Monocular 3D Object Detection

Computer Vision (CS6350)

TPA - 7

1. Problem Statement

This problem introduces the task of Monocular 3D Object Detection which is quite challenging considering that the only available input is an RGB image/sequence of frames in a video. The goal is to predict a 3D bounding box given an input image without extra information such as point clouds or Lidar Data.

2. Input

An image/video

3. Expected Output

- 3D bounding box enclosing each detected object in an image, For output demo, refer : https://github.com/dingmyu/D4LCN/blob/master/demo.gif (input is just a video/images)
- For Quantitative evaluation, report the $AP_{R_{40}}$ metric [4].

4. Dataset

- KITTI and its variants. Eg. KITTI Cars easy, KITTI Cars Moderate, etc.

Link for the dataset : http://www.cvlibs.net/datasets/kitti/eval_object.php?obj_benchmark=3d

5. References

1. Brazil, Garrick, et al. "Kinematic 3D Object Detection in Monocular Video.", ECCV 2020

2. Ding, Mingyu, et al. "Learning depth-guided convolutions for monocular 3d object detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.

3. Liu, Zechen, Zizhang Wu, and Roland Tth. "SMOKE: Single-Stage Monocular 3D Object Detection via Keypoint Estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.

4. Simonelli, Andrea, et al. "Disentangling monocular 3d object detection." Proceedings of the IEEE International Conference on Computer Vision. 2019.

5. Brazil, Garrick, and Xiaoming Liu. "M3d-rpn: Monocular 3d region proposal network for object detection." Proceedings of the IEEE International Conference on Computer Vision. 2019.