

RBE549 Computer Vision : HW2 Sexy Semantic Mapping

Thabsheer Jafer Machingal
Robotics Engineering (M.S)
tmachingal@wpi.edu
Submitted on: December 16 2022

I. INTRODUCTION

Semantic segmentation is the task of classifying a certain class of image and separating it from the scene. The idea of this assignment is to draw higher level semantic information from Lidar point clouds and RGB camera images collected, this often helps in self-driving cars for mapping and localization.

II. DATASET

For this Homework, we are using KITTI-360 dataset and it is a large scale dataset with sensory information and annotations. The dataset contains images from suburbs of Karlsruhe, Germany corresponding to over 320k images and 100k laser scans in a driving distance of 73.7km[1]. A sample of the Lidar points

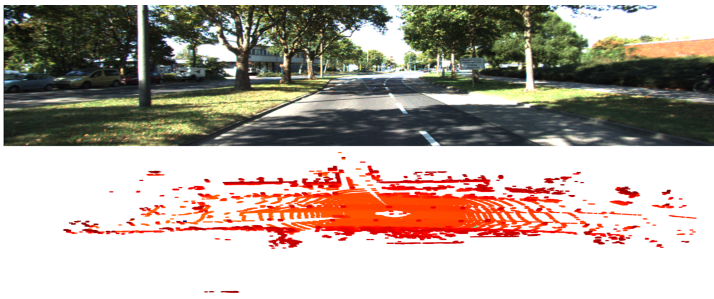


Fig. 1. An RGB image and corresponding Lidar point clouds

III. PIPELINE

The pipeline for segmenting different classes of objects in the scene is explained below. Since we are allowed to work on a third party code, I utilized this repository[2] as reference and the data from KITTI-360[1] is used. I used the following data from the dataset.

- RGB image
- Lidar data (binary file format)
- calibration of sensors (Intrinsics and extrinsics)

The pipeline is illustrated in the figure 1. Since I'm using the starter code and for this project, I'm supposed to build the function 'f' in figure 1 and is explained in the second step in the assignment is to paint the Lidar points using the semantic information from the network. below. From the dataset, we

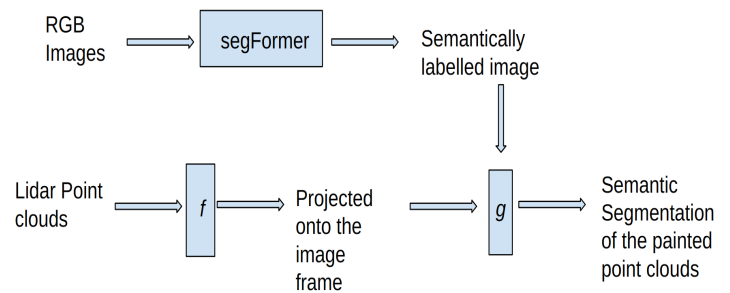


Fig. 2. The pipeline of sexy semantic segmentation

have RGB frames of the scene from the video. We are applying the pipeline frame by frame. The RGB images are fed into a neural network, here we are using a transformer network called 'segFormer' and we get the semantically labelled image out of it. The output is labelled in colors for different classes like cars, trucks, cyclists, road e.t.c. We also have the point clouds of Lidar data from the dataset and we project these Lidar point clouds onto the image and using the semantic labels, we paint the point clouds in 2D image and do that for each frame, which will yield a video of the semantically labelled Lidar data.

A. Building the Map

The first step is to build the map using the collected point cloud data from the dataset[1]. We are given the point cloud information as a binary file with x,y,z coordinates of the points in the Lidar coordinates and intensity value. I used the open3d library in python to convert the binary file into point cloud data (pcd). The next step is to project the point clouds(converted into numpy array of 'N' points) onto the image using the Lidar to camera projection matrix. The task of this assignment is to implement this part.

B. Semantic painting the points on the map

The second step in the assignment is to paint the Lidar points using the semantic information from the network. Using the semantic segmentation result from the network, the idea is to super impose these images with the 2D lidar points and create the segmented map.

The result from the pipeline should look like in figure 4



Fig. 3. The result from segFormer

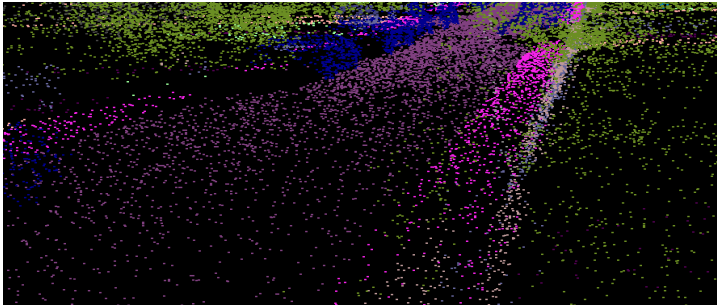


Fig. 4. The result from the pipeline

IV. RESULT

I implemented the functions as mentioned in the section "pipeline" and the results were different from expected.

REFERENCES

- [1] "KITTI-360." <https://www.cvlibs.net/datasets/kitti-360/download.php> (accessed Dec. 16, 2022).
- [2] Rajyaguru, Nov 15. <https://github.com/naitri/PointPainting>