Development of an effective method for detecting small buildings in satellite images

Anna Zawadzka[1[0009–0006–4744–5623]], **Przemysław Głomb**[1[0000–0002–0215–4674]]
Mateusz Żarski[1[0000–0003–0380–8411]]

[1]*Institute of Theoretical and Applied Informatics PAS*
Bałtycka 5, 44-100 Gliwice, Poland
azawadzka@iitis.pl pglomb@iitis.pl mzarski@iitis.pl

1. Introduction

Detecting and monitoring small-scale buildings is a critical challenge in various sectors and can be crucial for various important tasks. Buildings considered small are the ones with a footprint smaller than 10 x 10 meters. Such definition can be justified by resolution of Sentinel-2 images, which is 10 meters at best. Examples of tasks for which small buildings detection is essential include: assessment and monitoring of damage caused by natural disasters or war, identification of illegal settlements, detecting new buildings in key military areas and identification of structures in protected areas. Regrettably, widely available satellite imagery, such as Sentinel-2, provides very limited spatial resolution, making it impossible to efficiently detect small structures. Despite its importance, this issue remains unexplored, with very limited research directly concerning building segmentation and detecting small structures. Traditional segmentation and classification methods exhibit severe limitations, particularly due to small buildings being frequently positioned at pixel boundaries, essentially occupying an insignificant fraction of the original pixel. Additionally, in the context of buildings in complex environments (e.g. surrounded by vegetation or shadows), popular methods, such as U-Net with ResNet, often fail to deliver reliable results. In this study, we propose a novel approach to this issue, by integrating state-of-the-art object detection model with deep-learning-based super-resolution network.

2. Methodology

We employ Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) [1] to upsample Sentinel-2 imagery 4 times beyond its standard resolution, preserving and extracting small-scale details with high fidelity. The super-resolved images are passed to YOLOv8 model for object detection. The YOLO algorithm is successfully used for object detection in various types of images, and YOLOv8 demonstrates high effectiveness in detecting small objects in remote sensing images [2], adapting to specific types of objects visible in aerial photographs (e.g. cars [3]), and precisely detecting green structures [4]. These two state-of-the-art methods complement each other to provide accurate segmentation of buildings, at the level of precision that is not achievable when working at Sentinel-2's native resolution. The core research problem is to overcome the limitations coming from low resolution of available satellite imagery to accurately detect and segment small-scale buildings. One way to connect the SR and detection models is to allow them to train together by connecting them sequentially and allowing the gradient to flow from the detection network to the SR network. The general concept of integrating super-resolution and detection models has been explored in other works [5], however we propose a setup in which the detection model receives the last feature map of the SR model as an input, while the SR model is trained simultaneously on the high-resolution image and the gradients obtained by the detection model. This gives the SR model a chance to be fine-tuned in a way that increases the accuracy of the network performing detection on the image. A quantitative assessment is essential to ensure the possibility of optimization of our method and its reliability. There are several possible approaches to building segmentation evaluation. One is to treat the segmentation task as pixel-wise binary classification (building vs non-building) and consider metrics such as accuracy, precision and recall. Another approach is to calculate the degree of overlap between predicted building regions and ground-truth annotations and therefore use Intersection over Union (IoU) metric for evaluation.

3. Results

Our experimental results demonstrate that super-resolution, particularly ESRGAN with Visual-Transformer-based metric CLIPscore as loss, delivers very good results in terms of reconstructing small details. This provides us with reliable foundation for the YOLOv8 detection model to achieve accurate, state-of-the-art segmentation. This work highlights the feasibility and robustness of applying Super-Resolution generative networks with widely-used detection models, such as YOLO, in order to successfully adapt them to crucial and challenging tasks, e.g. detection of objects that are small in relation to image resolution.

Acknowledgment

The study is part of the project developed in collaboration of two institutions: ITAI PAS and Space Research Centre PAS. The datasets include Sentinel-2 images and Database of Topographic Objects (BDOT).

References

[1] Wolters, P., Bastani, F., and Kembhavi, A. Zooming out on zooming in: Advancing super-resolution for remote sensing. *arXiv preprint*, 2311.18082v1, 2023. URL `https://arxiv.org/abs/2311.18082`.

[2] Yu, L., Zhi, X., Zhang, S., Jiang, S., Hu, J., Zhang, W., and Huang, Y. A method for detecting aircraft small targets in remote sensing images by using cnns fused with handcrafted features. *IEEE Geoscience and Remote Sensing Letters*, PP:1–1, 2024. doi:10.1109/LGRS.2024.3403548. URL `https://doi.org/10.1109/LGRS.2024.3403548`.

[3] Zhao, C., Guo, D., Shao, C., Zhao, K., Sun, M., and Shuai, H. Satdetx-yolo: A more accurate method for vehicle target detection in satellite remote sensing imagery. *IEEE Access*, 12:46024–46041, 2024. doi:10.1109/ACCESS.2024.3382245. URL `https://doi.org/10.1109/ACCESS.2024.3382245`.

[4] Sapkota, R., Ahmed, D., and Karkee, M. Comparing yolov8 and mask r-cnn for instance segmentation in complex orchard environments. *Artificial Intelligence in Agriculture*, 13:84–99, 2024. doi:10.1016/j.aiia.2024.07.001. URL `https://doi.org/10.1016/j.aiia.2024.07.001`.

[5] Salgueiro, L., Marcello, J., and Vilaplana, V. Seg-esrgan: A multi-task network for super-resolution and semantic segmentation of remote sensing images. *Remote Sensing*, 14(22), 2022. doi:10.3390/rs14225862. URL `https://www.mdpi.com/2072-4292/14/22/5862`.