Extraction of Roads from Satellite Images

A thesis submitted

in partial fulfillment for the award of the degree of

Bachelor of Technology

in

Electronics and Communication Engineering (Avionics)

by

Rahul Velamala

Department of Avionics
Indian Institute of Space Science and Technology
Thiruvananthapuram, India

31st May - 19th July 2023

Certificate

This is to certify that the thesis titled ***Extraction of Roads from Satellite Images*** submitted by **Rahul Velamala**, to the Indian Institute of Space Science and Technology, Thiruvananthapuram, in partial fulfillment for the award of the degree of **Bachelor of Technology** in **Electronics and Communication Engineering (Avionics)** is a bonafide record of the original work carried out by him/her under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Shri.Anil Yadav
Sci/Eng-SE
APP.SOFTWARE AND COMPUTE INFRA DIVISION
National Remote Sensing Center, Hyderabad
Indian Space Research Organisation

Place: Hyderabad
Date: 18th July 2024

Declaration

I declare that this thesis titled ***Extraction of Roads from Satellite Images*** submitted in partial fulfillment for the award of the degree of **Bachelor of Technology** in **Electronics and Communication Engineering (Avionics)** is a record of the original work carried out by me under the supervision of **Shri.Anil Yadav**, and has not formed the basis for the award of any degree, diploma, associateship, fellowship, or other titles in this or any other Institution or University of higher learning. In keeping with the ethical practice in reporting scientific information, due acknowledgments have been made wherever the findings of others have been cited.

Place: Hyderabad
Date : 19th July 2024

Rahul Velamala
(IIST ID : SC21B127)

Acknowledgements

I would like to express my gratitude to my guide Shri.Anil Yadav (Sci/Eng-E) at National Remote Sensing Center, Hyderabad,Indian Space Research Organisation,for giving me this opportunity to work on this topic as my internship. I am very grateful for the tremendous support and guidance that they has provided me at every stage of my work, making my internship very interesting. I am also thankful to both of them for patiently clearing my doubts.

Rahul Velamala
College ID : SC21B127

Abstract

This report details the development and optimization of a U-Net model for accurate road extraction from satellite images. The model employs advanced deep learning techniques, including dropout, batch normalization, and a custom Jaccard-based loss function, to enhance segmentation performance. Key evaluation metrics—precision, recall, F1 score, and Intersection over Union (IoU)—demonstrate the model's effectiveness in delineating road networks. The integration of these techniques resulted in robust and accurate road segmentation, showcasing the potential of deep learning for remote sensing applications. Future work will explore the incorporation of attention mechanisms into the U-Net architecture and the use of sliding window techniques for patch extraction and stitching to handle larger and more varied image datasets. These advancements aim to further refine the road extraction process, improving scalability and efficiency for broader applications in urban planning and infrastructure monitoring.

Contents

List of Figures

Abbreviations

IoU	Intersection over Union
CNN	Convolutional Neural Network
RMSprop	Root Mean Square Propagation
ReLU	Rectified Linear Unit
TP	True Positives
FP	False Positives
FN	False Negatives
TN	True Negatives
mIoU	Mean Intersection over Union
LR	Learning Rate
RGB	Red, Green, Blue (color channels)
GIS	Geographic Information System
DL	Deep Learning
DNN	Deep Neural Network
BN	Batch Normalization

Chapter 1

Introduction

1.1 Problem Statement :

1.1.1 Extraction of Roads from Satellite Images

Extracting roads from satellite images is a crucial task in applications like navigation, urban planning, and disaster management. Traditional methods are often time-consuming and inaccurate due to the complexity and variability of satellite images. Recent advances in deep learning, specifically the U-Net model, offer a promising solution, but the classical U-Net architecture has some limitations that need to be addressed.

In this project, the aim is to enhance the performance of road extraction from satellite images by utilizing a modified U-Net model.

1.2 Objectives

The primary objective of this project is to enhance the performance of road extraction from satellite images using a modified U-Net model. This involves integrating batch normalization and dropout layers into the classical U-Net architecture to improve training stability and reduce overfitting. The DeepGlobe dataset will be utilized, with images and masks carefully preprocessed for optimal training. The modified U-Net model will be trained using TensorFlow and Keras, with a focus on selecting appropriate hyperparameters and optimizing the training process. The model's performance will be evaluated using standard segmentation metrics like Intersection over Union (IOU) . Visualizing the results will allow for qualitative assessment, comparing predicted road masks with the ground truth to identify areas for improvement. The aim is to provide a more accurate and reliable method for automated road extraction from satellite images.

1.3 Background

Extracting roads from satellite images is important for applications like navigation, urban planning, and disaster management. Accurate road mapping helps with transportation planning, emergency routing, and infrastructure development, which are crucial for modern cities and disaster preparedness.

Traditional methods for road extraction involve manual digitization, which is slow and labor-intensive. Automated image processing techniques also struggle with the complexity and variability of satellite images. They often face challenges like different resolutions, occlusions by trees or buildings, and shadows, leading to inconsistent and inaccurate results.

Recent advances in deep learning, especially convolutional neural networks (CNNs), have greatly improved image processing tasks. The U-Net model, originally designed for biomedical image segmentation, has shown great success in various segmentation tasks, including road extraction. Its encoder-decoder structure is good for effective feature extraction and precise localization.

However, the classical U-Net architecture has some limitations, such as sensitivity to image quality variations and potential overfitting with limited datasets. Enhancements like batch normalization and dropout layers can help improve training stability and generalization.

This project aims to improve road extraction from satellite images by using a modified U-Net model. By enhancing the U-Net architecture and optimizing the training process, this project seeks to create a more reliable and effective solution for automated road extraction, contributing to the field of remote sensing and image segmentation.

Chapter 2

Related Work

Several approaches have been explored for the task of road extraction from satellite images, ranging from traditional image processing techniques to modern deep learning methods.

2.1 Traditional Methods

Traditional methods for road extraction primarily rely on manual digitization and classical image processing techniques. Manual digitization is labor-intensive and time-consuming, often leading to inconsistencies due to human error. Classical methods involve edge detection, morphological operations, and thresholding. These methods struggle with the variability and complexity of satellite imagery, such as varying resolutions, occlusions, and shadows, leading to less accurate results. For instance, methods based on edge detection often fail in complex urban environments where roads are obscured by buildings or trees.

2.2 Machine Learning Approaches

Machine learning approaches introduced supervised and unsupervised techniques to automate road extraction. Early methods used handcrafted features and traditional classifiers like Support Vector Machines (SVM) and Random Forests. While these methods provided some improvement over purely manual approaches, they still struggled with the high variability in satellite images and often required extensive feature engineering.

2.3 Deep Learning Techniques

The advent of deep learning, particularly Convolutional Neural Networks (CNNs), has significantly advanced the field of image segmentation, including road extraction. CNNs automatically learn hierarchical feature representations from raw images, reducing the need for manual feature engineering.

2.4 U-Net with ResNet-50 Backbone

One modification involves integrating the U-Net architecture with a ResNet-50 backbone. ResNet-50, a deep residual network, provides robust feature extraction capabilities due to its deep architecture and residual connections that help mitigate the vanishing gradient problem. This combination leverages the strong feature extraction abilities of ResNet-50 with the precise localization capabilities of U-Net, resulting in improved performance on road extraction tasks. For example, a study by Zhang et al. (2018) introduced a U-Net model with a ResNet-50 backbone, which showed improved performance in road extraction due to better feature representation and reduced overfitting.

2.5 U-Net with Attention Mechanisms

Another significant enhancement is the incorporation of attention mechanisms into the U-Net architecture. Attention mechanisms allow the model to focus on the most relevant parts of the image, enhancing the segmentation accuracy by emphasizing important features while suppressing irrelevant background information. Attention U-Net models have demonstrated superior performance in various segmentation tasks, including road extraction, by providing better discrimination between road and non-road areas. For instance, Oktay et al. (2018) introduced Attention U-Net, which includes attention gates in the skip connections of the U-Net, allowing the model to learn where to focus more precisely.

2.6 SPIN (Spatial Information Network)

SPIN is an advanced deep learning framework designed to incorporate spatial context information into the road extraction process. By integrating spatial relationships and contextual information, SPIN models can better understand and segment complex road networks. The SPIN framework enhances traditional CNN architectures by including spatial dependencies, which improves the accuracy and reliability of road extraction from satellite images. Studies have shown that SPIN can effectively handle occlusions and varying image qualities by leveraging spatial context, making it a valuable addition to road extraction techniques.

2.7 Other Deep Learning Models

Other deep learning models, such as Fully Convolutional Networks (FCNs), SegNet, and DeepLab, have also been explored for road extraction. These models provide different architectural advantages, such as multi-scale feature extraction and improved boundary delineation. For example, DeepLab uses atrous convolutions to capture multi-scale context, which is beneficial for detecting roads at different scales.

2.8 Benchmark Datasets

Several benchmark datasets have been developed to facilitate the training and evaluation of road extraction models. The DeepGlobe Road Extraction Challenge dataset and the Wuhan University (WHU) Road Dataset are prominent examples. These datasets provide high-resolution satellite images with annotated road networks, enabling the development and benchmarking of various road extraction techniques.

Chapter 3

Data and Preprocessing

3.1 Datasets

Two primary datasets were used for the project: The Wuhan University (WHU) Road Dataset and the DeepGlobe Road Extraction Challenge dataset.

3.1.1 Wuhan University (WHU) Road Dataset

The satellite imagery dataset of roads in Southern China were manually annotated and termed as the Wuhan University (WHU) Road Dataset. The images in the dataset come from a variety of satellites, including Gaofen-II, WorldView-II, and ZY-III, with a spatial resolution ranging from 0.8 to 2 meters per pixel. This dataset covers Liuzhou City in Guangxi Province.

The road region annotations in the WHU Road Dataset were labeled by experts in remote sensing image interpretation using ArcGIS software to ensure high quality. In urban areas, various road types, including railways, motorways, primary roads, secondary roads, tertiary roads, and trunk roads, were annotated. To simplify the annotation process, trivial roads such as footways in residential areas were ignored. In rural areas, only roads wider than three meters and longer than five meters were labeled. The total area of valid data in the dataset exceeds 1145 km^2.

The WHU Road Dataset was cropped into image tiles of 512x512 pixels, resulting in a total of 6828 samples. For the project,i have used the data which includes remote sensing images with a resolution of 0.5 to 0.8 meters per pixel.

3.1.2 DeepGlobe Road Extraction Challenge Dataset

In addition to the WHU dataset, The DeepGlobe Road Extraction Challenge dataset, which provides a large and diverse set of satellite images for training and evaluation was also used for the project. The DeepGlobe dataset includes high-resolution satellite images at a resolution of 50 cm per pixel. The dataset consists of 6,803 training images and 2,419 validation images, each with a size of 1024x1024 pixels. The images are accompanied by binary masks that delineate the road regions, allowing for supervised training of road extraction models.

The DeepGlobe dataset was specifically designed to benchmark and advance road extraction techniques. It covers a variety of geographical regions, including urban, suburban, and rural areas, providing a robust and comprehensive dataset for evaluating the performance of our model across different environments and conditions. The images in the DeepGlobe dataset were collected from multiple regions, covering over 2,220 km², ensuring diverse and representative samples. The dataset's diversity in terms of geographical coverage and image quality makes it an ideal complement to the WHU Road Dataset for training and validating our road extraction model.

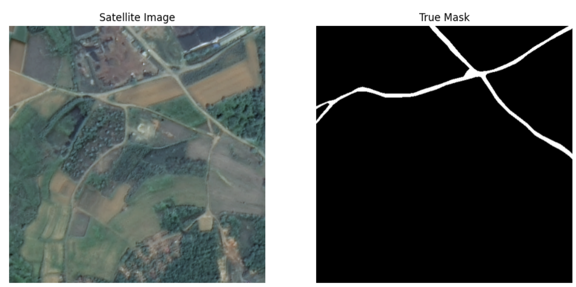

Figure 3.1: Example of training image along with its masks from WHU Road Datset

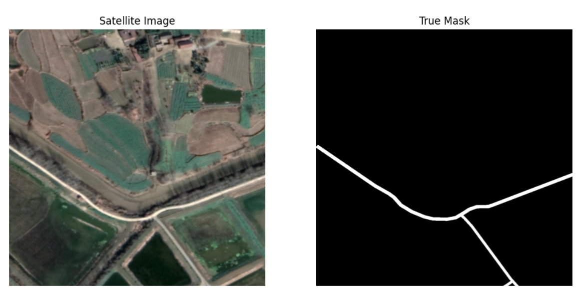

Figure 3.2: Example of training image along with its masks from DeepGlobe Road Extraction Datset

3.2 Data preprocessing

To efficiently train our model for road extraction from satellite images, A systematic data preprocessing pipeline was employed, including the use of a custom data generator and data augmentation techniques. These steps ensure that the model is exposed to a wide variety of training examples, enhancing its ability to generalize to new, unseen data.

3.2.1 Data Generator

The data generator is responsible for dynamically loading and preprocessing images and their corresponding masks in batches during model training. This approach is essential for handling large datasets that cannot fit into memory all at once. The generator ensures that each image is paired with its corresponding mask, resized to the desired input dimensions, and normalized to have pixel values in a consistent range.

3.2.2 Data Augmentation

To further improve the model's robustness and performance, Data augmentation techniques were incorporated into the preprocessing pipeline. Data augmentation artificially expands the training dataset by applying random transformations to the images and masks. These transformations include:

1. **Flipping**: Random horizontal and vertical flips to introduce orientation variability.

2. **Brightness Adjustment**: Random changes in brightness to simulate different lighting conditions.

3. **Contrast Adjustment**: Random modifications to the contrast to handle images with varying levels of contrast.

4. **Saturation Adjustment**: Random adjustments to saturation levels to account for different color intensities.

5. **Hue Adjustment**: Random changes in hue to simulate color variations.

These augmentations help the model learn to recognize roads under diverse conditions and appearances, which is crucial given the variability in satellite imagery due to factors like lighting, shadows, and occlusions.

3.2.3 Dataset Preparation and Splitting

3.2.3.1 WHU Road Dataset

The WHU Road Dataset, containing high-resolution images from various satellites, was divided into training, validation, and test sets. The dataset's images were split into patches of size 512x512 pixels, which were then used to train the U-Net model.

- **Training Set**: Consisted of 4,800 images
- **Validation Set**: Consisted of 1,000 images
- **Test Set**: Consisted of 1,028 images

3.2.3.2 DeepGlobe Road Extraction Challenge Dataset

The DeepGlobe dataset was also utilized for additional training and validation. The dataset includes high-resolution satellite images at 50 cm per pixel resolution, divided as follows:

- **Training Set**: Consisted of 6,803 images
- **Validation Set**: Consisted of 1,219 images
- **Test Set**: Consisted of 1,200 images

Chapter 4

Methods

In this chapter, the methods used in this project and their methodology will be illustrated in detail, including the network architecture of Unet, the processing of datasets, improvements performed to the network.

4.1 U-net Architecture

4.1.1 Overview

The U-Net architecture is a widely used convolutional neural network (CNN) model, originally designed for biomedical image segmentation. It has been successfully applied to various image segmentation tasks, including road extraction from satellite images. The architecture of U-Net is characterized by its encoder-decoder structure, which allows it to capture both the global context and fine details in the image.

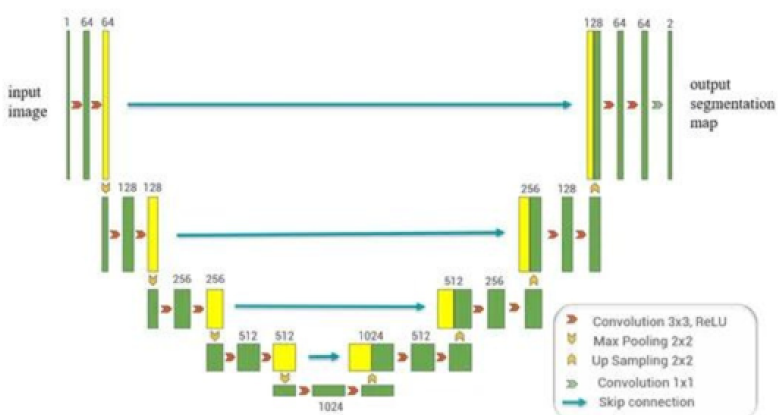

Figure 4.1: Unet Architecture

Encoder: The encoder path of the U-Net consists of several convolutional layers followed by max-pooling layers. Each encoder block contains two convolutional layers, each followed by batch normalization and ReLU activation. The max-pooling layers progressively reduce the spatial dimensions of the feature maps while increasing the number of filters, effectively capturing hierarchical features. Dropout layers are optionally added to prevent overfitting and improve generalization.

Bridge: At the bottleneck of the network, the bridge block connects the encoder and decoder paths. This block has the highest number of filters and performs the most complex feature extraction. It also includes dropout layers to handle overfitting.

Decoder: The decoder path mirrors the encoder but uses upsampling instead of downsampling. Each decoder block consists of a transposed convolution layer that upsamples the feature maps, followed by concatenation with the corresponding encoder block's output via skip connections. This helps retain the spatial information lost during downsampling. The concatenated feature maps are then processed through convolutional layers with batch normalization and ReLU activation. Dropout layers are also used in the decoder to enhance model robustness.

Output Layer: The final output layer of the U-Net model is a convolutional layer with a single filter and a sigmoid activation function. This layer produces a binary mask that delineates the road regions in the input image, with pixel values ranging between 0 and 1.

Skip Connections: A key feature of the U-Net architecture is the use of skip connections that link the corresponding layers in the encoder and decoder paths. These connections help preserve spatial information and allow the model to combine low-level and high-level features, resulting in more accurate segmentation.

In convolutional layers, we use learnable filters that require manual configuration. Zero padding is used throughout multiple convolution operations to maintain image boundary information and control output size conveniently. Convolutional layers aim for weight sparsity and weight sharing, unlike fully connected layers. The extracted features need additional processing, leading to computational challenges and an increased risk of overfitting due to the growing number of parameters in the network.

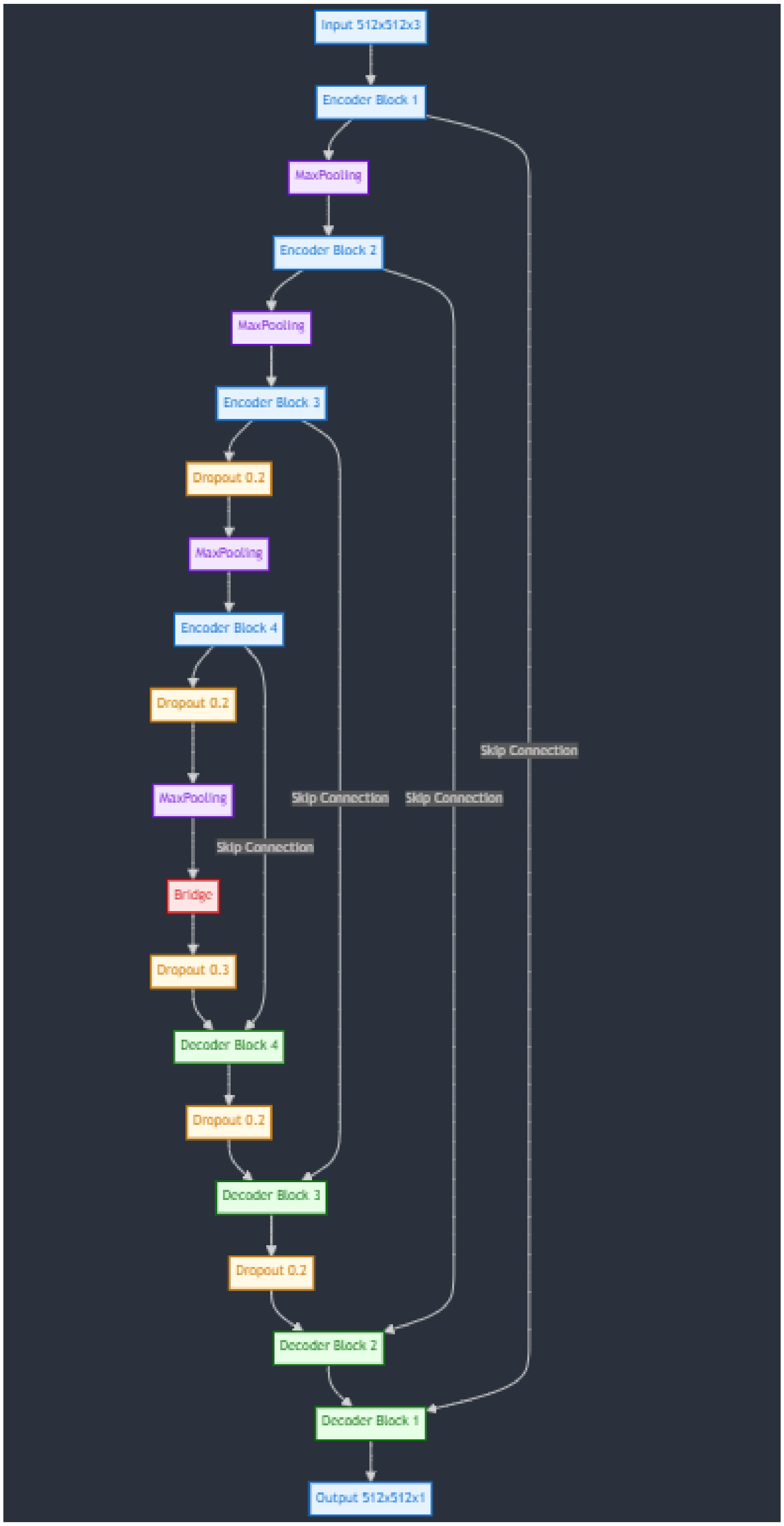

Figure 4.2: Architecture diagram of Unet in this Project

4.2 Metrics And Techniques Used

- **Loss Function**

  The primary loss function used was the Jaccard coefficient loss (also known as Intersection over Union loss), which is well-suited for segmentation tasks as it directly measures the overlap between the predicted and true masks. The Jaccard coefficient loss helps in balancing the optimization process by focusing on the regions of interest and improving the model's ability to delineate roads accurately.

- **Regularization**

  To mitigate overfitting, dropout layers were incorporated into the model. Dropout helps in regularizing the network by randomly setting a fraction of input units to zero at each update during training, thereby preventing the model from relying too heavily on specific neurons. This technique encourages the model to generalize better to unseen data.

- **Normalization**

  Batch normalization was integrated into the U-Net architecture. Batch normalization normalizes the input of each layer to have a mean of zero and a variance of one, which accelerates the training process and improves overall performance. It also acts as a form of regularization, reducing the need for other regularization methods such as dropout.

- **Optimizer**

  The RMSprop optimizer was selected for training the model. RMSprop is an adaptive learning rate optimization algorithm that adjusts the learning rate for each parameter, making it suitable for tasks with sparse gradients, such as image segmentation. The initial learning rate was set to 0.001, with a weight decay of 1e-4 to prevent overfitting.

- **Custom Metrics**

  To evaluate the performance of the model, several custom metrics were implemented:

  1. **Precision**: Measures the ratio of true positives to the sum of true positives and false positives, indicating the accuracy of the positive predictions.

2. **Recall**: Measures the ratio of true positives to the sum of true positives and false negatives, indicating the model's ability to capture all relevant instances.

3. **F1 Score**: The harmonic mean of precision and recall, providing a balance between the two.

4. **Intersection over Union (IOU)**: Measures the overlap between the predicted and true masks, providing a direct assessment of segmentation performance.

5. **Jaccard Coefficient**: Similar to IOU, this measures the similarity between the predicted and true masks.

- **Callbacks**

  Several callbacks were implemented to enhance the training process and monitor performance: ModelCheckpoint: Saves the best model based on the validation loss, ensuring the optimal model is retained. EarlyStopping: Stops training when the validation loss stops improving, preventing overfitting and reducing unnecessary computation. ReduceLROnPlateau: Reduces the learning rate when the validation loss plateaus, allowing for finer adjustments during later stages of training. CSVLogger: Logs training progress to a CSV file, enabling easy tracking and analysis of the training process.

- **Validation Visualization** A custom callback was developed to visualize validation predictions. This callback saves images of the input, true masks, and predicted masks after each epoch, allowing for qualitative assessment of the model's performance. This visual feedback helps in understanding how well the model is segmenting the roads and in identifying areas for potential improvement.

4.2.1 Mathematics Behind Metrics

4.2.1.1 Jaccard Coefficient and Loss

The *Jaccard Coefficient* (also known as the Intersection over Union, IoU) is a metric used to evaluate the similarity between the predicted segmentation mask and the ground truth mask. It is calculated as:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} = \frac{\sum(A \cdot B)}{\sum A + \sum B - \sum(A \cdot B)}$$

where A is the predicted binary mask, and B is the ground truth binary mask. The Jaccard coefficient ranges from 0 to 1, where 1 indicates perfect overlap and 0 indicates no overlap.

The *Jaccard Loss* is derived from the Jaccard coefficient and is defined as:

$$\text{Jaccard Loss} = 1 - \text{IoU}$$

or, in its negative form as used in the custom loss function:

$$\text{Jaccard Loss} = -\text{IoU}$$

This loss function directly optimizes the IoU, making it particularly suitable for segmentation tasks.

4.2.1.2 Precision

Precision is the ratio of true positive predictions to the total predicted positives. It is given by:

$$\text{Precision} = \frac{TP}{TP + FP}$$

where TP is the number of true positives and FP is the number of false positives. Precision measures the accuracy of the positive predictions.

4.2.1.3 Recall

Recall (or Sensitivity) is the ratio of true positive predictions to the total actual positives. It is given by:

$$\text{Recall} = \frac{TP}{TP + FN}$$

where TP is the number of true positives and FN is the number of false negatives. Recall measures the model's ability to identify all relevant instances.

4.2.1.4 F1 Score

The *F1 Score* is the harmonic mean of precision and recall. It provides a balance between precision and recall and is given by:

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

The F1 score ranges from 0 to 1, where 1 indicates perfect precision and recall.

4.2.1.5 Intersection over Union (IoU)

The *IoU* (as previously defined in the Jaccard coefficient) is given by:

$$\text{IoU} = \frac{\sum(A \cdot B)}{\sum A + \sum B - \sum(A \cdot B)}$$

where A and B are the predicted and ground truth binary masks, respectively. IoU is a common metric for evaluating segmentation performance.

4.2.1.6 RMSprop Optimizer

The *RMSprop* optimizer adjusts the learning rate for each parameter, making it effective for tasks with sparse gradients. The update rule for RMSprop is given by:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t$$

where:

- θ_t is the parameter at step t
- η is the learning rate
- $E[g^2]_t$ is the running average of the squared gradients
- ϵ is a small constant to prevent division by zero
- g_t is the gradient at step t

RMSprop helps in maintaining a balance between fast convergence and smooth training.

4.2.1.7 Batch Normalization

Batch Normalization normalizes the input of each layer to have a mean of zero and variance of one. The formula for batch normalization is:

$$\hat{x}^{(k)} = \frac{x^{(k)} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

where:

- $x^{(k)}$ is the input
- μ_B is the mini-batch mean
- σ_B^2 is the mini-batch variance
- ϵ is a small constant to prevent division by zero

After normalization, the output is scaled and shifted:

$$y^{(k)} = \gamma\hat{x}^{(k)} + \beta$$

where γ and β are learned parameters.

4.2.1.8 Dropout

Dropout is a regularization technique where a fraction of input units is set to zero during training to prevent overfitting. The formula for dropout is:

$$y = \frac{1}{1-p}x \cdot \text{mask}$$

where:

- p is the dropout rate
- x is the input
- mask is a binary mask with elements drawn from a Bernoulli distribution with probability $1 - p$

By combining these techniques and metrics, the model is optimized to accurately extract roads from satellite images, leveraging the strengths of each component to enhance overall performance.

4.3 Methodology Overview

The methodology of this project involved exploring various techniques to enhance the accuracy and efficiency of road extraction from high-resolution satellite images using the U-Net architecture. We investigated three primary approaches for handling high-resolution images, experimented with different filter sizes in the U-Net model, addressed class imbalance using Jaccard coefficient loss, and optimized batch sizes. The following subsections detail each aspect of the methodology.

4.3.1 Handling High-Resolution Images

High-resolution satellite images pose a unique challenge for segmentation tasks due to their large size and detailed content.Three different methods were explored to handle these high-resolution images:
Among these three methods, the first method proved to offer the optimal balance between computational resources and results.

1. **Dividing Images into Patches (512x512)**:
    - The high-resolution images were divided into smaller patches of size 512x512 pixels.
    - The U-Net model was trained on these patches, allowing it to focus on local features and improve segmentation accuracy.
    - This method is effective but requires careful handling of boundary regions to ensure seamless segmentation.

2. **U-Net for Large Images (1024x1024)**:
    - We designed a U-Net model that takes entire 1024x1024 images as input.
    - This approach leverages the full context of the image, potentially improving segmentation performance for larger structures.
    - However, this method demands significant computational resources and memory.

3. **Compressing Images to 512x512**:

    - High-resolution images were downsampled to 512x512 pixels and then used to train the U-Net model.
    - This method simplifies the processing pipeline but often results in loss of critical details and reduced segmentation accuracy.
    - This approach is generally not recommended due to its impact on the quality of segmentation.

4.3.2 Filter Size Experiments

In the U-Net architecture, the size of the convolutional filters plays a crucial role in feature extraction. We experimented with different filter sizes to identify the most effective configuration:

Filter Sizes Tested: Filter sizes of 3x3, 5x5, and 7x7 in the convolutional layers were tested.

Results: The experiments demonstrated that a filter size of 3x3 provided the best balance between capturing fine details and maintaining computational efficiency.

Conclusion: We adopted the 3x3 filter size for all subsequent experiments due to its superior performance.

4.3.3 Handling Class Imbalance

Class imbalance is a common issue in segmentation tasks, where the target class (roads) occupies a smaller portion of the image compared to the background. To address this, we employed the following strategies:

Jaccard Coefficient Loss:

The Jaccard coefficient (also known as Intersection over Union) as the loss function was used . This loss function focuses on the overlap between the predicted and ground truth masks, improving performance for imbalanced classes.

Weighted Loss Function:

The experimentation with weighted loss functions to give more importance to the road pixels during training. This approach helped to further mitigate the impact of class imbalance.

4.3.4 Batch Size Optimization

Batch size is a critical hyperparameter that affects both the training stability and the performance of the model. experiments were conducted to determine the optimal batch size:

Batch Sizes Tested: Batch sizes of 2, 4, 8, and 16 were tested.

Results: A batch size of 4 was found to offer the best trade-off between model accuracy and training time.

Conclusion: Batch size of 4 was adopted for all subsequent experiments, ensuring a stable and efficient training process.

Chapter 5

Experimental Results

5.1 Experimental Results

The experimental results section presents a comprehensive analysis of the performance of different methods, filter sizes, and optimization strategies implemented in the U-Net architecture for road extraction from high-resolution satellite images. This section is divided into several subsections to provide detailed insights into the outcomes of various experiments.

5.2 Experimental Results

5.2.1 Performance on Different Segmentation Approaches

Segmentation Approach	IoU	F1 Score	Precision	Recall
Patch-based (512x512)	0.6018	0.7250	0.7590	0.7234
Full image (1024x1024)	0.3983	0.5332	0.6916	0.4939
Compressed (256x256)	0.4319	0.6037	0.5982	0.5126

Table 5.1: Performance comparison of different segmentation approaches.

Our experiments showed that the patch-based segmentation approach using 512x512 patches yielded the optimal performance managing the computational resources and output in terms of IoU and F1 Score.

5.2.2 Effect of Different Filter Sizes

Our analysis indicated that a filter size of 3x3 was optimal for our U-Net model, providing the best balance between precision and recall.

Filter Size	IoU	F1 Score	Precision	Recall
3x3	0.6018	0.7250	0.7590	0.7234
5x5	0.5642	0.7290	0.7012	0.6584

Table 5.2: Performance comparison with different filter sizes in U-Net architecture.

5.2.3 Addressing Class Imbalance

To handle class imbalance in our dataset, we employed the Jaccard coefficient loss function. This significantly improved our model's performance:

- IoU increased from 0.5598 to 0.0.6018.
- F1 Score improved from 0.6921 to 0.7250.

5.2.4 Comparison with Benchmark Models

Our U-Net model was compared against several benchmark models to assess its effectiveness:

Model	IoU	F1 Score
SPIN Road Mapper	0.6702	0.84
LinkNet	0.6275	0.80
Batra et al	0.6721	0.8397
U-Net (ours)	0.6019	0.73

Table 5.3: Performance comparison with benchmark models.

5.2.5 Plots

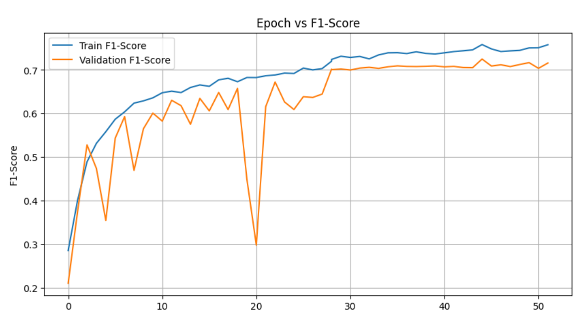

Figure 5.1: Plot of f1-score over epochs

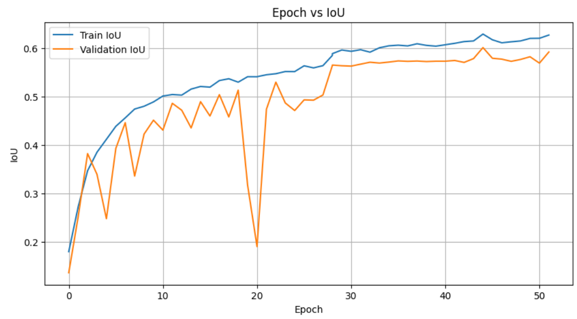

Figure 5.2: Plot of IOU over epochs

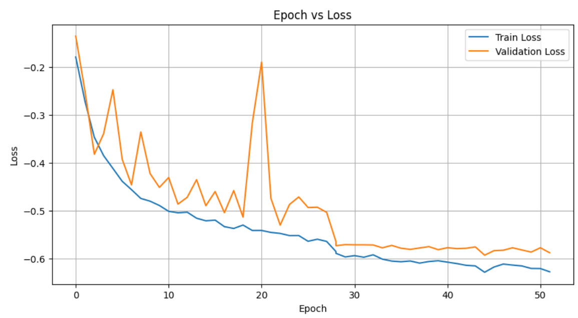

Figure 5.3: Plot of loss over epochs

5.2.6 Predictions

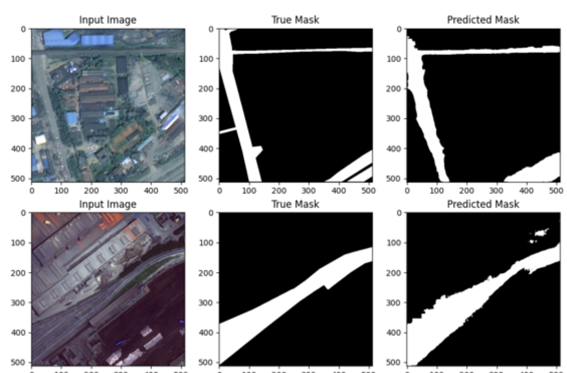

Figure 5.4: Predictions made by the model

Chapter 6

Conclusion and Future Work

6.1 Conclusion

In this project, we successfully developed and optimized a U-Net model for the extraction of roads from satellite images. The model was evaluated using various metrics, including precision, recall, F1 score, and Intersection over Union (IoU), demonstrating its effectiveness in accurately segmenting road networks from satellite imagery. The incorporation of advanced techniques such as dropout, batch normalization, and a customized loss function contributed significantly to the robustness and accuracy of the model. This project underscores the potential of deep learning models in the fields of remote sensing, urban planning, and infrastructure monitoring, providing a valuable tool for automating the analysis of satellite data.

6.2 Future Works

Future work will focus on further enhancing the model's performance and scalability. One promising direction is the integration of attention mechanisms into the U-Net architecture, allowing the model to better focus on relevant features in the images, thereby improving segmentation accuracy. Additionally, employing techniques such as sliding window for patch extraction and subsequent stitching can improve the model's ability to handle larger images and varied resolutions, making it more versatile for different datasets. Furthermore, exploring other deep learning architectures and hybrid models could yield improvements in processing speed and accuracy. These advancements will refine the road extraction process, making it more efficient, scalable, and applicable to a wider range of satellite imagery datasets, ultimately contributing to better urban planning and management.

Bibliography

[1] V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," in *European Conference on Computer Vision*. Springer, 2010, pp. 210–223.

[2] G. Mattyus, W. Luo, and R. Urtasun, "Deeproadmapper: Extracting road topology from aerial images," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3438–3446.

[3] Z. Xu, Y. Sun, and M. Liu, "icurb: Imitation learning-based detection of road curbs using aerial images for autonomous driving," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1097–1104, 2021.

[4] Z. Xu, Y. Sun, and M. Liu, "Topo-boundary: A benchmark dataset on topological road-boundary detection using aerial images for autonomous driving," *arXiv preprint arXiv:2103.17119*, 2021.

[5] M. Barzohar and D. B. Cooper, "Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, pp. 707–721, 1996.

[6] G. B. Vitor, A. C. Victorino, and J. V. Ferreira, "A probabilistic distribution approach for the classification of urban roads in complex environments," in *IEEE Workshop on International Conference on Robotics and Automation*, 2014.

[7] D. Chai, W. Forstner, and F. Lafarge, "Recovering line-networks in images by junction-point processes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1894–1901.

[8] I. Laptev, H. Mayer, T. Lindeberg, W. Eckstein, C. Steger, and A. Baumgartner, "Automatic extraction of roads from aerial images based on scale space and snakes," *Machine Vision and Applications*, vol. 12, no. 1, pp. 23–31, 2000.

[9] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234-241.

[10] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., and Rueckert, D., “Attention U-Net: Learning Where to Look for the Pancreas,” *arXiv preprint arXiv:1804.03999*, 2018.