引文格式：向浩,祁信舒,吕现伟,等. 基于提示信息的分割大模型建筑物提取方法研究[J]. 地理空间信息, 2024, 22(1): 29-32. doi:10. 3969/j. issn. 1672-4623.2024.01.007

基于提示信息的分割大模型建筑物提取方法研究

向　浩[1]，祁信舒[1]，吕现伟[2,3]，梁　思[1]，沈佳洁[1]
（1. 湖北省航测遥感院，湖北 武汉 430074；2. 华中科技大学，湖北 武汉 430074；
3. 武汉天远视科技有限责任公司，湖北 武汉 430070）

摘　要：针对现有方法训练的分割模型在测试集上提取建筑物效果不佳的问题，提出了一种基于分割大模型提示信息的建筑物精准提取方法。首先利用现有分割模型从原始影像中粗略提取建筑物区域，再从这些区域中生成建筑物目标框和关键点，作为分割大模型的提示信息，进而指导分割大模型精确分割建筑物目标。通过在多张遥感影像上进行分割实验验证了该方法的精度。结果表明，与现有分割方法相比，该方法的精确率和召回率分别从88.71%、58.38%提升至91.49%、83.74%。

关键词：目标分割；分割大模型；建筑物提取；深度学习
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2024）01-0029-04

Research on Building Extraction Method of Segmentation Model Based on Prompt Information

XIANG Hao[1], QI Xinshu[1], LYU Xianwei[2,3], LIANG Si[1], SHEN Jiajie[1]
(1. Hubei Institute of Photogrammetry and Remote Sensing, Wuhan 430074, China;
2. Huazhong University of Science and Technology, Wuhan 430074, China;
3. Wuhan Farsee Technology Co., Ltd., Wuhan 430070, China)

Abstract: To address the issue of poor building extraction performance on the test set by existing segmentation models, we proposed a precise building extraction method based on segmentation model prompt information. Firstly, we used the segmentation model to roughly extract the building areas from the original images. Then, we generated the buildings' target bounding boxes and key points from these areas to serve as prompt information for the segmentation model. Finally, we guided the precise segmentation of building targets by prompt information. To evaluate the proposed method, we conducted segmentation experiments on multiple remote sensing images. The results show that compared to existing segmentation methods, this method improves the precision and recall rates from 88.71%, 58.38% to 91.49%, 83.74%, respectively.

Key words: object segmentation, segmentation model, building extraction, deep learning

作为地理空间中最重要的人工要素信息，建筑物提取对城市建设、耕地保护、军事侦察等都具有重要意义[1]。因此，遥感影像中的建筑物提取一直都是人们研究的热点，早期学者主要利用遥感影像中的边界、区域等纹理信息识别建筑物，如明冬萍[2]等首先利用纹理信息在大尺度图像上粗略识别大目标区域，再利用区域大小、形状和位置关系得到基元特征，进而对这些区域进行精细化分割；刘正军[3]等首先利用建筑物特有的光谱、形态和纹理信息从遥感影像中提取建筑物，再将提取的建筑物边界变换为直线边界，从而得到建筑物的轮廓线；陈行[4]等基于建筑物形状和结构特点，从形态学的角度利用引导滤波改善了遥感影像中的建筑物提取性能，上述方法尽管在一定程度上实现了建筑物提取，但分割效率和精度还无法达到预期效果。

近年来，深度学习技术的快速发展极大地促进了以目标检测、识别与分割为代表的计算机视觉领域的进步。借助深度神经网络强大的特征提取能力，可显著提升计算机视觉技术的性能。FCN 网络[5]的提出拉开了基于深度学习的图像分割技术的序幕，实现了端到端的图像语义分割。Ronneberger O[6]等提出了一种 U 型网络模型 U-Net，通过不同网络层级的拼接，最大程度保留图像中的目标纹理和边缘等语义信息，从而改善了对目标的表征能力。Chen L[7]等提出的 DeepLabv3+模型在先前研究的基础上增加了一个解码器修正分割结果，从而进一步改善了分割边缘。鉴于 U-Net 和 DeepLabv3+在目标分割上取得的良好效果，吕道双[8]、于明洋[9]等分别将 U-Net 和 DeepLabv3+应用

收稿日期：2023-09-22。
项目来源：湖北省自然资源厅科技资助项目（ZRZY2023KJ08）。
第一作者简介：向浩（1981—），硕士，高级工程师，从事摄影测量与遥感生产管理和技术研究工作，E-mail：442493532@qq.com。

于遥感影像的建筑物提取中。虽然这些基于深度学习技术的模型可快速提取建筑物，但受限于训练样本数量和质量，其准确率仍难以满足“真实可靠准确”的质量要求。最近基于海量数据训练的分割大模型SAM引起了研究人员的广泛关注，尽管无法得到分割区域的语义信息，但其强大的分割能力以及可将目标上点、目标框等信息作为目标提示输入的特性被广泛应用于各种目标分割任务中。分割精度不高的建筑物提取模型可为SAM提供潜在的建筑物目标提示信息。

鉴于此，本文提出了基于SAM的建筑物分割方法，将建筑物分割模型与SAM相结合，建筑物分割模型提供语义信息，SAM提供准确的分割结果；并以LoveDA数据集中的遥感影像为实验数据，与直接使用DeepLabv3+得到的分割结果进行对比，以验证该方法的有效性。

1 研究方法

基于SAM的建筑物分割方法主要利用SAM的提示信息分割功能，首先利用DeepLabv3+得到遥感影像中建筑物粗分割区域；然后从建筑物粗分割区域中生成有效的建筑物目标点或目标框，作为SAM的提示信息；最后将提示信息与遥感影像一起输入SAM模型，利用SAM提示信息分割功能得到建筑物的精确分割结果（图1）。

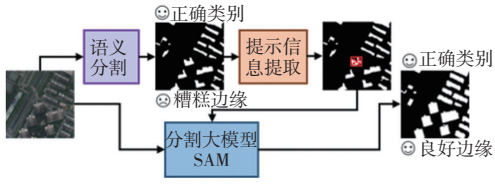


图1 基于提示信息SAM的建筑物分割方法研究框架

1.1 基于DeepLabv3+的建筑物粗分割

虽然基于深度学习的建筑物分割模型在准确度上仍难以满足生产质量要求，但可作为SAM分割模型重要的提示信息，将潜在的建筑物区域通过关键目标点或目标框与原始影像一起输入到SAM模型中，指导SAM模型精确提取建筑物，从而赋予SAM语义分割的能力。本文选择经典的DeepLabv3+作为建筑物粗分割模型，将训练好的DeepLabv3+模型直接输入原始遥感影像，输出为与原始影像大小相同的二值影像，其中白色表示潜在的建筑物区域。本文分别采用 $X \in R^{H\times W\times 3}$ 和 $I \in R^{H\times W}$ 表示分辨率为 $H\times W$ 的原始影像和对应的二值分割结果。

1.2 建筑物目标提示信息生成

SAM模型提供了基于目标关键点或目标框等提示信息的目标分割功能，因此需将DeepLabv3+输出的二值化分割结果转换为分割区域的目标关键点或目标框。由于DeepLabv3+的分割结果只包含建筑物的语义信息，不同建筑物的区域是无法区分的，因此不能直接将二值化分割结果转换为建筑物区域的提示信息。为解决该问题，本文首先利用开源程序库skimage.measure中的label函数将建筑物二值化结果中互不连通的区域转换为不同的建筑物目标，然后利用regionprops函数得到每个建筑物目标的最小矩形框。基于DeepLabv3+提取的建筑物目标框定义为：

$$\text{Boxes}=\text{regionprops}(\text{label}(I)) \tag{1}$$

二值化分割结果中不同的建筑物目标见图2，矩形框为基于上述方法得到的目标框，由于DeepLabv3+的分割结果中存在一些分散的小区域（图2中蓝色矩形框），其中并不包含建筑物目标，因此需进一步对建筑物目标面积小于一定阈值的区域进行过滤。基于DeepLabv3+提取的建筑物目标框定义修改为：

$$\text{Boxes}=\{b,\text{area}(b)>T\text{且}b\in\text{regionprops}(\text{label}(I))\} \tag{2}$$

式中，$b=(x,y,w,h)$ 为中心坐标为 (x,y) 、宽和高分别为 $2w$ 和 $2h$ 的目标框；$\text{area}(b)$ 为目标框 b 的面积；本文面积阈值 T 的取值为64。

最终得到的建筑物如图2中红色矩形框所示，这些建筑物目标框可有效指导SAM模型对该目标的精确分割。

为验证不同提示信息对SAM分割结果的影响，基于建筑物目标框生成目标关键点，进而利用目标提示点对SAM的建筑物分割进行指导。为避免DeepLabv3+模型分割结果在目标边界上的误差引起的错误提示，本文将建筑物目标框中心以及该中心与4个顶点的中心作为候选关键点；同时为了保证目标提示点的准确性，只有在建筑物目标上的候选点才能作为该建筑物目标的提示关键点，即二值化分割结果在候选点位置的值为1。如图3所示一个建筑物目标的5个候选关键点，只有4个红色点包含在DeepLabv3+二值化分割出的建筑物中，因此这4个点被选为该建筑物目标的提示关键点。给定一个建筑物目标框 $b=(x,y,w,h)$，5个候选关键点 points_c 和目标提示点 points 可定义为：

$$\text{points}_c=\left\{\begin{array}{c}(x,y),\left(x-\frac{w}{2},y-\frac{h}{2}\right),\left(x-\frac{w}{2},y+\frac{h}{2}\right),\\ \left(x+\frac{w}{2},y-\frac{h}{2}\right),\left(x+\frac{w}{2},y+\frac{h}{2}\right)\end{array}\right\} \tag{3}$$

$$\text{points}=\{p,I(p)=1\text{且}p\in\text{points}_c\} \tag{4}$$

式中，$I(p)$为DeepLabv3+二值分割结果在 p 点的取值。

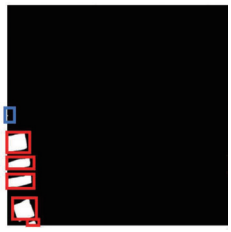


图2　建筑物目标框



图3　建筑物目标关键点

1.3　基于SAM的建筑分割方法

SAM分割模型提供了全图分割和提示分割两种方式，全图分割根据影像的纹理特征将整个影像分割为不同的区域，但无法输出这些区域的语义信息；提示分割则可根据给定的目标关键点、目标框等提示信息分割影像中的指定目标。本文采用SAM的提示分割功能提取给定遥感影像中的建筑物目标。根据DeepLabv3+建筑物粗分割结果中生成的建筑物目标框及其关键点，本文分别采用这两种提示信息指导SAM对遥感影像中的建筑物目标进行精准分割；再整合所有建筑物目标的分割结果，即可完成基于SAM的建筑物提取。

2　实验结果与分析

2.1　实验数据

为验证提出方法的有效性，本文采用武汉大学提供的遥感地表覆盖数据集LoveDA，包含农村和城市两种风格的遥感影像，影像分辨率为1 024×1 024。从该数据集的验证集部分随机选择100张包含建筑物目标的遥感影像进行实验评估。该数据集包含多个地物信息类别，本文进行实验评估时只考虑建筑物，其他视为背景。

2.2　评价方法

对于语义分割而言，影像的每一个像素分割均可视为二分类任务。因此，为评估该方法的建筑物提取效果，本文采用精确率（Prec）、召回率（Rec）、F_1–Score、交并比（IoU）和正确率（Acc）等语义分割常用评价标准，其中Prec表示模型预测为建筑物区域中真实建筑物区域像素点的占比，亦称查准率，反映模型预测结果的可信度；Rec表示所有建筑物区域对应的像素点被预测正确的比例，亦称查全率；F_1–Score为一个综合性评价指标，同时考虑了Prec和Rec；IoU表示模型预测为建筑物区域与真实建筑物区域在像素点上交集与并集的比值，理想情况下IoU为1，即预测为建筑物区域与真实建筑物区域完全重叠；Acc表示预测正确的像素点占比。

$$\mathrm{Prec}=\frac{TP}{TP+FP} \tag{5}$$

$$\mathrm{Rec}=\frac{TP}{TP+FN} \tag{6}$$

$$F_1=\frac{2\times\mathrm{Prec}\times\mathrm{Rec}}{\mathrm{Prec}+\mathrm{Rec}} \tag{7}$$

$$\mathrm{IoU}=\frac{TP}{TP+FP+FN} \tag{8}$$

$$\mathrm{Acc}=\frac{TP+TN}{TP+FP+FN+TN} \tag{9}$$

式中，TP 为将建筑物区域预测为建筑物区域的像素数；FP 为将非建筑物区域预测为建筑物区域的像素数；FN 为将建筑物区域预测为非建筑物区域的像素数；TN 为将非建筑物区域预测为非建筑物区域的像素数。

2.3　结果分析

SAM提供了vit_h、vit_l、vit_b等3个骨干网络预训练模型，其中vit_h拥有最多的模型参数，因此实验选取vit_h作为SAM的骨干网络。为对比不同提示信息对建筑物分割结果的影响，实验中分别采用单独使用目标框、单独使用目标关键点、同时使用目标框和关键点3种方法提取建筑物区域，并与直接基于DeepLabv3+的建筑物分割结果进行对比。不同方法的评价指标见表1。3种组合的分割结果均为基于目标框和关键点的最佳、基于目标框的次之、基于目标关键点的最差，这是由于目标关键点只是建筑物区域的一些代表点，与目标框相比，包含的目标区域信息量更少，而将二者组合后，在标定建筑物区域的同时可以指示目标框中建筑物区域与非建筑背景区域的具体差异，包含更具体的目标提示信息，因此目标框和关键点的组合可为SAM提供更好建筑物分割指导。另外，单独使用目标框或关键点方法的Prec比DeepLabv3+低，这是由于DeepLabv3+的分割结果区域较小且都属于建筑物区域，而单独使用目标框或关键点时，提示信息不够精确，导致SAM将提示信息周围纹理信息相近的区域分割出来，从而引入一部分非建筑物区域；而将目标框和关键点同时作为提示信息将会更精确地为SAM提供指导，因此提取的建筑物区域更加准确，与DeepLabv3+相比，Prec从88.71%提升至91.49%。对于其他的评价指标，相较于DeepLabv3+，基于3种提示信息组合的分割结果均有大幅提升，其中目标框和关键点的组合分别将Rec、F_1–Score、IoU和Acc从58.38%、70.42%、54.34%、94.67%提升至83.74%、87.44%、77.69%、97.38%。综上所述，基于提示信息的SAM建筑物提取方法可有效改善建筑物分割结果，

为SAM提供更多提示信息的目标框和关键点组合可更好地提升分割结果。不同方法在不同影像上的分割效果见图4。

表1 本文方法与基准方法的建筑物分割结果对比/%

算法		Prec	Rec	F_1	IoU	Acc
DeepLabv3+		88.71	58.38	70.42	54.34	94.67
SAM	提示点	81.66	75.88	78.67	64.84	95.52
	提示框	88.47	78.86	83.39	71.51	96.58
	提示框+点	91.49	83.74	87.44	77.69	97.38

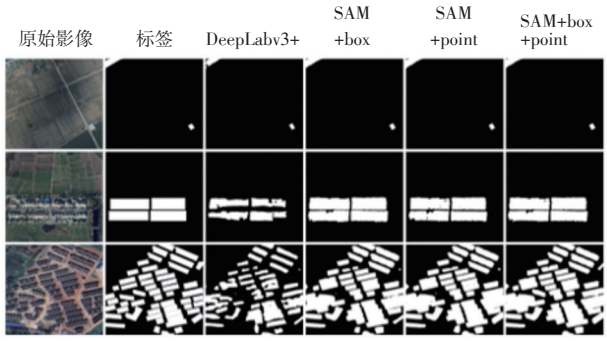


图4 不同方法建筑物提取结果对比

3 结 语

本文提出的基于提示信息的SAM建筑物提取方法可实现遥感影像中建筑物的精准提取。与DeepLabv3+相比，效果提升明显，其中Prec、Rec分别从88.71%、58.38%提升至91.49%、83.74%。该方法可显著改善建筑物边界，针对单个建筑物的分割效果更佳。未来将研究如何进一步改善基于SAM分割的效率问题。

参考文献

[1] 王俊,秦其明,叶昕,等. 高分辨率光学遥感图像建筑物提取研究进展[J]. 遥感技术与应用,2016,31(4):653-662

[2] 明冬萍,骆剑承,沈占锋,等. 高分辨率遥感影像信息提取与目标识别技术研究[J]. 测绘科学,2005,30(3):18-20

[3] 刘正军,张继贤,孟亚宾,等. 基于分类与形态综合的高分辨率影像建筑物提取方法研究[J]. 测绘科学,2007,32(3):38-39

[4] 陈行,卓莉,陶海燕. 基于MMBI的高分辨率影像建筑物提取研究[J]. 遥感技术与应用,2016,31(5):930-938

[5] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3 431-3 440

[6] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]. 18th International Conference on Medical Imaging Computing and Computer-assisted Intervention,2015:234-241

[7] Chen L, Zhu Y, Papandreou G, et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation[C]. 15th European Conference on Computer Vision,2018:833-851

[8] 吕道双,林娜,冯丽蓉,等. 基于改进型U-Net网络的高分辨率遥感影像建筑物提取[J]. 地理空间信息,2021,19(1):30-34

[9] 于明洋,张文焯,陈肖娴,等. 基于DeepLabv3+的高分辨率遥感影像建筑物自动提取[J]. 测绘工程,2022,31(4):1-10

(上接第12页)

[6] 王成,杨杰,姚辉,等. 四旋翼无人机飞行控制算法综述[J]. 电光与控制,2018,25(12):53-58

[7] 陈曦,谭鑫,高欢,等. 无人机不同航线倾斜摄影建模方法比较及改进[J]. 地理空间信息,2021,19(10):94-97

[8] 孙小雷,齐乃明,董程,等. 无人机任务分配与航迹规划协同控制方法[J]. 系统工程与电子技术,2015,37(12):2 772-2 776

[9] 尹高扬,周绍磊,吴青坡. 基于改进RRT算法的无人机航迹规划[J]. 电子学报,2017,45(7):1 764-1 769

[10] Chen C. CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature[J]. Journal of the Association for Information Science & Technology,2014,57(3):359-377

[11] 马子涵,姚奕,张靓. 基于科学知识图谱的装备保障效能可视化分析[J]. 舰船电子工程,2022,42(4):114-118

[12] 陈磊,李红霞. 中国煤矿安全评价知识图谱:基于CiteSpace V的可视化分析[J]. 科学技术与工程,2020,20(28):11 678-11 684

[13] Chen C. Searching for Intellectual Turning Points: Progressive Knowledge Domain Visualization[J]. Proccedings of the National Academy of Science of America(PNAS),2004,101(1):5 303-5 310