Comparative Analysis of Image Segmentation

Vandita Agarwal
M20MA208
agarwal.15@iitj.ac.in

Dattatreyo Roy
M22CS060
m22cs060@iitj.ac.in

Shilajit Banerjee
M22CS062
m22cs062@iitj.ac.in

Abstract

Image segmentation is an important problem in computer vision that involves dividing an image into meaningful regions or objects. In this report, we provide an overview of the different techniques and algorithms used for image segmentation where we apply a comparative analysis of different deep learning methods. We modified the U-Net standard architecture and compared these architectures. We have also generated image labels for this segmentation process.

1. Introduction

Image segmentation is a fundamental concept in computer vision that has a wide variety of applications in fields such as robotics, autonomous driving, and medical imaging. It is one of the most important aspects of the field. The objective of image segmentation is to break up a picture into meaningful portions or objects that can subsequently be put to use for further analysis or for activities further down the processing chain. Image segmentation can be achieved by the use of a wide number of methods and algorithms, some examples of which include threshold, clustering, edge detection, region growth, and methods that are based on deep learning. Other methods include a wide variety of image processing techniques. On the other hand, every strategy has its own advantages and disadvantages, and there is no one perfect solution that is applicable in every circumstance. There is no single tactic that can be utilised in every situation.

Figure 1. Sample Segmentation

2. Literature Survey

There has been a lot of research in the field of image segmentation, and many different approaches have been proposed. I have studied few papers on Image segmentation using KNN. In the paper ”KNN Image Segmentation Based on Distance Weighted Histograms” by J. K. Verma and R. S. Anand, the authors propose a KNN-based segmentation method that utilizes distance-weighted histograms to measure the similarity between pixels. The method is compared with several other segmentation techniques and is found to provide better segmentation results on several benchmark datasets. The paper ”U-Net: Convolutional Networks for Biomedical Image Segmentation” by Ronneberger et al. (2015) proposed a new architecture for biomedical image segmentation using convolutional neural networks (CNNs). Since its publication, the U-Net architecture has been widely adopted and applied to a variety of medical imaging tasks. Prior to the introduction of the U-Net architecture, the state-of-the-art methods for medical image segmentation typically relied on traditional image processing techniques, such as thresholding, region growing, and graph cuts. These methods often required manual parameter tuning and were not able to handle complex structures and variations in shape and texture. The U-Net architecture addressed these limitations by leveraging the power of deep learning, specifically CNNs.

In the years following the publication of the U-Net paper, numerous studies have applied and extended the U-Net architecture for various medical imaging tasks, including but not limited to, brain tumor segmentation, liver and lung segmentation, retinal vessel segmentation, and cell segmentation. These studies have reported significant improvements in segmentation accuracy and speed compared to traditional methods. One study by Çiçek et al. (2016) proposed a modified version of the U-Net architecture called "Residual U-Net" that incorporated residual connections to improve training and reduce vanishing gradients. Another study by Kamnitsas et al. (2017) introduced a multi-scale variant of the U-Net architecture called "3D Multi-Scale CNN (MS-CNN)" for brain tumor segmentation in 3D MRI scans.

3. METHODS of Image Segmentation

In our project, we are using different method of image segmentation. The methods used by us are:

1. **Clustering**: K-means clustering is a popular algorithm for image segmentation, which groups similar pixels into clusters. It works by iteratively assigning each pixel to the cluster whose mean is closest to it, updating the mean of each cluster based on the pixels assigned to it, and repeating until convergence.

2. **Edge-based segmentation**: Edge-based segmentation identifies image region boundaries using pixel intensity or gradient differences. Popular methods include Canny edge detector, Sobel operator, and LoG. Canny edge detection was used in our project.

3. **Selective segmentation**: Selective segmentation is a hybrid approach that combines both region-based and edge-based segmentation methods. The idea is to use edge-based segmentation to identify potential boundaries between regions and then use region-based segmentation to group the pixels within those boundaries into coherent regions.

4. **Semantic segmentation**: Semantic segmentation aims to classify each pixel in an image into a particular class or category. This is typically done using deep learning models such as U-NET, SegNET, and others. The models are trained on labeled datasets, where each pixel is labeled with its corresponding class. Once trained, the models can be used to segment new images by predicting the class of each pixel. For the semantic segmentation, we have used different models and trained them to get the results. The architecture used:

i. **Using the Original U-NET Architecture**: The U-NET architecture is a popular deep learning architecture for image segmentation that was introduced in 2015. It consists of a contracting path that down samples the image and a symmetric expanding path that up samples the image. The architecture uses skip connections between the contracting and expanding paths to preserve spatial information.

ii. **Using the DenseNet121 Model along with the U-NET architecture**: DenseNet is another deep learning architecture that has shown good performance for image segmentation. In this approach, the DenseNet121 model is used as the backbone for the U-NET architecture.

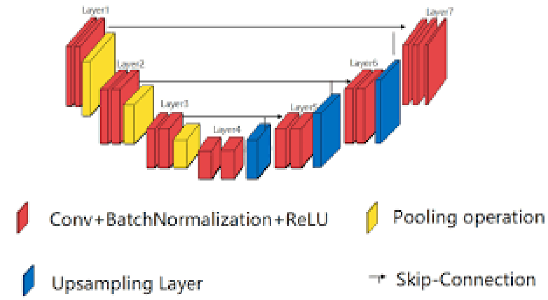

Figure 2. U-Net Architecture

iii. **Using the RESNET50 Model along with the U-NET architecture**: This method uses the RESNET50 model as the backbone for the U-NET architecture.

iv. **Segmentation with K-means clustering and U-NET architecture**: This approach combines the K-means clustering method with the U-NET architecture to perform image segmentation. The K-means algorithm is used to cluster the labels in the image into different regions, and then the U-NET model is trained to segment each region.

v. **Segmentation with DBSCAN clustering and U-NET architecture**: This approach combines the DBSCAN clustering method with the U-NET architecture to perform image segmentation. The DBSCAN algorithm is used to cluster the labels in the image into different regions, and then the U-NET model is trained to segment each region.

vi. **SegNET with VGG16 Architecture**: SegNET is another deep learning architecture that is specifically designed for image segmentation. It consists of an encoder network that downsamples the image and a decoder network that upsamples the image. The architecture uses pooling indices from the encoder network to perform upsampling in the decoder network. In this case VGG16 is used as backbone.

vii. **SegNET with Mobilenet Architecture**: In this architecture of Segnet, Mobilenet is used as backbone.

4. Work Done

K-Means clustering is an unsupervised method used for grouping and segmentation in computer vision and image processing. In image segmentation, K-Means splits an image into K clusters based on pixel color or intensity values obtained from the image.

We have used the K-Means algorithm to segment the image into K clusters (K=5) based on pixel color values. First, the image is converted into a 2D array of pixels, represented by a 3D vector of R, G, and B values. K-Means groups similar color vectors into K clusters, assigning each pixel a cluster label. The resulting segmented image is then displayed.

The **watershed algorithm** is a common approach for image segmentation that is based on the concept of flooding an image from its catchment basins. This method is quite effective and has been used extensively. The watershed algorithm is used in the context of image segmentation to separate the various objects that are present in an image by dividing the image into sections based on the local minima of the intensity or gradient values of the image. In other words, the watershed method segments the image.

The watershed algorithm is used to perform the segmentation of the input image in the code that has been provided. The grayscale image is first thresholded using Otsu's method to obtain a binary image. Morphological operations improve the foreground, and the distance transform determines object positions. Finally, the watershed algorithm segments the image, which is then displayed.

Another method is **selective segmentation**. Selective segmentation isolates regions of interest in an image, ignoring other parts. This method is useful when focusing on specific areas, and is commonly applied in computer vision for tasks like object recognition and image enhancement.

To perform selective segmentation, we load the image into memory and convert it to grayscale for better processing. A binary mask is created by thresholding the image to identify regions of interest. By applying this mask to the original image using bitwise operations, we can extract only the relevant parts. Selective segmentation is a powerful technique that allows us to focus on important areas and can be used for various applications.

The work done involved exploring different approaches for semantic segmentation, which aims to classify each pixel in an image into a particular class or category. To achieve this, various deep learning models were used, including U-NET, DenseNet121, RESNET50, and SegNET. The first approach is the **original U-NET architecture**, which consists of a contracting path that down-samples the image and an expanding path that up-samples the image. Skip connections were used between the two paths to preserve spatial information. The second approach combined the **U-NET architecture with the DenseNet121 model as the backbone**, while the third approach used the **RESNET50 model as the backbone for the U-NET architecture**.

The fourth approach combined **K-means clustering with the U-NET architecture to segment the image**. The K-means algorithm was used to cluster the labels in the image into different regions, and then the U-NET model was trained to segment each region. Similarly, the fifth approach combined the **DBSCAN clustering method with the U-NET architecture** to segment the image.

Finally, the sixth approach used the **SegNET architecture**, which is specifically designed for image segmentation. It consists of an encoder network that down-samples the image and a decoder network that up-samples the image. Pooling indices from the encoder network were used to perform upsampling in the decoder network. We have used two architectures, **SEGNET with VGG-16 as backbone** and **SEGNET with Mobilenet as backbone** to segment the images.

By this our work done involved comparing and evaluating the performance of these different approaches for semantic segmentation on various datasets. This included training and testing the models, measuring their accuracy, and visualizing the segmentation results.

5. Experiments

We have performed the image segmentation using the K-means algorithm on the Figure 1 and have obtained the result as shown in figure 2.

Figure 3. Original Image

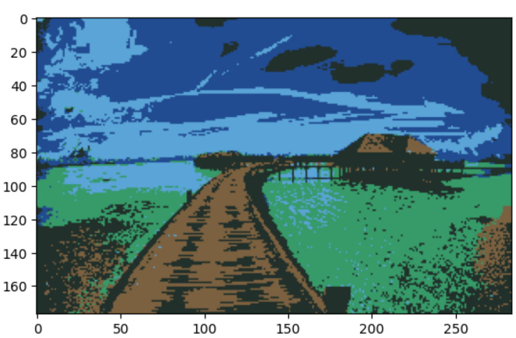

Figure 4. Segmented Image

We have performed the image segmentation using the K-means algorithm with different values of k and the results is shown on the Figure 4.

Another segmentation technique performed by us is watershed image segmentation. It is mainly used to separate foreground from the background in any image. So, it gives good results in images whose foreground or background has to be separated. It draws a line of separation to highlight foreground images.

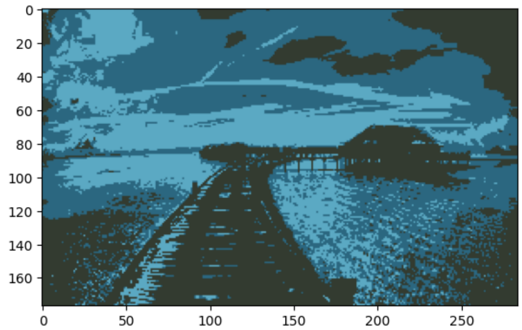

Figure 5. Segmented Image with different parameter

Here is the another set of image on which we have performed watershed segmentation with different threshold. Figure 7 and 8 are those.

Figure 6. Original Image

Figure 7. Resultant image after watersheding

For semantic segmentation, below are the results obtained.

Figure 9 and figure 10 is the Segmentation result with UNET as the model and the ground truth image. Figure 11 shows the Loss plot for the UNET model during training.

Figure 12 and figure 13 is the Segmentation result with ResNet50 as the backbone network for the UNET model. Figure 14 shows the Loss plot for the ResNet50-UNET model during training.

Figure 15 and Figure 16 is the visual of the segmentation result with DenseNet121 as the backbone network for the UNET model. The Figure 17 shows the Loss plot for the DenseNet121-UNET model during training.

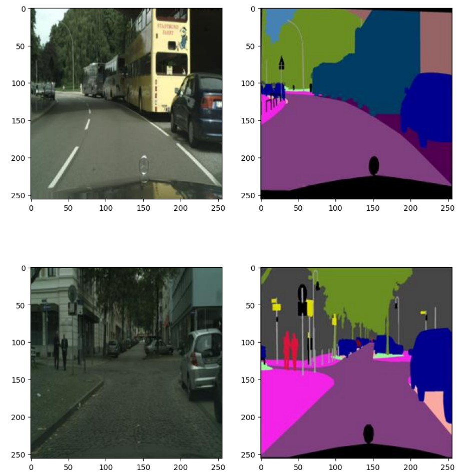

Figure 8. Sample Image Label

Figure 18 and Figure 19 is the segmentation result obtained by Kmeans clustering on the UNET model. Figure 20 shows the Loss plot for the Kmeans clustering with UNET model.

Figure 21 and Figure 22 shows the visual of the Segmentation result obtained by DBSCAN clustering on the UNET model.Figure 23 shows the Loss plot for DBSCAN clustering with UNET model.

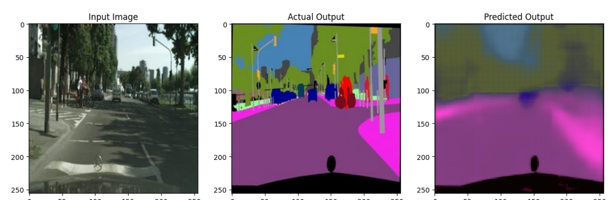

Figure 9. Segmentation with UNET as groundtruth

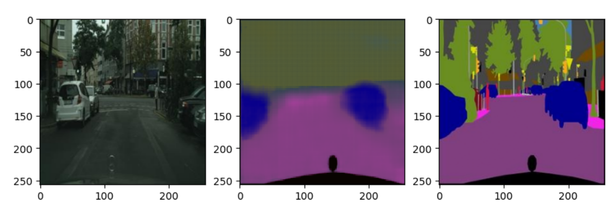

Figure 10. Segmentation with UNET as groundtruth

6. Evaluation

We have found out the evaluation metrices for the following technique mentioned in the methods. We have found out the metrices for the Semantic segmentation methods.

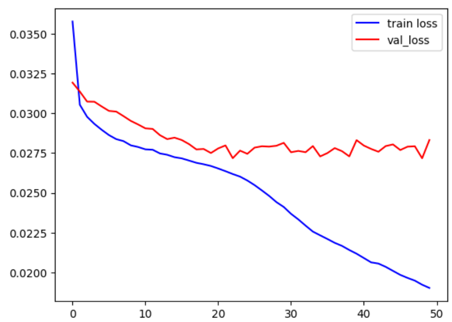

Figure 11. Loss Plot for Segmentation with UNET

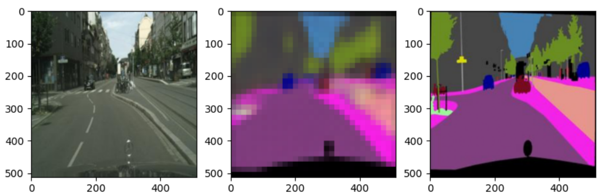

Figure 12. ResNEt50 as backbone output1

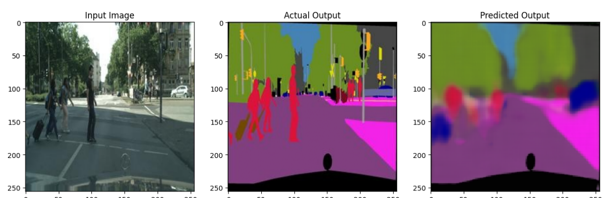

Figure 13. ResNEt50 as backbone output2

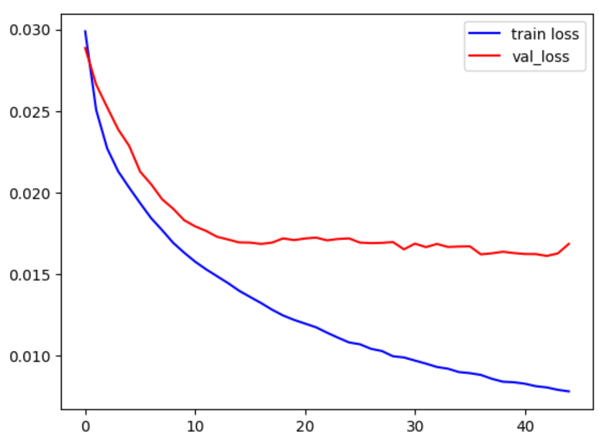

Figure 14. Loss plot for ResNEt50 as backbone

Segmentation is the process of dividing an image into different regions or segments based on the characteristics of the image pixels. The accuracy of segmentation is crucial for many computer vision applications, such as object detection, image recognition, and medical imaging. There-

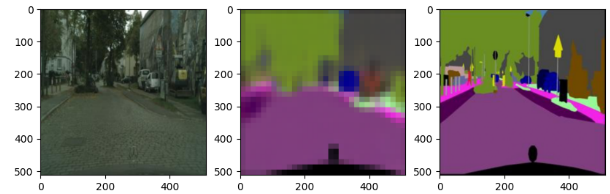

Figure 15. DenseNet121 as backbone output1

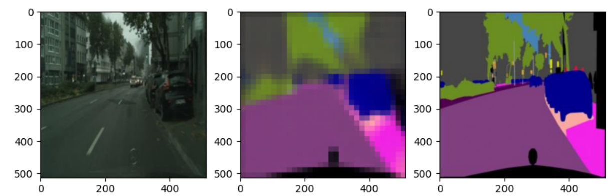

Figure 16. DenseNet121 as backbone output2

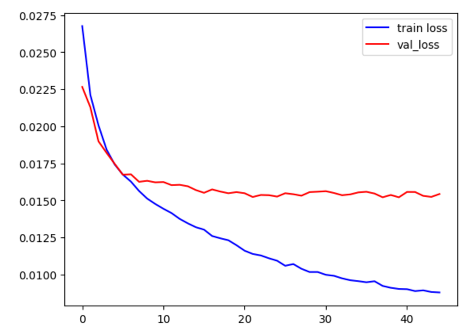

Figure 17. Loss plot for Densenet121 as backbone

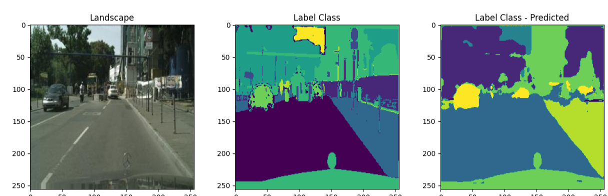

Figure 18. Segmentation by Kmeans with Kmeans output1

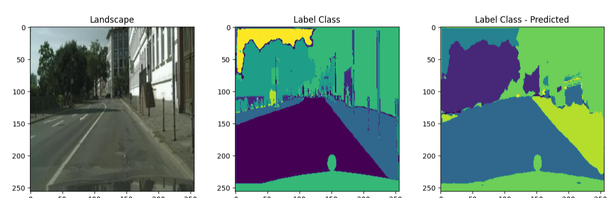

Figure 19. Segmentation by Kmeans with UNET output2

fore, it is essential to evaluate the performance of different segmentation models using appropriate evaluation metrics.

For the evaluation, we have found out the IOU score, Dice score and the Hausdorff distance. IOU score, also

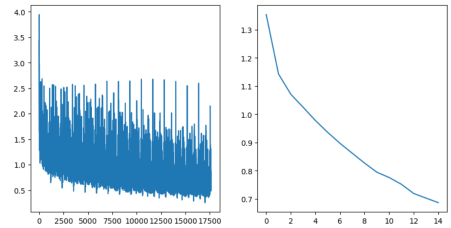

Figure 20. Loss plot for Segmentation by Kmeans with UNET

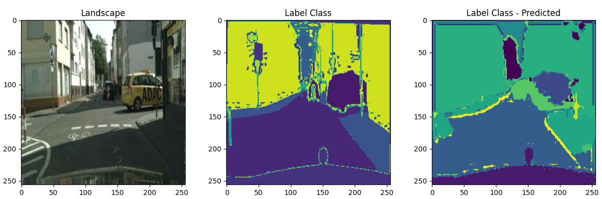

Figure 21. Segmentation by DBSCAN with UNET output1

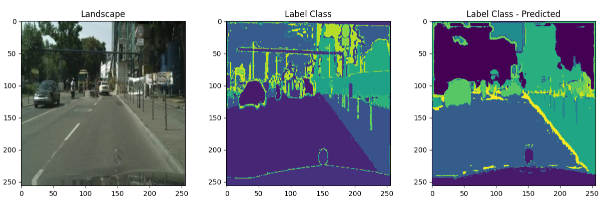

Figure 22. Segmentation by DBSCAN with UNET output2

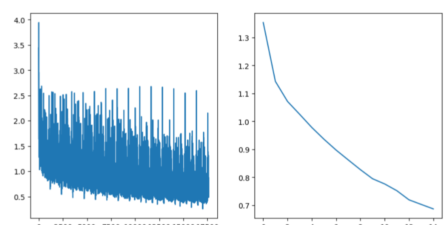

Figure 23. Loss plot for Segmentation by DBSCAN with UNET

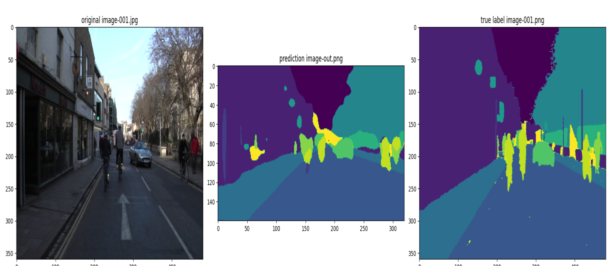

Figure 24. Segmentation by VGG16 with SEGNET output

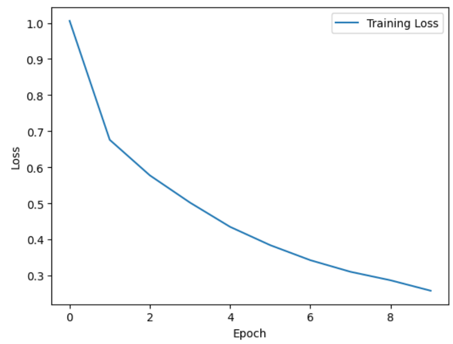

Figure 25. Loss plot for Segmentation by VGG16 with SEGNET

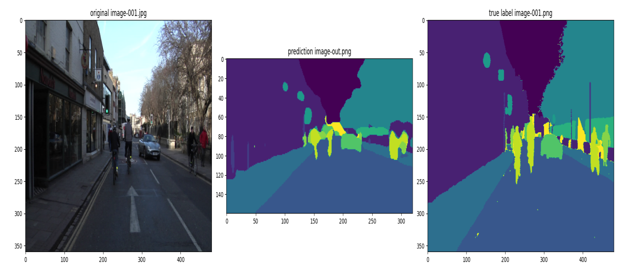

Figure 26. Segmentation by Mobilenet with SEGNET output

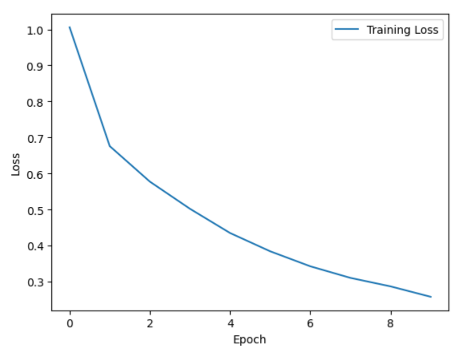

Figure 27. Loss plot for Segmentation by Mobilenet with SEGNET

known as Jaccard index, is a measure of overlap between the predicted segmentation and the ground truth segmentation. It is calculated as the ratio of the intersection of the predicted and ground truth segmentation to their union. The Dice score, also known as F1 score, is another measure of similarity between the predicted and ground truth segmentation. It is calculated as twice the product of the intersection and union of the predicted and ground truth segmentation divided by their sum of squares.

The IOU score and the Dice score are two commonly used evaluation metrics for image segmentation tasks. The IOU score measures the overlap between the predicted segmentation and the ground truth segmentation. A value of 1 indicates a perfect overlap, and a value of 0 indicates no overlap. The Dice score measures the similarity between the predicted segmentation and the ground truth segmentation. A value of 1 indicates a perfect match, and a value of 0 indicates no match. Both the IOU score and the Dice score are widely used in image segmentation tasks, and they provide a good indication of the segmentation accuracy.

The Hausdorff distance measures the maximum distance between the predicted segmentation and the ground truth segmentation. It is a measure of the largest deviation between the predicted and ground truth segmentation. These evaluation metrics help us to understand the accuracy of the segmentation results and to compare the performance of different models.

Results of the evaluation:

Table 1. Evaluation Metrics

Sl No	Avg IOU score	Med IOU score	Avg Dice score	Med Dice score	Avg Hausdorff distance	Med Hausdorff distance
1	0.4157	0.4189	0.325	0.302	46.281	42.168
2	0.9382	0.9408	0.453	0.413	31.138	29.176
3	0.8082	0.8109	0.438	0.426	25.689	30.256
4	0.6882	0.6908	0.312	0.296	41.544	43.514
5	0.4182	0.4203	0.296	0.215	45.148	47.134

Please follow the serial number of the following methods and refer in the table. Techniques: 1. Segmentation with normal UNET Architecture (Ground Truth) 2. Segmentation with DenseNet121 and UNet Architecture 3. Segmentation with Resnet50 and UNet Architecture 4. Segmentation with Kmeans and UNet Architecture 5. Segmentation with DBSCAN and UNet Architecture

7. Conclusion and Future Works

In conclusion, image segmentation is a crucial task in computer vision, and various approaches have been proposed in recent years. Deep learning-based approaches, such as fully convolutional networks and semantic segmentation, have shown promising results in accurately segmenting images. In this project, two approaches have been explored: K-Nearest Neighbors and Semantic Segmentation using DeepLab. Preliminary results indicate that both approaches have their strengths and weaknesses and require further experimentation and evaluation.

As for future work, exploring other deep learning-based approaches and evaluating their performance on the given dataset would be beneficial.

8. Reference

1. Long, J., Shelhamer, E., Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

2. Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 234-241).

3. He, K., Gkioxari, G., Dollar, P., Girshick, R. (2017). Mask R-CNN. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

4. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.

5. ”KNN Image Segmentation Based on Distance Weighted Histograms” by J. K. Verma and R. S. Anand, published in the International Journal of Computer Applications in 2013.

6. ”KNN Based Image Segmentation using Local Variance and Entropy” by H. S. Bhadauria and S. K. Gupta, published in the International Journal of Computer Applications in 2014.

7. ”Image Segmentation using Watershed Transform” by Amandeep Kaur, Aayushi, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-4, Issue-1, March 2014

8. ”Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In Medical Image Computing and Computer-Assisted Intervention” – MICCAI 2015 (pp. 234-241). Springer International Publishing.

9. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016 (pp. 424-432). Springer International Publishing.

10. Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., ... Rueckert, D. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical image analysis, 36, 61-78.

11. ”Badrinarayanan, V., Kendall, A., Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation.” IEEE transactions on pattern analysis and machine intelligence, 39(12), 2481-2495.

12. ”Mohanty, S. P., Hughes, D. P., Salathé, M. (2016). Using deep learning for image-based plant disease detection.” Frontiers in plant science, 7, 1419.