A Comparative Study of VoxelNet and PointNet for 3D Object Detection in Car by Using KITTI Benchmark

Harish S Gujjar, Department of Computer Science, SSA Government First Grade College, Ballari, India

ABSTRACT

In today’s world, 2D object recognition is a normal course of study in research. 3D objection recognition is more in demand and important in the present scenario. 3D object recognition has gained importance in areas such as navigation of vehicles, robotic vision, HoME, virtual reality, etc. This work reveals the two important methods, VoxelNet and PointNet, useful in 3D object recognition. In case of NetPoint, the recognition is good when used with segmentation of point clouds which are in small-scale. Whereas, in case of VoxelNet, scans are used directly on raw points of clouds which are directly operated on patterns. The above conclusion is arrived on KITTI car detection. The KITTI uses detection by using bird’s eye view. In this method of KITTI we compare two different methods called LiDAR and RGB-D. We arrive at a conclusion that pointNet is useful and has high performance when we are using small scenarios and VoxelNet is useful and has high performance when we are using large scenarios.

KEYWORDS

KITTI, LiDAR, PointNet, RGB-D, VoxelNet

1. INTRODUCTION

A Point Net is based on cloud for detecting 3D objects which is a very burning concept in a real-world application which includes autonomous guiding of the object (Engelcke, Rao, Wang, Tong, & Posner; 2017; Krizhevsky, Sutskever, & Hinton, 2012), robots used in housekeeping (Oh & Watanabe, 2002) and virtual reality/augmented (Park, Lepetit, & Woo, 2008). In comparison with the detection based on images, LiDAR helps in correctly localizing the objects and shapes (Li, 2017) are characterized by giving valid depth information of an object. Although, objects which are not similar, LiDAR has scattered point clouds and has different point density, because of this reason of non-uniform division of the 3D space, sensors having an effective range, pose related and occlusion.

In many applications there is a need and urge to know the 3D method such as driving autonomously and augmented reality (AR). As there is more demand for sensors which detects 3D used in the mobiles and vehicles of automatic, large amount of data is apprehended and processed. In this paper the author has studied the two methods of capturing 3D data and detection in the captured data. In the first method 3D sensors is used to record the data and classify the physical object category by boxes of 3D. 3D object detection is a burning problem, the presentation of point cloud data and the uses of deep net architectures are important. Large number of works converts 3D point cloud to an image with the help of projection (Dorai & Jain, 1997; Girshick, 2015) or through quantization of grids.
by volumetric (Riegler, Ulusoyys, & Geiger, 2016; Girshick, 2015) on the application of convolution networks.

A very short time ago Qi et al. (2017) PointNet was presented, actually an end-to-end intelligent neural network that follows the point wise property without deviation from the point clouds. This technique illustrates benchmarking results used on 3D recognition by dividing the object into parts and studying each part systematically.

In Qi, Yi, Su, and Guibas’s work (2017), an upgraded version of PointNet was unfolded which used to understand the existing object structures under different measures. To obtain more accurate results, these two methods have different characters on all the loaded data points (~1k points). Therefore, point clouds of the type acquire using LiDARs contain ~100k points, guiding the structures as in (Qi, Yi, Su, & Guibas, 2017; Qi, Su, Mo, & Guibas, 2017) outputs high calculation and storage space requirements by using region proposal network (RPN) (Ren, He, Girshick, & Sun, 2015).

Secondly, the author has put forth VoxelNet, a 3D recognition structure (Figure 1) that concurrently understand a difference property by using point clouds and forecasts exact 3D shaped boxes, in a point-to-point manner which is shown in Figure 2. The author has constructed a unique property converting into a VFE layer, which permits the engagement of conversation between the points within in a voxel, by the summation of the point wise features within the area by using their averages. Piling up of large number of VFE layers permits knowing difficult features for distinguishing 3D shape detail features within the particular area.

Exceptionally, VoxelNet segments the point cloud into the same sized 3D voxels, converts individual voxel using piled VFE layers. 3D part formed features are averaged points within the voxels. In the next phase a high dimensional volumetric representation by using the point cloud.

Lastly, a detection result is obtained by representing volumetric consumption by RPN. This systematic algorithm is useful for both from the structure of the point of sparse and a very good organization of parallel processing by using the voxel grid. By view of the Bird’s eye detection method is used to evaluate VoxelNet and overall detection tasks of 3D which was obtained by KITTI benchmarking (Geiger, Lenz, & Urtasun, 2012). On a large scale, the investigational result shows that 3D detection methods on start-of-the-art LiDAR are best. The author has also exhibit that VoxelNet attains very good supportive results in identifying cyclists and pedestrians from LiDAR cloud point.

PointNet is a new type deepnet structure, which exhibits a very high performance and competency if many 3D tasks through classification of object and semasiology segmentation. Whereas Point Nets are having potentials for point cloud to be grouped into sets on a white or for the prediction of logical value for each point in a point cloud, it much more ambiguous and difficult to understand this carefully designed structure used for different levels in 3D detection. To locate more accurately project feasible positions of 3D objects in a 3D space. Acting in the same manner as in case of image recognition, it effortlessly quantifies the object in the 3D boxes by using sliding window (Engelcke, Rao, Wang, Tong, & Posner, 2017) or by region proposal networks (RPN) as in (Song, Lichtenberg, & Xiao, 2015).

By and large, the method for calculating more difficult 3D seeking largely enlarges cubically in accordance with the resolution and becomes more expensive for massive scenes or when actual process takes place in utilization such as driving autonomously. The principle of dimension reduction is used to reduce the space search. The two variants of PointNet are used for conducting 3D box regression bounding by amodal and segmenting the 3D objects instantly. The 3D mask of the object of interest is predicted by the division of the network and the amodal 3D box shaped structured bounding forecasted by using the network of regression (the object is identified even though only a small part of the object is detected).

The 3D centric vision opens new potentials for examining 3D data in more efficient manner. Primarily in our technique of pipeline, a few modifications are carried out repeatedly on 3D coordinates, which inline could into a cycle of much conditioned and pertaining to frames. After in lining, the task
The Impact of Personal Electronic Communications on Work-Life Balance and Cognitive Absorption
[www.igi-global.com/article/impact-personal-electronic-communications-work/41721?camid=4v1a](www.igi-global.com/article/impact-personal-electronic-communications-work/41721?camid=4v1a)

Towards a Model of Employee Weblog Usage: A Process-Oriented Analysis of Antecedents and Consequences
[www.igi-global.com/chapter/towards-model-employee-weblog-usage/68139?camid=4v1a](www.igi-global.com/chapter/towards-model-employee-weblog-usage/68139?camid=4v1a)
Working Together to Improve Usability: Exploring Challenges and Successful Practices
[www.igi-global.com/article/working-together-improve-usability/39013?camid=4v1a](www.igi-global.com/article/working-together-improve-usability/39013?camid=4v1a)

Social Responses to Conversational TV VUI: Apology and Voice
[www.igi-global.com/article/social-responses-to-conversational-tv-vui/121635?camid=4v1a](www.igi-global.com/article/social-responses-to-conversational-tv-vui/121635?camid=4v1a)