MAULANA AZAD
NATIONAL INSTITUTE OF TECHNOLOGY
BHOPAL INDIA, 462003

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Semantic Image Segmentation

Minor Project Report
Semester VI

Submitted by:

Aditya Kumar	181112251
Pradhi Anil Kumar Das	181112252
Pushpak Manwani	181112286
Aaron Paul	181112287

Under the Guidance of
Dr. Dehalwar Vasudeo
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MAULANA AZAD
NATIONAL INSTITUTE OF TECHNOLOGY
BHOPAL INDIA, 462003

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that the project report carried out on
"Image Segmentation" by the 3rd year students:

Aditya Kumar	181112251
Pradhi Anil Kumar Das	181112252
Pushpak Manwani	181112286
Aaron Paul	181112287

Have successfully completed their project in partial fulfillment of their Degree in Bachelor of Technology in Computer Science and Engineering.

Dr. Dehalwar Vasudeo
(Minor Project Mentor)

DECLARATION

We, hereby declare that the following report which is being presented in the Minor Project Documentation Entitled as **"Semantic Image Segmentation"** is an authentic documentation of our own original work and to the best of our knowledge. The following project and its report, in part or whole, has not been presented or submitted by us for any purpose in any other institute or organization. Any contribution made to the research by others, with whom we have worked at Maulana Azad National Institute of Technology, Bhopal or elsewhere, is explicitly acknowledged in the report.

Aditya Kumar	181112251
Pradhi Anil Kumar Das	181112252
Pushpak Manwani	181112286
Aaron Paul	181112287

ACKNOWLEDGEMENT

With due respect, we express our deep sense of gratitude to our respected guide and coordinator **Dr. Dehalwar Vasudeo** , for his valuable help and guidance. We are thankful for the encouragement that he has given us in completing this project successfully.

It is imperative for us to mention the fact that the report of the minor project could not have been accomplished without the periodic suggestions and advice of our project guide **Dr. Dehalwar Vasudeo** and project coordinators Dr. Dhirendra Pratap Singh and Dr. Jaytrilok Choudhary.

We are also grateful to our respected director Dr. N. S. Raghuwanshi for permitting us to utilize all the necessary facilities of the college.

We are also thankful to all the other faculty, staff members and laboratory attendants of our department for their kind cooperation and help. Last but certainly not the least; we would like to express our deep appreciation towards our family members and batch mates for providing the much needed support and encouragement.

ABSTRACT

Image Segmentation is the process by which a digital image is partitioned into various subgroups (of pixels) called Image Objects. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze.Image segmentation is typically used to locate objects and boundaries in images.

Some Applications of Image Segmentation:

- Improved Quality of MRI/Medical images for better detection of diseases and problems
- Segmenting images can help to improve robot vision
- Image segmentation can also be applied to satellite images in order to get better object detection
- Self-driving vehicle autonomous cars must be able to perceive and understand their environment in order to drive safely. Relevant classes of objects include other vehicles, buildings, and pedestrians. Semantic segmentation enables self-driving cars to recognize which areas in an image are safe to drive.

- Iris recognition is a form of biometric identification that recognizes the complex patterns of an iris. It uses automated pattern recognition to analyze video images of a person’s eye.

- Face recognition identifies an individual in a frame from a video source. This technology compares selected facial features from an input image with faces in a database.

As well as many other applications are also possible to be implemented with the help of image segmentation.Our group will try to implement some of these applications of image segmentation.

TABLE OF CONTENTS

1.Introduction

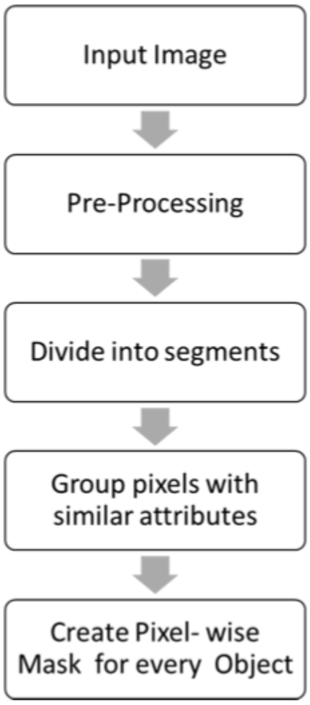
In digital image processing and computer vision, image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as image objects). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics.

The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image (see edge detection). Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s).When applied to a stack of images, typical in medical imaging, the resulting contours after image segmentation can be used to create 3D reconstructions with the help of interpolation algorithms like marching cubes.

Basically segmentation is not just a simple procedure, it includes and executes multiple chores. The images that are laid open to the process of segmentation are segregated into manifold divisions, and the valuable insights would be obtained as a result either as texture, intensity or color. In addition from this the segmentation takes an important part in recognizing the images.

Several techniques are available and more new techniques are devised to perform the procedures of segmentation. But the available and the devised methodologies prove to be less efficient as they are practically incompatible over all sorts of images. So there arises a necessity to identify an effective segmentation procedure, to enhance the performance of the image processing.

The segmentation procedure scopes in simplifying the examination process into eloquent and easily analyzable form. It is vital and the initial procedure in analyzing an image. The main purpose of the segmentation process is to segregate the images to have the similar attributes. It is an essential first step in many of the application like object detection and recognition, retrieval of image based on the content, video surveillance, medical imaging etc. Apart from this the segmentation is the first and the foremost step in developing a multi label classifier construction to classify more than one object found in the image. For this it works by segregating the image into multiple sections and groups the same attributes pixels and creates a pixel-wise mask for all the objects in the image. The figure below shows the stages in the **segmentation process**.



Basically the segmentation process could be classified into two fundamental kinds as local and global segmentation and further based on the image properties as discontinuity detection strategy and similarity detection strategy.The former types of segmentations mentioned are concerned about the particular section or the area of an image and whole image that is encompassed with the huge set of pixels. The later segmentation types based on the image properties segments the images into regions based on the discontinuity and similarity respectively. The several segmentation methods of the image are listed below in figure.2

There are more than a few prevailing practices that are utilized for segmenting images. These methods have their own significance. These modus operandi could come within reach of from dualistic elementary methods of segmentation i.e. region based or the edge based, every single procedures can be applied on dissimilar images to accomplish the necessary segmentation. All these techniques could be classified based on three categories. As (i) structural: works on the structure of the region, (ii) Stochastic: works on the discrete pixel values and (iii) hybrid: Combines both the structural and the stochastic.

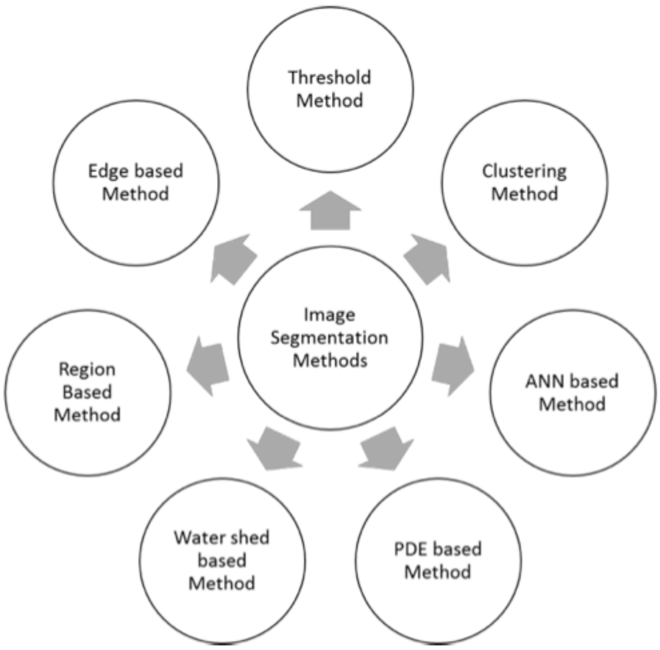


Figure. Image Segmentation Methods

As the digital images acquired are projected to multitudes of distortions that deteriorates the quality of the image, an effective segmentation process becomes a necessary requisite for retaining the valuable data of the images especially in medical imaging.

Some Applications of Image Segmentation:

- Improved Quality of MRI/Medical images for better detection of diseases and problems
- Segmenting images can help to improve robot vision
- Image segmentation can also be applied to satellite images in order to get better object detection
- Self-driving vehicle autonomous cars must be able to perceive and understand their environment in order to drive safely. Relevant classes of objects include other vehicles, buildings, and pedestrians. Semantic segmentation enables self-driving cars to recognize which areas in an image are safe to drive.

- Iris recognition a form of biometric identification that recognizes the complex patterns of an iris. It uses automated pattern recognition to analyze video images of a person's eye.

- Face recognition identifies an individual in a frame from a video source. This technology compares selected facial features from an input image with faces in a database.

As well as many other application are also possible to be implemented with the help of image segmentation.Our group will try to implement some of these applications of image segmentation.

2.Literature Review and Survey

There have been many works done in the area of image segmentation by using different methods. And many are done based on different application of image segmentation.K-means algorithm is the one of the simplest clustering algorithm and there are many methods implemented so far with different method to initialize the centre. And many researchers are also trying to produce new methods which are more efficient than the existing methods, and shows better segmented result.

Some of the existing recent works are discussed here.

Pallavi Purohit and Ritesh Joshi 4 introduced a new efficient approach towardsK-means clustering algorithm. They proposed a new method for generating the cluster center by reducing the mean square error of the final cluster without large increment in the execution time. It reduced the means square error without sacrificing the execution time. Many comparisons have been done and it can conclude that accuracy is more for dense dataset rather than sparse dataset.

Madhu Yedla, Srinivasa Rao Pathakota, T. M. Srinivasa 6 proposed EnhancingK-means clustering algorithm with improved initial center. A new method for finding the initial centroid is introduced and it provides an effective way of assigning the data points to suitable clusters with reduced time complexity. They proved their proposed algorithm has more accuracy with less computational time comparatively original k-means clustering algorithm. This algorithm does not require any additional input like threshold value. But this algorithm still initializes the number of cluster k and suggested determination of value of k as one of the future work.

K.A. Abdul Nazeer, M. P. Sebastian 7 proposed an enhanced algorithm to improve the accuracy and efficiency ofthe k-means clustering algorithm. They present an enhanced k-means algorithm which combines a systematic method consisting two approaches. First one is finding the initial centroid and another is assigning the data point to the clusters.They have taken different initial centroid and tested execution time and accuracy. From the result it can be conclude that the proposed algorithm reduced the time complexity without sacrificing the accuracy of clusters.

Koresh, M. H et al [9] presented the "Computer vision based traffic sign sensing for smart transport." Chandy, Abraham et al [10] elaborated the "A Review on Iot Based Medical Imaging Technology for Healthcare Applications." Bindhu, et al [11] presents the "Biomedical Image Analysis Using Semantic Segmentation." V Kumar, T. Senthil et al [12] proposed the "A Novel Method for HDR Video Encoding, Compression and Quality Evaluation."

3. Proposed work and Methodology

In this section we will mention how each of our application works and the principle behind it.

3.1 Generic Object Segmentation

This folder contains a python file that will convert the given images into segmented images with the help of K-Means Clustering Algorithm

The K-means algorithm is an iterative technique that is used to partition an image into K clusters.The basic algorithm is

1.Pick K cluster centers, either randomly or based on some heuristic method.
2.Assign each pixel in the image to the cluster that minimizes the distance between the pixel and the cluster center
3.Re-compute the cluster centers by averaging all of the pixels in the cluster
4.Repeat steps 2 and 3 until convergence is attained (i.e. no pixels change clusters)

In this case, distance is the squared or absolute difference between a pixel and a cluster center. The difference is typically based on pixel color, intensity, texture, and location, or a weighted combination of these factors. K can be selected manually, randomly, or by a heuristic. This algorithm is guaranteed to converge, but it may not return the optimal solution. The quality of the solution depends on the initial set of clusters and the value of K.

3.2 Medical Image Segmentation

The main principle involved in this project is that the images are segmented with K-Nearest Neighbour algorithm with N=3 as the most common colors in a MRI scan are various shades of black ,white and grey and with just 3 colors the images of the scan will be segmented properly which will provide a lot of useful information to the doctor,such as the place where the tumor is and the place where the mass is most dense etc.

Main Algorithm working:
1.Choosing the number of Clusters(n) and using k-means to make the classes

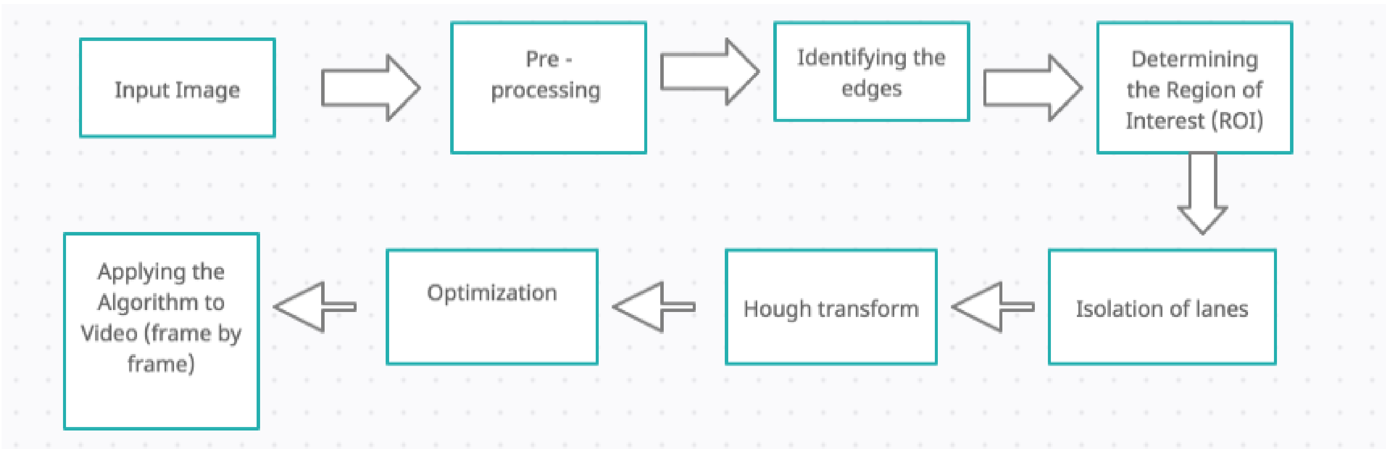
2.kmeans centers points for centroid.

3.Assigning each Data point or as we say each pixel value to the above centroid using KNN that further gives us clusters.

4.In our case 41 was passed as the number of neighbors(Odd number to break tie)

5.On the last step we just do the concatenation of original and segmented image

3.3 Lane Detection Program



A) Pre - processing of the Image : This includes converting the input image frame (RGB) to gray scale for reducing the channels. Here we also remove Image noise for better and accurate edge detection. Then we smoothen the Image.

B) Identification of edges: We use one of the edge detection techniques called the Canny Technique, which will help us identify the edges in the image.

C) Determining the region of Interest (ROI) : Before we start detection, we need to decide the particular lane area in the image that we should work-on to write the program. For this we create a ROI function which returns the enclosed region of our field of view, in the case of a road, triangle region would be best fit. We do so by creating a polygon and applying it on a black mask.

D) Isolation of lanes : Now, we will need to isolate the lanes in the masked portion of the image. We use binaries to isolate. We apply Bitwise AND between the existing image and the gradient image. Doing so will give us perfectly masked out lanes on the road.

E) Hough Transform : Hough Transform will help us to detect the straight lines in the image and help identifying Lane lines. Hough space is the representation of the **b,m** axis in the Line equation **y=mx+b.** We use the Hough space to find the series of points.

Understanding:-

- First, we split our Hough space in to a grid
- Each **Bin** inside our grid corresponding to the slope and y-intercept value of a particular line
- All of these points intersection in Hough space are inside of a single bin
- For every point of intersection, votes will be casted for that particular bin.
- The bin with the maximum number of votes, will be our line.
- whatever **m and b** value that **bin** belongs to that's the line which we will draw. Since, it was voted as the line of best fit in describing our data.

But the above doesn't take vertical lines into account because when we try to compute the slope of a vertical line the change in x is zero, which ultimately will always evaluate to a slope of infinity which is not something that we can represent.

Therefore we use a more robust representation of lines, so that we don't encounter any numeric problems. We can express it in the **polar** coordinates system **rho** and **theta**, such that our line equation can be

$$\rho = xcos\theta + ysin\theta$$

F) Optimization : We see that the resulting image contains multiple lines having breakpoints in between certain lanes, We can fix this by taking average of their slope and the y intercept to form a single line that will trace a whole lane, thereby optimizing the preview.

G) Applying the Algorithm to a video frame by frame : We will use opencv's VideoCapture function, which will allow us to process the video frame by frame by reading every frame and then applying the algorithm to that particular frame.

Here we also combine the current frame with the line image by giving them particular weights. We then display the combined images frame by frame.

3.4 Road Segmentation using Satellite Images

Get_Data_Generators.py:Builds and returns data generators.

acquire_data.py:Downloads the Massachusetts Roads Dataset or the Massachusetts Buildings Dataset. By changing "link_file" to point at a custom list of links, you can download any other dataset too,but we worked with this dataset.

build_dataset.py; Crops the images into small 256x256 images and divides the data set into training and testing set.

Road_detection_GPU.ipynb:define our model and make a custom loss function and finally testing our model and getting the output.Finally our model is made and it can be used similarly with other data sets.So Detection of road is possible with the help of satellite imagery with the help of our model.

Dataset

For this project, we will be using the Massachusetts Roads Dataset. This dataset contains 1171 aerial images, along with their respective maps. They are 1500 x 1500 in dimension and are in .tiff format.

Preprocessing

1. Removed images where more than 25% of the map was missing.
2. **Cropping instead of resizing**: Training our model on large images is not only resource-intensive but is bound to take a lot of time as well. Resizing images to lower dimensions can be an answer, but resizing comes at a cost. Regardless of the interpolation method we choose while resizing, we end up losing information. Therefore, we will crop out smaller, 256 x 256 images from the large images. Doing so leaves us with about 22,000 useful images and maps
3. **Thresholding and binarizing the maps:** Binarized the mask so that the pixel value is always between 0 & 1.

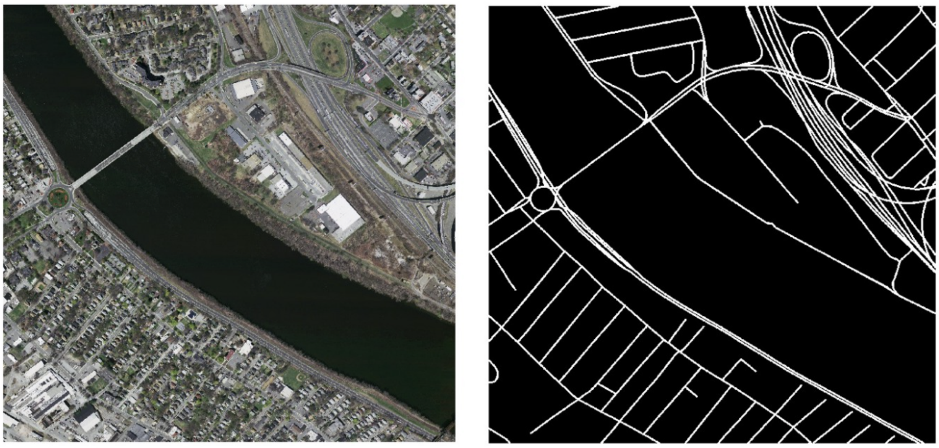


Image Mask

Neural Modelling

Now that we have dealt with the data, its time we start modelling our Neural network. To accomplish our segmentation task, we will be using a Fully Convolutional Network. These kinds of networks are mostly composed of convolutional layers, and unlike the more traditional neural networks, fully connected layers are absent.

i. About F.C.N

Fully Convolutional Network was developed for biomedical image segmentation at the Computer Science Department of the University of Freiburg, Germany [1]. It was later realised that the scope of these networks is well beyond the medical realm. These networks can perform multiclass segmentation of any kind of object — be it segmenting people, cars or even buildings.

ii. Network Architecture

This project uses U-net, a fully convolutional neural network which is quite intuitively named. This network takes a 256x256 multichannel image and outputs a single-channel map of the same dimension.

A U-net has two parts — The encoder or the downsampling section, and the decoder or the up-sampling section. Just have a look at the following image.

Encoder: It is a.k.a. the downsampling section. This segment uses convolutional layers to learn the temporal features in an image and uses the pooling layers to downsample it. This part is accountable for learning about the objects in an image. In this case, this segment learns how a road looks like and can detect it. I added dropout layers which will randomly ignore neurons to prevent overfitting, and I added BatchNormalization to ensure that each layer can learn independently of the previous one.

Decoder: It is a.k.a. the Upsampling segment. Continuous pooling operations result in the loss of spatial information of the image. The model does know about the contents of the image, but it doesn't know where it is. The whole idea behind the decoder network is to reconstruct the spatial data using the feature maps which we extracted in the previous step. We use Transposed convolutions to upsample the image. Unlike plain interpolation, Conv2DTranspose has learnable parameters.

Skip Connections: Direct connections between the layers in the encoder segment to the layers in the decoder section are called Skip connections. They are called skip connections because they bridge two layers while ignoring all the intermediate layers. Skip connections provide the spatial information to the upsampling layers and help them reconstruct the image and "put things into place" (quite literally).

Training the Model

i. Loss Function and Hyper-parameters

At a pixel level, this segmentation challenge can be considered as a binary classification problem where the model classifies whether each pixel is white(road) or black(not road). But we need a balanced dataset to facilitate proper segmentation, and since the number of black pixels in these images greatly outnumbers the white one, we have an imbalanced dataset.

There are a few different approaches to deal with the imbalanced data issue. In this challenge, we will use the Soft Dice Loss as it is based on the Dice Coefficient. Dice Coefficient is the measure of overlap between the predicted sample and the ground truth sample, and this value ranges between 0 and 1. Where 0 represents no overlap and 1 represents complete overlap.

$$\frac{2 * |X \cap Y|}{|X| + |Y|}$$

Smooth Dice Loss is simply *1 — Dice Coefficient*, this is done to create a minimizable Loss Function[2]. Please have a look at the following code for Dice Loss.

Accuracy Metric: Accuracy metrics tell us about the correctness of the generated segmentation maps. We will be using the Jaccard Index, aka Intersection over Union, to tell us how accurate the generated maps are. As the name suggests, Intersection over Union is the measure of the correctness of the segmentation maps. The numerator is the intersection between the predicted map and the ground truth label, while the denominator is the total area of both the ground truth label and segmentation map (calculated using Union operation).

4.Gaps Identified and their Solutions

4.1 Generic Object Segmentation

the gaps identified in this program are for a large value of K (number of cluster) the output image is over-segmented ,an example is given below

(i) Original Image used for Segmentation (ii) Output image that is over-segmented due to a large K value

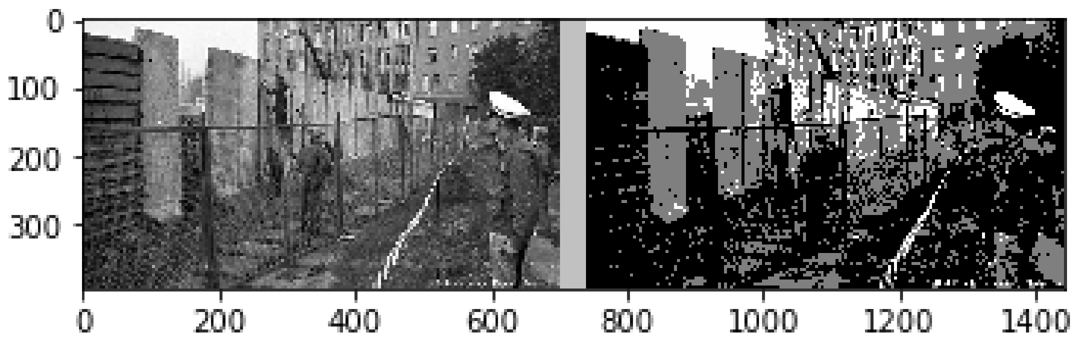
This problem can be easily corrected by changing the value of K to a lower value ,but if the value is too low also the segmentation will be not done properly.example is given below when K is too small

(i) Original Image used for Segmentation (ii) Output image that is improperly segmented due to a small K and iteration value

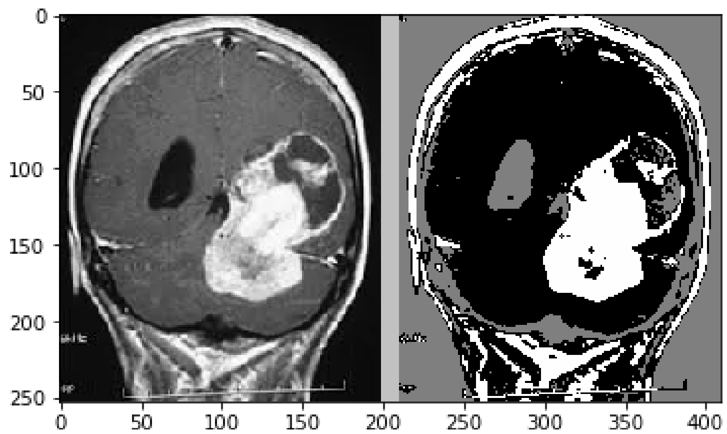
Another problem other than deciding the correct value of K and number of iteration, is that it takes a lot of time to segment a single image (approximately 30 to 40 minutes depending on the value of K and iterations) .To overcome this problem the next program uses KNN algorithm and is implemented in Jupiter notebook to directly show the results instead of waiting for the result during the presentation.

4.2 Medical Image Segmentation

This program works well when the images passed to it are MRI scans and X-rays because it was made to be implemented in this domain and passing any other image will not provide the same quality of result as it does with medical images.



This is an image of Berlin during world war 2,as you can clearly see the image is not segmented properly and it more distorted than the original image , This is because the application was made with medical images in mind. Below is an example of a medical imaged passed to the program:



We can clearly see than the medical image is properly segmented and we can identify what the injury is and the places where the masses in the brain can be found and where it is most dense. Since it is implemented by Jupiter notebook we can directly see the result instead of waiting for the processing time to get the results(which was a problem in the previous program,its equivalent python code is also provided just in case).

4.3 Lane Detection Program

This program works well when the area is properly lit (daytime) compared to when it is dark (night time). The program won't function with curved lanes, however it could work if perspective warp and curve fitting is applied. Another problem we encountered was deciding the lane segment and the lane segment properties. The properties we have taken into consideration were:

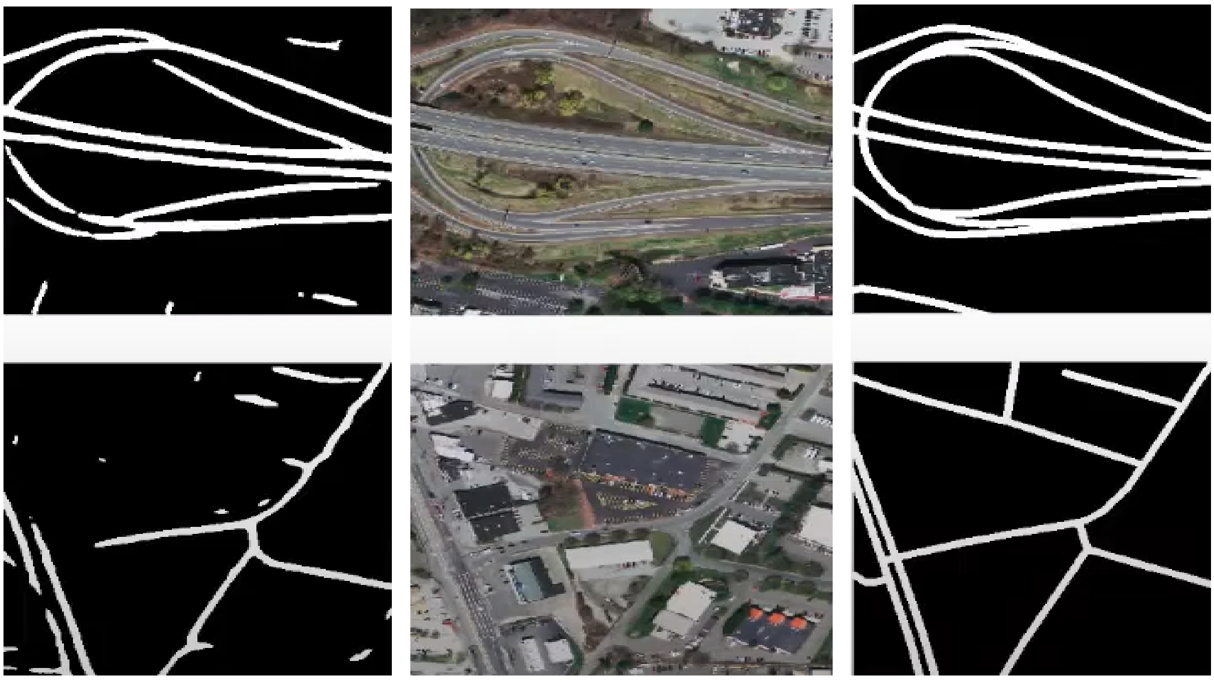
1) Canny function threshold : We can change this depending on how strict we want the edge to be, we can do so by changing the low threshold and the high threshold but the general ratio is 1:3.

2) Segment area: A 3 sided polygon on a certain area was taken, as it is the most general case.

3) Orientation : Lane segments by the virtue of their nature are never close to horizontal (unless extremely steep turns are encountered). This property helped us in removing highlighted vehicle bumpers and other segments which were otherwise being treated as false positives. Minor false positives still remained but those can be easily avoided by further process.

4.4 Road Segmentation using Satellite Images

One thing that we observed when making our project,we noticed that roads that were surrounded by grass had better performance than the roads that we near parking lots.this can be seen below:



In the above image the first column contains the predicted road, the second column contains the actual image being processed and the final column contains the actual location of the roads.

5.Tools and Technology to be used

The main technologies that was used in each project is mentioned below:

1. Generic Object Segmentation:Python
2. Medical Image Segmentation:Python and Jupyter Notebook
3. Lane Detection Program: Python
4. Road Segmentation using Satellite Images: Jupyter Notebook and Python
5. Website:Anvil

Some of the tools used by our group are:

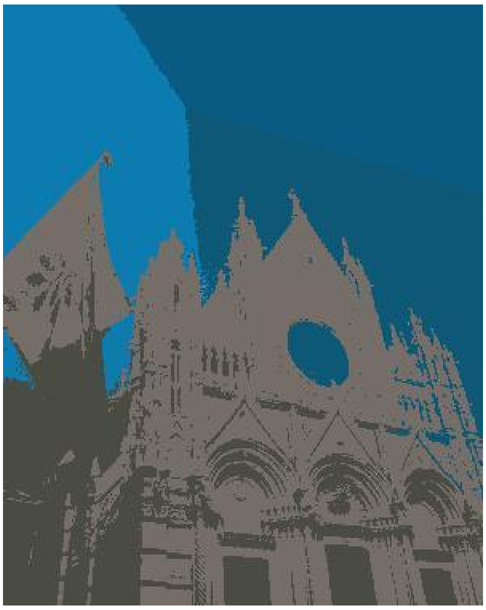
1. Visual Studio Code- LiveShare
2. GitHub
3. Microsoft Teams

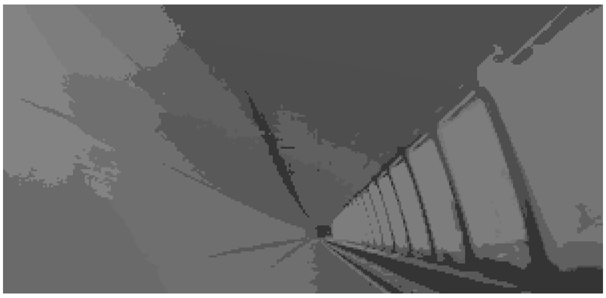
6.Conclusion and Results

The conclusion of our Minor Project is that there are a lot of applications and projects what can be made in the field of image segmentation that can be use to humankind and our project just scratches the surface of what is possible with the technology available in this field ,it provides a starting point for further development in the future for research in this field. We would again like to thanks our parents and teacher without whom this project could not be completed.

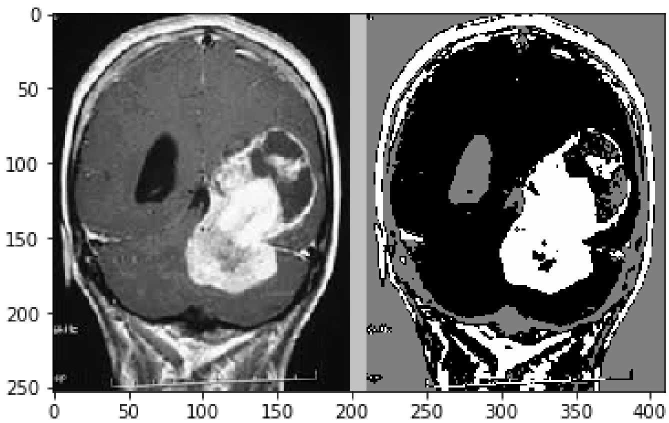
The result of each of our Program are given below:

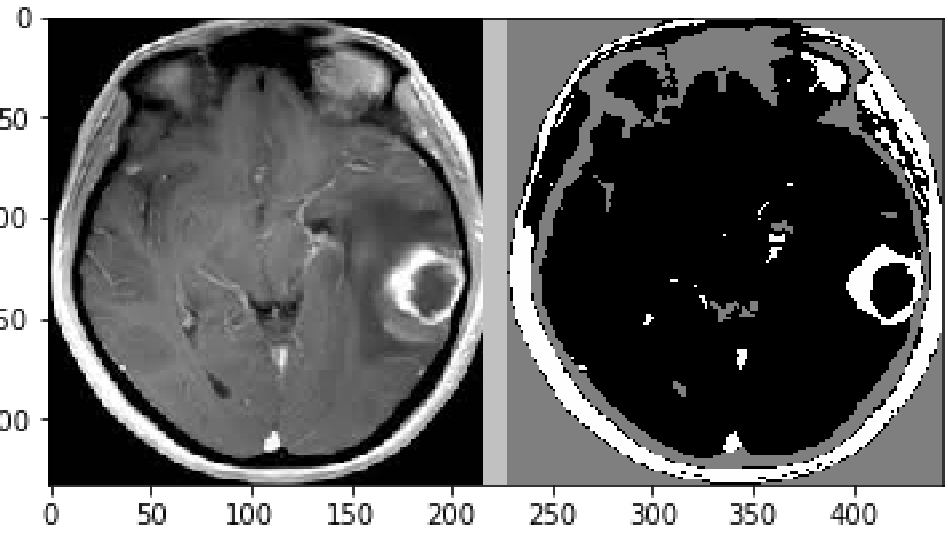
Results of Generic object segmentation program:





Results of Medical Image Segmentation Program:



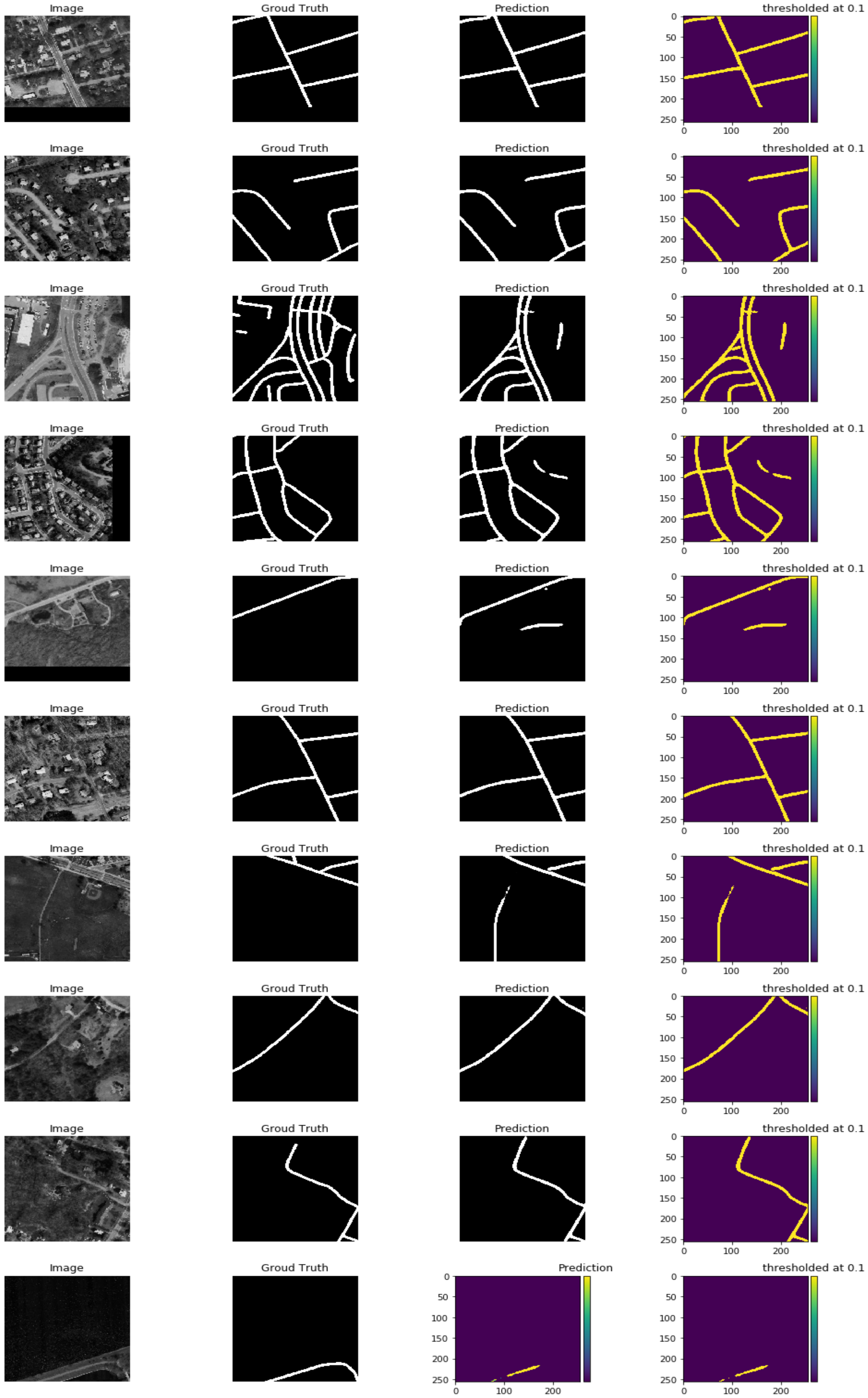


Results of Lane Detection Program:

The Output video can has been uploaded on this link :

https://www.youtube.com/watch?v=kikCMDAUhKU

<u>Result of Road Segmentation using Satellite Images Program:</u>



7.References

https://en.wikipedia.org/wiki/Image_segmentation#:~:text=Image%20segmentation%20is%20typically%20used,same%20label%20share%20certain%20characteristics.

https://www.researchgate.net/publication/340087951_Image_segmentation_Techniques_and_its_application

https://www.cs.toronto.edu/~vmnih/docs/Mnih_Volodymyr_PhD_Thesis.pdf

https://arxiv.org/pdf/1505.04597.pdf

https://www.youtube.com/watch?v=Gs5HlHKqAYQ&ab_channel=MadhurBehl

https://link.springer.com/article/10.1007/s10278-019-00227-x

https://link.springer.com/article/10.1007/s11263-019-01247-4#:~:text=Generic%20object%20detection%20is%20closely,semantic%20segmentation%20which%20does%20not.

https://youtu.be/1PKx28Ndyp8

https://youtu.be/guPS_3dauo8