Summer Internship Report

Chinni Jyothi Prakash
Student ID: 202111020
Organization: Indian Institute of Technology Tirupati
Mentor: Dr. Chalavadi Vishnu
Assistant Professor, Department of CSE
Indian Institute of Technology Tirupati

***Abstract*—This report is a summary of the work I have done during my summer internship at IIT Tirupari under the guidance of Dr. Chalavadi Vishnu. The main focus of the project is to use semantic segmentation in self-driving vehicles using computer vision. To achieve that, we started learning the basics of image segmentation, semantic segmentation, domain generalization, and a lot of concepts related to it. After that, we implemented a model that shows a validation loss of 0.1484 and a validation accuracy of 0.9420. We used U-net architecture for implementing the model. The complete information about the tasks assigned, work I carried out, challenges I faced and their solutions, my experience, and future scope are discussed in this report.**

I. Introduction

THE project work started by learning the basics required for the main project, i.e., the basics of image segmentation and it's types. I studied the differences between image classification and image segmentation. With this basic knowledge, I started implementing them on basic sample images. After understanding them better, I had the opportunity to study a research paper on "Domain Generalization of 3D Semantic Segmentation in Self-Driving Vehicles." I learned about LIDAR and its uses, the 3DLabelProp algorithm, and challenges faced in domain generalization through the research paper. Understanding the main problem statement and its necessity, we started implementing a U-Net architecture model for performing semantic segmentation in self-driving vehicles. We achieved a validation accuracy of 0.9420 that shows significant results. We studied other outcomes of the model execution, and optimizing the model is our future goal of the project.

II. Tasks Assigned

During the internship, I was assigned to some important tasks, which helped me understand the project better. The main tasks were:

1) Initially, I was assigned to explore the articles on image segmentation, image classification, and their differences. The purpose of initially assigned tasks is to understand the basics of the project and visualize their use in real-life scenarios.
2) I was assigned to explore a research paper on domain generalization of 3D semantic segmentation in self-driving vehicles. The main purpose of this task is to understand the latest technologies used in this field and the already existing algorithms for this problem.
3) The main task of this project is to implement a working U-Net architecture model on semantic segmentation. The purpose of this task is to implement a working model on real-life data.

III. Work Carried Out

During the internship, to complete the tasks assigned to me, I carried out the following work:

1) **Exploring Image Segmentation articles:** To understand the basics of the project, I explored some articles on image segmentation. I understood the main differences between image classification and image segmentation. I learned about the segmentation types and implemented them on a sample image. I learned about threshold segmentation, edge-based segmentation and region-based segmentation. Some of the results are shown in the Figure 2 and Figure 3.
2) **Studying a research paper on Domain generalization:** After gaining the required basic knowledge, I explored a research paper on "Domain Generalization of 3D Semantic Segmentation in Self-Driving Vehicles." I studied the use of LIDAR (light detection and ranging) in this field. I learned about the challenges we will face during the segmentation, i.e., domain shifts. I explored different types of domain shifts like appearance shift, scene shift, and sensor shift and how they affect the segmentation process. After understanding the main issue, I studied the algorithm proposed in the paper, i.e., the 3D LabelProp Algorithm. It mainly focuses on propagating the labels from the previous data to the current scene and making the segmentation process easier. I studied about the different types of datasets like SemanticKITTI, SemanticKITTI32, SemanticPOSS, etc. to study the effects of domain shifts in the segmentation process. Upon performing this task, I understood

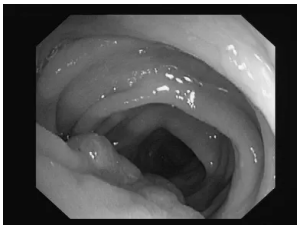


Fig. 1: Original Picture

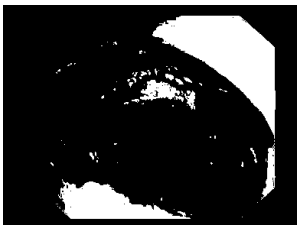


Fig. 2: Thresholding

Fig. 3: Edge-based segmentation

the already available technologies and solutions to the problem and the areas on which we have to focus for a better result.

3) **Implemented Semantic Segmentation Using U-Net:** The main goal of the project is to implement a working model on real-life data with good accuracy. To achieve this, we worked on a dataset of 2500 sample images and observed the outcomes.

A. Dataset

We worked on a dataset that contains 5 different types of image data. We took 500 random images from each type, i.e., a total of 2500 images, and fed them into the model. There are 13 classes in our dataset. We fed the data with a batch size of 32 into the model. As we are considering RGB images, there are 3 channels in each image. We split the dataset into the training dataset (80%) and validation dataset (20%). We stored the images and masks in separate lists and generated the mask images. A sample is shown in the Figure 4.

B. Model

We have used the U-Net architecture model. The model consists of 4 encoding layers and 4 decoding layers. Each encoding layer is followed by a max-pooling layer, and the decoding layer

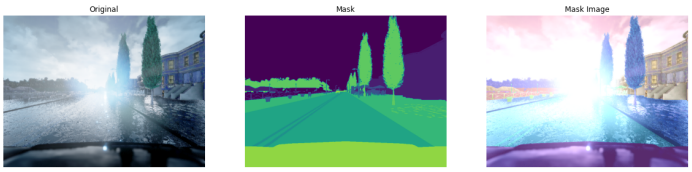


Fig. 4: Original Image-Mask-Mask Image

is followed by upsampling. Each convolutional block in both the encoding and decoding layers is followed by a LeakyReLU activation function.

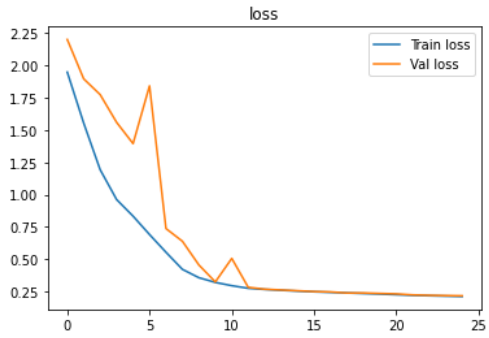
C. Results and Observations

On training and validating the model, we have achieved a validation accuracy of 0.9420 and a validation loss of 0.1484. The comparison between the training and validating accuracies and losses is shown in Figure 5b and Figure 5a. We have run the model for 25 epochs. Initially, it showed a huge difference between training and validation results. But, eventually, at the end, we achieved similar results for both training and validation with good accuracy and minimum loss.

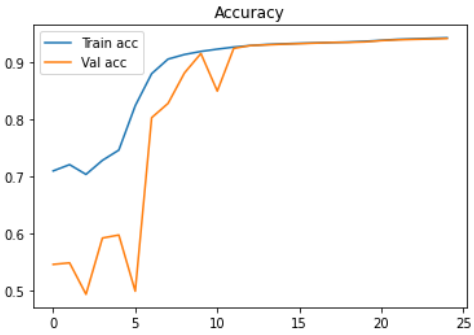
Epoch	Training Accuracy	Validation Accuracy
1	0.7094	0.5453
2	0.7204	0.5479
3	0.7033	0.4928
4	0.7279	0.5916
5	0.7459	0.5969
6	0.8237	0.4982
7	0.9056	0.8279
8	0.9137	0.8814
9	0.9191	0.9154
10	0.9231	0.8495
11	0.9267	0.9244
12	0.9294	0.9294
13	0.9313	0.9306
14	0.9325	0.9319
15	0.9335	0.9327
16	0.9341	0.9338
17	0.9349	0.9348
18	0.9357	0.9351
19	0.9369	0.9359
20	0.9386	0.9381
21	0.9404	0.9395
22	0.9412	0.9403
23	0.9421	0.9410
24	0.9430	0.9418
25	0.9435	0.9420

TABLE I: Training and Validation Accuracy Over 25 Epochs

4) **Federated Image Segmentation:** I learned about a new method of implementing deep learning models, i.e., federated image segmentation. The main purpose of this process is to overcome the lack of sensitive data in the field of medicine and other fields. It ensures data privacy and allows us to work on sensitive data also.



(a) Training Vs Validation Loss



(b) Training Vs Validation Accuracy

Fig. 5: Comparison of Training Vs Validation Loss and Accuracy

IV. Challenges and Solutions

Initially, I didn't have proper resources to gain the required knowledge about the main concepts of the project. As my mentor, Dr. Chalavadi Vishnu, provided me the necessary resources, I have gone through them and gained a good foundation on the basic topics. It helped me understand the complex concepts in a better way. After gaining the fundamental knowledge, the main challenge is to understand the complex architecture of the model and the workings behind it. A lot of new terms are introduced, and it became difficult for me to understand. A sample UNet architecture model is shown in the Figure 6. But my mentor allowed me to learn the new concepts at my own pace and gave me freedom to explore the concepts until I was confident with them. By implementing small levels of models, I gained some confidence and worked on the main model.

V. Experience

This internship gave me confidence and made me understand that, having a good foundation and proper understanding of the main problem of the work, we can approach the problem step-by-step by gaining knowledge. Even small learnings can be helpful in performing the main project. I found use of many such small concepts in the main project. I would like

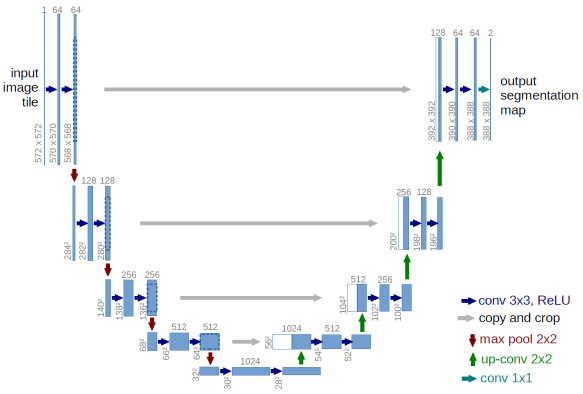


Fig. 6: U-Net architecture

to express my sincere thanks to my mentor, Dr. . Chalavadi Vishnu, for giving me freedom to explore the concepts and learn them at my own pace. He did not just provide me the resources, but he made me understand the underlying necessary concepts for implementing the project. His constant response to my doubts helped me to work better. He maintained a friendly atmosphere, which allowed me to express my own thoughts and ask any doubts without any issue.

VI. Future Scope

The current model is accurate as we expected. We will be trying to perform our project on much larger datasets and improve it based on the results. We will be working on the model to optimize it further for much better results. We are now ready to explore Federated Image Segmentation, which will help us to perform the real-life application of this project in the unavailability of some sensitive and superficial information without causing any data privacy issues. I am looking forward to exploring new concepts, algorithms, and technologies used in Federated Image Segmentation models.

Acknowledgment

I would like to sincerely thank Dr. Chalavadi Vishnu, Assistant Professor in the Department of Computer Science and Engineering at the Indian Institute of Technology Tirupati, for his invaluable guidance and support throughout my internship. I look forward to continuing to work under his guidance to further develop and improve this project.

VII. Conclusion

My internship at IIT Tirupati was a enjoyable and enriching experience. I had the opportunity to learn and explore new concepts in deep learning, particularly in the areas of image segmentation and domain generalization. The absence of work pressure allowed me the freedom to dive deep into these topics, taking the time to understand them thoroughly and implement them effectively.