This is a template submission for the midterm course : 3D Object Detection (Midterm).

3D Object detection

We have used the Waymo dataset real-world data and used 3d point cloud for lidar based object detection.

- Configuring the ranges channel to 8 bit and view the range /intensity image (ID_S1_EX1)
- Use the Open3D library to display the lidar point cloud on a 3d viewer and identifying 6 images from point cloud.(ID_S1_EX2)
- Create Birds Eye View perspective (BEV) of the point cloud,assign lidar intensity values to BEV,normalize the heightmap of each BEV (ID_S2_EX1,ID_S2_EX2,ID_S2_EX3)
- In addition to YOLO, use the repository [https://github.com/maudzung/SFA3D.git]and add parameters ,instantiate fpn resnet model(ID_S3_EX1)
- Convert BEV coordinates into pixel coordinates and convert model output to bounding box format (ID_S3_EX2)
- Compute intersection over union, assign detected objects to label if IOU exceeds threshold (ID_S4_EX1)
- Compute false positives and false negatives, precision and recall(ID_S4_EX2,ID_S4_EX3)

The project can be run by running

python loop_over_dataset.py

All training/inference is done on udacity workspace.

Step-1: Compute Lidar point cloud from Range Image

In this we are first previewing the range image and convert range and intensity channels to 8 bit format. After that, we use the openCV library to stack the range and intensity channel vertically to visualize the image.

- Convert "range" channel to 8 bit
- Convert "intensity" channel to 8 bit
- Crop range image to +/- 90 degrees left and right of forward facing x axis
- Stack up range and intensity channels vertically in openCV

The changes are made in 'loop_over_dataset.py'

```
## Set parameters and perform initializations

## Select Waymo Open Dataset file and frame numbers
# data_filename = 'training_segment-1005081002024129653_5313_150_5333_150_with_camera_labels.tfrecord' # Sequence 1
# data_filename = 'training_segment-10072231702153043603_5725_000_5745_000_with_camera_labels.tfrecord' # Sequence 2
data_filename = 'training_segment-10963653239323173269_1924_000_1944_000_with_camera_labels.tfrecord' # Sequence 3
show_only_frames = [1,10] # show only frames in interval for debugging

## Prepare Waymo Open Dataset file for loading
data_fullpath = os.path.join(os.path.dirname(os.path.realpath(__file__)), 'dataset', data_filename) # adjustable path
called from another working directory
model = "darknet"
sequence = "3"
results_fullpath = os.path.join(os.path.dirname(os.path.realpath(__file__)), 'results/' + model + '/results_sequence_'
model)
```

misc
results
student
tools
.student_bashrc
DepthCamera_2022-01-10-17-!
DepthCapture_2022-01-10-17-
LICENSE.md
README.md

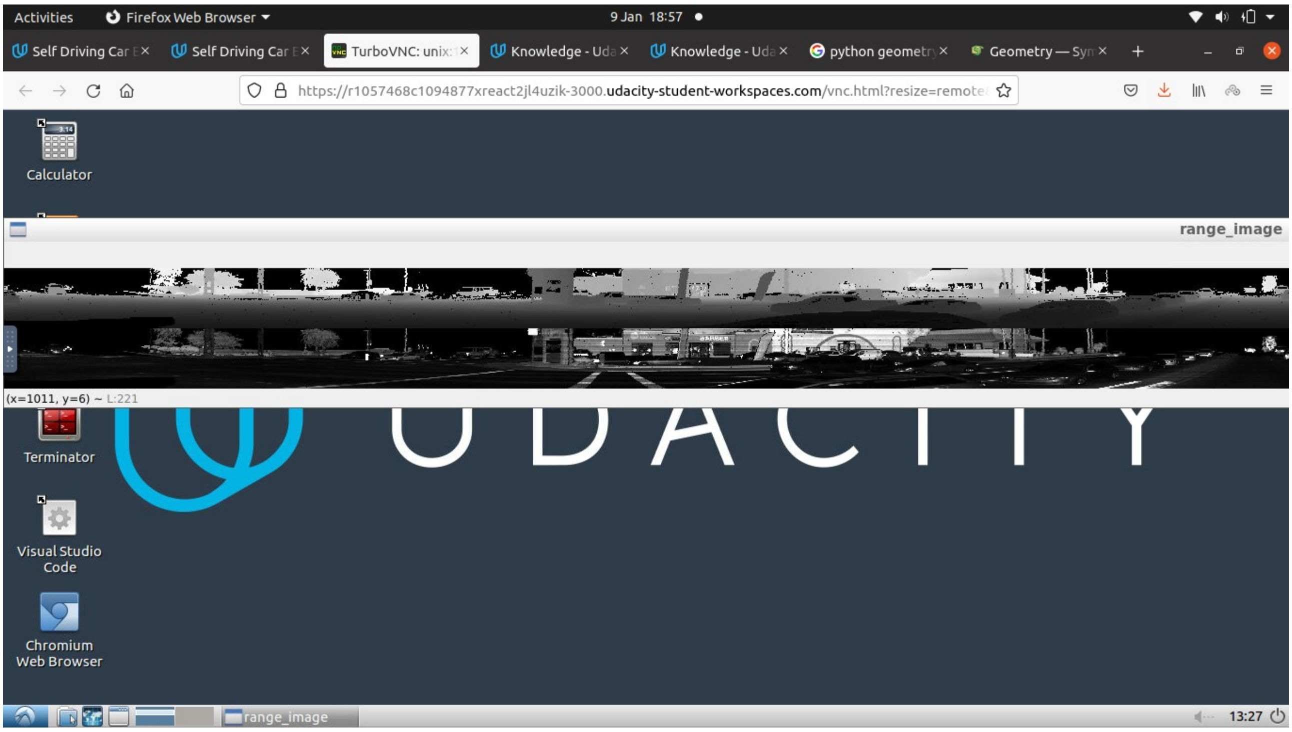
```
np.random.seed(10) # make random values predictable

## Selective execution and visualization
exe_data=[]  #option 'pcl_from_rangeimage','load_image'
exec_detection = [] # options are 'bev_from_pcl', 'detect_objects', 'validate_object_labels', 'measure_detection_performance'; options not in the list will be loaded from file
exec_tracking = [] # options are 'perform_tracking'
exec_visualization = ['show_range_image'] # options are 'show_range_image', 'show_bev', 'show_pcl', 'show_labels_in_image', 'show_objects_and_labels_in_bev', 'show_objects_in_bev_labels_in_camera', 'show_tracks', 'show_detection_performance', 'make_tracking_movie'
exec_list = make_exec_list(exec_detection, exec_tracking, exec_visualization)
vis_pause_time = 0 # set pause time between frames in ms (0 = stop between frames until key is pressed)
```

The changes are made in "objdet_pcl.py"

README.md × loop_over_datase... × objdet_eval.py × objdet_pcl.py ×

```
# visualize range image
def show_range_image(frame, lidar_name):

    ####### ID_S1_EX1 START #######
    #######
    print("student task ID_S1_EX1")

    # step 1 : extract lidar data and range image for the roof-mounted lidar
    lidar = [obj for obj in frame.lasers if obj.name == lidar_name][0]

    # step 2 : extract the range and the intensity channel from the range image
    ri = dataset_pb2.MatrixFloat()
    ri.ParseFromString(zlib.decompress(lidar.ri_return1.range_image_compressed))
    ri = np.array(ri.data).reshape(ri.shape.dims)

    # step 3 : set values <0 to zero
    ri[ri<0]=0.0

    # step 4 : map the range channel onto an 8-bit scale and make sure that the full range of values is appropriately considered
    ri_range = ri[:,:,0]
    ri_range = ri_range * 255 / (np.amax(ri_range) - np.amin(ri_range))
    img_range = ri_range.astype(np.uint8)

    # step 5 : map the intensity channel onto an 8-bit scale and normalize with the difference between the 1- and 99-percentile to mi
outliers
    ri_intensity = ri[:,:,1]
    percentile_1, percentile_99 = percentile(ri_intensity,1), percentile(ri_intensity,99)
    ri_intensity = 255 * np.clip(ri_intensity,percentile_1,percentile_99)/percentile_99
    img_intensity = ri_intensity.astype(np.uint8)

    # step 6 : stack the range and intensity image vertically using np.vstack and convert the result to an unsigned 8-bit integer
    img_range_intensity = np.vstack((img_range,img_intensity))
    img_range_intensity = img_range_intensity.astype(np.uint8)
```

The range image sample:



For the next part, we use the Open3D library to display the lidar point cloud on a 3D viewer and identify 10 images from point cloud

- Visualize the point cloud in Open3D
- 10 examples from point cloud with varying degrees of visibility

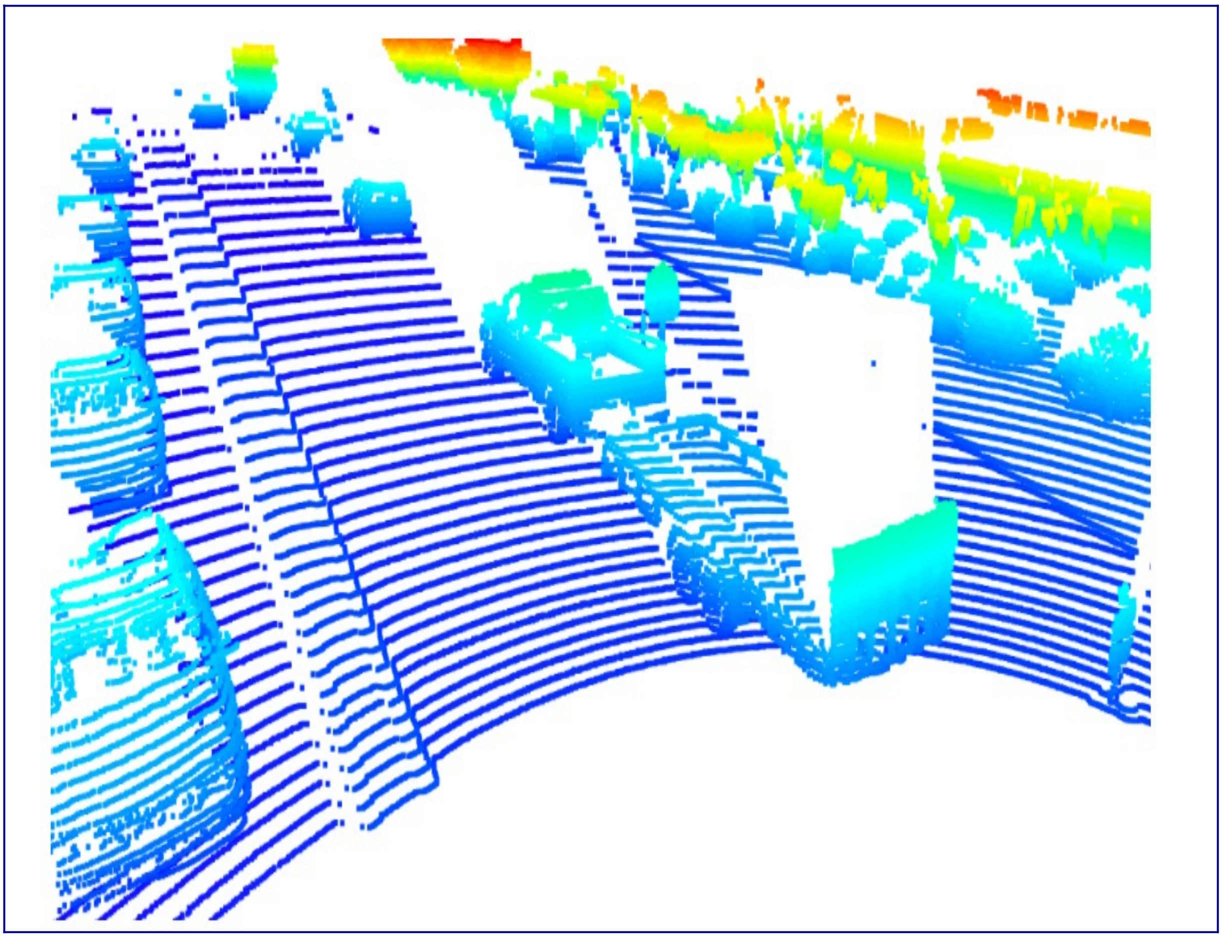
The changes are made in 'loop_over_dataset.py'

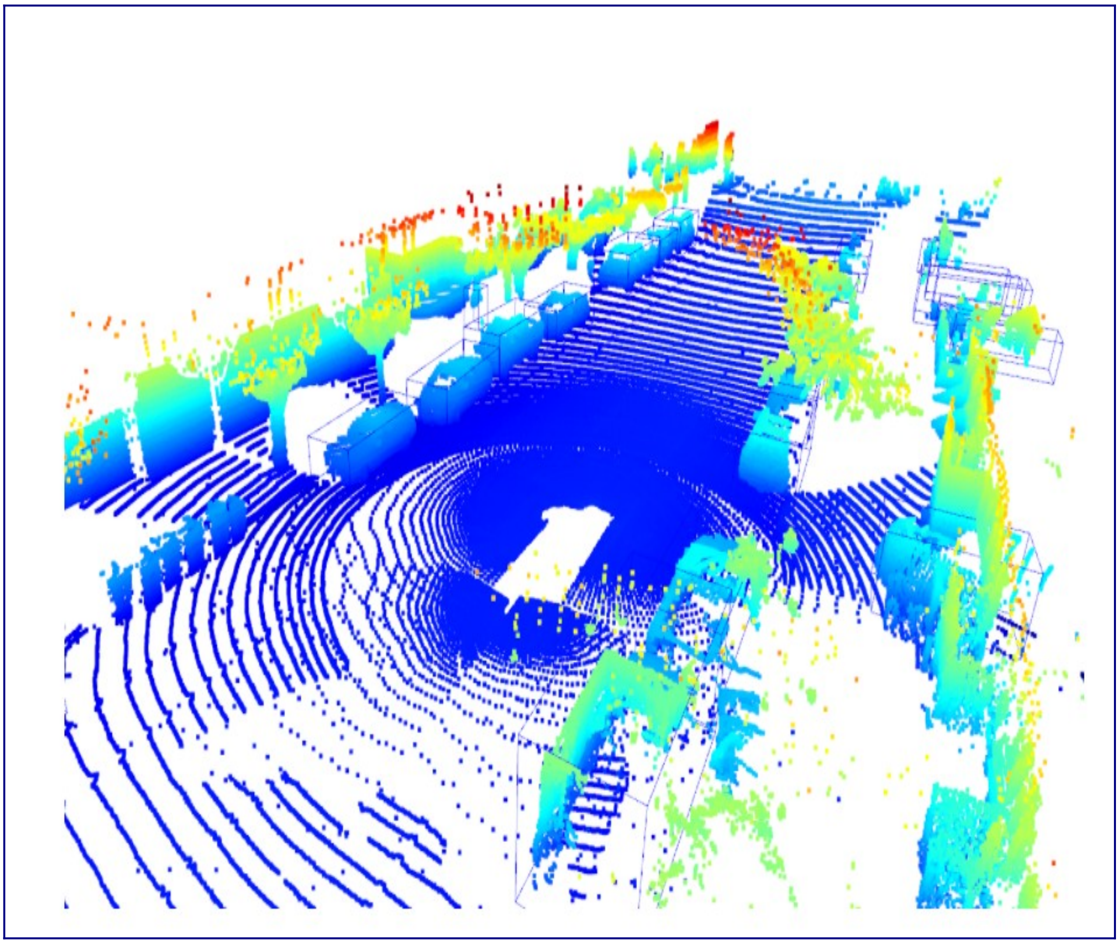
```
3D Object Detection Workspace
README.md    loop_over_datas...    objdet_pcl.py

## initialize tracking
KF = Filter() # set up Kalman filter
association = Association() # init data association
manager = Trackmanagement() # init track manager
lidar = None # init lidar sensor object
camera = None # init camera sensor object
np.random.seed(10) # make random values predictable

## Selective execution and visualization
exe_data=[]  #option 'pcl_from rangeimage','load_image'
exec_detection = [] # options are 'bev_from_pcl', 'detect_objects', 'validate_object_labels', 'measure_detect
not in the list will be loaded from file
exec_tracking = [] # options are 'perform_tracking'
exec_visualization = ['show_pcl'] # options are 'show_range_image', 'show_bev', 'show_pcl', 'show_labels_in_i
'show_objects_and_labels_in_bev', 'show_objects_in_bev_labels_in_camera', 'show_tracks', 'show_detection_perf
'make_tracking_movie'
exec_list = make_exec_list(exec_detection, exec_tracking, exec_visualization)
vis_pause_time = 0 # set pause time between frames in ms (0 = stop between frames until key is pressed)
```

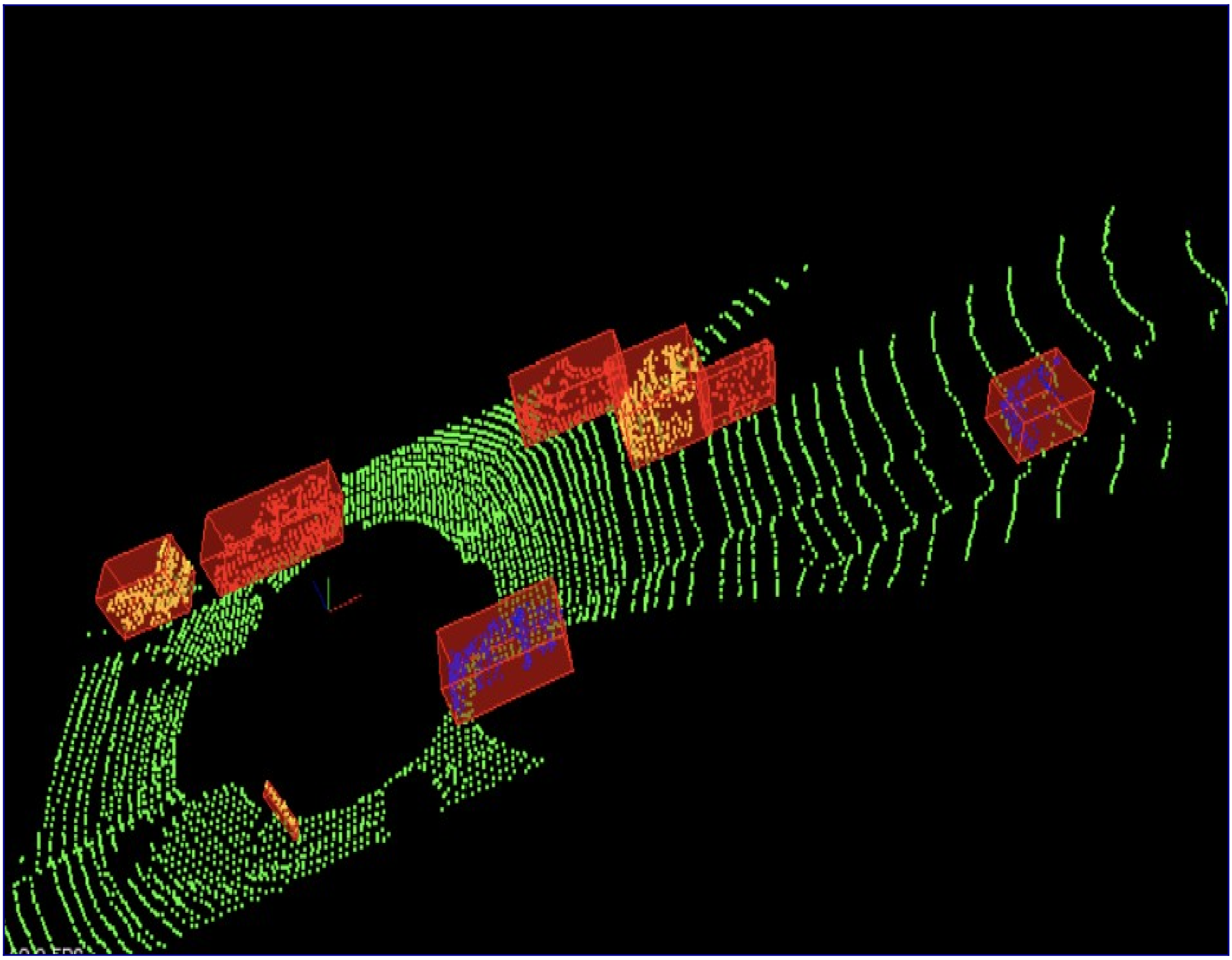
The changes are made in "objdet_pcl.py"

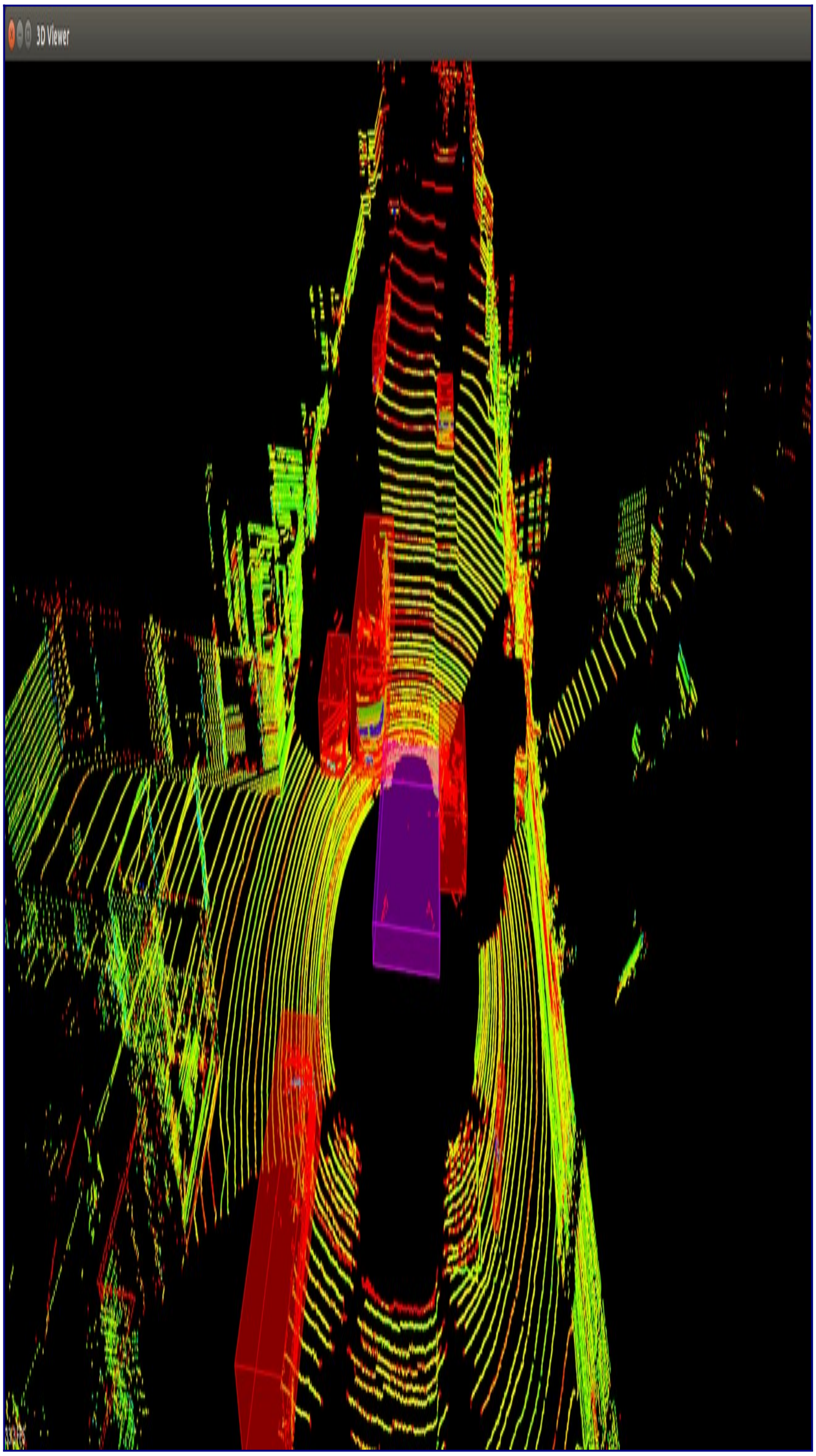
```
README.md    loop_over_datase...    objdet_eval.py    objdet_pcl.py

# visualize lidar point-cloud
def show_pcl(pcl):

    ####### ID_S1_EX2 START #######
    #######
    print("student task ID_S1_EX2")

    # step 1 : initialize open3d with key callback and create window
    viz = o3d.visualization.VisualizerWithKeyCallback()
    viz.create_window(window_name='Open3D', width=1280, height=1080, left=50, top=50, visible=True)

    global ptr
    ptr= True
    def Window(viz):
        global ptr
        print('right arrow pressed')
        ptr= False
        return
    viz.register_key_callback(262,Window)

    # step 2 : create instance of open3d point-cloud class
    pcd = o3d.geometry.PointCloud()

    # step 3 : set points in pcd instance by converting the point-cloud into 3d vectors (using open3d function Vector3dVector)
    pcd.points = o3d.utility.Vector3dVector(pcl[:,:3])

    # step 4 : for the first frame, add the pcd instance to visualization using add_geometry; for all other frames, use update_geomet
    viz.add_geometry(pcd)

    # step 5 : visualize point cloud and keep window open until right-arrow is pressed (key-code 262)
    while ptr:
        viz.poll_events()
        viz.update_renderer()
```

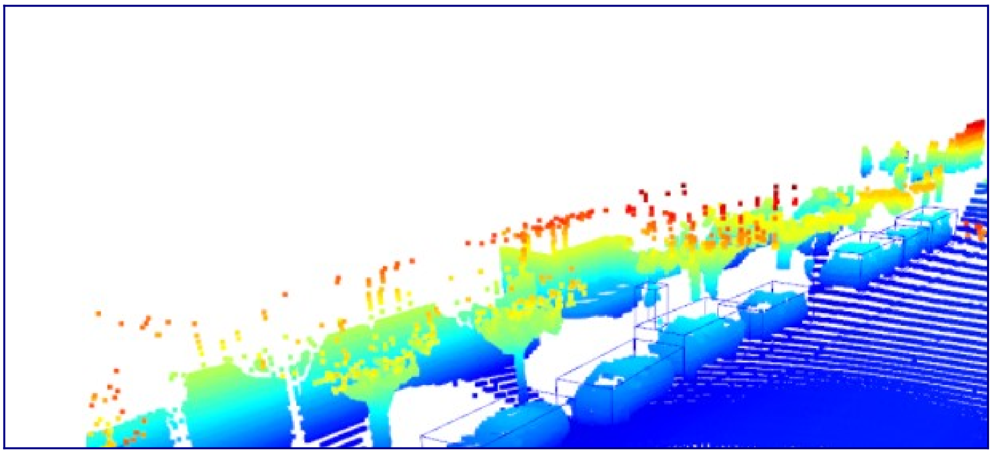
Point cloud images

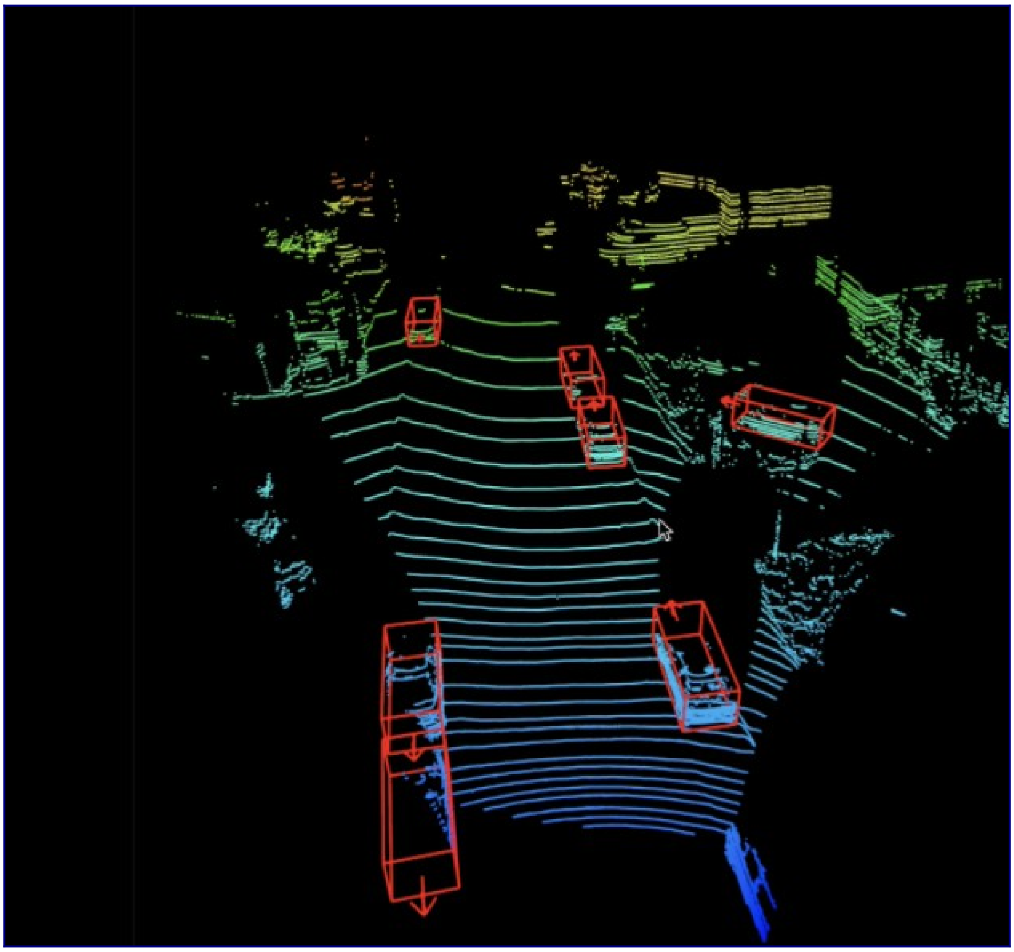


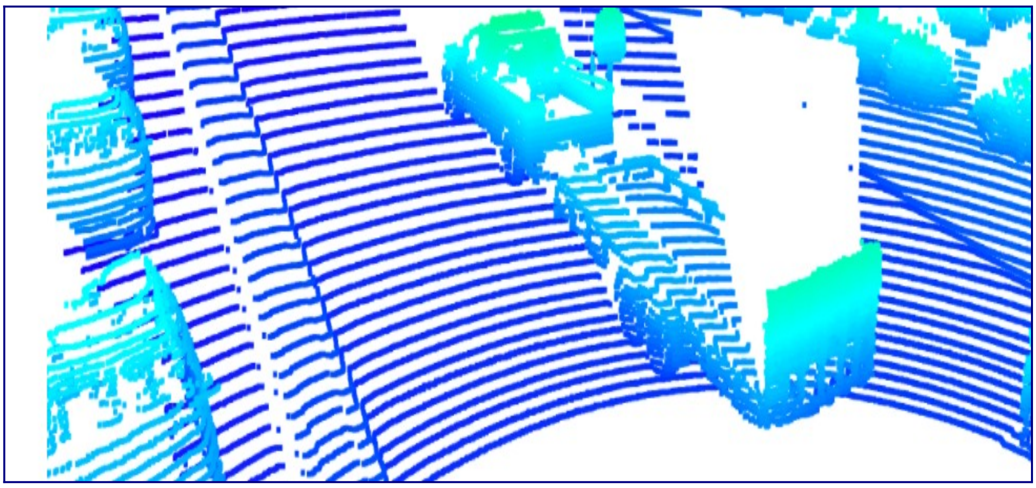


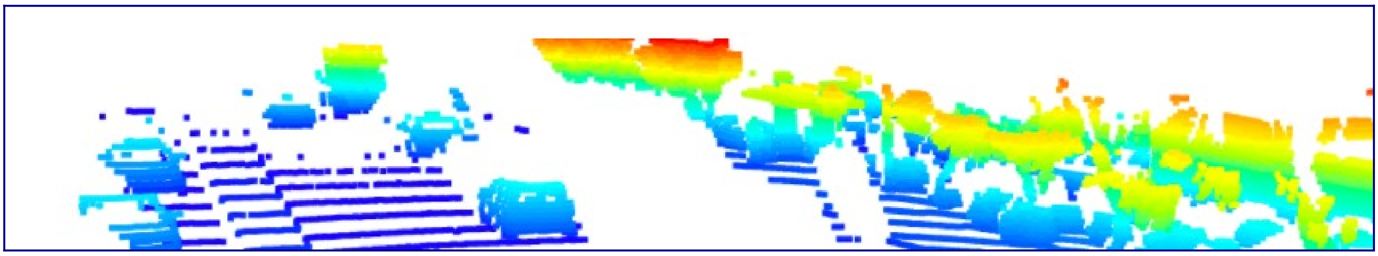



3D Viewer









Stable features include the tail lights, the rear bumper majorly. In some cases the additional features include the headover lights, car front lights, rear window shields. These are identified through the intensity channels . The chassis of the car is the most prominent identifiable feature from the lidar perspective. The images are analysed with different settings and the rear lights are the major stable components, also the bounding boxes are correctly assigned to the cars (used from Step-3).

Step-2: Creaate BEV from Lidar PCL

In this case, we are:

- Converting the coordinates to pixel values
- Assigning lidar intensity values to the birds eye view BEV mapping
- Using sorted and pruned point cloud lidar from the previous task
- Normalizing the height map in the BEV
- Compute and map the intensity values

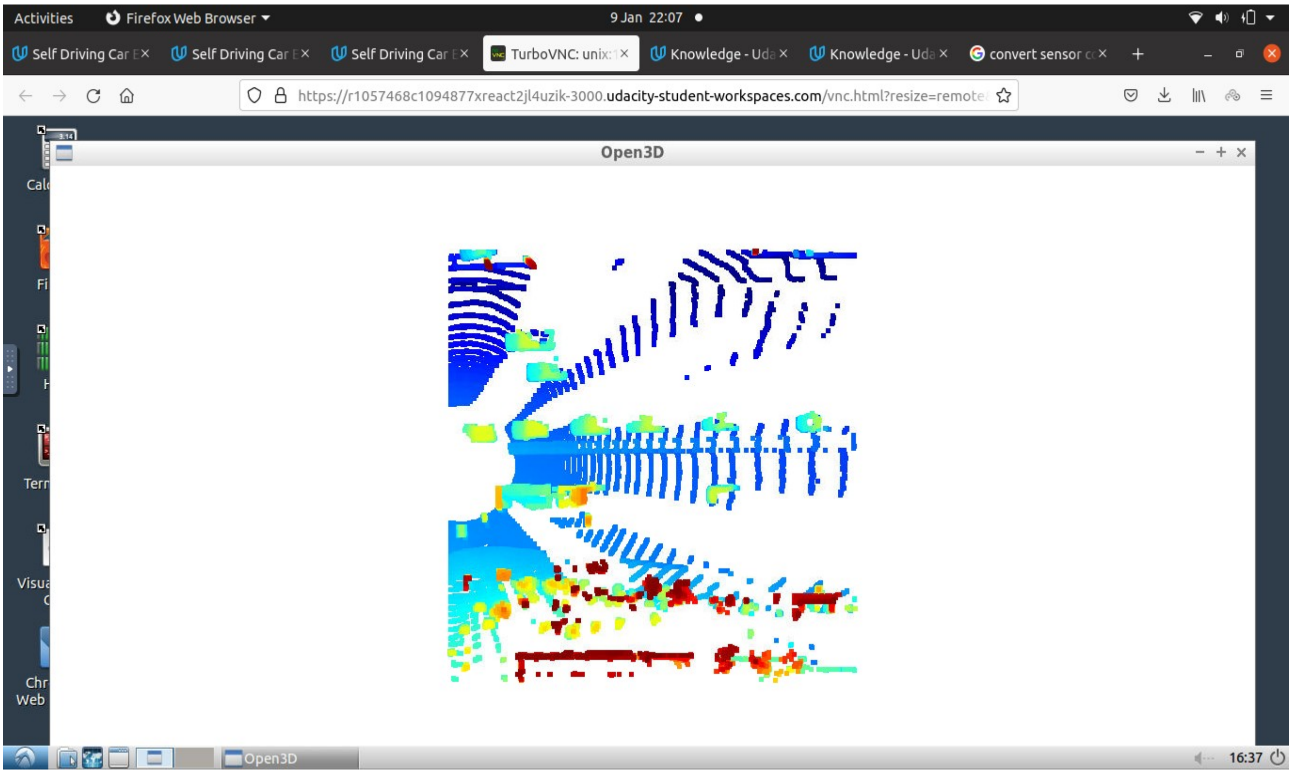
The changes are in the 'loop_over_dataset.py'

```
## Selective execution and visualization
exe_data=['pcl_from_rangeimage']  #option 'pcl_from_rangeimage','load_image'
exec_detection = ['bev_from_pcl'] # options are 'bev_from_pcl', 'detect_objects', 'validate_object_labels', 'mea
loaded from file
exec_tracking = [] # options are 'perform_tracking'
exec_visualization = [] # options are 'show_range_image', 'show_bev', 'show_pcl', 'show_labels_in_image', 'show_
'show_objects_in_bev_labels_in_camera', 'show_tracks', 'show_detection_performance', 'make_tracking_movie'
exec_list = make_exec_list(exec_detection, exec_tracking, exec_visualization)
vis_pause_time = 0 # set pause time between frames in ms (0 = stop between frames until key is pressed)
```

The changes are also in the "objdet_pcl.py"

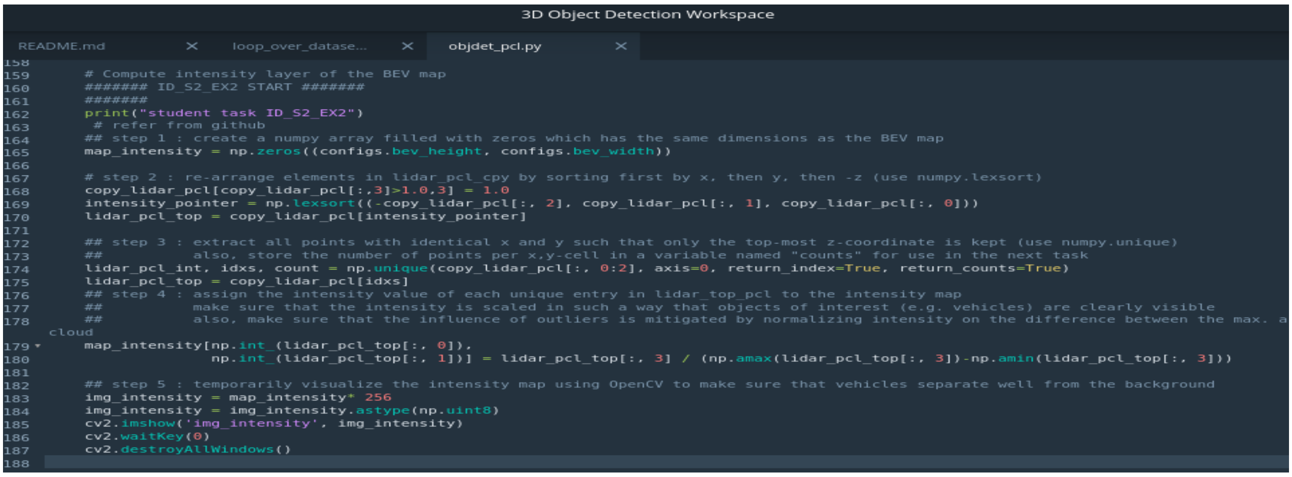
```
3D Object Detection Workspace
README.md    loop_over_datase...    objdet_pcl.py

# create birds-eye view of lidar data
def bev_from_pcl(lidar_pcl, configs):

    # remove lidar points outside detection area and with too low reflectivity
    mask = np.where((lidar_pcl[:, 0] >= configs.lim_x[0]) & (lidar_pcl[:, 0] <= configs.lim_x[1]) &
                    (lidar_pcl[:, 1] >= configs.lim_y[0]) & (lidar_pcl[:, 1] <= configs.lim_y[1]) &
                    (lidar_pcl[:, 2] >= configs.lim_z[0]) & (lidar_pcl[:, 2] <= configs.lim_z[1]))
    lidar_pcl = lidar_pcl[mask]

    # shift level of ground plane to avoid flipping from 0 to 255 for neighboring pixels
    lidar_pcl[:, 2] = lidar_pcl[:, 2] - configs.lim_z[0]

    # convert sensor coordinates to bev-map coordinates (center is bottom-middle)
    ####### ID_S2_EX1 START #######
    #######
    print("student task ID_S2_EX1")


    ## step 1 :  compute bev-map discretization by dividing x-range by the bev-image height (see configs)
    bev_discret = (configs.lim_x[1] - configs.lim_x[0]) / configs.bev_height


    ## step 2 : create a copy of the lidar pcl and transform all metrix x-coordinates into bev-image coordinates
    copy_lidar_pcl = np.copy(lidar_pcl)
    copy_lidar_pcl[:, 0] = np.int_(np.floor(copy_lidar_pcl[:, 0] / bev_discret))

    # step 3 : perform the same operation as in step 2 for the y-coordinates but make sure that no negative bev-coordinates occur
    copy_lidar_pcl[:, 1] = np.int_(np.floor(copy_lidar_pcl[:, 1] / bev_discret) + (configs.bev_width + 1) / 2)
    copy_lidar_pcl[:, 1] = np.abs(copy_lidar_pcl[:,1])

    # step 4 : visualize point-cloud using the function show_pcl from a previous task
```

A sample preview of the BEV:



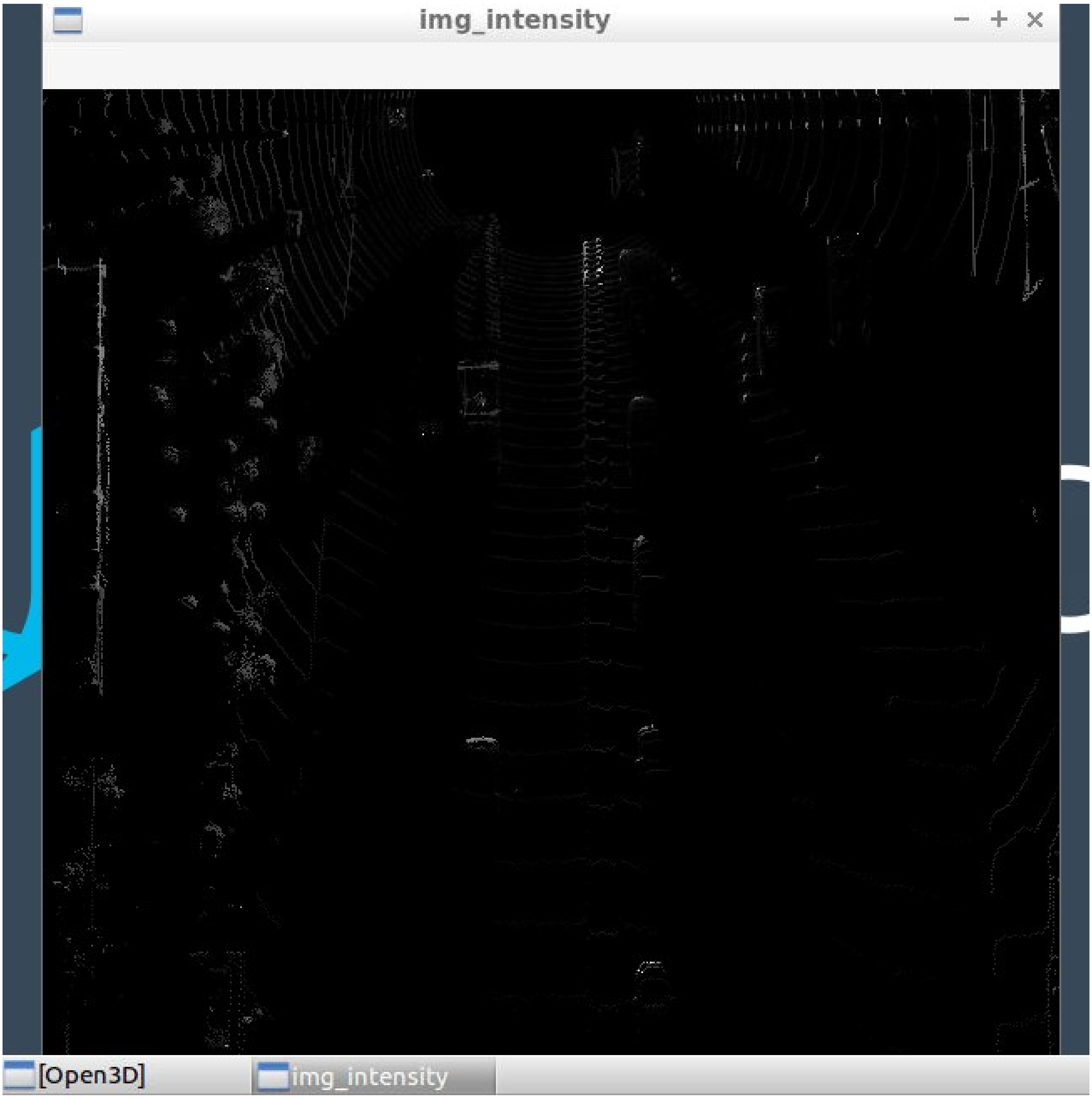
A preview of the intensity layer:

The 'lidar_pcl_top' is used in this case, shown in the Figure:

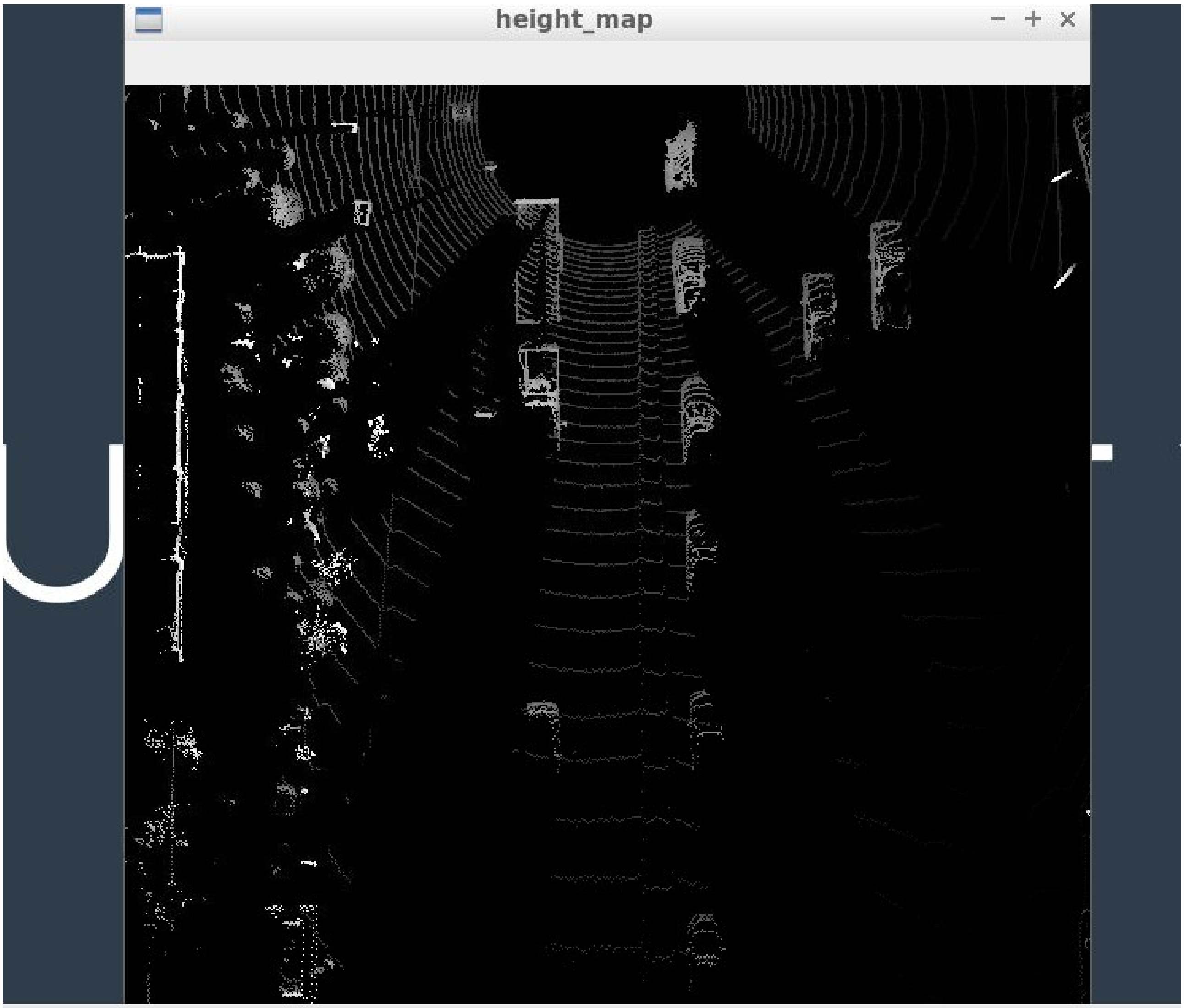



```
        # Compute intensity layer of the BEV map
        ####### ID_S2_EX2 START #######
        #######
        print("student task ID_S2_EX2")
         # refer from github
        ## step 1 : create a numpy array filled with zeros which has the same dimensions as the BEV map
        map_intensity = np.zeros((configs.bev_height, configs.bev_width))

        # step 2 : re-arrange elements in lidar_pcl_cpy by sorting first by x, then y, then -z (use numpy.lexsort)
        copy_lidar_pcl[copy_lidar_pcl[:,3]>1.0,3] = 1.0
        intensity_pointer = np.lexsort((-copy_lidar_pcl[:, 2], copy_lidar_pcl[:, 1], copy_lidar_pcl[:, 0]))
        lidar_pcl_top = copy_lidar_pcl[intensity_pointer]

        ## step 3 : extract all points with identical x and y such that only the top-most z-coordinate is kept (use numpy.unique)
        ##          also, store the number of points per x,y-cell in a variable named "counts" for use in the next task
        lidar_pcl_int, idxs, count = np.unique(copy_lidar_pcl[:, 0:2], axis=0, return_index=True, return_counts=True)
        lidar_pcl_top = copy_lidar_pcl[idxs]
        ## step 4 : assign the intensity value of each unique entry in lidar_top_pcl to the intensity map
        ##          make sure that the intensity is scaled in such a way that objects of interest (e.g. vehicles) are clearly visible
        ##          also, make sure that the influence of outliers is mitigated by normalizing intensity on the difference between the max. a
    cloud
        map_intensity[np.int_(lidar_pcl_top[:, 0]),
                      np.int_(lidar_pcl_top[:, 1])] = lidar_pcl_top[:, 3] / (np.amax(lidar_pcl_top[:, 3])-np.amin(lidar_pcl_top[:, 3]))

        ## step 5 : temporarily visualize the intensity map using OpenCV to make sure that vehicles separate well from the background
        img_intensity = map_intensity* 256
        img_intensity = img_intensity.astype(np.uint8)
        cv2.imshow('img_intensity', img_intensity)
        cv2.waitKey(0)
        cv2.destroyAllWindows()
```

The corresponding intensity channel:



The corresponding normalized height channel:



Step-3: Model Based Object Detection in BEV Image

Here, particularly the test.py file and extracting the relevant configurations from 'parse_test_configs()' and added them in the 'load_configs_model' config structure by looking into this repo [https://github.com/maudzung/SFA3D.git].

- Instantiating the fpn resnet model from the cloned repository configs
- Extracting 3d bounding boxes from the responses
- Transforming the pixel to vehicle coordinates
- Model output tuned to the bounding box format [class-id, x, y, z, h, w, l, yaw]

The changes are in "loop_over_dataset.py"

```python
## Selective execution and visualization
exe_data=['pcl_from rangeimage','load_image']  #option 'pcl_from rangeimage','load_image'
exec_detection = ['bev_from_pcl','detect_objects'] # options are 'bev_from_pcl', 'detect_objects', 'validate_object_labels', 'measure_det
the list will be loaded from file
exec_tracking = [] # options are 'perform_tracking'
exec_visualization = ['show_objects_in_bev_labels_in_camera',] # options are 'show_range_image', 'show_bev', 'show_pcl', 'show_labels_in_
'show_objects_and_labels_in_bev', 'show_objects_in_bev_labels_in_camera', 'show_tracks', 'show_detection_performance', 'make_tracking_mov
exec_list = make_exec_list(exec_detection, exec_tracking, exec_visualization)
vis_pause_time = 0 # set pause time between frames in ms (0 = stop between frames until key is pressed)
```

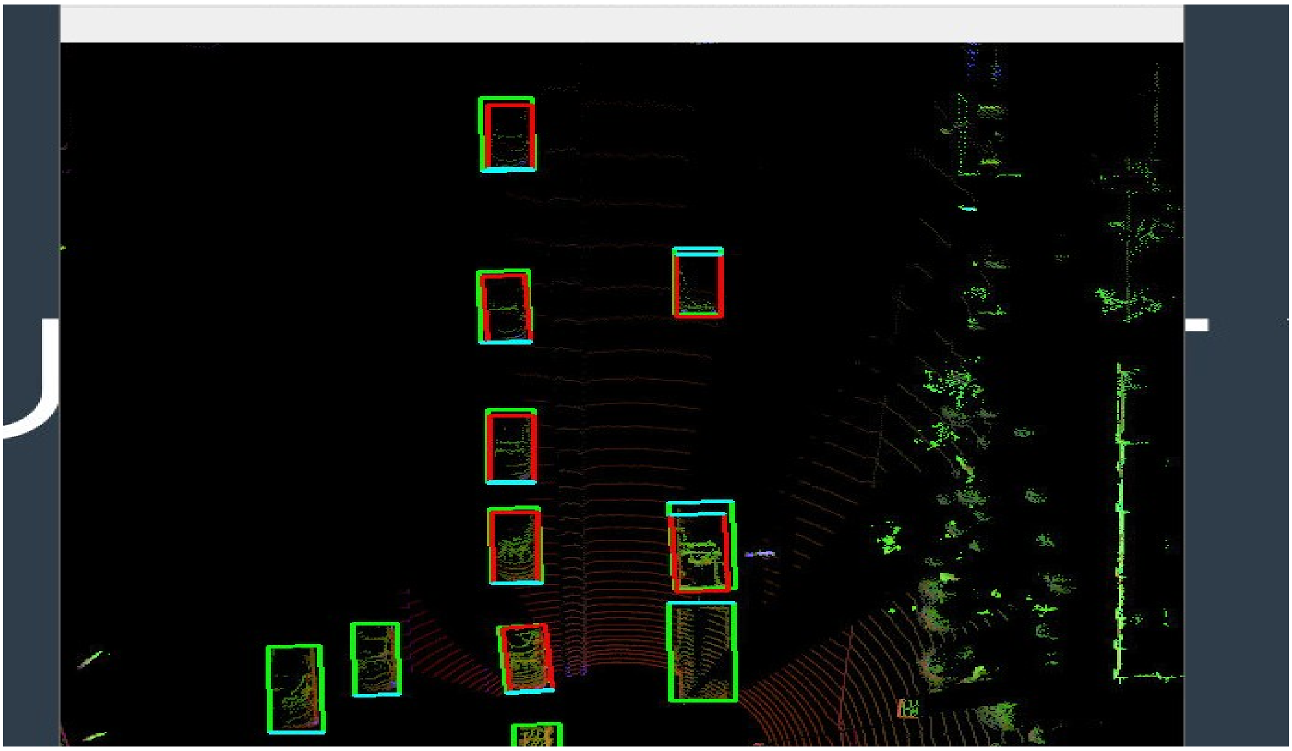
The changes for the detection are inside the "objdet_detect.py" file:

README.md × loop_over_datase... × objdet_detect.py ×

```python
# detect trained objects in birds-eye view
def detect_objects(input_bev_maps, model, configs):

    # deactivate autograd engine during test to reduce memory usage and speed up computations
    with torch.no_grad():

        # perform inference
        outputs = model(input_bev_maps)

        # decode model output into target object format
        if 'darknet' in configs.arch:

            # perform post-processing
            output_post = post_processing_v2(outputs, conf_thresh=configs.conf_thresh, nms_thresh=configs.nms_thresh)
            detections = []
            for sample_i in range(len(output_post)):
                if output_post[sample_i] is None:
                    continue
                detection = output_post[sample_i]
                for obj in detection:
                    x, y, w, l, im, re, _, _, _ = obj
                    yaw = np.arctan2(im, re)
                    detections.append([1, x, y, 0.0, 1.50, w, l, yaw])

        elif 'fpn_resnet' in configs.arch:
            # decode output and perform post-processing

            ####### ID_S3_EX1-5 START #######
            #######
            print("student task ID_S3_EX1-5")
            outputs['hm_cen'] = _sigmoid(outputs['hm_cen'])
            outputs['cen_offset'] = _sigmoid(outputs['cen_offset'])
            # detections size (batch_size, K, 10)
            detections = decode(outputs['hm_cen'], outputs['cen_offset'], outputs['direction'], outputs['z_coor'],
                                outputs['dim'], K=40) #K=configs.k
            detections = detections.cpu().numpy().astype(np.float32)
            # print(detections)
            detections = post_processing(detections, configs)
            detections = detections[0][1]
```

As the model input is a three-channel BEV map, the detected objects will be returned with coordinates and properties in the BEV coordinate space. Thus, before the detections can move along in the processing pipeline, they need to be converted into metric coordinates in vehicle space.

A sample preview of the bounding box images:



Step-4: Performance detection for 3D Object Detection

In this step, the performance is computed by getting the IOU between labels and detections to get the false positive and false negative values.The task is to compute the geometric overlap between the bounding boxes of labels and the detected objects:

- Assigning a detected object to a label if IOU exceeds threshold
- Computing the degree of geometric overlap
- For multiple matches objects/detections pair with maximum IOU are kept
- Computing the false negative and false positive values
- Computing precision and recall over the false positive and false negative values

The changes in the code are:

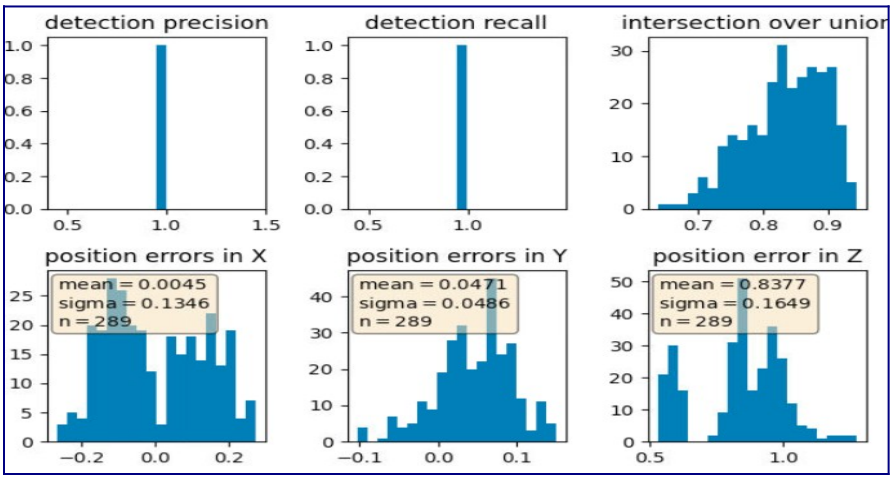
```
## Selective execution and visualization
exe_data=['pcl_from rangeimage']  #option 'pcl_from rangeimage','load_image'
exec_detection = ['bev_from_pcl','detect_objects','validate_object_labels', 'measure_detection_performance'] # option
'detect_objects', 'validate_object_labels', 'measure_detection_performance'; options not in the list will be loaded f
exec_tracking = [] # options are 'perform_tracking'
exec_visualization = ['show_detection_performance'] # options are 'show_range_image', 'show_bev', 'show_pcl', 'show_l
'show_objects_and_labels_in_bev', 'show_objects_in_bev_labels_in_camera', 'show_tracks', 'show_detection_performance'
exec_list = make_exec_list(exec_detection, exec_tracking, exec_visualization)
vis_pause_time = 0 # set pause time between frames in ms (0 = stop between frames until key is pressed)
```

The changes for "objdet_eval.py" where the precision and recall are calculated as functions of false positives and negatives:

```python
            center_devs.append(best_match[1:])


    ####### ID_S4_EX2 START #######
    #######
    print("student task ID_S4_EX2")

    # compute positives and negatives for precision/recall

    ## step 1 : compute the total number of positives present in the scene
    all_positives = labels_valid.sum()

    ## step 2 : compute the number of false negatives
    true_positives= len(ious)
    false_negatives = all_positives-true_positives

    ## step 3 : compute the number of false positives
    false_positives = len(detections)-all_positives

    #######
    ####### ID_S4_EX2 END #######

    pos_negs = [all_positives, true_positives, false_negatives, false_positives]
    det_performance = [ious, center_devs, pos_negs]

    return det_performance
```

```python
            ## step 1 : extract the four corners of the current label bounding-box
            box=label.box
            box_1=tools.compute_box_corners(box.center_x, box.center_y, box.width, box.length, box.heading)

            ## step 2 : loop over all detected objects
            for objects in detections:

                ## step 3 : extract the four corners of the current detection
                _id, x, y, z, _h, w, l, yaw= objects
                box_2=tools.compute_box_corners(x, y, w, l, yaw)

                ## step 4 : computer the center distance between label and detection bounding-box in x, y, and z
                x_dist=np.array(box.center_x-x).item()
                y_dist=np.array(box.center_y-y).item()
                z_dist=np.array(box.center_z-z).item()

                ## step 5 : compute the intersection over union (IOU) between label and detection bounding-box
                rect1=Polygon(box_1)
                rect2=Polygon(box_2)
                intersection=rect1.intersection(rect2).area
                union=rect1.union(rect2).area
                iou=intersection/union


                ## step 6 : if IOU exceeds min_iou threshold, store [iou,dist_x, dist_y, dist_z] in matches_lab_det and increase the TP

                if iou>min_iou:
                    matches_lab_det.append([iou, dist_x, dist_y, dist_z])

            #######
            ####### ID_S4_EX1 END #######
```

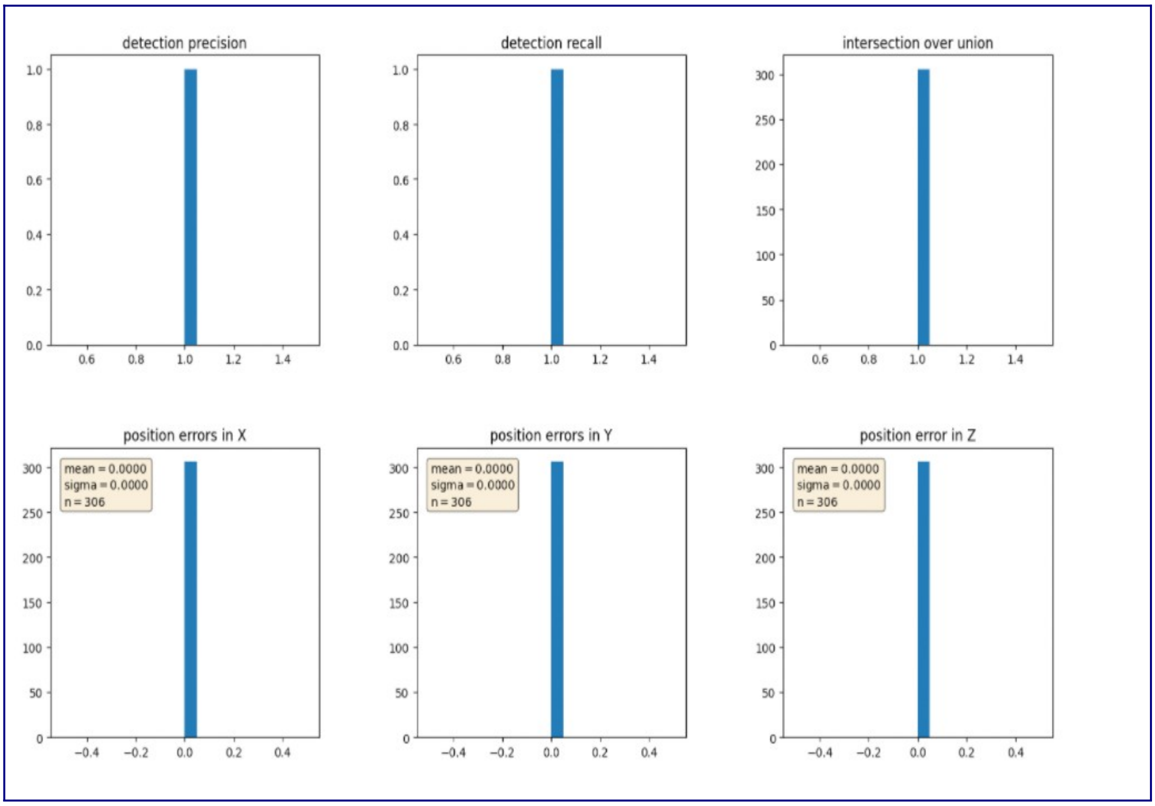
The precision recall curve is plotted showing similar results of precision =0.986 and recall=0.81372



In the next step, we set the

configs_det.use_labels_as_objects=True

which results in precision and recall values as 1.This is shown in the following image:



Summary of Lidar based 3D Object Detection

From the project, it is understandable that for a stabilized tracking, lidar should be used . The conversion of range data to point cloud through spatial volumes, or points (or CNN networks) are important for further analysis. The usage of resnet/darknet and YOLO to convert these high dimensional point cloud representations to object detections through bounding boxes is essential for 3D object detection. Evaluating the performance with help of maximal IOU mapping ,mAP, and representing the precision/recall of the bounding boxes are essential to understand the effectiveness of Lidar based detection.