Semantic Segmentation for Land Use/Land Cover

Benjamin Koob | **Mentor: Jacob Arndt**

Geospatial Science and Human Security Division

Introduction

Accurate **land use/cover (LULC)** data has several important applications, including environmental monitoring, urban planning, and population estimation.

We sought to develop a **high-accuracy**, **globally scalable**, and **fast inference** land use segmentation model.

Data

Imagery: ESA Sentinel-2 (25 diverse cities, 2 scenes each)

Labels: ESA WorldCover V200 (11 classes)

Validation: Hand-labelling (2 scenes)

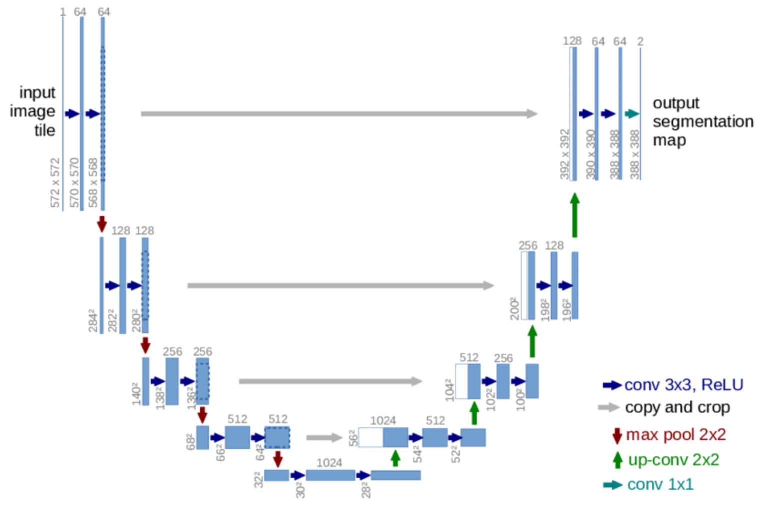
Methods

Semantic segmentation is a **computer vision task** that assigns class labels to every pixel of an image, enabling accurate and granular classification.

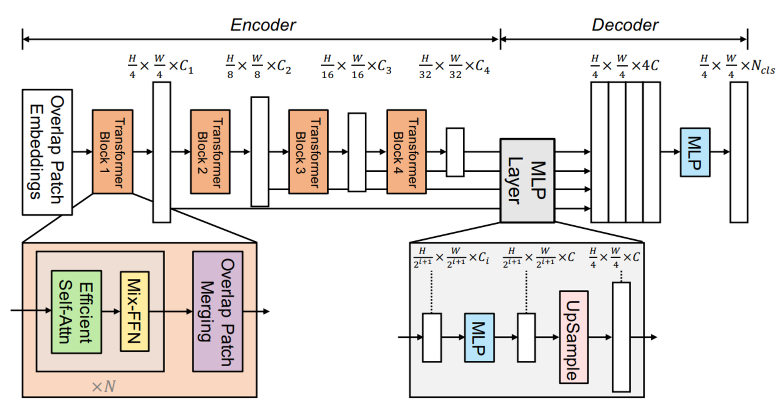
We trained and tested the following deep learning models:

- **U-Net**: Accepted baseline, **CNN** architecture [1]
- **SegFormer**: Efficient **transformer**-based architecture [2]
- **Prithvi EO2**: State-of-the-art **foundation model** [3]

U-Net Architecture [1]



SegFormer Architecture [2]

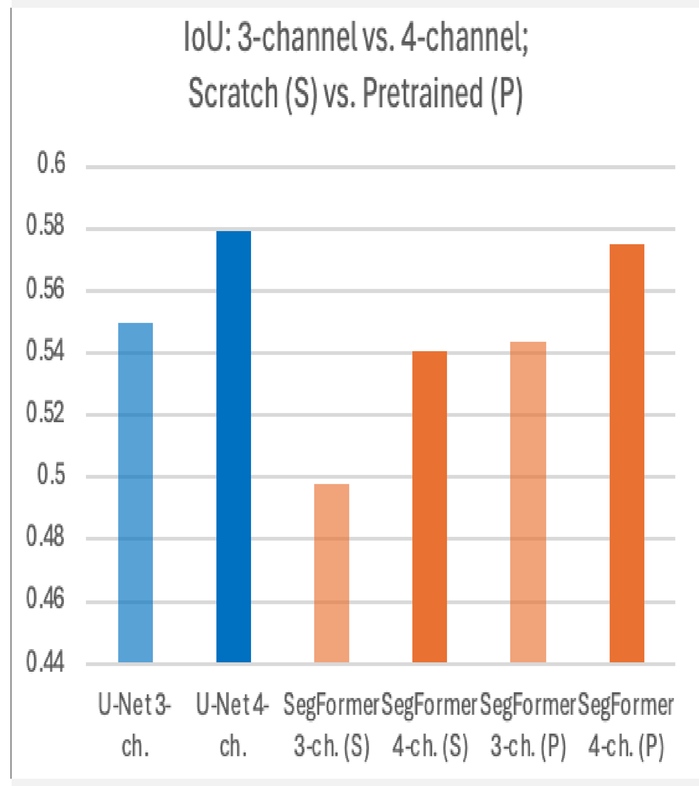


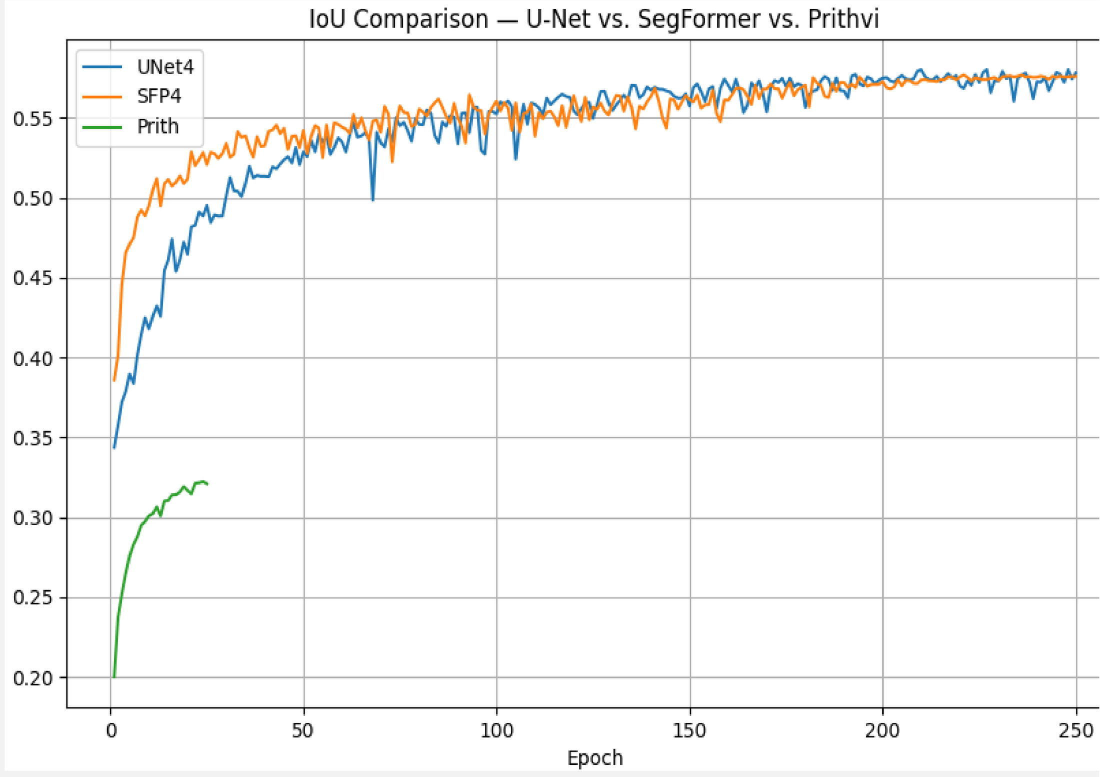
Experiments.

- *Input Data:* 3-channel (RGB imagery) vs. 4-channel (RGB + Near-Infrared).
- *Initialization & Training:* Training from scratch vs. finetuning from pretrained weights.
- *Model Size:* For pretrained models, testing small (3.7M params) vs. large (82M params) models.

Results

- **4-channel** models outperformed 3-channel models.
- **Pretrained** models outperformed from-scratch models.
- **Larger weight** models outperformed smaller weight models.





Confusion Matrices:

U-Net (4-channel): 86.05% overall accuracy

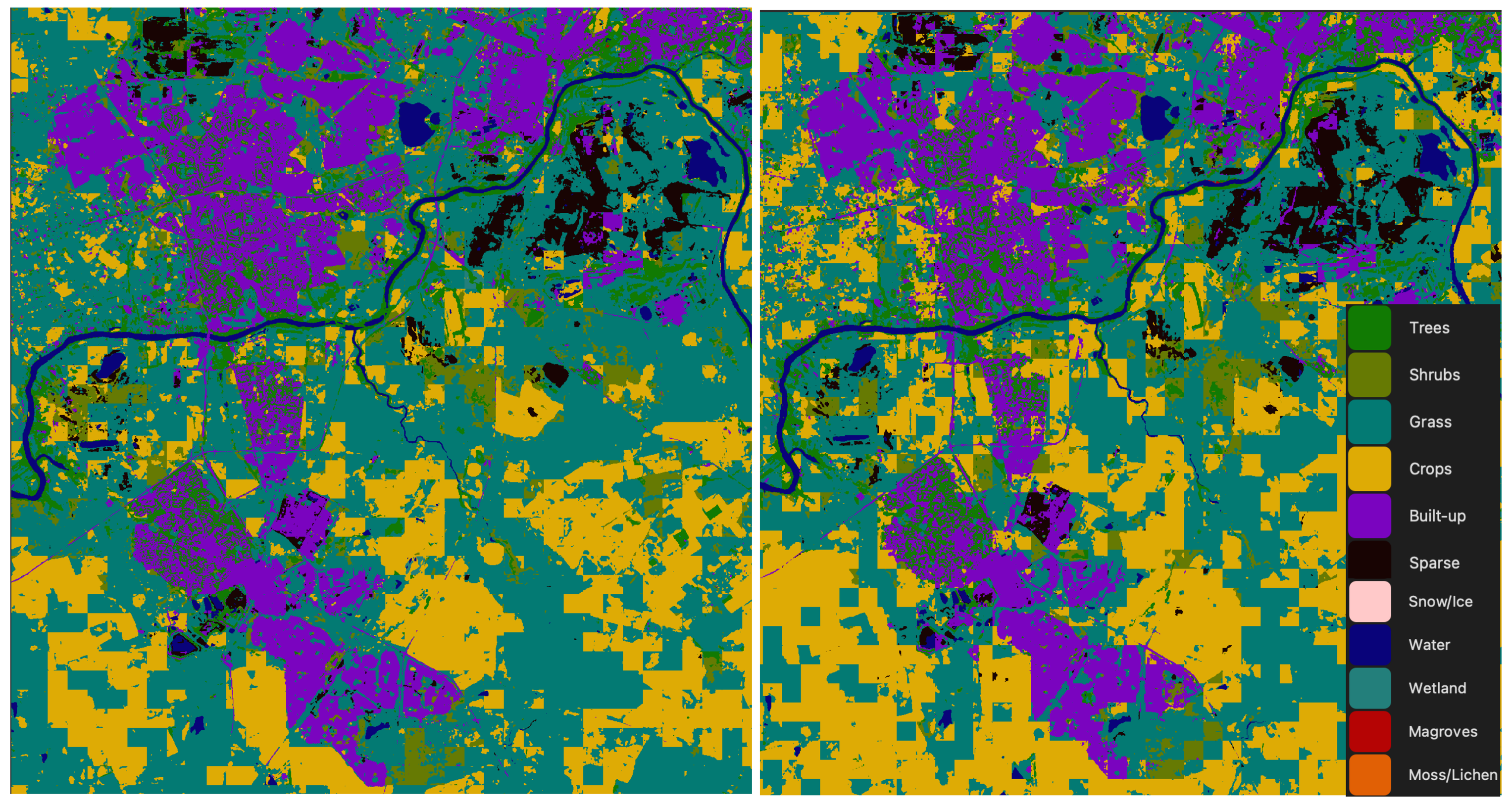
	Trees	Shrubs	Grass	Crops	Built-up	Bare	Water	Wetlands	Mangroves	Moss/Lichen
Trees	**91.66**	0.79	2.05	1.59	1.43	0.03	2.29	0.07	0.06	0.02
Shrubs	18.76	**70.63**	6.81	2.52	1.00	0.14	0.11	0.02	0.02	0.00
Grass	14.19	2.15	**73.80**	6.56	1.93	0.57	0.28	0.26	0.04	0.21
Crops	7.59	0.64	5.97	**84.10**	1.05	0.21	0.23	0.18	0.03	0.00
Built-up	7.14	0.39	2.24	2.08	**86.65**	1.27	0.18	0.00	0.04	0.00
Bare	2.45	0.16	5.60	1.59	6.31	**82.89**	0.84	0.01	0.02	0.15
Water	8.42	0.00	0.23	0.29	0.13	0.07	**90.54**	0.26	0.05	0.00
Wetlands	15.23	0.06	15.91	11.21	0.27	0.07	9.16	**47.34**	0.75	0.00
Mangroves	12.78	0.00	1.24	3.68	0.29	0.00	4.16	0.17	**77.69**	0.00
Moss/Lichen	11.56	0.00	45.12	0.18	0.09	0.15	0.55	0.25	0.01	**42.09**

SegFormer (4-channel, pretrained): 85.84% overall accuracy

	Trees	Shrubs	Grass	Crops	Built-up	Bare	Water	Wetlands	Mangroves	Moss/Lichen
Trees	**90.36**	1.01	2.68	2.15	1.54	0.05	1.99	0.14	0.05	0.04
Shrubs	16.77	**71.07**	7.84	2.97	1.04	0.17	0.13	0.01	0.01	0.00
Grass	13.40	2.40	**73.02**	7.45	1.94	0.89	0.31	0.30	0.03	0.26
Crops	7.07	0.79	5.93	**84.46**	1.06	0.25	0.24	0.16	0.03	0.01
Built-up	7.23	0.50	2.81	2.54	**85.13**	1.49	0.25	0.01	0.05	0.00
Bare	2.33	0.17	4.73	1.55	5.42	**84.94**	0.71	0.02	0.01	0.12
Water	6.12	0.00	0.25	0.27	0.10	0.12	**92.88**	0.22	0.03	0.00
Wetlands	11.94	0.03	15.59	12.82	0.21	0.07	10.98	**47.71**	0.64	0.02
Mangroves	14.76	0.00	1.77	4.81	0.26	0.15	6.32	0.60	**71.33**	0.00
Moss/Lichen	8.84	0.00	42.90	0.22	0.05	0.67	0.57	0.45	0.00	**46.30**

Prithvi (4-channel, less epochs): 70.07% overall accuracy

Inference (U-Net vs. SegFormer)



Conclusions

- U-Net and SegFormer achieved very similar performance.
- When evaluated on **unseen** cities, performance **drops dramatically**, indicating need for a larger training base.
- Compared to our prior model, **granularity** is much higher, and **inference** is 1,000x faster.
- In the future, we plan to explore **few-shot** experiments to understand data efficiency and model adaptation to distinguishing residential vs. non-residential built-up land.

Citations

[1] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer international publishing.

[2] Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, *34*, 12077-12090.

[3] Szwarcman, D., Roy, S., Fraccaro, P., Gíslason, Þ. E., Blumenstiel, B., Ghosal, R., ... & Moreno, J. B. (2024). Prithvi-eo-2.0: A versatile multi-temporal foundation model for earth observation applications. *arXiv preprint arXiv:2412.02732*.

Acknowledgement

This research was supported in part by an appointment to the Education Collaboration at ORNL, sponsored by the U.S. Department of Energy and administered by ORISE.

Oak Ridge National Laboratory
Oak Ridge Institute for Science and Education — Shaping the Future of Science
U.S. Department of Energy
Georgia Institute of Technology