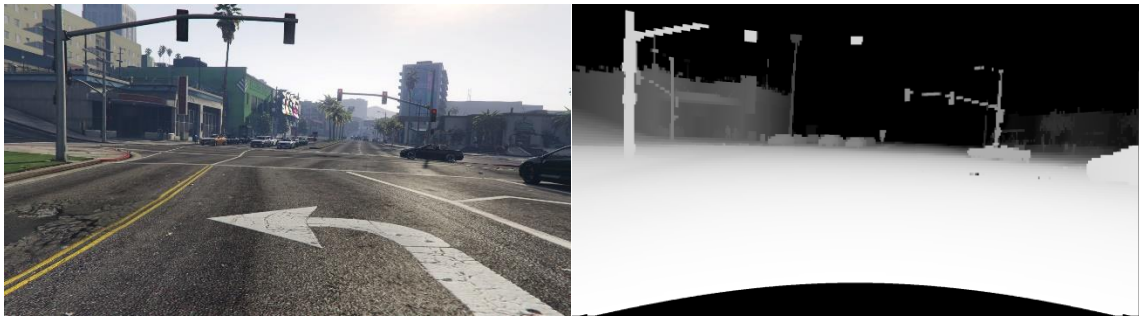
ROB 599 Preception Report

Team 13: DRIVERUNKNOWN'S RACINGGROUNDS

1. Data pre-processing

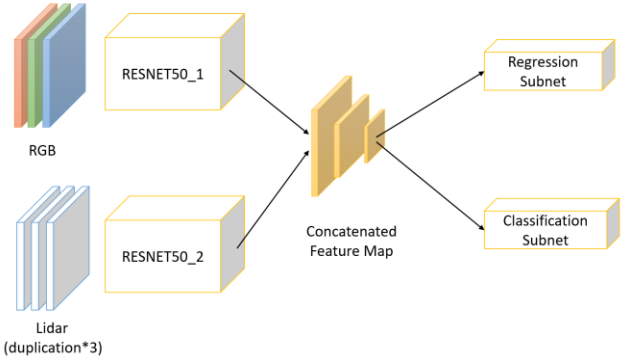
To use lidar information together with RGB info inside image, the point clouds are preprocessed into grayscale images, where each pixel denotes the distance between the 3D point and origin. In order to render the depth image continuously, the affected area of every point is calculated based on the distance and the horizontal and vertical resolution of lidar, and applied onto a blank image in order of distance decending. Here's an example:



Also, we need 2D bbox instead of 3D bbox for our network, thus transformation is necessary. The 2D bbox is given by the bounding box of the projected 3d bbox on image plane.

2. Basic model

Referring to Keras RetinaNet, which is designed for drawing 2D bounding box for objects in given RGB image, we developed an enhanced network with Lidar input. In order to fuse LIDAR information into the RetinaNet, LIDAR front view is duplicated into 3 channels as input to the Resnet50. There will be four different scales of feature output for both RGB image and Lidar image. Then the two feature map are concatenated at the same depth and cascaded into the two subnets to get the final result.



3. Repo

The zip file of repo will be uploaded with the report. The main branch is **6dinput**