

DOI:10.13203/j.whugis20200445



文章编号:1671-8860(2021)05-0659-13

街景影像下的临街建筑风格映射及地图生成方法

徐 虹¹ 王禄斌² 方志祥² 何明辉¹ 侯学成¹ 左 亮¹
管昉立² 熊 策¹ 龚毅宇¹ 庞晴霖¹ 张 涵¹
孙树藤¹ 娜迪热·艾麦尔¹

1 武汉科技大学城市建设学院,湖北 武汉,430065

2 武汉大学测绘遥感信息工程国家重点实验室,湖北 武汉,430079

摘要:精细化的城市建筑风格地图已成为古建筑保护、城市规划、旅游资源开发的重要参考依据。但城市建筑众多,信息采集困难,仅靠人工难以实现成图,因此提出了面向街景影像建筑区域匹配的建筑风格地图生成方法。首先,在提取特征建筑风格影像的基础上,结合球形全景影像的空间几何约束和图像特征,通过匹配同名建筑区域构建双像建筑区域点位映射;然后,利用街景采集点到建筑俯视轮廓的方位范围,提出单像建筑区域方位映射,建立街景建筑区域与单体建筑俯视轮廓的空间匹配关系;最后,综合判定各单体建筑的风格属性,生成精细尺度的建筑风格地图。实验结果表明,基于单、双像位置映射的建筑区域匹配正确率分别达80.3%和85.1%,且19类建筑风格地图的分类精确率为55.1%,召回率为76.4%,在一定程度上能反映大范围的城市建筑风格的地理分布特征。

关键词:街景影像;建筑风格分类;深度学习;街景影像匹配;建筑视觉定位

中图分类号:P283;P208 文献标志码:A

每个地域往往具有特定的建筑风格特征,精细化探究建筑风格的地理分布有利于历史古建筑的保护、特色旅游资源的开发以及城市建筑区的科学规划。然而城市建筑数量多、区域广,人工采集和调查不满足大范围作业的需求。近年来,以谷歌为代表的互联网公司推出城市街景影像(street view images, SVI)。街景影像分辨率高,含有全方位的城市街道景观以及精确的位置和姿态信息,为探究城市建筑风格的地理分布提供了可能。

街景建筑区域的风格识别与地图匹配是生成城市建筑风格地图的技术难点。一方面,建筑风格特征复杂多样,图像分类困难。相比传统分类模型^[1-4],卷积神经网络^[5](convolutional neural network, CNN)在建筑风格分类上的精度更高。另一方面,建筑图像的跨视角匹配成为建筑地理定位和土地利用分类^[6]等领域的重要手段,现有

研究主要从单一建筑的空间几何特征考量^[7-11],不适合处理场景复杂、建筑众多的全景像片。因此,本文借助深度学习实现街景建筑区域的风格分类,并建立建筑影像与建筑俯视轮廓的映射关系,精细化探究大范围城市建筑风格地图的生成方法。

1 面向街景影像建筑区域匹配的建筑风格地图生成方法

更快的基于区域CNN(faster region-based CNN, Faster R-CNN)目标检测模型^[12]可以提取各类风格的建筑目标区域。在此基础上,采用双像建筑区域点位映射方法,通过匹配相邻两张街景影像的同名建筑区域来估算每个单体建筑的实地位置,建立与建筑轮廓的点位映射关系。针对无同名建筑区域的建筑影像,本文提出了单像建筑区域方位映射方法,通过街景与地图中建筑

收稿日期:2020-08-25

项目资助:国家自然科学基金(41771473)。

第一作者:徐虹,博士,副教授,主要从事城乡规划与设计、城市与建筑遗产保护、数字城市与建筑等方面的研究。xuhong@wust.edu.cn

通讯作者:王禄斌,硕士生。lbwang@whu.edu.cn

物方位角的对应关系实现单体建筑影像与对应建筑轮廓的匹配。对成功匹配的建筑轮廓获取其建筑风格属性,生成城市建筑风格地图。本文方法的流程见图1。图1中,PASCAL VOC是国际顶级的计算机视觉竞赛提供的标准格式的目标检测数据集。

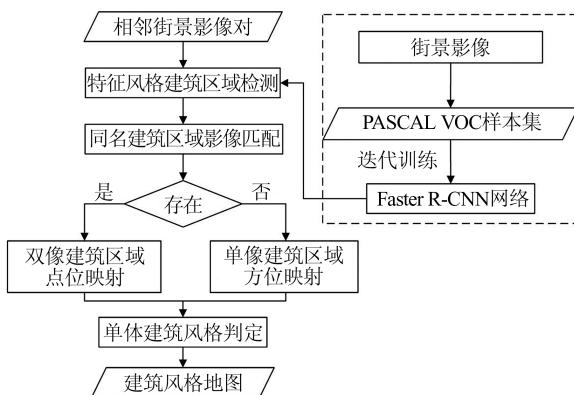


图1 建筑风格地图生成方法流程图

Fig.1 Flowchart of Architectural Style Map Production Method

1.1 相邻街景中同名建筑区域的匹配方法

同名建筑区域是指相邻两张街景中代表同一建筑物的影像区域,是进行双像建筑区域点位映射的必要条件(见图2)。本文首先通过建筑影像的纹理和颜色特征剔除相似度小的单体建筑区域影像对,然后对少量的候选影像对进行核线约束下的加速稳健特征^[13](speeded up robust features, SURF)匹配,得到最佳同名建筑区域及核线距离最近的一个同名点对,用于前方交会定位,匹配流程如图3所示。



图2 同名建筑区域示例

Fig.2 An Example of Two Corresponding Architectural Images

街景单体建筑影像的纹理特征采用不同方向上的灰度共生矩阵^[14](gray level co-occurrence matrix, GLCM)的熵、二阶矩、对比度和逆差矩表示,本文选取0°、45°、90°和135°这4个方向的灰度共生矩阵,构成16维的纹理特征向量。色调、饱和度、亮度(hue, saturation, value, HSV)颜色空间的低阶矩能很好地反映图像明暗程度、颜色分布

范围和分布对称性,共生成9维的颜色特征向量。将两个特征向量融合,得到代表建筑区域的25维全局特征。

两个单体建筑区域特征向量的差向量可作为判定相似程度的特征向量。首先,将差向量作为后验概率的支持向量机^[15](posteriori probability support vector machine, PPSVM)的输入,计算同名建筑区域的概率值,将概率值大于阈值的同类风格建筑区域对作为候选匹配对。然后,在候选匹配对中,以左片的单体建筑区域为基准,分别与右片的每个单体建筑区域进行球面核线约束^[16]下的SURF匹配,选取同名点核线距离均值最小的右片建筑区域作为最佳匹配对。最后,在每个最佳匹配对中,取核线距离最小的一对同名点,通过前方交会得到建筑物的坐标信息。

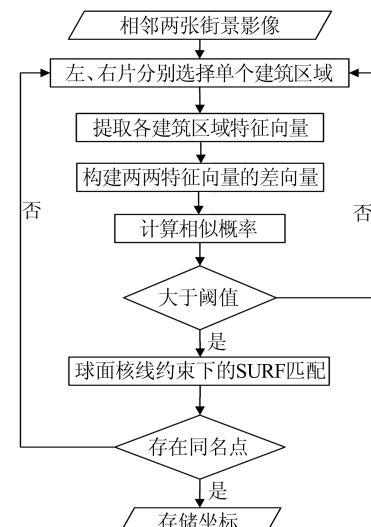


图3 同名建筑区域匹配流程图

Fig.3 Flowchart of Matching Two Corresponding Architectural Images

街景建筑影像形变较大,存在同名点匹配困难、误匹配较多的问题。对此,本文通过检验左、右片单体建筑影像区域中心点的核线距离是否小于阈值来辅助判别是否为同名建筑区域。

1.2 街景影像中建筑区域的位置映射

建立街景影像中建筑区域与地图中建筑轮廓的对应关系,是生成建筑风格地图的关键。本文提出了双像建筑区域点位映射和单像建筑区域方位映射方法,流程如图4所示。

1.2.1 双像建筑区域点位映射

双像建筑区域点位映射需要满足两个条件:(1)相邻两张街景影像存在同名单体建筑区域;(2)两张影像可以完成相对定向。图5显示了同名建筑区域点位映射的过程。图5中, P_1 和 P_2 为

同名建筑区域的最佳同名点对; S_1 和 S_2 是相邻两张街景影像的摄影中心, 分别构成球形全景坐标

系 $S_1-X_1Y_1Z_1$ 和 $S_2-X_2Y_2Z_2$; P 是 P_1 和 P_2 对应的前方交会点。

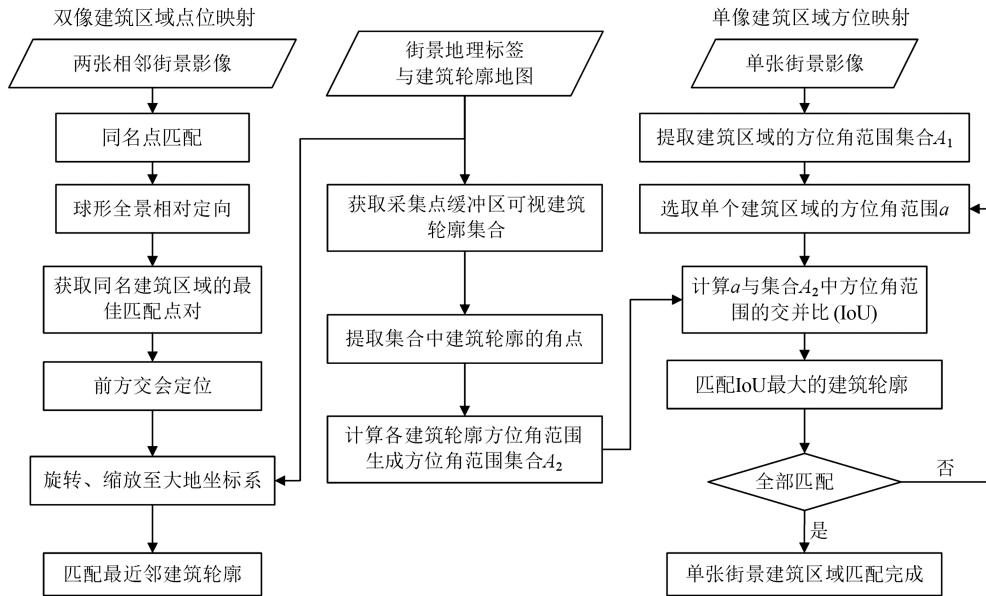


图 4 街景建筑区域映射方法流程图

Fig.4 Flowchart of Building Outline Mapping Methods Based on Architectural Area in Street View

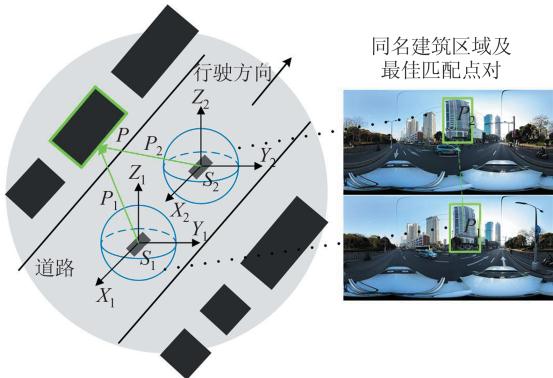


图 5 双像建筑区域点位映射示意图

Fig.5 Location Mapping Method Based on a Panoramic Image Pair

完成两张街景的相对定向需要将平面街景影像转换成球形全景影像^[17-18]。首先, 采用随机抽样一致性 (random sample consensus, RANSAC) 的 8 点法^[16, 19], 由式(1)解算相邻两张街景的旋转矩阵和平移向量。然后, 将同名点对进行前方交会, 利用街景影像的坐标和姿态参数完成绝对定向^[17], 得到对应单体建筑的实地坐标。

$$\begin{cases} P_2 \cdot E \cdot P_1 = 0 \\ E = [T] \times \cdot R_s \end{cases} \quad (1)$$

式中, E 为本质矩阵; P_1 和 P_2 表示至少 8 对同名点球面坐标的矩阵; $[T] \times$ 表示平移向量 T 的反对称矩阵; R_s 为像对的旋转矩阵。

由于无法进行严格的绝对定向, 估算的单体建筑位置点与建筑轮廓存在偏差。首先, 通过限制前方交会距离小于 150 m 和交会角大于 10° 提高定位精度^[20-21]。然后, 采用空间近邻分析, 使每个位置点匹配距离最近的单体建筑轮廓, 且限制匹配距离小于 10 m, 提高单体建筑的匹配准确率。

1.2.2 单像建筑区域方位映射

当球形全景像空间坐标系旋转至与地图坐标系平行时, 单体建筑区域的方位范围与街景采集点到对应建筑轮廓的方位范围相同。图 6 为建筑区域方位映射示意图, S_1 为街景影像的摄影中心, 构成球形全景坐标系 $S_1-X_1Y_1Z_1$ 。

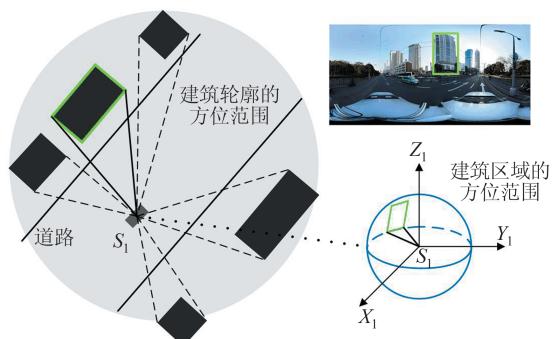


图 6 单像建筑区域方位映射示意图

Fig.6 Diagram of Azimuth Mapping Method Based on a Panoramic Image

在实际场景中, 检测出的单体建筑区域范围存在偏差, 地图中对应的建筑俯视轮廓也可能存

在位置偏移,导致两个方位范围不严格相同。定义两个方位范围的交并比(intersection over union, IoU),用于描述单体建筑区域与建筑俯视轮廓的方位相似度(见图7)。

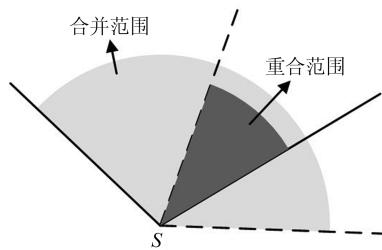


图7 两个方位范围交并比的定义

Fig.7 Definition of Two Azimuth Coverage's IoU

交并比 α 越接近1,则相似度越大;越接近0,则相似度越小。计算公式为:

$$\alpha = \frac{I[(a_i, b_i), (c_j, d_j)]}{U[(a_i, b_i), (c_j, d_j)]} \quad (i, j = 1, 2, 3 \dots) \quad (2)$$

式中, (a_i, b_i) 表示第*i*个建筑区域的方位角范围; (c_j, d_j) 表示街景采集点到第*j*个建筑轮廓的方位角范围; I 表示两个方位范围的重合角度; U 表示两个方位范围的合并角度。

所有单体建筑区域的方位范围构成集合 $A_1 = \{(a_i, b_i) | i = 1, 2, 3 \dots\}$ 。采集点到单个建筑

轮廓的方位范围可以由采集点到建筑轮廓的两个边界方位角顺时针组成(见图8),获取缓冲区内所有单体建筑轮廓的方位角范围,构成集合 $A_2 = \{(c_j, d_j) | j \in B\}$, B 表示缓冲区内的单体建筑轮廓集合。利用式(2)选取IoU最大的匹配对,完成单像建筑区域方位映射。

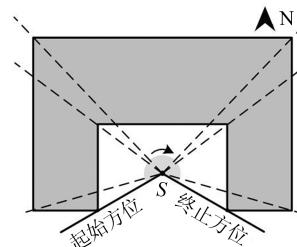


图8 点到建筑俯视轮廓的方位范围示意图

Fig.8 Diagram of Azimuth Coverage from One Position to Building Outline

1.3 基于街景建筑区域匹配的单体建筑风格判定

图9描述了单个建筑轮廓匹配多个不同建筑区域影像的情形,其中 S_i ($i = 1, 2 \dots 6$)代表各街景采集点位置, P_1 和 P_2 代表两个建筑区域的映射点位置。单个建筑轮廓可能映射多类建筑风格(简称多映射问题),故需要确定每个建筑轮廓的唯一风格属性。

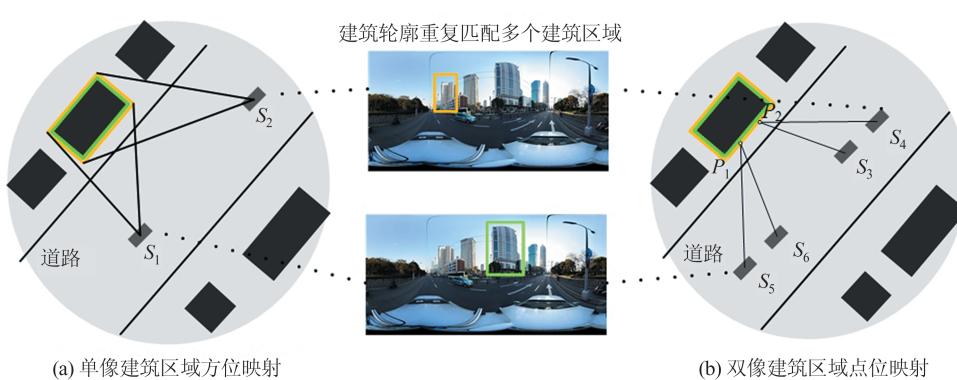


图9 单体建筑轮廓匹配多个建筑区域影像的示意图

Fig.9 Diagram of One Building Outline Matching Multiple Architectural Images

对于双像建筑区域点位映射结果,每个建筑轮廓可能匹配到多个建筑区域位置点。针对一个单体建筑轮廓,将匹配到的每类建筑风格的位置点各构成一个集合,考虑每个集合中所有位置点的数量 m_1 、Faster R-CNN分类概率均值 m_2 、到建筑轮廓的距离均值 m_3 以及用于交会的同名点对的核线距离均值 m_4 ,用这4个参数构成每类建筑风格的评价指标,且权重相同。当一类建筑风格的 m_1 和 m_2 越大, m_3 和 m_4 越小,该类风格的得分越高。将指标归一化、正向化后,构成的标准矩阵 M 为:

$$M = \begin{bmatrix} m_{1,1} & m_{1,2} & m_{1,3} & m_{1,4} \\ m_{2,1} & m_{2,2} & m_{2,3} & m_{2,4} \\ \vdots & \vdots & \vdots & \vdots \\ m_{19,1} & m_{19,2} & m_{19,3} & m_{19,4} \end{bmatrix} \quad (3)$$

式中, $m_{i,j}$ 表示第*i*类建筑风格的第*j*个指标的标准化值。

结合优劣解距离方法^[22](technique for order preference by similarity to an ideal solution, TOPSIS),第*i*类建筑风格的得分计算式如式(4)所示,取得分最高的建筑风格为该单体建筑轮廓的风格属性 s_i 。

$$s_i = \frac{\sqrt{\sum_{j=1}^4 (m_j^+ - m_{i,j})^2}}{\sqrt{\sum_{j=1}^4 (m_j^+ - m_{i,j})^2} + \sqrt{\sum_{j=1}^4 (m_j^- - m_{i,j})^2}} \quad (4)$$

式中, m^+ 为 4 维向量, 且 $m^+ = [\max(m_{1,1}, m_{2,1} \dots m_{19,1}), \max(m_{1,2}, m_{2,2} \dots m_{19,2}) \dots]$; m^- 也为 4 维向量, 且 $m^- = [\min(m_{1,1}, m_{2,1} \dots m_{19,1}), \min(m_{1,2}, m_{2,2} \dots m_{19,2}) \dots]$ 。

针对单像建筑区域方位映射结果, 考虑每个建筑区域影像与该建筑俯视轮廓的方位交并比和建筑区域影像的风格分类概率, 利用式(4)确定建筑轮廓的唯一风格属性。

2 建筑风格映射及地图生成实验与结果

2.1 实验数据与环境

本文对北京、上海、武汉、西安 4 座城市进行

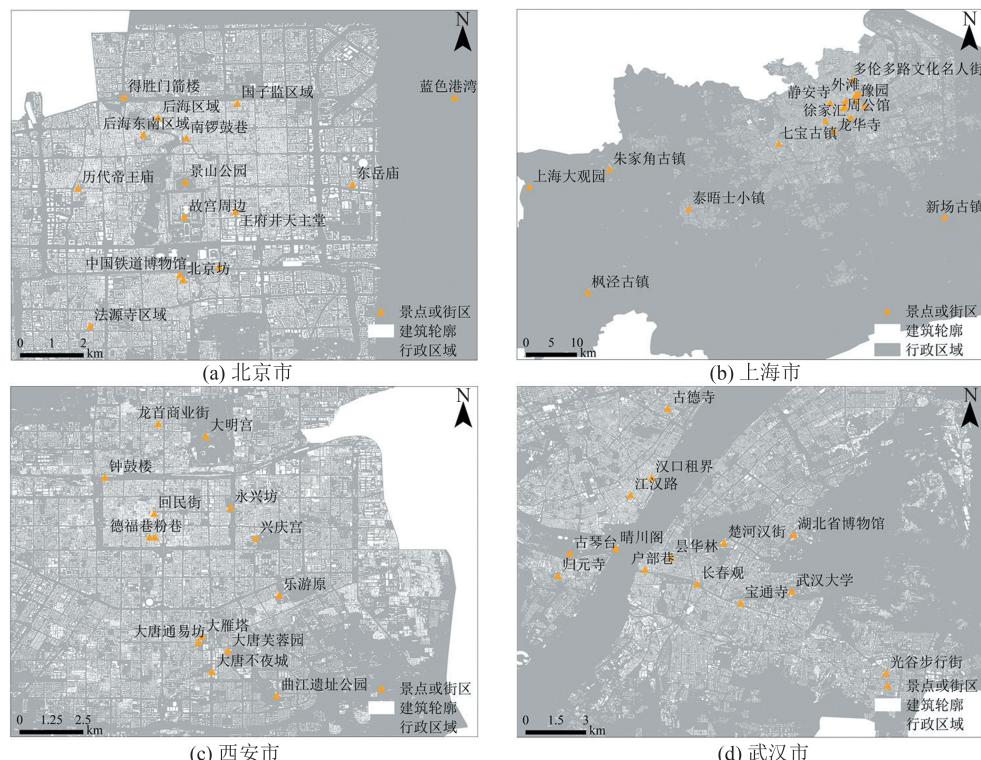


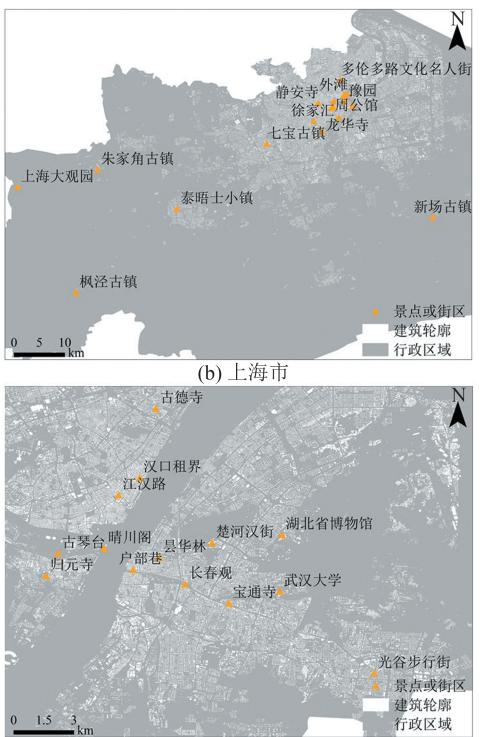
图 10 实验区域及代表性街区或景点的位置分布

Fig.10 Experimental Region and Location Distribution of Representative Blocks or Scenic Spots

为训练匹配同名建筑区域的支持向量机模型, 人工截取 200 对同名建筑区域影像, 每对同名建筑区域影像作为一个正样本, 共获得 200 个正样本数据。同时, 任意不同名建筑区域影像两两配对, 共组成 200 个负样本。训练集和测试集的

建筑风格地图生成实验, 实验区域选取各类建筑风格分布密集的中心城区, 并获取了 2017 年间共 73 420 张百度街景影像, 以及共含约 26.29 万个建筑俯视轮廓的矢量地图数据(.shapfile 格式)。每张街景影像分辨率为 2048×1024 像素, 相邻两张影像的实地距离为 8~20 m, 两个街景影像对的实地距离为 50~150 m, 每张街景影像带有地理信息标签, 包括影像拍摄地的经纬度、正北方向在影像中的方位角、影像的拍摄姿态信息和相邻街景的唯一标识符。

为了训练检测街景建筑风格的 Faster R-CNN 网络, 收集了 4 座城市中含有 19 类建筑风格的 65 个代表性景点或街区, 如图 10 所示。人工标注生成 7 296 张影像的 PASCAL VOC 样本集, 训练样本和测试样本的数量比例为 4:1。图 11 显示了各类建筑风格的原始标定数量, 为了使训练样本数量均衡, 将 12 类数量较少的建筑区域影像水平翻转, 使样本数量增加一倍。



数量比例为 3:1。实验所采用的计算硬件配置为 Intel 8 GB CPU, NVIDIA 1660Ti 6 GB GPU, 算法由 Python 3.6 实现, 深度学习框架采用 TensorFlow-GPU 1.13, 机器学习框架利用 scikit-learn 库, 图像处理通过 OpenCV 库完成。

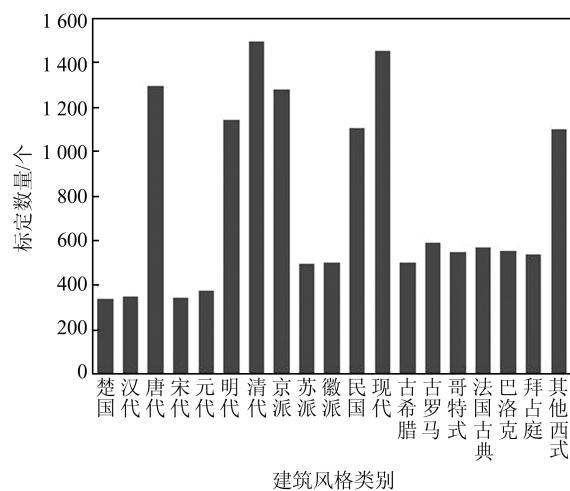


图 11 各类建筑风格的原始标定数量

Fig.11 Number of Calibration of Different Architectural Styles

2.2 匹配与位置映射精度

2.2.1 各类建筑风格的检测精度

Faster R-CNN 网络的参数设置为: 权重衰减为 0.0005, 学习率为 0.001, 动量参数为 0.9, 学习率衰减参数为 0.1, 区域建议网络层生成的候选检测框的数量为 300, IoU 为 0.7, 网络迭代次数为 50 000 次。采用测试集上的平均精度值(average precision, AP)作为 Faster R-CNN 的精度指标, 结果见表 1, 检测结果的总体精度(mean average precision, mAP)为各建筑风格类别检测的 AP 值

的算术平均值。IoU 设定为 0.7 时的测试集检测结果示例见图 12, 图 12 中记录了建筑风格类型和检测概率值。

表 1 各类风格建筑区域的检测精度表

Tab.1 Detection Precision of Architectural Area of Different Styles on Test Set

| 建筑区域的风格类别 | AP/% |
|------------|-------|
| 战国时期楚国建筑风格 | 69.23 |
| 汉代建筑风格 | 57.17 |
| 唐代建筑风格 | 89.43 |
| 宋代建筑风格 | 86.73 |
| 元代建筑风格 | 73.61 |
| 明代建筑风格 | 78.53 |
| 清代建筑风格 | 79.59 |
| 京派民居 | 88.53 |
| 苏派民居 | 72.19 |
| 徽派民居 | 84.37 |
| 民国民居 | 78.61 |
| 现代建筑风格 | 84.13 |
| 古希腊建筑风格 | 76.97 |
| 古罗马建筑风格 | 73.37 |
| 哥特式建筑风格 | 63.17 |
| 法国古典风格 | 43.85 |
| 巴洛克建筑风格 | 48.29 |
| 拜占庭建筑风格 | 89.20 |
| 其他西式风格 | 65.40 |
| mAP | 73.81 |



图 12 测试集的建筑区域检测结果示例

Fig.12 Selected Examples of Architectural Area Detection Results on Test Set

2.2.2 同名建筑区域匹配精度

支持向量机选取高斯核函数。通过网格搜

索确定模型最佳参数, 经过十折交叉验证得到测试集上的 AP 为 92%。最佳参数值包括核函数次

数为 1.0, Gamma 为 0.001, 最大迭代次数为 1 000。为真实反映匹配方法的精度, 随机选取 100 个相邻街景像对, 利用 Faster R-CNN 检测建筑区域, 得到 808 个建筑区域匹配对, 经人工核验, 其中有 206 对真实同名建筑区域, 如图 13 所示。利用相同的检验数据, 将本文方法与仅采用附加核线约束的 SURF 匹配结果的精确率、召回率和 F1 分数^[23]进行对比, 结果如表 2 所示。

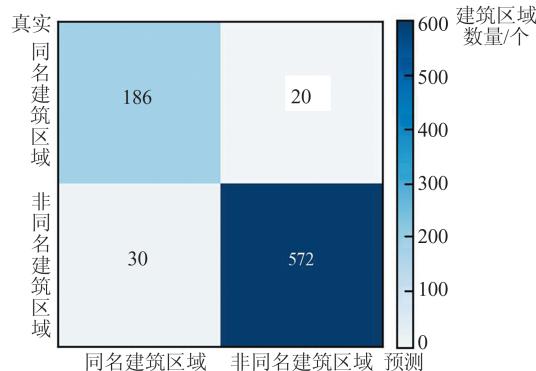


图 13 同名建筑区域匹配结果的混淆矩阵

Fig.13 Confusion Matrix of Matching Results with the Same Name Architectural Area

由表 2 可知, 本文方法的 F1 分数相比 SURF 算法提高 10.1%, 匹配一个建筑区域对的平均耗时减少 25.4%。召回率下降说明 PPSVM 在筛选候选匹配对时, 存在将同名建筑区域误判成非同名建筑区域的情况, 而核线约束下的 SURF 匹配能通过定量的核线距离关系辅助判别两建筑区域是否同名, 在一定程度上减少对影像特征的依赖, 但核线约束不具有唯一性, 精确率降低。因此, 融合影像特征和核线几何约束可以提高建筑区域匹配的精确率和时间效率。图 14 展示了相邻两张街景进行同名建筑区域匹配的过程, 左、右两张街景记录了建筑风格类型和检测概率值。

表 2 两种同名建筑区域匹配方法的精度对比结果

Tab.2 Accuracy Comparison of Two Matching Methods

| 同名建筑区域匹配方法 | 精确率/% | 召回率/% | F1 分数 | 耗时/ms |
|----------------|-------|-------|-------|-------|
| 本文方法 | 86.1 | 90.3 | 0.882 | 180.1 |
| 核线约束下的 SURF 匹配 | 66.7 | 94.2 | 0.781 | 241.6 |



图 14 相邻两张街景的同名建筑区域匹配过程

Fig.14 Matching Process of Two Corresponding Architectural Areas in an Image Pair

2.2.3 两种位置映射方法对比与分析

为检验两种位置映射方法的准确率,实验选取150个不同的单体建筑区域影像,分别通过本文的两种方法映射至建筑俯视轮廓。结合街景影像,人工统计正确映射的单体建筑数量,得到表3的结果。从表3可以看出,对于单像方位映射方法,单体建筑区域与对应俯视轮廓的IoU均值为0.682,匹配正确率达80.3%,单个建筑区域完成方位映射的平均时间为0.017 s;双像点位映射得到的单体建筑位置点与对应俯视轮廓的匹配距离均值为5.498 m,匹配正确率为85.1%,一对同名建筑区域完成点位映射的平均时间为2.741 s。两种方法的时间效率对比如图15所示。

表3 两种位置映射方法准确率的对比结果

Tab.3 Accuracy Results of Two Mapping Methods

| 位置映射方法 | 正确映射/个 | 错误映射/个 | 映射失败/个 |
|--------|--------|--------|--------|
| 单像方位映射 | 106 | 26 | 18 |
| 双像点位映射 | 63 | 11 | 76 |

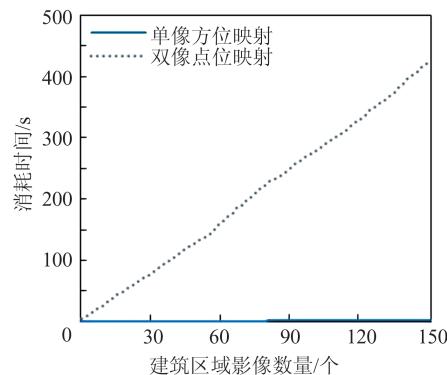


图15 两种映射方法的耗时对比

Fig.15 Time Consumption Comparison of Two Mapping Methods

从表3和图15可以看出,双像点位映射方法的正确率较高,但满足映射条件的街景像对的比例仅为49.33%,丢失了较多不同风格的建筑区域影像;同时,双像点位映射方法涉及同名建筑区域匹配和相对定向,计算耗时长,不适合处理海量街景数据。单像方位映射方法在保证一定精度的同时,成功率高,耗时短,但不适合处理过小的建筑区域。因此,在实际处理大量街景数据时,一般采用单像方位映射匹配建筑轮廓。当检测到过小的单体建筑区域时,可采用双像点位映

射进行辅助定位匹配。

表4显示了在同一个区域两种映射方法导致多映射的单体建筑数量。在两种映射方法下,存在多映射的建筑数量占比分别为42.9%和23.4%。可见,单像方位映射出现多映射的比例高于双像点位映射,主要因为单像方位映射的成功率高,更多单体建筑影像匹配到建筑轮廓,导致出现较多的多匹配现象,这也进一步说明了建筑风格综合判定方法的必要性。

表4 存在多映射问题的单体建筑数量

Tab.4 Number of Buildings with Multiple Mapping Problems

| 位置映射方法 | 成功映射的建筑/个 | 存在多映射的建筑/个 |
|--------|-----------|------------|
| 单像方位映射 | 13 522 | 5 805 |
| 双像点位映射 | 9 595 | 2 245 |

2.3 建筑风格地图准确率

图16~19为实验生成的北京、西安、上海、武汉4个城市的建筑风格地图。为检验地图的准确率,本文在生成结果中选取了331个单体建筑轮廓,人工标注其建筑风格特征作为先验信息。将实验结果与先验类别进行验证,得到图20所示的建筑风格地图分类精度图。

由于实验区域中没有用于验证的元代、宋代以及法国古典风格的先验建筑轮廓信息,在忽略这3类建筑风格的情况下,建筑轮廓分类精确率的平均值为55.1%,召回率的平均值为76.4%,F1得分均值为0.601。结果表明,单体建筑轮廓的分类平均精确率不高,但能基本反映建筑风格的地理分布。此外,楚国风格、汉代风格、明代风格、民国民居风格、拜占庭风格的精确率远低于平均值。结合Faster R-CNN模型测试结果,分类精确率低的建筑风格对应的训练样本数量少。从建筑风格特征上考量,一方面,同一类建筑风格在不同地域有着内在差异,例如上海和武汉的民国民居在建筑立面外观和建筑结构上存在不同。另一方面,不同类的建筑风格有相似性,在建筑结构或构件上有相同或相似的元素,例如民国民居往往会采用欧式建筑立面,古典主义的欧式建筑会采用相同的窗形或柱式。上述两个因素会导致建筑风格的影像分类较困难,影响建筑风格地图的分类准确率。

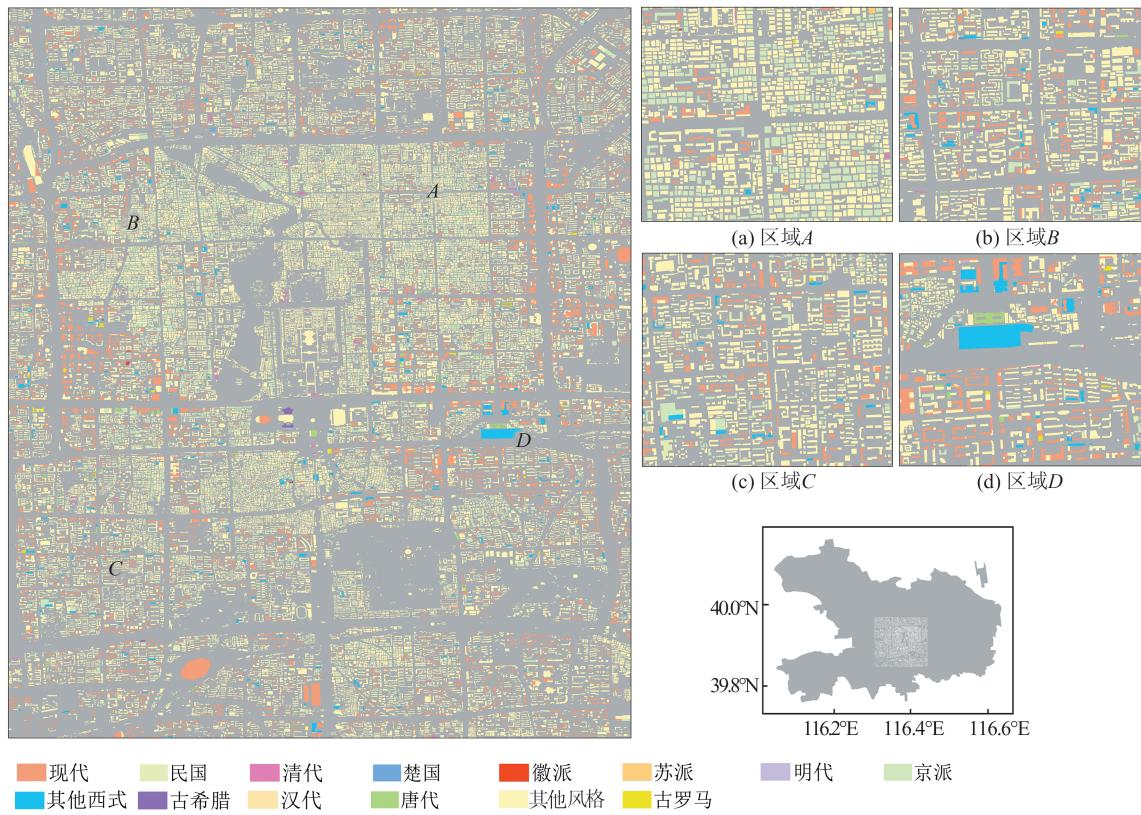


图 16 北京市建筑风格地图
Fig.16 Architectural Style Map of Beijing

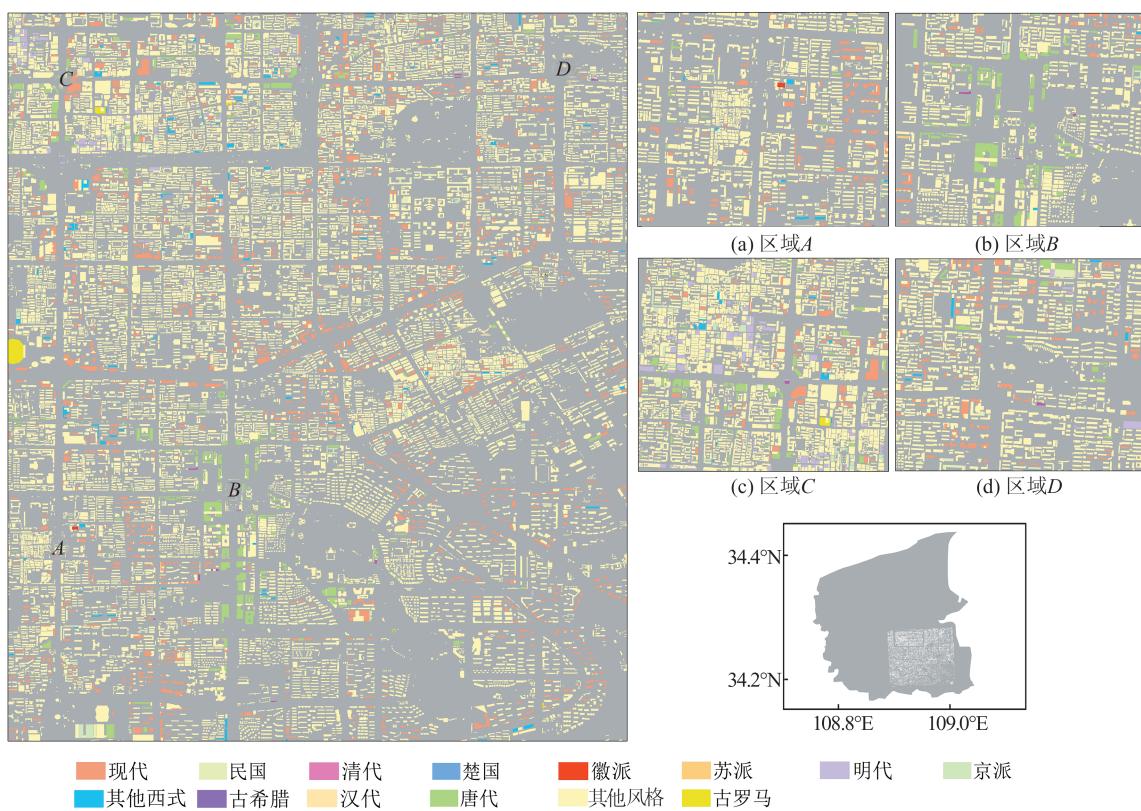


图 17 西安市建筑风格地图
Fig.17 Architectural Style Map of Xi'an

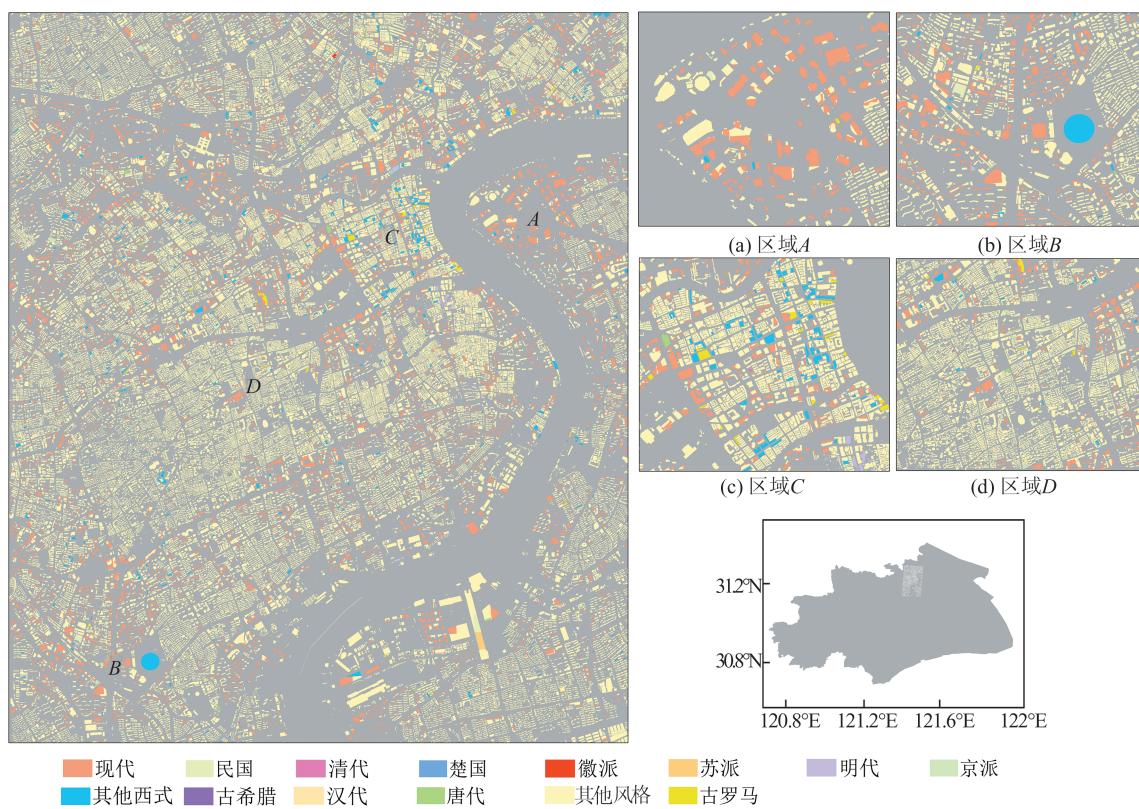


图18 上海市建筑风格地图

Fig.18 Architectural Style Map of Shanghai

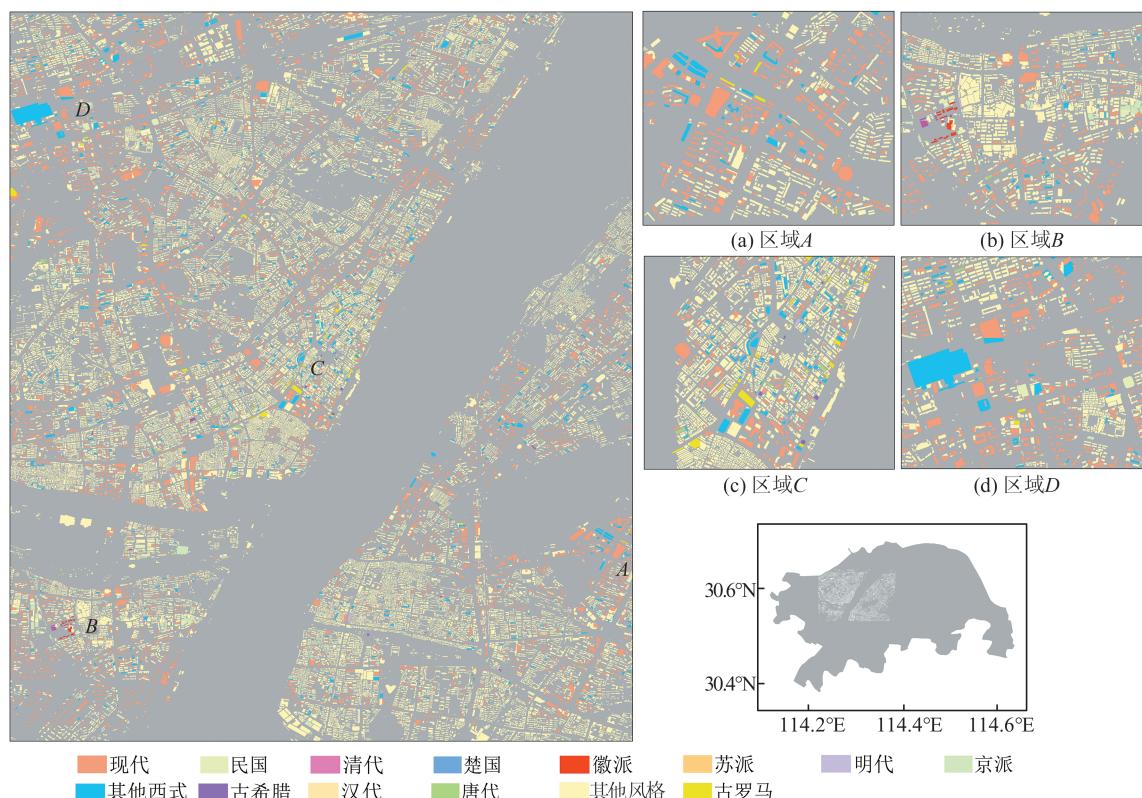


图19 武汉市建筑风格地图

Fig.19 Architectural Style Map of Wuhan

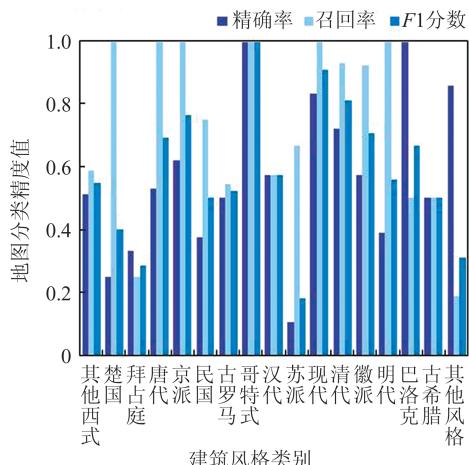


图 20 建筑风格地图分类精度条形图

Fig.20 Bar Chart of Classification Results of Architectural Styles

3 结语

本文以城市街景影像为研究对象,在完成 19 类建筑风格影像提取的基础上,结合球形全景影像的空间几何与图像特征,探究同名建筑区域的匹配方法,建立单、双像建筑区域的位置映射关系,实现街景建筑区域影像与单体建筑俯视轮廓的匹配,实际生成了 4 座城市中心区域的建筑风格地图。在未来仍需进一步提高风格分类精度和地图的精细度。

参 考 文 献

- [1] Goel A, Juneja M, Jawahar C V. Are Buildings Only Instances?: Exploration in Architectural Style Categories [C]//The Eighth Indian Conference on Computer Vision, Graphics and Image Processing, New York, USA, 2012
- [2] Zhang Luming, Song Mingli, Liu Xiao, et al. Recognizing Architecture Styles by Hierarchical Sparse Coding of Blocklets [J]. *Information Sciences*, 2014, 254:141-154
- [3] Zhao Peipei, Miao Qiguang, Song Jianfeng, et al. Architectural Style Classification Based on Feature Extraction Module [J]. *IEEE Access*, 2018, 6: 52 598-52 606
- [4] Shen Jiajie, Pan Li, Hu Xiangyun. Building Detection from High Resolution Remote Sensing Imagery Based on a Deformable Part Model [J]. *Geomatics and Information Science of Wuhan University*, 2017, 42(9): 1 285-1 291(沈佳洁, 潘励, 胡翔云. 可变形部件模型在高分辨率遥感影像建筑物检测中的应用[J]. 武汉大学学报·信息科学版, 2017, 42(9):1 285-1 291)
- [5] Zhao Peipei. Architectural Style Classification Algorithms Research Based on Ensemble Projection and Convolution Neural Network [D]. Xi'an: Xidian University, 2015(赵佩佩. 基于集成投影及卷积神经网络的建筑风格分类算法研究[D]. 西安:西安电子科技大学, 2015)
- [6] Cao Rui, Zhu Jiasong, Tu Wei, et al. Integrating Aerial and Street View Images for Urban Land Use Classification [J]. *Remote Sensing*, 2018, 10(10): 1 553-1 575
- [7] Wolff M, Collins R T, Liu Yanxi. Regularity-Driven Building Facade Matching Between Aerial and Street Views [C]//IEEE Conference on Computer Vision and Pattern Recognition, Washington D C, USA, 2016
- [8] Song Weigang. A Visual Localization Technique Based on Street View and Aerial Image Registration [D]. Suzhou: Soochow University, 2018(宋为刚. 基于街景与航拍图像配准的视觉定位技术[D]. 苏州:苏州大学, 2018)
- [9] Sun Bin, Chen Chen, Zhu Yingying, et al. GEOCAPSNET: Ground to Aerial View Image Geo-Localization Using Capsule Network [C]//IEEE International Conference on Multimedia and Expo, Washington D C, SA, 2019
- [10] Xiong Xi, Research on Fast Single-Image-Based Building Localization with a Smartphone [D]. Beijing: Tsinghua University, 2015(熊曦. 基于智能手机单张照片的建筑物快速定位算法[D]. 北京:清华大学, 2015)
- [11] Chen Yun, Cai Zhongliang, Li Bozhao, et al. A Geo-location Computation Method of Objects in the Photo[J]. *Journal of Geomatics*, 2020, 45(5): 142-145(陈运, 蔡忠亮, 李伯钊, 等. 一种拍摄目标的地理位置标注方法[J]. 测绘地理信息, 2020, 45(5):142-145)
- [12] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6):1 137-1 149
- [13] Bay H, Ess A, Tuytelaars T. Speeded Up Robust Features (SURF) [J]. *Computer Vision and Image Understanding*, 2008, 110(3): 346-359
- [14] Lü Fenghua, Shu Ning, Gong Yan, et al. Regular Building Extraction from High Resolution Image Based on Multilevel-Features [J]. *Geomatics and Information Science of Wuhan University*, 2017, 42(5): 656-660(吕凤华, 舒宁, 龚龑, 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(5):656-660)
- [15] Wu Gaowei, Tao Qing, Wang Jue, et al. Support Vector Machines Based on Posteriori Probability [J]. *Journal of Computer Research and Development*, 2005, 42(2):196-202(吴高巍, 陶卿, 王珏, 等. 基于后验概率的支持向量机[J]. 计算机研究与发展, 2005, 42(2):196-202)
- [16] Xie Donghai, Zhong Ruofei, Wu Yu, et al. Relative Pose Estimation and Accuracy Verification of Spherical Panoramic Image [J]. *Acta Geodaetica et Cartographica Sinica*, 2017, 46(11):1 822-1 829(谢东海, 钟若飞, 吴俣, 等. 球面全景影像相对定向与精度验证[J]. 测绘学报, 2017, 46(11):1 822-1 829)
- [17] Liu Shuai, Chen Jun, Sun Min, et al. Measurable Panorama Construction Based on Binocular Spherical Projective Geometry [J]. *Journal of Computer Aided Design and Computer Graphics*, 2015, 27(4):657-665(刘帅, 陈军, 孙敏, 等. 双球面投影几何可量测全景模型的构建[J]. 计算机辅助设计与图形学学报, 2015, 27(4):657-665)
- [18] Zhang Chunsen, Wang Xiqi, Guo Bingxuan. Space Location of Image in Urban Environments Based on C/S Structure[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(7): 978-983(张春森, 王西旗, 郭丙轩. 城市环境下基于C/S架构
- 的影像空间定位[J]. 武汉大学学报·信息科学版, 2018, 43(7):978-983)
- [19] Wu Yousi. Position and Orientation Estimation of Spherical Panorama Image[D]. Wuhan: Wuhan University, 2017(吴幼丝. 球形全景影像位姿估计[D]. 武汉: 武汉大学, 2017)
- [20] Guan Fangli, Fang Zhixiang, Yu Tao, et al. Detecting Visually Salient Scene Areas and Deriving Their Relative Spatial Relations from Continuous Street-View Panoramas[J]. *International Journal of Digital Earth*, 2020, 13(12):1 504-1 531
- [21] Wang Zhixuan, Zhong Ruofei, Xie Donghai. Automatically Measuring the Coordinates of Streetlights in Vehicle-Borne Spherical Images [J]. *Journal of Image and Graphics*, 2018, 23(9):1 371-1 381(王志旋, 钟若飞, 谢东海. 球面全景影像自动测量路灯坐标的方法[J]. 中国图象图形学报, 2018, 23(9):1 371-1 381)
- [22] Hwang C L, Yoon K, Hwang C L, et al. Multiple Attribute Decision Making[J]. *Lecture Notes in Economics and Mathematical Systems*, 1981, 404(4): 287-288
- [23] Kang Jian, Körner M, Wang Yuanyuan, et al. Building Instance Classification Using Street View Images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, 145: 44-59

Street-Facing Architectural Image Mapping and Architectural Style Map Generation Method Using Street View Images

XU Hong¹ WANG Lubin² FANG Zhixiang² HE Minghui¹ HOU Xuecheng¹ ZUO Liang¹
 GUAN Fangli² XIONG Ce¹ GONG Yiyu¹ PANG Qinglin¹ ZHANG Han¹
 SUN Shuteng¹ NADIRE Aimaier¹

1 School of Urban Construction, Wuhan University of Science and Technology, Wuhan 430065, China

2 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

Abstract: Objectives: Each region has specific characteristics of architectural styles, and a detailed investigation of the geographical distribution of architectural styles is conducive to the protection of historic buildings, the development of special tourism resources and the scientific planning of urban architectural areas. However, the number of urban buildings is large, manual collection and investigation cannot meet the needs of large-scale operations. In recent years, Google and other Internet companies have launched street view images. Street view images are high resolution, containing a full range of urban street views as well as precise location and posture information, which provide a possibility to explore the geographic distribution of urban architectural styles. Therefore, we use deep learning to identify and match the styles of street view building areas, and establish a mapping relationship between the building area images and building outlines,

so as to construct the generation method of a large-scale urban architectural style map in detail. **Methods:** The style identification and map matching of architectural areas in street view images are the key and difficulty in generating urban architectural style maps. Firstly, we extract the building area images of various styles through Faster R-CNN. In order to establish the mapping relationship between building area images and single building outlines, we construct a building location mapping method by matching the same name building area in two adjacent street view images, then the building can be located by forward intersection. Secondly, for the single building image without a same-name area, we also propose a building azimuth mapping method, which combines the spatial azimuth relationship between the street view building area and building outlines in a digital map. The intersection of union (IoU) of the single building image azimuth range and the building outline azimuth range can help match the building area in a street view image and building outlines in a digital map. Finally, Technique for order preference by similarity to an ideal solution is used to determine the unique style attribute of each map building outline to solve the multiple mapping problem of a single building and generate a fine-grained architectural style map. **Results:** The experimental results of the proposed method are as follows: (1) The average accuracy of Faster R-CNN detection of 19 types of architectural style areas on the test set is 73.81%. (2) The accuracy of matching two adjacent street images with the same name architectural area is 86.1%, the recall is 90.3%, and the average time to match an architectural region pair is 180.1 ms, which is 25.4% less than the time using SURF(speeded up robust features) under spherical epipolar geometry constraint and an accuracy improvement of 19.4%; (3) The accuracy of a building location mapping method is 85.1%, the mapping success rate is only 49.33%, and the average time for two corresponding building area to complete location mapping is 2.741 s; the accuracy of the building azimuth mapping method is 80.3%, the mapping success rate is 88.0%, and the average time for a single building area to complete azimuth mapping is 0.017 s. (4) In the test region, the building azimuth mapping method is more likely to cause multiple mapping problems, with 42.9% of the building outlines matching to multiple building images compared to 23.4% for the building location mapping method. (5) By verifying the style attributes of 331 building outlines in a digital map, we obtain a mean classification accuracy of 55.1%, a mean recall of 76.4%, and a mean F1 score of 0.601 for the architectural style maps. **Conclusions:** Under the two architectural area mapping methods, the generation time of architectural style maps is short, and the F1 score of classification is 0.601, which can basically reflect the geographic distribution characteristics of a large range of urban architectural styles. In addition, the regional and similarity of architectural styles is the main reason for the difficulty in classifying architectural style images, which affects the classification accuracy of architectural style maps and can be studied in more depth in the future.

Key words: street view images; architectural style classification; deep learning; street view image matching; building visual localization

First author: XU Hong, PhD, associate professor, specializes in urban and rural planning and design, urban and built heritage preservation.
E-mail: xuhong@wust.edu.cn

Corresponding author: WANG Lubin, postgraduate. E-mail: lbwang@whu.edu.cn

Foundation support: The National Natural Science Foundation of China (41771473).

引文格式: XU Hong, WANG Lubin, FANG Zhixiang, et al. Street-Facing Architectural Image Mapping and Architectural Style Map Generation Method Using Street View Images[J]. Geomatics and Information Science of Wuhan University, 2021, 46(5):659-671. DOI:10.13203/j.whugis20200445(徐虹,王禄斌,方志祥,等.街景影像下的临街建筑风格映射及地图生成方法[J].武汉大学学报·信息科学版,2021,46(5):659-671. DOI:10.13203/j.whugis20200445)