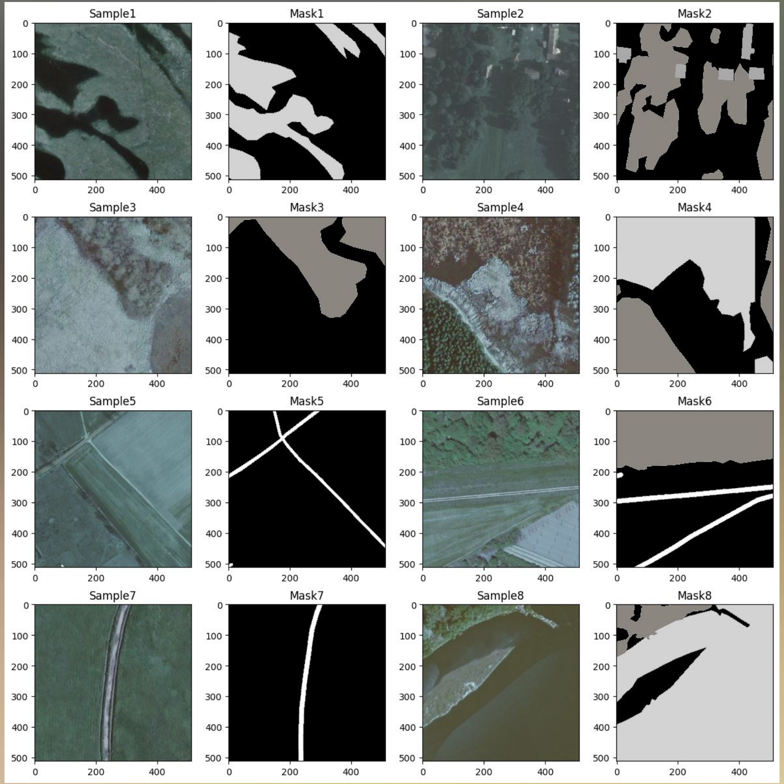
Skynetai: Deep Learning For Multi-Source Semantic Segmentation Of High-Resolution Aerial Imagery

Team Members:

1. Hardik Soni (20CS30023)
2. Rahul Mandal (20CS30039)



I. Introduction

Advancements in remote sensing have enabled the capture of high-resolution aerial images, offering detailed insights for land cover and land use (LCLU) classification. Traditional image analysis methods, however, face limitations due to inefficiencies and high computational demands. Deep learning (DL) methods, particularly convolutional neural networks (CNNs), have gained traction in semantic segmentation tasks, significantly improving both the accuracy and speed of image analysis.

II. Objectives

1

1. Develop a Hybrid Deep Learning Architecture

Design a novel deep learning model, integrating U-Net-based architectures with self-attention mechanisms and separable convolutions.

2

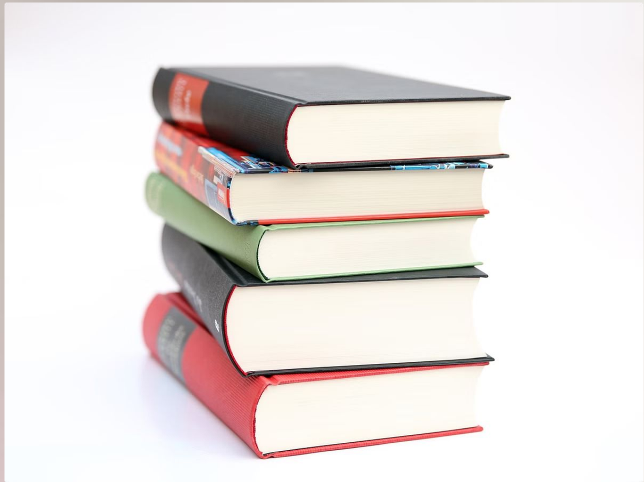
2. Incorporate Multi-Source Imagery

Leverage various remote sensing data sources, such as multispectral, LiDAR, and SAR imagery, in a deep learning framework to improve the precision and robustness of land feature classification.

3

3. Evaluate the Impact of Attention Mechanisms

Investigate how attention mechanisms, including self-attention and channel/spatial attention, contribute to refining deep learning models for more effective segmentation of aerial images.



III. Literature Review

Recent advancements in deep learning for remote sensing and semantic segmentation of high-resolution aerial imagery have focused on improving accuracy, efficiency, and the integration of multi-source data. In the work by Latif et al. (2023), a hybrid approach combining CNNs and attention mechanisms is presented, which significantly enhances the segmentation of complex land cover structures by focusing on key areas while filtering irrelevant data. Similarly, Deng et al. (2019) explore the application of deep neural networks, emphasizing the importance of high-resolution data from sources like Sentinel-2 for land-use classification, while proposing a novel U-Net-based model.

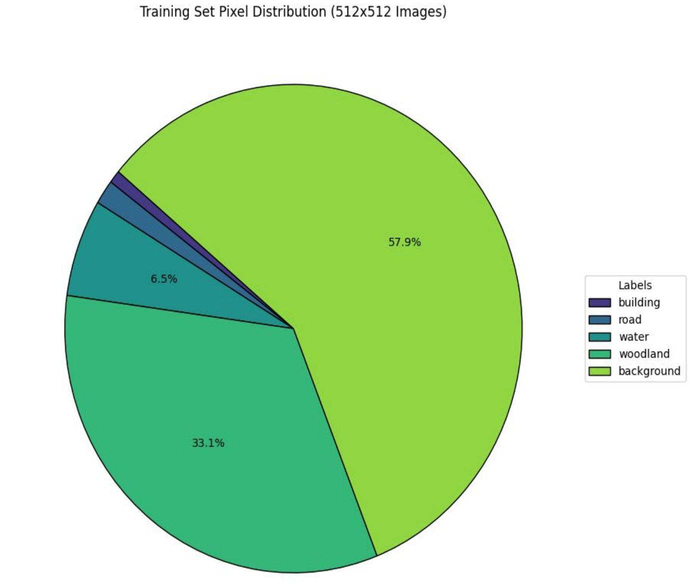
IV. Dataset Description: LandCover.ai

Significance

The LandCover.ai dataset is particularly relevant for critical applications in natural resource management, urban planning, and environmental protection.

Composition

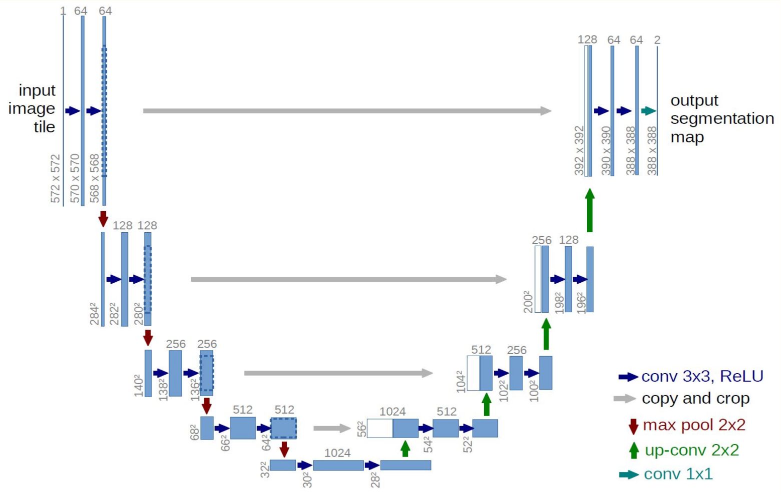
The dataset spans a total area of 216.27 square kilometers across rural Poland. The aerial imagery was acquired at two resolutions: 25 cm per pixel and 50 cm per pixel.

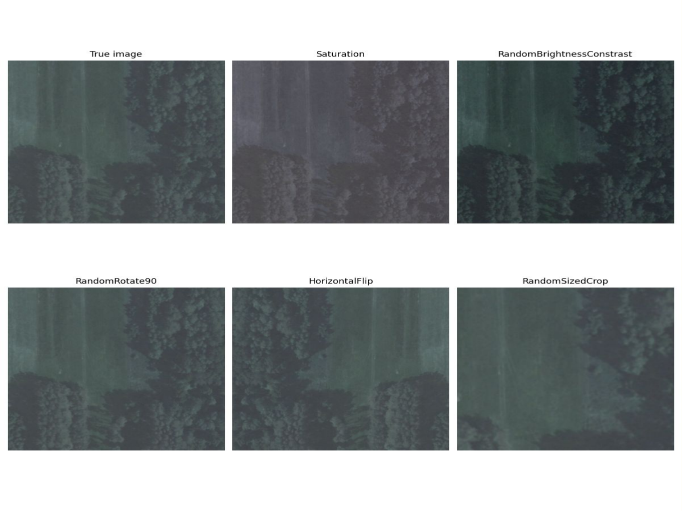


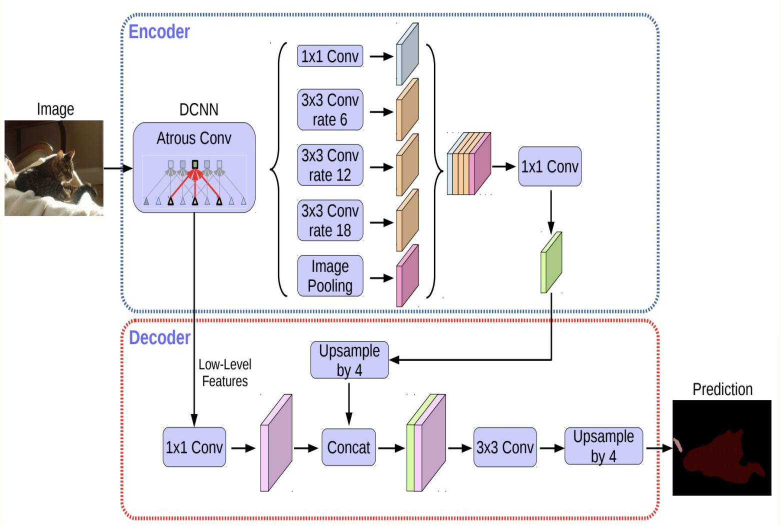
V. Experiments

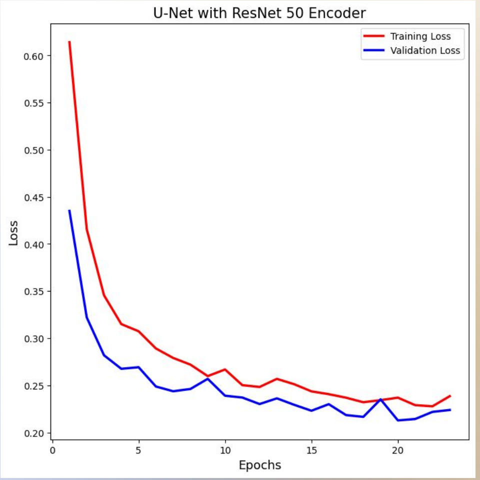
The experiments mainly focused on using augmented images from the dataset and analysing their training and performance on the three main models we have used:

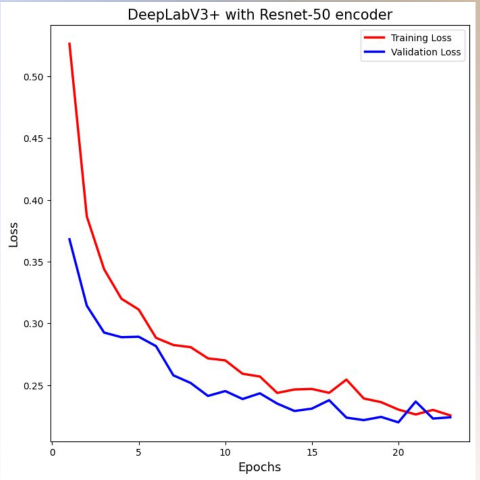
1. U-Net
2. U-Net with Resnet-50 Encoder
3. DeepLabV3+ with Resnet-50 Encoder











VI. Results And Observations

U-Net Model

The U-Net model demonstrates a high overall accuracy of 92.39% across all classes, highlighting its effectiveness for high-resolution segmentation tasks.

U-Net With ResNet 50

The ResNet-U-Net model demonstrates a high overall accuracy of 94.78% across all classes, indicating significant improvements over the Vanilla U-Net model.

DeepLab V3+ With ResNet 50

The DeepLabV3+ with ResNet50 encoder model demonstrates excellent overall performance, with high precision, recall, and F1-scores across all classes.

Class	Precision	Recall	F1-score	Support
* background	0.95	0.93	0.94	239926712
* building	0.85	0.79	0.82	3956554
* woodland	0.90	0.95	0.92	144170112
* water	0.93	0.87	0.90	24388122
* road	0.76	0.61	0.68	7513188

- Total accuracy: 0.9239
- Mean IoU (Jaccard Index): 0.7510

- Class probabilities
* background: 0.998
* building : 0.992
* woodland : 0.998
* water : 0.998
* road : 0.990

Metrics for Vanilla U-Net

Class	Precision	Recall	F1-score	Support
* background	0.96	0.95	0.96	239926712
* building	0.90	0.83	0.86	3956554
* woodland	0.93	0.96	0.94	144170112
* water	0.97	0.95	0.96	24388122
* road	0.83	0.70	0.76	7513188

- Total accuracy: 0.9478
- Mean IoU (Jaccard Index): 0.8205

- Class probabilities
* background: 0.999
* building : 0.994
* woodland : 0.999
* water : 0.999
* road : 0.991

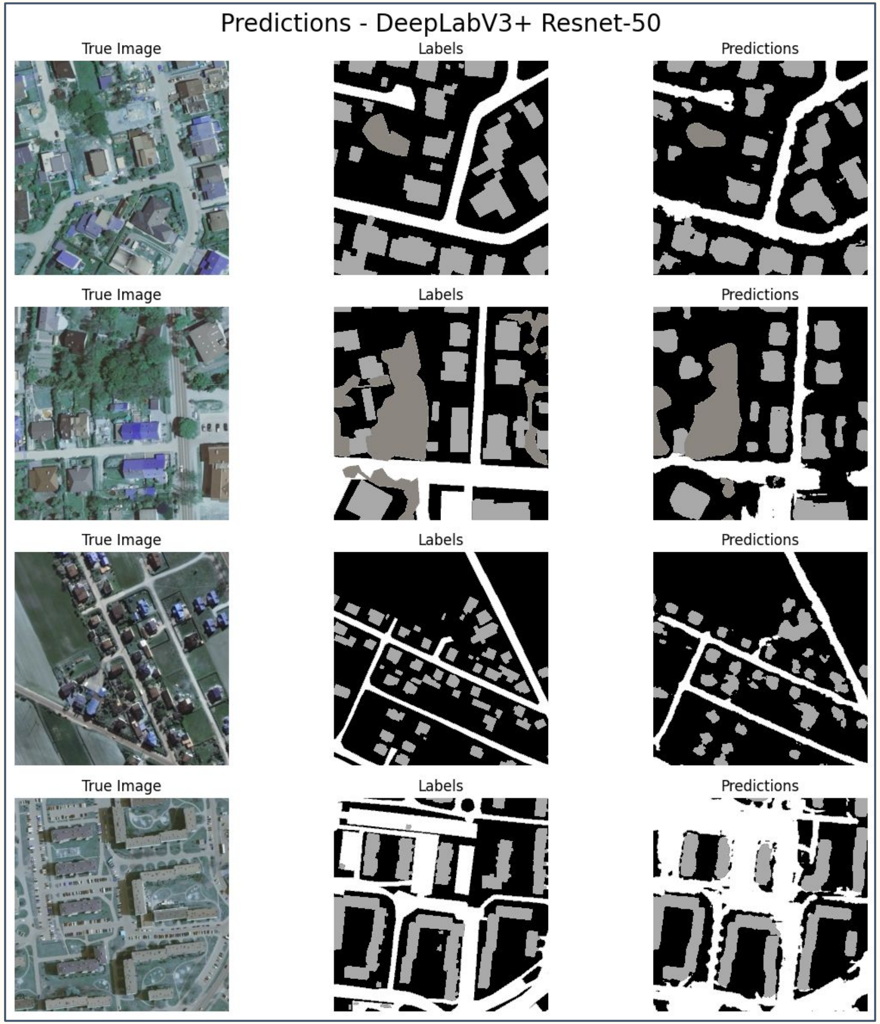
Metrics for DeepLab V3+ with Resnet50 Encoder

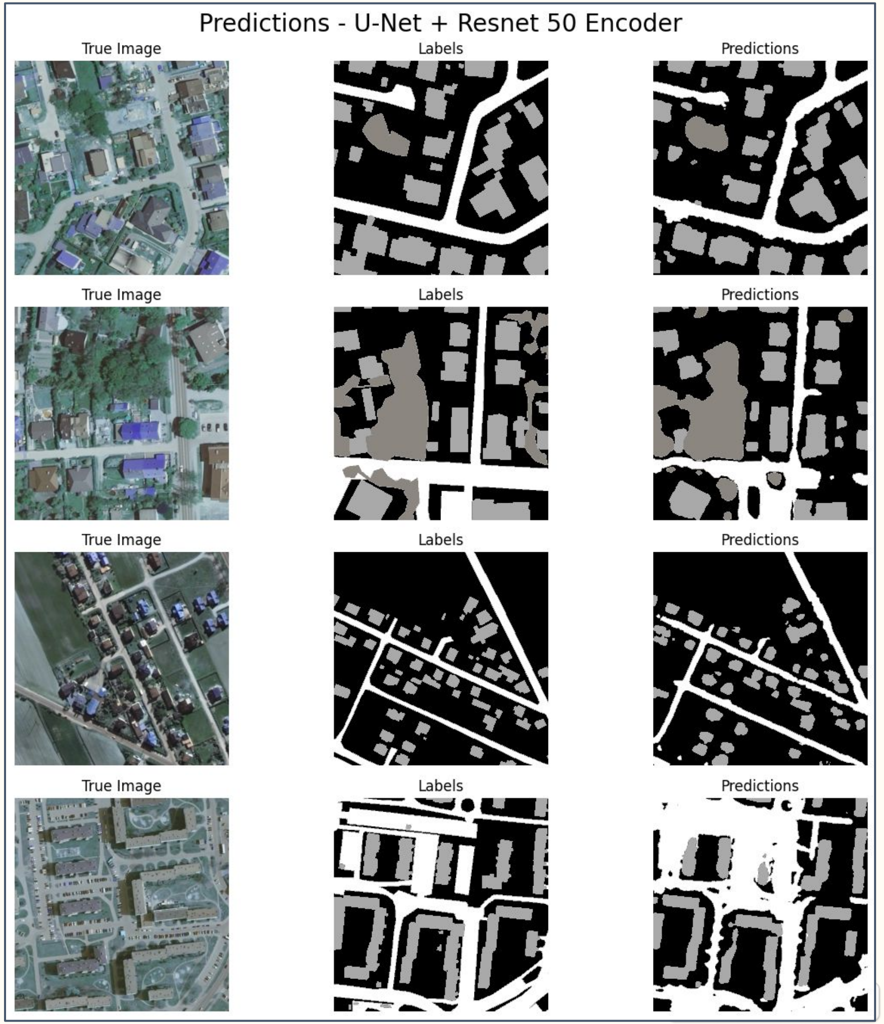
Class	Precision	Recall	F1-score	Support
* background	0.95	0.96	0.96	239926712
* building	0.88	0.82	0.85	3956554
* woodland	0.94	0.94	0.94	144170112
* water	0.97	0.95	0.96	24388122
* road	0.81	0.72	0.76	7513188

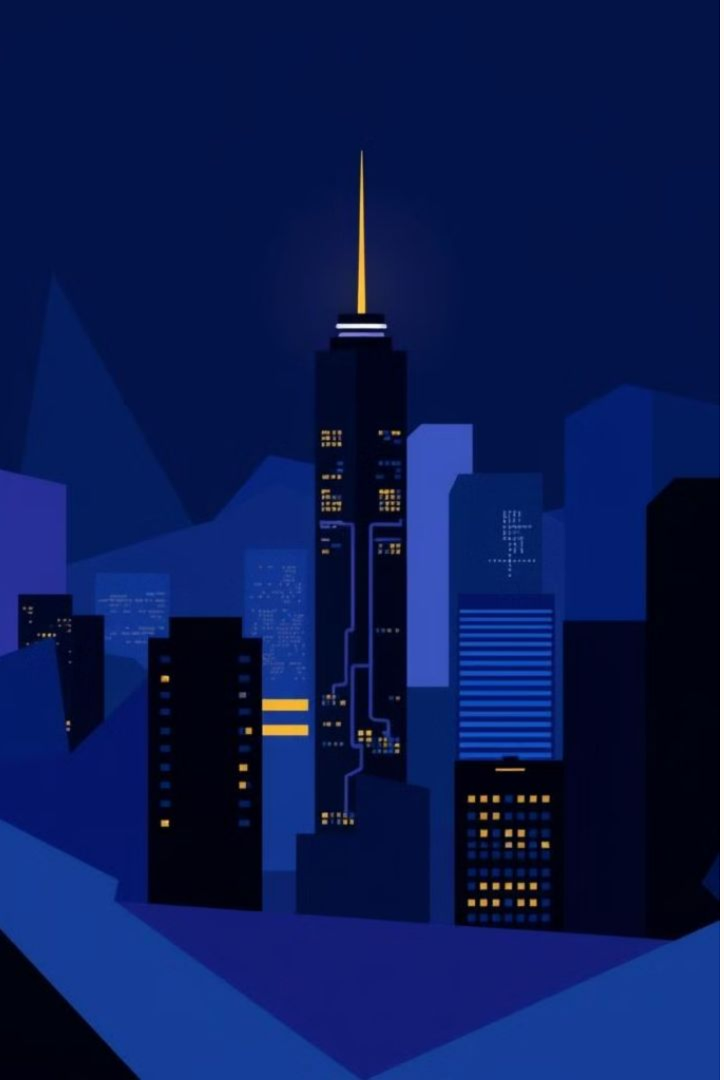
- Total accuracy: 0.9459
- Mean IoU (Jaccard Index): 0.8155

- Class probabilities
* background: 0.996
* building : 0.982
* woodland : 0.995
* water : 0.995
* road : 0.978

Metrics for U-Net with Resnet 50 Encoder


Predictions - DeepLabV3+ Resnet-50
True Image
Labels
Predictions
True Image
Labels
Predictions
True Image
Labels
Predictions
True Image
Labels
Predictions


Predictions - U-Net + Resnet 50 Encoder
True Image
Labels
Predictions
True Image
Labels
Predictions
True Image
Labels
Predictions
True Image
Labels
Predictions



Conclusion

This research has shown the potential of deep learning models for semantic segmentation of aerial imagery, leveraging the rich information from a variety of geospatial datasets. The results demonstrate that **DeepLabV3+** with **ResNet50** encoder achieved the highest performance in segmenting high-resolution aerial imagery, showcasing its ability to effectively handle complex and intricate structures. Future research directions include further exploration of multi-source data integration and the development of more efficient deep learning architectures.