Comparative Analysis of Deep Learning Architectures for Aerial Image Segmentation

Evaluating CNN-based and U-Net models
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

Table of contents

Introduction

Accurate segmentation of aerial images is essential for generating up-to-date maps and understanding urban dynamics. Traditional manual mapping is slow and costly, while deep learning models offer a scalable solution. However, different architectures vary in their ability to handle challenges like shadows, small objects, and high intra-class variability. By systematically comparing these approaches, this project aims to identify the most effective model for reliable and efficient aerial image segmentation, contributing to improved automated mapping systems.

This project focuses on aerial image segmentation, a key task in remote sensing used for mapping, urban planning, and environmental monitoring. The goal is to compare the performance of different deep learning architectures, ranging from classic convolutional neural networks (CNNs) to advanced segmentation models such as U-Net, MultiRes-UNet, and CSE-UNet. Using publicly available datasets such as iSAID and Dubai Aerial Imagery, the project will evaluate model performance in terms of accuracy, generalization, and boundary precision.

Learning Aerial Image Segmentation From Online Maps [2] – This paper shows that CNNs can learn to segment aerial images using noisy labels from online maps like OpenStreetMap. The results demonstrate that large, imperfect datasets can reduce the need for manual annotation while maintaining good performance.

A context and semantic enhanced UNet for semantic segmentation of high-resolution aerial imagery [3] – This paper introduces CSE-UNet, a semantic segmentation network designed for high-resolution aerial imagery. It addresses labeling challenges caused by intra-class heterogeneity and inter-class homogeneity using multi-level receptive field blocks and a dual-path encoder for richer contextual and semantic features. Experiments on the ISPRS Potsdam and Vaihingen datasets show that CSE-UNet outperforms UNet and other deep networks.

Integrating semantic edges and segmentation information for building extraction from aerial images using UNet [1] – This paper proposes MultiRes-UNet, an enhanced deep-learning model for accurate building extraction from aerial imagery. By integrating multi-scale feature learning, improved skip connections, and semantic edge information, the network refines building boundaries and handles challenges like shadows and vegetation. Experiments show that MultiRes-UNet outperforms models such as UNet, DeeplabV3, and ResNet in roof segmentation accuracy.

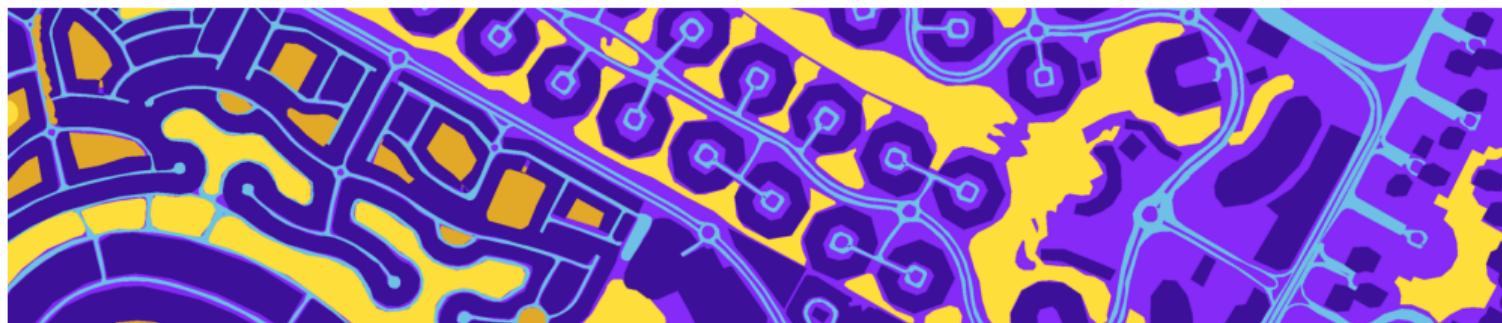**1** Satellite images of Dubai, the UAE segmented into 6 classes

Figure 1: Sample mask for the dataset

2. iSAID: A Large-scale Dataset for Instance Segmentation in Aerial Images – 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

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.