Cooper Simpson
Final Project Proposal
December 7, 2020

Recurrent 3D LiDAR Object Detection

For my Fall 2020 Applied Deep Learning final project I propose to extend and adapt the work accomplished in *An LSTM Approach to Temporal 3D Object Detection in LiDAR Point Clouds* by Rui Huang and other Google researchers. This work, and my proposal, seeks to use recurrent neural network (RNN) architectures to improve the process of object detection in multi-frame LiDAR data. Robotics – including self-driving cars, drones, and many other sub-fields – stands to benefit greatly from advances in such techniques.

1 Context and Objectives

In the work described above, the researchers build off of a previous model for 3D LiDAR object detection that uses voxelization and sparse convolutions. Using this as a backbone, they add a sparse convolutional LSTM module into the framework to take advantage of the temporal dimension of the data. A recently available LiDAR dataset from Waymo is used for training and evaluating their model. They find that their method achieves state-of-the-art performance while reducing computational cost and increasing memory efficiency.

I propose to expand off of this exceptional foundation by altering the architecture and expanding its capabilities. The core idea of leveraging the temporal aspect of multi-frame LiDAR data using an RNN is intuitive and central to their great results. However, I believe there is significant room for exploration in the specifics of their model architecture, evaluation, and functionality. With this I hope to learn more about cutting edge deep learning techniques, extend my technical skills in model development, and add something meaningful to this research field.

2 Approach

I propose a number of more detailed specific goals anchored by a core fundamental objective – this being to recreate the model architecture described in the paper. While this may not immediately seem like that much of an achievement, it will be quite a task in itself. A number of the technical components used by the researchers are not widely available in packages such as TensorFlow or PyTorch. Furthermore, their code is not publicly available and cannot be used as an aid. This means the model will need to be cobbled together from somewhat disparate sources or custom developed – most likely with a combination of the two. This is not all bad as it will serve as a great learning opportunity, and put me in a better position for expanding it later. For this expansion I will list a number of goals, ordered by their priority in approach:

1. Recurrent Layer: I propose to alter the RNN architecture to use a GRU instead of an LSTM module. My motivation being that GRUs have less memory overhead and can be easier to train.

2. Backbone Network: I propose to alter the backbone network being used throughout the model. This original network is a U-Net style sparse 3D CNN, and I suggest replacing this with a Dilated Encoder-Decoder sparse CNN (like DeepLabv3+).

3. Voxelization: I propose to reduce the number of direct and inverse voxelizations that take place within the network. This consists of determining a way to maintain the voxelized format throughout the network, and de-voxelizing only at the end.

4. Object Detection: I propose to alter the object detection portion of the network. Specifically, I would like to consider replacing the non-maximum suppression, and would also consider changing other components.

I would like to achieve all the goals presented here, but I realize that what I seek to do is challenging. I believe that whatever the outcome I will have gained valuable knowledge and experience.

References

[1] Benjamin Graham and Laurens van der Maaten. Submanifold sparse convolutional networks, 2017.

[2] Rui Huang, Wanyue Zhang, Abhijit Kundu, Caroline Pantofaru, David A Ross, Thomas Funkhouser, and Alireza Fathi. An lstm approach to temporal 3d object detection in lidar point clouds, 2020.

[3] Mahyar Najibi, Guangda Lai, Abhijit Kundu, Zhichao Lu, Vivek Rathod, Thomas Funkhouser, Caroline Pantofaru, David Ross, Larry S. Davis, and Alireza Fathi. Dops: Learning to detect 3d objects and predict their 3d shapes, 2020.

[4] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection, 2017.