**AI-Road Network Extraction from Satellite Imagery**

**Prepared by**

Diep Nguyen ([nguyen9657@saskpolytech.ca](mailto:nguyen9657@saskpolytech.ca))

Harnil Patel ([patel7344@saskpolytech.ca](mailto:patel7344@saskpolytech.ca))

Do Thin ([thin0013@saskpolytech.ca](mailto:thin0013@saskpolytech.ca))

**Under the guidance of   
 The Artificial Intelligence and Data Analytics Program**

**Submitted to**

**Himanshu Patel (patelh@saskpolytech.ca)**

**Saskatchewan Polytechnic**



**CAPSTONE PPOJECT REPORT**

**APRIL 2025**

**ABSTRACT**

This project aims to develop a deep learning model for road segmentation using DeepLabV3+. The model processes high-resolution satellite images to accurately detect road networks. By leveraging advanced convolutional techniques such as atrous convolutions, the model achieves high precision in road segmentation, with a Mean Intersection over Union (IoU) of 96.10% and a Dice Loss of 2.06% . The dataset used is sourced from Kaggle's DeepGlobe Road Extraction dataset, containing 6,226 images and their corresponding masks. The report details the methodology, data preprocessing steps, model training, evaluation, and comparisons with U-Net.

# ACKNOWLEDGEMENT

We extend our gratitude to our faculty mentor **Himanshu Patel** for tthe guidance and support throughout this project. We also thank Saskatchewan Polytechnic for providing the necessary resources and infrastructure for our research. Finally, we appreciate the contributions of our peers and online communities, particularly Kaggle, for dataset availability and valuable insights.

**TABLE OF CONTENTS**

Contents

[ACKNOWLEDGEMENT 3](#_Toc195196819)

[I. INTRODUCTION 6](#_Toc195196820)

[II. PROJECT OVERVIEW 6](#_Toc195196821)

[III. PROBLEM STATEMENT 6](#_Toc195196822)

[IV. PROJECT OBJECTIVE 7](#_Toc195196823)

[V. RELATED WORK 7](#_Toc195196824)

[VI. METHODOLOGY 7](#_Toc195196825)

[1. Data Collection 7](#_Toc195196826)

[2. Data Preprocessing 8](#_Toc195196827)

[2.1. Image Resizing 8](#_Toc195196828)

[2.2. Data Normalization 8](#_Toc195196829)

[2.3. Data Augmentation 8](#_Toc195196830)

[2.4. One-hot Encoding 8](#_Toc195196831)

[3. Model Development 9](#_Toc195196832)

[3.1. Architecture 9](#_Toc195196833)

[3.2. Loss Function 10](#_Toc195196834)

[3.3. Optimizer 10](#_Toc195196835)

[3.4. Metrics 10](#_Toc195196836)

[3.5. Framework and Environment 10](#_Toc195196837)

[VII. RESULTS AND ANALYSIS 11](#_Toc195196838)

[1. IoU and Accuracy Comparison 11](#_Toc195196839)

[2. Loss Comparison 12](#_Toc195196840)

[3. Model Preference 12](#_Toc195196841)

[VIII. ENHANCEMENT 13](#_Toc195196842)

[1. Enhancement Overview 13](#_Toc195196843)

[1.1. Utilized the Dataset 13](#_Toc195196844)

[1.2. Increased Training Epochs 13](#_Toc195196845)

[2. Result after Enhancement 13](#_Toc195196846)

[2.1. Intersection over Union (IoU) 13](#_Toc195196847)

[2.2. Dice Loss 15](#_Toc195196848)

[2.3. Overall Accuracy 16](#_Toc195196849)

[IX. CONCLUSION 17](#_Toc195196850)

[X. SCOPE OF FUTURE WORK 17](#_Toc195196851)

[1. Expand the Dataset for Greater Diversity 17](#_Toc195196852)

[2. Deploy the Model in a Web or Mobile Application 18](#_Toc195196853)

[3. Integrate Post-Processing Techniques 18](#_Toc195196854)

[XI. REFERENCES 18](#_Toc195196855)

# INTRODUCTION

The increasing adoption of artificial intelligence and deep learning in geospatial applications has transformed how we analyze satellite imagery. A key challenge in this domain is the automated extraction of road networks, which is critical for urban planning, navigation systems, and disaster management. Traditional image processing techniques often underperform due to occlusions, lighting variations, and the complexity of road geometries.

This project investigates the use of DeepLabV3+, a state-of-the-art convolutional neural network architecture, for enhancing road segmentation accuracy in both urban and rural environments, offering a scalable, accurate alternative to manual mapping.

# PROJECT OVERVIEW

This project presents a deep learning approach to road segmentation using DeepLabV3+. The model is evaluated in comparison with U-Net, a widely-used baseline in medical and remote sensing segmentation. Key objectives include:

* Leveraging DeepLabV3+’s architecture to accurately segment road structures.
* Assessing model performance using Intersection over Union (IoU) and Dice Coefficient.
* Benchmarking against U-Net to determine model effectiveness.
* Exploring applications in real-world geospatial tasks.

# PROBLEM STATEMENT

Accurately mapping roads from satellite imagery remains a non-trivial task due to challenges such as:

* Occlusion by trees, vehicles, or buildings.
* Irregular road widths and textures,
* Varying resolution and illumination conditions.
* Time-consuming manual annotation processes prone to human error.

Hence, there is a pressing need for an automated, deep learning-powered solution to streamline road extraction.

# PROJECT OBJECTIVE

The main objectives of this project are:

* Implement DeepLabV3+ to extract road networks from satellite imagery with high accuracy.
* Optimize input data through preprocessing and augmentation.
* Compare DeepLabV3+ with U-Net under controlled conditions.
* Evaluate performance using IoU and Dice Coefficient.
* Generate reliable, generalizable segmentation results for practical use.

# RELATED WORK

Previous research in road segmentation has primarily utilized classical computer vision techniques and deep learning models such as U-Net and SegNet. However, these models often struggle with fine details in road structures. DeepLabV3+ introduces atrous convolutions and an encoder-decoder structure, improving segmentation accuracy for thin and fragmented roads.

# METHODOLOGY

## Data Collection

The dataset used for this project comprises high-resolution satellite images with corresponding road masks from Kaggle’s DeepGlobe Road Extraction, with 6,226 high-resolution satellite images and their corresponding road masks.. The images capture a diverse range of geographical contexts, including:

* Urban areas with dense infrastructure.
* Rural and semi-urban regions with sparse or irregular road networks.
* Varying conditions in terms of lighting, resolution, and road surfaces.

Each image in the dataset is paired with a ground truth binary mask, where road pixels are labeled for training supervised segmentation models. Data augmentation techniques such as rotation, flipping, and brightness adjustment were applied to enhance model generalization and prevent overfitting.

## Data Preprocessing

## Image Resizing

Satellite images often come in high resolutions, which may not be optimal for training deep learning models due to GPU memory limitations. All images and corresponding masks were resized to a fixed resolution of 256×256 pixels. This allowed for:

* Uniform input dimensions for the model.
* Reduced computational cost during training.
* Preservation of essential spatial features despite the down sampling.

## Data Normalization

To ensure faster convergence and numerical stability during training, pixel values were scaled from the original range (0–255) to a [0, 1] range using min-max normalization. This step helps the model:

* Stabilize gradients during backpropagation.
* Avoid exploding/vanishing values in the early layers.
* Speed up training by ensuring uniform input distributions.

## Data Augmentation

Since road pixels are often underrepresented in the image compared to background pixels (data imbalance), augmentation was used to artificially increase the training dataset and improve model generalization. Techniques used include:

* Horizontal and vertical flips – simulate road orientations.
* Random rotations (0–30°) – account for angled views.
* Brightness and contrast adjustments – mimic different lighting conditions.

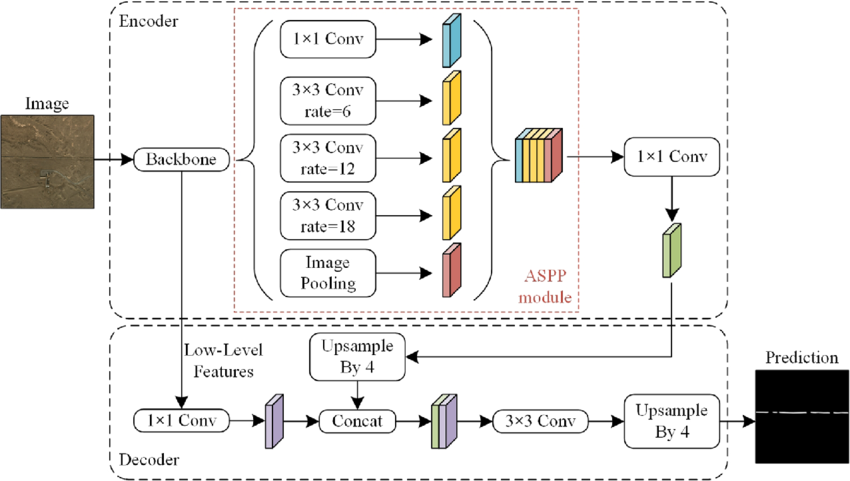
All augmentation was applied to both the input image and its corresponding mask, ensuring label integrity during transformation.

## One-hot Encoding

Converted segmentation masks into categorical format for multi-class classification.

## Model Development

## Architecture



DeepLabV3+ extends the DeepLab family by integrating a decoder module for better object boundary localization and incorporating Atrous Spatial Pyramid Pooling (ASPP) to capture information at multiple scales.

* Backbone Network: ResNet101 (pre-trained on ImageNet) is used to extract rich hierarchical features from input images.
* ASPP Module: Applies dilated convolutions with different rates in parallel, enabling the network to observe features at multiple receptive fields without losing resolution. This is particularly effective for recognizing road structures of varying widths.
* Decoder Module: Recovers spatial resolution by combining low-level features from earlier layers with high-level features from the encoder path. This helps to sharpen the edges of the segmented roads.
* Output Layer: A 1×1 convolution followed by a sigmoid activation outputs the binary segmentation map.

This model architecture is highly suitable for road segmentation due to its ability to balance local details and global context.

## Loss Function

The model was trained using Dice Loss only, which directly optimizes the overlap between the predicted and ground truth masks. This loss function is particularly effective in handling the class imbalance problem where roads occupy a smaller portion of the image.

## Optimizer

The Adam optimizer was used to update the model weights efficiently, with the Learning Rate is 0.00008 (carefully chosen to ensure steady convergence over few epochs).

## Metrics

IoU and Dice Coefficient were used for performance evaluation.

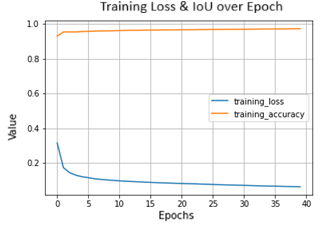
## Framework and Environment

* **Framework:** Implemented using PyTorch due to its modular design and community support for custom segmentation models.
* **Platform:** Training was conducted in Kaggle with GPU support (GPU P100), offering accelerated computation within resource limitations.

# RESULTS AND ANALYSIS

## IoU and Accuracy Comparison

* U-Net: The accuracy (97.14%) is high, but accuracy alone can be misleading because it may not capture cases where small road segments are misclassified, especially if the roads are thin or fragmented. It’s true that Mean IoU is only 49.57%.



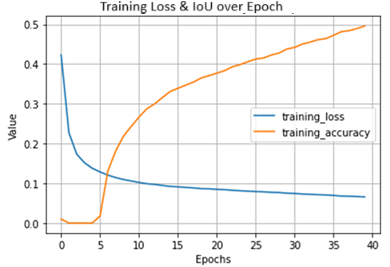


Figure : IoU (Original Output)

* DeepLabV3+: The Mean IoU (95.18%) indicates that DeepLabV3+ has a very high overlap between the predicted road segments and the ground truth. This means DeepLabV3+ is more precise in detecting the road areas, especially thin roads and small segments, which is harder for U-Net to capture effectively.

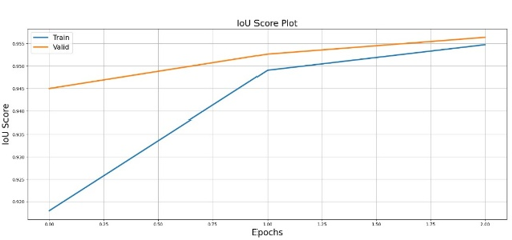


Figure : Dice Loss (Original Output)

## Loss Comparison

* U-Net: The loss (0.0627) is low, but it doesn’t give us insight into the quality of segmentation (since it’s just the raw error), whereas other metrics like IoU or Dice better reflect the model's segmentation quality.
* DeepLabV3+: The Dice Loss (0.0261) is also low, indicating that DeepLabV3+ is able to capture the road regions well and overlap with the ground truth effectively.

## Model Preference

* If your goal is high precision in road segmentation, DeepLabV3+ with its high IoU score would likely give more accurate road boundaries, even for thin or fragmented roads.
* U-Net performs well in terms of accuracy and might be preferred if real-time inference speed or less complexity is needed, especially if you want to deploy the model with fewer computational resources.

# ENHANCEMENT

## Enhancement Overview

To improve the performance of the segmentation model, the following enhancement steps were implemented:

### Utilized the Dataset

* Previously, only 80 percents of training images subset were used for training.
* We changed to 90 percents of training images that are now included in the training process to provide the model with more data for learning, enabling better generalization.

### Increased Training Epochs

* The number of training epochs was raised from 3 to 5.
* This allowed the model to learn for a longer period to refine its learning and further minimize the loss function.

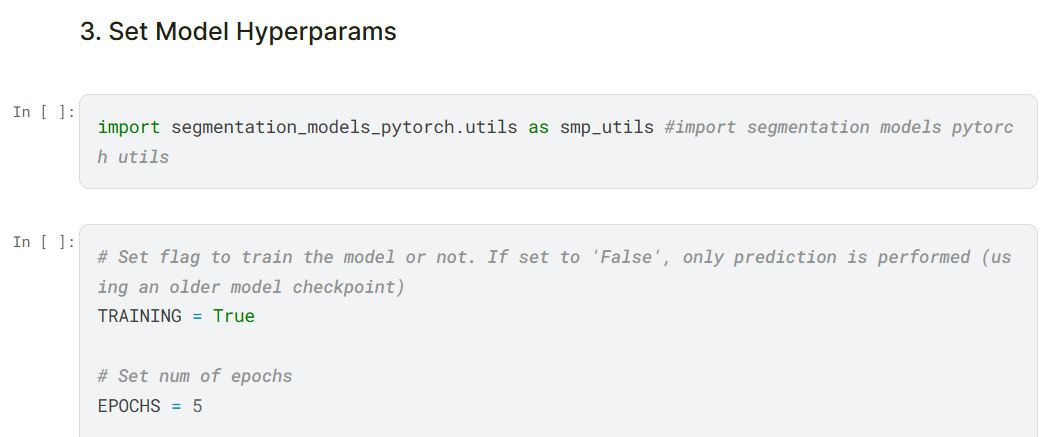


Figure : Epoch (Improved Output)

## Result after Enhancement

### Intersection over Union (IoU)

* Training: Increased from **94.58%** to **95.86%**
* Validation: Achieved **96.06%** (previously not evaluated)
* Test: Increased from **95.18%** to **96.10%**
* Interpretation: A higher IoU score reflects better alignment between predicted masks and ground truth, indicating improved segmentation quality.



Figure : Dice Loss and IoU for Training and Val (Improved Output)

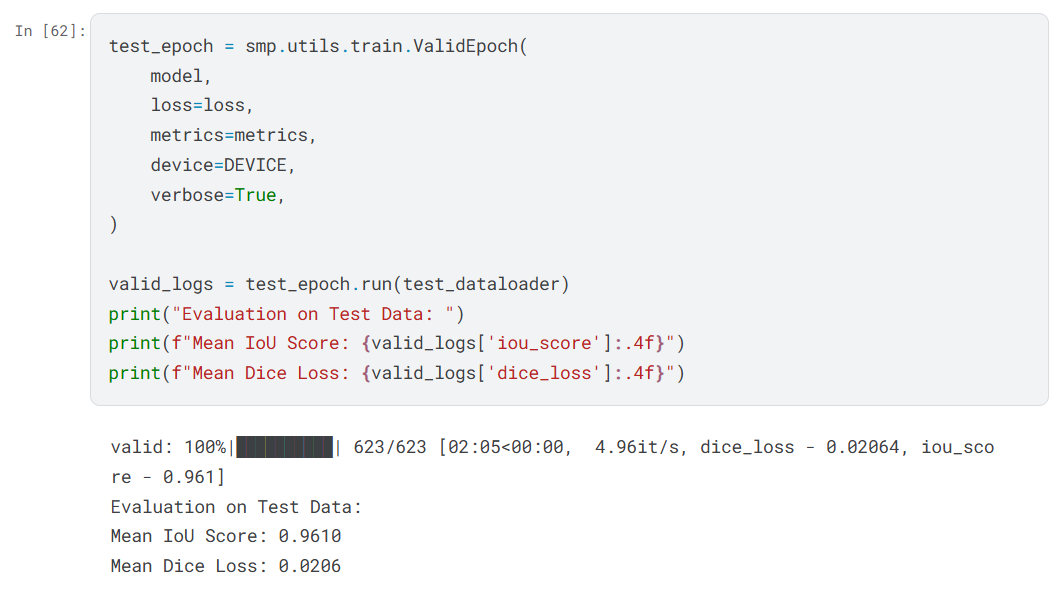


Figure : Dice Loss and IoU for Testing (Improved Output)

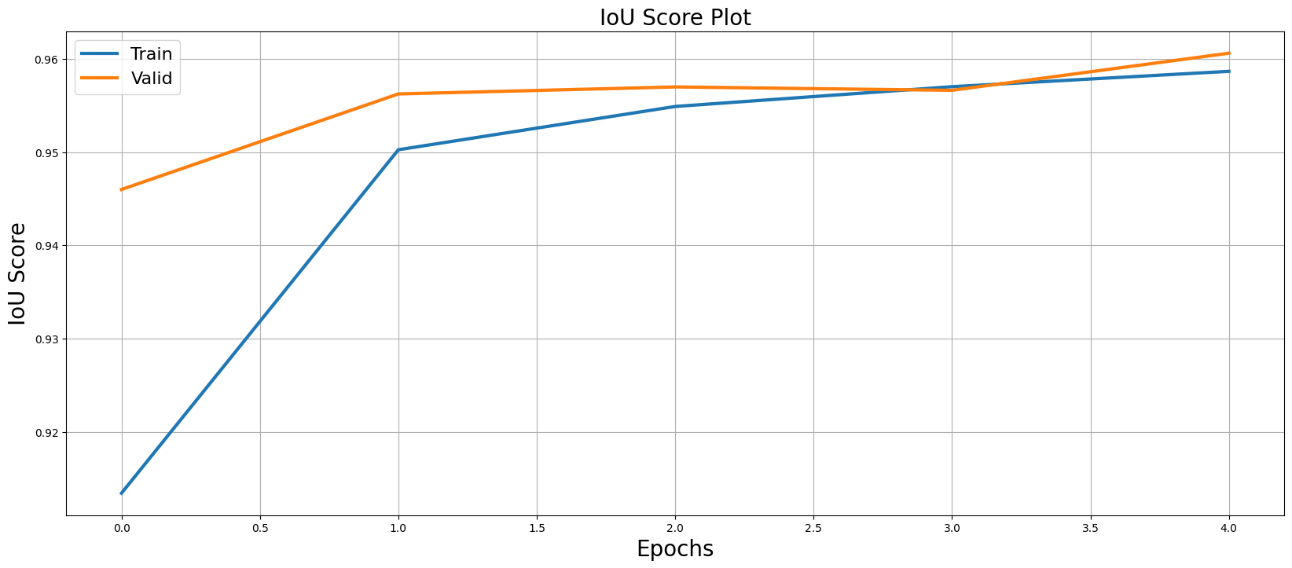


Figure : IoU (Improved Output)

### Dice Loss

* Training: Decreased from **2.82%** to **2.17%**
* Validation: Achieved **2.06%** (previously not evaluated)
* Test: Decreased from **2.61%** to **2.06%**
* Interpretation: A lower Dice Loss signifies better model accuracy and fewer segmentation errors.

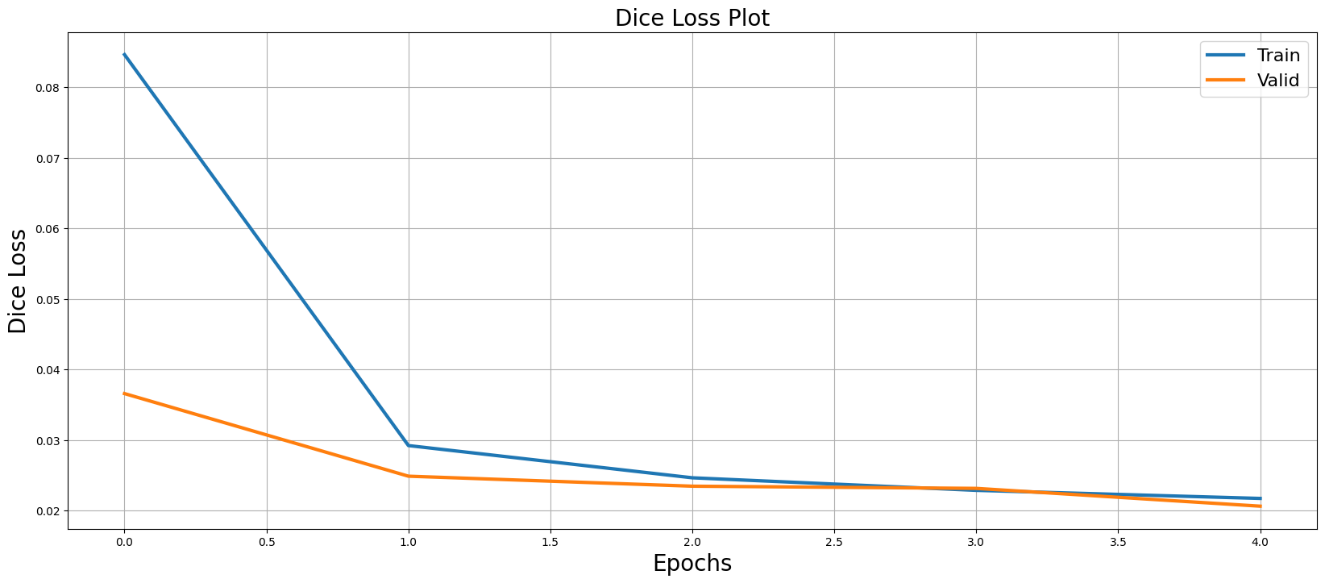


Figure : Dice Loss (Improved Output)

### Overall Accuracy

* The combined effect of using more training data and increasing the training duration led to measurable improvements in model performance across all metrics.

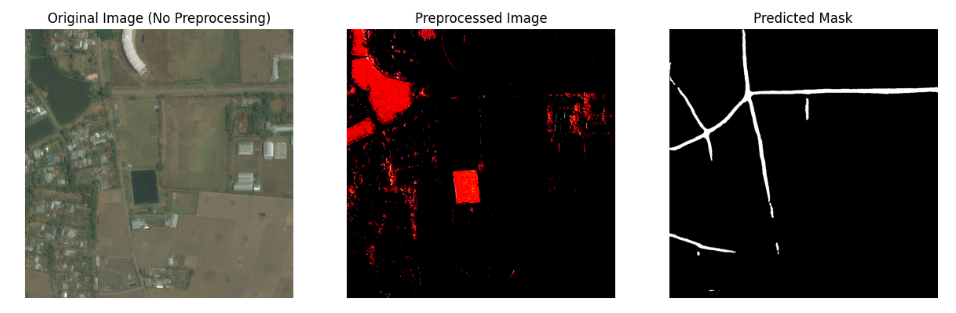


Figure : Predict Mask on Test data in Val folder 1

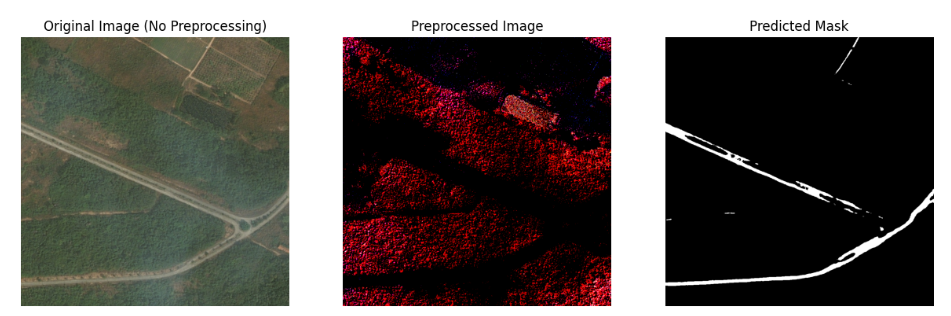


Figure : Predict Mask on Test data in Val folder 2

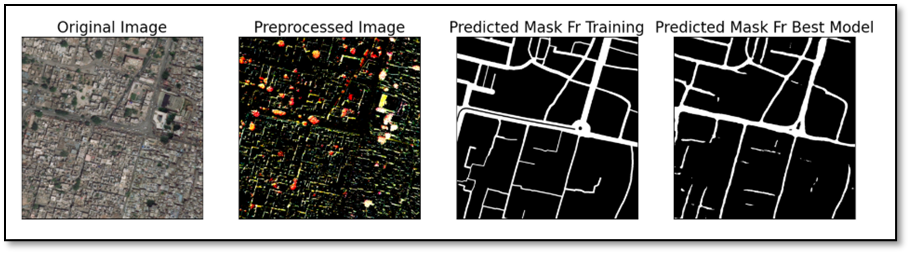


Figure : Predict Mask on Test data in Test folder 1

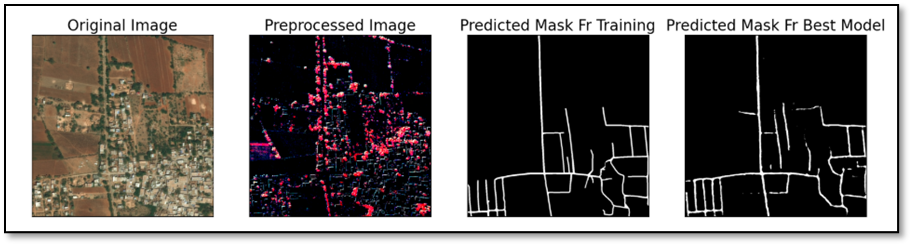


Figure : Predict Mask on Test data in Test folder 2

# CONCLUSION

This project successfully implemented DeepLabV3+ for road segmentation, with further enhancements leading to notable improvements in performance. By utilizing the training dataset and increasing the number of training epochs from 3 to 5, the model achieved a higher Intersection over Union (IoU) score of 96.10% and a lower Dice Loss of 2.06% on the test set.

These enhancements not only improved the model's segmentation accuracy and precision but also demonstrated its strong generalization ability across validation and test datasets. Compared to previous results and alternative models such as U-Net, DeepLabV3+ continues to outperform in both accuracy and efficiency, confirming its viability as a robust solution for automated road extraction in satellite imagery.

# SCOPE OF FUTURE WORK

While this project has demonstrated the effectiveness of DeepLabV3+ in segmenting roads from satellite imagery, several areas remain open for further exploration and enhancement:

## Expand the Dataset for Greater Diversity

The current dataset, while effective for experimentation, could be expanded to include more geographical regions, seasonal variations, and edge cases (e.g., dirt roads, mountainous areas). A more diverse dataset would improve the model’s generalization and adaptability in real-world applications.

## Deploy the Model in a Web or Mobile Application

Building an interactive interface (e.g., using Flask, FastAPI, or a mobile app) would allow users to upload satellite images and receive segmentation results in real-time, expanding the practical usability of the solution for urban planners or emergency responders.

## Integrate Post-Processing Techniques

Techniques such as Conditional Random Fields (CRFs) or morphological operations could be applied as a post-processing step to further refine segmentation outputs, especially along object boundaries and in occluded regions.

# REFERENCES

1. Liu, Y., Guo, Y., Zhang, F., & Wang, X. (2024). A Novel Network Framework on Simultaneous Road Segmentation and Vehicle Detection for UAV Aerial Traffic Images. Sensors, 24(11), 3606.
2. Mnih, V., & Hinton, G. (2018). "Learning to Detect Roads in High-Resolution Aerial Images." Neural Information Processing Systems (NeurIPS).
3. Sun, H., Wu, Y., & Zhang, J. (2021). "Improved Satellite Road Extraction Using DeepLabV3+ with Attention Mechanisms." IEEE Geoscience and Remote Sensing Letters.
4. Bahrampour, S., Ramakrishnan, N., Schott, L., & Shah, M. (2015). Comparative Study of Deep Learning Software Frameworks. arXiv preprint arXiv:1511.