Deep Learning Algorithms for Building Segmentation from Aerial Images

Mid-Term Minor Project Report

Submitted in partial fulfilment of the requirements for the degree of

M. TECH. in SIGNAL PROCESSING AND MACHINE LEARNING

Submitted by

YOGESH RAJU N R (232SP029)

&

NEERAJ KUMAR JHA (232SP012)

Department Of Electronics & Communication Engineering

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA (NITK)

SURATHKAL, KARNATAKA, INDIA – 575025

February 19, 2024

Table of Contents

List of Figures

ABSTRACT

This project centers on implementing and examining the performance of deep learning algorithms designed specifically for building segmentation within aerial images. Deep learning has revolutionized the field of satellite and aerial image analysis, and image classification holds particular significance when leveraging satellite imagery.

The task of building segmentation within aerial images stands as both a critical and complex challenge. Variations in building textures, backgrounds, and image acquisition conditions all contribute to the difficulty. These distinct hurdles, coupled with typically large image sizes and potentially diverse object categories, open a wide array of opportunities within the deep learning domain.

A core focus of this project lies in the way deep learning techniques facilitate the extraction of contextual information. By understanding the broader context surrounding a pixel, rather than considering it in isolation, these models better distinguish buildings from other elements within their visual environments. This enhanced scene understanding ultimately improves building segmentation outcomes.

CHAPTER 1

INTRODUCTION

Detecting buildings in aerial and satellite images is a multifaceted challenge within the fields of remote sensing and machine vision. Factors such as complex backgrounds, shadows, variations in perspective due to shooting angles, and occlusions from both structures and natural elements (like trees) can severely impede accurate building identification. The ability to precisely extract and delineate buildings holds immense value across several applications, including but not limited to:

- **Urban Planning and Development:** Building detection facilitates informed decision-making for managing urban growth, optimizing infrastructure, and implementing land-use policies.
- **Disaster Response and Assessment:** Rapidly identifying affected buildings supports targeted rescue efforts and enables swift damage quantification during natural disasters or crises.
- **Geographic Information Systems (GIS):** Automated building extraction streamlines the creation and updating of detailed maps used in geo-analysis and navigation systems.
- **Defense and Intelligence:** Understanding building distribution and patterns can hold strategic importance for both defensive and reconnaissance operations.

Aerial and satellite imagery represent indispensable resources for gathering large-scale information about Earth's surface. A primary concern of image analysis in these contexts is classifying objects. The need for effective object classification extends to humanitarian purposes in rescue operations and to military applications regarding the monitoring of strategic activities. The past reliance on manual image analysis – the careful study of images across seasons by humans – proves costly, laborious, and ultimately unsustainable when the sheer volume of available data is considered.

Automatic building detection in aerial images remains a major area of research. Other tasks reliant on this capability include creating comprehensive maps for geographic information systems and supporting urban development programs. Ideally, automated systems would identify, recognize, and segment buildings – essentially understanding the content of these images. These abilities would underpin higher-level analysis such as the assessment of structural damage following an event, the tracking of demographic changes through construction monitoring, and potentially even the analysis of military deployment patterns. The fundamental geometric characteristics of buildings, coupled with the variability of

surrounding factors and image acquisition methods, presents the ongoing challenge for image-based detection. Our central objective remains precise building classification as a foundation for segmenting building objects within backgrounds, offering a representation compatible with advanced analysis.

1.1 OBJECTIVE

The main objective of building segmentation from aerial images is to

- Design a building segmentation algorithm that isolates buildings from diverse landscape features within aerial/satellite images for precise urban mapping applications. This focuses on achieving detailed results that would directly benefit a specific use case.

- To learn about new techniques for pre-processing the raw data, image segmentation and deep learning algorithms build a new approach towards segmentation of buildings from aerial images.

CHAPTER 2

METHODOLOGY

The methodology followed to implement building segmentation for aerial images is as shown in the below block diagram, fig.1.

This implementation of building segmentation for aerial mainly follows 5 stages, namely–

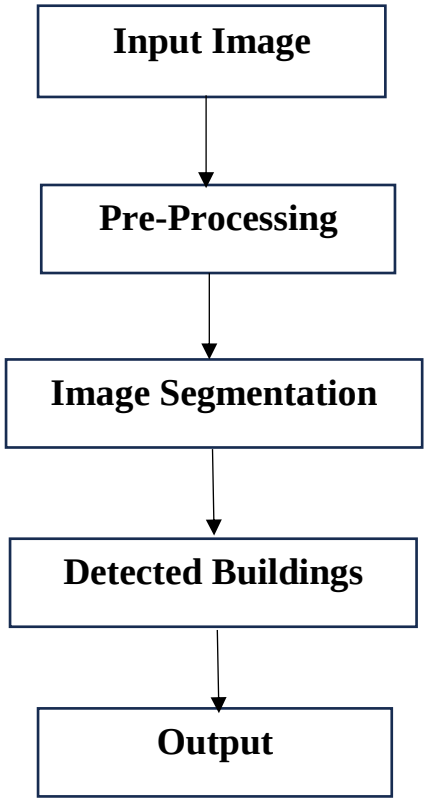

Fig 1: Methodology

CHAPTER 3

Work Done

3.1 Input Image:

The aerial images or satellite images which is most probably a high-resolution image is obtained from the satellites. There exist datasets containing various sets of aerial images taken from different places.

The image we have in this part focuses on an area featuring buildings. The number, size, and density of buildings, also defines the complexity of the surrounding environment (urban, rural, etc.) will shape the difficulty of the task.

The input images are obtained from the dataset and this input images are used for the next pre-processing stage which we will discuss briefly in the next sub-chapter.

The brief information about the dataset is described below.

3.1.1 Dataset:

Massachusetts building dataset is used to implement this project i.e. building segmentation from aerial images.

A dataset of images is the important part for training and estimation of quality for different machine learning algorithms. Now there exist some available databases of aerial photos.

The Massachusetts Buildings Dataset provides a detailed look at urban and suburban areas around Boston. It contains 151 high-resolution aerial images (1500x1500 pixels), each covering 2.25 square kilometres, for a total area of roughly 340 square kilometres. The dataset includes building footprint masks derived from OpenStreetMap data, making it ideal for building segmentation tasks.

It's divided into training, testing and validation sets. The training set consists of 137 images, testing set consists of 10 images and validation set consists of 4 images.

Examples of images from the Massachusetts database are shown in Fig. 2.

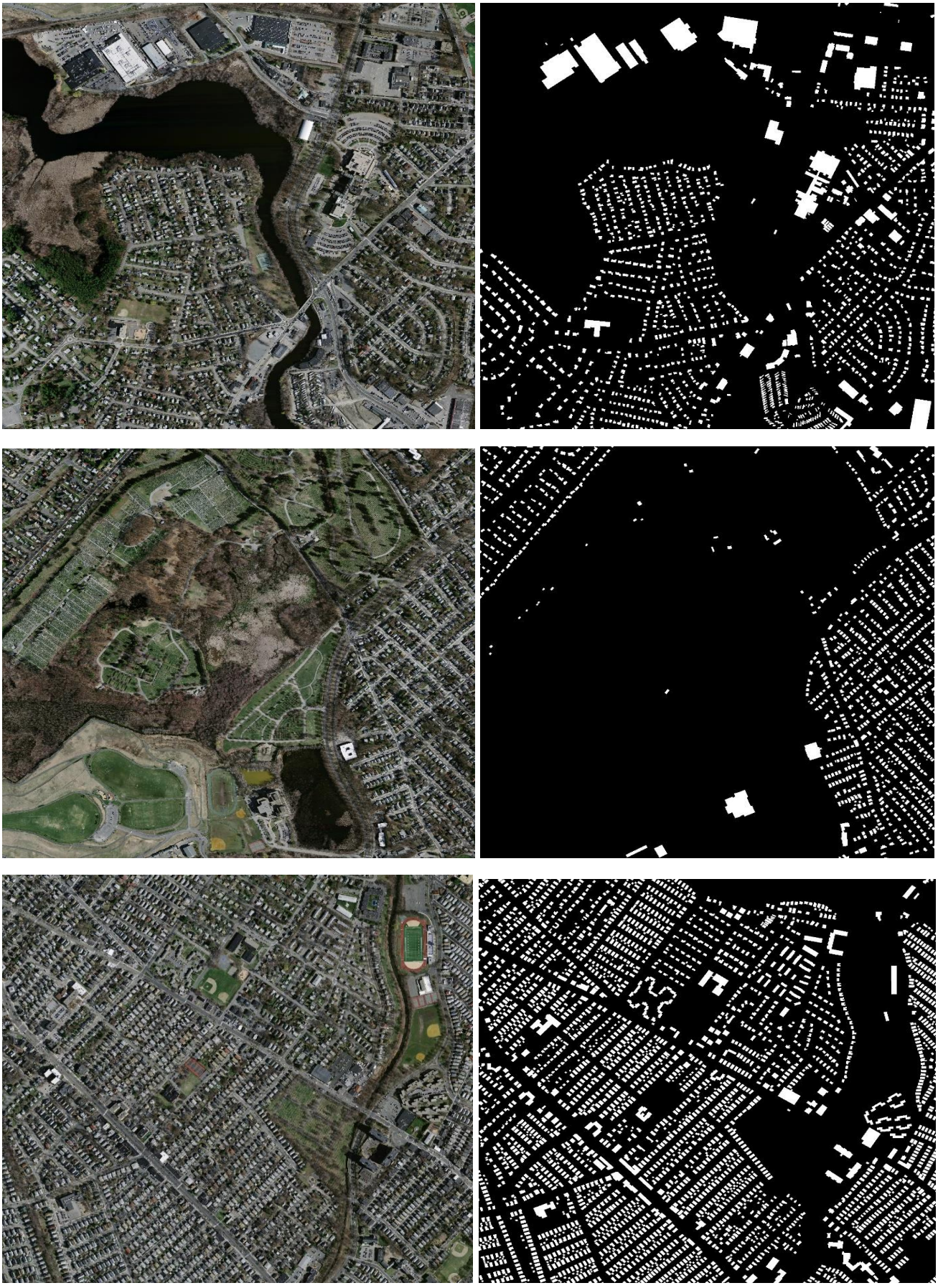

Fig 2: Massachusetts building dataset examples

3.2 Pre-Processing

The pre-processing stage aims to enhance image quality while preserving crucial building edges for later detection.

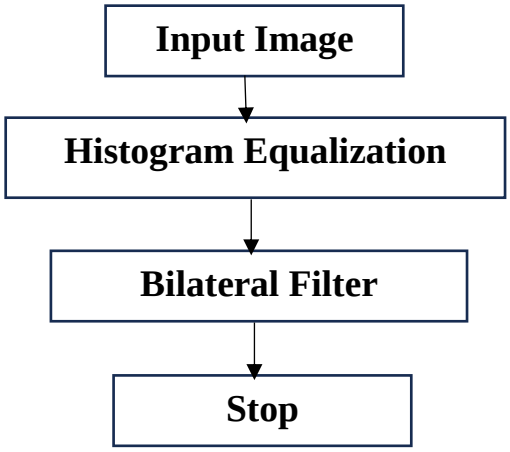

Fig 3: Pre-Processing Stages

3.2.1. Noise Reduction and Edge Preservation (Bilateral Filter):

Aerial images can suffer from noise (grainy or speckled appearance) due to sensor imperfections or atmospheric conditions. Noise obscures detail and impedes building detection. However, traditional smoothing filters often blur edges along with noise.

- Bilateral Filter, Unlike standard filters that average pixels purely based on spatial closeness, a bilateral filter considers both spatial proximity and intensity similarity. Pixels that are nearby and share a similar color/brightness receive greater weight in the averaging process.
- The filter effectively smooths homogeneous regions (e.g., a roof surface) while avoiding blurring across sharp transitions (e.g., the building's edge against the sky). This is crucial for later algorithms that rely on clear boundaries for building detection.

3.2.2. Contrast Enhancement (Histogram Equalization):

Images with poor contrast (weak differentiation between the darkest and brightest regions) make identifying buildings difficult. Buildings might blend into their backgrounds, hindering accurate detection.

- Histogram Equalization technique redistributes pixel intensities across the available brightness spectrum. Essentially, it stretches out the contrast, making bright areas brighter, dark areas darker, and consequently revealing

previously subtle details.

- Higher contrast generally aids segmentation algorithms. Buildings and their boundaries will visually "pop" against their surroundings more prominently due to this contrast stretching.

In the below images Fig 3 and Fig 4, we can identify the change in color from high resolution RGB to black and white image.

This process is mainly carried by the histogram equalization technique.

Fig 4: Before applying Bilateral filter and Histogram Equalization

Fig 5: After applying Bilateral Filter and Histogram Equalization

3.3 Image Segmentation and Detecting Buildings

Image segmentation is the process of partitioning a digital image into meaningful regions or segments. Each segment corresponds to an object, a specific area, or parts with similar characteristics.

This segmentation process is done by identifying regions within an image that share common characteristics, such as color, texture, intensity and other features. Pixels within each segment are grouped together based on these similarities, ultimately simplifying the image and making it more amenable to analysis.

Image segmentation plays a pivotal role in numerous applications.

- Building Segmentation - Building Detection and Extraction, Roof Segmentation, Change detection, urban planning, reconstruction
- Medical imaging – Used for delineation of anatomical structures, tumors, or abnormalities, aiding diagnosis and treatment planning
- Remote Sensing and geospatial analysis - classification of land cover, urban areas, and other environmental features
- Autonomous Vehicles - self-driving cars to understand their environment, separating roads, pedestrians, vehicles, and traffic signs.

There are numerous applications of image segmentation but here we just focus on building segmentation application.

There are various methods to do image segmentation by both Deep learning methods and Non Deep-learning methods, some are listed below:

Convolutional Neural Networks (CNNs): Powerful models that learn complex patterns from data; widely used for image segmentation. Key architectures include:

- **U-Net**: The main characteristic of U-Net is about it's U-shaped architecture. It has two paths mainly – encoder path and decoder path

  The encoder path processes the input image, capturing its features and progressively reducing its spatial resolution. It typically uses repeated stacks of convolutional layers with pooling operations to achieve this

  The decoder path up-samples the extracted features and combines them with high-resolution information from the contracting path. This allows for precise localization of the segmented objects. It often uses up-convolutional layers and concatenation operations to achieve this.
  The architecture of U-Net describing encoder and decoder part is as shown in the below fig.6.

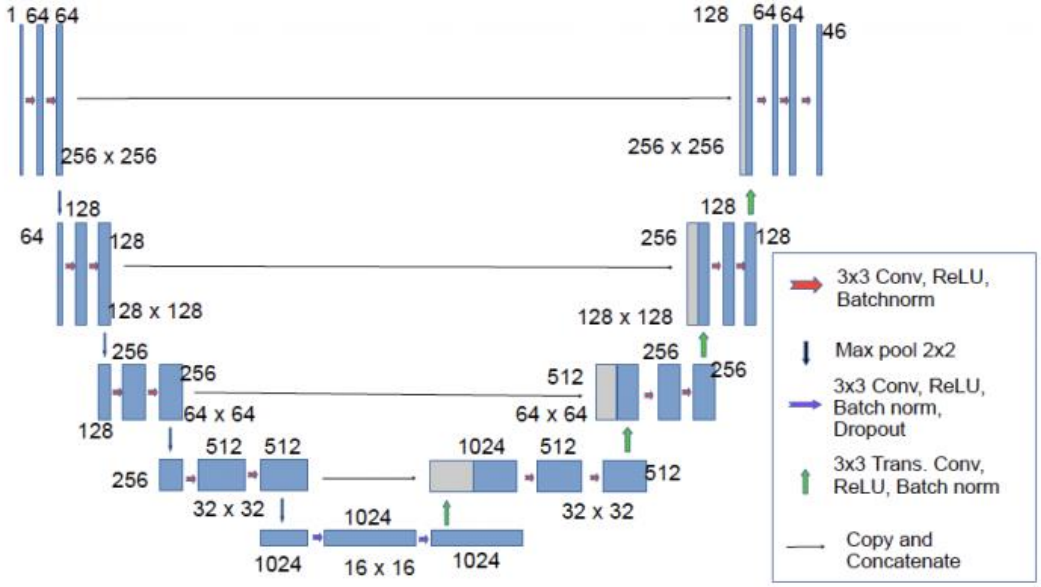

Fig. 6: U-Net Architecture

- **K-means Clustering:** K-means clustering is a fundamental unsupervised machine learning algorithm used to partition data points into a predefined number of clusters (groups). It's a simple yet effective technique for grouping similar data points together based on their features.

- **Otsu Thresholding Method:** Otsu's method aims to find an optimal threshold value that divides an image's pixels into two classes (foreground and background) by minimizing the within-class variance or equivalently maximizing the between-class variance. Think of it as finding the ideal cut-off point on the image's brightness histogram.

- **Chan-Vese Method:** The Chan-Vese method is an active contour model that effectively segments images with well-defined or blurry boundaries into distinct regions. It's based on the idea of energy minimization, seeking to find a contour (a curve) that best separates these regions.

  From above definitions, we know that U-Net is a Deep learning method whereas, K-means clustering, Osu Thresholding, Chan-Vese Method are Non Deep learning method. Thus, we see the comparisons of all the three above methods and find out which model segments the buildings better.

  Firstly, Let’s see the results obtained by Non Deep learning Methods, i.e. Chan-Vese Methos and Otsu Thresholding method.

3.3.1 Chan-Vese Method:

The following process was undergone to obtain the results as shown in Fig.7 and Fig.8 by Chan-Vese method for building segmentation.

Firstly, the dataset is loaded from the drive, then the color image is converted to Black and white image.
The energy function is calculated with the help of curve lengths, image gradients.
The contours are formed iteratively through the method called level set evolution.
Finally, the contours represent the segmentation boundary, highlighting the object of interest.

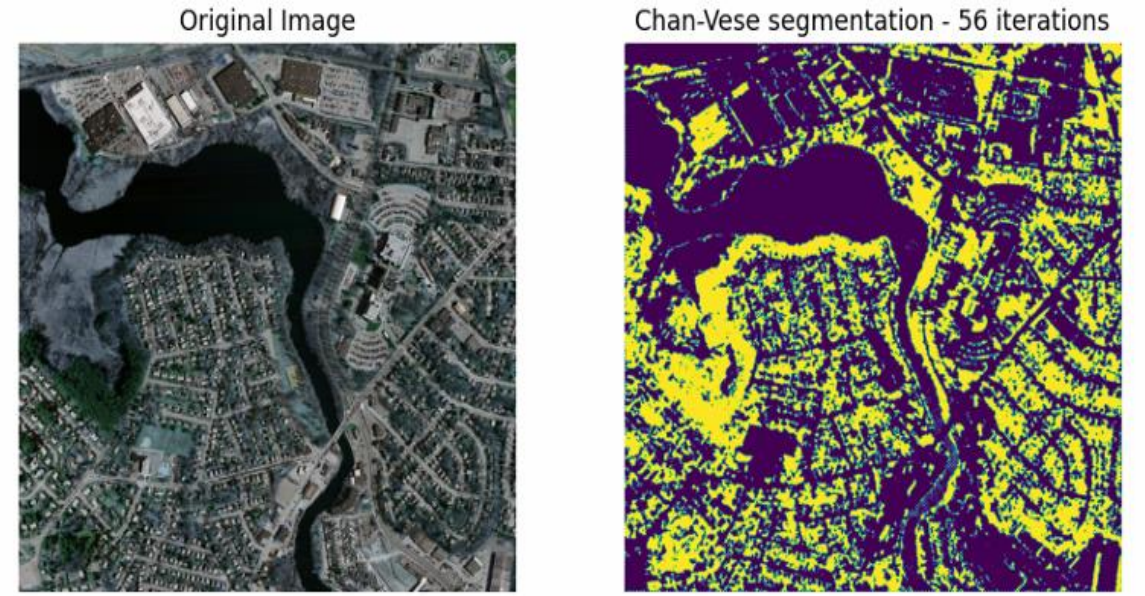

Fig.7. Chan-Vese Segmentation Output

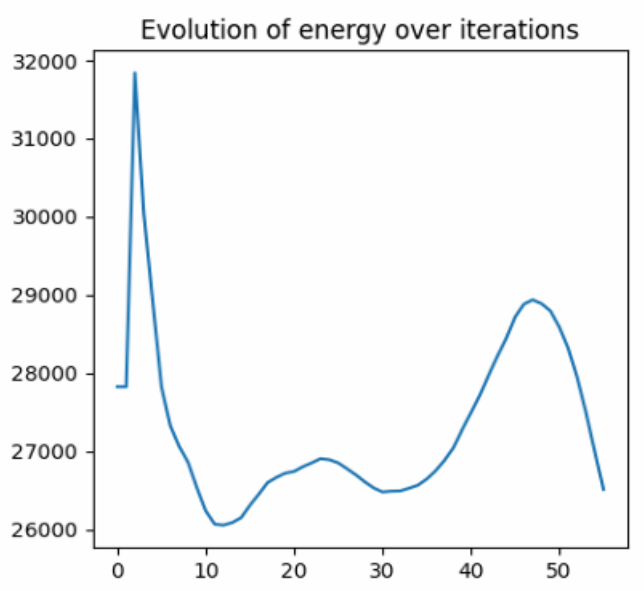

Fig. 8. Energy versus iterations in Chan-Vese segmentation

The accuracy obtained by the prediction of output i.e segmentation of buildings in Chan-Vese Segmentation is observed to be 53.10% at maximum.

3.3.2 Otsu Thresholding:

The following process was undergone to obtain the results as shown in Fig.9. by Otsu Thresholding method for building segmentation.

Firstly, the histogram of pixel intensity for the grayscale image is calculated.
Then the iteration method tests all the possible values that could separate the foreground and background
The variance is calculated
Finally, the pixels with intensities below the threshold are classifies as foreground and the pixels above are classifies as background

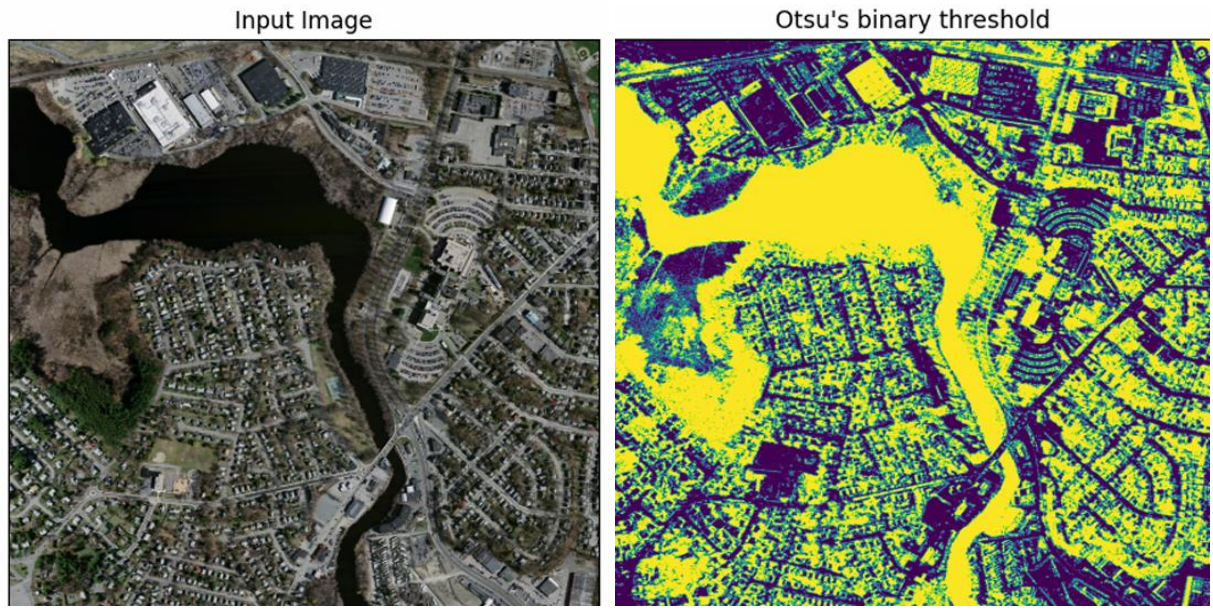

Fig.9. Otsu's Binary Threshold Output

The accuracy obtained by the prediction of output i.e segmentation of buildings in Otsu's Binary threshold output is observed to be 58.10% at maximum.

Now, coming to the deep learning method.

Let's see the performance of U-Net segmentation model

3.3.3 U-Net

Firstly, the aerial image is given as input for the u-net, then a series of convolution layers extract the important features required.

The pooling layers reduces the image dimensions, i.e. capturing the large-scale information into the required dimensions. The base of the U-net architecture is highly compressed with essential features.

The decoder parthelps recover the original image resolution. Finally, the output image will be in the same size as the input image and each pixel is classified as either “building” or “not building”. The building part is highlighted by a specific color and the non-building part is omitted.
The output of the U-net segmentation can be shown in the below Fig.10.

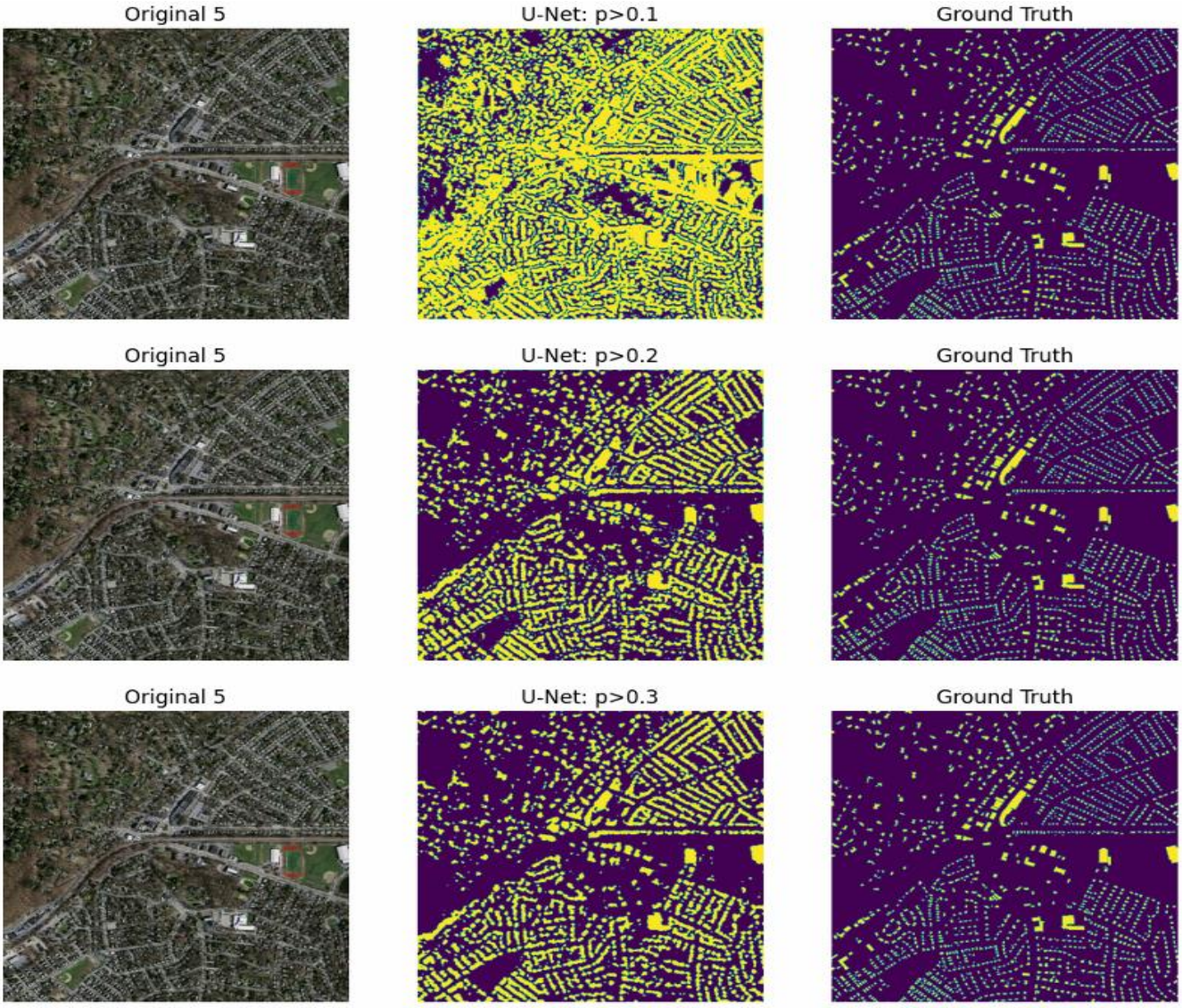

Fig.10: U-Net segmentation output - 1

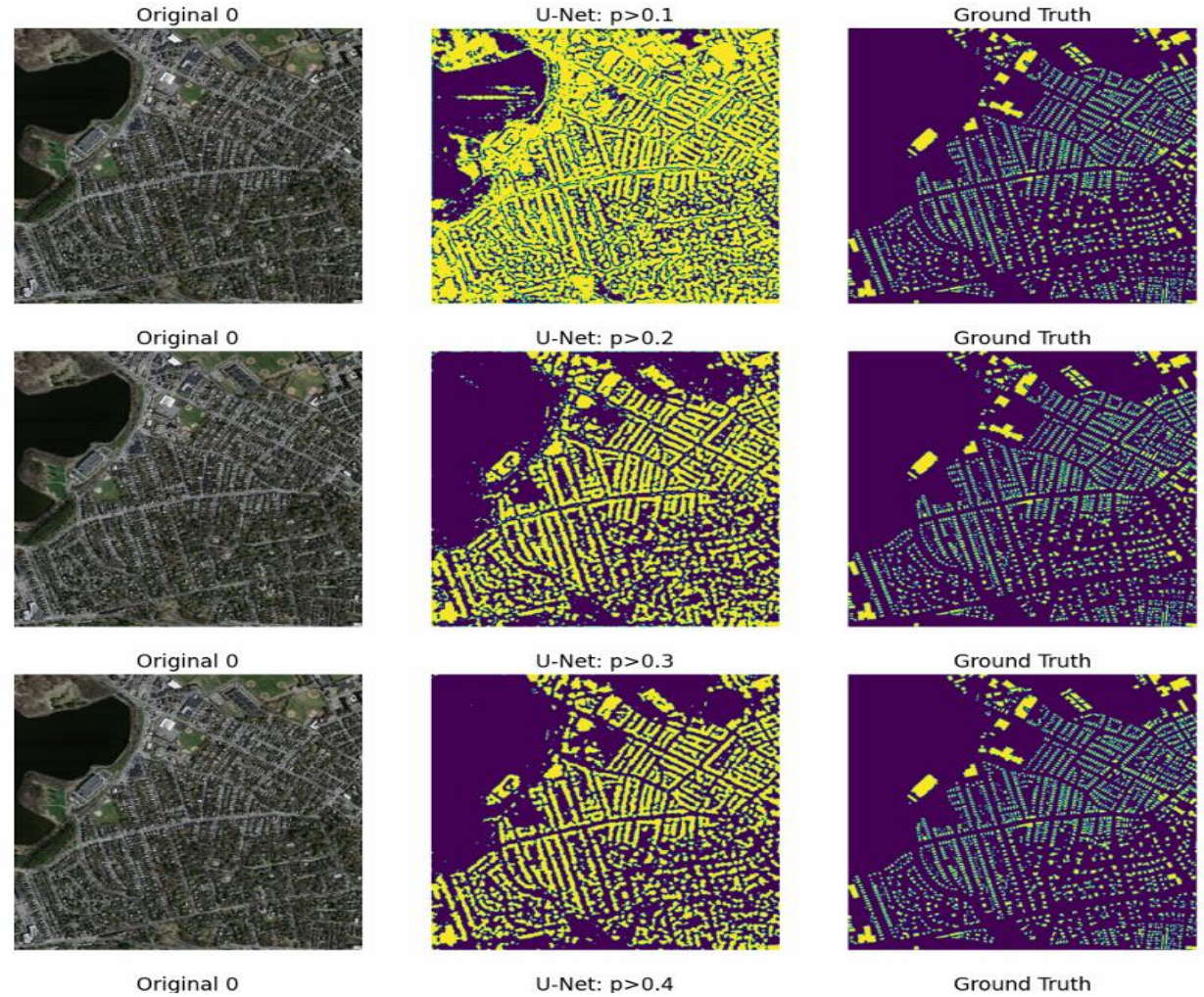

Fig.11: U-Net Segmentation Output – 2

The training Accuracy was also calculated and represented graphically as shown in Fig.12.

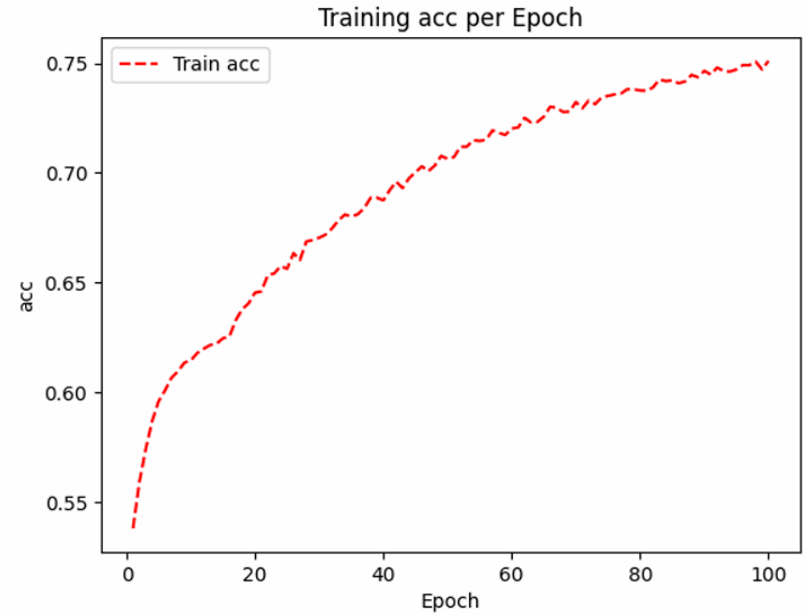

Fig.12. Training accuracy per Epoch

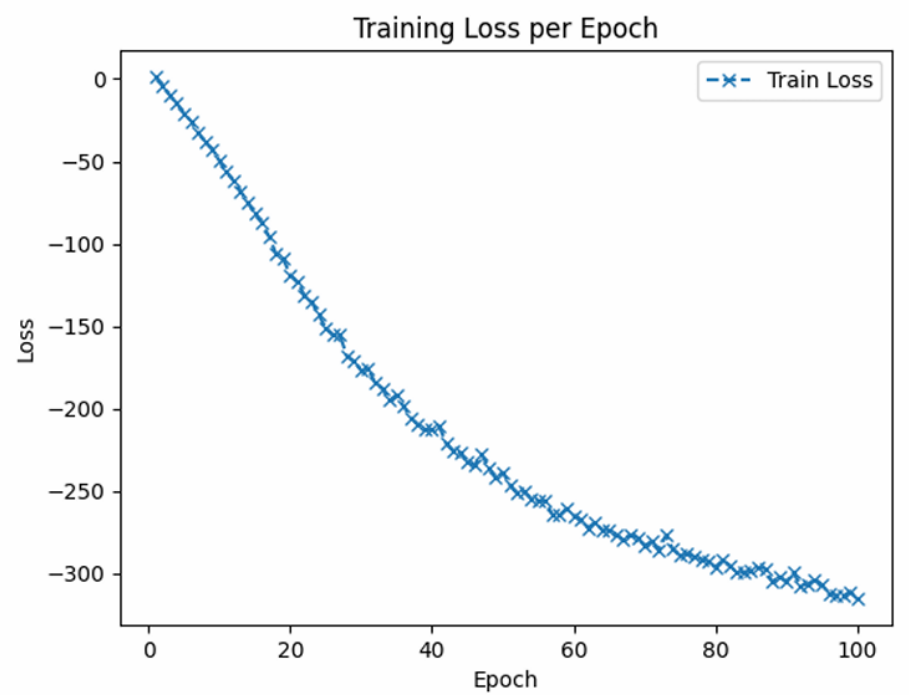

Fig.13. Training Loss per Epoch

The accuracy obtained by the prediction of output i.e segmentation of buildings in U-Net is observed to be 75.10% at maximum.

3.3.4 SegNet

SegNet is another famous convolution neural network which is used most commonly is segmentation tasks. It is similar to the U-Net architecture but have some distinctions from it. SegNet assigns every single pixel in an image to a class label, therefore this turns the image into a detailed segmentation map.

Like U-Net, SegNet also have encoder and decoder layers. The encoder layer consists of 13 convolution layers like VGG16, this extract features and reduces image resolution. The decoder layer maps low-resolution feature maps back to the original input resolution, producing a dense segmentation prediction.

SegNet's unique aspect is its upsampling method in the decoder. Instead of learned upsampling layers (as in U-Net), SegNet reuses the pooling indices from the max-pooling operations in the encoder to perform a more efficient upsampling. This preserves some high-frequency details and helps with localization accuracy.

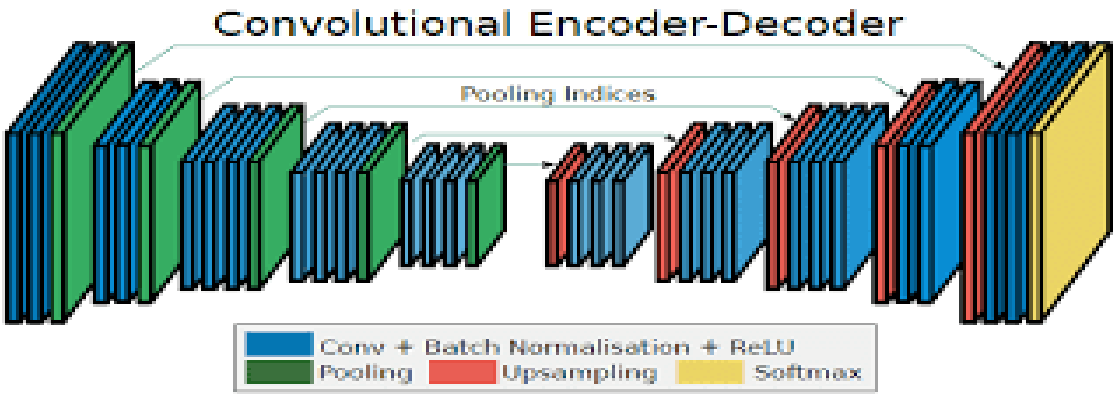

Fig 14: SegNet Architecture

The Fig. 14. represents the architecture of SegNet with all its encoder and decoder parts

The results obtained by SegNet Model is as shown in below figures

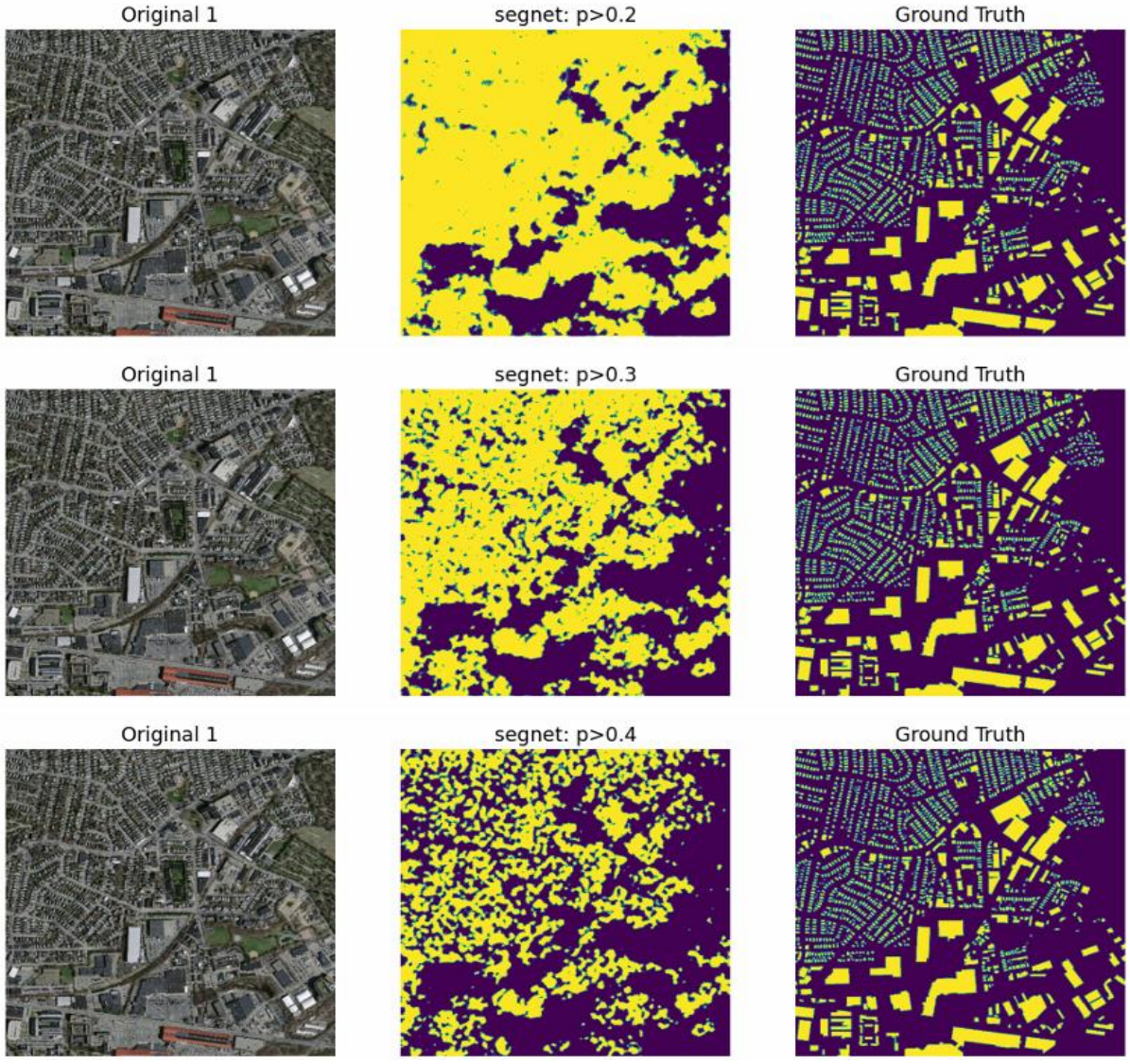

Fig. 15: SegNet Segmentation Output – 1

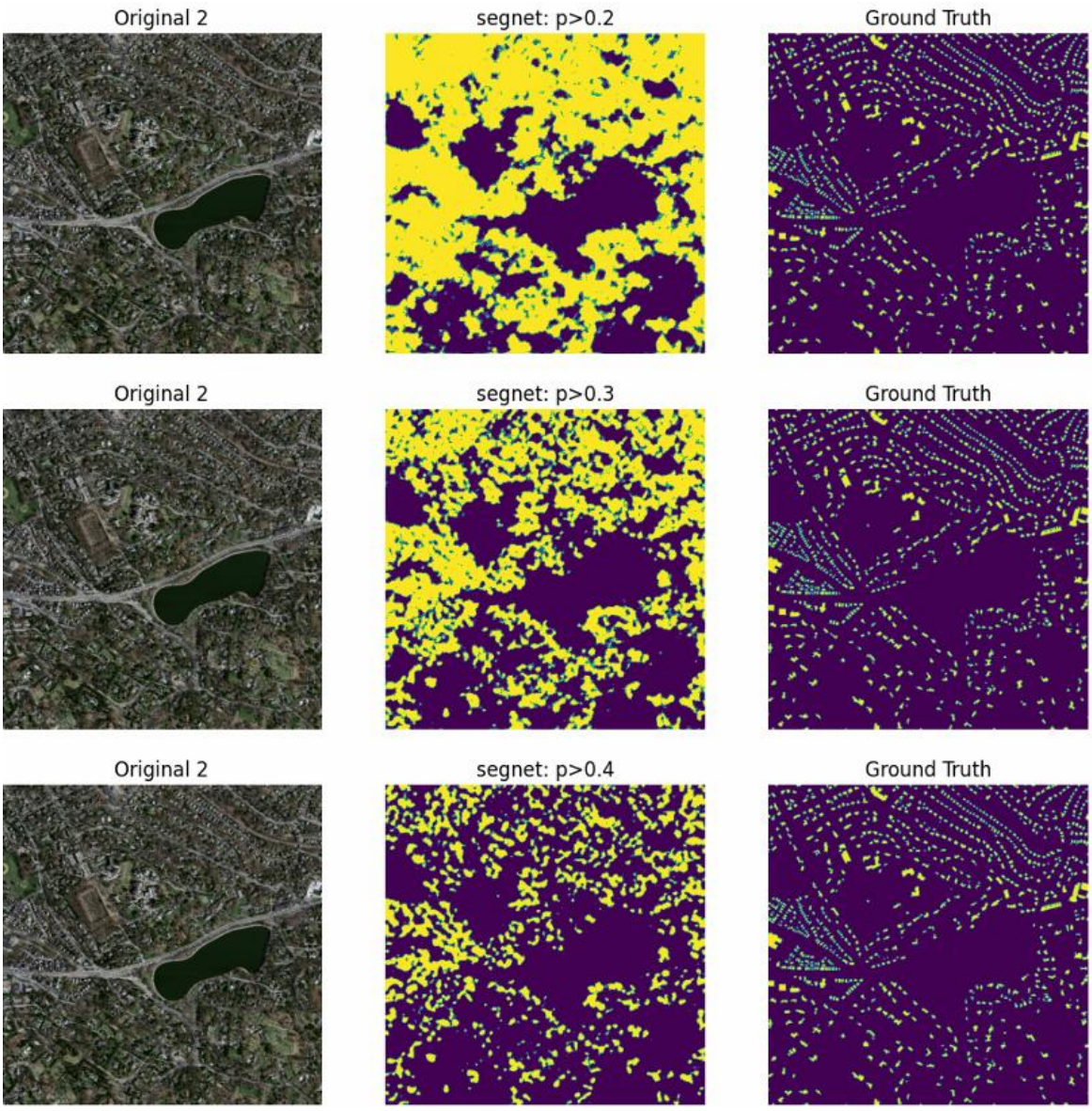

Fig. 16: SegNet Segmentation Output – 2

The training Accuracy was also calculated and represented graphically as shown in Fig.17 and Fig. 18

The accuracy obtained by the prediction of output i.e segmentation of buildings in U-Net is observed to be 71.36% at maximum.

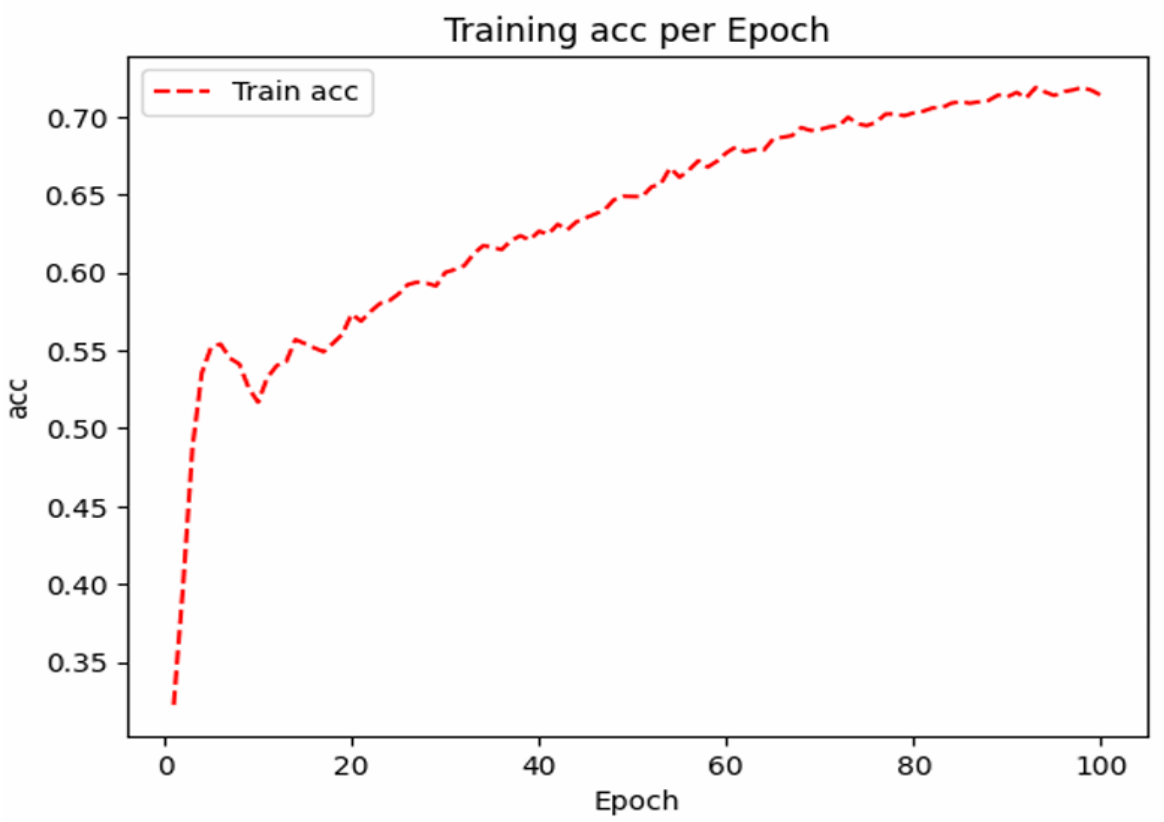

Fig.17: Training Accuracy of SegNet

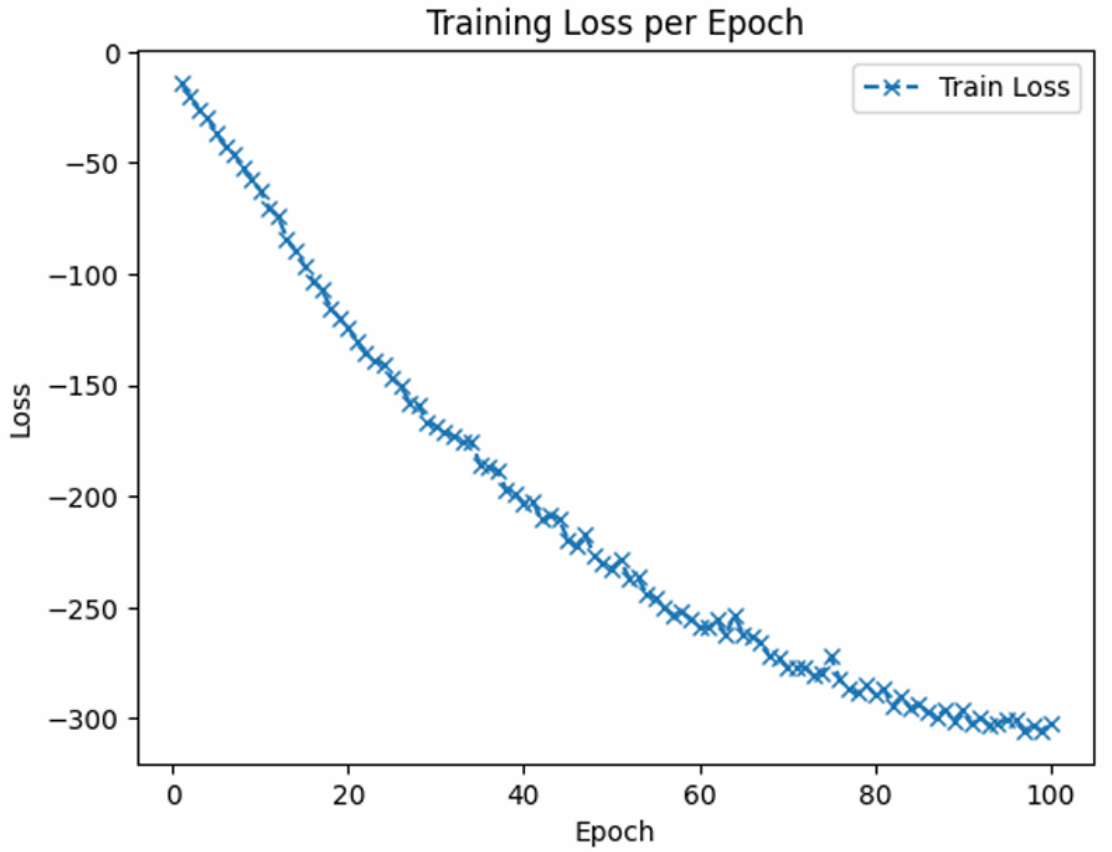

Fig. 18: Training Loss of SegNet

CHAPTER – 5

CONCLUSION

The comparison between both Non Deep-learning method and Deep Learning method was implemented and represented as shown in the above chapters.

From above observations of all three methods, we can conclude that the accuracy of the Deep Learning method i.e U-Net and SegNet was found to be much higher at 75.10% and 71% than the Non Deep-Learning Methods i.e Chan-vese and otsu thresholding method at 58% and 53% respectively.

Thus, using a deep learning model (like U-net and SegNet) for the semantic segmentation of buildings from aerial data performs with highest accuracy and can be implemented for various application.

Thus, Deep Learning Algorithms for Building Segmentation from Aerial Images is successfully implemented.

CHAPTER 6

REFERENCES

1. R. E., R. Hebbar and S. P., "Buildings Detection from Very High Resolution Satellite Images Using Segmentation and Morphological Operations," 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, India, 2018, pp. 106-110, doi: 10.1109/ICDI3C.2018.

2. Z. -q. WANG, W. -w. GONG, Y. -f. JIAO, Q. -l. WU and W. -y. LI, "Fully Convolutional Network for Recognition of Small Buildings in Aerial Images," 2018 Fifth International Workshop on Earth Observation and Remote Sensing Applications (EORSA), Xi'an, China, 2018, pp. 1-5, doi: 10.1109/EORSA.2018.8598558.