

多尺度 SLIC-GMRF 与 FCNSVM 联合的高分影像建筑物提取

井然^{1,2,3}, 宫兆宁^{1,2,3}, 朱文定^{1,2,3}, 关鸿亮^{1,2,3}, 赵文吉^{1,2,3}, 张涛⁴

1. 首都师范大学 资源环境与旅游学院, 北京 100048;
2. 三维信息获取与应用教育部重点实验室, 北京 100048;
3. 资源环境与地理信息系统北京市重点实验室, 北京 100048;
4. 国家测绘地理信息局卫星测绘应用中心, 北京 100048

摘要: 遥感影像建筑物提取具有重要的应用价值。然而, 高分辨率遥感影像中细节信息繁多、特征复杂, 增加了建筑物提取难度。针对这一问题, 本文提出一种基于多尺度 SLIC-GMRF 和 FCNSVM 的建筑物提取方法, 一定程度上提高了高分辨率遥感影像建筑物提取能力。首先, 利用多尺度 SLIC-GMRF 分割算法确定初始建筑物区域, 然后, 充分利用 FCN 神经网络在语义分割中的优势抽取建筑物特征, 最后, 结合提取出的建筑物特征训练 SVM 分类器细化建筑物提取结果, 通过 3 种控制实验, 两种对比方法得出以下结论: SLIC 分割算法影响初始分割结果; SVM 分类器影响建筑物细部提取; FCN 特征影响 SVM 分类器性能。对于特征清晰、遮挡干扰较少的研究区, 本文方法能够较好提取影像中的建筑物, 查准率、查全率、质量指标均优于对比方法, 对建筑物复杂分布的研究区同样能够取得较好的提取效果。

关键词: 遥感, 建筑物提取, 影像分割, FCN 神经网络, 支持向量机, 高分辨率遥感影像

引用格式: 井然, 宫兆宁, 朱文定, 关鸿亮, 赵文吉, 张涛. 2020. 多尺度 SLIC-GMRF 与 FCNSVM 联合的高分影像建筑物提取. 遥感学报, 24(1): 11-26

Jing R, Gong Z N, Zhu W D, Guan H L, Zhao W J and Zhang T. 2020. Extraction of buildings from remote sensing imagery based on multi-scale SLIC-GMRF and FCNSVM. *Journal of Remote Sensing(Chinese)*, 24(1): 11-26[DOI: 10.11834/jrs.20208221]

1 引言

建筑物的精确识别在城市规划、人口估算、景观分析等应用中具有重要的意义。近年来, 随着一系列高分辨率卫星的发射, 卫星影像的空间分辨率及光谱分辨率得到了极大提高, 使得建筑物的精确识别定位成为可能。然而在高分影像建筑物提取中, 基于像元的提取方法由于建筑物斑块像元表现出的光谱异质性降低了可区分度, 对提取造成不利影响, 因此影像分析的基本单元由面向像元向基于对象的方法改变。超像素分割是将多个同质像素合并为影像对象的方法, 最早由 Ren 和 Malik(2003)提出。基于超像素分割的方法主

要有基于图的方法(Felzenszwalb 和 Huttenlocher, 2004)、Ncut 方法(Shi 和 Malik, 2000)、Superpixel lattice 方法(Moore 等, 2008)、Turbopixels 方法(Levinshtein 等, 2009)以及 SLIC 方法(Achanta 等, 2010), 其中 SLIC 影像分割方法与其他方法相比实现较为简便, 分割效率高(Achanta 等, 2012)。然而由于遥感影像具有复杂的场景特征, 传统分割方法分割精度较低, 难以直接应用于遥感影像中。基于概率图模型的影像分类方法能够结合邻近像素间的语义信息, 因此在建筑物提取中得到了广泛的应用(张宏毅等, 2013)。Grinias 等(2016)以城市边缘区域为研究区, 提出基于马尔可夫随机场语义模型的建筑物分割算法。Liasis 和 Stavrou

收稿日期: 2018-05-16; 预印本: 2018-09-28

基金项目: 国家高分专项高分遥感测绘应用示范系统项目(编号: AH1601-8)

第一作者简介: 井然, 1991年生, 男, 博士研究生, 研究方向为遥感技术及地学应用。E-mail: 15911157479@163.com

通信作者简介: 宫兆宁, 1976年生, 女, 副教授, 研究方向为遥感技术及地学分析应用。E-mail: gongzhn@163.com

(2016)基于建筑物的先验知识提出改进的主动轮廓模型,将影像由RGB转换至HSV色彩空间,提高了模型的建筑物提取能力。Benedek等(2015)利用基于多层MRF模型对高空间分辨率遥感影像进行变化检测,并对多种模型进行对比实验。然而传统基于概率图模型的提取方法存在对先验知识依赖度高,局部不连接的情况,提取地物对象周边存在误分及孤立的建筑物类型。近年来,深度学习被广泛应用于自然语言处理(Collobert和Weston, 2008; Hinton等, 2012; Socher等, 2011)以及计算机视觉等领域(He等, 2016; Krizhevsky等, 2012; Sermanet等, 2013),通过构建深度神经网络多层次结构,逐层提取影像的低层次特征到高层次特征,因此适用于遥感影像复杂地物信息提取。Sharma等(2017)利用仅包含卷积层的卷积神经网络对分块的中分辨率Landsat 8遥感影像中的建筑物及其他类型地物进行分类,相比于基于像素的神经网络分类,基于像素的卷积神经网络分类及分块的神经网络分类总体精度分别提升了24.36%, 24.23%及11.52%。Gong等(2017)利用卷积神经网络对SAR影像中的建筑物进行三元变化检测。Volpi和Tuia(2017)利用3种卷积网络结构(斑块尺

度,小斑块尺度以及训练的上采样全尺度)对高分辨率遥感影像中的建筑物进行提取。然而传统基于神经网络的建筑物提取存在以下缺陷,由于神经网络隐层会对影像进行降采样及升采样等操作,使得建筑物的细节信息被合并,从而导致边缘信息缺失,同时其作为分类器的性能存在缺陷。

针对高分影像建筑物提取相关研究存在的不足,本文综合SLIC影像分割执行效率高的优势,与充分应用上下文语义信息的GMRF(Gauss-MRF)模型相结合,提出多尺度SLIC-GMRF建筑物初始分割算法。并利用全卷积网络FCN(Fully Convolutional Networks),将其微调后的权重系数,结合支持向量机(SVM)设计FCNSVM分类器。利用多尺度SLIC-GMRF算法对研究区影像进行初始分割聚类,得到建筑物的初始分类结果,然后提取FCN网络特征对SVM分类器进行训练,对建筑物初始分类结果进行二次分类,从而增强高分影像建筑物提取能力。

2 研究方法

本文方法技术流程见图1:

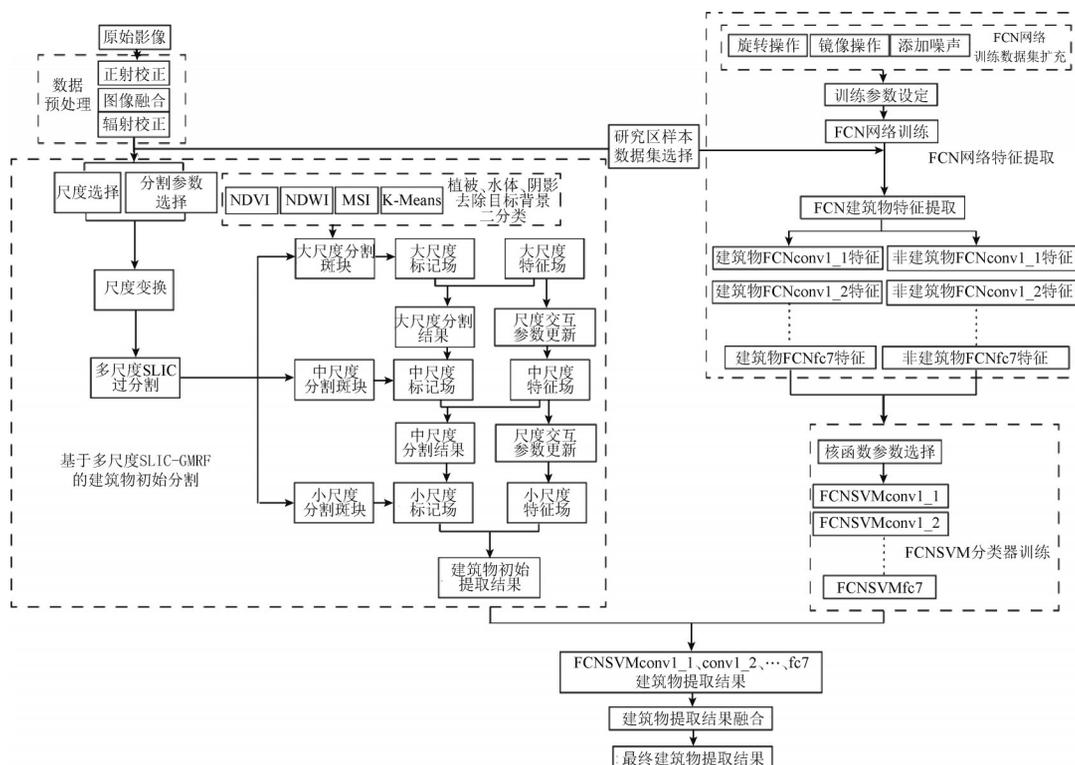


图1 技术路线

Fig.1 Technical flowchart

技术流程主要包括3个部分: (1)对原始影像进行预处理后, 首先对其进行基于多尺度 SLIC-GMRF 的建筑物初始分割; (2)FCN 网络特征提取及 FCNSVM 分类器训练; (3)结合初始分割结果及 FCNSVM 分类器的建筑物提取。

2.1 基于多尺度 SLIC-GMRF 的建筑物初始分割

基于结构的马尔可夫模型在图像分割中应用广泛(Besag, 1986; Bouman 和 Liu, 1991; Geman 和 Geman, 1984), 根据马尔可夫随机场的定义及其特性, 影像中像元对某个像元产生的影响等价于该像元的邻域对其影响(Li, 1994)。将建筑物斑块分割问题置于贝叶斯理论框架下, 计算公式为

$$P(\eta|y) \propto P(y|\eta)P(\eta) \quad (1)$$

式中, y 为特征场, η 为标记场, $P(\eta|y)$ 为给定特征场时标记场的后验概率分布, $P(y|\eta)$ 为特征场 y 的条件概率分布, $P(\eta)$ 为标记场的先验概率分布。

利用影像尺度变换后的大尺度提取结果作为小尺度影像下的初始分割结果, 能够充分利用建筑物在不同尺度影像下的特征信息, 从而提高建筑物提取的完整性(Bouman 和 Shapiro, 1994)。因此本文提出基于多尺度 SLIC-GMRF 的建筑物初始分割。首先利用尺度变换得到多尺度影像序列 $\sigma = [0, \dots, 3]$; 然后利用 SLIC 影像分割算法得到的大尺度影像预分割结果作为小尺度影像层的初始分割, 计算多尺度标记场的先验概率分布; 利用 GMRF 模型对特征场建模, 获得多尺度似然。最后根据 SMAP 准则获得最终的分割结果。

2.1.1 基于多尺度 SLIC 的标记场建模

初始分割对最终分割结果有很大影响, 传统基于像元的初始分割方法忽略像元间的邻域关系, 仅依据 DN 值进行分割, 分割结果破碎。因此本文利用 SLIC 算法获得初始分割结果, 如式(2)–(4)

$$d_{lab} = \sqrt{(l_i - l_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2} \quad (2)$$

$$d_{xy} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (3)$$

$$D_s = d_{lab} + \frac{m}{s} d_{xy} \quad (4)$$

式中, SLIC 算法在 LAB 色彩空间中运算, d_{lab} 代表颜色距离, d_{xy} 代表空间距离, D_s 代表结合 d_{lab} 及 d_{xy} 的最终距离度量结果。 l 表示亮度, a 、 b 皆为颜色通道, a 表示区间从低到高次序为深绿色(低亮度)、灰色(中亮度)、亮粉红色(高亮度), b 表示区间从低

到高次序为亮蓝色(低亮度)、灰色(中亮度)、黄色(高亮度)。基于 K 均值初始聚类结果, 依据距离测度生成过分割对象。 i 、 j 为分割对象编号, m 为紧凑因子, 其控制分割斑块颜色及形状特征所占比重, 较高参数值会为形状特征赋更高比重, 使分割对象尺寸更为方整, 引入 SLIC0 算法(Achanta 等, 2012)对 m 自适应估计, 使各分割斑块具有不同的紧凑因子, 取各聚类斑块范围像素与其聚类中心颜色距离最大值 m_c 作为该分割斑块的参数值; $s = \sqrt{N/k}$ 为步长, 由影像像素数量 N 及定义分割斑块数目 k 决定。

对过分割后的影像斑块首先利用归一化植被指数 NDVI(Normalized Difference Vegetation Index)(Rouse, 1974)去除植被地物, 归一化水体指数 NDWI(Normalized Difference Water Index)(Gao, 1996)去除水体, 形态学阴影指数 MSI(Morphological Shadow Index)(Huang 和 Zhang, 2012)去除分割斑块中的阴影部分。再利用基于城市距离测度的 K-均值算法, 将最高尺度分割结果进行目标背景二分类(Kaufman 和 Rousseeuw, 1990), 生成初始标记场。

标记场中当前节点的标记类型受其尺度间父节点、叔父节点的影响。采用因果的一阶 Markov 链进行建模(Bouman 和 Shapiro, 1994), 如式(5)

$$P(x_s^n | x_{f(s)}^{n+1}, x_{u(s)}^{n+1}) = P(x_s^n | x_{f(s)}^{n+1}, x_{u(s1)}^{n+1}, x_{u(s2)}^{n+1}, x_{u(s3)}^{n+1}) = \frac{\theta^n}{5} \left(2\delta(x_s^n, x_{f(s)}^{n+1}) + \delta(x_s^n, x_{u(s1)}^{n+1}) + \delta(x_s^n, x_{u(s2)}^{n+1}) + \delta(x_s^n, x_{u(s3)}^{n+1}) + \frac{1 - \theta^n}{C} \right) \quad (5)$$

式中, $P(x_s^n | x_{f(s)}^{n+1}, x_{u(s)}^{n+1})$ 为当前节点标记类型的条件概率, x 为节点标记类型, n 为尺度编号, $\delta()$ 为采样函数, $f(s)$ 为父节点, $u(s_i)$ ($i=1, 2, 3$) 为 3 个叔父节点, $\theta^n \in [0, 1]$ 表示节点与父节点及叔父节点具有相同标记的概率, C 为分类数。

尺度内当前节点的标记同时受二阶邻域标记的影响, 采用非因果的 MLL 模型进行建模

$$P(x_s^n | x_{N(s)}^n) = \frac{\exp\left(-\sum_{\tau \in N} V_c(x_s^n, x_{s+\tau}^n)\right)}{\sum_{x_s} \exp\left(-\sum_{\tau \in N} V_c(x_s^n, x_{s+\tau}^n)\right)} \quad (6)$$

式中, $P(x_s^n | x_{N(s)}^n)$ 为尺度内当前节点标记类型的条件概率, x 为节点标记类型, s 表示当前节点, n 为尺度序号, τ 表示二阶邻域相对于中心节点的偏移

量, 其中 V_c 为

$$V_c(x_s^n, x_{s+\tau}^n) = \begin{cases} -\beta, & x_s^n = x_{s+\tau}^n \\ \beta, & x_s^n \neq x_{s+\tau}^n \end{cases} \quad (7)$$

式中, β 为势函数。结合式(5)及式(6)共同确定标记场的先验分布, 得到多尺度标记场。

2.1.2 GMRF 模型特征场建模

为充分利用建筑物多尺度纹理信息及其相互关系, 采用 GMRF 模型 (Krishnamachari 和 Chellappa, 1997) 对特征场向量建模, 考虑二阶邻域位置影响, 如图2所示。

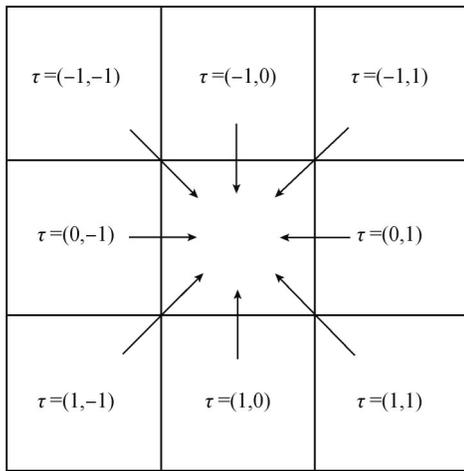


图2 二阶邻域系统交互关系

Fig.2 Interaction of second-order neighboring system

图2中, $\tau \in N$, 为尺度内节点二阶邻域位置关系, 考虑到邻域节点对中心节点的贡献, 似然函数定义为

$$f(\omega_s^n | \mathbf{nei}_s^n, x_s^n = k) = \frac{1}{(\sqrt{2\pi})^{dim} |\mathbf{cov}_k^n|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\mathbf{e}_s^n)^T (\mathbf{cov}_k^n)^{-1} \mathbf{e}_s^n\right) \quad (8)$$

式中, ω_s^n 为节点在尺度 n 上的特征向量, \mathbf{nei}_s^n 为二阶邻域位置对应特征向量集合, dim 为向量的维数, x_s^n 为标记类型, \mathbf{cov}_k^n 为 k 类标记的噪声协方差矩阵。 $\mathbf{e}_s^n = \omega_s^n - \mu_k^n - \mathbf{cov}_k^n \times (\omega_{s+\tau}^n - \mu_k^n)$, μ_k^n 为特征均值向量。

得到 GMRF 模型为

$$P(X = x, W = \omega) = \prod_{n \in \{0, 1, \dots, J-1\}} \prod_{s \in S^n} (P(\omega_s^n | \mathbf{nei}_s^n, x_s^n) P(x_s^n | x_{N(s)}^n)) \quad (9)$$

式中, $P(X = x, W = \omega)$ 为结合多尺度特征与多尺度标记的联合概率密度, 为提高算法运行效率,

特征场模型建立在最高分辨率影像上, 根据 SMAP 准则对多尺度影像进行最优分割, 提取出候选建筑物对象。

2.2 建筑物 FCN 特征提取

在建筑物初始分割的基础上, 利用经微调后的全卷积网络 FCN (Fully Convolutional Networks) (Long 等, 2015) 抽取建筑物相关特征层特征, 输入 SVM 模型进行训练获得 FCNSVM 分类器, 并对建筑物初始分割结果进行二次分类得到建筑物最终提取结果。

2.2.1 FCN 网络结构

本文使用的网络结构如图3所示, 由于高分辨率遥感影像包含可见光谱段, 并且空间分辨率小于 1 m, 因此采用基于 VGG16 (Simonyan 和 Zisserman, 2015) 的网络结构, VGG16 网络基于 ImageNet 数据集 (Russakovsky 等, 2015) 训练得到, 其预训练获得的权重对于本文网络结构的微调具有重要的意义。FCN 网络将传统卷积神经网络 (CNN) 中的全连接层 (Fully connected layer) 替换为卷积层 (Convolution layer), 网络中包括卷积层、激活函数层 (Activation function layer)、池化层 (Pooling layer)、反卷积层 (Deconvolution layer)、裁剪层 (Crop layer) 以及 Eltwise 层。

卷积层是卷积神经网络的核心层, 其对输入影像进行一系列操作, 与图像滤波类似

$$O_{ic,jc} = f\left(\sum_{uk=1}^{m_k} \sum_{vk=1}^{n_k} F_{uk,vk} \cdot I_{ic-uk+1, jc-vk+1} + bias_{uk,vk}\right) \quad (10)$$

$$\begin{aligned} & (1 \leq ic \leq M_c, 1 \leq jc \leq N_c) \\ & (1 \leq uk \leq m_k, 1 \leq vk \leq n_k) \end{aligned}$$

式中, M_c 、 N_c 为影像的行列数, m_k 、 n_k 为滤波器 F 尺寸, $bias$ 为偏置项。给定输入影像 I , 影像 $I_{ic-uk+1, jc-vk+1}$ 处的像元与对应 (uk, vk) 处的滤波器 F 相乘, 并与 (uk, vk) 偏置项相加, 求和得到对应 $O_{ic,jc}$ 处的输出像元。 f 一般不取任何操作, 由之后的激活函数层来产生非线性映射。

激活函数层的主要作用是将输入数据中加入非线性因素, 使得模型能够解决更多非线性问题。常用的激活函数包括 Sigmoid 函数、tanh 函数、ReLU 函数 (Canziani 等, 2016)。

池化层对输入的特征进行压缩, 减小特征图尺寸, 突出影像中的主要特征, 使得网络计算复杂度降低。按照缩放后的取值规则分为极大值采

样(maxpooling), 平均采样(avgpooling), 以及随机采样(randpooling)等(Laptev 等, 2016)。

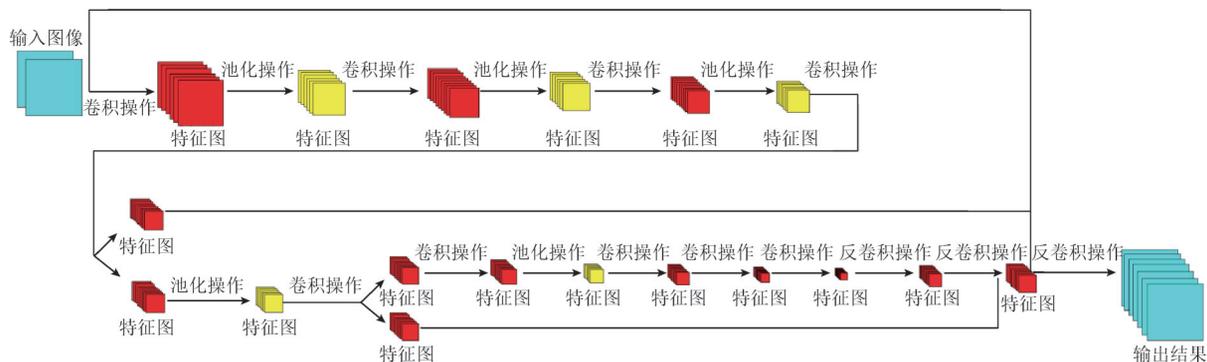


图3 FCN 网络结构

Fig.3 FCN architecture

反卷积层用于将经过卷积及池化操作后的特征影像进行上采样操作, 从而恢复特征影像的尺寸, 反卷积层可以使得网络学习到更加复杂的特征(Bengio 等, 2013)。

裁剪层的主要作用是对特征影像进行裁剪, 使其满足要求的尺寸大小。FCN 网络对输入数据进行了填充操作, 因此利用裁剪层使处理后的特征数据恢复原始尺寸(Jia 等, 2014)。

Eltwise层对特征层进行操作并合并为一层, 主要操作包括(点乘、相加减、取最大值)(Jia 等, 2014)。

2.2.2 建筑物特征提取

FCN 网络能够直接用于语义分割, 从而得到建筑物的提取结果, 然而FCN 网络在多次上采样下采样过程中使得提取结果细节信息缺失严重, 因此直接将其作为分类器并不能获得理想的建筑物提取结果。同时 FCN 网络是一个多层的神经网络

结构, 每一层都由多个2维平面构成, 每一个2维平面都由若干独立的神经元组成, 网络中的浅层结构能够得到色调、边缘等浅层信息, 随着层次增加, 非线性程度增强, 卷积得到的特征会更加抽象, 并具备局部特征, 因此利用FCN 网络作为特征提取器能够获得建筑物的多元层次特征, 利于建筑物的提取。

将网络数据集输入微调后得到的FCN 网络, 能够得到对应卷积层输出的影像特征, 不同卷积层输出的影像特征结构不同。

如图4所示, 底层特征(如卷积层1特征)更多地提取了边缘及角点等低层次特征, 同时道路与建筑物的区分度较低, 随着层数加深, 相似光谱纹理特征像元逐渐被合并, 形状特征逐渐成为主导特征, 不同细节的建筑物特征被合并成为整体建筑物特征, 由此抽象出建筑物的高层次特征。

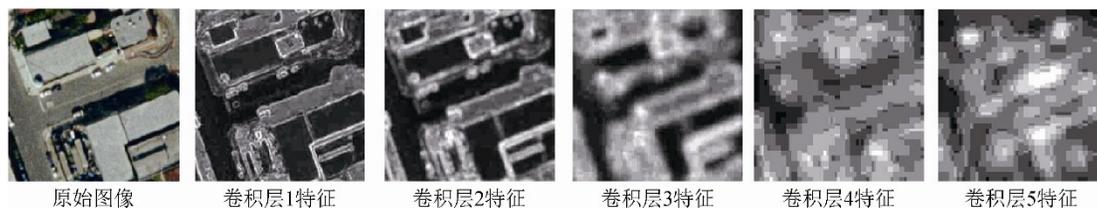


图4 建筑物 FCN 特征

Fig.4 FCN features of buildings

2.3 多尺度 SLIC-GMRF 与 FCNSVM 联合的建筑物提取

2.3.1 FCNSVM 分类器构建

结合 FCN 网络输出得到的各卷积层对应地物特征, 并同时生成地物类标签, 利用 SVM 分类器

进行建模, SVM 是经典的机器学习算法, 最早被用于解决分类问题, 其核心思想为寻找能够区分不同地物类型的最优超平面, 能够最大化不同类别训练样本间的分类距离获得最佳的分类结果(Cortes 和 Vapnik, 1995)。

本文利用FCN网络各卷积层输出的建筑物特征及其他地物特征构建FCNSVM模型。流程如图5所示。

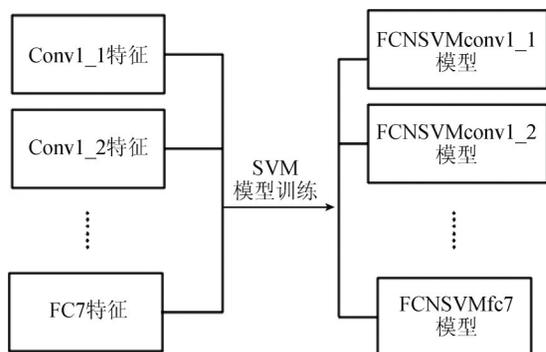


图5 FCNSVM模型建立

Fig.5 FCNSVM model

如图5所示,由于不同卷积层输出的特征维度不同,因此利用libsvm分类器(Chang和Lin, 2011)分别对每个卷积层特征建模,从而得到建

筑物对应卷积层特征结构的FCNSVM分类器。结合SLIC-GMRF建筑物初始分割,对其利用FCNSVM分类器进一步分割,从而获得建筑物提取结果。

2.3.2 FCNSVM建筑物提取

由于建筑物初始分割结果中存在人工地物及裸土混分现象,因此需要结合训练得到的FCNSVM分类器进行进一步的划分提取,得到最终建筑物提取结果。

如图6所示,建筑物初始分割结果为覆盖研究区域的二值影像图,其中影像中的白色区域值为1,代表建筑物候选区域;黑色区域值为0,表示其他地物的背景区域。为减小建筑物识别的运算量,将原始影像与建筑物初始分割结果进行掩膜运算,对应背景区域划归空值不参与到建筑物的最终提取中。

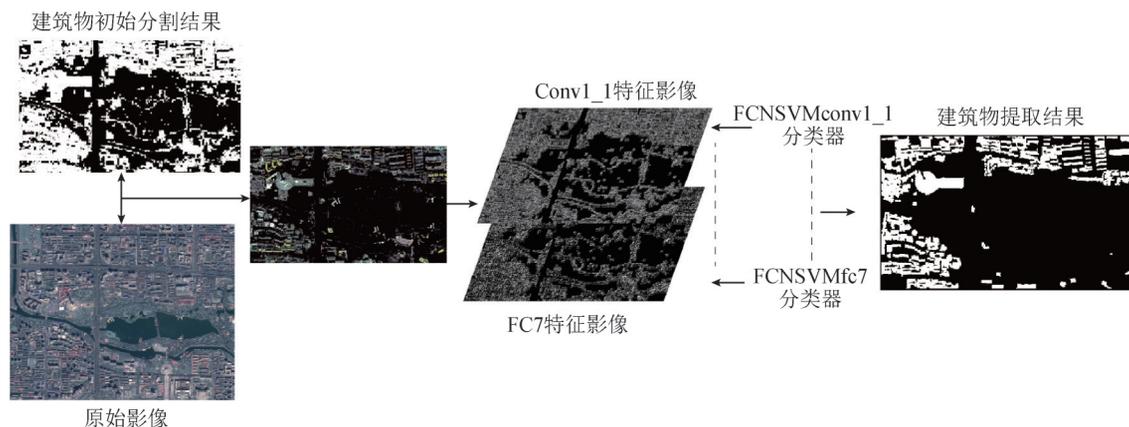


图6 多尺度SLIC-GMRF与FCNSVM联合的建筑物提取

Fig.6 Extraction of buildings in remote sensing imagery based on multi-scale SLIC-GMRF and FCNSVM

若将掩膜运算后的影像直接输入到FCN网络生成特征影像,在经过网络卷积层及池化层的运算处理后,影像中地物特征会受到极大的影响,造成地物细节缺失,最终影响建筑物提取结果。因此本文将掩膜运算后的影像分块输入FCN网络,以获取较小窗口尺寸下的各卷积层特征影像,保留地物细节信息,同时特征影像块维度与对应各卷积层特征维度相同。由于影像通过各卷积层会使得输出影像尺寸小于原始尺寸,为保证影像尺寸的一致性,利用双线性插值法将卷积层运算后的影像块恢复至原始尺寸,将其拼接后得到各卷积层特征影像,利用训练完成的FCNSVM分类器对其进行

建筑物提取。

FCNSVM分类器是根据研究区样本影像不同卷积层特征训练得到的,利用各卷积层的FCNSVM分类器分别对特征影像进行计算,其中背景区域不参与提取识别。由于提取目标为建筑物,因此将建筑物子类型合并为建筑物大类,同时将其地物类型划归为背景。由于分别利用各层分类器对影像进行提取,因此需要解决各层提取结果的不一致性,深层次FCNSVM分类器能够依据建筑物的抽象结构特征进行分类,因此以最深层次的FCNSVMfc7分类器的提取结果作为基础分类结果,然而其地物边缘特征较为概括,本文

利用其他层分类结果进行修正。若基础分类结果为非建筑物类型, 则将该地物对象划分为背景, 不考虑上层分类结果; 如果各层分类结果皆为建筑物, 则该斑块对象为建筑物内部斑块; 若基础分类结果为建筑物, 上层分类结果存在非建筑物分类结果, 则为建筑物边界区域, 为减小判别误差, 上层分类结果中建筑物大于 70% 的, 地物斑块类别不发生改变, 否则将该地物斑块类型划分

为背景。依此规则得到最终建筑物提取结果。

3 实验及评价

3.1 数据及参数选择

3.1.1 实验数据

本文选择北京市、武汉市、越南岷港作为研究区, 数据源为高分二号卫星影像, 数据信息见表 1。

表 1 研究区影像数据源

Table 1 Image dataset of study areas

研究区	获取时间	全色波段空间分辨率/m	多光谱波段空间分辨率/m	像幅
北京	2015-09-12	0.8	4	6725×4207
武汉	2016-02-07	0.8	4	4548×4503
岷港	2015-08-04	0.8	4	7300×6908

3 个研究区具有不同的建筑物特征, 北京、武汉研究区分布较多的高层建筑, 岷港研究区则以低矮建筑物为主。为对建筑物提取精度进行定量评价, 本文采用查准率(Precision)、查全率(Recall)及综合分数(F_1), 见式(11)–(13)。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (11)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (12)$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \quad (13)$$

式中, TP 为正确提取出的建筑物(单位为像素, 下同), FP 为虚警, 即与参考结果无匹配的建筑物提取结果, FN 为漏分, 即未提取出的建筑物, F_1 值越大, 表明算法提取效果越好(Heipke 等, 1997)。

3.1.2 参数选择

多尺度 SLIC-GMRF 模型中需要对参数序列 $p = [\sigma, k, \theta, \beta]$ 进行估计, 即分别为尺度参数 σ , 分割对象数目 k , 尺度间交互参数 θ 以及势函数参数 β 。由于本文选择具有不同密度及高度的建筑物类型影像(北京为中密度高层建筑; 武汉为高密度高层建筑; 岷港为高密度低矮建筑)进行参数的选择估计, 因此依据其得到的参数具有一定适用性, 能够应用于其他区域。

不同尺寸建筑物在尺度序列影像中具有差异性(李小文和王祎婷, 2013), 因此对研究区影像进行尺度变换, 以突出不同影像尺度下的建筑物特征。利用研究区实地踏勘采样数据, 计算建筑物样本同其他地物的统计可分性, 生成尺度—离散

度曲线, 如图 7 所示。

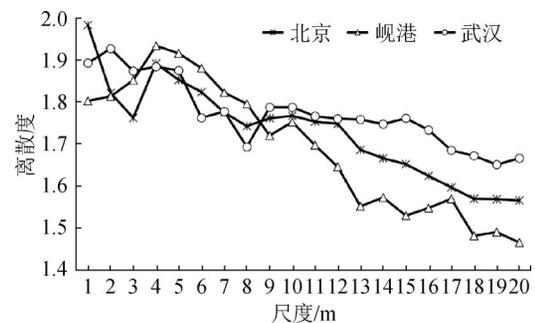


图 7 尺度—离散度曲线

Fig.7 Scale-dispersion curve

在北京研究区中, 离散度指标随尺度增大, 呈逐渐降低趋势, 北京研究区存在高层建筑物, 植被覆盖度较高, 影像空间分辨率为 1 m 时, 建筑物与周边地物类别具有较高的统计可分性, 离散度值较高, 随着影像尺度增大, 区域边界逐渐表现为混合像元, 离散度值减小, 类别可分性降低; 岷港研究区中, 离散度指标随尺度增大表现为先上升后降低的趋势, 具有明显峰值。研究区中分布大量邻接低矮建筑物, 当影像尺度增加时, 建筑物内部的像元光谱异质性降低, 对离散度的贡献高于边界混和像元, 离散度值增加; 随着影像尺度进一步升高, 建筑物与周围地物表现为混合像元, 离散度值降低; 武汉研究区样本统计可分性随影像尺度升高基本保持一致, 研究区影像存在高大建筑群, 边界清晰, 当影像尺度上升时, 建筑物内部像元光谱异质性降低, 混和像元对离散度的贡献有限, 曲线变化较为平缓。结合各研

研究区尺度—离散度曲线变化特征及各研究区影像地物分布实际情况, 确定影像研究尺度为 1 m、4 m、10 m, 利用双三次插值算法对 1 m 融合影像进行重采样, 生成 10 m 分辨率影像。对多尺度变换后的研究区影像利用 SLIC 算法进行过分割处理。

SLIC 算法中, 需要指定分割对象数目 k , 过小的参数值会导致影像欠分割, 不能充分利用影像的光谱空间特征, 并弱化建筑物斑块与非建筑物斑块的边界, 对建筑物提取造成不利影响; 过高的参数值会影响算法的执行效率, 因此本文利用局部方差变化率 (Rate of Change, ROC) (Drăguț 等, 2010) 对 k 值进行筛选优化。

设定初始参数 $k=8000$, 以 500 为增长间隔, 对空间分辨率 1 m、4 m 及 10 m 的研究区影像进行过分割实验, 生成 ROC 曲线, 如图 8 所示。

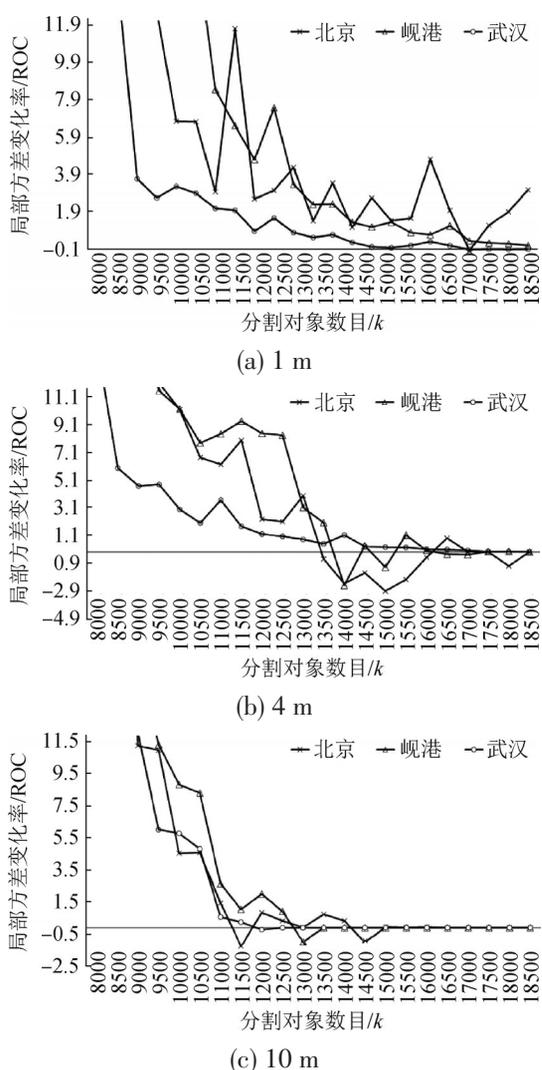


图 8 局部方差变化率曲线

Fig.8 ROC curve

局部方差反映了分割对象间的同质性, 随着分割对象数目增加, 对象间的差异性不断减小, 局部方差值不断降低。局部方差变化率放大局部方差变化, 其极大值处代表局部方差变化突变点, 表明在该点两侧局部方差变化剧烈, 作为参数候选值可使分割对象间保持较大的差异性 (Drăguț 等, 2014)。

由图 8 可知, 武汉研究区 ROC 曲线与其他研究区相比, 对 k 值变化更为敏感, 分割对象数目增加使得局部方差变化率迅速变化。通过对 3 个研究区 ROC 曲线的分析比较, 结合局部方差变化率的增减趋势, 确定空间分辨率为 1 m 时, 参数 $k=12500$, 空间分辨率为 4 m 时, 参数 $k=11500$, 空间分辨率为 10 m 时, 参数 $k=10000$, 对 3 个研究区多尺度影像进行过分割, 如图 9 所示。

如图 9 所示, 随着影像尺度逐渐升高, 影像中的地物细节被概括, 研究区影像空间分辨率为 1 m 时, 地物边界包括地物内部依照同质性规则划分为若干区域, 建筑物顶部细节突出, 内部光谱异质性较高, 导致其顶部分割较为破碎; 影像空间分辨率为 4 m 时, 建筑物顶部类内光谱异质性逐渐降低, 影像分割对象内部斑块数目减少; 影像空间分辨率为 10 m 时, 各研究区中建筑物内部光谱纹理特征均一, 细小道路被概括为混合像元, 便于减轻非建筑物地物对分割的影响。统计生成各研究区 10 m 空间尺度影像斑块的 NDVI、NDWI、MSI, 剔除其植被, 及低反射率地物如水体、阴影的影响。然后利用基于城市距离测度的 K-means 算法对结果进行目标背景二分类, 生成初始聚类标记场, 其余尺度标记场通过尺度间一阶 Markov 链的依赖关系以及 MLL 模型确定, 以提高算法的运行效率。

结合多尺度 SLIC 过分割生成的初始聚类结果, 一阶 Markov 链及 MLL 模型中存在参数 θ^n 及 β 需要确定。对于 θ^n 采用如下公式进行计算 (Bouman 和 Shapiro, 1994)。

$$\hat{\theta}^n = \operatorname{argmax}_{\theta^n} (E(\lg(f(y_d^n | x^n, y))) \quad (14)$$

$$P(x^n | x^{n+1}) | y, x^{n+1}, \theta^n)$$

式中, θ^n 为当前迭代次数的尺度间交互参数, $f(y_d^n | x^n, y)$ 为各尺度的似然函数, y_d^n 为各尺度间的特征向量树, x^n 为各尺度标记类型, $P(x^n | x^{n+1})$ 为各尺度标记先验概率。



图9 SLIC多尺度过分割结果

Fig.9 Over-segment results by SLIC

势函数参数 β 能够影响分割结果的区域性, 通常通过经验确定(Yu 和 Clausi, 2006), 例如, Rignot 和 Chellappa(1992)对极化 SAR 影像进行分割, 确定最适宜 β 参数范围为[1.0,1.6]。Bruzzone 和 Prieto(2000)设定 β 参数为[1.6,1.8]以满足变化检测精度要求。由于本文采用影像数据空间分辨率为 1 m, 并且建筑物提取对分割细节要求较高, 本文设定 β 参数值为 0.9。采用设定参数序列得到建筑物初始分割结果, 如图 10 所示。

如图 10 所示, 3 个研究区中由于建筑物类型及密度不同, 初始分割效果具有差异, 北京市研究区同其他两个研究区相比, 建筑物密度较低, 建筑物间距较大, 建筑物能够较为完整地提取出来, 在武汉市及岷港研究区中, 由于建筑物密度高于北京市, 因此存在部分道路与裸地错误分割为建筑物的情况, 需要通过后续步骤修正。

FCN 网络训练阶段在 VGG16 权重的基础上进行, VGG16 在 ImageNet 数据集基础上(包含 120 万张图片)训练而来, 为使其能较好地应用于遥感影像分类应用中, 本文利用 UC Merced Land Use Dataset 数据集对其进行微调(Yang 和 Newsam, 2010)。UC Merced Land Use Dataset 数据集包括 21 个地物类型(农业用地、飞机、棒球场、海滩、

建筑物、灌木、高密度居民地、森林、高速公路、高尔夫球场、港口、道路交叉口、中密度居民地、活动房区、高架桥、停车场、河流、机场跑道、低密度居民地、储油罐、网球场), 每一地类包含 100 张影像, 影像空间分辨率为 0.3 m, 尺寸 256×256, 涵盖高分辨率卫星影像能够识别的地物类型, 通过对影像数据集进行旋转、镜像、添加噪声操作对数据集进行扩充, 最终将每一地类影像数目扩充至 1000 张, 共 21000 张地物影像。

将 21000 张影像以 4 : 1 比例划分为训练集及测试集, 以保证训练集影像的数目, 避免训练模型过拟合(Romero 等, 2016)。定义批尺寸(Batch Size)参数值时考虑到过小的参数会导致模型难以收敛, 过大的参数值会导致内存消耗过大, 因此确定 Batch Size 为 10(Bengio, 2012), 即每次从训练影像集中取出 10 幅图像进行训练。初始学习率设定为 10^{-5} , 学习率每 2000 次迭代除以 10。权值衰减(Weight decay)参数值设定为 0.02, 以防止参数过拟合。Momentum 参数值设定为 0.9 以加快梯度下降的速度。生成 Loss 曲线, 如图 11 所示。

由图 11 可得, 随着训练迭代次数增加, 学习率的调整, 网络逐渐收敛。在 3 个研究区影像中随机选择相同数目的训练样本及测试样本, 各研究区样本数目如表 2 所示。

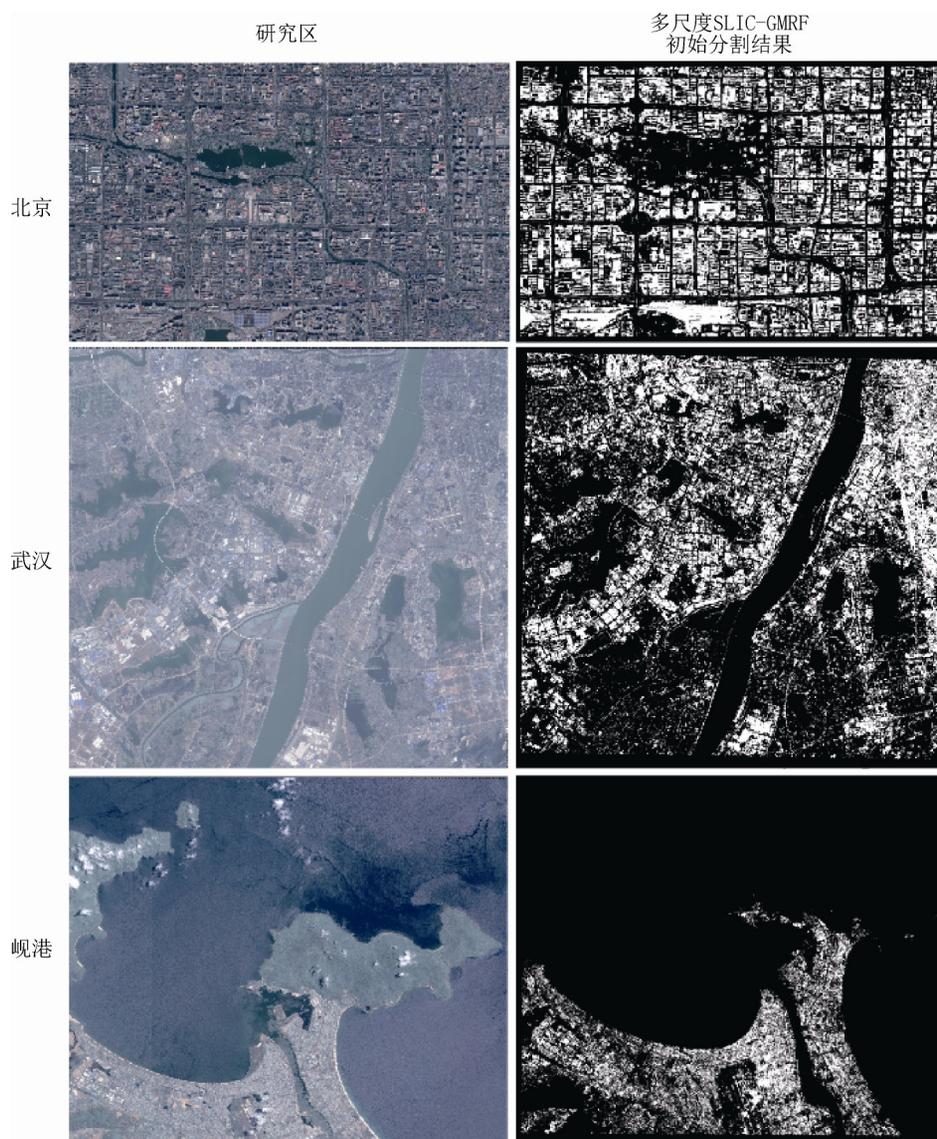


图10 多尺度SLIC-GMRF建筑物初始分割结果

Fig.10 Initial segmentation result of buildings by multi-scale SLIC-GMRF

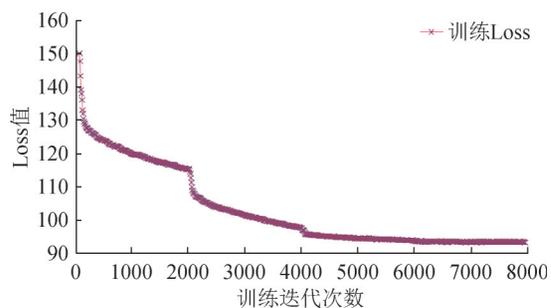


图11 FCN网络 epoch-loss 曲线

Fig.11 The epoch-loss curve of FCN network

将样本图像重新输入到FCN网络中,得到各卷积层提取的影像特征,其中获取影像特征数目见表3。

由于本文目标为建筑物提取,因此在训练

FCNSVM分类器时,将对应于FCN网络中的建筑物、高密度居民地、中密度居民地、活动房区、低密度居民地5类地物与建筑物类相对应,其他地物类型合并为背景地物类,以对应研究区样本类型。将各层卷积层影像特征分别建立FCNSVM模型,由表3所示,卷积层输出特征随着网络深度加深,特征维度迅速增加,利用传统训练方式会导致训练复杂度增长,不利于模型的构建,因此利用MapReduce模型进行分布式运算,以提高模型的训练效率(Wickham, 2011)。其中核函数的选择对SVM分类器的性能具有重要的影响,常用的核函数包括线性核、多项式核、径向基函数(RBF)、sigmoid核(Drăguț等, 2010)。由于构建分类器的训练样本维度高,并且为增强非线性表达能力,选择

RBF 函数作为核函数进行训练, 利用五折交叉验证 (5-fold cross validation) 选择最高准确率的核函数参数 ($\gamma=6.8564, c=0.00028$)。将各研究区建筑物初始分割区域影像输入到微调完成的 FCN 网络特征提取器中, 生成各卷积层 FCN 影像特征, 并利用训练完成的 FCNSVM 模型对其进行二次分类。

表 2 样本数目表
Table 2 Table of samples

	样本数量		
	建筑物	非建筑物	合计
训练样本	2364	2363	4727
测试样本	1013	1012	2025

表 3 FCN 特征数目表
Table 3 Feature maps of FCN

卷积层名称	影像特征数目	卷积层名称	影像特征数目
Conv1_1	64	Conv4_2	512
Conv1_2	64	Conv4_3	512
Conv2_1	128	Conv5_1	512
Conv2_2	128	Conv5_2	512
Conv3_1	256	Conv5_3	512
Conv3_2	256	Fc6	4096
Conv3_3	256	Fc7	4096
Conv4_1	512		

3.2 结果分析及对比

3.2.1 控制实验对比

为了对本文方法进行验证, 首先进行以下控制实验:

表 4 建筑物提取精度评价表

Table 4 Evaluation table of building extracting accuracy

	北京			武汉			岷港		
	R	P	F_1	R	P	F_1	R	P	F_1
对比 1	81.24	89.26	85.06	81.54	87.21	84.28	79.54	94.28	86.29
对比 2	89.36	91.51	90.42	92.71	86.00	89.23	92.86	91.65	92.25
对比 3	91.72	74.65	82.31	88.73	74.32	80.89	84.58	76.78	80.49
本文方法	97.26	96.68	96.97	94.45	95.97	95.20	97.13	93.24	95.15

注: R(Recall), P(Precision), F_1 分别为查全率、查准率及综合分数指标。

表 4 中 R、P、 F_1 分别为查全率、查准率及综合分数指标, 本文方法提取精度高于其他 3 种对比实验结果。对比 3 提取精度在 4 种方法中最低, 由于 SVM 分类器在训练过程中仅利用到训练数据的

对比 1: 不采用影像分割, 利用多尺度 GMRF 模型进行建筑物初始分割, 其他部分相同的对照试验, 对影像分割的影响进行评判。

对比 2: 在多尺度 SLIC-GMRF 建筑物初始分割的基础上, 建筑物二次分类时利用训练完成的 FCN 网络直接分类, 以评价 FCN 网络分类器性能。

对比 3: 在多尺度 SLIC-GMRF 建筑物初始分割的基础上, 利用 UC Merced Land Use Dataset 数据集及研究区样本数据中的光谱纹理信息训练 SVM 分类器并进行直接分类, 以分析建筑物 FCN 影像特征对建筑物提取的影响。

以 Google earth 20 级影像 (空间分辨率: 0.27 m) 作为参考影像数据, 对研究区内的建筑物人工矢量化生成参考数据, 以定量评价建筑物提取结果。提取结果如图 12 所示灰度图为提取得到的建筑物二值影像, 彩色影像为提取结果与原影像叠加结果绿色为与参考结果相同部分, 蓝色为满分, 红色虚警。

如图 12 所示, 本文方法效果优于其他 3 种对比实验方法, 产生较少的虚警及漏分, 与参考结果的一致性更佳。对比 3 在建筑物初始分割基础上直接利用 SVM 分类器进行提取, 在 3 个研究区中产生了较多的虚警及漏分。对比 1 采用基于像元的 GMRF 算法对建筑物进行初始分割, 在各研究区建筑物边缘处产生较多虚警。对比 2 利用 FCN 网络在建筑物初始分割结果基础上直接分类, 如图所示其在各研究区密集建筑物区域存在较多的虚警现象, 将道路错误提取为建筑物。生成建筑物提取精度评价表对提取结果进行定量评价, 如表 4 所示。

光谱及纹理特征, 使得分类器的性能较差, 直接应用于建筑物初始分割结果导致查准率指标较低。对比 1 在生成初始分割结果时未进行影像分割, 因此使得其建筑物候选区域的精度劣于其他几种方

法, 其查全率指标较低。对比2在建筑物密集区域提取结果较差, 由于FCN网络作为分类器时对影

像进行一系列降尺度升尺度操作, 概括了大量地物细节。

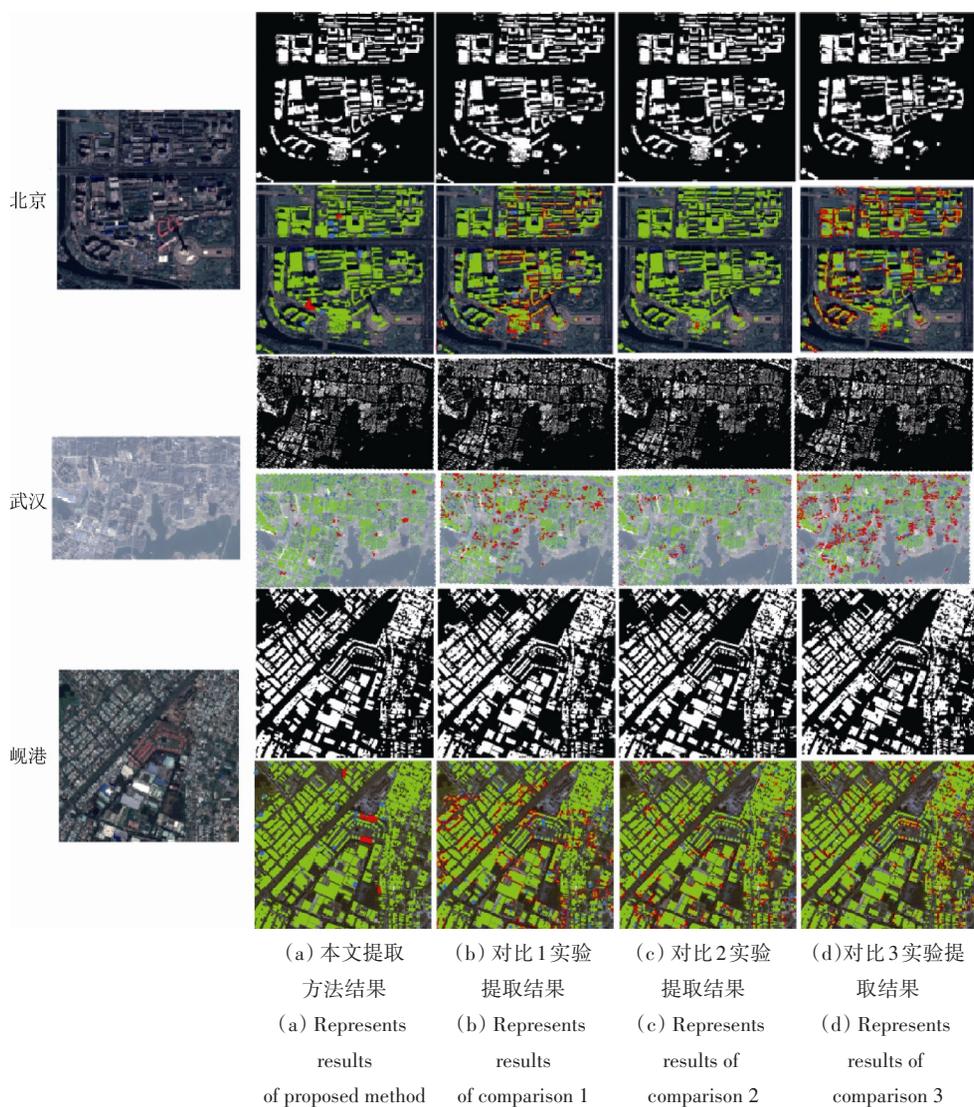


图12 建筑物提取结果

Fig.12 Results of building extraction

采用配对卡方检验评价以上对比实验及本文方法对建筑物提取结果的影响差异是否显著。生成配对卡方检验表如表5所示。

表5 配对卡方检验评价表

Table 5 McNemar-Bowker test table

方法	本文方法			显著性
	正确	错误		
对比1	正确	3570593	217482	0.000
	错误	771261	16049319	
对比2	正确	4075879	325342	0.000
	错误	364957	15913165	
对比3	正确	3617826	298968	0.000
	错误	317463	16388675	

如表5所示, 统计比较本文方法前后3种对比实验结果像元标签的转化情况, 分为正确—正确, 即采用本文方法前后不变的正确分类像元数、正确—错误, 即由正确像元转化至错误像元数、错误—正确, 即由错误像元转换为正确像元数, 错误—错误, 即采用本文方法前后不变的错误分类像元数。4种方法的显著性均小于0.01, 具有极显著的统计学意义, 表明采用以上对比实验与本文方法相比, 结果具有显著的差异性。

3.2.2 同其他方法对比

同时与两种基于CNN的建筑物提取方法进行对比:

(1)Mnih(2013)利用 CNN 网络模型分别提取道路及建筑物, 采用的 CNN 网络模型输入层结构为 $3 \times 64 \times 64$, 并包括 3 层卷积层($64 \times 13 \times 13$ 、 $112 \times 9 \times 9$ 、 $80 \times 7 \times 7$)及两个 4096、256 维全连接层。

(2)Saito 等(2016)进行建筑物提取采用的 CNN 网络模型输入层结构为 $3 \times 64 \times 64$, 包括 3 个卷积层($64 \times 9 \times 9$ 、 $128 \times 7 \times 7$ 、 $128 \times 5 \times 5$)及两个 4096、256 维的全连接层。

为同对比方法结果进行比较, 采用相同数据集、训练参数对文中 CNN 网络进行实现, 并利用相同 Massachusetts 数据集(Mnih, 2013)进行建筑物提取比较。

如图 13 所示, 与前文研究区相比, Massachusetts 数据集中的建筑物密度及植被覆盖度较低, 地物间反差更为明显, 利用多尺度 SLIC-GMRF 算法能够得到完整的建筑物初始分割斑块, 结合 FCNSVM 模型取得较好的目视提取结果, 漏分及虚警的像元数均较少。为对本文提取方法进行评价, 引入 P-R 曲线, 横轴对应 Precision 值, 纵轴对应 Recall 值, 利用 P-R 曲线能够较好地反映出模型之间建筑物提取能力的优劣, P-R 曲线包括一系列不同阈值对应的 Precision 以及对应的 Recall 值, 如图 14 所示。

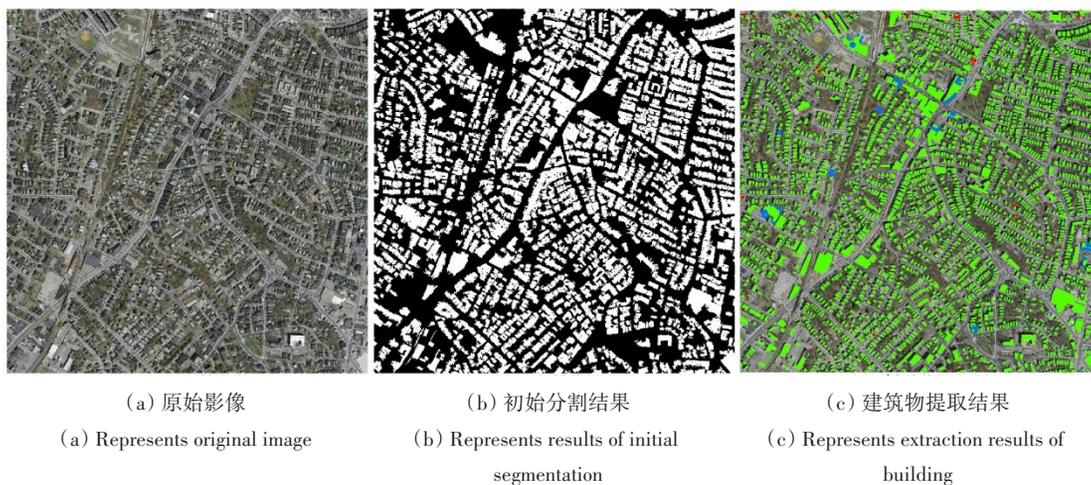


图 13 Massachusetts 数据集提取结果

Fig.13 Extraction results of Massachusetts dataset

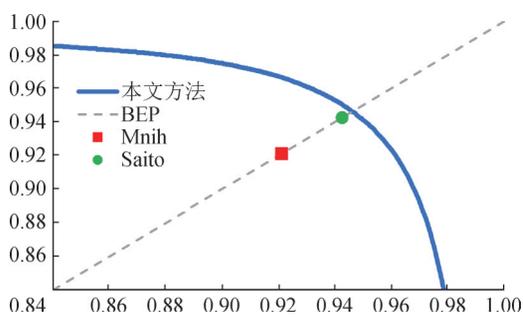


图 14 P-R 曲线

Fig.14 P-R curve

如图 14 所示, 本文方法对应的 P-R 曲线将两对比方法 BEP(Break-Even-Point)值包含在内, BEP 曲线代表该曲线上的点 Precision 与 Recall 的值相同, 亦即查准及查全的平衡点。为定量评价提取精度生成 BEP 点提取精度表, 如表 6 所示。

如表 6 所示, 本文方法对应的 BEP 精度高于两

种利用 CNN 网络作为分类器的对比方法。本文方法同对比方法相比提取了建筑物候选区域, 而准确的建筑物候选区域能够有效避免直接利用 CNN 网络进行分类导致的建筑物边缘精度较差的缺陷, 并且本文采用的神经网络深度结构高于两种对比方法, 得到与 Liu 和 Deng(2016)相一致的结论, 即更深的网络结构有助于得到更精确的提取结果。同样生成配对卡方检验表进行评价, 如表 7 所示。

表 6 BEP 点处提取精度

Table 6 Extraction accuracy on BEP points

方法	Massachusetts Dataset	
	R	
Mnih	92	
Saito	94.3	
本文方法	95.4	

表7 配对卡方检验评价表
Table 7 McNemar-Bowker test table

方法	本文方法			显著性
		正确	错误	
Mnih	正确	290084	94327	0.000
	错误	66841	1287632	
Saito	正确	345329	108725	0.000
	错误	54527	1508646	

如表7所示, 基于配对卡方检验评价结果, 对于 Massachusetts 数据集, 3种不同建筑物提取方法结果具有本质不同, 本文方法建筑物提取结果优于两种对比方法。

4 结论

本文提出基于多尺度 SLIC-GMRF 和 FCNSVM 的建筑物提取方法能有效解决以高分影像为数据源的建筑物提取难题。相比于现有的建筑物提取方法, 本文方法利用 FCN 网络作为特征提取器, 综合影像的多尺度特征、光谱特征、几何特征、纹理特征以及深层结构特征, 使得影像中的建筑物得到增强。同时, 计算机视觉领域中的 SLIC 影像分割及 GMRF 模型提高了建筑物提取结果的完整性及准确性, 避免了传统建筑物提取方法中由于直接分类导致提取完整性较差的缺陷。经控制实验结果表明, 不采用影像分割, 会使建筑物候选区域的提取精度较差, 查全率指标较低, 表明 SLIC 影像分割影响到初始分割结果的完整性; 二次分类时直接利用 FCN 网络分类会导致建筑物密集区域提取结果较差, 表明 SVM 分类器影响到建筑物细部提取结果; SVM 分类器直接应用于建筑物初始分割结果导致查准率指标较低, 表明 FCN 特征影响到 SVM 模型的性能。对比实验结果表明, 在建筑物特征清晰、遮挡干扰较少的区域, 本文方法能够取得较好的提取结果; 同时, 较深的 FCN 网络结构有助于增强建筑物的提取能力, 与对比建筑物提取方法相比, 本文方法提高了建筑物提取精度。然而, 本文方法仍未能解决遮挡物对建筑物提取造成的干扰问题; 同时, 实验结果中仍然存在边缘细节丢失以及误提取现象, 需要在下一步研究中进行改善。

参考文献(References)

- Achanta R, Shaji A, Smith K, Lucchi A, Fua P and Süsstrunk S. 2010. SLIC Superpixels. EPFL Technical Report no. 149300. EPFL
- Achanta R, Shaji A, Smith K, Lucchi A, Fua P and Süsstrunk S. 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11): 2274-2282 [DOI: 10.1109/TPAMI.2012.120]
- Benedek C, Shadaydeh M, Kato Z, Szirányi T and Zerubia J. 2015. Multilayer Markov random field models for change detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 107: 22-37 [DOI: 10.1016/j.isprsjprs.2015.02.006]
- Bengio Y. 2012. Practical recommendations for gradient-based training of deep architectures//Montavon G, Orr G B and Müller K B, eds. *Neural Networks: Tricks of the Trade*. Berlin, Heidelberg: Springer: 437-478 [DOI: 10.1007/978-3-642-35289-8_26]
- Bengio Y, Courville A and Vincent P. 2013. Representation learning: a review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8): 1798-1828 [DOI: 10.1109/TPAMI.2013.50]
- Besag J. 1986. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B: Methodological*, 48(3): 259-302
- Bouman C and Liu B. 1991. Multiple resolution segmentation of textured images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2): 99-113 [DOI: 10.1109/34.67641]
- Bouman C A and Shapiro M. 1994. A multiscale random field model for bayesian image segmentation. *IEEE Transactions on Image Processing*, 3(2): 162-177 [DOI: 10.1109/83.277898]
- Bruzzone L and Prieto D F. 2000. Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3): 1171-1182 [DOI: 10.1109/36.843009]
- Canziani A, Paszke A and Cullurciello E. 2016. An analysis of deep neural network models for practical applications//*Proceedings of 2016 ICLR*. San Juan: ICLR
- Chang C C and Lin C J. 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3): 27 [DOI: 10.1145/1961189.1961199]
- Collobert R and Weston J. 2008. A unified architecture for natural language processing: deep neural networks with multitask learning//*Proceedings of the 25th International Conference on Machine Learning*. Helsinki, Finland: ACM: 160-167 [DOI: 10.1145/1390156.1390177]
- Cortes C and Vapnik V. 1995. Support-vector networks. *Machine Learning*, 20(3): 273-297 [DOI: 10.1007/BF00994018]
- Drăguț L, Tiede D and Levick S R. 2010. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science*, 24(6): 859-871 [DOI: 10.1080/13658810903174803]

- Drăguț L, Csillik O, Eisank C and Tiede D. 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88: 119-127 [DOI: 10.1016/j.isprsjprs.2013.11.018]
- Felzenszwalb P F and Huttenlocher D P. 2004. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2): 167-181 [DOI: 10.1023/B:VISI.0000022288.19776.77]
- Gao B C. 1996. NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3): 257-266 [DOI: 10.1016/S0034-4257(96)00067-3]
- Geman S and Geman D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6): 721-741 [DOI: 10.1109/TPAMI.1984.4767596]
- Gong M G, Yang H L and Zhang P Z. 2017. Feature learning and change feature classification based on deep learning for ternary change detection in SAR images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 129: 212-225 [DOI: 10.1016/j.isprsjprs.2017.05.001]
- Grinias I, Panagiotakis C and Tziritas G. 2016. MRF-based segmentation and unsupervised classification for building and road detection in peri-urban areas of high-resolution satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 122: 145-166 [DOI: 10.1016/j.isprsjprs.2016.10.010]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, United States: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Heipke C, Mayer H, Wiedemann C and Jamet O. 1997. Evaluation of automatic road extraction. *International Archives of Photogrammetry and Remote Sensing*, 32(3-2W3): 47-56
- Hinton G, Li D, Dong Y, Dahl G E, Mohamed A R, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath T N and Kingsbury B. 2012. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6): 82-97 [DOI: 10.1109/MSP.2012.2205597]
- Huang X and Zhang L P. 2012. Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(1): 161-172 [DOI: 10.1109/JSTARS.2011.2168195]
- Jia Y Q, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S and Darrell T. 2014. Caffe: convolutional architecture for fast feature embedding//*Proceedings of the 22nd ACM International Conference on Multimedia*. Orlando: ACM: 675-678 [DOI: 10.1145/2647868.2654889]
- Kaufman L and Rousseeuw P J. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley and Sons: 111-112 [DOI: 10.1002/9780470316801]
- Krishnamachari S and Chellappa R. 1997. Multiresolution gauss-Markov random field models for texture segmentation. *IEEE Transactions on Image Processing*, 6(2): 251-267 [DOI: 10.1109/83.551696]
- Krizhevsky A, Sutskever I and Hinton G E. 2012. ImageNet classification with deep convolutional neural networks//*Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe: ACM: 1097-1105
- Laptev D, Savinov N, Buhmann J M and Pollefeys M. 2016. TI-POOLING: transformation-invariant pooling for feature learning in Convolutional Neural Networks//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV: IEEE: 289-297 [DOI: 10.1109/CVPR.2016.38]
- Levinshstein A, Stere A, Kutulakos K N, Fleet D J, Dickinson S J and Siddiqi K. 2009. TurboPixels: fast superpixels using geometric flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12): 2290-2297 [DOI: 10.1109/TPAMI.2009.96]
- Li S Z. 1994. Markov random field models in computer vision//*Proceedings of the 3rd European Conference on Computer Vision*. Stockholm, Sweden: Springer-Verlag: 361-370 [DOI: 10.1007/BFb0028368]
- Li X W and Wang Y T. 2013. Prospects on future developments of quantitative remote sensing. *Acta Geographica Sinica*, 68(9): 1163-1169(李小文, 王祎婷. 2013. 定量遥感尺度效应刍议. *地理学报*, 68(9): 1163-1169) [DOI: 10.11821/dlxb201309001]
- Liasis G and Stavrou S. 2016. Building extraction in satellite images using active contours and colour features. *International Journal of Remote Sensing*, 37(5): 1127-1153 [DOI: 10.1080/01431161.2016.1148283]
- Liu S Y and Deng W H. 2016. Very deep convolutional neural network based image classification using small training sample size//*Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition*. Kuala Lumpur, Malaysia: IEEE: 730-734 [DOI: 10.1109/ACPR.2015.7486599]
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, Massachusetts: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Mnih V. 2013. *Machine Learning for Aerial Image Labeling*. Toronto: University of Toronto
- Moore A P, Prince S J D, Warrell J, Mohammed U and Jones G. 2008. Superpixel lattices//*Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition*. Anchorage, AK: IEEE: 1-8 [DOI: 10.1109/CVPR.2008.4587471]
- Ren X F and Malik J. 2003. Learning a classification model for segmentation//*Proceedings of the 9th IEEE International Conference on Computer Vision*. Nice, France: IEEE: 10-17 [DOI: 10.1109/ICCV.2003.1238308]
- Rignot E and Chellappa R. 1992. Segmentation of polarimetric synthetic aperture radar data. *IEEE Transactions on Image Processing*, 1(3): 281-300 [DOI: 10.1109/83.148603]
- Romero A, Gatta C and Camps-Valls G. 2016. Unsupervised deep feature extraction for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 54(3): 1349-

- 1362 [DOI: 10.1109/TGRS.2015.2478379]
- Rouse J W, Hass R H, Schell J A, Deering D W and Harlan J C. 1974. Monitoring the vernal advancement and retrogradation (green-wave effect) of natural vegetation. NASA/GSFC Type III Final Report. Greenbelt, MD
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z H, Karpathy A, Khosla A, Bernstein M, Berg A C and Li F F. 2015. ImageNet large scale visual recognition challenge. International Journal of Computer Vision, 115(3): 211-252 [DOI: 10.1007/s11263-015-0816-y]
- Saito S, Yamashita T and Aoki Y. 2016. Multiple object extraction from aerial imagery with convolutional neural networks. Journal of Imaging Science and Technology, 60(1): 10402-1-10402-9 [DOI: 10.2352/J.ImagingSci.Technol.2016.60.1.010402]
- Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R and LeCun Y. 2013. OverFeat: integrated recognition, localization and detection using convolutional networks. arXiv:1312.6229
- Sharma A, Liu X W, Yang X J and Shi D. 2017. A patch-based convolutional neural network for remote sensing image classification. Neural Networks, 95: 19-28 [DOI: 10.1016/j.neunet.2017.07.017]
- Shi J B and Malik J. 2000. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8): 888-905 [DOI: 10.1109/34.868688]
- Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition//Proceedings of 2015 ICLR. San Diego, California: [s.n.]: 1409-1556
- Socher R, Lin C C Y, Ng A Y and Manning C D. 2011. Parsing natural scenes and natural language with recursive neural networks//Proceedings of the 28th International Conference on Machine Learning. Bellevue, USA: Omnipress: 129-136
- Volpi M and Tuia D. 2017. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing, 55(2): 881-893 [DOI: 10.1109/TGRS.2016.2616585]
- Wickham H. 2011. The split-apply-combine strategy for data analysis. Journal of Statistical Software, 40(1): 1-29
- Yang Y and Newsam S. 2010. Bag-of-visual-words and spatial extensions for land-use classification//Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. San Jose, California: ACM: 270-279 [DOI: 10.1145/1869790.1869829]
- Yu Q Y and Claudi D A. 2006. Filament preserving model (FPM) segmentation applied to SAR sea-ice imagery. IEEE Transactions on Geoscience and Remote Sensing, 44(12): 3687-3694 [DOI: 10.1109/TGRS.2006.885046]
- Zhang H Y, Wang L W and Chen Y X. 2013. Research progress of probabilistic graphical models: a survey. Journal of Software, 24(11): 2476-2497(张宏毅, 王立威, 陈瑜希. 2013. 概率图模型研究进展综述. 软件学报, 24(11): 2476-2497) [DOI: 10.3724/SP.1001.2013.04486]

Extraction of buildings from remote sensing imagery based on multi-scale SLIC-GMRF and FCNSVM

JING Ran^{1,2,3}, GONG Zhaoning^{1,2,3}, ZHU Wending^{1,2,3}, GUAN Hongliang^{1,2,3}, ZHAO Wenji^{1,2,3}, ZHANG Tao⁴

1. College of Resources Environment & Tourism, Capital Normal University, Beijing 100048, China;

2. Key Laboratory of 3D Information Acquisition and Application of Ministry, Beijing 100048, China;

3. Key Laboratory of Resources Environment and GIS of Beijing Municipal, Beijing 100048, China;

4. Satellite Surveying and Mapping Application Center, State Bureau of Surveying and Mapping, Beijing 100048, China

Abstract: The extraction of buildings from remote sensing imagery has an important application value. However, high-resolution images contain detailed information and complex features that hinder the difficulty of building extraction process.

To address this problem, we propose a building extraction method of building extraction based on multi-scale SLIC-GMRF and FCNSVM that demonstrates an improved ability of extracting buildings from high-resolution remote sensing images to some extent. First, a multi-scale SLIC-GMRF segmentation algorithm is applied to determine the initial building area, and then the advantages of the FCN neural network in semantic segmentation are utilized to extract the building features. Second, the extracted building features are used to train an SVM classifier to refine the building extraction results of building.

The results of three control experiments and two comparative tests reveal that the SLIC segmentation algorithm affects the initial segmentation results, the SVM classifier affects the extraction of building details, and the FCN features influence the performance of the SVM classifier. The precision rate, recall rate, and quality index of the proposed method are all better than the compared methods.

The following conclusions can be drawn from the experimental results. For the study area with clear features and minimal obstructions, the proposed method can effectively extract buildings from an image. This method can also obtain ideal results for areas with a complex distribution of buildings can also get ideal results.

Key words: remote sensing, building extraction, image segmentation, FCN neural network, SVM, high-resolution remote sensing image

Supported by: Demonstration System of High-resolution Remote Sensing Application in Surveying and Mapping (No.AH1601-8)