Precise Building Boundary Extraction Using Deep Learning Method

Lin Y. H.[1], Tsai F.[2*]

[1]*Student, Department of Civil Engineering, National Central University, Taiwan.*
0123kan@gmail.com
[2*]*Professor, Center for Space and Remote Sensing Research, National Central University, Taiwan.*
ftsai@csrsr.ncu.edu.tw

Abstract: Buildings are key features in urban areas and can be clearly identified in high-resolution satellite imagery. Compared with LiDAR, satellite images provide lower cost and more frequent updates, making them suitable for large-scale building extraction. Deep learning, particularly Mask R-CNN, has shown strong performance in this mission, with recent advances improving boundary regularity and generalization. Our preliminary experiments, fine-tuning WHU pretrained weights on Pleiades imagery in Taichung, Taiwan, achieved an F1 score of 0.85, detecting most buildings but with limited boundary precision. Future work will enhance detection accuracy by integrating nDSM data and applying boundary regularization algorithms to yield more accurate building footprints.

Keywords: Building boundary extraction, Deep learning, Mask R-CNN

1. Introduction

Building extraction from satellite imagery plays a crucial role in applications in many research fields. With the advancement of high-resolution satellite imaging, there has been a growing demand for automated and accurate methods to detect and visualize building structures. These technological developments not only enhance the precision of spatial data analysis but also assist timely decision-making in critical scenarios.

Deep learning methods, particularly convolutional neural networks (CNNs), have shown outstanding performance in computer vision tasks and are recognized as one of the most influential deep learning architectures (Gu et al., 2015). Based on it, instance segmentation models such as Mask R-CNN (He et al., 2017) have emerged as powerful tools for object detection and segmentation. By extending Faster R-CNN with an additional branch for predicting segmentation masks, Mask R-CNN achieves high precision in distinguishing object boundaries, making it capability for the task of building boundary extraction.

Previous studies have extended Mask R-CNN through various techniques to improve boundary precision. Zhao et al. (2018) applied a post-processing algorithm to generate more regularized polygons. Han et al. (2021) incorporated shadow removal in preprocessing to improve detection in disaster-affected regions. Chen et al. (2021) introduced Mask Height R-CNN by adding a 3D-RPN branch to predict building heights.

This work aims to investigate the effectiveness of Mask R-CNN for the task, emphasizes its ability whether it can capture precise building outlines with high accuracy.

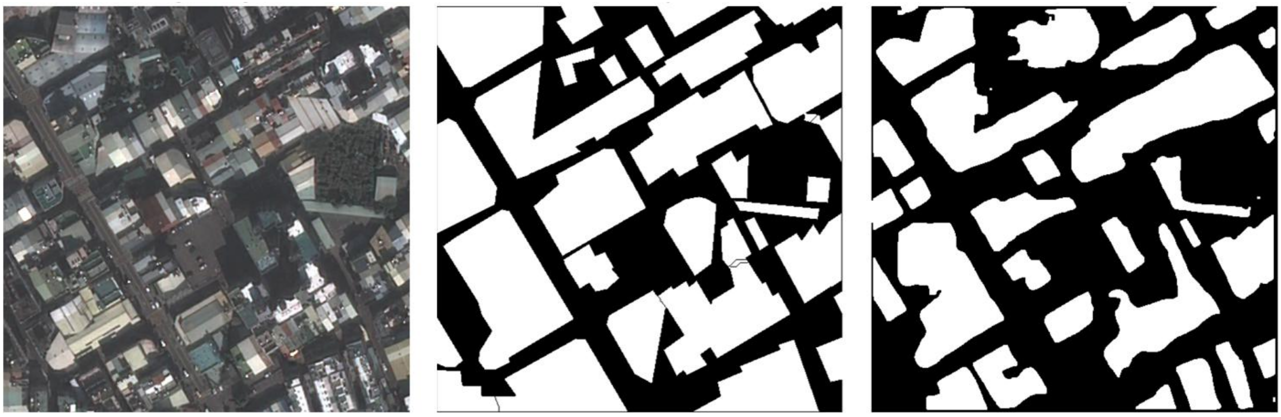
2. Methodology

The detection model is Mask R-CNN, extended with PointRend (Kirillov et al., 2019) to refine mask predictions and improve boundary regularity. The backbone network employed ResNet-101 with FPN for feature extraction.

First, the WHU Aerial Imagery Dataset (Ji et al., 2019) was employed for pretraining, comprising 2,943 training images, 627 validation images, and 2,220 test images, all resized to 512×512 patches. Second, a custom dataset was created from Pleiades satellite imagery of Taichung, Taiwan. After pan-sharpening and preprocessing, images were cut into overlapping 512×512 patches, yielding 136 training, 29 validation, and 30 test images. Pre-processing involved converting images to 8-bit format and normalizing brightness to ensure consistency. Fine-tuning was then performed using pretrained WHU weights.

3. Results/Findings

Results after fine-tuning step visualized in binary mask. Figure 1 and 2 show the result within different types of regions (urban and rural areas).



(a) (b) (c)

Figure 1: (a) Pleaides satellite image (b) Annotation of building in binary mask (c) Prediction of building in binary mask.

(a) (b) (c)

Figure 2: (a) Pleaides satellite image (b) Annotation of building in binary mask (c) Prediction of building in binary mask.

The predictions indicate that the model still requires improvement in dense urban areas, whereas in rural regions the generated masks exhibit more regularized shapes. The evaluation metrics, including precision, recall, and F1 score, obtained during the fine-tuning stage are summarized in Table 1.

Table 1: The precision, recall and F1 score in fine-tuning step.

	Precision	**Recall**	**F1 Score**
Validation	0.8512	0.8523	0.8517
Test	0.8490	0.8609	0.8549

4. Conclusion

The study confirmed that Mask R-CNN, when pretrained on large open datasets and fine-tuned on local satellite imagery, retains its ability to detect buildings. While the results demonstrate promising accuracy, reducing false negatives remains an essential step before reliable vectorized footprints can be produced. Addressing these challenges will allow the method to be further integrated into automated mapping and achieve LoD-2 modeling. Future work will focus on post-processing techniques such as the Douglas–Peucker algorithm (Douglas and Peucker, 1973) will be applied to vectorize building footprints, while incorporating nDSM data to filter non-building detections.

References

Chen, S., Mou, L., Li, Q., Sun, Y. and Zhu, X., 2021. Mask-Height R-CNN: An End-to-End Network for 3D Building Reconstruction from Monocular Remote Sensing Imagery, 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 1202-1205.

Cheng, B., Girshick, R.B., Dollár, P., Berg, A.C. and Kirillov, A., 2021. Boundary IoU: Improving Object-Centric Image Segmentation Evaluation, 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 15329-15337.

Douglas, D. H. and Peucker, T. K., 1973. *Cartographica*, 10(2), pp. 112-122.

Fujita, S. and Han, X., 2020. Cell Detection and Segmentation in Microscopy Images with Improved Mask R-CNN, *ACCV Workshops*.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. and Chen, T., 2015. Recent advances in convolutional neural networks, *Pattern Recognit*, 77, pp. 354-377.

Han, Q., Yin, Q., Zheng, X. and Chen, Z., 2021. Remote sensing image building detection method based on Mask R-CNN, *Complex & Intelligent Systems*, 8(3), pp. 1847-1855.

He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017. Mask R-CNN, *Proceedings of the IEEE international conference on computer vision*, pp. 2961-2969.

Hinton, G.E. and Mnih, V., 2013. Machine Learning for Aerial Image Labeling.

Ji, S., Wei, S. and Lu, M., 2019. Fully Convolutional Networks for Multi-Source Building Extraction from An Open Aerial and Satellite Imagery Dataset, *IEEE Transactions on geoscience and remote sensing*, 57(1), pp. 574-586.

Kirillov, A., Wu, Y., He, K. and Girshick, R.B., 2019. PointRend: Image Segmentation as Rendering, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9796-9805.

Lin, T. Y., Goyal, P., Girshick, R., He, K. and Dollar, P., 2018. Focal Loss for Dense Object Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2), pp. 318–327.

Xia, L., Zhang, J., Zhang, X., Yang, H. and Xu, M., 2021. Precise Extraction of Buildings from High-Resolution Remote-Sensing Images Based on Semantic Edges and Segmentation, *Remote Sensing*, 13(16), pp. 3083.

Zhao, K., Kang, J., Jung, J. and Sohn, G., 2018. Building Extraction from Satellite Images Using Mask R-CNN with Building Boundary Regularization, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 242-2424. Salt Lake City, UT, USA.