Vision Model for Environmental Monitoring Using Drone Footage Analysis

EL BATTAH Ahmed • EL ALAOUI Oumaima • KARRAKCHOU Taoufiq

Intelligence and Data Science Group, IPS Team, Faculty of Science of Rabat, Mohammed V University, Rabat, Morocco.

Contributing authors: ahmed_elbattah@um5.ac.ma; oumiama_elalaoui4@um5.ac.ma; taoufiq_karrakchou@um5.ac.ma;

Abstract

Environmental monitoring from UAV imagery enables rapid assessment of land-cover dynamics such as vegetation evolution and water-surface variations. This paper presents an end-to-end semantic segmentation pipeline built on OpenEarthMap [7] and a baseline U-Net model. We detail (i) dataset preparation with strict RGB-mask decoding into class IDs, (ii) construction of paired train/validation/test splits, and (iii) training and evaluation using standard segmentation metrics. On the validation set, the baseline U-Net achieves a mean Intersection-over-Union (mIoU) of 0.4416 and pixel accuracy of 0.6385. Per-class IoU is strongest for *Building* (0.6072) and *Tree* (0.5704), while *Water* reaches 0.3974. Beyond single-date mapping, we also demonstrate an environmental-monitoring proxy for change detection by comparing vegetation and water coverage between synthetic "before/after" masks, motivated by remote sensing change detection practice [6]. The presented baseline provides a reproducible foundation for future improvements such as multi-temporal change detection, multispectral features (e.g., NDWI/EVI [2], [3]), and stronger architectures.

1 Introduction

Environmental monitoring is essential for protecting ecosystems such as forests, rivers, and lakes. Traditional methods rely heavily on manual field visits, which are slow, expensive, and limited in spatial coverage. Drones (UAVs) enable rapid acquisition of high-resolution imagery and video; however, manual analysis of large-scale drone footage is impractical, motivating automated computer vision systems that can detect environmental changes (e.g., vegetation loss, water-level variation, and other visible indicators). This motivation and scope are stated in the project proposal.

Project goal. The objective of this project is to develop a vision model to analyze drone footage for environmental monitoring, focusing on detecting changes in vegetation and water levels using semantic segmentation. We implement a complete workflow aligned with the course expectations (dataset preparation, model training, quantitative evaluation, and report-ready visualizations).

Scope decision and limitations. The original proposal also mentioned pollution detection and object detection (e.g., YOLO). In the delivered work, pollution detection is not implemented because a task-aligned, labeled dataset for "signs of pollution" compatible with land-cover segmentation was not available within the project constraints (labeling requirements and domain mismatch). Object detection is complementary but distinct from dense land-cover mapping and would require additional annotations and evaluation. We therefore focus on land-cover segmentation (vegetation/water indicators) as the core deliverable, and we outline pollution detection and object detection as future work in Section 6.

2 Related Work

Drone-based environmental monitoring integrates remote sensing, environmental science, and computer vision. Srivastava et al. [5] survey drone-based monitoring and image processing workflows, emphasizing the practical value of UAV imagery and the importance of robust preprocessing and analysis pipelines for real deployments.

Recent reviews also highlight the role of UAV remote sensing in vegetation- and water-related monitoring tasks. Yang et al. [8] review UAV remote sensing for crop water and nutrient monitoring, describing end-to-end pipelines (sensor selection, preprocessing, modeling) and stressing that standardized processing and segmentation contribute to reliable environmental inference. For vegetation identification, Chang et al. [1] provide a review and meta-analysis of UAV remote sensing studies, showing that performance depends on sensor configuration, resolution, and the choice of learning models (including segmentation).

Water body detection is often addressed with segmentation models such as U-Net variants. Ngo et al. [4] show that deep learning with UAV multispectral imagery improves detection of small water bodies in challenging environments, and they discuss the benefit of integrating water-sensitive indices such as NDWI. Although the present project uses RGB land-cover segmentation, this line of work motivates water-surface indicators and future multispectral/index-aware extensions.

Finally, environmental monitoring frequently requires *change detection* across time. Willis [6] synthesizes best practices for remote sensing change detection for ecological indicators, reinforcing the importance of careful experimental design and validation. While our main experiments operate on single-date segmentation, the produced land-cover maps provide a prerequisite representation for multi-temporal analysis.

Index background (contextual). Two classical indices widely used in environmental monitoring are NDWI for enhancing open-water presence [3] and EVI for improved vegetation sensitivity under varying atmospheric and soil backgrounds [2]. Even though our implementation does not compute spectral indices (RGB-only), these indices motivate future work when multispectral UAV sensors are available.

3 Methods

3.1 Problem definition

Given an aerial RGB image $I \in \mathbb{R}^{H\times W\times 3}$, the objective is to predict a dense semantic label map $\hat{Y} \in \{0,\dots,7\}^{H\times W}$ over the eight OpenEarthMap land-cover classes [7].

3.2 Mask decoding (RGB → class IDs)

OpenEarthMap provides ground-truth masks encoded as RGB color images. A critical implementation step is a deterministic mapping from exact RGB values to integer class IDs. In our data preparation workflow, we validate decoding by visual inspection of (i) raw RGB masks, (ii) intermediate representations (e.g., grayscale IDs), and (iii) re-colored decoded masks rendered using the official class palette. Any unexpected/unmapped colors can be mapped to an ignore index to avoid corrupting supervision.

3.3 Model architecture

We implement a baseline U-Net encoder–decoder architecture with skip connections. The encoder downsamples feature maps to capture context, while the decoder upsamples and fuses multi-scale features from corresponding encoder stages to recover spatial detail. U-Net-style models are widely used in UAV and remote sensing segmentation tasks and provide a strong baseline for land-cover mapping [4].

3.4 Loss and optimization

Training uses a standard segmentation objective combining a pixel-wise classification loss with a region-overlap term (Dice). We additionally incorporate class balancing using weights computed from training-set pixel frequencies to mitigate dominance by frequent classes.

Optimization uses AdamW with learning rate 3.0000×10^{-4} and weight decay 1.0000×10^{-4}. Mixed precision (AMP) is enabled for efficiency, and the best checkpoint is selected based on validation mIoU.

3.5 Evaluation metrics

We report:

- **Pixel accuracy** (overall fraction of correctly labeled pixels),
- **Per-class IoU** $\mathrm{IoU}_c = \frac{TP_c}{TP_c+FP_c+FN_c}$,
- **mIoU** (mean IoU over the eight classes),
- **Superclass IoU for vegetation**: IoU of the union of vegetation classes (rangeland + tree + agriculture) vs. all other pixels,
- **Water IoU**: IoU for the water class (class 5).

The superclass IoU is useful for the project goal because confusion within vegetation subclasses (e.g., tree vs. rangeland) is less critical than separating vegetation from non-vegetation in monitoring scenarios.

4 Data Description

4.1 Dataset: OpenEarthMap

OpenEarthMap is a high-resolution land-cover mapping benchmark designed for semantic segmentation at global scale [7]. The dataset defines eight classes with a fixed mask color palette. Table 1 lists the official class proportions and HEX colors used in label encoding (as required by the project scope).

Table 1: OpenEarthMap classes, official proportions (%) and HEX color codes [7].

ID	Class	%	HEX	Color
0	Bareland	1.5	`#800000`	
1	Rangeland	22.9	`#00FF24`	
2	Developed space	16.1	`#949494`	
3	Road	6.7	`#FFFFFF`	
4	Tree	20.2	`#226126`	
5	Water	3.3	`#0045FF`	
6	Agriculture land	13.7	`#4BB549`	
7	Building	15.6	`#DE1F07`	

4.2 Prepared dataset and splits

The data preparation notebook constructs valid image–mask pairs and exports prepared splits used by training. The labeled splits contain **2149** training samples, **268** validation samples, **270** test samples, plus an **infer_test** split of **1151** images for inference-only evaluation.

For training, images are resized to 256×256 and normalized using mean and standard deviation computed and stored during preparation:

$$\mu = [0.4719643,\ 0.4814581,\ 0.4312912], \quad \sigma = [0.2019147,\ 0.1916494,\ 0.2061472].$$

4.3 Class imbalance (training split)

Pixel-frequency analysis shows substantial imbalance, with rangeland and tree dominating the training pixels while bareland and water are relatively rare. The computed class proportions in the prepared training split (in %) are: Bareland 2.01, Rangeland 35.7, Developed 11.5, Road 5.33, Tree 12.7, Water 2.51, Agriculture 14.9, Building 15.4. This motivates the use of class weighting during optimization.

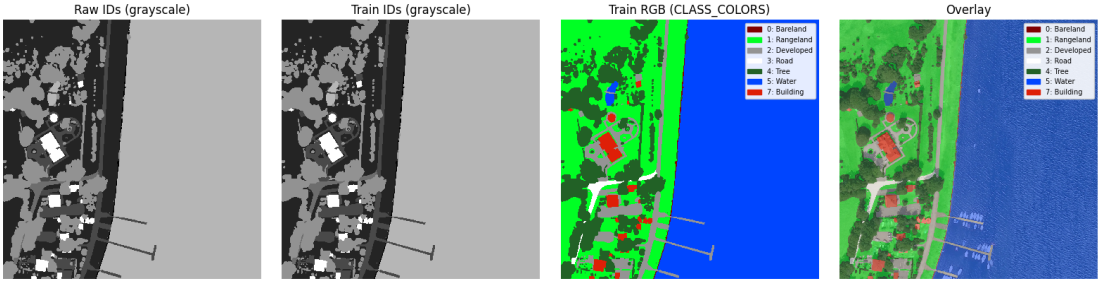
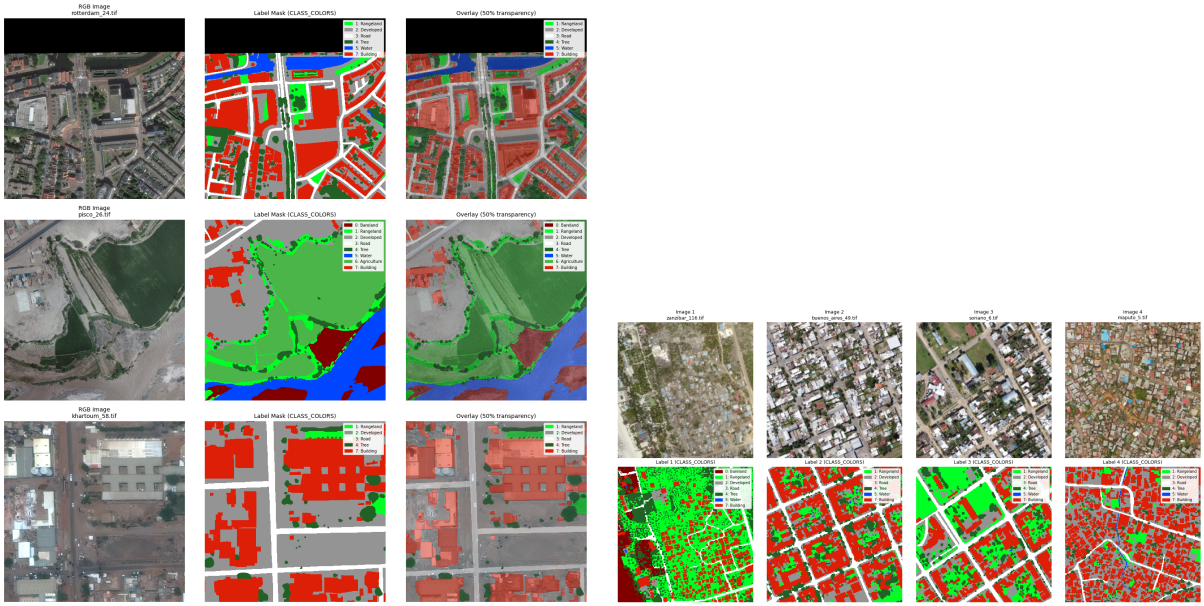


Figure 1: Mask processing views during data preparation (raw IDs, train IDs, RGB palette, and overlay).



(a) Sample images and masks (grid).

(b) Batch visualization (paired samples).

Figure 2: Qualitative inspection of paired image–mask samples during preparation.

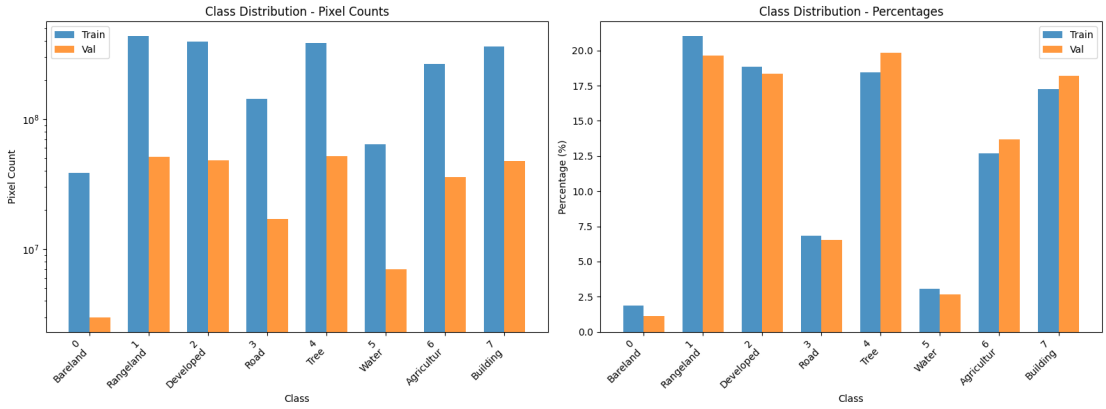
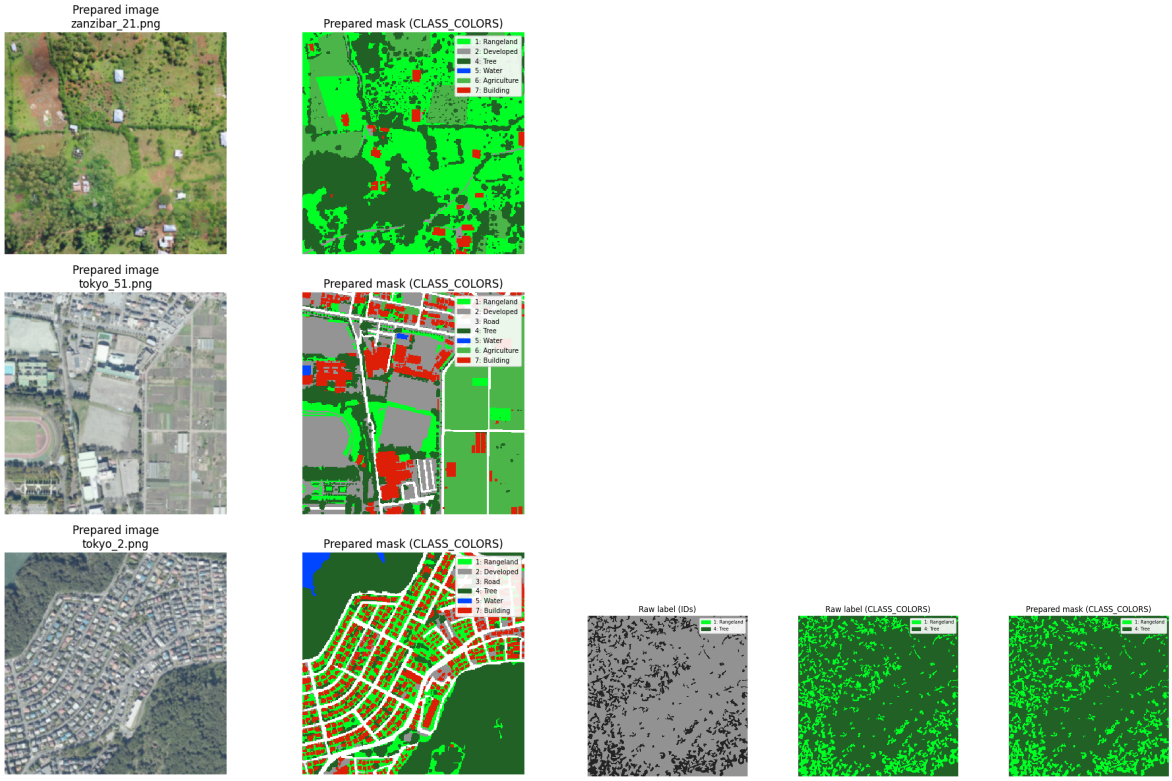


Figure 3: Class distribution diagnostics computed during data preparation.



(a) Prepared sample visualization.

(b) Label decoding sanity check.

Figure 4: Report-ready preparation artifacts exported from the notebook.

5 Experiments and Results

5.1 Experimental setup

Training protocol. Models are trained on the prepared labeled splits with input size 256×256, batch size 4, and up to 30 epochs. The optimizer is AdamW (learning rate 3.0000×10^{-4}, weight decay 1.0000×10^{-4}), and AMP is enabled. The best checkpoint is selected using validation mIoU.

Reproducibility. Experiments are run in a GPU environment (e.g., Colab T4) using PyTorch. Dataset normalization statistics are stored and re-used consistently across training and evaluation. Where applicable, random seeds are fixed.

Runtime notes. A benign PyTorch DataLoader worker cleanup warning may appear during interpreter

shutdown (e.g., `Exception ignored in: __del__ ... _shutdown_workers`). This warning is known to occur at process teardown and does not invalidate the computed metrics.

5.2 Quantitative segmentation performance

Baseline U-Net (required). Table 2 reports per-class IoU for the baseline U-Net on the validation set. The mean IoU is computed as the average across the eight per-class IoUs, yielding mIoU 0.4416. The baseline also reports pixel accuracy 0.6385 on the validation set.

We additionally report (i) water IoU (0.3974) and (ii) vegetation macro-IoU (mean of rangeland/tree/agriculture IoUs: 0.4936). A vegetation *superclass* IoU can be higher because it does not penalize confusions among vegetation subclasses; the notebook reports vegetation-superclass IoU 0.8135 for the baseline.

Table 2: Baseline U-Net validation performance: per-class IoU (8 classes).

Class	IoU	Class	IoU
Bareland (0)	0.2057	Tree (4)	0.5704
Rangeland (1)	0.3569	Water (5)	0.3974
Developed space (2)	0.3663	Agriculture (6)	0.5536
Road (3)	0.4749	Building (7)	0.6072
mIoU (mean over 8 classes)		**0.4416**	
Pixel accuracy		**0.6385**	
Water IoU (class 5)		0.3974	
Vegetation macro-IoU (mean of 1,4,6)		0.4936	

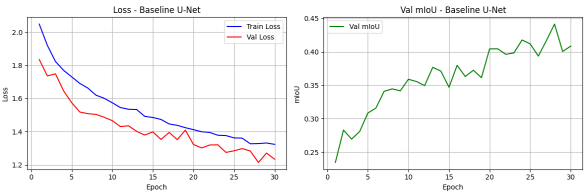
Optional improved model. In addition to the required baseline, the notebook evaluates an improved U-Net variant . On validation, the improved model achieves mIoU 0.4961 and pixel accuracy 0.6691, with a marked gain on water IoU (0.6045).

Table 3: Validation comparison between baseline U-Net and improved model.

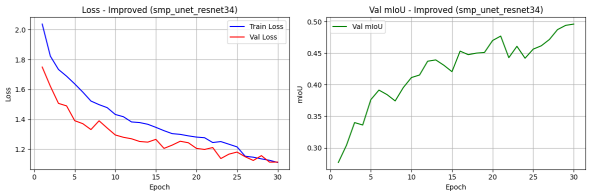
Model	mIoU	Pixel Acc.	Water IoU	Vegetation superclass IoU
Baseline U-Net	0.4416	0.6385	0.3974	0.8135
Improved (Option A)	0.4961	0.6691	0.6045	0.8545

5.3 Training dynamics and qualitative segmentation results

We visualize training curves and qualitative predictions to diagnose convergence and to assess whether vegetation and water regions are segmented plausibly. Qualitative overlays are particularly important for environmental monitoring because small water bodies and thin roads can be under-represented in global metrics.



(a) Baseline U-Net training curves.



(b) Improved model training curves.

Figure 5: Training dynamics (loss and mIoU) exported by the training notebook.

Figure 6: Baseline qualitative predictions (images, ground truth masks, and predictions).

5.4 Environmental monitoring proxy: synthetic change detection

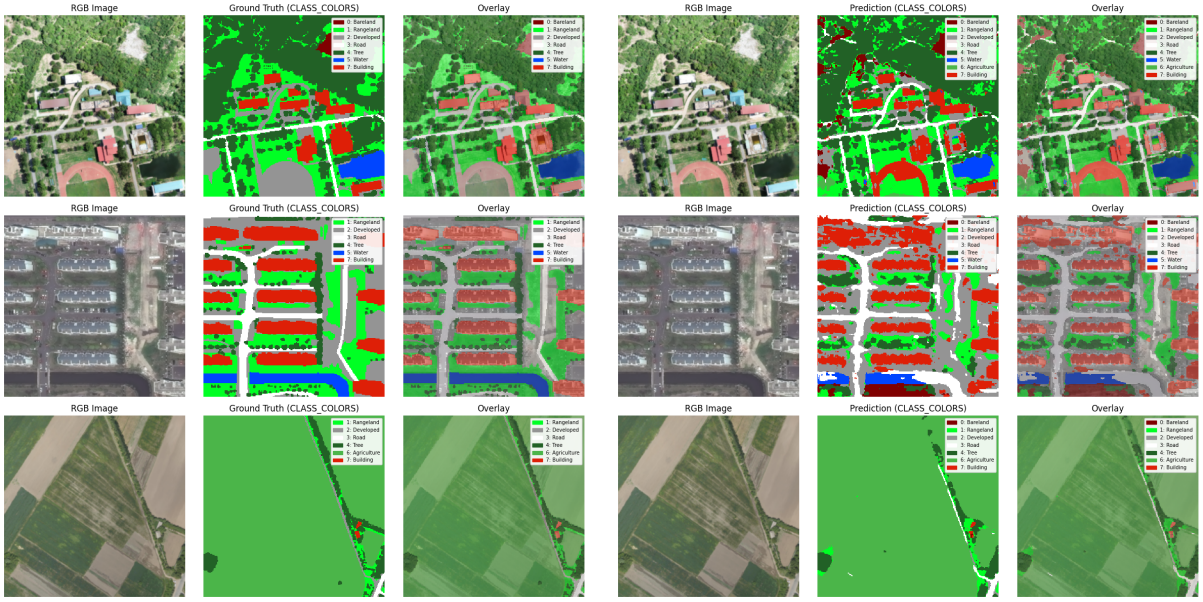
OpenEarthMap is not time-paired; therefore, the notebook demonstrates change detection using *synthetic before/after* pairs created via controlled perturbations. The change proxy compares (i) water surface area (percent of pixels predicted as water) and (ii) vegetation coverage (percent of pixels in the union of vegetation classes). This demonstration is aligned with standard remote sensing practice where consistent land-cover maps across time enable downstream change analysis [6].

In the synthetic demonstration, the notebook reports mean ground-truth changes of $-0.6900\,\%$ for vegetation and $0.1050\,\%$ for water, while predicted mean changes are $-0.5270\,\%$ for vegetation and $0.3350\,\%$ for water.

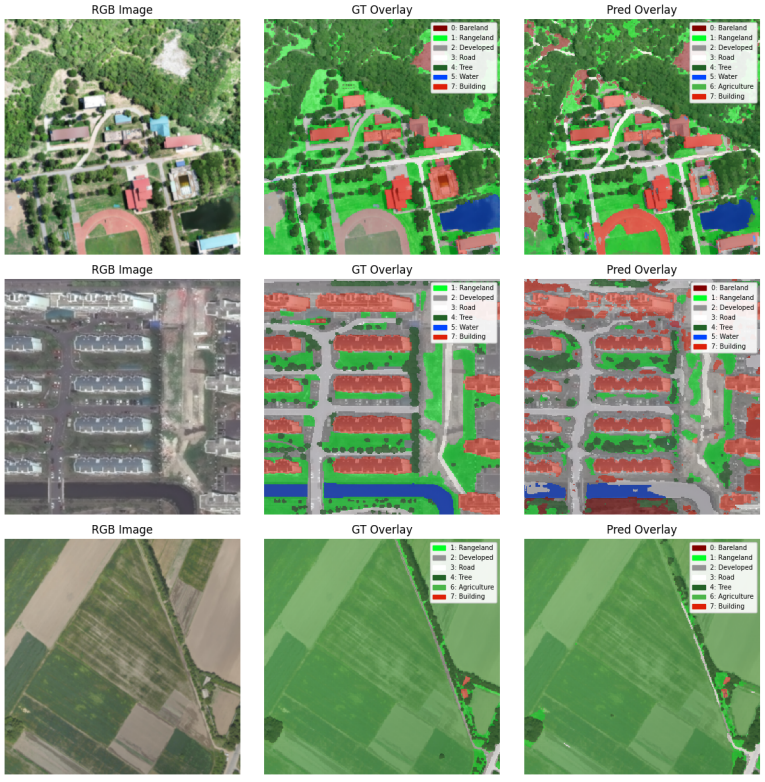
5.5 Discussion

The baseline U-Net provides a solid starting point for OpenEarthMap land-cover segmentation, reaching mIoU 0.4416 with strong performance on visually distinctive classes such as buildings and trees. Water segmentation remains challenging due to limited class prevalence and diverse appearance; prior work suggests that additional spectral information and indices such as NDWI can improve water delineation [3], [4]. Vegetation monitoring can similarly benefit from multispectral indices such as EVI [2]. These observations are consistent with UAV monitoring reviews emphasizing that sensor selection and preprocessing are key determinants of downstream performance [1], [8].

The synthetic change detection experiment demonstrates how consistent semantic maps can be converted into monitoring indicators (coverage changes). While synthetic, it provides a bridge from single-date segmentation to time-based monitoring, a direction supported by established change detection practice [6].

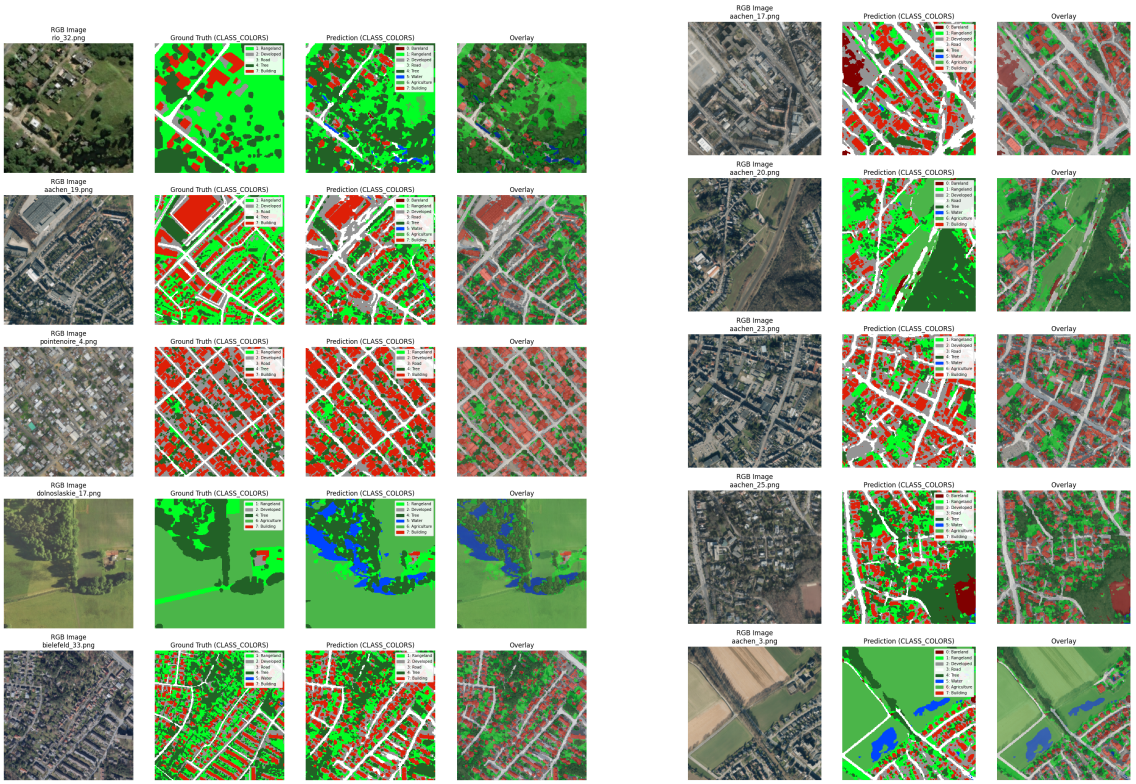


(a) Ground truth overlay (CLASS_COLORS). (b) Prediction overlay (CLASS_COLORS).



(c) Overlay comparison grid (GT vs. Pred across multiple samples).

Figure 7: Baseline overlays for qualitative assessment.



(a) Baseline predictions on test split.

(b) Baseline inference on infer_test split.

Figure 8: Baseline predictions beyond validation: test and inference-only splits.

(a) Improved model qualitative predictions (Option A).

(b) Baseline vs. improved qualitative comparison.

Figure 9: Qualitative results for the improved model and comparison against the baseline.

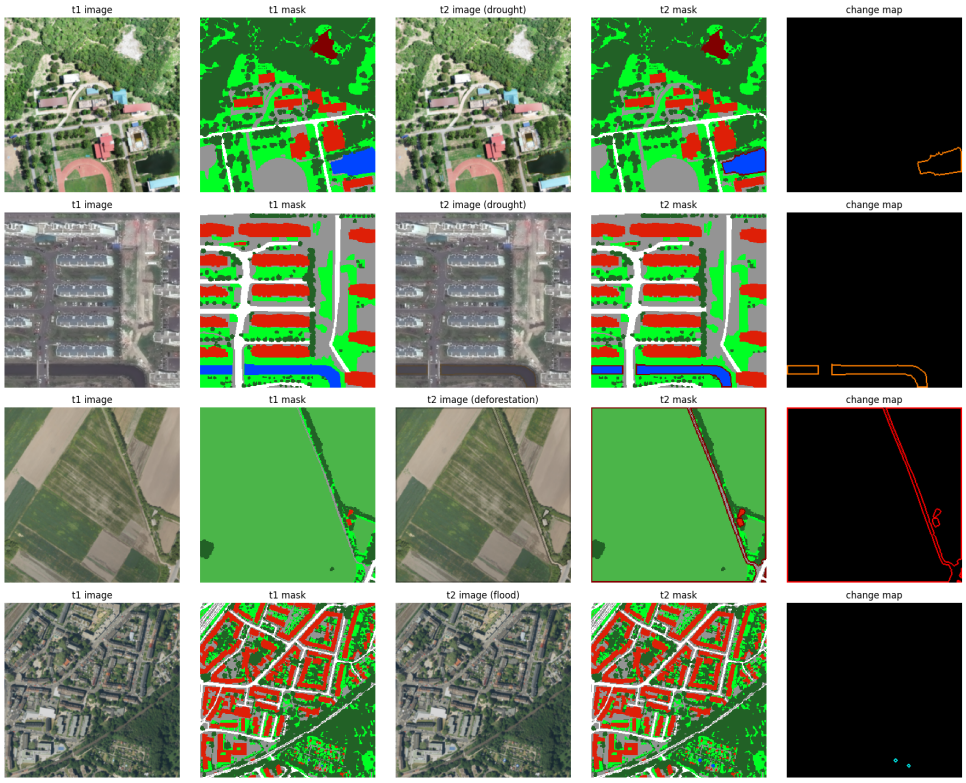


Figure 10: Synthetic change detection examples (before/after segmentation and coverage deltas).

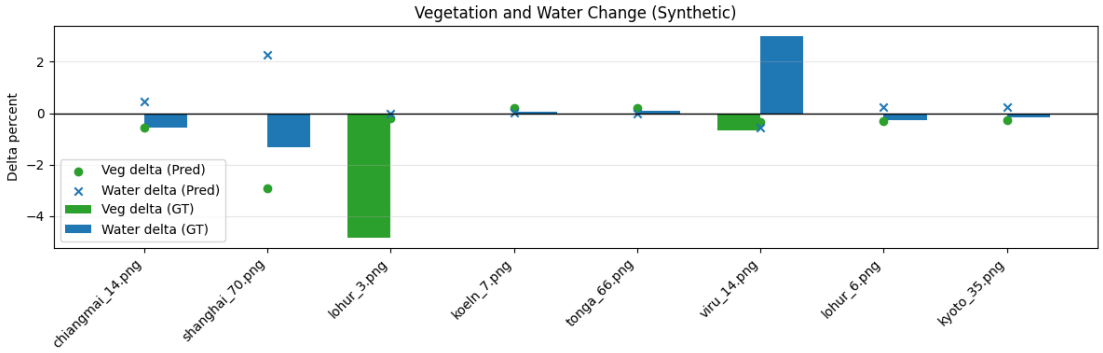


Figure 11: Change deltas plot for vegetation and water coverage (synthetic pairs).

6 Conclusion

This work presented an end-to-end semantic segmentation pipeline for environmental monitoring using OpenEarthMap aerial imagery and a baseline U-Net. We described robust data preparation (paired splits, RGB mask decoding, normalization), trained a baseline U-Net, and evaluated it quantitatively (per-class IoU, mIoU, pixel accuracy) and qualitatively (prediction grids and overlays). The baseline achieves mIoU 0.4416 and pixel accuracy 0.6385 on validation, with water IoU 0.3974 and strong performance on vegetation-related classes. We additionally demonstrated a monitoring proxy for vegetation and water "change" using synthetic before/after segmentation pairs.

Planned vs. achieved. The proposal included broader ambitions (pollution detection and object detection). These were not implemented due to dataset/label mismatch and the need for dedicated supervision for pollution categories. The achieved deliverable is a reproducible land-cover segmentation baseline targeting vegetation and water-surface indicators, matching the course guideline emphasis on a complete DL pipeline and reportable results.

Future work. Several extensions follow naturally: (i) multi-temporal change detection using real time-paired UAV data [6]; (ii) integrating multispectral imagery and index features (NDWI/EVI) [2], [3]; (iii) stronger architectures (DeepLab variants, transformer-based segmenters) and improved losses; (iv) domain adaptation from OpenEarthMap to real drone video frames with temporal smoothing and uncertainty estimation; (v) adding a pollution detection module with task-aligned datasets and/or YOLO-based object detection for complementary monitoring tasks.

References

[1] B. Chang, F. Li, Y. Hu, H. Yin, Z. Feng, and L. Zhao, "Application of UAV Remote Sensing for Vegetation Identification: A Review and Meta-analysis," *Frontiers in Plant Science*, vol. 16, 2025. DOI: 10.3389/fpls.2025.1452053

[2] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, "Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices," *Remote Sensing of Environment*, vol. 83 no. 1–2, pp. 195–213, 2002. DOI: 10.1016/S0034-4257(02)00096-2

[3] S. K. McFeeters, "The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features," *International Journal of Remote Sensing*, vol. 17 no. 7, pp. 1425–1432, 1996. DOI: 10.1080/01431169608948714

[4] P. L. Ngo, V. H. Pham, N. L. Bui, H. A. T. Phan, H. B. Vo, T. P. Velavan, and D. K. Tran, "Detection of Small Water Bodies for Vector Control Using Deep Learning on Multispectral Imagery from Unmanned Aerial Vehicles," *Discover Artificial Intelligence*, vol. 5, p. 170, 2025. DOI: 10.1007/s44163-025-00422-6

[5] S. K. Srivastava, R. Kaluri, S. Singh, et al., "Drone-Based Environmental Monitoring and Image Processing," *Sensors*, vol. 22 no. 20, p. 7872, 2022. DOI: 10.3390/s22207872

[6] K. S. Willis, "Remote Sensing Change Detection for Ecological Monitoring in United States Protected Areas," *Biological Conservation*, vol. 182, pp. 233–242, 2015. DOI: 10.1016/j.biocon.2014.12.006

[7] J. Xia, B. Adriano, C. Broni-Bediako, and N. Yokoya, "OpenEarthMap: A Benchmark Dataset for Global High-Resolution Land Cover Mapping," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Dataset paper introducing OpenEarthMap and land-cover color palette., IEEE/CVF, 2023, pp. 1–11. [Online]. Available: https://openaccess.thecvf.com/

[8] X. Yang, J. Chen, X. Lu, H. Liu, Y. Liu, X. Bai, L. Qian, and Z. Zhang, "Advances in UAV Remote Sensing for Monitoring Crop Water and Nutrient Status: Modeling Methods, Influencing Factors, and Challenges," *Plants*, vol. 14 no. 16, p. 2544, 2025. DOI: 10.3390/plants14162544