但是特征信息的简单融合不能够充分利用高级特征，同时由于低级特征包含丰富的空间信息不同于高级抽象特征，高级特征和低级特征简单的融合会降低分类的精度。为了进一步减少空间信息的丢失、增强特征信息和提高分类精度，注意力机制结构被应用到深度卷积神经网络中。注意力机制结构主要通过压缩高级特征作为特征向量增强特征信息和减少信息丢失。特征信息的增强主要分为两类：增强高级特征信息、利用高级特征指导低级特征信息。本文中的注意力机制主要利用高级特征指导低级特征，充分利用高级特征提高像素分类的一致性并降低空间信息的冗余避免影像的过分割。结构如图 3 示：

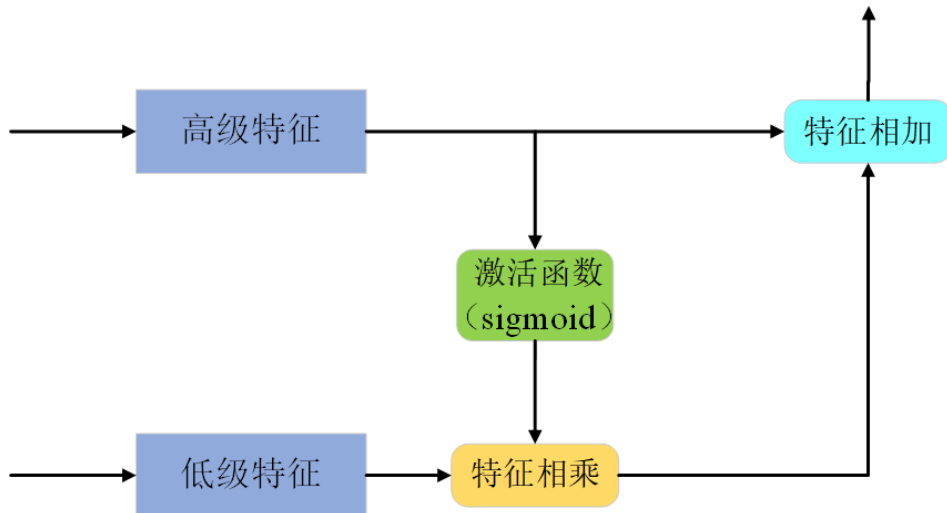


图3 注意力机制结构

3 实验

3.1 实验数据及模型训练

本文利用飞马 D200 无人机为空中飞行平台，搭载可见光相机，对安徽省某区域进行数据采集，获取无人机建筑 DOM 和 DSM 影像，分辨率为 2cm，其中 DOM 影像为蓝、绿、红三通道数据。所有影像均经过几何校正去除环境因素的影响。受计算机硬件影响，本文将样本数据大小设定为 448×448。为保证样本的充足，我们随机裁剪所有影像数据，得到 23667 个样本，其中训练样本 20034 个，测试样本 3633 个。此外，本文直接测试大小为 2560×2560 的未经预处理的无人机数据。



图4 飞马 D200 无人机

本文采用 Tensorflow 深度学习框架训练模型。其中模型训练总轮数为 60 轮，批处理个数为 12，每轮迭代次数为 1200 次，初始学习率为 0.01。为更好的训练模型，本文利用训练轮数自动调整学习率，即将训练轮数分为不同等级，当轮数达到相应等级时学习率自动缩小 10 倍，最终的学习率为 0.000001。由于 Adam 优化器计算高效、低内存需求，本文采用 Adam 优化器更新所有参数。此外，本实验中模型训练总时间约 50 个小时，每幅大小为 2560×2560 的测试数据测试的时间为 1.2 秒。

3.2 精度评价指标

为了定量评价模型的效果，本文采用像素精度 PA (pixel accuracy) ,F1-score 以及 IoU (Intersection over Union) 评估模型。其中像素精度 PA 表示预测值和真实值匹配像素占总

像素的比例。其值越大表示预测值和真实值间像素匹配度越高，精度越高。计算公式如下：

$$OA = \frac{\sum_{i=1}^k p_{ii}}{\sum_{i=1}^k \sum_{j=1}^k p_{ij}} \quad (3)$$

其中 p_{ii} 表示匹配的像素个数， p_{ij} 表示 i 类被分为 j 类的个数。匹配像素包括前景和背景像素，当前景较小时，该评价方式不能有效的表示模型分割的效果。F1-score 是评估二分类模型精确度的一种指标，它同时兼顾了分类模型的精确率和召回率，是两类精度的一种调和平均。其计算公式如下：

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (4)$$

$$recall = \frac{TP}{TP + FN}, precision = \frac{TP}{TP + FP} \quad (5)$$

其中 TP, FP 和 FN 分别表示真正 (true positive)、假正 (false positive)、假负 (false negative)。召回率指标表示正确预测像素个数和真实值像素数 (前景) 的比值，精度指标表示正确预测像素个数和预测值像素数 (前景) 的比值。这些数值可以通过基于像素的混淆矩阵计算。此外，IoU 评价指标表示预测值和真实值交集数与并集数的比值。计算公式如下：

$$IoU = \frac{TP}{FN + TP + FP} \quad (6)$$

3.3 实验结果与分析

模型训练完成后，本文测试三幅大小为 2560×2560 的无人机数据 (如图 5)。由图 5 可知测试结果中建筑物分割完整，提取效果较好，基本与真值数据一致。此外，为进一步量化本文方法的优越性，三幅测试数据的 PA、F1-score 和 IoU 列于表 1。对于所有测试数据而言，平均像素精度 (PA)、平均 F1-score 和平均 IoU 分别达到 95.90%，88.89% 和 80.55%。在测试数据中，数据 2 取得了较好的效果，其 PA、F1-score 和 IoU 均高于 90%，这与对应预测结果的效果保持一致。尽管数据 1 和数据 3 中的测试精度低于数据 2，数据 1、3 中仍然将建筑完整的分割出来。表明本文方法能够有效的、完整的提取无人机建筑物。

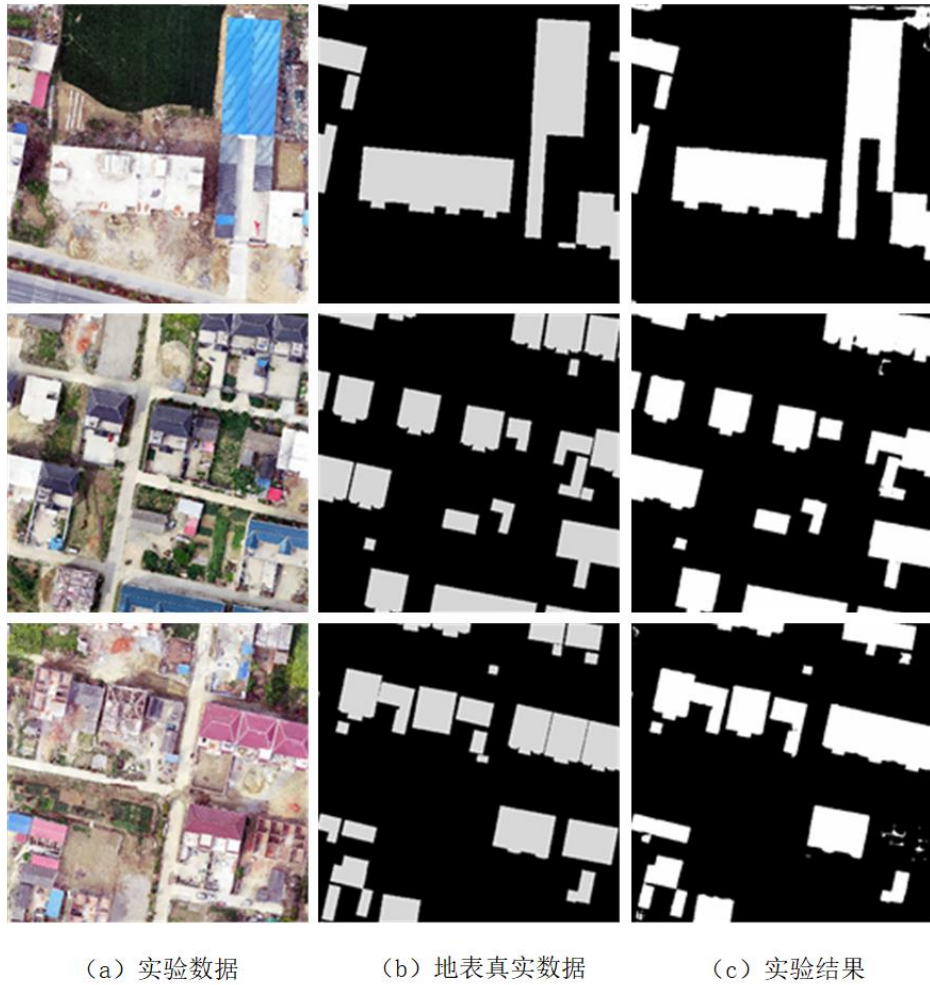


图 5 测试数据预测结果

表 1 测试数据精度评价表

测试数据	PA	F1-score	IoU
所有数据	95.90%	88.89%	80.55%
数据 1	92.75%	81.05%	68.14%
数据 2	98.24%	96.03%	92.36%
数据 3	96.71%	89.60%	81.16%

尽管本文方法能够有效的提取无人机建筑物,但模型识别的精度仍然受到一些因素影响。本文总结出影响模型精度因素为:(1) 异物同谱的干扰,(2) 树木遮蔽的干扰。

实际测试中,建筑物和道路均由相同材料构成,虽然为不同地物类型,但其光谱信息基本一致,因而容易造成误分的现象。图 6 为部分道路误分为建筑物。



(a) 实验案例 1

(b) 案例 1 检测结果

图 6 部分道路误分为建筑

建筑物的高度一般低于树木，大多数建筑物均受到树木遮蔽干扰。深度学习模型在识别过程中，仅仅将裸露出的房屋建筑分割提取，而被树木遮蔽部分无法被正确识别。而真值数据在标定中不受树木遮蔽干扰，因而造成预测值和真值间的差异，影响了识别精度。如图 7 所示：



(a) 实验案例 2

(b) 案例 2 检测结果

图 7 树木遮蔽为别正确识别

针对不同因素的干扰，本文的下一步工作能够有效的解决。构建样本中目标物的形态特征作为一新的特征支，通过高级特征指导形态支中的低级特征，能够降低噪音的同时有效的增强了形态支中的相关边缘特征信息，提高分割精度。受此启发，异物同谱和树木遮蔽情况也可以通过增加建筑物的几何特征来降低。道路的细长（包含宽度信息）特征不同于建筑物规则的几何特征，因而可有效区分；尽管树木遮蔽造成多数建筑不连续，但是建筑本身几何特征的连贯性在一定程度上降低了树木遮挡的干扰。

4 结论

本文针对建筑物自动化提取程度低、卫星遥感数据建筑物分辨率低不易提取等问题，提出一种基于深度学习的无人机建筑物提取方法。该方法通过浅层孪生结构融合建筑物的空间信息和光谱信息、利用密集连接结构充分利用特征信息同时使用注意力机制增强特征信息，有效的提高了建筑分割的精度。实验结果中 OA（像素精度）达到 95.90%，F1-score 达到 88.89%，IoU（交并比）达到 80.55%，表明本文方法在无人机建筑物提取中取得较好的效果，能够高效的提取建筑物。相比人工建筑物提取方法，本文方法在提高精度的同时更加快速化和自动化。

参考文献

- [1] 王俊博. 融合 LIDAR 点云与正射影像的城区建筑物细化提取研究[D]. 吉林大学, 2019.
- [2] 邓媛媛, 巫兆聪, 易俐娜. 面向对象的高分辨率影像农用地分类[J]. 国土资源遥感, 2010, 22(4):117-121.
- [3] 王旭东, 段福洲, 屈新原, 李丹, 余攀锋. 面向对象和 SVM 结合的无人机数据建筑物提取[J]. 国土资源遥感, 2017, 29(01):97-103.
- [4] Antonis K, and S. Hichem. A Stochastic Framework for the Identification of Building Rooftops Using a Single Remote Sensing Image[J]. IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, 2008, 46(1): 259-271.
- [5] 陶超, 谭毅华, 蔡华杰, 杜博, 田金文. 面向对象的高分辨率遥感影像城区建筑物分级提取方法[J]. 测绘学报, 2010, 39(1):39-45.
- [6] 刘正军, 张继贤. 基于分类与形态综合的高分辨率影像建筑物提取方法研究[J]. 测绘科学, 2007, 32(3):38-40.
- [7] 刘海飞, 常庆瑞, 李粉玲. 高分辨率影像城区建筑物提取研究[J]. 西北农林科技大学学报(自然科学版), 2013, 41(10):221-227+234.
- [8] Tao C, Tan Y, Yu J, et al. Urban area detection using multiple Kernel Learning and graph cut[C]. Geoscience & Remote Sensing Symposium. IEEE, 2012.
- [9] 吕凤华, 舒宁, 龚龔, 等. 利用多特征进行航空影像建筑物提取[J]. 武汉大学学报(信息科学版), 2017, 42(05):656-660.
- [10] 崔卫红, 熊宝玉, 张丽瑶. 多尺度全卷积神经网络建筑物提取[J]. 测绘学报, 2019, 48(05):597-608.
- [11] Alshehhi, Rasha, et al. Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks[J]. Isprs Journal of Photogrammetry & Remote Sensing. 2017, (130):139-149.
- [11] Audebert N, Le Saux B, Lefevre S. Semantic segmentation of earth observation data using multimodal and multi-scale deep networks. In Computer vision - accv 2016, pt i, Lai, S.H.; Lepetit, V.; Nishino, K.; Sato, Y, Eds. 2017; Vol. 10111, pp 180-196.
- [12] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation[C]. International Conference on Medical Image Computing & Computer - assisted Intervention, 2015.
- [13] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017:1-1.
- [14] Lin G, Milan A, Shen C, et al. RefineNet: Multi-Path Refinement Networks for High - Resolution Semantic Segmentation[J]. 2016.arXiv 2016,arXiv:1611.06612.
- [15] Chen L C, Papandreou G, Kokkinos I, et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs[J]. Computer Science, 2014.
- [16] Sherrah J. Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery. CoRR 2016, abs/1606.02585.
- [17] Nogueira K, Penatti, Otávio A. B, Santos J A D. Towards better exploiting convolutional neural networks for remote sensing scene classification[J]. Pattern Recognition, 2017, 61:539-556.
- [18] Li Da, Li Lin, Li Xiang. Classification of remote sensing images based on densely connected convolutional networks[C]. // IEEE Conference on Computer Vision & Pattern Recognition.

2017.

- [19] Chopra S, Hadsell R, Lecun Y. Learning a similarity metric discriminatively, with application to face verification[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005.
- [20] Huang G, Liu Z, Laurens V D M. Densely Connected Convolutional Networks[J]. 2016.