HUST

ĐẠI HỌC BÁCH KHOA HÀ NỘI

HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.

ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

# Capstone project

## Road Segmentation for Aerial Images

**Project Class : 157213 — Computer Vision**

**Group 8**


| | |
|---|---|
| Phùng Tiến Thành | 20225531 |
| Đinh Bảo Hưng | 20225446 |
| Nguyễn Phan Thắng | 20225529 |


ONE LOVE. ONE FUTURE.

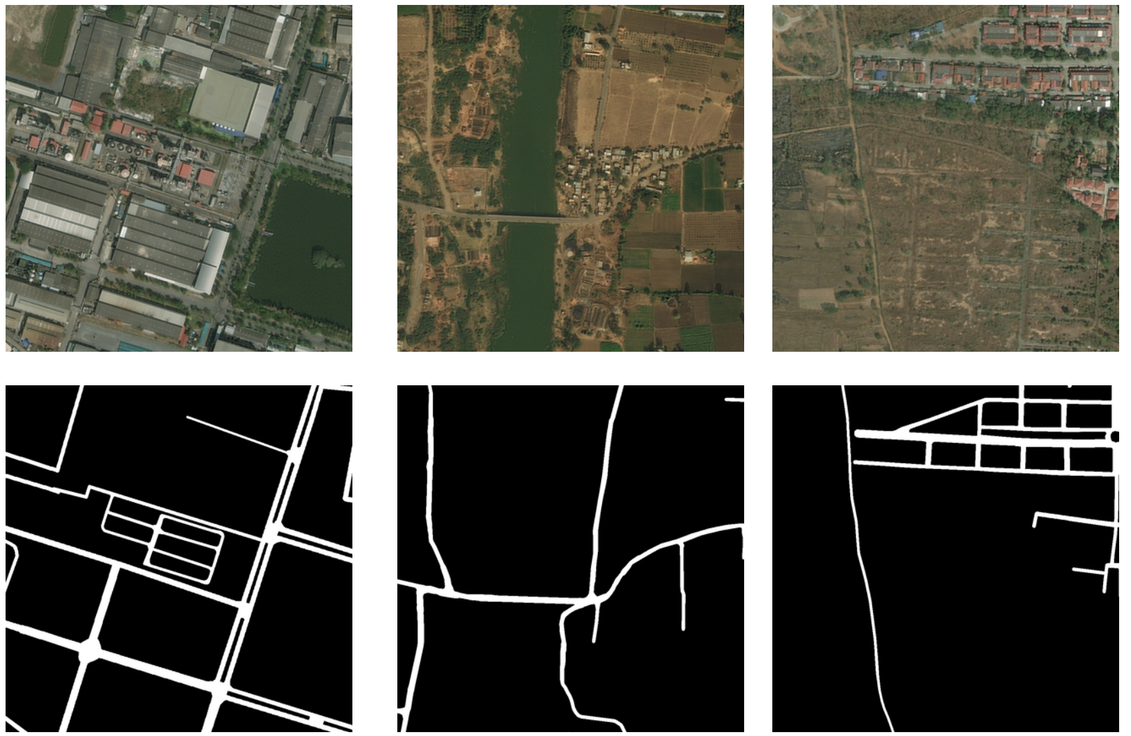
# Table of content

# I. Problem

# I. Problem

## 1.1. Problem statement

- **Objective:** Automatically segment roads from aerial/satellite images using computer vision and deep learning.
- **Input:** Aerial images (RGB, multispectral, or LiDAR).
- **Output:** A binary mask highlighting road pixels.



## 1.2. Challenge

- Roads may be **occluded** (by trees, shadows, clouds).
- **Diverse shapes** (straight, curved, and intersections).
- **Varying widths** (highways vs. narrow alleys).
- **Visual noise** (similar textures to rooftops, parking lots).
- **Lighting/weather conditions** (haze, low resolution, shadows).

# II. Solution

# II. Solution

**Datasets Used:**

- Massachusetts Roads Dataset
- DeepGlobe Road Extraction Dataset

Both datasets consist of high-resolution satellite images paired with binary masks indicating road locations. Fortunately, they share the same labeling convention:

- **White pixels**: Roads
- **Black pixels**: Background

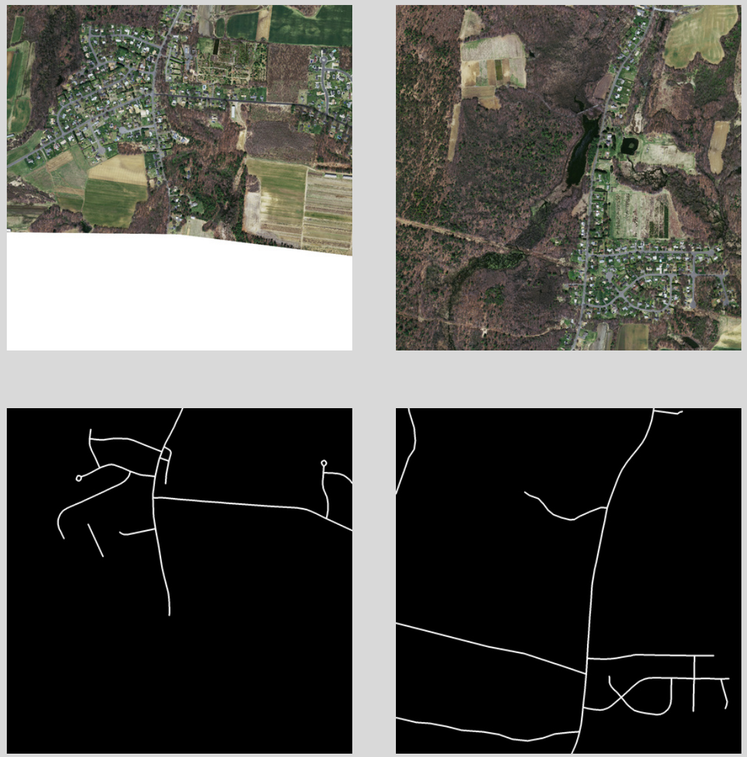
This consistency eliminates the need for additional label standardization, simplifying the integration process and ensuring direct compatibility between the two datasets.

**Models:**

- D-LinkNet: Uses ResNet-34 + dilated convolution for improved feature retention.
- SegFormer (fine-tune).

# 2.1 Data Acquisition & Preprocessing

**Massachusetts Roads Dataset:** 1171 aerial images (with available labels) of the state of Massachusetts,  1500×1500 pixels in size. Resolution is 1 pixel per square meter.

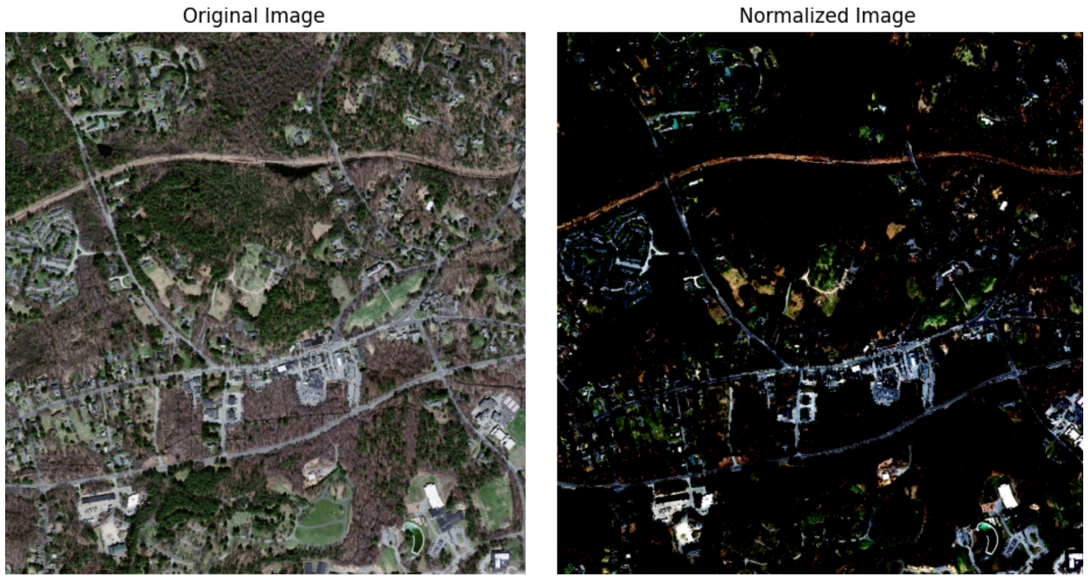


**DeepGlobe Road Extraction Dataset:** 6226 satellite images (1243 validation and 1101 test images without masks) in RGB, size 1024x1024, 50cm pixel resolution.

# 2.1 Data Acquisition & Preprocessing

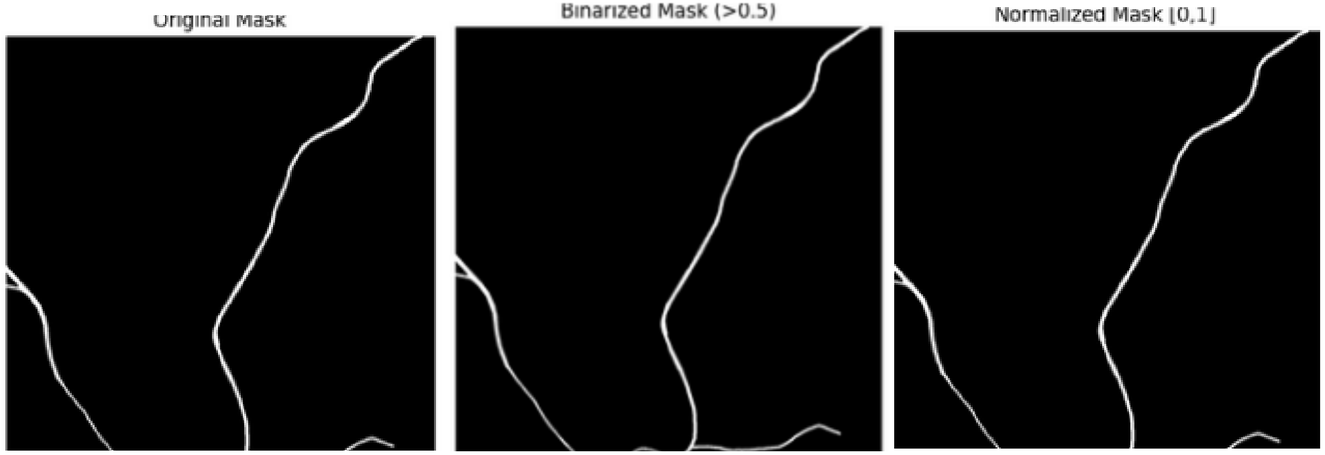
## Image Processing:

- Combine the 2 datasets and resize all the images to 1024x1024 pixels.
- Normalization formula: $I_{norm} = (I \div 255) \times 3.2 - 1.6$




## Masks Processing:

- Combine the 2 datasets and resize all the masks to 1024x1024 pixels.
- Scale pixel values from [0, 255] to [0, 1], applying the threshold = 0.5



# 2.1 Data Acquisition & Preprocessing

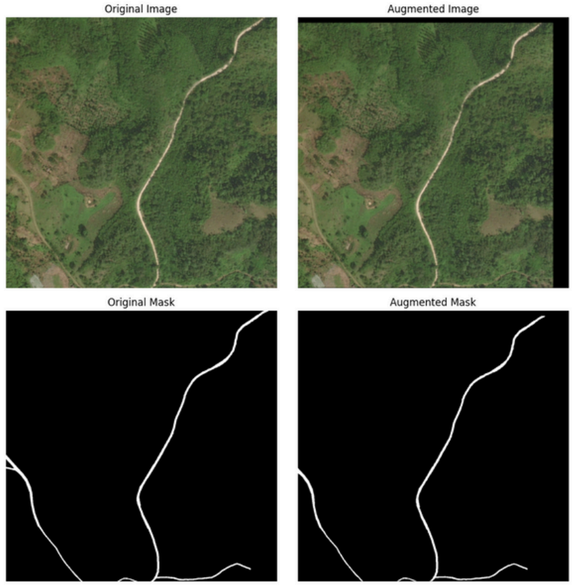
**Augmentations:**

- These augmentations introduced variations in color, orientation, and geometry of training images, simulating diverse imaging conditions and mitigating overfitting risks.

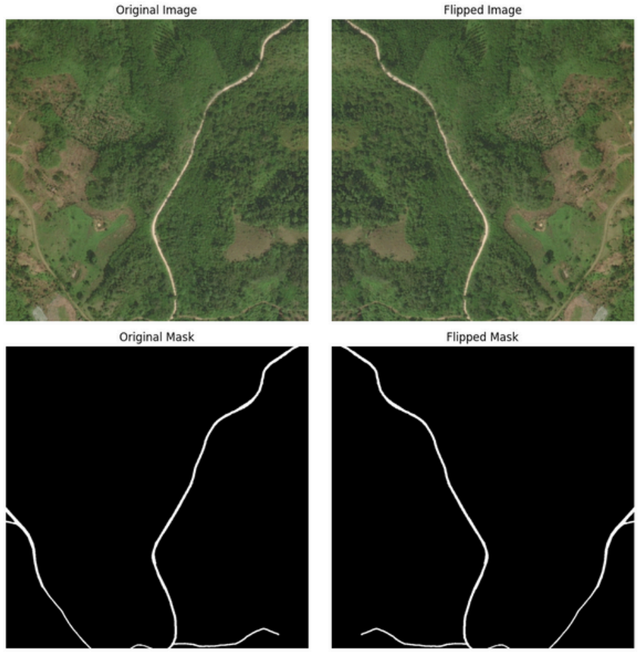
- All augmentations were dynamically applied during data loading using OpenCV functions within a custom PyTorch Dataset class, ensuring unique image variations in each training epoch.

# 2.1 Data Acquisition & Preprocessing

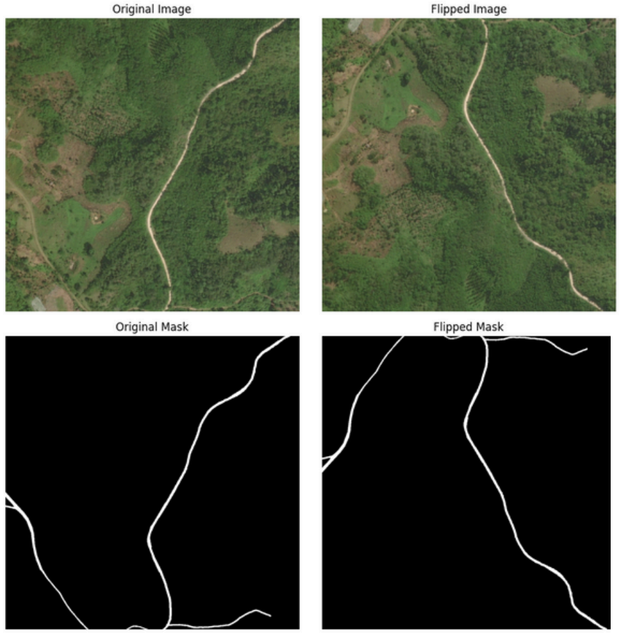
**Augmentations:** list specific techniques



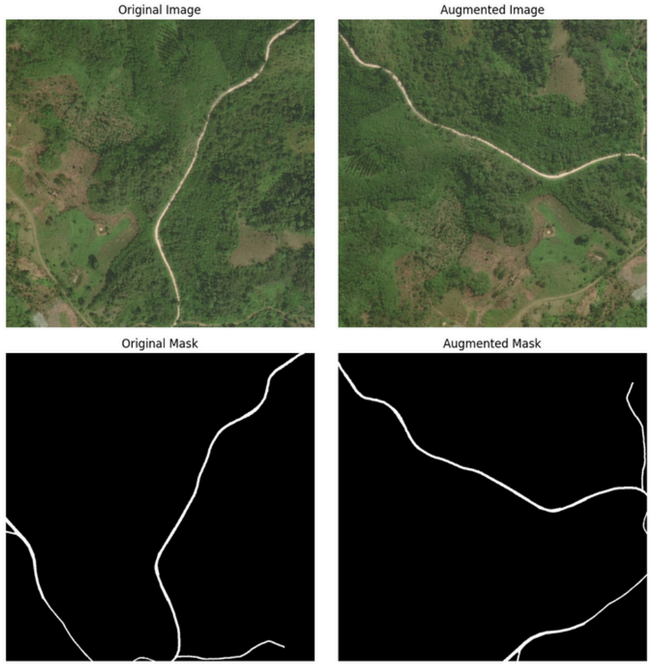

Random Shift, Scale, and Rotate




Horizontal flip

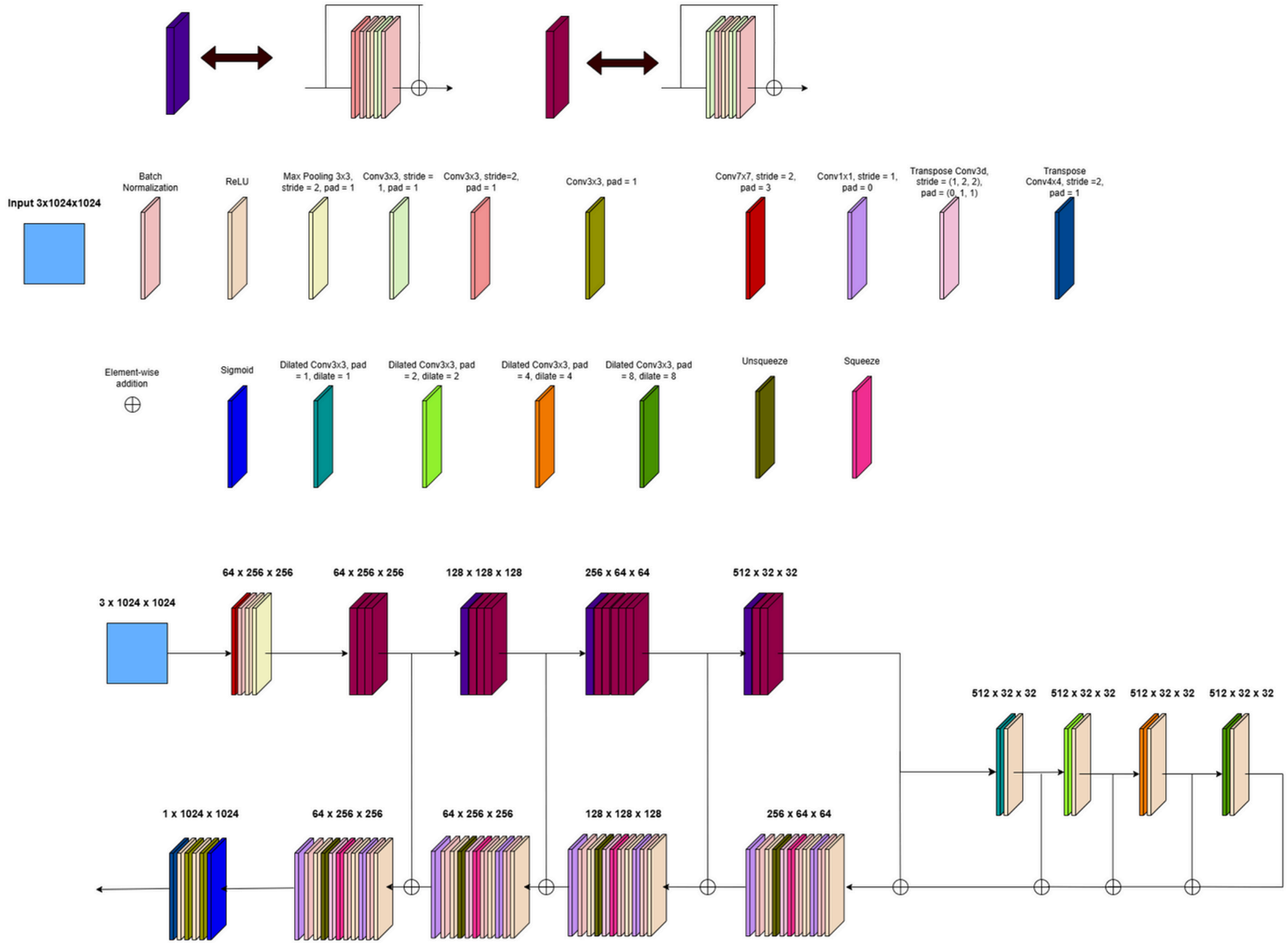



Vertical flip




Rotate 90 degree

# 2.2 Model Selection & Implementation

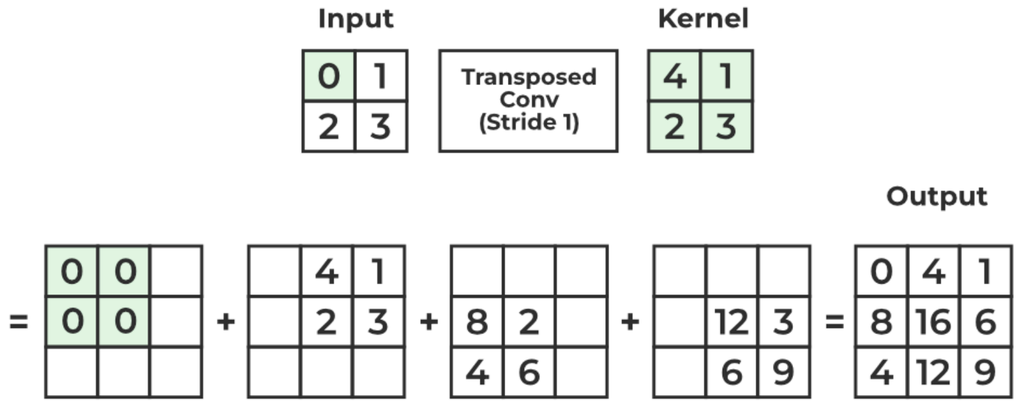



**D-LinkNet architecture**

# 2.2 Model Selection & Implementation

**Transpose Convolution Operation:**

- Regular convolution applies a sliding kernel to compress input data into a smaller output.
- Transposed convolution expands a small input into a larger output by inserting zeros/spaces between pixels before applying the kernel, effectively "reconstructing" higher-resolution data.

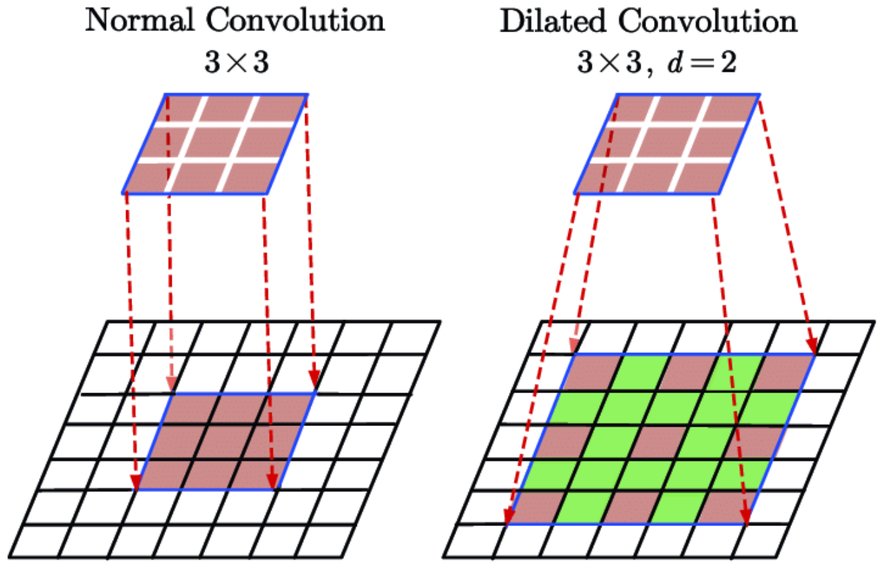



Transpose Convolution

**Dilated Convolution Operation:**

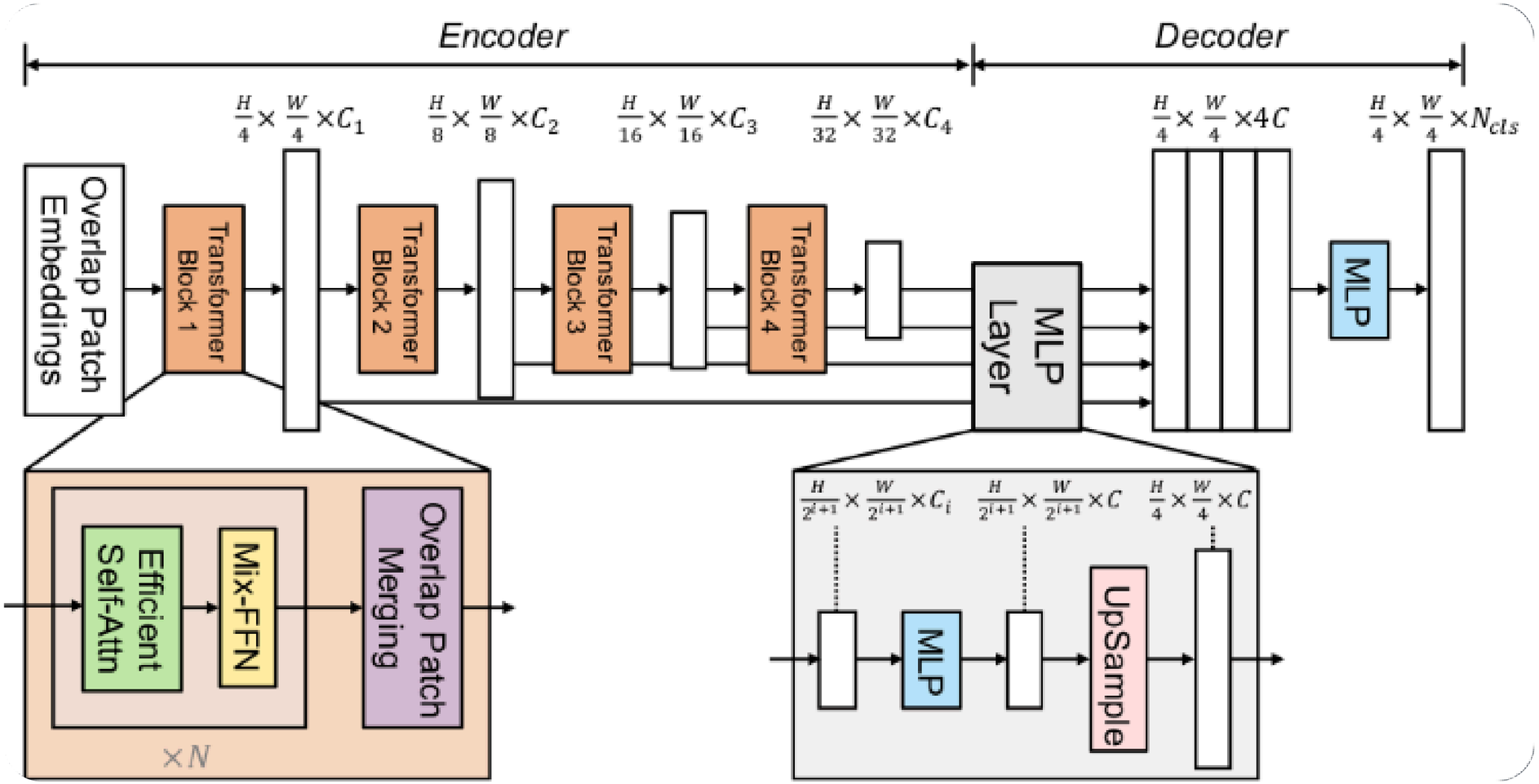
This approach is especially valuable for segmentation tasks, as it:

- Captures broader contextual information
- Maintains spatial resolution
- Avoids parameter inflation




Dilated Convolution

# 2.2 Model Selection & Implementation




**Segformer architecture**

# III. Experimental Results

# 3. Experimental Results

## Evaluation Metrics

- Dice coefficient
- Intersection over Union (IoU)
- Confusion matrix, precision, recall, f1-score

$$\text{Dice} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}}$$

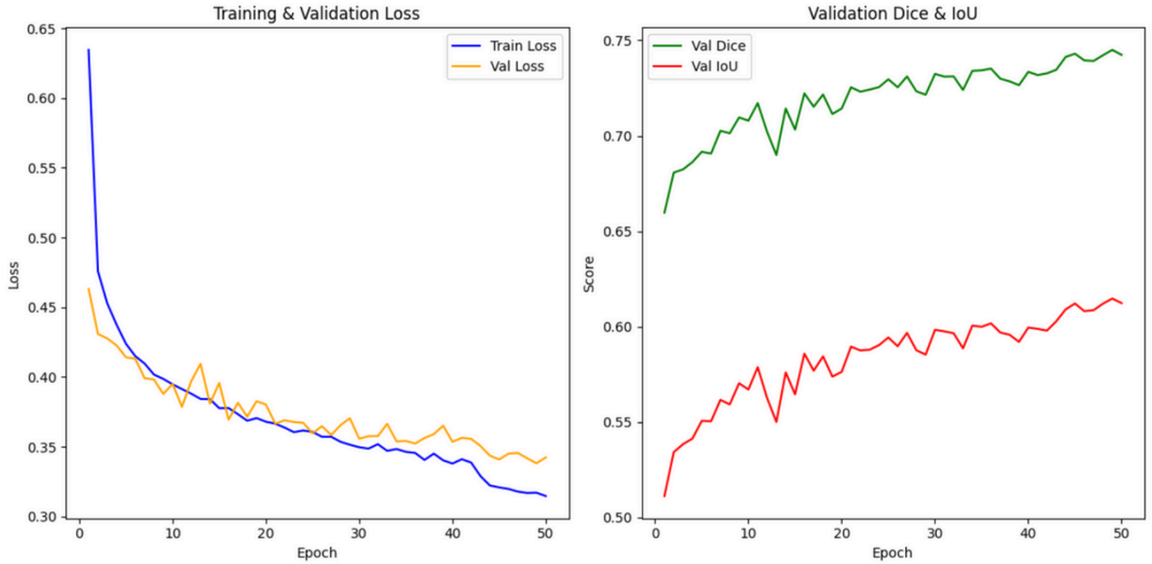
**Dice coefficient**

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

**Intersection over Union (IoU)**

# 3. Experimental Results

## Quantitative Results




Training & Validation Loss, Validation Dice & IoU over epochs

| Metrics | Best Dice | Best IoU | Final Val Loss | Final Train Loss | Epoch Reached |
|---|---|---|---|---|---|
| Value | 0.7404 | 0.6095 | 0.3437 | 0.3097 | 50 |

Best results after 50 epochs training:

- The graphs suggest a stable training and evaluation process.
- There are no clear signs of overfitting or underfitting.
- The model shows continuous improvement in validation accuracy, particularly in segmentation metrics like Dice and IoU.

# 3. Experimental Results

## Quantitative Results

### Result on test set with Post-process

| Kernel Size | Dice | IoU | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| 1 | 0.7488 | 0.6204 | 0.7489 | 0.7884 | 0.7682 |
| 2 | 0.7439 | 0.6136 | 0.7406 | 0.7894 | 0.7642 |
| 3 | 0.7311 | 0.5977 | 0.6925 | 0.8187 | 0.7503 |
| 4 | 0.7082 | 0.5687 | 0.6447 | 0.8369 | 0.7283 |
| 5 | 0.6855 | 0.5413 | 0.6023 | 0.8514 | 0.7055 |

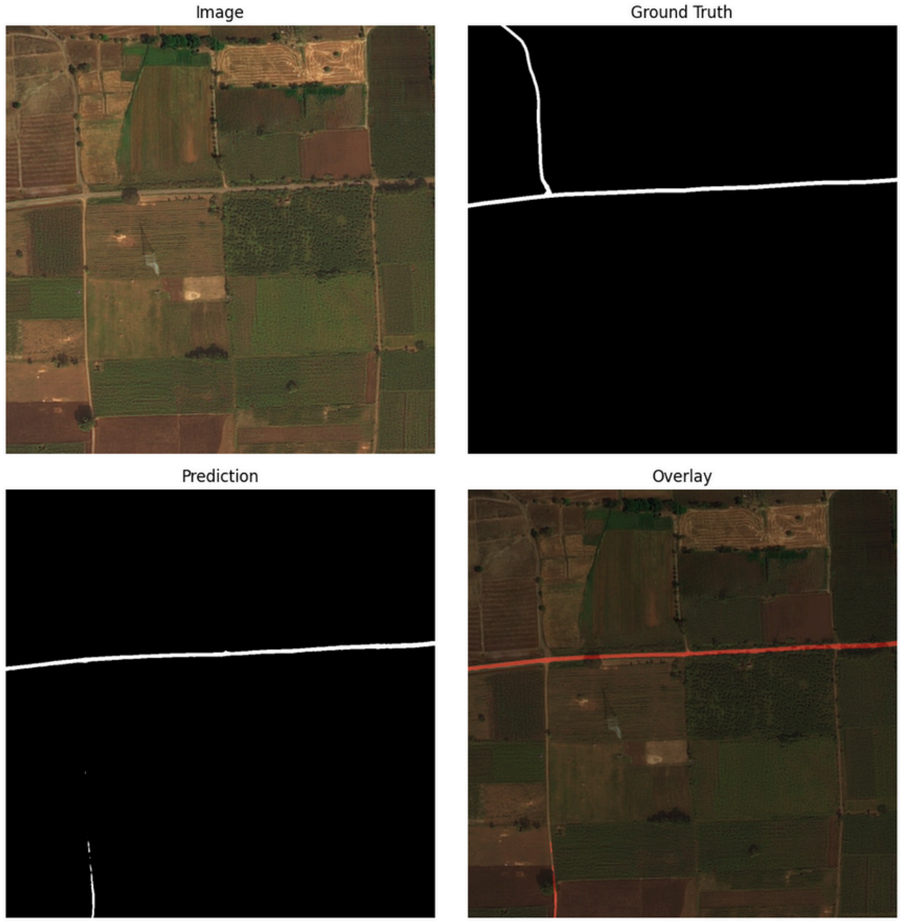
Fixed threshold = 0.5, Kernel Size sweep

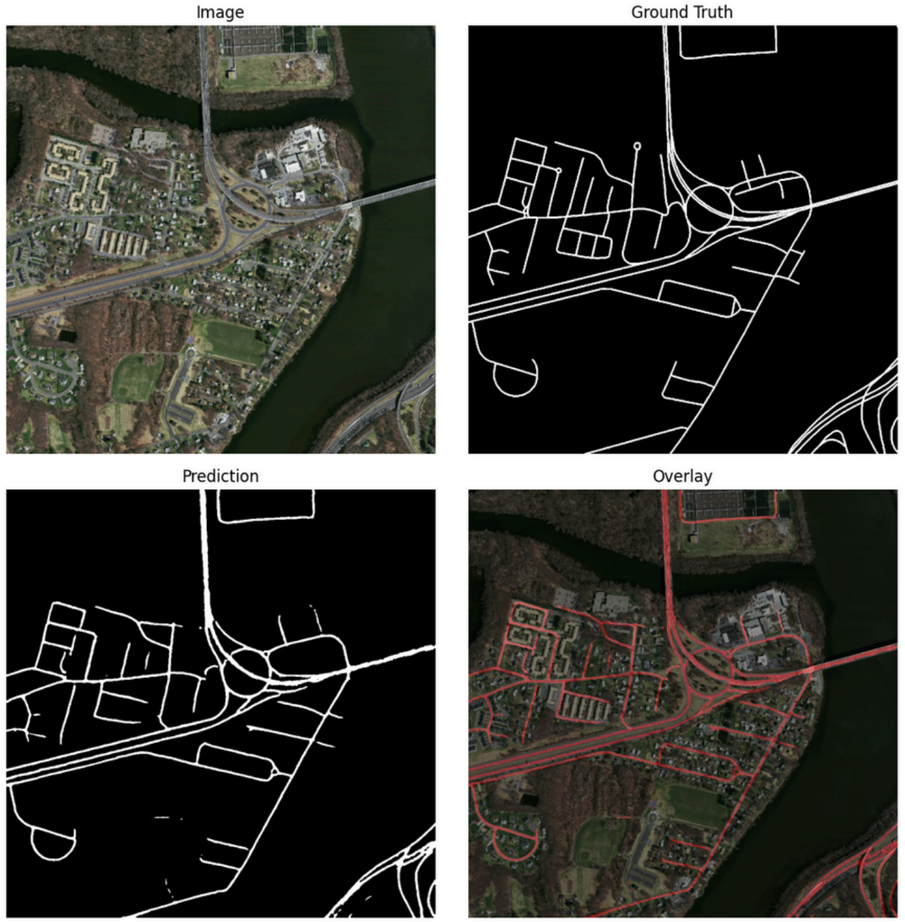
| Threshold | Dice | IoU | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| 0.2 | 0.7497 | 0.6213 | 0.7491 | 0.7884 | 0.7682 |
| 0.3 | 0.7500 | 0.6218 | 0.7666 | 0.7727 | 0.7696 |
| 0.4 | 0.7496 | 0.6214 | 0.7788 | 0.7607 | 0.7697 |
| 0.5 | 0.7488 | 0.6204 | 0.7895 | 0.7496 | 0.7690 |
| 0.6 | 0.7475 | 0.6187 | 0.8004 | 0.7374 | 0.7676 |
| 0.7 | 0.7452 | 0.6158 | 0.8135 | 0.7216 | 0.7648 |
| 0.8 | 0.7405 | 0.6099 | 0.8314 | 0.6970 | 0.7583 |

Fixed Kernel Size = 1, Threshold sweep

# 3. Experimental Results

## Qualitative Results



# 3. Experimental Results

## Additional Discussion

### Result on test set without Post-process

| Threshold | Dice | IoU | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| 0.2 | 0.7497 | 0.6214 | 0.7489 | 0.7884 | 0.7682 |
| 0.3 | 0.7500 | 0.6218 | 0.7665 | 0.7727 | 0.7696 |
| 0.4 | 0.7496 | 0.6214 | 0.7787 | 0.7608 | 0.7696 |
| 0.5 | 0.7488 | 0.6204 | 0.7894 | 0.7497 | 0.7690 |

⟶ Without post-process, threshold sweep

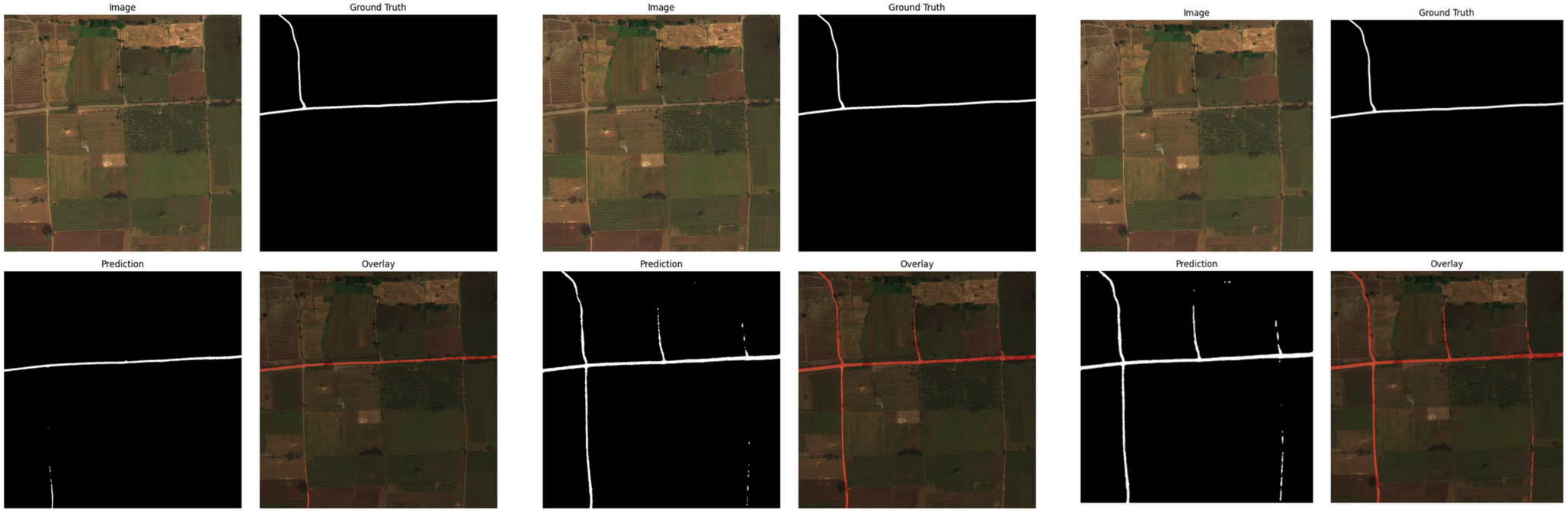
| Threshold | Dice | IoU | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| 0.2 | 0.7497 | 0.6213 | 0.7491 | 0.7884 | 0.7682 |
| 0.3 | 0.7500 | 0.6218 | 0.7666 | 0.7727 | 0.7696 |
| 0.4 | 0.7496 | 0.6214 | 0.7788 | 0.7607 | 0.7697 |
| 0.5 | 0.7488 | 0.6204 | 0.7895 | 0.7496 | 0.7690 |
| 0.6 | 0.7475 | 0.6187 | 0.8004 | 0.7374 | 0.7676 |
| 0.7 | 0.7452 | 0.6158 | 0.8135 | 0.7216 | 0.7648 |
| 0.8 | 0.7405 | 0.6099 | 0.8314 | 0.6970 | 0.7583 |

⟶ Kernel Size = 1, threshold sweep

# 3. Experimental Results

## Additional Discussion

### Threshold Selection




Threshold = 0.3 | Threshold = 0.001 | Threshold = 0.0005

# IV. Conclusion

# 4. Conclusion

## Limitation

- Road pixels occluded by shadows or trees can sometimes be missed.
- Complex intersections may be under-segmented.
- Predictions in visually ambiguous regions occasionally include false positives.

## Future work

- Finish SegFormer.
- Experiment other post-process methods like Conditional Random Fields (CRFs).

HUST

# THANK YOU !