CMSC 691 Digital Image Processing Final Project: Image Segmentation

Krishna Vamsi Kurumaddali, Department of Computer Science and Electrical Engineering,
University of Maryland Baltimore County

***Abstract*— Through this project on image segmentation, it is divided into implementation and analysis parts. The first part will be on implementing methods and techniques on Multiple Thresholding, K-Means clustering, SLIC (Simple Linear Iterative Clustering). Also, implementations on Multilayer feed forward network and Convolutional Neural Networks (CNN) are done for Handwritten Character Recognition. The second part of the project involves a comparative analysis for the above algorithm which lays out multiple results obtained throughout this project.**

***Index Terms* — Multiple-Thresholding, K – Means Clustering, Superpixels, Multi-Layer feed forward Networks, Convolutional Neural Networks (CNN).**

I. Introduction

THIS project is concentrated on Image Segmentation which utilizes various techniques aimed towards segmenting an image with the help of Multiple Thresholding, K-Means Clustering, SLIC (Simple Linear Iterative Clustering). And with the help of deep learning-based methods such as Multi-Layer Feed Forward Networks and Convolutional Neural Networks for recognizing handwritten character recognition.

In the implementation part, Firstly I implement the multiple thresholding based on Otsu's threshold methods[2]. Then taking the number of clusters as K=3,4,5, a K-Means Clustering technique is implemented. After K- Means, based on the concept of super pixels, SLIC (Simple Linear Iterative Clustering)[1] is implemented with a set of super pixel values. Then comes the analysis part where I will be comparing and analyzing the results obtained from the above techniques and create a statistical analysis to understand how each algorithm's approach towards segmentation of image is done. The visualized results help to know more about image segmentation at a greater detail.

When it comes to second part of the project, I focus on handwritten character recognition as it is now one of the major emerging problems needed to be solved with better techniques and efficiency. For this, the implementation of a Multi-Layer Feed Forward Network is done at first with various parameters involved to get better efficient results. The next method to be implemented is based on Convolutional Neural Networks (CNN). As I am working on Handwritten characters in form of images, CNN is highly useful and efficient to work. Various hyper parameters are given to fine-tune the model.

II. Methodology

A. Image Segmentation

Image segmentation consists of dividing the image in multiple regions set of pixels with similar variables [4]. It is generally used to find objects of an image such as lines, shapes, curves and many more. Some of the applications of Image Segmentation in real-world scenario are medical imagery, satellite imagery analysis, surveillance systems, fingerprint and iris recognition, etc.

B. Multiple - Thresholding

The multiple thresholding involves separation of pixels from the given input image into multiple classes with accordance to the gray level's intensity recorded in the image. This algorithm produces multiple thresholds to perform separation of pixels.

C. K-Means Clustering

K-Means clustering is an unsupervised algorithm based on clustering. It is used to segment area of interests from background. The partitions it creates are called clusters and these clusters are made based on a data point which is called as centroid. Here, similar set of image pixels are grouped together and are formed as clusters.

D. SLIC (Simple Linear Iterative Clustering)

When it comes to SLIC, we need to know about superpixels. A superpixel is comprised of group of pixels which have an image characteristic in common which in this case is the pixel intensity. SLIC tries to create clusters based on color similarity and also the pixel's proximity with each other.

E. Multi-Layer Feed Forward Network

The multi-layer feed forward network is comprised of interconnected perceptrons where the information flows in unidirectional manner. The layers consist of neurons with weights and the results are computed with the help of activation functions. Three major layers defines the structure of multi-layer feed forward network which are input layer, hidden layer and output layer.

F. Convolutional Neural Networks (CNN)

The convolutional neural network is based on deep learning where the model takes image as input and gives learnable parameters like weights, bias to various objects contained in

the image. CNN is also called as ConvNet and its architecture is based on the human brain's neuron connections.

Here the concept of receptive field comes into picture where individual neurons in ConvNet respond to the input image in restricted region of the whole image. To cover the entire image, collective receptive fields overlap. The important layers in the ConvNet are Convolutional layer, Pooling layer and Fully-Connected layer.

Convolutional layer is the first layer in the CNN where majority of the computation on images happen. It utilizes input data, image filter and a feature map.

Pooling layer performs dimensionality reduction which means reducing number of parameters. Two types of pooling involved which are max pooling and average pooling which can be used based on requirement.

The final layer is fully connected layer, this layer is responsible for connecting the pixel values generated by previous layer to the output layer through the help of nodes.

III. Implementation

A. Image selected

The image selected for the project is a photo of the buildings in a sunny morning of New York City with a faint rainbow in the sky.

Fig 1. Buildings in New York City Image

B. Multiple thresholding using Otsu Method

The process involves giving a gray scale image as input and using threshold values generate regions. In our case, we generate three regions to get two thresholds which is optimal and generates results more efficiently.

Histogram is also plotted to track the threshold values which are generated. At the end of processing, we get the generated image as the multi threshold Otsu image.

C. K – Means Clustering

The K – Means Clustering, I defined the number of clusters to be formed. For the implementation of model, I have chosen cluster K value as 3,4 and 5. Then assigned each data point (pixel value) to the closest centroid. After all assignments of datapoints to the nearest centroid, the model is ready for generating the segmented image. The clusters of 3, 4 & 5 images are visualized.

D. SLIC (Simple Linear Iterative Clustering)

The SLIC is computed by using clusters which are spaced in regularity and moving them with lowest gradient value. Then I computed the color and intensity. This process gets repeated till the pixels come to nearest cluster and finally converge. 4 superpixels are considered for the implementation of this algorithm are 64,128,256 and 400 respectively.

E. MNIST Dataset Description

MNIST[3] stands for Modified National Institute of Standards and Technology dataset. It comprises of 60,000 images of size 28 X 28 pixel grayscale handwritten single digits numbered between 0 and 9.

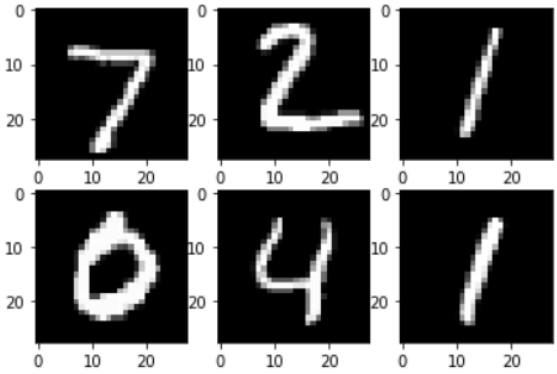

Fig 2. Sample images from MNIST Dataset

F. Multi-layer Feed Forward Network

The Hyperparameters set for the model are.

1. Image size is 28 X 28, so the input size is set to 784.
2. Total number of characters are 10 so the number of classes are set to 10.
3. Number of training epochs are set to 10 with step size as 100.
4. Hidden size 500, Batch size of 100 and learning rate of 0.001.

The optimizer used is Adam optimizer which is popular in deep learning space and suitable for this implementation. The Loss is calculated with the help of cross entropy.

When it comes to the layers, the model consists of

- 1 Input Layer
- 2 Hidden Layer
- 1 Output Layer

Loss is tracked for each epoch and the final accuracy of the model is obtained after testing the model.

G. Convolutional Neural Network (CNN)

The model is divided into two parts which are feature extraction and classifier for prediction. The image input shape is of size 28 X 28 and for weights, He initialization is used.

The layers are as follows:

1. Layer-1: Convolutional Layer with filter size of 3 X 3 with 32 filters used.
2. Layer-2: Pooling layer which utilizes max pooling. The

feature map is of size 2 X 2.

3. Layer-3: Convolutional Layer with filter size of 3 X 3 with 64 filters used.
4. Layer-4: Convolutional Layer with filter size of 3 X 3 with 64 filters used.
5. Layer-5: Pooling Layer which utilizes max pooling. The feature map is of size 2 X 2.
6. Layer-6: Flattening layer is used to convert 2D array obtained from pooling layer which are feature map into a single continuous linear vector. This layer is given as input for fully-connected layer.
7. Layer-7: Fully connected layer with 100 nodes.
8. Layer-8: Fully connected layer with 10 nodes.

The size of the receptive field is 3 X 3. Subsampling size is of 2 X 2. All the layers except the last fully connected layers use ReLu as the activation function and the last layer uses Softmax as the activation function.

The optimizer used is stochastic gradient descent with categorical cross entropy for loss, learning rate of 0.01 and momentum of 0.9 respectively. The number of epochs is 5.

IV. Results & Analysis

A. *Multiple Thresholding using Otsu method*

The original image given as input for the Otsu method[2] is figure 2 below:

Fig 3. Grayscale image of bulidings picture

The histogram obtained from the multiple thresholding using otsu method is given below in the figure 3.

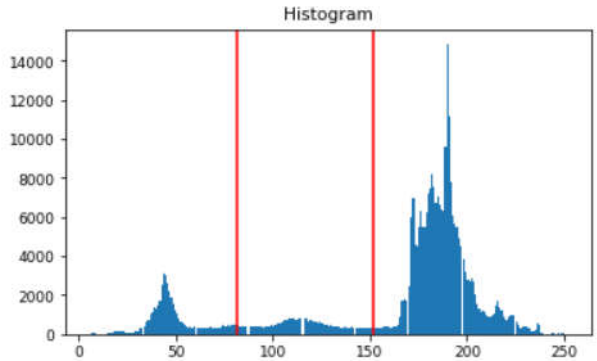

Fig 4. Histogram of the image obtained from multiple thresholding method with 2 thresholds shown by a red line.

The obtained thresholds from the method are:

- Threshold – 1: 82
- Threshold – 2: 152

The final obtained result from multiple thresholding using otsu method is depected in figure 4.

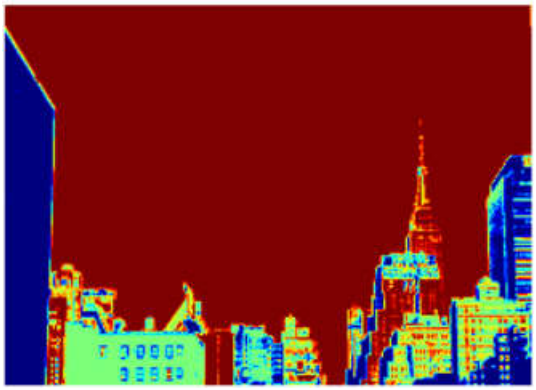

Fig 5. Final image obtained from Multiple thresholding using Otsu's Method.

B. *K – Means Clustering*

The k-means clustering is implemented with 3 cluster values of k = 3, 4 and 5.

1. For K = 3:

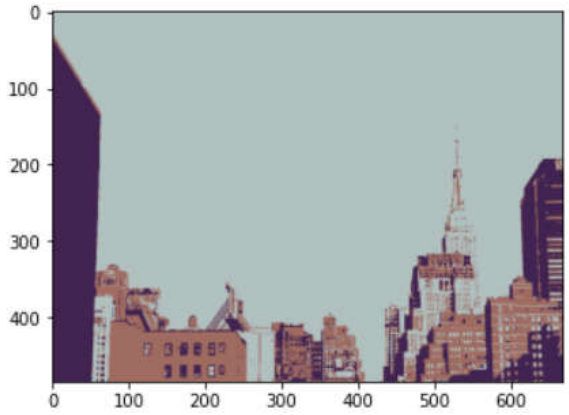

Fig 6. Image obtained from K-Means clustering for value K=3

2. For K = 4:

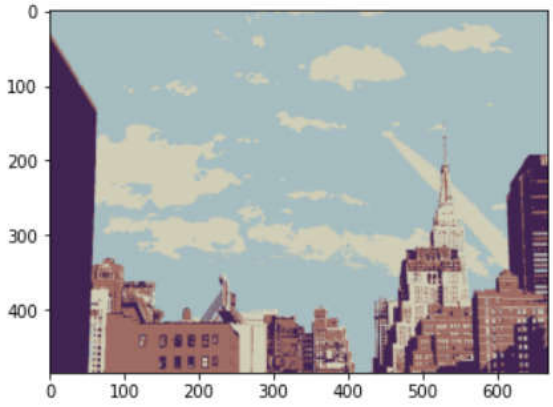

Fig 7. Image obtained from K-Means clustering for value

K=4

3. For K = 5:

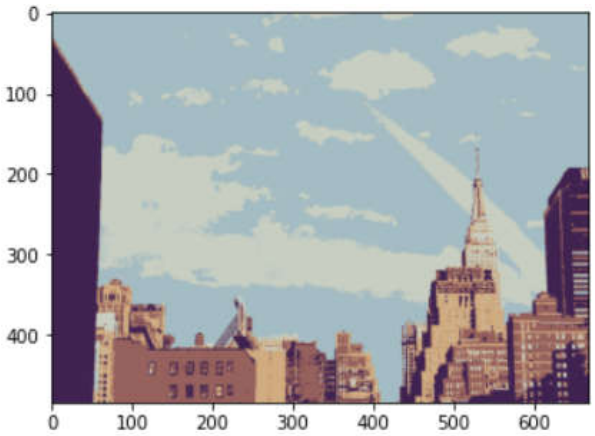

Fig 8. Image obtained from K-Means clustering for value K=5

C. SLIC (Simple Linear Iterative Clustering)

The images obtained from SLIC[1] technique are depicted below with the cluster centroids as black dots.

Fig 9. Image obtained from SLIC method for Superpixel value 64

Fig 10. Image obtained from SLIC method for Superpixel value 128

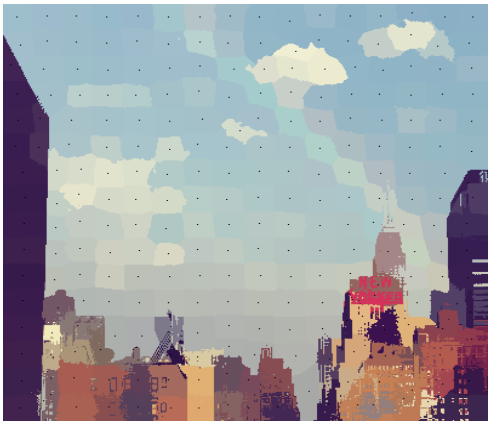

Fig 11. Image obtained from SLIC method for Superpixel value 256

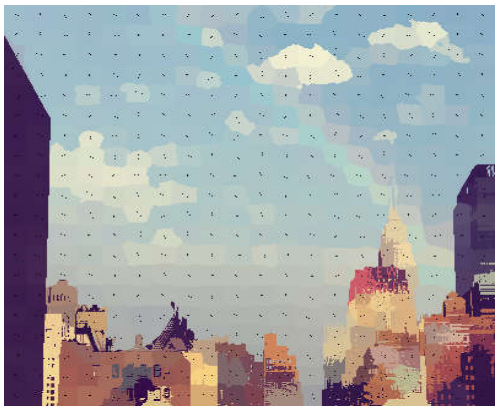

Fig 12. Image obtained from SLIC method for Superpixel value 400

D. Analysis between Multiple Thresholding, K-Means Clustering and SLIC

From the figures (3), (5), (8), (11) & (12), we can observe that Multiple thresholding has given better edges than the other two methods. Whereas K-means clustering of value K=3 has given a better segmented image.

The implemented SLIC using 4 superpixel values of 64,128,256 and 400 and of them the fig. 11 with superpixel value 64 has given better contrast for text present in the image which is segmented. But for overall the K-means and Multiple thresholding has done better segmentation.

The drawback which we can see in SLIC is when the pixel assigning is done to the centroid, the convergence of the pixel intensity causing to lose some valuable information in the image.

E. Analysis between K-Means Clustering and SLIC

When it comes to SLIC the superpixel which lie between the extreme values are yielding better results which are 128 and 256. For example, the rainbow can be seen in fig 11 and fig 12 is no where to be seen in other methods.

The trend which can be seen in K-Means clustering is that as we increase the number of clusters, the image gets more enhanced and only few information is lost.

F. *Multi-Layer Feed Forward Network*

The images which are used to train the multi-layer feed forward network are show in the figure below. The accuracy of the networks when tested on 10,000 images is 98.05%. Loss recorded when training the model is shown in the figure 12.

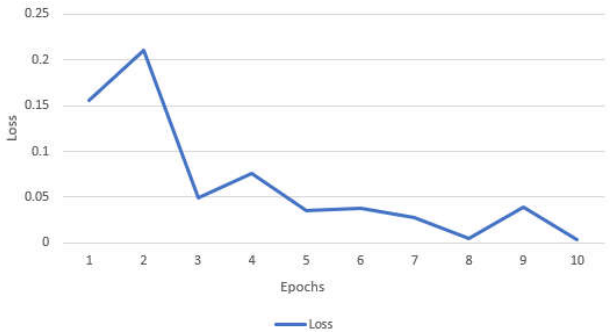

Fig 13. Epochs vs Loss recorded in training.

G. *Convolutional Neural Networks (CNN)*

The CNN model summary is shown below with total and trainable parameters.

Layer (type)	Output Shape	Param #
conv2d_15 (Conv2D)	(None, 26, 26, 32)	320
max_pooling2d_10 (MaxPooling2D)	(None, 13, 13, 32)	0
conv2d_16 (Conv2D)	(None, 11, 11, 64)	18496
conv2d_17 (Conv2D)	(None, 9, 9, 64)	36928
max_pooling2d_11 (MaxPooling2D)	(None, 4, 4, 64)	0
flatten_5 (Flatten)	(None, 1024)	0
dense_10 (Dense)	(None, 100)	102500
dense_11 (Dense)	(None, 10)	1010

Total params: 159,254
Trainable params: 159,254
Non-trainable params: 0

Fig 13. Generated CNN model summary

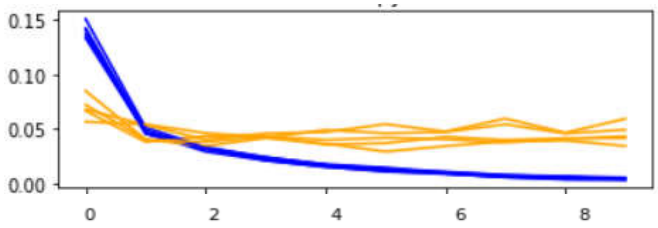

Fig 14. Cross Entropy Loss

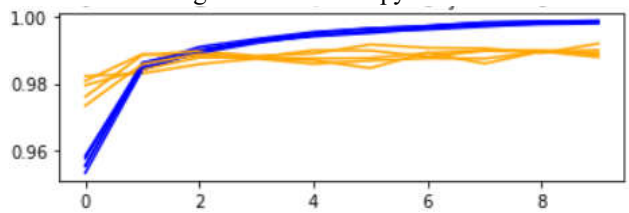

Fig 15. Accuracy

The blue line in the graph indicates the training loss and orange line indicates the validation loss. The final accuracy of the model is 98.96% with standard deviation of 0.139.

H. *Analysis between Multi-layer feed forward network and Convolutional Neural Networks (CNN)*

The feed forward network has reduced loss than CNN. The CNN has better accuracy than feed forward network where the prediction of characters is more accurate in CNN due to more complex layers incorporated into the model.

Feed forward network excels in execution as it takes less time to train the model due to its more naïve layers than CNN. The CNN utilizes each and every parameter to get itself trained and thus yielding better result when taking more time to train.

V. Conclusion

In conclusion, the project has given various results which have their pros and cons and tells us that there always exists a better model or an existing model with a better fine-tuning. In the multiple thresholding using otsu method, the distinguishing of foreground and background is better and K-means clustering is yielding better image when clusters are increased. The SLIC method gives better understanding of the pixel intensity and grouping around the centroid more effectively.

When it comes to multi-layer feed forward network, it is fast, easy to implement and efficient in producing results but lacks flexibility in terms of near perfect prediction. The CNN model takes time to get trained but is more reliable for predicting the characters which are handwritten. It also utilizes all the available parameters to get trained. Overall, image segmentation can really be helpful in real-world cases like medical imagery, satellite imagery, computer vision-based tasks and many more.

Acknowledgment

I would like to thank my professor Dr. Chein -I-Chang and Teaching Assistant Mr. Mehedi Galib for their constant support. This project could not be completed without their able guidance. This project has made me learn new concepts in image processing. I am once again thankful for the opportunity to learn under professor Chang.

References

[1] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, *34*(11), 2274-2282.

[2] Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics, 9(1), 62-66.

[3] LeCun, Y. (1998). The MNIST database of handwritten digits. *http://yann.lecun..com/exdb/mnist/*.

[4] Li, J., Erdt, M., Janoos, F., Chang, T. C., & Egger, J. (2021). Medical image segmentation in oral-maxillofacial surgery. *Computer-Aided Oral and Maxillofacial Surgery*, 1-27.