

Comparative Analysis of Deep Learning Architectures for Aerial Image Segmentation

Evaluating CNN- and U-Net-based architectures
for accurate semantic segmentation of aerial imagery.



Mateusz Mazur

Faculty of Electrical Engineering, Automation, Computer Science and Biomedical Engineering, AGH

Field of study: Computer Science and Intelligent Systems

Specialization: Artificial Intelligence and Data Analysis

October 22, 2025

1 Introduction

Introduction

Accurate aerial image segmentation is vital for up-to-date mapping and urban analysis, yet manual methods are slow and costly. Deep learning offers scalable solutions, but architectures differ in handling challenges like shadows, small objects, and class variability. By comparing key models, this project seeks the most effective approach for reliable segmentation—directly supporting my geospatial research, where high-quality segmented imagery is essential for deeper analysis and better results.

This project focuses on **aerial image segmentation**, a key task in remote sensing with applications in mapping, urban planning, and environmental monitoring. The objective is to **compare the performance of different deep learning architectures**, ranging from classic convolutional neural networks (CNNs) to advanced models such as **U-Net** and U-Net-derived architectures (e.g, MultiRes-UNet, and CSE-UNet). Using publicly available datasets like **iSAID** or the **Dubai Aerial Imagery dataset**, the project evaluates model performance to identify the most reliable and efficient one.

■ Learning Aerial Image Segmentation From Online Maps [2]

Shows that **CNNs** can learn to segment aerial images using noisy labels from online maps like OpenStreetMap. The study demonstrates that large, imperfect datasets can reduce manual annotation needs while maintaining strong performance.

■ A Context and Semantic Enhanced UNet for Semantic Segmentation of High-Resolution Aerial Imagery [3]

Introduces **CSE-UNet**, which enhances segmentation in high-resolution aerial imagery using multi-level receptive field blocks and a dual-path encoder. The model effectively addresses intra-class heterogeneity and inter-class homogeneity, outperforming UNet and other baselines.

■ Integrating Semantic Edges and Segmentation Information for Building Extraction from Aerial Images Using UNet [1]

Proposes **MultiRes-UNet**, an improved model for building extraction that integrates multi-scale feature learning and semantic edge information. Results show superior boundary accuracy compared to UNet, DeeplabV3, and ResNet.

To evaluate and compare different segmentation architectures, two benchmark datasets are considered:

1 Semantic segmentation dataset – The Humans in the Loop (Kaggle)

The Humans in the Loop dataset contains aerial images of Dubai annotated with pixel-wise semantic segmentation across six classes. It is publicly available under a CC0 1.0 license, making it free for use in research and analysis.



Figure 1: Sample mask for the dataset

- 1 iSAID: A Large-scale Dataset for Instance Segmentation in Aerial Images [4] This paper introduces iSAID, the first large-scale benchmark dataset for instance segmentation in aerial imagery, featuring 655,451 objects across 15 categories. It combines object detection and pixel-level segmentation, addressing challenges like dense scenes and tiny objects. Results show that standard models like Mask R-CNN and PANet perform suboptimally, highlighting the need for specialized methods for aerial images.



Figure 2: Sample images from the iSAID dataset

Thank you for your attention

Questions

References

- [1] Abdollahi, A. and Pradhan, B. 2021. Integrating semantic edges and segmentation information for building extraction from aerial images using UNet. *Machine Learning with Applications*. 6, (2021), 100194.
- [2] Kaiser, P., Wegner, J.D., Lucchi, A., Jaggi, M., Hofmann, T. and Schindler, K. 2017. Learning aerial image segmentation from online maps. *IEEE Transactions on Geoscience and Remote Sensing*. 55, 11 (2017), 6054–6068. DOI:<https://doi.org/10.1109/TGRS.2017.2719738>.
- [3] Wang, F. and Xie, J. 2020. A context and semantic enhanced UNet for semantic segmentation of high-resolution aerial imagery. *Journal of physics: Conference series* (2020), 012083.
- [4] Waqas Zamir, S., Arora, A., Gupta, A., Khan, S., Sun, G., Shahbaz Khan, F., Zhu, F., Shao, L., Xia, G.-S. and Bai, X. 2019. Isaid: A large-scale dataset for instance segmentation in aerial images. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (2019), 28–37.

