Developing a more frequent and faster method for segmentation of urban roads from satellite images

Project ID: SM21D001

BY

Md. Rakibul Hasan
ID: 183-15-11826

AND

Miraziz Salehin
ID: 183-15-11901

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Dr. Sheak Rashed Haider Noori
Associate Professor & Associate Head
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Dr. Md Zahid Hasan
Associate Professor & Coordinator MIS
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

SEPTEMBER 2022

APPROVAL

This Project titled "**Developing a more frequent and faster method for segmentation of urban roads from satellite based images'**'", submitted by *Md. Rakibul Hasan* and *Miraziz Salehin* to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *12 September 2022*.

BOARD OF EXAMINERS

Chairman

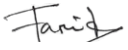
Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Md. Monzur Morshed [DMM]
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Ms. Samia Nawshin [SN]
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



External Examiner

Dr. Dewan Md Farid
Professor
Department of Computer Science and Engineering
United International University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Dr. Sheak Rashed Haider Noori, Associate Professor & Associate Head, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Dr. Sheak Rashed Haider Noori
Associate Professor & Associate Head
Department of CSE
Daffodil International University

Co-Supervised by:

Dr. Md Zahid Hasan
Associate Professor & Coordinator MIS
Department of CSE
Daffodil International University

Submitted by:

(Md. Rakibul Hasan)
ID: 183-15-11826
Department of CSE
Daffodil International University

(Miraziz Salehin)
ID: 183-15-11901
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First of all, and initially, we convey our profound appreciation to Almighty God for His wonderful gift, that has enabled us to complete the final year proposal successfully. **Dr. Sheak Rashed Haider Noori**, Associate Professor & Associate Head, Department of CSE, Daffodil International University, Dhaka, deserves our deep gratitude and appreciation. Our supervisor is an expert in the field of computer vision and deep learning, which was employed to finish this research, and he has a keen interest in it. We were able to finish the project thanks to his unwavering compassion, academic guidelines, continuous motivation, consistent and energized oversight, suggestions for improvement, valuable suggestions, reciting many inadequate manuscripts and adjusting them at all phases, and perusing many lesser documents and clarifying them at all stages. We would really like to extend our sincere thanks to our Co-Supervisor, Md. Zahid Hasan, Associate Professor, Department of CSE, for his invaluable assistance in completing our project, as well as to the other academic members and employees of Daffodil International University's CSE department.

ABSTRACT

In the recent era of computer vision research and development, the researchers still struggle in the case of an upkeep in contribution to this genre. Image segmentation is the technique of dividing a picture into useful regions and items. It may be applied to comprehending and recognizing scenes in a variety of industries, including ecology, healthcare, automation, and aerial photographs. So here we propose, an automatic method for recognizing the urban network of highways on high resolution satellite images has been developed and implemented. Then we also did several experiments regarding our satellite-based image dataset with the U-Net, U-Net with ResNet50 as an encoder, DeepLabV3+ ResNet50 as an encoder and the DeepLabV3+ ResNet101 for the comparison experimentation. Our main purpose of this research work was to apprehend between the last decade of research to our own accorded model. A comparison between the results was then introduced in our research article. In such circumstance we can say that our result was divided into various contexts of the models in order of our result convention. In case of our result estimation the best performance was acquired by the DeepLabV3+ ResNet101, the dice loss is 0.05 and the mean IoU 90%. The U-Net ResNet50 got the same IoU of 90%, but the dice loss was 0.06. Which and why we would prefer our model of the DeepLabV3+ ResNet101 to be the best in case. We prepared out dataset in order of accustom to our models structure. We hope our work would be enlightened in case of the image segmentation and computer visions massive indenture.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

In the worldwide consolation and numerous works uphold the world's technological simulations and modifications, computer vision has proven to change not only one's perspective while looking forward to the technology, but also it has made the technologies preferable to use for all of us. Now, the dataset that we selected had the resolution of 1500x1500 pixels in each image, which were quite enough for our training and testing purposes of training and testing our model. The main purpose of our work is to propose a more stable model when it comes to road segmentation. It is thought to be a difficult semantic task to identify and organize uniform regions for study by image segmentation, which is the division of an image into a set of well - defined parts [1]. [1] asserts that a properly segmented image should exhibit a number of basic characteristics, including I region consistency and similarity in its characteristics, including such gray level, color, or texture; (ii) geographic area consistency, without gaps; (iii) major difference among both neighbor nodes; and (iv) geographic accuracy to seamless and very well boundaries. In order to investigate relevant behavior inside the learning process, we suggest certain ways in this work to integrate the side-outputs from various layers by employing straightforward merging functions. We also investigate the total number of side-outputs required to develop a workable regional offer. In order to exclude some undesired short segments, we further suggest using an article filtering based on mathematical morphological idempotent functions [2]. Convolutional networks are frequently employed for classification tasks in where the outputs to a visual is a single course label. The desired output, or the assignment of a class label to each pixel, should incorporate localization in many visual tasks, particularly in biomedical image processing. Thousands of training photos are typically out of reach for biomedical jobs as well. In order to predict the classifier of each pixel, Ciresan et al. [3] built a system in a sliding-window arrangement using the local region (patch) surrounding each individual pixel as input. The main reason a U-Net is called as a U to begin with its because of its "U" shaped architecture. The updated DeepLabV3+ system we present in this paper features several improvements compared to its first version

reported in our original conference publication [4]. Actually, Highway Network [5] offered gated shortcut connections before Resent even used shortcut connections. The amount of information that can pass through the shortcut is controlled by these parameterized gates. A parameterized forget gate inside the Long-Term Short Memory (LSTM) [6] unit, which regulates how much information flows to the following time step, uses a similar concept. Resent can therefore be viewed as a particular instance of Highway Network. However, tests reveal that Highway Network does not outperform Resent, which is odd because Resent is present in the optimal solution of Highway Network [7], therefore it ought to perform at least as well as Resent.

1.2 Motivation

In case of creating or working with Deep Learning models, the scientists have always faced quite natural affection towards dissolving their curiosity towards solving real life-based problems. Basically, the DL approaches were meant to rule over the era of machine learning models. We wanted to complete our work while having similar skeptical thoughts, if our knowledge in this module could help us to do proper contribution towards our global or per-say, the problematic consults of human-kind, it would be sentiment of proud and honor for us. Which and why we proposed our case of doing a work that's come with tough handling and a larger parameter of understanding towards computer vision. We hope to full-fill our perception towards our ability to adjust our knowledge in the betterment of our society.

1.3 Rational of the study

Today we live in a smart world. We always try to solve our daily life problem using technology. Nowadays road detection & segmentation is great problem. So, our proposal is we develop a better model than can segmented road from satellite-based image than previous model [8]. There are many available convolutional Neural Network Model, available in terms of the work with computer visions numerous research projects [9]. This models also compels towards many wayward methods which can sometimes row as a wave

of per say, ups and downs in parameters of detection and segmentation performance. Which and why the understanding of this methods in crucial for our work towards this research. We also need to determine about the images we are willing to execute for this kind of research. As the images varies a big amount of contrast in case of detection and segmentation accuracy panel.

1.4 Research Questions

- What is road segmentation
- What is CNN?
- What is Deep Learning?
- What is Neural Network?
- Why Deep Learning is important?
- How to Deep learning works?
- What is U-Net
- What is ResNet50 and ResNet101
- What is DeepLabV3+

1.5 Expected Output

The predicted result was to segment the annotated parts of images which were roads. The segmentation speed and accuracy were the condemned using various curves and charts. Which were then later presented in this report. As we first predicted before beginning our work, it is highly sufficient towards achieving the height we were expecting of by proving over previous works.

1.6 Report Layout

This report in varied in total of six different chapters. Which are capable of extending the understanding of "Road Segmentation" more briefly.

Chapter 1: Introduction

The source of inspiration is explained, as well as the proposition's goal and introduction.

Chapter 2: Background Studies

The relevant work is discussed, and major popular strategies are introduced in relation to the work.

Chapter 3: Research Methodology

The approach for data collection, data pre-handling, and element determination is presented.

Chapter 4: Experimental Results and Discussion

The evaluation grouping philosophies are defined, and the results are reviewed.

Chapter 5: Implementation

The inquiry, the accuracy assessment, and the assessment plan are all introduced.

Chapter 6: Conclusion and Future Scope

The conclusion has been reached, and promises have been depicted.

CHAPTER 2

Background Studies

2.1 Introduction

While there was a deliberate debate between the recent decade in the structural premonition of computer visions most known researches in image classification and various object detection. In the 90s [10] CNN wasn't just a fascination, but also in the high payload in cases of computer vision research modules. But with the continuous high performances given by the SVM (Support Vector Machine), it was out of the league. However, in 2012, [11] the ILSVRC (ImageNet Large Scale Visual Recognition Challenge) [12] changed the perspective on this term, when Krizhevsky et al. showed that there is more to think on CNN, by gaining excruciating high accuracy on image classification. But there were still many flaws in the case that are still to be proven as it was the solution for the recent eras of technological trends. For example, CNN's localization problem, known as a troublemaker in the case of object detection, was "Recognition Using Regions" paradigm [13], however it was solved in the cases of both object detection [14] and the semantic segmentation [15]. As the time passed, more and more improvements were done which made the basic CNN to transform into a most famous in CV research. In the [16], Dong et al. used around 80 images to train a very large scale of random images for Pascal Visual Object Classes consisting of 20 categories FSD (Few-Shot Detector). Their main purpose was to prepare a number of models, as each of the models which will generate training labels for the case of image detection with high accuracy from those are the randomly unlabeled images. And then there was the LSTD (Low-Shot Transfer Detector) [17] which works as like tuner for the object detectors, the LSTD has two unique features when it comes to the regularization task, one of them for the depression in the background and another one is for to link up between the target domain to the transfer domain. There was quite very interesting research one the YOLO V2 [18] as Kang et. al. [19] made an extension to detection in the case of a single object. It works by reweighting the feature which can be predicted via a meta model. And there was also the Schwartz et. al. [20] who

modified Faster R-CNN using a metric learning module, which was capable of using a handful of examples in case of evaluating the similarities from each predicted box. And in recent works to be mentioned when it comes to terms with our model, there was a zero-shot detection [21,22,23,24] process, in case they were able to use a approach to train a joint embedding for acquiring the query image and texted elaboration while skipping the visual elaboration in the novel categorization. While in our case we were able to collect the visual description deliberately.

2.2 Related Work

In the recent decade the computer vision has made its way toward a lot of prosperity in the recent decade consequential improvements in computer science. Computer vision mainly consists of the work of localizing and individualizing a certain labeled object from visualized data [25]. As we drove through the existing works behind the road detection from various different works through the last consistency using CNN, we stumbled upon the work done by Henry et al. [26] where they used the FCNN (Fully Convolutional Neural Network), modifying with consistency in order of road segmentation. The FCNN could be proven to be quite a very circumstantial model for the road detection. In such cases they added a tolerance rule in case of handling significant mistakes and in the case of enhancing the quality of the extraction of the roads. In a similar study Xin et al. [27] used the Dense U-Net [28] which is a multiphoton image segmentation model based on CNN. In their work they used Encoder and decoder in order to comprehend the model towards training and getting the output phase of the segmentation. When they combined the dense connection mode with the U-Net, they were able to overcome the issue of the occurrence of trees and shadows. And in the process to emphasize on the foreground pixels they used a weighted loss function in order to operate properly. When we did our significant digging of the dense utilized processes, Chen et al. [29] used DFPN (Dense Feature Pyramid Network) Used a more deliberate approach towards Road segmentation process. In such a case, they needed to preprocess the data for the deep learning model which is based upon DFPN [30]. By setting up a framework they fulfilled the feature extraction procedure. Their main purpose was to introduce a more frequent focal loss function which could come in handy when a

researcher is working towards hard classified samples that have less pixel foreground [31]. compared fully convolutional neural networks (FCNNs) to traditional neural networks for the purpose of segmenting roads in high-resolution synthetic aperture radar pictures. The authors improved FCNNs by adding a tolerance rule for spatially modest defects in order to extract roadways. To get a 44% IoU throughout the test set, Deeplabv3+ was updated with a class-weighted mean-squared error loss[35]. A CNN-based method called Dense U-Net was created by Xin et al. [36] to extract the road network from RS photos with a minimal number of parameters and resilient properties. In order to address the issues of tree and shadow occlusion and highlight foreground pixels, the model mixes dense connection with U-Net. Two datasets of high-resolution photos were used for the experiments, which contrasted three traditional semantic segmentation techniques.

To take into account the specificity and complexity of instance segmentation of roadways, Chen et al. [37] proposed a DL-based model employing a dense feature pyramid network. In order to employ in-depth features for the shallow feature maps with high-resolution pictures, shallow feature channels and deep feature channels in this study were concatenated. The hard-classified samples with the less-pixel foreground should be taken into account by computing mask loss in the DL model using the focus loss function, according to the authors. In comparison to state-of-the-art approaches, the trials showed that the suggested method improved the instance segmentation of road markings [38,39,40,41].

2.3 Research Summary

The images were collected in a sense of determining the quality, the pixel count and the cleared base. To create a circumstance where we could determine our model's perception for such dataset, we needed to go through the models' different parts, such as regressors, features, classifiers etc., such in case so that these portions are not particularly affected by the dataset's variation. As working in the field of computer vision we come to learn that, when the dataset is fully conclusive, the other parts are quite shorthand to overcome.

2.4 Scope of the problem

Our study is conservative towards the assertion of road data segmentation. In which case the images were then divided towards the class of only roads highlighting. The other portions which were not marked were considered as to be null value or not to be assigned towards our models learning process. In such basis models learning ratio and detection capability was then optimized to a level so that no underfitting or overfitting consequence doesn't occur.

2.5 Challenges

In this study, we used an angle analysis on voyager inputs to examine at carrier organizations. The component identification and over-inspecting procedures are similarly vital in enhancing overall findings, according to the suggested method. But when it comes to determining the performance evaluation according to other skewered approaches, we came across cutting edge methods and transformation algorithms, which worked vital role towards determining the improvement of our model. There are many consequences where the model seems to be confused with the classification process, in more deliberate term the model figures sufficient amount of annotated portion and unable to maintain the phase while segmentation of the role of the road's structural details.

CHAPTER 3

Research Methodology

3.1 Introduction

In this section, I'll briefly outline the processes I took to complete our research assignment. Road photos serve as the foundation for many naturally occurring image-based applications, including autonomous vehicles, hence the automatic identification and categorization of from road photographs has arisen as an essential priority. We used multiple Python programming languages to investigate the complicated Deep learning techniques. Because the dataset processing is the most significant component of the Deep learning technique, the algorithms were chosen and implemented on the sample on a constant schedule. We considered processing cost, memory needs, and simplicity while deciding which options to start with. Although we only used a few example, approaches to create our benchmarks, several of the elements we take into account can be swapped out for or merged with many more sophisticated techniques. The method that's been employed was U-Net, U-Net ResNet50, DeepLabV3+ ResNet101, DeepLabV3+ ResNet50. The goal of this research was to discover a better segmentation method. Figure 1 demonstrates how we think about the procedure when looking at a flow chart. As from the start the main procedure starts with the pre-processing the dataset in order to go through the models training and testing procedure. The models were accustomated to the use the data augmentations protests. What data augmentation is, that the dataset should be manipulated in order to re-demonstrate the training images to produce authentic dataset which is basically in larger scale then the existing dataset. What it does is, it improves the model performance. The models were then thoroughly designed in order to perform through our presented dataset with the high-quality performance. The model explanation given later in the article. And then the model tuned and applied with the testing data in order to evaluate the model in proper accustomating. This step uses two fundamental aspects of a polygon, such as the ratios between of cube of the border and the size, and the amount of edges, to verify the lumps that have been acquired from the segmentation phase according to their

shape. Any form for a road should fit into one of the one design classification which is polygon.

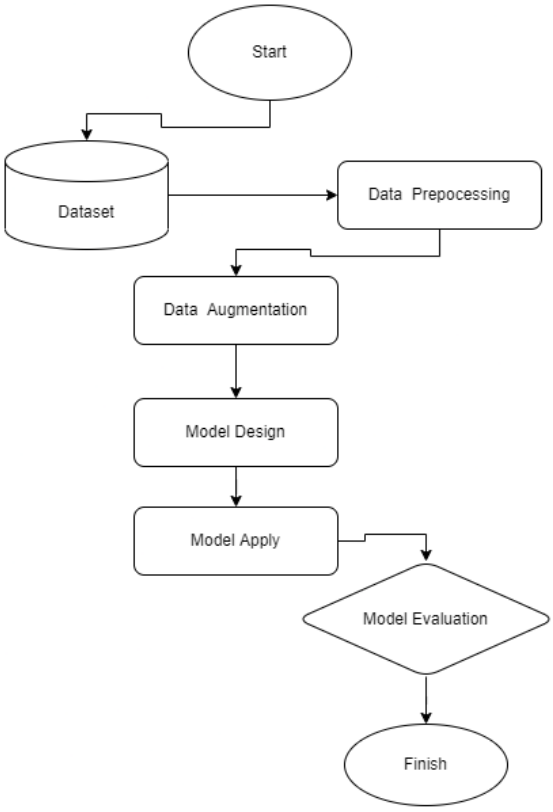


Figure 1.1: Workflow Procedure

3.2 U-NET

The upgraded deep convolutional network model known as U-Net has a shape-U-like topology [31]. Figure depicts the U-Net design in detail. U-Net has a higher segmentation accuracy and fewer training sets than other convolutional neural networks. As shown by the Figure 1.1, it is made up of an encoder and a decoder that are equal with the intermediate layer's symmetry axis. Through convolution layer down-sampling (sometimes known as pooling) layers, the encoder extracts image features. In contrast, the decoder performs feature image up sampling, and Sensors 2021, 2153 Between the matching encoder and

decoder levels in layer 4 of 21 there are cross-layer connections, which can assist the up-sampling layer in recovering the image's details. The U-net is proportional in design. It has an embedding layer and a decoder, two structures. By means of skip connections, these components are linked to one another. Additionally, it can maintain U-net feature maps in their original size. Figure 1 depicts the Ronneberger and research team's creation of the U-Net structure. The cornerstone is the architecture and design component that specifies how these layers are organized in the encoder network and dictates how the decoder network should be constructed. Vanilla CNNs like Convolution layer, ResNet50, ResNet101, Formation, and Efficient Net are frequently used as backbones because they can handle encoding and down sampling on their own. To create the final U-Net, these networks are extracted and their equivalents are created to execute decoding and up sampling. The photos are rather huge at 1500 x1500 x 3, which negatively impacts performance. The photos are divided into different sections with a size of 256 x 256 x 3 and provided in parts to enable a better performance of a model. Thus, the validity of the system was improved.

Specifically, the 3 3 convolution operation, RELU function, and 2 2 max-pooling layer make up the picture features extracted that the coder extracts using the convolutional layer. The down sampling process is repeated four times. After each pooling process, the size of feature images shrinks and also the amount of channels doubles. The decoder conducts a 2 2 deconvolution layer up sampling (also known as transposing convolution) in order to gradually retrieve the image data. The decoder section completes up-sampling four times, matching the encoder part. After each up-sampling, the size of a feature images grows larger while the number of outlets decreases by half. Through the conjunction of both the corresponding feature patterns of the encoder and decoder, segmentation is made possible by the detailed location information that can be more efficiently saved with shallow networks. There are 23 convolutional layers in the U-Net.

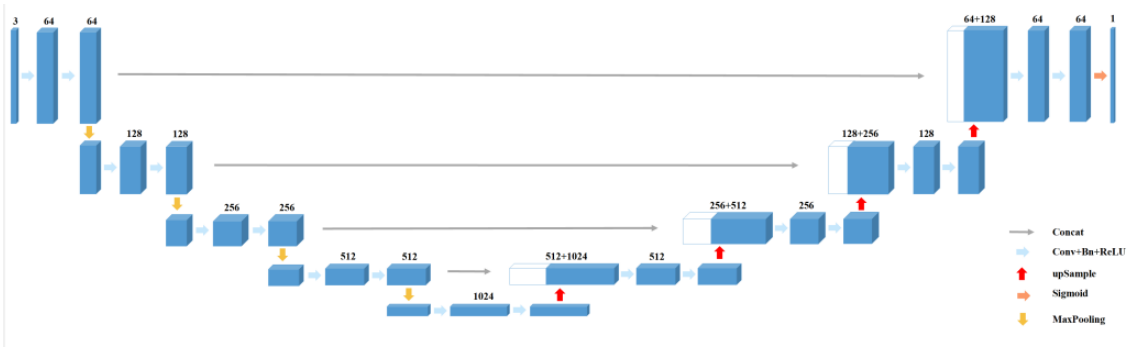


Figure 1.2: U-NET Architecture

3.3 ResNet50 and ResNet101

Res-Net uses workarounds to address this issue.) It is a method known as skip connection. By skipping multiple levels, a connection is made directly to the output. This prevents the gradient from expanding or vanishing.) There are numerous variations of Res-Net architectures. The names are given based on the depth levels. The Res-Net architecture comes in various forms. Typically, it is named based on the number of layers.

The three Res-Net architectures with the greatest usage are ResNet34, ResNet50, and ResNet152. The model weights trained on tens of datasets can be attained using the ImageNet dataset. The performance is greatly improved when the parameters from the ImageNet dataset are used as the starting weight for training. Many machine vision models use the neural network Res-Net (Residual Network), which serves as their foundation. Deeper neural network training is relatively difficult under normal conditions.) The vanishing gradient issue appears to be causing higher mistake rates while training deep neural networks. Theoretically, stacking Convolution Neural networks should reduce training error. However, it appears that the Convolution Neural Network's training error increases as more layers are added in practice. Here, there may be issues with corruption or optimization.

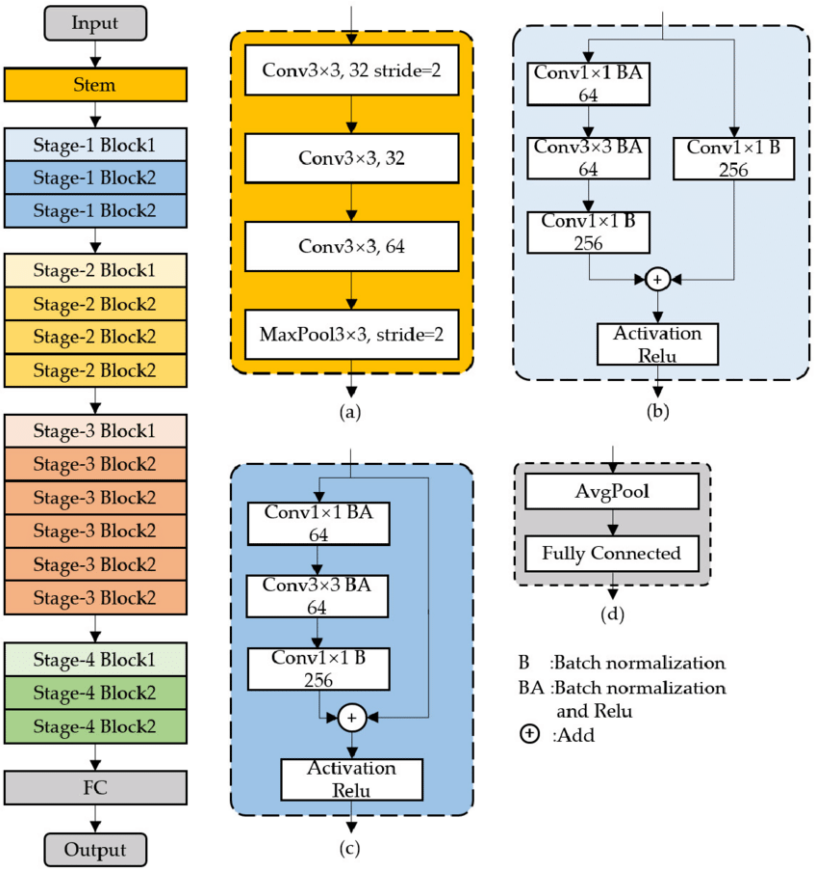


Figure 1.3: The architecture of RESNET

3.4 DeepLabV3+

Xception serves as the network's backbone for Deeplab-V3+. The Res Net-like residual connection technique, which Xception has incorporated, considerably speeds up the convergence of Xception and improves deep feature extraction; To achieve the fusing of shallow characteristic details and deep semantic features, DeepLabV3+ network adds Encoder-Decoder Architecture, the feature map created in the Encoder, and uses concatenation for the subsequent stage of fusion.

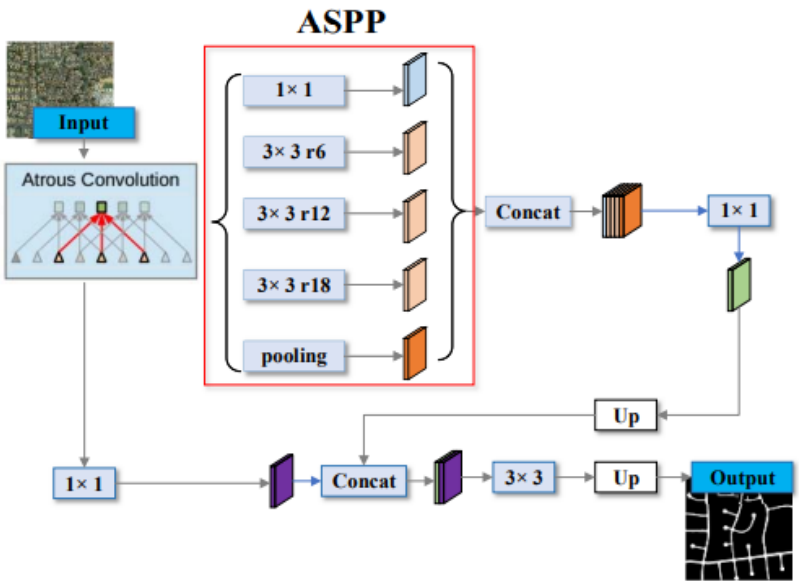


Figure 1.4: The core architecture of DeepLabv3+

To obtain the relevant data, the knowledge area and global average pooling of the feature map are employed. The knowledge area of various scales acquired by the ASPP module is then combined by convolution.

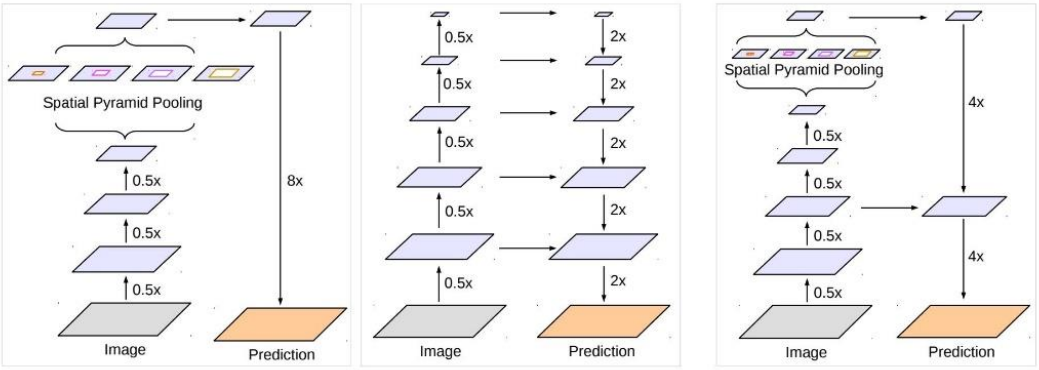


Figure 1.5: Spatial Pyramid Pooling, Encoder and Decoder, Atrous Conv

The suggested model, DeepLabv3+, recovers the precise bounding boxes from of the encoder modules, which also contributes rich semantic information to the model. By using atrous convolution, the encoder module enables us to automatically extract at any resolution.

3.5 Research Subject and Instrumentation

We chose road image segmentation as the phrase to represent the entire idea. So, because field of DL is increasing as a consequence of technical developments, this research will follow up a lot more ground in comparison to earlier attempts. In order to accommodate such a lengthy period of rigorous study, the computer I needed in order to complete my research model was built to a high specification.

Hardware and Software:

- 16GB RAM and Intel core i5 8th generation.
- 1 TB Hard Disk.

Tools:

- Windows 10
- Python 3.8
- Matplotlib
- U-Net
- ResNet 50 and ResNet 101
- DeeplabV3+
- Google Colab Notebook

3.6 Road Dataset

The dataset we selected contains 1171 satellite images. Each image contains a binary classification that divides it into road sections and non-road segments. The photos have a 1.2 m2 area and a resolution of roughly 1500 x 1500 pixels. The collection includes flora, houses, roads, rivers, and vehicles over a total area of more than 2600 km2.The labels are binary images where roads are represented as constant 7-pixel thick lines generated by rasterizing and dilating Open Street Map vector centerlines. Validation and test splits are

thus statistically under-representative of the whole dataset. To increase our results' consistency, we perform our ablation study on a re-split with 1108 images in train data 49 images in test data and 14 images in validation data and report our final results on the official split in.

The dataset we used was consistently in better resolution then we would have expected to have. The images had 1500 per pixel in the aspect of X and Y angle from the display perspective. In the case of using the neural network models [32,33,34] where the images taken from satellites of cities filled with roads were annotated. The pixels were then divided into one particular class, which were then labeled as "Roads". Basically, the visualization was determined by highlighting the roads, and other parts were not to be counted as roads. As computer vision is one known to be the vastest sector of research area, the annotation process is one of the most significant works in the case of the work zone. Basically, when it comes to training machines the visualization data, such as video or image fragments, can become more complicated than the text string data. So, annotation can be also interpreted as labeling the data in case of pre-processing it for the supervised learning methods. As our target object for detection, we don't actually attain any certain type of shapes.

3.7 Dataset Distribution

After all of the scrubbing and removing of all the untamed data, the dataset was finally ready to be divided into classes for training and testing. The whole quantity of data was distributed in case of the most afform way so that the models could perform at their highest as possible. The train data size was 1108, whereas the test data size was 49. The frequency distribution of the dataset is shown below.

3.8 Preprocessing

As computer vision is one known to be the most vast sector of research area, the annotation process is one of the most significant work in the case of the work zone. Basically, when it comes to training machines the visualization data, such as video or image fragments, can become more complicated than the text string data. So, annotation can be also interpreted as labeling the data in case of pre-processing it for the supervised learning methods. As our

target object for detection, we don't actually attain any certain type of shapes. That's why we used polygon segmentation in case of highlighting the streets. In other cases of experiment via U-Net, U-Net ResNet50, DeepLabV3+ ResNet50 and the DeepLabV3+ ResNet101 we simply used the separated masks for training and testing.

3.9 Statistical Analysis

1. In the dataset total 1172 data is presented.
2. Our target class is Road
3. 1108 data is used for the train.
4. 49 data is used for the test.
5. 14 data is used for validation
6. Highest accuracy achieved 90%.

3.10 Implementation Requirements

The deep learning model was implemented using Python as a scripting language. The dataset is loaded using the Panda package, and prepping is done using the web-based annotation tool. The whole process is carried out in a Google Colab notebook.

CHAPTER 4

Experimental Results and Discussion

4.1 Introduction

As we implemented our models the evaluation basis is in to order of contrast to come by. The results in case were then divided into different parameter evaluation, now we have discussed the performance measurement in this section thoroughly. The performance description were given according to the graphs, figures, alignment. We discussed the measurement of our model and the models that rather performed in order. As our base model was customized according we also created the measurement comparison table in case of the models comparison concealing.

4.2 Model Performance

A network of roads is highlighted in the photographs after the neural network has run, however there are roads as well as other imperfections in the image. The final image is subjected to extra processing in order to remove mistake. Mathematical morphology, which analyzes and analyzes geometric structures in images, is built on set theory. After processing, the binary image is seen as a set of pixels. Some primitives use mathematical morphological approaches to carry out one or more actions on pixel-by-pixel image processing. The two basic processes are dilation and erosion (expansion). As the titles suggest, the lines in the image grow thinner after erosion and thicker after dilatation. This study employs the closure procedure, which calls for carrying out dilation and eroding in that order. To use the model was possible to considerably enhance the outcome of identifying roads on satellite pictures by post-processing.

As for the U-NET model, we also implemented the same dataset. The training loss in the figure is regarded as the capacity of the model that consisted of the training data and the validation loss indicated the model's capacity to comprehend the ability of the model to cope up with the new data.

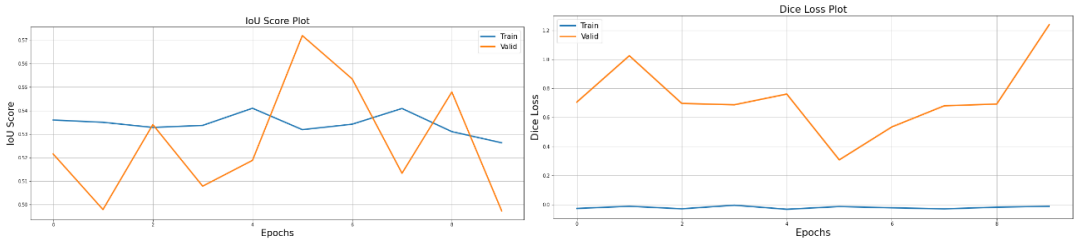


Figure 1.6: Training and Validation IoU and Loss of U-Net

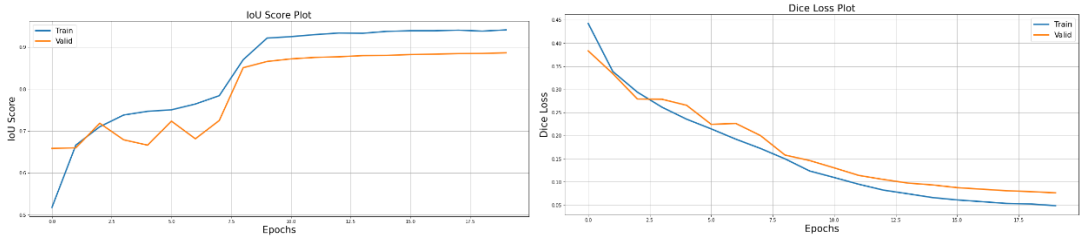


Figure 1.7: Training and Validation IoU and Dice Loss U-Net ResNet50

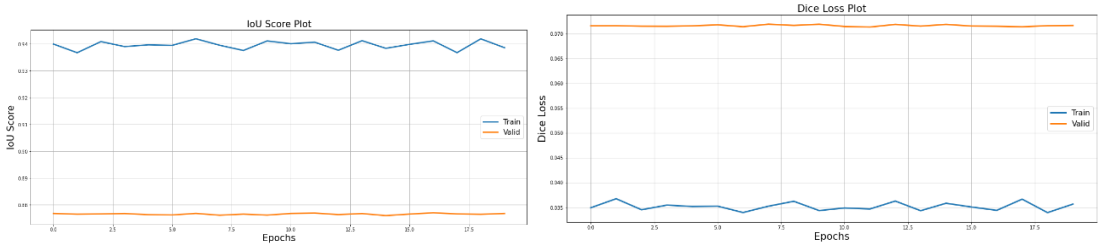


Figure 1.8: Training and Validation IoU and Dice Loss DeepLabV3+ ResNet101

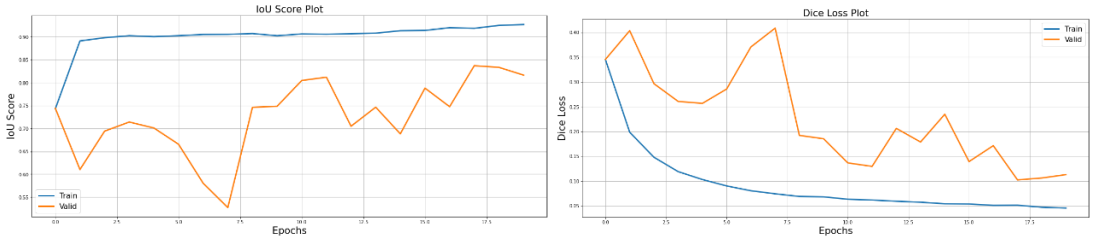


Figure 1.9: Training and Validation IoU and Dice Loss of DeepLabV3+ ResNet50

The final segmentation of the U-Net ResNet50 model is proven to be quite as we expected it to be. Particularly in two-lane roadways, the suggested model exhibits clear, clean outcomes with minimal noise. As demonstrated by the red rectangle in the second row, the model is able to accurately and effectively detect each lane as well as the intricate road features. The model can also discriminate between things with a similar structure to highways, such as parking spaces and building rooftops.

The segmentation via U-Net ResNet50 was more frequent and time consuming than the rest of the models. Which would then also take a high toll also on our system rig, when the actual experimentation was complete.

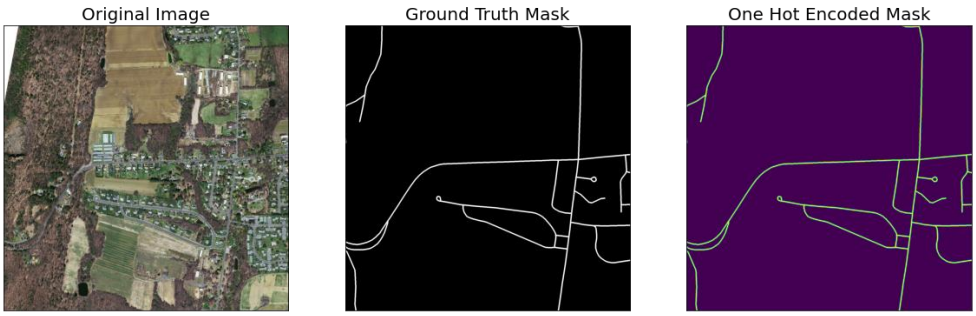


Figure 1.10: The original, The ground truth mask, The hot encoded mask of U-Net

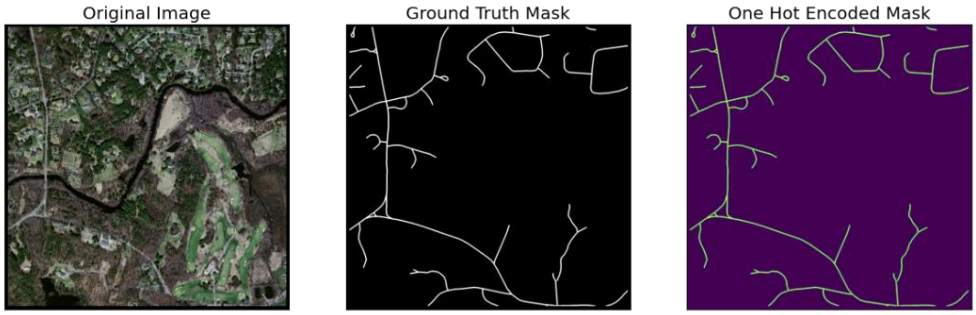


Figure 1.11: The original, The ground truth mask, The hot encoded mask of U-Net ResNet50

Figure 1.12: The original, The ground truth mask, The hot encoded mask of DeepLabV3+ ResNet101

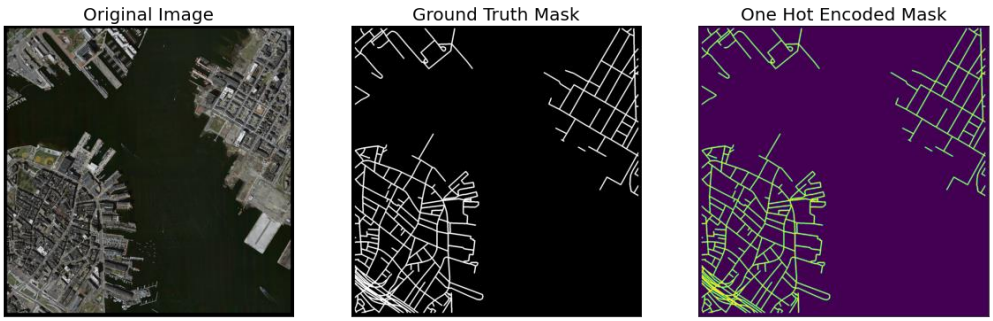


Figure 1.13: The original, The ground truth mask, The hot encoded mask of DeepLabV3+ ResNet50

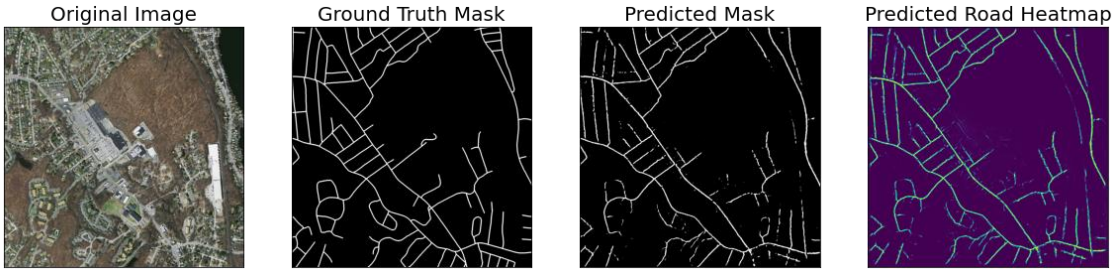


Figure 1.14: The original image, The ground truth mask and the predicted road heatmap by DeepLabV3+ Res-net101

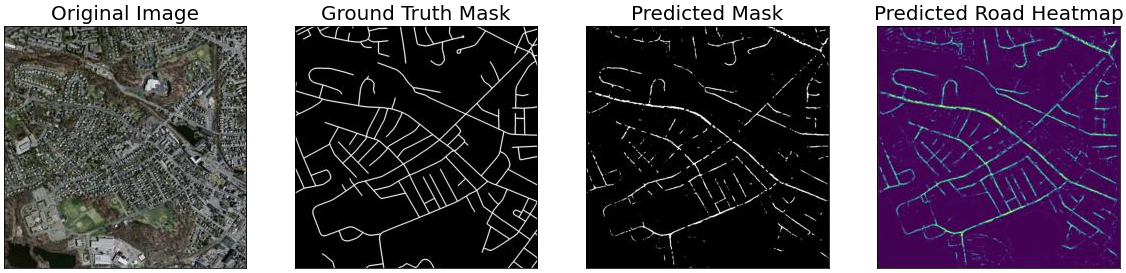


Figure 1.15: The original image, The ground truth mask and the predicted road heatmap by DeepLabV3+ Res-net50

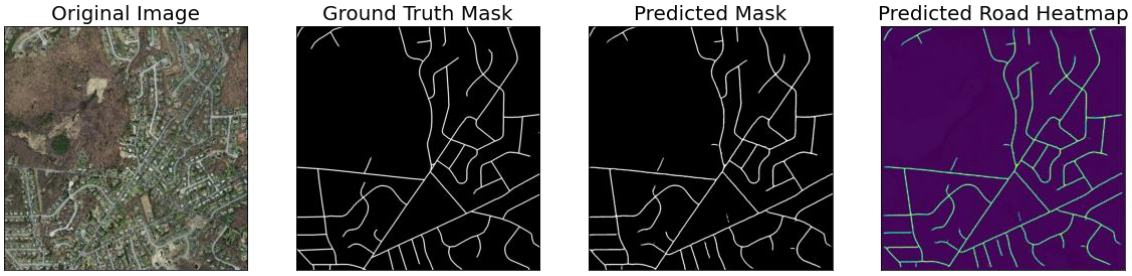


Figure 1.16: The original image, The ground truth mask and the predicted road heatmap by U-Net Res-Net50

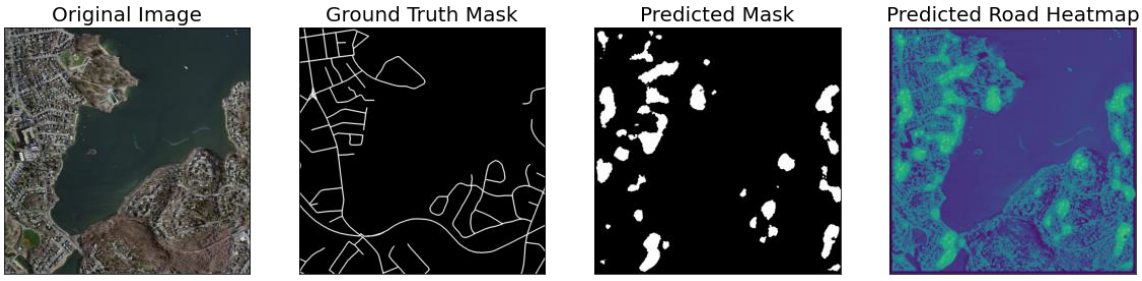


FIGURE 1.17: THE ORIGINAL IMAGE, THE GROUND TRUTH MASK AND THE PREDICTED ROAD HEATMAP BY U-NET

TABLE 4.1: COMPSRISON BETWEEN MODELS

Model	Train IoU	Test IoU	Mean IoU	Mean Dice Loss
U-Net	52%	49%	57%	1.59
U-Net ResNet50	94%	88%	90%	0.06
DeepLabV3+ ResNet50	92%	81%	87%	0.08
DeepLabV3+ ResNet101	93%	87%	90%	0.05

TABLE 4.2: COMPSRISON WITH OTHERS MODEL

Model	IoU
Our Model (DeepLabV3+ ResNet101)	87%
Et. Lichen [42] used a D-Linknet called network structure based upon the Deep U-Net With 7 pooling layers at the images size of 1024 x 1024 pixels. Their Link Net 34 structure was able to perform a little bit better than the UNet base structure.	64%
Et. Buslaev [43] used a ResNet pretrained decoder from the well known Vanilla U-Net family for road extraction sequence.	64%

IOU is indeed utilized in non-maximal suppression, that seeks to remove several boxes that are located around a single object dependent about which box has the highest confidence.

$$IOU = \frac{Area\ Of\ Intersection\ Of\ Two\ Boxes}{Area\ Of\ Union\ Of\ Two\ Boxes}$$

One of the most fundamental techniques for comparing data samples in machine learning is indeed the intersection over union (IOU) method. In statistics, the Correlation Index is often referred to as IoU. IoU is used as an evaluation technique for computer vision tasks such as objects recognition, machine vision, classification techniques, and others.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

IoU determines how closely two groups of elements overlap in general. We can create the IoU depending on the left circle, width, and altitude of A's bounding box using a boundary

box extrapolation, where A stands for the bounding box while B for the prediction box. The models were able to perform as according to our estimation of the performance result. The U-Net base structure was able to achieve in mean IoU of 57% while we added the ResNet50 as an encoder in case, and the mean IoU achieved of 90%. And if we shade light on the DeepLabV3+ with the encoder of ResNet50 the achieved mean IoU was 87% and with the encoder of ResNet101 the IoU was achieved as per 90%. As we can see that the both DeepLabV3+ ResNet101 and the U-Net ResNet50 was able to perform at the same height, but when we see the mean dice loss the DeepLabV3+ ResNet101 outperformed the other existing models.

4.3 Summary

In the circumstances, We had to come up with a new strategy by re-processing the dataset into a more class-like environment. The main purpose was to accustomate a single base from existing models and then tune it to the existing datasets preference so that it can segment to the highest accuracy measure. A several other deep learning model evaluations was also presented in previous section to clarify existing model evaluation and figure out the best outcome.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

Every human expression might be related to the phrases we see on a regular basis in the virtual environment on numerous internet platforms. In this situation, having a system in place to distinguish between genuine impulses and or before hostility is crucial for these platforms. That's why we've chosen to focus in our efforts on sensing many of the most intriguing genres of all period, including such Craftsmanship, Scientific research, Economic history, Climate, Foreign, Car crashes, Viewpoints, Violent act, Enjoyment, Athletics, Training, Diplomacy, or Mishaps, Views, Violent assault, Amusements, Games, Youth development, and Current events. We may expect a more decisive and diversified digital era as a result of this.

5.2 Impact on Environment

On numerous online sites these days, it is pretty common to communicate a sociopolitical topic with a highly definite term of mixed scorn. In a range of areas, classification, on the other hand, provides for the depiction of a social perspective on terminology. As a result, we must exercise greater caution when it comes to scientific corroboration, so that systems may fully know which is which. Because there have been several cases of people being able to understand various environmental paths through the use of maps as the primary strategy. As a corollary, segmentation has been found to have both positive and negative effects on people's environmental perceptions.

5.3 Ethical Aspects

People of any age may now access the media outlets on the internet. As both a result, the user limits criteria are no longer relevant. Because security mechanisms to discriminate across individual and spiritual viewpoints are insufficient. The full context of a concept given across platforms must be understood. This has been found to be destructive to

people's social standards in a variety of situations. If the algorithm does not understand the parameters, it may prohibit a genuine transmission while allowing a detrimental one. Road mapping may be interesting in the context of a poignant situation, but more often it might push a user away from actual path owing to the numerous regulatory burdens enforced.

5.4 Sustainability

- Worldwide, there are more than 1.3 billion people who uses google map on daily basis.
- Many well-known companies and brands are now authorized on their products delivery and other accustoms on the mapping system.
- 60% of those people are probable to suffer in order of their daily work consent in order if they are unable to attain their daily schedule entry to the web platforms.

CHAPTER 6

Conclusion and Future Work

6.1 Conclusion

A method for segmenting roads in satellite SAR pictures developed as a result of a deep neural network has demonstrated its great efficiency. The quality of the terrain marking produced is equivalent to that of hand marking. The inclusion of this framework as one of the parts of an intelligent vehicle system will enable automatic appraisal of the beaten - path in order to create and organize new transportation networks, as well as serve as a backbone route for unmanned air surveillance of rural settlements. It was feasible to increase the quality of road analysis using the constructed neural network. For driving safety and performance to be improved, an accurate real-time traffic sign identification technique with a low percentage of wrongful convictions is crucial. As the best performance was achieved via the DeepLabV3+ with the encoded as ResNet101. In which case as we preferred the models to have an outstanding performance as expected. The dataset was a major influence in this case. An algorithm for the identification of road traffic signals has been put forward in this research. The capacity of the segmentation-based detection method to identify a ROI on a street sign is determined to be reliable. The form classifier subsequently reduces the amount of time required for computing in the subsequent step of identification and excludes items that are not traffic signs but have colors that are similar to those of road signs (such as vehicles and houses) with the high likelihood. The classification/recognition phase solely works the with ROIs which are designated as prospective candidate areas after the initial identification. And we hope in future our work will shed light on the aspect of this subject with more lucrative procedures.

6.2 Recommendations

- Annotate the required label for the model.
- The percentage of data categories must be divided in a manner to models understanding.

- To get a better outcome, create a more frequent concerned neural network.
- For segmentation, parameter adjustment is required.

6.3 Implication for Further Study

Image segmentation is a key field of research in image application of machine vision, with automated driving being the most common application scenario. The great majority of existing image segmentation algorithms based on the deep network due to the multiple restrictions of power supply and connectivity in in-vehicle systems. Despite the high level of classifier performance, the problem of excessive mesh distortions and segmentation is clear, and the high cost, computing, and energy consumption devices required is challenging to implement in real-world applications.

REFERENCES

[1] D. Dom´ınguez and R. R. Morales, Image Segmentation: Advances. Magnum Publishing LLC, 2016, vol. 1, no. 1.

[2] L. Najman and H. Talbot, Mathematical morphology: from theory to applications. Hoboken, USA: John Wiley & Sons, 2013.

[3] Ciresan, D.C., Gambardella, L.M., Giusti, A., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: NIPS,pp. 2852–2860 (2012)

[4] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," in ICLR, 2015.

[5] R. Srivastava, K. Greff and J. Schmidhuber. Training Very Deep Networks. arXiv preprint arXiv:1507.06228v2,2015.

[6] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural Comput., 9(8):1735–1780, Nov. 1997.

[7] Mnih V 2013 Machine learning for aerial image labeling Doctor of Philosophy Graduat

[8] network IEEE-International Conference on Recent Trends in Information Technology 1061-

[9] Mnih, V. (2013). *Machine learning for aerial image labeling*. University of Toronto (Canada).

[10] Mnih V and Hinton G E 2010 Learning to detect roads in high-resolution aerial images Computer

[11] Vision – ECCV 2010, Lecture Notes in Computer Science 6316 210-23

[12] Mnih V 2013 Machine learning for aerial image labeling Doctor of Philosophy Graduate

[13] Department of Computer Science University of Toronto

[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradientbased learning applied to document recognition. Proc. of the IEEE, 1998.

[15] A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. In NIPS, 2012.

[16] J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. FeiFei. ImageNet Large Scale Visual Recognition Competition 2012 (ILSVRC2012).

[17] C. Gu, J. J. Lim, P. Arbelaez, and J. Malik. Recognition using regions. In CVPR, 2009.

[18] J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders. Selective search for object recognition. IJCV, 2013.

[19] J. Carreira and C. Sminchisescu. CPMC: Automatic object segmentation using constrained parametric min-cuts. TPAMI, 2012.

[20] Xuanyi Dong, Liang Zheng, Fan Ma, Yi Yang, and Deyu Meng. Few-Example Object Detection with Model Communication. TPAMI, 2018.

[21] Hao Chen, Yali Wang, Guoyou Wang, and Yu Qiao. LSTD: A Low-Shot Transfer Detector for Object Detection. AAAI, 2018.

[22] Joseph Redmon and Ali Farhadi. YOLO9000: Better, Faster, Stronger. In CVPR, 2017.

[23] Bingyi Kang, Zhuang Liu, Xin Wang, Fisher Yu, Jiashi Feng, and Trevor Darrell. Few-shot object detection via feature reweighting. arXiv:1812.01866, 2018.

[24] Eli Schwartz, Leonid Karlinsky, Joseph Shtok, Sivan Harary, Mattias Marder, Sharathchandra Pankanti, Rogerio Feris, Abhishek Kumar, Raja Giryes, and Alex M. Bronstein. RepMet: Representative-based metric learning for classification and one-shot object detection. In CVPR, 2019.

[25] Ankan Bansal, Karan Sikka, Gaurav Sharma, Rama Chellappa, and Ajay Divakaran. Zero-Shot Object Detection. ECCV, 2018.

[26] Shafin Rahman, Salman Khan, and Fatih Porikli. Zero-Shot Object Detection: Learning to Simultaneously Recognize and Localize Novel Concepts. In ACCV, 2018.

[27] Berkan Demirel, Ramazan Gokberk Cinbis, and Nazli Ikizler-Cinbis. Zero-shot object detection by hybrid region embedding. In BMVC, 2018.

[28] Pengkai Zhu, Hanxiao Wang, and Venkatesh Saligrama. Zero shot detection. In TCSVT, 2019.

[29] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, 2010.

[30] Henry, C.; Azimi, S.M.; Merkle, N. Road segmentation in SAR satellite images with deep fully convolutional neural networks. IEEE Geosci. Remote Sens. Lett. 2018, 15, 1867–1871.

[31] Xin, J.; Zhang, X.; Zhang, Z.; Fang, W. Road extraction of high-resolution remote sensing images derived from DenseUNet. Remote Sens. 2019, 11, 2499.

[32] Cai, S., Tian, Y., Lui, H., Zeng, H., Wu, Y., & Chen, G. (2020). Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. Quantitative imaging in medicine and surgery, 10(6), 1275.

[33] Chen, S.; Zhang, Z.; Zhong, R.; Zhang, L.; Ma, H.; Liu, L. A dense feature pyramid network-based deep learning model for road marking instance segmentation using MLS point clouds. IEEE Trans. Geosci. Remote Sens. 2020, 59, 784–800.

[34] Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin, Germany, 2015; pp. 234–241.

[35] Xin, J.; Zhang, X.; Zhang, Z.; Fang, W. Road extraction of high-resolution remote sensing images derived from DenseUNet. *Remote Sens.* **2019**, *11*, 2499. [**Google Scholar**] [**CrossRef**][**Green Version**]

[36] Chen, S.; Zhang, Z.; Zhong, R.; Zhang, L.; Ma, H.; Liu, L. A dense feature pyramid network-based deep learning model for road marking instance segmentation using MLS point clouds. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 784–800. [**Google Scholar**] [**CrossRef**]

[37] Emara, T.; Munim, H.E.A.E.; Abbas, H.M. LiteSeg: A Novel Lightweight ConvNet for Semantic Segmentation. In 2019 Digital Image Computing: Techniques and Applications (DICTA). 2019. Available online: **https://ieeexplore.ieee.org/abstract/document/8945975** (accessed on 19 September 2021). [**CrossRef**][**Green Version**]

[38] Paszke, A.; Chaurasia, A.; Kim, S.; Culurciello, E. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. *arXiv* **2016**, arXiv:1606.02147. [**Google Scholar**]

[39] Aich, S.; van der Kamp, W.; Stavness, I. Semantic Binary Segmentation Using Convolutional Networks without Decoders. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, UT, USA, 18–23 June 2018. [**Google Scholar**]

[40] Sovetkin, E.; Achterberg, E.J.; Weber, T.; Pieters, B.E. Encoder–Decoder Semantic Segmentation Models for Electroluminescence Images of Thin-Film Photovoltaic Modules. *IEEE J. Photovolt.* **2021**, *11*, 444–452. [**Google Scholar**] [**CrossRef**]

[41] Hamaguchi, R.; Fujita, A.; Nemoto, K.; Imaizumi, T.; Hikosaka, S. Effective Use of Dilated Convolutions for Segmenting Small Object Instances in Remote Sensing Imagery. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1442–1450. [**Google Scholar**] [**CrossRef**][**Green Version**]

[42] Zhou, L., Zhang, C., & Wu, M. (2018). D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 182-186).

[43] Buslaev, A., Seferbekov, S., Iglovikov, V., & Shvets, A. (2018). Fully convolutional network for automatic road extraction from satellite imagery. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 207-210).

abcd

ORIGINALITY REPORT

18%	12%	9%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	5%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	4%
3	Waleed Alsabhan, Turky Alotaiby. "Automatic Building Extraction on Satellite Images Using Unet and ResNet50", Computational Intelligence and Neuroscience, 2022 Publication	1%
4	www.mdpi.com Internet Source	1%
5	"Computer Vision – ECCV 2018", Springer Science and Business Media LLC, 2018 Publication	1%
6	Yuewu Hou, Zhaoying Liu, Ting Zhang, Yujian Li. "C-UNet: Complement UNet for Remote Sensing Road Extraction", Sensors, 2021 Publication	1%
7	elib.dlr.de Internet Source	1%