

# 基于卷积神经网络的遥感图像语义分割方法研究

朱双玲, 古丽娜孜·艾力木江\*, 苏金善, 乎西旦·居马洪, 帕孜来提·努尔买提

伊犁师范大学电子与信息工程学院, 新疆 伊宁

Email: \*dxfy\_glnz@ylnu.edu.cn

收稿日期: 2021年1月17日; 录用日期: 2021年2月12日; 发布日期: 2021年2月19日

## 摘要

高分辨率遥感图像大部分情况下包含相对来说较为复杂的语义信息以及容易混淆的目标, 对高分辨率遥感图像进行语义分割是一项很重要并且具有挑战性的任务。近几年来, 深度卷积神经网络(Deep Convolutional Neural Network, DCNN)为代表并结合条件随机场(Conditional Random Field, CRF)的算法在图像分割领域中有着杰出的表现。本文基于DeepLap V3+网络结构, 结合DCNN, 设计出了一种针对高分辨率遥感图像的语义分割网络, 仿真实验结果验证了该方法的有效性和鲁棒性。

## 关键词

卷积神经网络, 条件随机场, 图像语义分割, DeepLap V3+, 深度学习

# Semantic Segmentation of Remote Sensing Image Based on Convolutional Neural Network

Shuangling Zhu, Gulnaz·Alimjan\*, Jinshan Su, Hurxida·Jumahun, Pazilat·Nurmat

School of Electronics and Information Engineering, Yili Normal University, Yining Xinjiang

Email: \*dxfy\_glnz@ylnu.edu.cn

Received: Jan. 17<sup>th</sup>, 2021; accepted: Feb. 12<sup>th</sup>, 2021; published: Feb. 19<sup>th</sup>, 2021

## Abstract

In most cases, high-resolution remote sensing images contain relatively complex semantic infor-

\*通讯作者。

文章引用: 朱双玲, 古丽娜孜·艾力木江, 苏金善, 乎西旦·居马洪, 帕孜来提·努尔买提. 基于卷积神经网络的遥感图像语义分割方法研究[J]. 计算机科学与应用, 2021, 11(2): 356-369. DOI: 10.12677/csa.2021.112036

mation and easily confused targets. Semantic segmentation of high-resolution remote sensing images is a very important and challenging task. In recent years, deep convolutional neural network (DCNN) as a representative and combined with Conditional Random Field (CRF) algorithm has outstanding performance in the field of image segmentation. Based on the DeepLab V3+ network structure and combined with the DCNN, this paper designs a semantic segmentation network for high-resolution remote sensing images. The results of simulation experiments verify the effectiveness and robustness of the method.

## Keywords

Convolutional Neural Network, Conditional Random Field, Image Semantic Segmentation, DeepLab V3+, Deep Learning

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

近年来,人类探索宇宙空间的能力在我国航空航天技术的快速发展下取得了一定程度下的提高,鉴于遥感卫星能够拍摄高分辨率的地球遥感图像,结果,越来越多的通信卫星被送入环绕地球的轨道。高像素遥感图像的语义分割的日常任务是指为遥感图像的每个定义分配语义标记[1],简单来说,遥感图像用于对高线图像进行语义分割,例如山脉,森林,道路,大城市和海域,常规图像分割无法识别建筑物,道路,河流和其他商品,而语义分割则可以做到。遥感图像的语义分割已广泛应用于城市规划建设[2]和土地资源覆盖[3]等行业。

最早是 J. Shotton 提出图像的语义分割[4],定义了一个条件随机场,分类器选用提升决策树,来完成语义分割。Ren X 等明确提出了简单的线性迭代聚类[5],即超像素分割 SLIC (Simple Linear Iterative Clustering)优化算法,若超像素的分割不稳定,分类很可能会发生错误,Xiao J 等人提出了针对街景图像的多视图语义分割框架[2],这种方法可以用于街景语义理解的许多日常任务。

深度神经网络的语义分割是利用神经网络自动获取图像中每个像素的含义,紧接着对像素进行分类和识别,可以合理地节省人力。众所周知,遥感影像的情况相对来说更为复杂,不同类型的物体(例如河流,道路)通常在中间具有高度相似的外观,有时物体会被遮挡,在采集图像和传输图像的过程中也会有各种噪声的存在,噪声在图像里面被认为是无效的信息内容,因此输入图像也会对分割结果造成损害。此文在 U-Net 网络模型的基础上,采用了一种基于 DCNN 与 CRF 的算法,弥补图像在传输和获取过程中存在的噪声问题。

## 2. 语义分割的网络结构

语义分割的准确率往往不高,原因在于在高分辨率遥感图像中进行分割任务时,在空间分布上的类别数据易混淆且相邻交错分布,除此之外,一般的网络模型很难进行有效的特征表示,本文为了改善易混淆数据类别分割准确率低的问题,该设计方案的网络模型基于 DeepLab V3 结构进行了改进。Internet 的主要结构如图 1 所示。

### 2.1. DeepLab V3+算法

DeepLab V3+是 DeepLab V3 优化算法的改进。以 DeepLab V3+作为编码结构可以很好地管理分段结

果。根据简单合理的编解码器模块的改进,基础 Internet 选择了改进的 Xception 和 mobilenet 模块。DeepLab V1, DeepLab V2 [6] [7] [8]应用了空心卷积,并明确提出在空间维度上完成金字塔的空洞池化(Atrous SpatialPyramidPooling, ASPP)。DeepLab V3+改进了空间维度上的 ASPP,该模块级联了多个空洞卷积结构,见图 2。

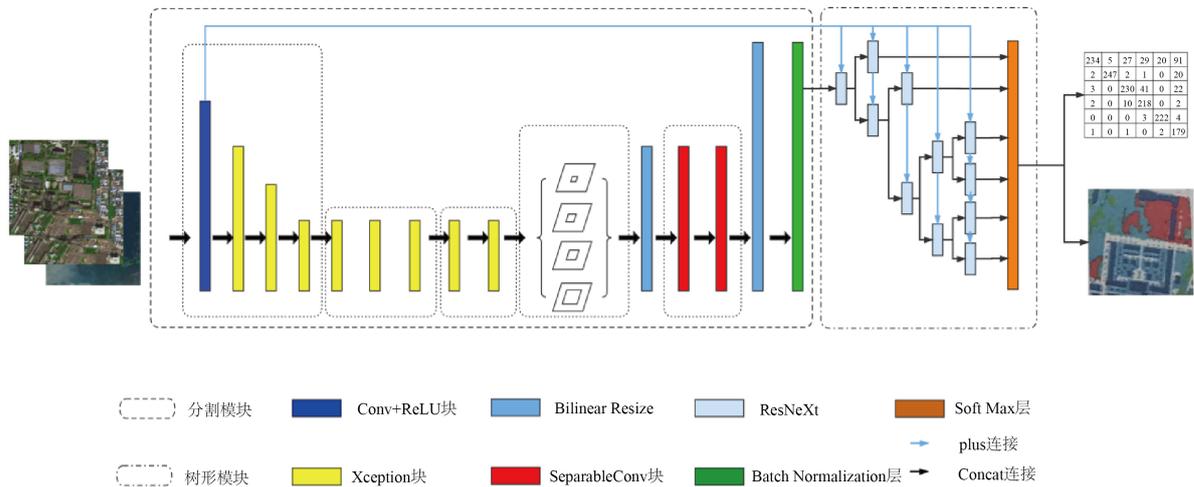


Figure 1. The overall schematic diagram of the network model  
图 1. 网络模型的整体示意图

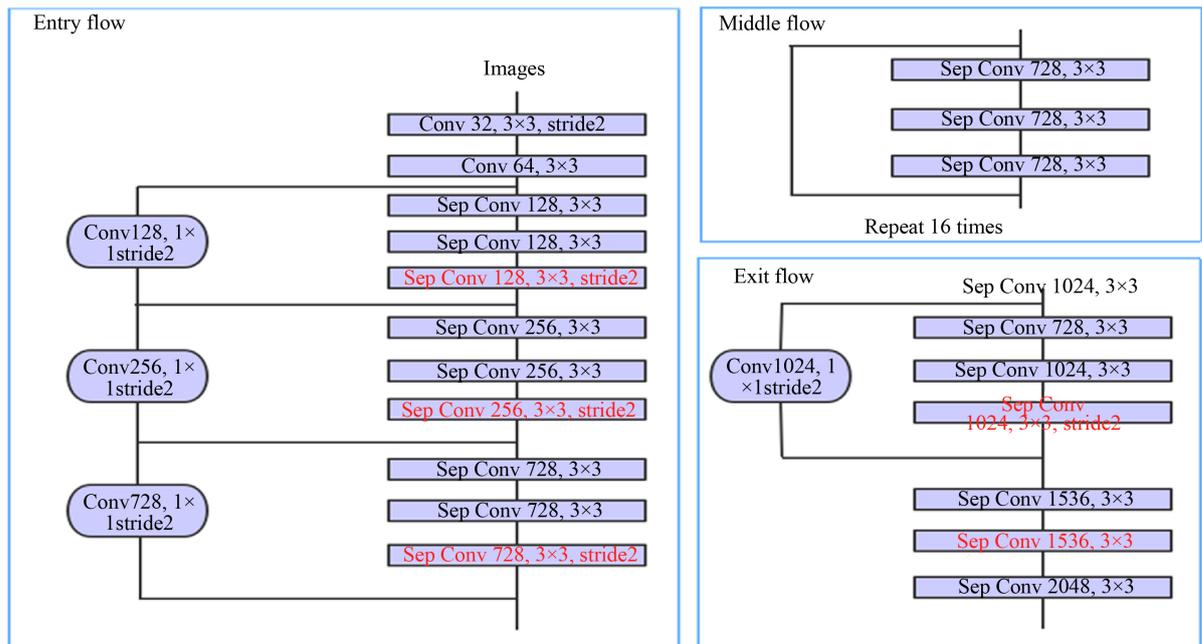


Figure 2. Improved Xception network  
图 2. 改进的 Xception 网络

DeepLab V3+的网络模型主要由两部分组成:编码器部分和解码器部分。解码器由双层 SeparableConv 块组成,其中几个 Xception 块包含残差。

## 2.2. DeepLab V3+的 Keras 实现

DeepLabV3+被认为是语义分割的新高峰，主要是因为这个模型的效果非常的好，DeepLabv3+主要在模型的架构上进行优化改进，为了融合多尺度信息，其引入了语义分割常用的 encoder-decoder 形式，在 encoder-decoder 体系结构中，根据空洞卷积平衡精度和耗时，介绍了可以任意控制编码器提取特征的分辨率，DeepLabV3 的关键构造见图 3。

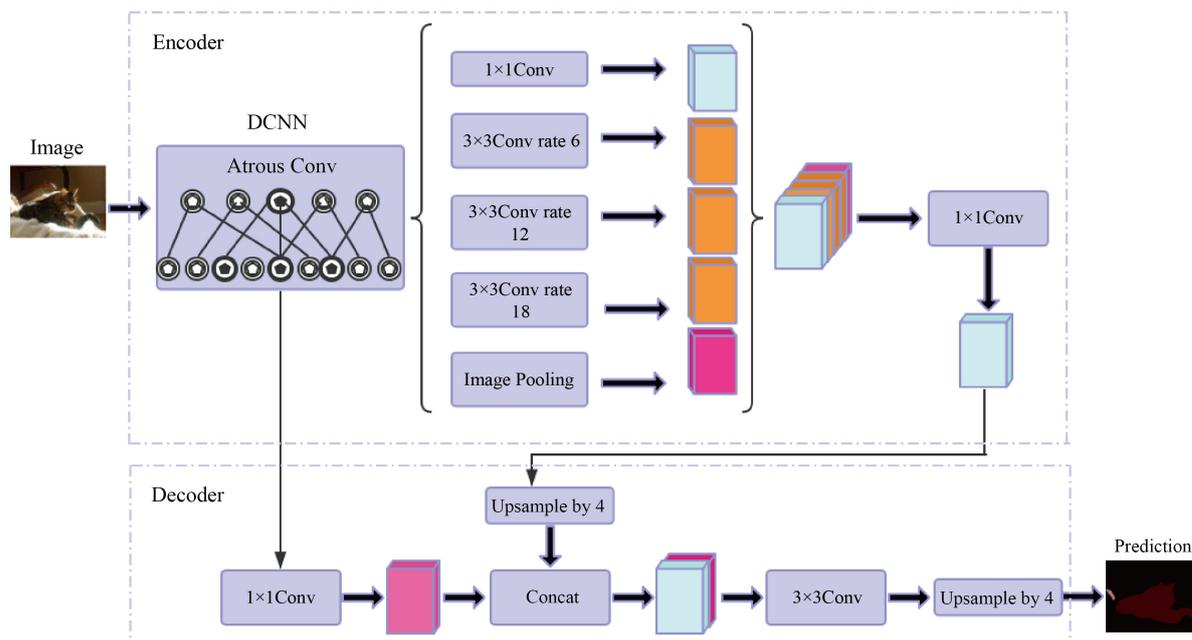


Figure 3. The main structure of DeepLab V3+  
图 3. DeepLab V3+主要结构图

其实，从 Deeplab V3+的网络可以看出，这个网络简单优美，没有那么多复杂的组合等等，四个空洞卷积块则是最为核心的内容，其中卷积核分别是 1、6、12、18。由这幅图我们可以发现，其实 deeplabV3+模型仍然是两个部分，一个部分是 Encoder，一个部分是 Decoder。空洞就是特征点提取的时候会跨像素，空洞卷积的目的其实也就是提取更有效的特征，所以它位于 Encoder 网络中用于特征提取。

编码器 Encoder 有两个核心点：1) 串行通信的 Atrous 卷积应用于骨干 DCNN 深度卷积神经网络。串行的意思就是一层又一层，普通的深度卷积神经网络的结构就是串行结构；2) 在图片经过主干 DCNN 深度卷积神经网络之后的结果分为两部分，一部分直接传入 Decoder，另一部分经过并行的 Atrous Convolution，分别用不同 rate 的 Atrous Convolution 进行特征提取，再进行合并，再进行  $1 \times 1$  卷积压缩特征。

Decoder 看起来就简单多了，其输入有两部分，一部分是 DCNN 的输出，一部分是 DCNN 输出经过并行空洞卷积后的结果。这两个结果经过一定的处理后 Concat 在一起，在 DeeplabV3 中，Upsample 的方式是双线性插值。

## 3. U-Net 模型

U-Net 能被划分为三个部分：第一部分是主干特征提取部分，我们可以利用主干部分获得一个又一个的特征层，U-Net 的主干特征提取部分与 VGG(Visual Geometry Group)相似，为卷积和最大池化的堆叠。利用主干特征提取部分我们可以获得五个初步有效特征层，在第二步中，我们会利用这五个有效特征层

可以进行特征融合；第二部分是加强特征提取部分，我们可以利用主干部分获取到的五个初步有效特征层进行上采样，并且进行特征融合，获得一个最终的，融合了所有特征的有效特征层；第三部分是预测部分，我们会利用最终获得的最后一个有效特征层对每一个特征点进行筛选，它的原理见图 4。

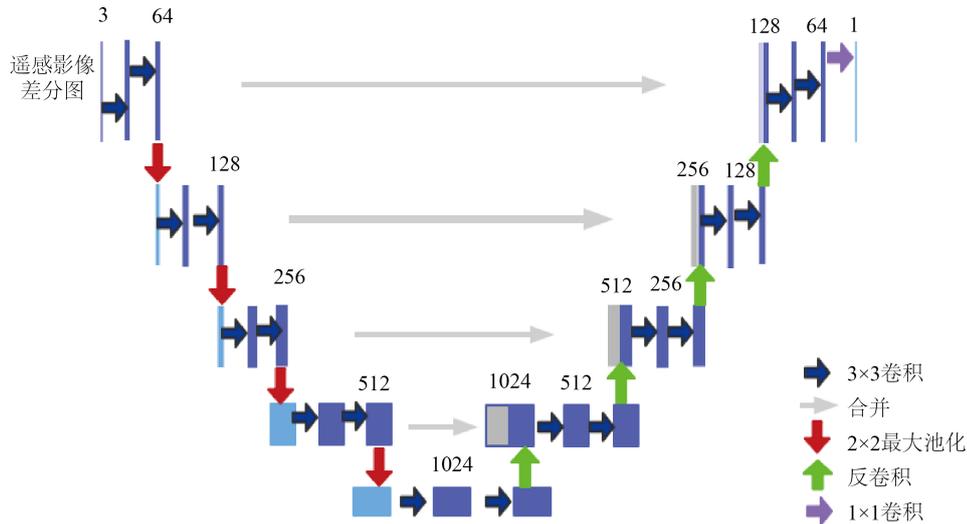


Figure 4. U-Net network model diagram

图 4. U-Net 网络模型图

根据这个模型的形状能观察到它结构上的优美组合，我们可以在这个基础上可进行大量的改进，比如说，特征提取块能够使用残差网络(resnet)进行替换，除了避免梯度消失外，它还可以学习和训练更多的特征，有利于提高精度。在特征融合层，大家普遍采用两种方式，第一种直接相加，即将编码层与解码层的特征直接相加，另外一种就是常用的 concat，关于这两种有什么优缺点，本文认为 concat 可以融合更多特征，其实就是以前的向量相加(vector stacking)。

#### 4. 条件随机场(Conditional Random Field, CRF)

传统的图像分析中，CRF 主要用来做平滑处理，又因为 short-range CRFs 可能会对我们恢复局部信息的目标起到反作用，所以使用全连接 CRF，考虑全局信息。全连接条件随机场的能量函数如公式(1)所示：

$$E(x) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, y_i) \quad (1)$$

这个能量函数计算的是所有像素点的量和能量，其中后半部分 pairwise 部分显示，特征数量是  $n(n-1)$  [ $n$  指像素个数]，所以全连接条件随机场也被称为 Dense CRF。能量函数的前半部分是一个一元函数，如公式(2)所示：

$$\theta_i(x_i) = -\log P(x_i) \quad (2)$$

还要注意的，观测点是你要标注的这些词本身和它们对应的特征，例如说词性是不是专有名词、语义角色是不是主语之类的。隐节点，是这些词的标签，比如是不是人名结尾，是不是地名的开头这样。这些隐节点(就是这些标签)，依次排开，相邻的节点中间有条边，跨节点没有边(线性链、二阶)。然后所有观测节点(特征)同时作用于所有这些隐节点(标签)。至于观测节点之间有没有依赖关系，这些已经不再重要了，因为它们已经被观测到了，是固定的。线性链的条件随机场跟线性链的隐马尔科夫模型一

样,一般推断用的都是维特比算法,该算法是一个简单的动态规划。

首先我们推断的目标是给定一个  $X$ , 找到使  $P(Y|X)$  最大的那个  $Y$ , 然后这个  $Z(X)$ , 一个  $X$  就对应一个  $Z$ , 所以  $X$  固定的话这个项是常量, 优化跟他没关系( $Y$  的取值不影响  $Z$ )。然后  $\exp$  也是单调递增的, 也不带它, 直接优化  $\exp$  里面。所以最后优化目标就变成了里面那个线性的和的形式, 就是对每个位置的每个特征加权求和。比如说两个状态的话, 匹配的概率是逐渐过渡到第一状态的概率加上从第一状态过渡到第二状态的概率, 这里概率是只  $\exp$  里面的加权和。那么这种关系下就可以用维特比了, 首先算出第一个状态取每个标签的概率, 然后再计算到第二个状态取每个标签得概率的最大值, 这个最大值是指从哪个状态标签转移到这个标签的概率最大, 值是多少, 并且记住这个转移(也就是上一个标签)。然后再计算第三个取哪个标签概率最大, 取最大的话上一个标签应该是哪个。以此类推。整条链计算完之后, 就知道最后一个词去哪个标签最可能, 以及去这个标签的话上一个状态的标签是什么、取上一个标签之前状态的标签又是什么。这里说的概率都是  $\exp$  里面的加权和, 因为两个概率相乘其实就对应着两个加权和相加, 其它部分都没有变。

## 5. 实验方法

### 5.1. 基于卷积神经网络的算法

要有效地对图像进行语义分割重点要先解决两个部分, 首先要训练网络模型, 我们要先确定欲分割图像的类型, 然后选择大量与欲分割图像类型一致的图像作为数据集, 与此同时制作相应的标签图, 这样就能够进行网络模拟的训练, 紧接着是对输入的图像进行预处理, 在得到分割结果之前, 将处理好的图像输入训练后的网络模型中即可。在语义分割任务中被广泛采取的基本是深度学习[9] [10]。2017年 Shelhamer 等人[11]建立了一种可以接受任意大小图像的全卷积神经网络(Fully Convolutional Networks, FCN), 诠释了 FCN 在空间密集型预测任务上的应用并且给出了他与之前其他网络之间的联系, 在图像语义分割领域具有重要的意义。图像分割的整体流程如图 5 所示。

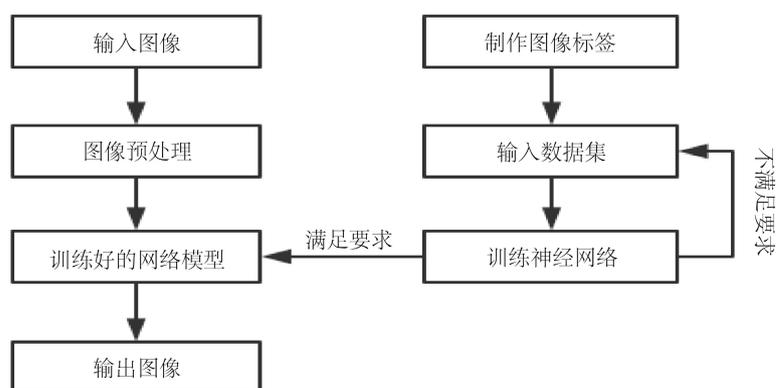


Figure 5. Flow chart of overall segmentation

图 5. 整体分割流程图

### 5.2. K-均值聚类法

由于 K-均值聚类法也被称为 K-means 算法, K-means 是一群无标记数据间的因为自我相似的聚拢, 所谓物以类聚人以群分, 相似的人们总是相互吸引在一起, 数据也是一样, 而 K-means 的目标为簇内密集而簇间稀疏。其原理如图 6 所示。

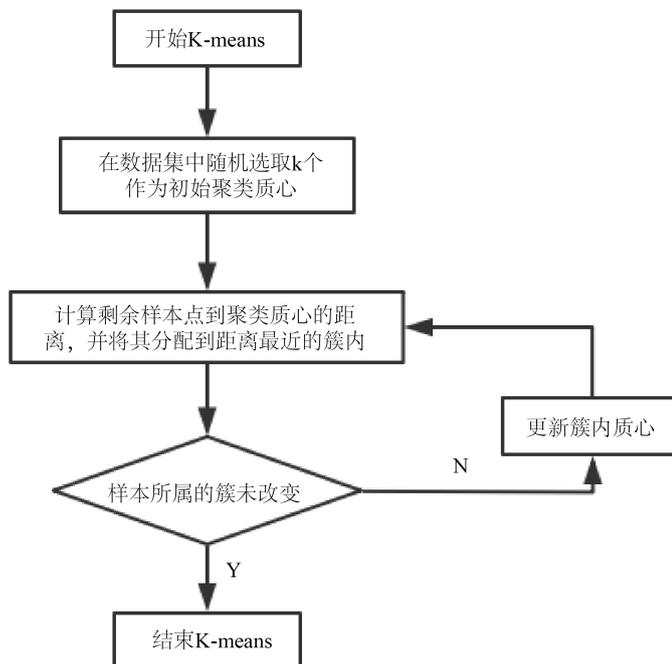


Figure 6. K-means algorithm flow chart  
图 6. K-均值算法流程图

K-均值优化算法简单, 快速且高效。当数据是聚集的, 特别是球形或块状的聚类, 再加上聚类与聚类之间的差距相当明显时, 聚类算法的实际效果是很好的, 但是它也有很多问题, 仅当均值时才可以使用 K-means 是定义的, 不适用于某些分类特性的数据信息。例如, 如果每种隐式类型的数据信息不平衡, 例如每种隐式类型的信息量更加严重, 对齐不正确, 或者每种隐式类型的标准差都不相同, 那么聚类算法的实际效果就不好, 最终结果与原始点的选择有关, 很容易陷入局部优化, 仍然需要科学研究。李全武等人[12]针对高速公路卡口抓拍的车辆图像, 研究了一种利用 K-means 聚类算法进行道路背景检测的方法, 并把占比最大的类别作为道路背景类别。

### 5.3. 自组织迭代数据分析法

必须事先人为地弄清 K 在 K 均值中的值, 并且不能在所有优化算法的整个过程中篡改 K 的值。一旦接触到高维和较多数据时, 通常很难准确确定 K 的大小。迭代自组织数据分析技术算法(ISODATA)优化改善了此问题, 其概念也是很直观: 如果属于某种类型的样本数量太少时, 应该去除该类型; 若是属于某种类别的样本数目太大并且分散程度大时, 此类别应当划分为两个子类别类型。基于经验的启发式迭代过程被称作是 ISODATA, 即算法整体运行中的许多步骤都来自于实验中得到的经验值来递推运行。王鹏[13]等人对 ISODATA 分类算法进行了研究, 在算法流程可并行性分析的基础上进行完善。赵泉华[14]等人将基于隐马尔可夫随机场(Hidden Markov Random Field, HMRF)方法加入 ISODATA 框架, 在分裂与合并操作后添加了优化操作。在此阶段, 大多数遥感影像解决方案软件工具(例如 ENVI 和 ERDAS)都具有这种分类方法。

ISODATA 优化算法基于 K 平均优化算法。它改进了“组合”和“分解”聚类算法结果的两个实际操作, 并设置了优化算法的主要参数。“合并”的实际操作: 当聚类算法的某个类别中的样本数太少或两种类型之间的距离太近时, 请合并。“分解”的实际操作: 当聚类算法得出某个类别时, 样本的某个特征类别内的标准偏差会很大, 并且这种发展会被分解。ISODATA 算法主体部分的描述如下表 1 所示:

**Table 1.** ISODATA algorithm**表 1.** ISODATA 算法

- 
- Step 1: 从集中数据信息中随机选择  $K_0$  个样本作为原始聚类算法管理中心  $C = \{C_1, C_2, \dots, C_{k_0}\}$ ;
- Step 2: 对于每个集中的数据信息模板, 将其计算到聚类算法管理中心  $C$ . 将其间距分类到由聚类算法管理中心以最小间距匹配的聚类中
- Step 3: 要区分每个类别中的元素数量是否少于该数量, 必须丢弃该类别, 设置  $K = K - 1$ , 并将该类别中的样本分配给其余类别中距离最小的类;
- Step 4: 对于每种类型  $c_i$ , 重新计算它的聚类中心  $c_i = \frac{1}{|c_i|} \sum_{x \in c_i} x$  (即属于该类的所有样本的质心)
- Step 5: 如果当前  $K \leq \frac{K_0}{2}$ , 表示当前类型的数量太少, 则销毁实际操作;
- Step 6: 如果当前  $K \geq 2K_0$ , 表示今天有太多类型, 请分解实际操作;
- Step 7: 如果迭代更新频率较高, 则停止, 否则返回第 2 步;
- 

## 6. 实验及其处理分析结果

### 6.1. 实验配置环境

Ubuntu 系统是本实验采用的主要系统, U-Net 网络实现工具为 Keras. 该模型的训练和测试是在 Tensorflow-GPU1.10 和 Ubuntu18.04 和 Cuda10.0 以及 Keras2.1.5 配置环境下开展的, 与此同时使用了 ENVI4.8 进行相关实验, 实验过程中对网络模型中的参数进行反复的迭代与调整, 调整到训练网络模型达到预期效果为止。

### 6.2. 评价指标

在此实验中, Jaccard 索引(也称为 Jaccard 相似系数)用于比较一组有限模式之间的相似性和差异。Jaccard 索引等于模式集和组合模式集的交集之比。可以说, 这体现了集中模式的相似性和分散性的可能性。Jaccard 系统标准值越大, 图的相似度值越高, 其表达式如公式(3)所示:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

### 6.3. 结果分析

为了测试深度卷积神经网络结合 CRF 算法模型的性能, 对于所有方法的内部参数设置成最优值, 对遥感影像实验数据产生专题分类效果图。

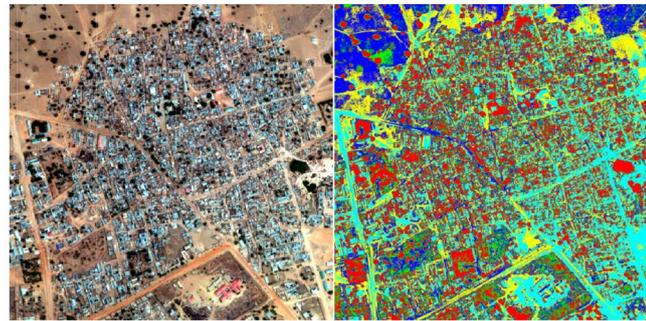
#### 6.3.1. 第一数据集及实验结果

本研究采用的第一数据集的训练数据集的图像具有 20 个通道 {RGB 波段(3 个通道) + A 波段(8 个通道) + M 波段(8 个通道) + P 波段(1 个通道)} 以及相应的对象标签。共有 8 种标记轮廓的重叠对象(wkt 数据类型), 包括建筑物、道路、轨道、树木、作物、死水、大型车辆、小型车辆等。分割效果见图 7。

图 7 是在基于 ENVI4.8 的实验环境下进行实验得到的分割效果图, 在此基础上使用 K-means 算法与 ISODATA 形成对比实验, 本文搭建的网络模型使用 U-Net 方法对实验数据图进行语义分割, 分割波段图如图 8 及预测图见图 9。

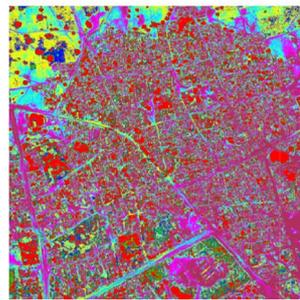
本实验训练集有 3345 个样本, 测试验证集 val 有 990 个样本, 根据实验结果能了解到网络对 building、crop、Man-made structures、Track、Trees 等类别分割效果较好, 对 road、Standing water、Vehicle Small 等目标分割效果不完整, 出现这种结果的主要原因可能是训练集中像素类别分布不均匀, 准确率高的类

别所占比例比较多，而准确率低的类别所占的比例小。也有可能是在遥感图像类别中占比比较小的在网络性能不够强的时候，对正确与错误这两种类别的理解不够充分，这将导致 Internet 收敛和分段特性降低。



(a) 实验数据图原图

(b) k-means 分割效果图



(c) ISODATA 分割效果图

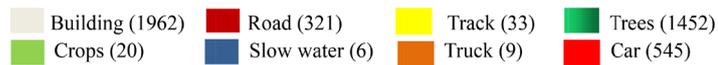


Figure 7. Classification effect diagram

图 7. 分类效果图

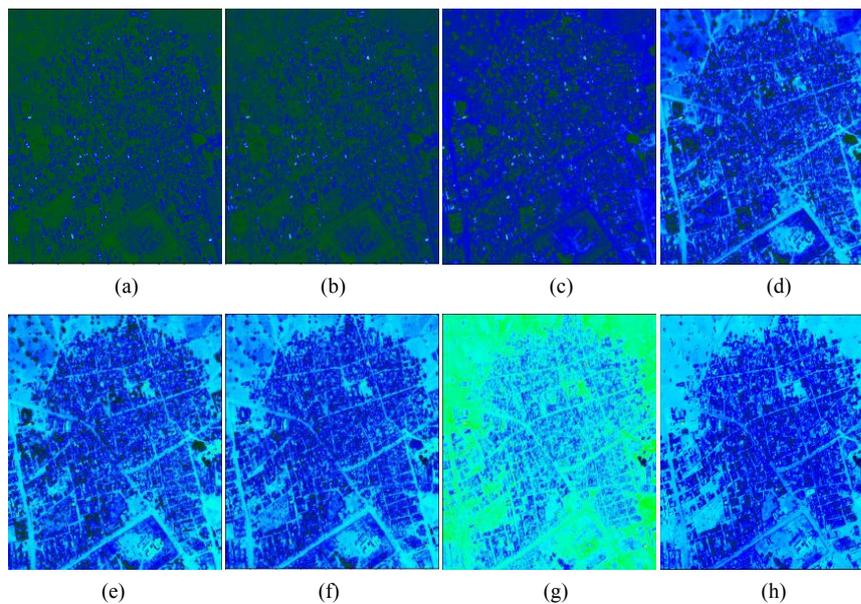
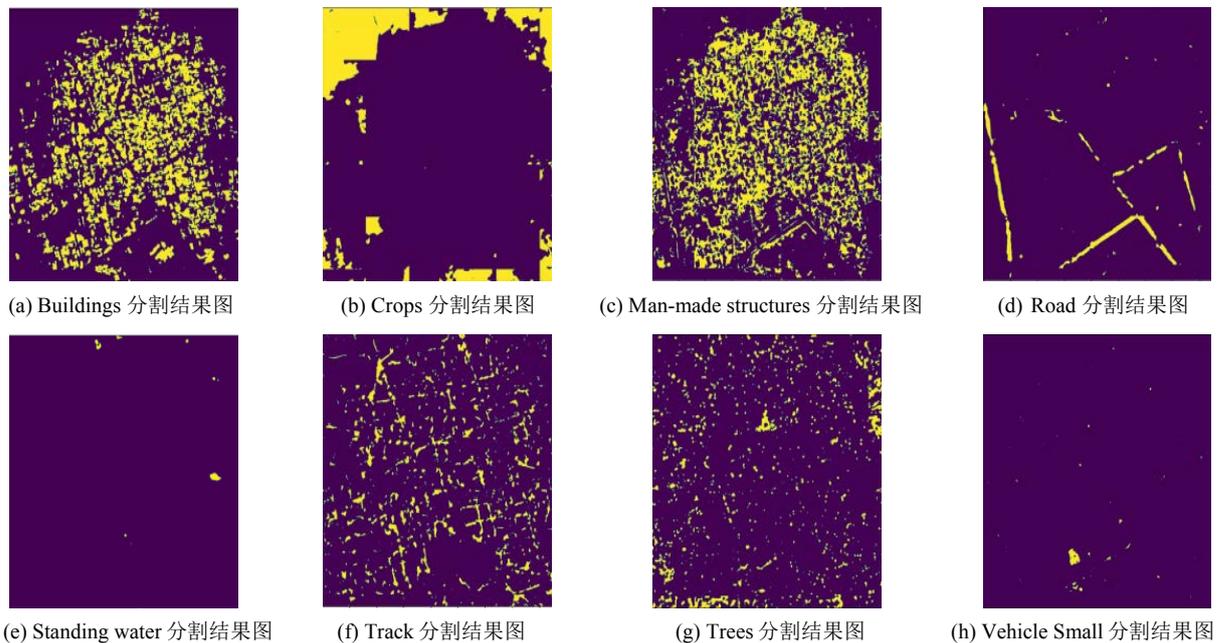


Figure 8. Image semantic segmentation band diagram

图 8. 图像语义分割波段图

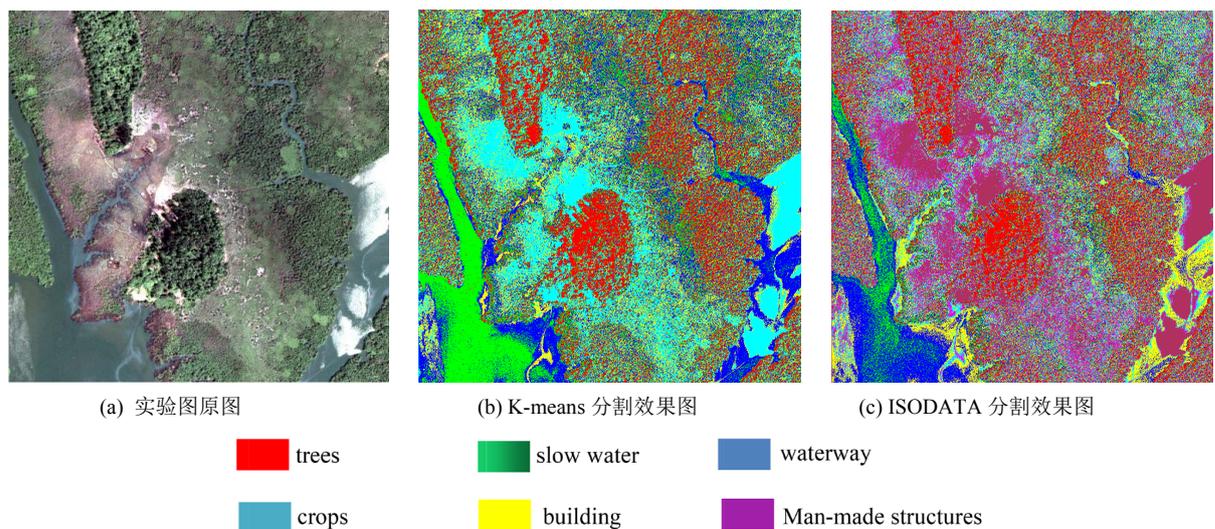


**Figure 9.** Image semantic segmentation prediction diagram

**图 9.** 图像语义分割预测图

### 6.3.2. 第二数据集及实验结果

本研究第二个实验数据集是由 823 行, 824 个, 3 个波段([BSQ]式)组成, 含 2, 034, 456 字节样本。图像分割效果如下图 10 所示。



**Figure 10.** Classification effect diagram

**图 10.** 分类效果图

本文算法在测试集上的分割效果如图 11 和图 12 所示, 表 2 是本文算法在分割时的网络训练参数。

根据实验可观察到网络对 Waterway、Man-made structures、Trees、Slow water 等类别分割效果较好, 对 road、Buildings、Crops、Track 等目标分割效果不明显, 出现这种结果的原因大概是实验数据图里相关像素类别分布不均匀, 也有可能是迭代次数不够导致, 提高这种分割方法的实际效果也是未来研究的

内容之一。

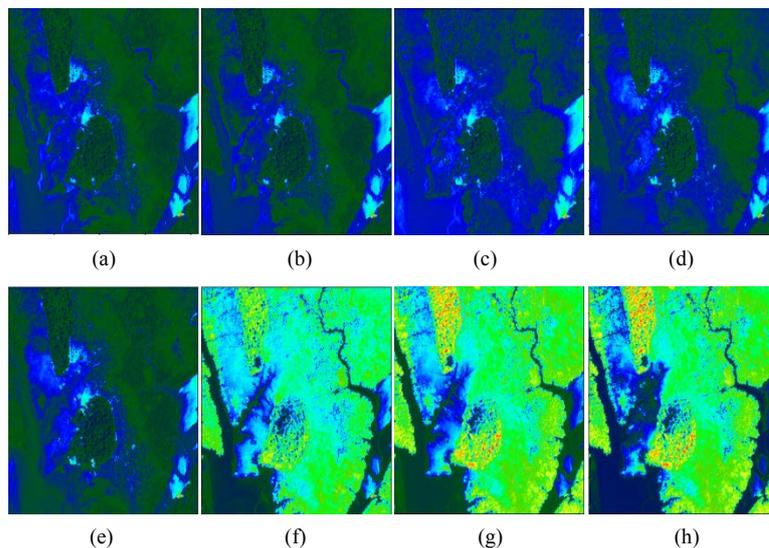


Figure 11. Image semantic segmentation band diagram

图 11. 图像语义分割波段图

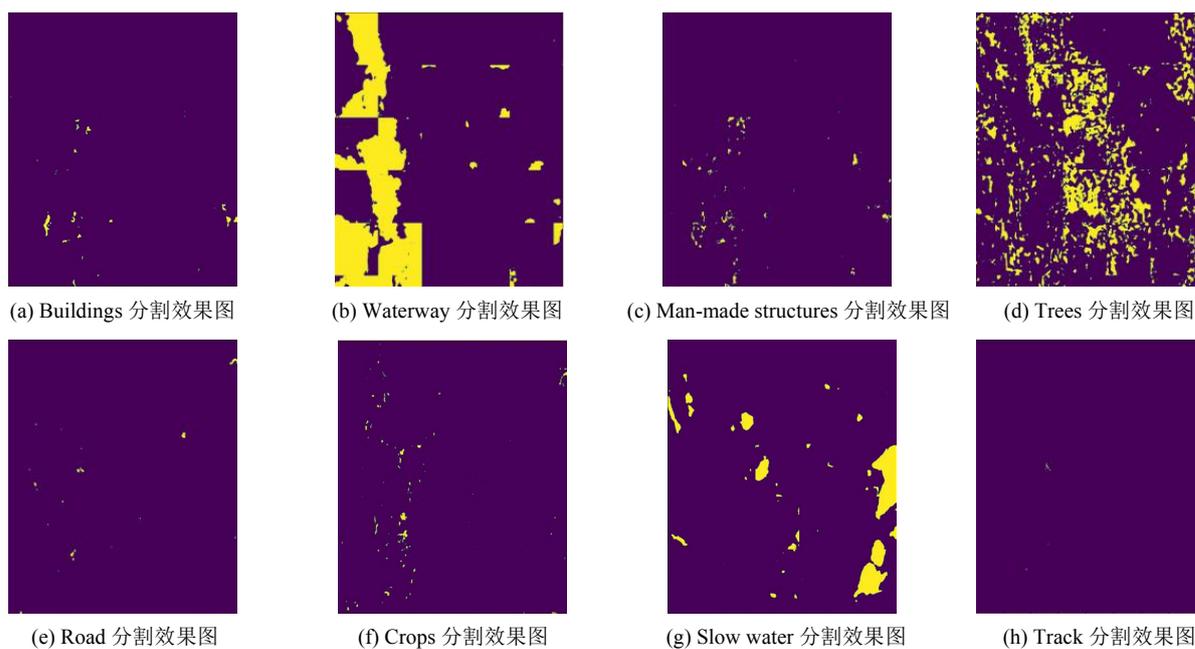


Figure 12. Image semantic segmentation prediction map

图 12. 图像语义分割预测图

Table 2. Network training parameters

表 2. 网络训练参数

参数	数值
Epoch	85
Batch size	64

图 13~15 是本文算法从第 1 轮至第 85 轮迭代后 loss 函数与 jaccard 系数的相关曲线图, 由图可知, loss 函数基本上一直处于下降状态, 神经网络在训练之前, 我们需要给其赋予一个初值, 但是如何选择这个初始值, 则要参考相关资料, 选择一个最合适的初始化方案。常用的初始化方案有全零初始化、随机正态分布初始化和随机均匀分布初始化等。合适的初始化方案很重要, 用对了, 事半功倍, 用不对, 模型训练状况不忍直视。loss 损失函数相当于模型拟合程度的一个评价指标, 这个指标的结果越小越好, 一个好的损失函数, 可以在神经网络优化时, 产生更好的模型参数。

神经网络的优化器一般选取 Adam, 网络训练的速率和学习率密切相关, 但它不是越高越佳, 当网络趋近于收敛时应该选择较小的学习率来保证找到更好的点, 为了挑选一个适宜的初始学习率, 我们需要调整学习率, 等到训练不了时, 再调低学习率, 然后再训练一段时间, 这时候基本上就完全收敛了, 一般学习率的调整是乘以或除以 10 的倍数, 不过现在也有一些自动调整学习率的方案了, 不过, 我们也要知道如何手动调整到合适的学习率。还可以将学习率设置为在整个训练过程中连续降低, 以防止学习过程中的梯度下降损害某个值范围内的往返波动。无论是培训还是测试, 深层的 CNN 互联网都是非常耗时的整个过程。在 Internet 上进行培训和测试大约需要花费几天的时间, 如果要获得满意的结果, 则必须连续调整参数。另外, 很有可能必须添加或删除 Internet 层的一部分以优化网络实体模型。因此, 如何调整 Internet 的主要参数是本文方法中的关键问题。

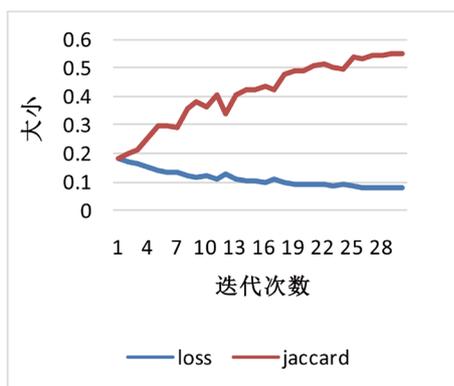


Figure 13. Iterations 1~30

图 13. 第 1~30 轮迭代

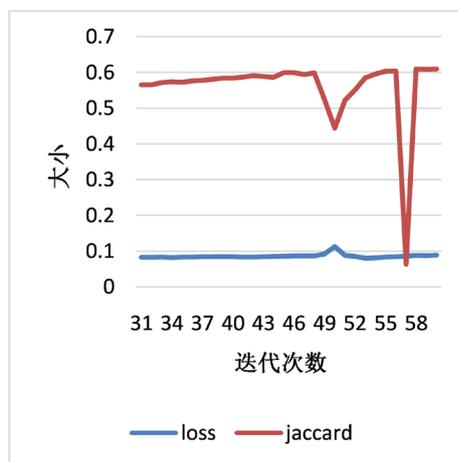


Figure 14. Iterations 31~60

图 14. 第 31~60 轮迭代

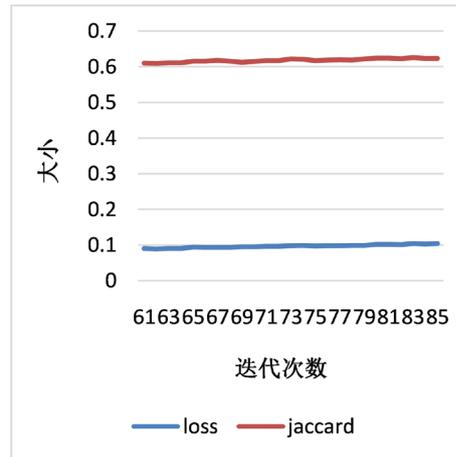


Figure 15. Iterations 61~85  
图 15. 第 61~85 轮迭代

## 7. 结语

本文的关键是使用 Pytorch 深度神经网络体系结构来完成 Internet 建设。经过 85 次迭代的更新训练，互联网的泛化能力得到了提高。图像语义分割是人工智能算法行业中的难题。传统的图像语义分割在技术上太逻辑化，难以理解，并且不使用大量图像数据来驱动语义分割的日常任务，无法完成大规模的图像解决方案，从而限制了其在特定应用中的市场推广。随着深层神经网络的基本理论和应用的发展趋势，深层 CNN 互联网继续得到清晰的提出。本文将 CNN 引入到遥感图像语义分割中，并根据结构卷积和神经网络完成了遥感图像的语义分割。可以通过测试结果看到，对于遥感图像中某些及其细微的区域分割结果不太理想，在今后的研究中，也应该对此进行进一步的探索。

## 基金项目

新疆自然科学基金项目(2019D01C337); 国家自然科学基金项目(61663045); 校级重点课题(2020YSZD01)。

## 参考文献

- [1] Csurka, G. and Perronnin, F. (2011) An Efficient Approach to Semantic Segmentation. *International Journal of Computer Vision*, **95**, 198-212. <https://doi.org/10.1007/s11263-010-0344-8>
- [2] Xiao, J.X. and Quan, L. (2009) Multiple View Semantic Segmentation for Street View Images. *International Conference on Computer Vision*, Kyoto, 29 September-2 October 2009, 686-693.
- [3] 许新征, 丁世飞, 史忠植, 等. 图像分割的新理论和新方法[J]. 电子学报, 2010, 38(2): 76-82.
- [4] Shotton, J., Winn, J. and Bother, C. (2006) Textonboost: Joint Appearance, Shape and Context Modeling for Multiclass Object Recognition and Segmentation. *ECCV*, **1**, 1-15. [https://doi.org/10.1007/11744023\\_1](https://doi.org/10.1007/11744023_1)
- [5] Ren, X. and Malik, J. (2003) Learning a Classification Model for Segmentation. *International Conference on Computer Vision*, **1**, 10-17. <https://doi.org/10.1109/ICCV.2003.1238308>
- [6] Chen, L.C., Papandreou, G., Kokkinos, I., et al. (2016) Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **40**, 834-848. <https://doi.org/10.1109/TPAMI.2017.2699184>
- [7] Chen, L.C., Papandreou, G., Schroff, F., et al. (2017) Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv:1706.05587.
- [8] Noh, H., Hong, S. and Han, B. (2015) Learning Deconvolution Network for Semantic Segmentation. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, 18 February 2015, 1520-1528. <https://doi.org/10.1109/ICCV.2015.178>
- [9] Zheng, S., Jayasumana, S., Romera-Paredes, B., et al. (2015) Conditional Random Fields as Recurrent Neural Net-

---

works. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, 1529-1537.  
<https://doi.org/10.1109/ICCV.2015.179>

- [10] Ren, X., Bo, L. and Fox, D. (2012) Rgb-(d) Scene Labeling: Features and Algorithms. *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, 2759-2766.
- [11] Shelhamer, E., Long, J. and Darrell, T. (2017) Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **39**, 640. <https://doi.org/10.1109/TPAMI.2016.2572683>
- [12] 李全武. 基于 K-means 的道路背景检测算法研究[J]. 信息与电脑(理论版), 2020, 32(15): 48-50.
- [13] 王鹏. 遥感影像的 ISODATA 分类算法的并行化研究[D]: [硕士学位论文]. 开封: 河南大学, 2016.
- [14] 赵泉华, 赵雪梅, 李玉. 结合 HMRF 模型的模糊 ISODATA 高分辨率遥感图像分割[J]. 信号处理, 2016, 32(2): 157-166.