3D Object Detection in Point Clouds using Deep Hough Voting

Mohammad Reza Ramezani (Student ID: 9812762394)
Kimia Mahdinezhad (Student ID: 9812762349)

October 15, 2023

Abstract

We aim to have an object detector application for the autonomous drive mobile robot Eddie, by using the 3D point cloud that is presented by the SLAM algorithm that the robot uses to navigate. we use VoteNet, an end-to-end 3D object detection network based on a synergy of deep point set networks and Hough voting.

Introduction

More specifically the goal of this proposal is to estimate oriented 3D bounding boxes as well as semantic classes of objects from point clouds. Those can be achieved by VoteNet with the use of PointNet++ a 3D deep learning model for point clouds, the network is inspired by generalized Hough voting processes for object detection.

This can be achieved by passing the input point cloud though a PointNet++ network then the VoteNet samples a set of seed points and generate votes from their features. Votes are targeted to reach object centers so can be detected. One of the obstacles that we are facing is that kinect camera only captures surface of the robot so the result of the point based networks have difficulty aggregating scene context, to cover this our vote clusters emerge near object centers somewhere is empty space and in turn can be aggregated through a learned module to generate box proposals.The result is a powerful 3D object detector that is purely geometric and can be applied directly to point clouds.

Methodology and Related Works

Using Hough voting for object detection

The Hough transform was introduced in 1950s, it translates the problem of detecting simple patterns in point samples to detecting peaks in a parametric space. This algorithm later extends to feature extraction technique used in image analysis, computer vision, and digital image processing. The purpose of the technique is to find imperfect instances of objects within a certain class of shapes by a voting procedure. There also has been combined algorithms with the Hough voting over the years.

Deep learning on point clouds

Recently we see a surge of interest in designing deep network architectures suited for point clouds which showed remarkable performance in 3D object classification, object part segmentation, as well as scene segmentation.

Deep Hough Voting

Tradition Hough voting

It is a 2D detector which has an offline and online step.
Offline step uses a collection of images with annotated object bounding boxes and a constructed mapping between images features and their offset to their centers, called codebook.
Online step we select some interest points then compare them to the features in codebook to retrieve

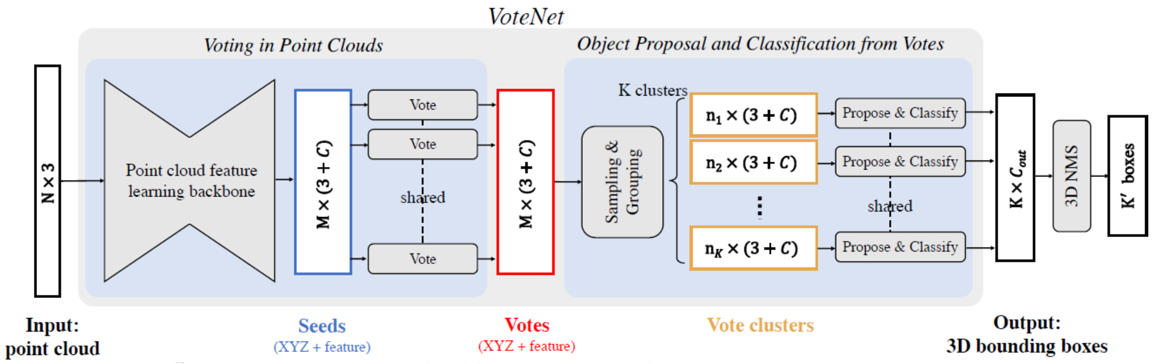

Figure 1: Illustration of the VoteNet architecture for 3D object detection in point clouds.

offsets and compute votes. As object patches will tend to vote in agreement, clusters will form near object centers.

Our Hough voting method

We use region proposal networks (RPN) to generate centers for objects which is likely to be in an empty space. To use it in proper way some adoptions are propesed as follow:
Interest points selected by deep neural networks instead of depending on hand-crafted features.
Vote generation is learned by a network instead of using a codebook. In addition, a vote location can be augmented with a feature vector allowing for better aggregation.
Vote aggregation is realized through point cloud processing layers with trainable parameters. Utilizing the vote features, the network can potentially filter out low quality votes and generate improved proposals.
Object proposals location, dimensions, orientation and even semantic classes can be directly generated from the aggregated features, mitigating the need to trace back votes' origins.

VoteNet Architecture

Learning to Vote in Point Clouds

Our goal is to generate votes where each vote has both a 3D coordinate and a high dimensional feature vector. There are two major steps:
point cloud feature learning through a backbone network by adopting PointNet++ which has several set-abstraction layers and feature propagation, it outputs a subset of inputs in size of M with XYZ and an enriched C-dimensional feature vector.
Hough voting with deep networks a shared voting module generates votes from each seed independently. The MLP takes seed feature f_i and outputs the Euclidean space offset $\Delta x_i \in R^3$ and a feature offset $\Delta f_i \in R^C$ such that the vote $v_i = [y_i; g_i]$ generated from the seed s_i has $y_i = x_i + \Delta x_i$ and $g_i = f_i + \Delta f_i$.

Expected Outcomes

By considering the 3D map output of the SLAM, our VoteNet application should be run without problem in real-time manner. Also beacause of the done researches which shows the application is the state-of-the-art we expect that our robot will detect objects of room correctly in simulator or in real test.

References

- Dana H Ballard. Generalizing the hough transform to detect arbitrary shapes. Pattern recognition, 13(2):111–122, 1981.
- Bastian Leibe, Ales Leonardis, and Bernt Schiele. Combined object categorization and segmentation with an implicit shape model. In Workshop on statistical learning in computer vision, ECCV, volume 2, page 7, 2004.
- Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pages 91–99, 2015.
- Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. arXiv preprint arXiv:1706.02413, 2017.
- Paul VC Hough. Machine analysis of bubble chamber pictures. In Conf. Proc., volume 590914, pages 554–558, 1959.
- Bastian Leibe, Ale˘s Leonardis, and Bernt Schiele. Robust object detection with interleaved categorization and segmentation. International journal of computer vision, 77(13):259–289, 2008.