Abstract

Name of the student: **Prajwal Kumar Singh** Roll No: **19210048**

Degree and Department: **M.Tech. Computer Science & Engineering**

Thesis title: **Generating X From The Embedding of Images And Text**

Thesis supervisor: **Dr. Shanmuganathan Raman**

Point cloud is an efficient way of storing geometric data, and due to this, deep learning on point cloud is time and memory efficient. Several methods are proposed for processing point cloud using deep learning, e.g., PointNet is for classification and segmentation, and FoldingNet model is for point cloud encoding-decoding. In this work, we proposed an efficient encoder-decoder model that can use graph convolution and hierarchical information of point clouds to generate information preserving embedding better than the previous methods. We have also shown its application on single view image to 3D point cloud reconstruction.

Synthesising the realistic image using text is a practical problem. There are several methods proposed for it in recent years. In this work, we proposed a small modification to the architecture of StackGAN and different loss function to train the architecture of the network. Our experiment shows that modifying architecture and adding a new loss function cause improvement in the StackGAN model's performance. These improvements cause the proposed model to generate semantically meaningful images that are close to text descriptions.

Most of the deep learning based methods follow the data-driven approach, which means we require a robust dataset to solve any problem using deep learning. We have proposed two such datasets in this work. The first dataset we created is APEX-1M for solving the problem of data extraction from the images of graph plots. The second dataset we created is HDIB1M which is more robust than the existing datasets for handwritten document image binarization.