Comparative Analysis of Deep Learning and Classical Computer Vision Techniques for Semantic Segmentation

Haardik Ravat
B20EE021

Chakshu Dhannawat
B20AI006

Abstract

Semantic segmentation plays a crucial role in computer vision tasks, enabling pixel-level understanding of images. In this study, we present a comparative analysis of popular methods for semantic segmentation: Markov Random Fields along with Graph Cut, Recurrent Neural Networks (RNN), Fully Convolutional Neural Networks (FCN), and U-Net. We explore their strengths and weaknesses and evaluate their performance on a benchmark dataset. Our results provide insights into the effectiveness of each method and offer guidance for selecting an appropriate approach based on specific application requirements.

1. Introduction

Semantic segmentation is a crucial task in computer vision with numerous applications such as medical image analysis, autonomous driving, and video surveillance. Deep learning methods have shown remarkable performance in semantic segmentation tasks, but they require a large amount of annotated data and computing resources to train and deploy. In contrast, classical computer vision techniques are computationally efficient and require fewer annotated data points. In this project, we aim to compare the performance of deep learning and classical computer vision techniques for semantic segmentation on various datasets and analyze their differences.

2. Methods

To conduct our comparative analysis, we will first review and select relevant papers and datasets related to deep learning and classical computer vision techniques for semantic segmentation. We will implement representative methods from both, deep learning and non-deep learning based categories. We will then evaluate and compare the performance of these methods on selected datasets using standard evaluation metrics such as Intersection over Union (IoU) and Pixel Accuracy. We will also analyze the computational requirements, robustness, and generalizability of these methods.

Overall, this project aims to provide insights into the strengths and weaknesses of deep learning and classical computer vision techniques for semantic segmentation and help researchers and practitioners select appropriate methods for their specific applications.

2.1. Research Papers

Given the extensive literature available in the field of semantic segmentation, we have judiciously handpicked a select few research papers to implement and gain a deeper understanding of the underlying algorithms employed. They are mentioned below:

1.Region classification with Markov field aspect models

The paper presents a new approach to recognize and localize visual object classes using aspect-based spatial field models that combine two complementary approaches: traditional spatial models and aspect models. The models, which can be trained using either patch-level labels or image-level keywords, use factored observation models combining texture, color, and position cues.

2.Reseg: A recurrent neural network-based model for semantic segmentation

The paper presents a novel architecture called ReSeg for semantic segmentation that combines CNNs and RNNs to exploit local and global information. ReSeg achieves state-of-the-art performance on several widely-used semantic segmentation datasets and is efficient, flexible, and suitable for a variety of structured prediction tasks.

3.U-Net: Convolutional Networks for Biomedical Image Segmentation
This paper presents a novel network and training strategy that efficiently uses annotated samples for deep network training. The proposed architecture includes a contracting path to capture context and a symmetric expanding path for

precise localization, achieving superior results on the ISBI challenge for segmentation of neuronal structures and cell tracking challenges in transmitted light microscopy images, with fast processing times.

4. Fully Convolutional Networks for Semantic Segmentation

This paper introduces fully convolutional networks (FCNs) that surpass the state-of-the-art in semantic segmentation by utilizing deep learning and end-to-end training. Its significance lies in their ability to efficiently process images of any size and produce accurate segmentations, achieving superior results on benchmark datasets while reducing inference time.CNs allow for efficient processing of images of varying sizes, resulting in accurate and detailed segmentations, thereby advancing the field of semantic segmentation.

2.2. Datasets

For the purpose of conducting a comprehensive comparative analysis, we have selected the following datasets as the primary subjects of our experimentation.

1. PASCAL VOC (PASCAL Visual Object Classes Challenge)

The PASCAL Visual Object Classes (VOC) 2012 dataset contains 20 object categories including vehicles, household, animals, and others. Each image in this dataset has pixel-level segmentation annotations, bounding box annotations, and object class annotations. This dataset has been widely used as a benchmark for object detection, semantic segmentation, and classification tasks.

2. DRIVE

The Digital Retinal Images for Vessel Extraction (DRIVE) dataset is a dataset for retinal vessel segmentation. The DRIVE dataset consists of 40 fundus images, all with a resolution of 584 × 565, with eight bits per colour channel(3 channels). These images were randomly selected from a diabetic retinopathy screening set of about 400 Dutch participants suffering from diabetes. In this subset of 40 images, 33 of them are considered healthy, and the remaining 7 have early signs of diabetic retinopathy.

We have primararily used the PASCAL VOC dataset to perform a comparative analysis among the various methods, and the DRIVE dataset, to test the application of the U-Net model.

2.3. Metrics Used

We have used the following metrics to perform a comparative analysis among the various approaches used in semantic segmentation.

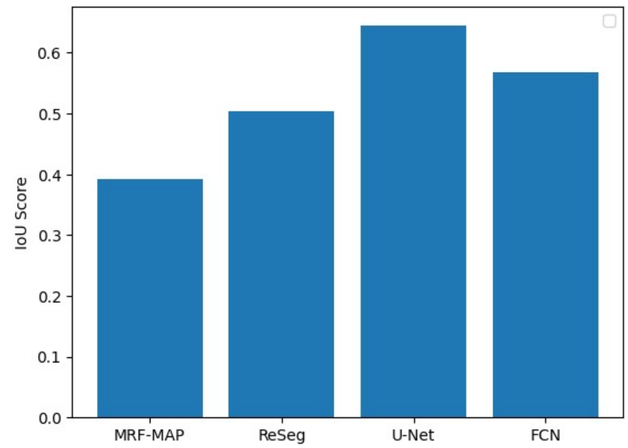

Figure 1. Comparing IoU scores of the methods

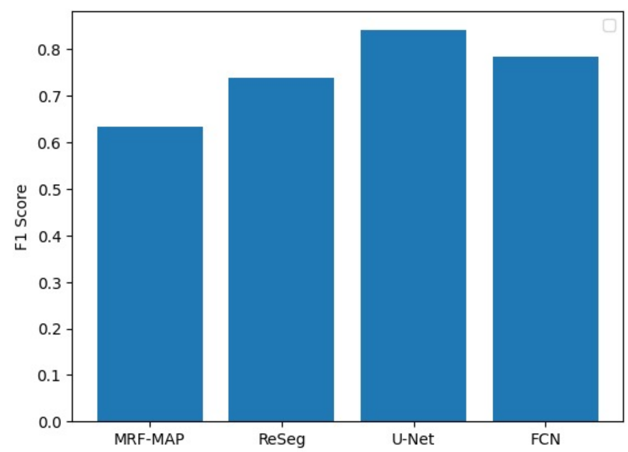

Figure 2. Comparing F1 scores of the methods

2.3.1 Pixel Accuracy

Pixel accuracy measures the percentage of pixels that are correctly classified in the predicted segmentation mask compared to the ground-truth segmentation mask.

2.3.2 IoU score

IoU measures the similarity between the predicted and ground-truth segmentation masks by computing the ratio of their intersection and union. IoU ranges from 0 to 1, with higher values indicating better performance.

2.3.3 F1-score

The F1 score in semantic segmentation is a balanced metric that combines precision and recall to evaluate the model's performance. It indicates the overall accuracy and completeness of the predicted segmentation masks, with a

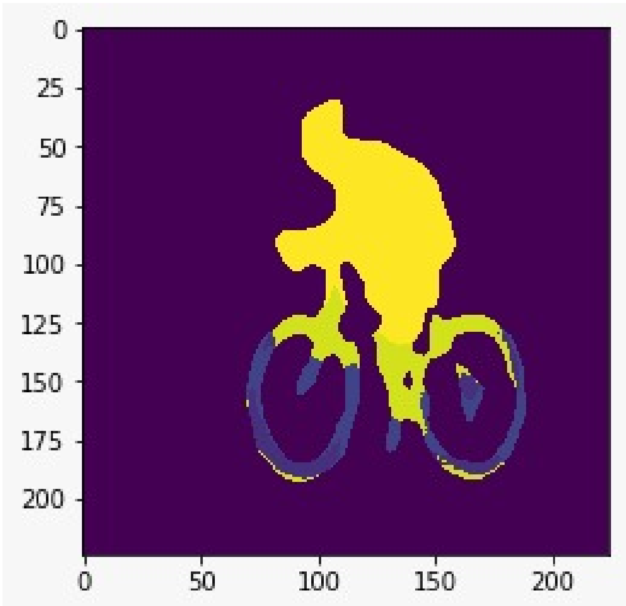

Figure 3. Segmentation map generated from MRF

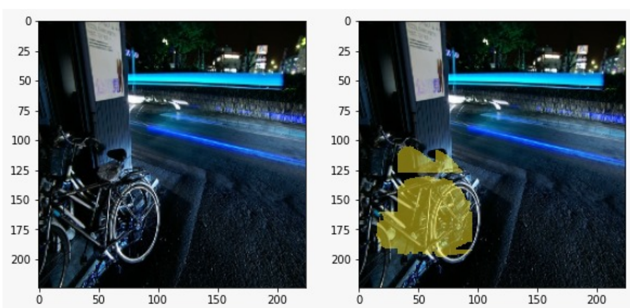

Figure 4. Segmentation map generated from ReSeg

higher value indicating better performance. The F1 score ranges from 0 to 1, where 1 represents perfect precision and recall.

3. Observations and Comparative analysis

3.1. Quantitative analysis

For comparing the models quantitatively we have compared the IoU scores, F1 scores, Pixel Accuracy scores of all the methods when testing on the test dataset of Pascal VOC. You can see the plots in Figure 1 and 2.

3.2. Qualitative analysis

For qualitative analysis, we have visually compared various random segmented images from the methods. The results of which are shown in Figure 3,4,5,6.

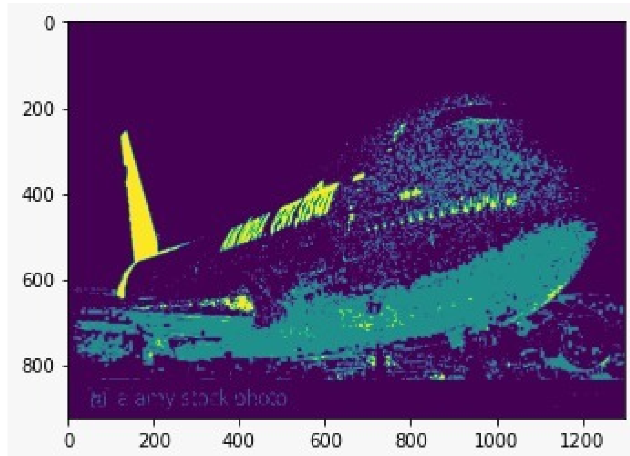

Figure 5. Segmentation map generated from FCN

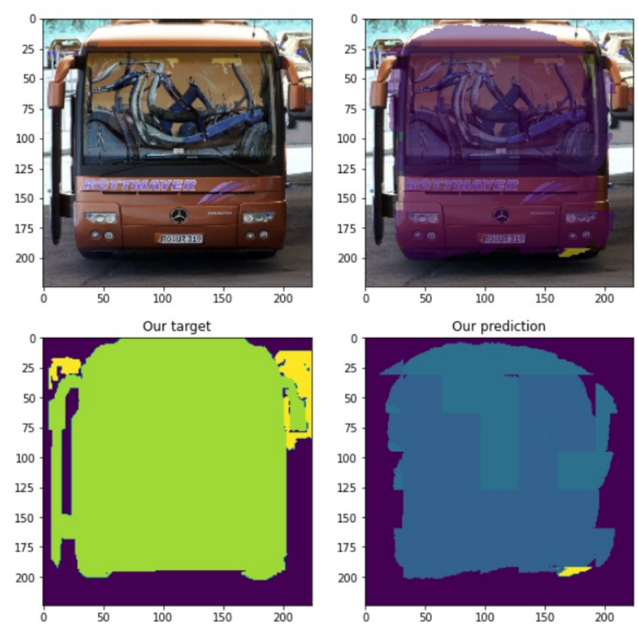

Figure 6. Segmentation map generated from U-Net

3.3. Overall analysis

In terms of qualitative analysis, the four semantic segmentation techniques—Unet, FCN, Reseg, and MRF-MAP—exhibited varying levels of performance on the Pascal VOC dataset. Among them, Unet stood out as the best performer, consistently producing accurate and well-defined object boundaries. Unet's architecture, featuring skip connections, effectively captured both local and global information, resulting in superior segmentation results. FCN also demonstrated good qualitative performance, although the object boundaries were slightly less precise compared to Unet. FCN's utilization of fully convo-

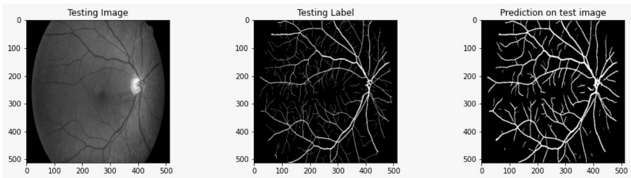

Figure 7. Figure representing segmentation using U-Net. Ground truth(center). Generated Segmented image(right)

lutional layers and multi-scale processing contributed to its overall effectiveness. Reseg, on the other hand, fell slightly behind Unet and FCN but still provided satisfactory qualitative results. Finally, MRF-MAP performed the worst among the techniques, indicating limitations in its ability to accurately segment objects in the Pascal VOC dataset. Based on this analysis, Unet was chosen as the preferred technique for further models due to its superior qualitative performance.

4. Exploring Potential Applications

We have further extended the project to explore the application of U-Net for retinal image segmentation on the DRIVE dataset. We have observed that the pixel accuracy was about 94 percent and the IoU score was around 0.64 percent. You can refer to Figure 7 for the same.

5. Conclusion

In order to achieve high-quality semantic segmentation, there are two commonly concerned questions: how to design efficient feature representations to differentiate objects of various classes, and how to exploit contextual information to ensure the consistency between the labels of pixels. For the first question, most early methods benefit from using the hand-engineered features, such as Scale Invariant Feature Transform (SIFT) and Histograms of Oriented Gradient(HOG). With the development of deep learning, the use of learned features in computer vision tasks, has achieved great success in past few years. As a result, the semantic segmentation community recently paid lots of attention to the learned features, which usually refer to Convolutional Neural Networks(CNN or ConvNets). For the second issue, the most common strategy, no matter the feature used, is to use contextual models such as Markov Random Field(MRF). These graphical models make it very easy to leverage a variety of relationships between classes via setting links between adjacent pixels. More recently, the use of Recurrent Neural Networks(RNN) is more commonly seen in retrieving contextual information. Also, recent use of U-Nets and FCN's for semantic segmentation has proven to be very successful.

Although there are many strategies available for addressing the problems mentioned above, these strategies are not yet mature. For example, there are still no universally accepted hand-engineered features while research on learned features has become a focus again only in recent few years.Thus, new and creative semantic segmentation methods are being developed and reported continuously.

The main motivation of this project is to provide a comprehensive survey of semantic segmentation methods, focus on analyzing the commonly concerned problems as well as the corresponding strategies adopted.

References

[1] Link to Github

[2] Link to Video Demo

[3] Region classification with Markov field aspect models *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2007), pp. 1-8, J. Verbeek, B. Triggs*

[4] Exact Maximum A Posteriori Estimation for Binary Images *In Journal of the Royal Statistical Society: Series B (Methodological). D. M. Greig, B. T. Porteous, A. H. Seheult*

[5] Reseg: A recurrent neural network-based model for semantic segmentation *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2016), pp. 41-48, F. Visin, M. Ciccone, A. Romero, K. Kastner, Cho K., Y. Bengio, M. Matteucci, A. Courville*

[6] U-Net: Convolutional Networks for Biomedical Image Segmentation *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*

[7] Fully Convolutional Networks for Semantic Segmentation *Jonathan Long, Evan Shelhamer, Trevor Darrell*

[8] Methods and datasets on semantic segmentation: A review *Hongshan Yu, Zhengeng Yang, Lei Tan, Yaonan Wang, Wei Sun, Mingui Sun, Yandong Tang*