

## Abstract

Recent deep learning breakthrough have allowed to effective classify and segment point clouds by convert them into graph and feed them to *Geometric Deep Learning* models. Point clouds are able to represent any shape in space making them one of the most versatile data-structure for non-euclidean dataset. The architecture that archives *state of the art* results in booth classify and segment those cloud is Dynamic Graph CNN [?]. It utilizes a special Convolution operator applied directly on the graph's edge called *EdgeConv* using booth local and global information. The aim of this project is to first reproduce the results on the famous ModelNet40 [?] dataset with more than ten thousand meshed. Reproducing the paper's results ensure our architecture's correctness. Then, we test DGCNN against a vanilla CNN on a dataset composed by heightmaps, images where each pixel is the height value of a terrain region, labeled as traversable or not traversable. Those images were generated letting a legged crocodile-like robot, *Krock*, walking into a simulated environment on synthetic maps and cropping the corresponding terrain patch around each stored trajectory pose. This dataset represents a really interesting playground to explore with architecture is able to extract the most information from the geometry of the terrain and correctly predict their traversability. s

# Traversability Estimation with Dynamic Graph CNN

Francesco Saverio Zuppichini

June 8, 2019

## 1 Introduction

Graphs can express a wide array of data structures using a set of nodes and the relationship between them, edges. Graphs are non-euclidean data, thus standard deep learning methods, like CNN, fails. Recently, a new set of machine learning models, *Graph Deep Learning* have flourished and successfully applied on a wide array of graphs structure such as social networks `To do (??)`, proteins `To do (??)`, and on point clouds. A Point cloud is a set of data points in space that often represents a specific 3D object by defined its points coordinates. They are extremely flexible and can be easily gather using consumer hardware such as a Kinect or a Lidar. Even if they are not graphs by definition, we can easily convert them connecting to each point  $k$  neighbors.

As with images, we are interested in classify or segment those point clouds. The current architecture that achieves the *state of the art* in those tasks is *Dynamic Graph CNN* proposed by Wang et al [?]. It utilizes a special convolution layer, *EdgeConv*, to classify or segment point clouds.

This this project will focus on explain, test and evaluate this architecture. First, we will reproduce the results of the original paper on the classification task. Then to test its effectiveness on predicting traversability estimation using ground's patches for a legged robot by comparing it to a classic CNN approach. The discussion will only scrape the surface of the topic and will not treat in detail other architectures. If interested, we suggest to the reader the review by Zhou et al [?].

## 1.1 Model

In this section we summarized the original paper's architecture, starting by the backbone of DGCNN, the *EdgeConv*.

## 1.2 Dynamic Graph CNN

Dynamic Graph CNN is composed by multiple layer of *EdgeConv* stacked one after the other similar to how Convolution are integrated into CNN. Figure 1 shows the two architecture proposed by the Authors, for classification and segmentation.

### 1.2.1 EdgeConv

Given a point cloud with  $n$  denoted as  $X = \{x_1, \dots, x_n\} \subseteq R^F$  and a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  representing the local structure of the cloud, the  $i$ -th output of *EdgeConv* operation is defined as

$$x'_i = \square_{j:(i,j) \in \mathcal{E}} h_{\Theta}(x_i, x_j) \quad (1)$$

Where  $\square$  is a channel wise aggregation function and  $h_{\Theta}$  is the edge function. The authors choose  $\square = \max$  and  $h_{\Theta} = h_{\Theta}(x_i, x_j - x_i)$ . Where  $x_i$  and  $x_j$  are the coordinates of the center of the patch and its neighbors respectively. Intuitively, this allows the layer to use both global and local information about the mesh. In addition, the edge function has a learnable parameter  $\Theta$  that is a classic feed-forward neural network.

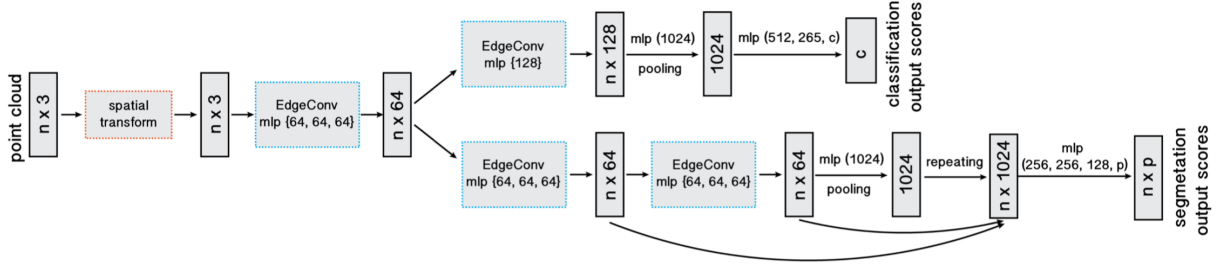


Figure 1: The two Dynamic Graph CNN architecture divided into upper and lower branch for classification and segmentation respectively, from the original paper [?]. In each edge conv,  $\text{mlp}\{n_0, \dots, n_i\}$  represents the number of output neurons of each dense layer. For example, in the first edge conv there are three dense layers with outputs size of 64.

### 1.2.2 Dynamic EdgeConv

### 1.3 Similarities to other architectures

The authors shown that DGCNN is related to booth two categories of approaches: PointNet and graph CNN. First, PointNet is just a special case of DGCNN where  $k = 1$  resulting in an empty graphs where the edge convolution functions utilises only the global information of each patch,  $h_{\Theta} = h_{\Theta}(x_i)$

While, the common denominator with graph CNN methods, MoNet, ECC and Graph Attention Network and DGCNN is the notion of local patch. However, one crucial different is that all previous models work on a *static* graph.



Figure 2: Some of the *chairs* CAD meshed in ModelNet40 [?].

## 2 Experiments

### 2.1 Setup

### 2.2 Mesh Classification

Exactly as in the original paper, we tested the model's score on the *ModelNet40* dataset, a collection of more than ten thousand CAD meshed. The samples are divided into 9,843 for training 2,468 for testing with 40 categories. Figure 2 shows some of the meshed for the category *chair*.

### 2.3 Traversability Estimation

## 3 Results

## 4 Conclusion