COMP9444 Project Summary

Image Segmentation in Autonomous Driving

Team Nexa

<z5187710 Neil Liao>

<z5442663 Jingjie Xu>

<z5499015 Hangwei Liang>

<z5449848 Ze Chen>

<z5451234 Nuo Chen>

I. Introduction

The primary objective of our project is to explore the efficacy of various image segmentation techniques, specifically Unet, Segnet, and DeeplabV3, in the context of autonomous driving. The significance of this study lies in the crucial role that precise image segmentation plays in augmenting the perception capabilities of autonomous vehicles. It is essential for their safe and effective operation in complex real-world scenarios.

Accurate identification and segmentation of critical elements, such as lanes, vehicles, pedestrians, and traffic signs are vital for the functionality of autonomous driving systems. By evaluating the performance of these advanced models—Unet, Segnet, and DeeplabV3, our project aims to offer valuable contributions to the enhancement of current practices in the field. Through this project, we seek to provide insights into which of these segmentation models is most effective in dealing with the diverse challenges posed by real-world driving environments, ultimately aiding in the development of more reliable and efficient autonomous driving technologies.

II. Literature Review

This project explores three pivotal models for image segmentation in the field of computer vision: U-Net, SegNet, and DeeplabV3. Each of these architectures has made a significant impact in various application domains.

The U-Net architecture, first introduced in the paper "U-Net: Convolutional Networks for Biomedical Image Segmentation" by Ronneberger et al. in 2015 [1], has been particularly successful in biomedical image segmentation. Its encoder-decoder structure, enhanced with skip connections, is known for its effectiveness and simplicity. This model has been adapted and improved for different applications, such as natural and satellite image segmentation, leading to variations like Attention U-Net and Nested U-Net.

SegNet, on the other hand, is renowned for its efficiency in processing noisy and low-resolution images. Presented in "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation" by Badrinarayanan et al. in 2017 [2], its memory-efficient design makes it well-suited for real-time applications.

The DeeplabV3, a model that has significantly advanced the field of semantic image segmentation. Introduced in the paper "Rethinking Atrous Convolution for Semantic Image Segmentation" by Chen et al. in 2017 [3], DeeplabV3 incorporates atrous convolutions to capture multi-scale context by employing multiple parallel filters with different rates. This approach allows DeeplabV3 to effectively segment images with fine details and complex structures, making it highly effective in tasks like autonomous driving and urban scene understanding.

Each of these models—U-Net, SegNet, and DeeplabV3—demonstrates unique strengths and applications, collectively pushing the boundaries of image segmentation techniques in various domains. This project aims to delve deeper into their architectures, applications, and the enhancements they bring to the field of computer vision.

III. Methods

In this project, we use Unet, Segnet, and DeeplabV3 as image segmentation methods. These methods were carefully adapted and optimized for the specific challenges and requirements of autonomous driving scenarios. The rationale for choosing these methods lies in their proven effectiveness in handling complex spatial relationships, noise reduction, and capturing long-range dependencies. To expedite the training process, pretrained models were utilized, and fine-tuning was conducted to ensure adaptability to the nuances of the autonomous driving dataset [4]. The implementation was carried out using the PyTorch framework, leveraging its extensive deep learning libraries and computational capabilities. Additionally, we also use data augmentation to add around 3,000 images to the dataset to enhance the data quality.

IV. Experimental Setup

The dataset used in this study comprises a comprehensive collection of high-resolution images relevant to various aspects of autonomous driving, including road scenes, traffic signs, pedestrians, and vehicles. The dataset encompasses 5,967 images sourced from publicly available repositories from Kaggle, and we implemented a data augmentation to increase approximately 3,000 images to ensure the quality of data. Exploratory data analysis revealed the presence of diverse environmental conditions, complex road geometries, and varying lighting scenarios, which posed significant challenges for the image segmentation models.

Our dataset covers a wide array of elements critical to autonomous driving systems. This includes varied road scenes capturing different traffic conditions, a spectrum of traffic signs in various settings, pedestrians in multiple postures and actions, and a range of vehicles in different sizes and types. The diversity in the dataset is crucial, as it exposes the autonomous driving models to a wide range of real-world scenarios, they are likely to encounter.

The preprocessing steps are an integral part of preparing this dataset for effective use in model training. Firstly, all images are resized to maintain consistency in input dimensions for the models [5]. This resizing is done carefully to preserve the integrity of the image data and to ensure that crucial details are not lost.

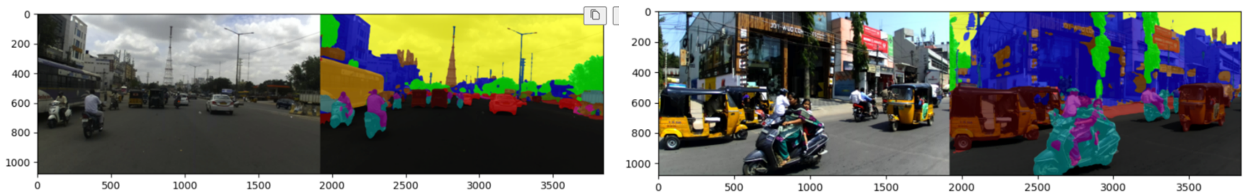
Furthermore, the images are converted into tensors, which are multi-dimensional arrays that are particularly suited for feeding into neural networks, the backbone of most autonomous driving systems. This conversion is a key step in making the dataset compatible with the deep learning models used in autonomous driving.

A unique aspect of the preprocessing involves handling special cases, notably the pixel value of 255. In this dataset, the value 255 is used to represent "unlabeled/out of Region of Interest (ROI)" data. This is significant as it helps the training models to distinguish between relevant and non-relevant parts of the image, focusing their learning on the areas that matter most for autonomous driving [6].

V. Results

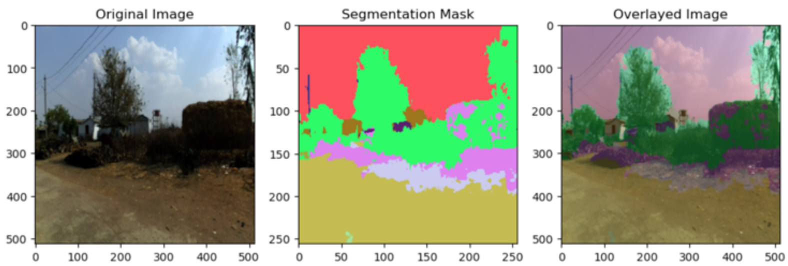
In the analysis of the three different models for autonomous driving datasets, we observe distinct performance metrics.

U-Net, with its encoder-decoder architecture and skip connections, demonstrates the ability to capture multi-scale semantic features and preserve spatial information, achieving over 70% mean Intersection over Union (mIoU) and above 80% accuracy. The code of Unet model was found from Github and we use this model as our baseline model. This performance suggests a strong capability in segmenting road scenes and identifying various objects, despite challenges like scale variation and occlusions. However, U-Net may not be as efficient as SegNet in processing very large images due to its structure. Moreover, it might not perform as well on non-medical images or require more training data for generalization.



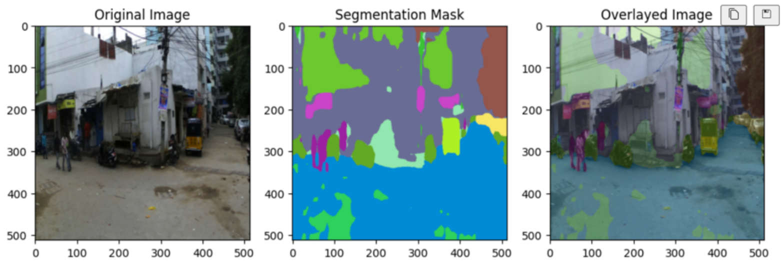
On the other hand, SegNet showcases a commendable accuracy of 83% on the test set, with a precision of 0.6827 and recall of 0.5231, indicating a strong segmentation capability. The color-coded visualization of the segmentation masks overlaid on the original images suggests effective delineation, even without the benefit of pre-training. However, a comparison with pre-trained networks like ResNet18 and ResNet50 indicates room for improvement in SegNet's performance. The lower F1 scores across certain classes suggest potential issues with class imbalance or the model's ability to generalize across diverse scenarios.

```
Test set: Average loss: 0.1303, Accuracy: 55274622/66584576 (83%)
Precision: 0.6827, Recall: 0.5231, F1 Score: 0.5333
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.92      0.94  16716294
           1       0.66      0.80      0.72   4063060
           2       0.62      0.56      0.59    209395
           3       0.54      0.56      0.55    966046
           4       0.59      0.67      0.62    647699
           5       0.71      0.52      0.60    448535
           6       0.71      0.80      0.75    875874
           7       0.38      0.06      0.10     23437
           8       0.79      0.81      0.80    949721
           9       0.87      0.84      0.85   1505188
          10       0.54      0.61      0.58    761102
          11       0.42      0.57      0.49    235629
          12       0.65      0.08      0.15    262711
          13       0.82      0.79      0.80    948792
          14       0.44      0.51      0.47    906122
          15       0.50      0.20      0.29    433762
          16       0.83      0.27      0.41     98617
          17       0.59      0.41      0.48   1082493
          18       0.66      0.07      0.12     59085
...

    accuracy                           0.83  66584576
   macro avg       0.68      0.52      0.53  66584576
weighted avg       0.83      0.83      0.82  66584576
```



DeeplabV3 showcases a remarkable performance, distinguishing itself with its advanced segmentation capabilities. Achieving an impressive accuracy of 90.5% on the training set and stabilizing at 87% on the validation set, DeeplabV3 outperforms many of its counterparts. Its precision index stands at 0.754, with a recall rate of 0.636 and an F1 score of 0.651, indicating a balanced proficiency in both precision and recall. The effectiveness of DeeplabV3 is further evidenced by the visualization of segmented results, where different colors assigned to each segmented object allow for a clear and detailed observation of the segmentation effectiveness. This high level of performance is attributed to the model's use of pre-trained weights and the fine-tuning of its optimizer and loss function, making it a highly competent model for the nuanced requirements of autonomous driving.

```
Test set: Average loss: 0.0985, Accuracy: 3619323/4161536 (87%)
Precision: 0.7547, Recall: 0.6360, F1 Score: 0.6514
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.95      0.95   1045043
           1       0.73      0.77      0.75    253891
           2       0.69      0.69      0.69     13090
           3       0.57      0.67      0.61     60359
           4       0.76      0.76      0.76     40470
           5       0.75      0.74      0.74     27833
           6       0.74      0.88      0.80     54774
           7       0.42      0.27      0.33      1481
           8       0.86      0.90      0.88     59337
           9       0.91      0.88      0.90     93971
          10       0.65      0.85      0.74     47646
          11       0.93      0.70      0.80     14775
          12       0.61      0.32      0.42     16388
          13       0.89      0.81      0.85     59295
          14       0.67      0.46      0.54     56660
          15       0.46      0.40      0.43     26998
          16       0.72      0.55      0.62      6210
          17       0.70      0.59      0.65     67534
          18       0.62      0.38      0.47      3688
...
    accuracy                           0.87   4161536
   macro avg       0.75      0.64      0.65   4161536
weighted avg       0.87      0.87      0.87   4161536
```

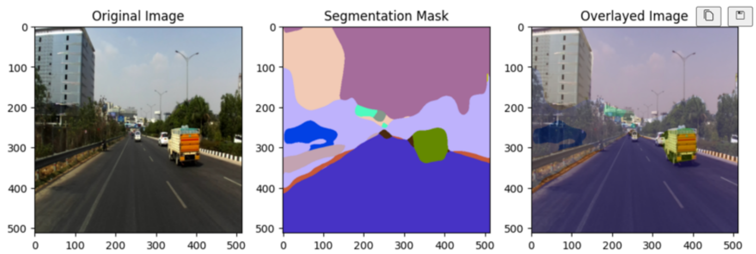


VI. Conclusions

All three models – U-Net, SegNet, and DeeplabV3 – exhibit strengths in segmenting road conditions. However, there are disparities in recognition rates across classes in SegNet, hinting at possible class imbalance issues. DeeplabV3, with its higher accuracy and effective segmentation capabilities, shows promise in overcoming some limitations faced by the other models.

In our latest research, we have introduced a new ResNet model, and all models—including SegNet and DeepLabV3—have outperformed our baseline U-Net model. The ResNet model has achieved an accuracy rate of 82% on the same dataset. These results not only affirm the superiority of our models but also pave new pathways for the advancement of image segmentation technology in the realm of autonomous driving.

```
Test set: Average loss: 0.1415, Accuracy: 3406213/4161536 (82%)
Precision: 0.6698, Recall: 0.4799, F1 Score: 0.5007
Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.96      0.94   1045043
           1       0.75      0.67      0.71    253891
           2       0.50      0.57      0.53     13090
           3       0.63      0.43      0.51     60359
           4       0.60      0.49      0.54     40470
           5       0.53      0.53      0.53     27833
           6       0.69      0.62      0.65     54774
           7       1.00      0.00      0.00      1481
           8       0.85      0.72      0.78     59337
           9       0.80      0.83      0.81     93971
          10       0.69      0.59      0.63     47646
          11       0.51      0.52      0.51     14775
          12       0.35      0.07      0.12     16388
          13       0.73      0.81      0.77     59295
          14       0.38      0.51      0.43     56660
          15       0.42      0.17      0.24     26998
          16       0.45      0.28      0.35      6210
          17       0.60      0.42      0.49     67534
          18       0.54      0.10      0.17      3688
...
    accuracy                           0.82   4161536
   macro avg       0.67      0.48      0.50   4161536
weighted avg       0.81      0.82      0.81   4161536
```



Future enhancements for these models could include the integration of attention mechanisms and multi-modal sensor fusion to refine segmentation further and increase robustness, particularly in challenging scenarios involving occluded or non-rigid objects. Adapting learning rates and incorporating early stopping could mitigate overfitting and improve validation performance for SegNet [7].

By addressing these aspects, we can expect improved accuracy and efficiency from all models, thereby inching closer to the stringent accuracy requirements of autonomous driving perception systems. The potential for the three models in the autonomous driving landscape remains significant, with continuous advancements in model architecture and training strategies promising to overcome current limitations [8].

References

[1] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 234-241). Springer, Cham. DOI:10.1007/978-3-319-24574-4_28.

[2] 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. DOI:10.1109/TPAMI.2016.2644615.

[3] Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv preprint arXiv:1706.05587.

[4] Wang, W., et al. (2023). An Improved Deeplabv3+ Model for Semantic Segmentation of Urban Environments Targeting Autonomous Driving. International Journal of Computers, Communications & Control (IJCCC), 18(6). https://doi.org/10.15837/ijccc.2023.6.5879.

[5] Zhang, H., et al. (2023). Research on Object Detection and Recognition Method for UAV Aerial Images Based on Improved YOLOv5. Drones, 7(6), 402. https://doi.org/10.3390/drones7060402.

[6] Ghasemieh, A., et al. (2022). 3D object detection for autonomous driving: Methods, models, sensors, data, and challenges. Transportation Engineering. https://doi.org/10.1016/j.treng.2022.100115.

[7] Amosa, T. I., et al. (2023). Multi-camera multi-object tracking: A review of current trends and future advances. Neurocomputing, 126558. https://doi.org/10.1016/j.neucom.2023.126558.

[8] Elallid, B. B., et al. (2022). A Comprehensive Survey on the Application of Deep and Reinforcement Learning Approaches in Autonomous Driving. Journal of King Saud University – Computer and Information Sciences, 34(9), 7366-7390. https://doi.org/10.1016/j.jksuci.2022.09.004.