

引文格式:王椰,张新长,姜明,等. CSG-Net:一种融合域适应与视觉基础模型 SAM 的遥感影像建筑物足迹提取方法[J]. 测绘通报,2026(3): 57-61. DOI:10. 13474/j. cnki. 11-2246. 2026. 0310.

CSG-Net:一种融合域适应与视觉基础模型 SAM 的遥感影像建筑物足迹提取方法

王 椰¹,张新长¹,姜 明²,阮永俭¹

(1. 广州大学地理科学与遥感学院,广东 广州 510006; 2. 中山大学地理科学与规划学院,
广东 广州 510006)

摘要:针对深度学习建筑物足迹提取模型在跨平台与跨分辨率应用中因域间分布不一致导致的泛化能力显著下降问题,本文提出了一种跨尺度几何精炼网络(CSG-Net),构建了一个“概率-几何”串联的伪标签精炼框架,旨在提升模型在无标签目标域中的适应性与提取精度。首先,通过计算模型双预测分支的 Jensen-Shannon 散度(JSD),实现对伪标签的不确定性度量与概率加权,以软性方式抑制不可靠区域的噪声;然后,引入基于 segment anything model(SAM)分割结果的几何先验,通过重叠率分析对初始伪标签的边界进行硬性几何修正,从而生成高质量的训练目标。在跨尺度建筑物提取任务上的试验表明,CSG-Net 的交并比(IoU)达到 73.05%,显著优于 Baseline(52.49%)及其他先进域适应方法,验证了本文框架在提升跨域稳健性和提取精度方面的有效性。

关键词:遥感影像;建筑物足迹;语义分割;域适应;segment anything model(SAM)

中图分类号:P237

文献标识码:A

文章编号:0494-0911(2026)03-0057-05

CSG-Net: a method for building footprint extraction from remote sensing images by integrating domain adaptation and the visual foundation model SAM

WANG Ye¹,ZHANG Xinchang¹,JIANG Ming²,RUAN Yongjian¹

(1. School of Geography and Remote Sensing, Guangzhou University, Guangzhou 510006, China; 2. School of Geography and Planing,
Sun at-sen University, Guangzhou University, Guangzhou 510006, China)

Abstract: To address the substantial decline in generalization performance of deep learning models for building footprint extraction across different platforms and spatial resolutions caused by inter-domain distribution inconsistency, this study introduces a cross-scale geometric-refined network (CSG-Net) aimed at enhancing both adaptability and extraction accuracy in unlabeled target domains. The proposed approach establishes a cascaded probabilistic-geometric pseudo-label refinement framework. Firstly, pseudo-label uncertainty is quantified by computing the Jensen-Shannon divergence (JSD) between the dual prediction branches of the model, and a probabilistic weighting scheme is applied to suppress noise in unreliable regions. Subsequently, geometric priors derived from the segmentation results of the segment anything model (SAM) are incorporated to perform explicit geometry-constrained refinement of pseudo-label boundaries based on overlap ratio analysis, thereby generating high-quality training targets. Extensive experiments on challenging cross-scale building extraction tasks demonstrate that CSG-Net achieves an intersection over union (IoU) of 73.05%, markedly surpassing the Baseline (52.49%) and outperforming other state-of-the-art domain adaptation approaches. These findings confirm the effectiveness of the proposed framework in improving cross-domain robustness and extraction accuracy.

Key words: remote sensing imagery; building footprint; semantic segmentation; domain adaptation; segment anything model (SAM)

建筑物足迹作为地理信息的重要组成部分,其精确提取对于光伏潜力评估、灾害损失评估及违章建筑监测等方面具有重要应用价值^[1-3]。随着遥感技术的快速发展,大量高分辨率遥感影像被应用于地物识别,而如何从中准确提取建筑物足迹一直是该领域的研究热点和难点^[4-6]。近年来,基于深度

学习的语义分割方法在建筑物足迹提取方面取得了显著进展,尤其在具备大规模同源、同尺度标注数据的场景下表现出优异性能。然而,这种使用完全监督的方法不仅依赖于庞大的人工标注数据集,还需假设不同数据间存在较小的特征差异。实际上,不同空间分辨率与不同区域影像之间往往存在

收稿日期:2025-10-09

基金项目:国家自然科学基金(42371406;42401518;42571533);广东省自然科学基金(2025A1515011066)

作者简介:王 椰(2001—),男,硕士生,主要研究方向为智慧城市与城市遥感。E-mail:2112301066@e. gzh. edu. cn

通信作者:张新长。E-mail:zhangxc@gzh. edu. cn

显著差异^[7]。在跨平台数据迁移过程中,分辨率不一致与光谱差异严重制约了模型的泛化与可迁移性^[8]。为应对上述挑战,域适应方法受到广泛关注,其核心思想是利用源域的有标签数据辅助目标域的无标签数据进行联合学习,从而在跨域任务中提升模型的适应性与稳健性^[9]。

传统域适应方法多基于对抗学习、自训练策略和对比学习,通过特征对齐或伪标签迭代优化缩小域间分布差异^[10-15]。然而,在建筑物足迹提取任务中,面对更大的尺度差异和复杂的形态背景,这些方法依然面临初始伪标签噪声大,且在迭代自训练中易产生误差累积等问题。文献[16]提出通过双分支预测的 Jensen-Shannon 散度对伪标签进行概率加权,一定程度上抑制了伪标签噪声问题,但未能有效解决因尺度剧变导致的建筑物边界模糊、形态不规则问题。

近年来,视觉基础模型的突破为遥感解译提供了新的研究范式。其中,Meta 提出的 segment anything model (SAM) 具备强大的零样本通用分割能力,能生成边界精准的掩码^[17],为提升伪标签的几何质量提供了可靠的外部先验。然而,SAM 本身不具备语义判别能力,且其固定输入尺寸和全局自注意力机制在处理大幅面遥感影像时面临计算资源消耗大和效率低等挑战^[18]。因此,如何在无标签目标域中,将分割模型生成的语义预测结果与 SAM 所提供的高精度几何先验有效融合,成为一个亟待研究的问题。

基于上述背景,为系统性地应对跨尺度建筑物提取中的伪标签质量低下、模型适应失效及边界精准性 3 大难题,本文结合文献[16]的方法,提出一种新颖的跨尺度几何精炼网络 (cross-scale geometric-refined network, CSG-Net), 在一个经典的单模型自训练范式下,创新性地集成一套多策略协同的优化流程。

1 CSG-Net 方法

1.1 网络结构

CSG-Net 的整体网络架构如图 1 所示。该框架遵循自训练范式,包含源域监督训练与目标域伪标签精炼两个循环迭代的阶段。在源域监督训练阶段,利用带标签的源域数据对基础分割网络进行全监督训练;在目标域伪标签精炼阶段,对象是无标签的目标域图像。首先,利用训练好的网络生成初始伪标签及预测概率图;然后,伪标签依次通过两个核心模块:概率加权模块利用网络双分支的预测差异 Jensen-Shannon divergence (JSD) 评估不确定

性,并对训练损失进行加权,几何修正模块将初始伪标签与 SAM 生成的掩码进行叠加分析,基于重叠率准则优化物体边界;最后,精炼后的高质量伪标签用于后续网络模型的迭代训练。

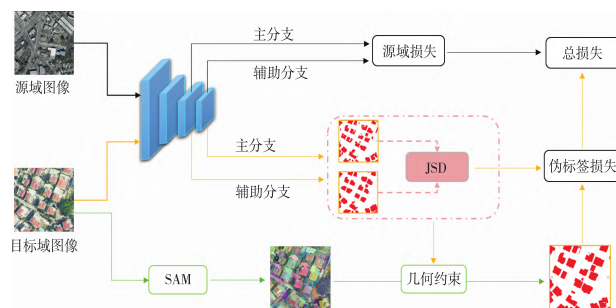


图 1 CSG-Net 网络结构

1.2 基础分割网络

该网络以 UperNet^[19] 作为解码器,并配置 ConvNeXt^[20] 作为编码器以进行特征提取。UperNet 通过结合特征金字塔网络 (FPN) 与金字塔池化模块 (PPM),能够有效地融合网络不同层级的特征,在扩大感受野、捕捉丰富上下文信息的同时,保留精细的空间细节,这对于处理尺寸变化多样的建筑物目标至关重要。

1.3 概率-几何串联伪标签精炼模块

1.3.1 基于预测不一致性的概率加权

采用一个带有主、辅两个预测分支的深度监督网络结构进行预测不一致性的概率加权。在对无标签的目标域样本 x_i^t 进行预测时,主分支 $F_m(x_i^t | \theta)$ 接收较精细的深层特征,而辅助分支 $F_a(x_i^t | \theta)$ 接收较浅层的特征,两个预测分支的预测结果的不一致性被视为模型对该预测的不确定性的体现,并通过 $D_{JS}(x_i^t)$ 进行定量度量。本文采用 JSD 度量两个预测分布之间的不一致性,以此作为不确定性的量化指标。JSD 的计算基于两个分布及其平均分布 M 之间的 KL 散度,其定义如下

$$M = \frac{F_m(x_i^t | \theta) + F_a(x_i^t | \theta)}{2} \quad (1)$$

$$D_{JS}(x_i^t) = \frac{1}{2} D_{KL}(F_m(x_i^t | \theta) \| M) + \frac{1}{2} D_{KL}(F_a(x_i^t | \theta) \| M) \quad (2)$$

式中, D_{KL} 表示 KL 散度; D_{JS} 的值域为 $[0, 1]$, 值越大,表明模型在该位置的预测不确定性越高。基于此不确定度量,引入一个自适应权重 $\exp(-D_{JS}(x_i^t))$ 。该权重与不确定性负相关,使得模型在训练过程中能够动态地聚焦于高置信度的预测,有效抑制不确定性区域产生的噪声。加权后的伪标签损失 $\mathcal{L}_w$ 可表示为 $\mathcal{L}_w(x_i^t) = \exp(-D_{JS}(x_i^t)) \cdot \mathcal{L}_{ce}(x_i^t) + D_{JS}(x_i^t)$ (3)

式中, $\mathcal{L}_{ce}(x_i^t)$ 为标准交叉熵损失。

1.3.2 基于 SAM 几何先验的硬修正

引入一种基于外部几何先验的伪标签修正策略,即几何约束算法。该方法的核心思想为,利用 SAM 生成的精细边界掩膜,对分割模型生成的初始伪标签进行修正,以缓解伪标签边界模糊和形状不规则的问题。如图 2 所示,算法首先接收模型预测的初始伪标签,以及由 SAM 生成的无语义但边界精确的掩膜;然后,遍历每个掩膜并计算其与对应初始伪标签的交并比(IoU)。基于预设的阈值,当 IoU 高于该阈值时,采用 SAM 掩膜替代初始伪标签作为修正结果;否则保留原始伪标签。

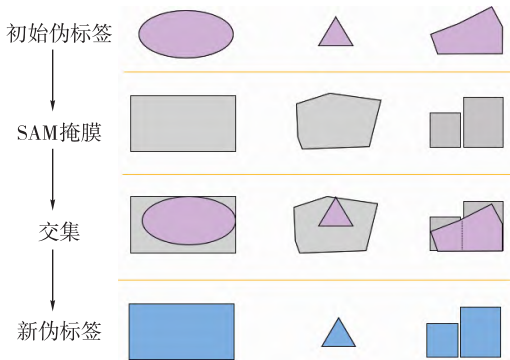


图 2 利用 SAM 对伪标签进行几何修正

2 试验分析

2.1 试验环境

所有模型在 1 台搭载 NVIDIA RTX 6000 GPU 的工作站上进行,均基于 PyTorch 深度学习框架实现。训练采用 AdamW 优化器,初始学习率为 0.000 01,动量值为 0.9,权重衰减值为 0.005,批量大小设置为 4,并迭代训练 20 000 次。

2.2 试验数据集

源域数据集采用 Inria 建筑物数据集^[21],空间分辨率为 0.3 m,从原训练数据集中,随机选取不同城市的 63 张尺寸为 5000×5000 像素的影像,将其无重叠地裁剪为 1008 张尺寸为 1024×1024 像素的影像。目标域数据集采用无人机航拍数据集,其空间分辨率为 0.03 m,影像尺寸为 1024×1024 像素,数据标注采用 ArcGIS pro 的深度学习影像标注功能进行人工手动标注,按照像素级标准进行标注,共 824 张训练图像、412 测试图像。

2.3 评价指标

本文模型评估采用交并比(IoU)、精确率(P)、召回率(R)、F1 得分等指标进行,公式分别为

$$IoU = \frac{TP}{TP + FP + FN} \times 100\% \quad (4)$$

$$P = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

$$F1 = \frac{2PR}{P + R} \times 100\% \quad (7)$$

式中,TP、FP 和 FN 分别为真阳性、假阳性和假阴性。

3 试验结果

3.1 对比试验

表 1 展示了 CSG-Net 与多种先进域适应方法在该任务上的性能对比。所有模型均在 Inria 数据集(源域)上进行训练,并在广州无人机数据集(目标域)上进行了评估。传统的自训练方法如 PyCDA、CBST、Iast 及 DCA 均展现出了一定的适应能力,其中 CBST 取得了 72.08% 的 IoU,在对比方法中表现最佳。而本文提出的 CSG-Net 取得了 73.05% 的 IoU 和 84.43% 的 F1 分数,在所有对比方法中性能最优。相较于同样基于自训练且表现强劲的 CBST,CSG-Net 依然实现了性能的超越。这充分说明,在通过概率加权等策略解决了伪标签的基础语义噪声后,几何约束算法也发挥了一定作用,通过对伪标签进行边界层面的硬性修正,本文方法有效克服了因尺度剧变导致的建筑物形态错位和轮廓模糊问题,从而在像素级分类精度上取得了最终的优势。此外,为全面评估模型的实际计算开销,在统一的试验设置下,测定了各模型的训练时间和测试时间。训练时间定义为完成 20 000 次迭代所需的总耗时,测试时间则定义为模型在整个验证集上完成一次完整预测及评估所需的总时间。由表可见,CSG-Net 的总训练时间(10.19 h)略高于其他自训练方法,这是因为其更复杂的精炼所付出的必要开销。但在更关键的测试时间上,CSG-Net 耗时 546 s,在所有对比方法中耗时最短,显著优于 DCA(591 s)和 PyCDA(706 s)。这证明了 CSG-Net 在取得最高精度的同时,也具备较高水平的推理效率。

表 1 不同模型试验精度及开销对比

模型	IoU/ (%)	F1/ (%)	R/ (%)	P/ (%)	训练 时间/h	测试 时间/s
PyCDA	65.63	79.25	72.75	87.01	5.34	706
CBST	72.08	83.78	82.22	85.39	5.91	699
Iast	66.18	79.65	72.71	88.06	4.30	714
DCA	60.33	75.25	67.88	84.43	4.72	591
CSG-Net	73.05	84.43	76.40	94.33	10.19	546

3.2 可视化结果分析

为更直观地对比各网络在不同场景下的提取效果,在测试集中选取了5个具有代表性的场景进行可视化,如图3所示。通过将各方法的提取结果(图3(c)~(g))与真实标签(图3(b))进行比较,可以直观地反映出不同网络在面对巨大域差异时的性能表现。

在场景1和场景3这样建筑物密集且形态复杂的区域,大部分对比方法均出现了严重的粘连和断裂现象,难以清晰地分离出独立的建筑个体。相比

之下,本文提出的 CSG-Net 能够更好地保持建筑物的独立性和结构完整性,其提取结果在密集区域的区分度最高。在场景2和场景5这样建筑物轮廓相对规整的区域,CSG-Net 在边界锐利度和几何形态上展现出明显优势。其他方法(如 PyCDA、DCA)的预测结果边缘较为模糊,而 CSG-Net 生成的建筑物足迹轮廓更平滑、更完整,与真实标签的几何形态高度一致。在场景4这样包含大型建筑的区域,CSG-Net 同样表现出色,有效避免了在大型目标内部出现缺漏的问题,并保持了其规整的矩形轮廓。

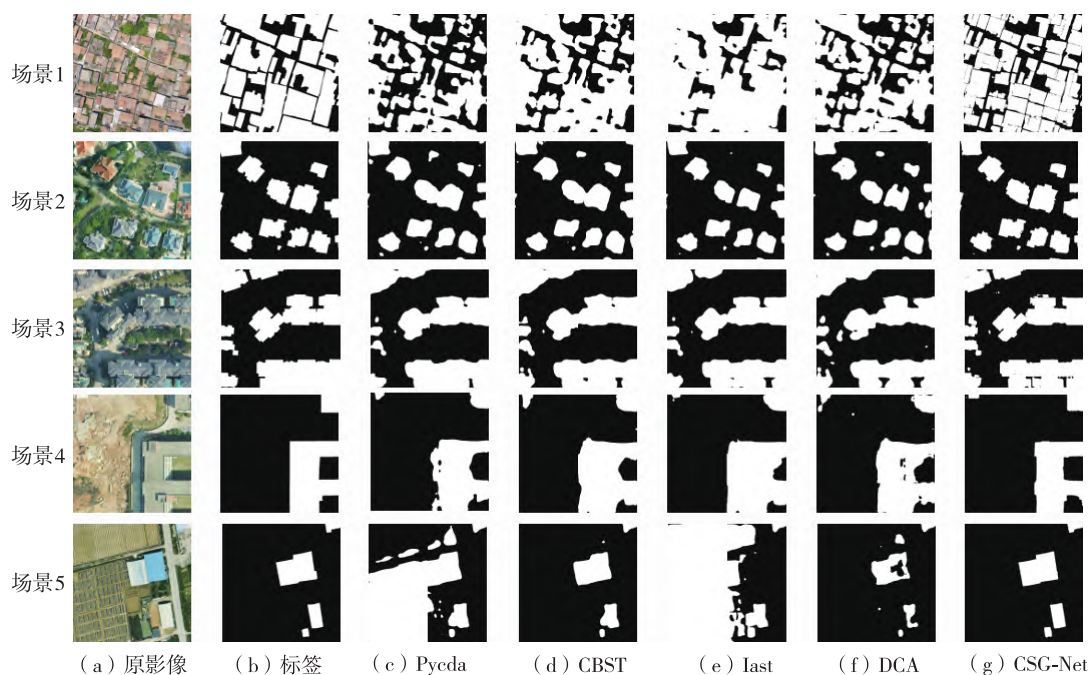


图3 不同模型在测试集上的建筑物提取可视化结果

3.3 消融试验

消融试验结果(见表2)验证了各个模块的有效性。①在基线模型上加入概率加权, IoU 从 52.49% 提升至 70.85%, 这说明通过 JSD 不确定性加权,能有效过滤噪声伪标签,防止模型在不可靠区域过拟合,是性能提升的关键。②在①的基础上加入几何修正: IoU 进一步提升至 73.05%。这表明,在概率软约束的基础上,引入基于 SAM 的几何硬修正,能进一步细化建筑物边界,带来可观的精度增益。因此,概率加权与几何修正两个模块相辅相成,共同构建了 CSG-Net 高效的伪标签精炼框架。

表2 模型消融试验结果 (%)

方法	IoU	F1	R	P
基线模型	52.49	68.85	56.24	88.75
+概率加权	70.85	82.94	73.98	94.37
+几何修正	73.05	84.43	76.40	94.33

4 结 语

本文提出了一种跨尺度几何精炼网络——CSG-Net,用于解决遥感影像建筑物提取中的无监督域适应难题。试验结果表明,CSG-Net 在跨尺度建筑物提取任务中表现出显著优势,相较于仅在源域训练的基线模型,其 IoU 提升了 20.56 个百分点,并优于包括 CBST 在内的多种主流域适应方法。消融试验进一步验证了概率加权与几何修正两个核心模块在性能提升中的关键作用。综合结果表明,CSG-Net 能够有效生成并利用高质量伪标签,从而在存在大规模域差异的情况下实现高精度建筑物的提取。尽管如此,本文研究仍存在一定的局限性。目前的方法仅限于建筑物的语义分割,尚未实现对独立建筑单体的实例化提取。未来研究将考虑引入更先进的实例分割网络,以实现单体建筑物

的精细提取,并探索将本文方法拓展至其他目标类别的提取任务(如道路、水体等),以进一步验证其通用性与应用潜力。

参考文献:

- [1] 吴剑,陈鹏,刘耀林,等. 震害损毁建筑物高分辨率遥感信息提取方法[J]. 地理与地理信息科学,2013,29(3):35-38.
- [2] 陈婕,刘纪平,徐胜华. 增强边缘信息的全卷积神经网络遥感影像建筑物变化检测[J]. 测绘通报,2023(6):61-67.
- [3] 方超,廖运茂,刘飞,等. 改进 Mask R-CNN 的无人机影像建筑物提取[J]. 北京测绘,2024,38(1):97-101.
- [4] 范荣双,陈洋,徐启恒,等. 基于深度学习的高分辨率遥感影像建筑物提取方法[J]. 测绘学报,2019,48(1):34-41.
- [5] 张萍,李必军,李垠,等. 面向震后快速灾害评估的遥感影像房屋数据空间化构建方法[J]. 测绘通报,2024(2):69-73.
- [6] 孟月波,苏世龙,黄欣羽,等. 细节增强与跨尺度几何特征融合的遥感影像建筑物提取网络[J]. 地球信息科学学报,2025,27(4):930-945.
- [7] YUAN Jiangye. Learning building extraction in aerial scenes with convolutional networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2018,40(11):2793-2798.
- [8] WANG Junjue, MA Ailong, ZHONG Yanfei, et al. Cross-sensor domain adaptation for high spatial resolution urban land-cover mapping: from airborne to spaceborne imagery[J]. Remote Sensing of Environment,2022,277:113058.
- [9] LI Qingyu, MOU Lichao, SUN Yao, et al. A review of building extraction from remote sensing imagery: geometrical structures and semantic attributes[J]. IEEE Transactions on Geoscience and Remote Sensing,2024,62:4702315.
- [10] LI Xue, LUO Muying, JI Shunping, et al. Evaluating generative adversarial networks based image-level domain transfer for multi-source remote sensing image segmentation and object detection [J]. International Journal of Remote Sensing,2020,41(19):7343-7367.
- [11] SHI Lukui, WANG Ziyuan, PAN Bin, et al. An end-to-end network for remote sensing imagery semantic segmentation via joint pixel-and representation-level domain adaptation[J]. IEEE Geoscience and Remote Sensing Letters,2021,18(11):1896-1900.
- [12] LIAN Qing, DUAN Lixin, LÜ Fengmao, et al. Constructing self-motivated pyramid curriculums for cross-domain semantic segmentation: a non-adversarial approach [C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul:IEEE,2019:6757-6766.
- [13] ZOU Yang, YU Zhiding, VIJAYA KUMAR B V K, et al. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training [M]//Computer Vision-ECCV 2018. Cham: Springer International Publishing,2018:297-313.
- [14] MEI Ke, ZHU Chuang, ZOU Jiaqi, et al. Instance adaptive self-training for unsupervised domain adaptation [M]//Computer Vision-ECCV 2020. Cham: Springer International Publishing,2020:415-430.
- [15] WU Linshan, LU Ming, FANG Leyuan. Deep covariance alignment for domain adaptive remote sensing image segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing,2022,60:5620811.
- [16] LIU Wei, LIU Jiawei, LUO Zhipeng, et al. Weakly supervised high spatial resolution land cover mapping based on self-training with weighted pseudo-labels[J]. International Journal of Applied Earth Observation and Geoinformation,2022,112:102931.
- [17] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything [C]//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris:IEEE,2024:3992-4003.
- [18] 赫晓慧,吴凯旋,李盼乐,等. 基于大幅面 SAM 的遥感影像建筑物提取研究[J]. 时空信息学报,2025,32(2):148-157.
- [19] XIAO Tete, LIU Yingcheng, ZHOU Bolei, et al. Unified perceptual parsing for scene understanding [C]//Proceedings of 2018 Computer Vision. Cham: Springer, 2018:432-448.
- [20] LIU Zhuang, MAO Hanzi, WU Chaoyuan, et al. A ConvNet for the 2020s[C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022: 11966-11976.
- [21] MAGGIORI E, TARABALKA Y, CHARPIAT G, et al. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark [C]//Proceedings of 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Fort Worth: IEEE,2017:3226-3229.

(责任编辑:马保卫)