Pacific Graphics (2023)
R. Chaine, Z. Deng, and M. H. Kim (Editors)

Poster

TreeGCN-ED: A Tree-Structured Graph-Based Autoencoder Framework For Point Cloud Processing

Prajwal Singh, Ashish Tiwari, Kaustubh Sadekar & Shanmuganathan Raman

CVIG Lab, Indian Institute of Technology Gandhinagar

Abstract

Point cloud is a widely used technique for representing and storing 3D geometric data. Several methods have been proposed for processing point clouds for tasks such as 3D shape classification and clustering. This work presents a tree-structured autoencoder framework to generate robust embeddings of point clouds through hierarchical information aggregation using graph convolution. We visualize the t-SNE map to highlight the ability of learned embeddings to distinguish between different object classes. We further demonstrate the robustness of these embeddings in applications such as point cloud interpolation, completion, and single image-based point cloud reconstruction. The anonymized code is available here for research purposes.

CCS Concepts

• ***Computing methodologies*** → ***Shape representations;***

1. Introduction

Point cloud is an important data structure that can be used to store information about the geometry of any 3D shape. Several applications require efficient embeddings of point clouds, such as point cloud classification and segmentation [QSMG17], point cloud completion [CFG*15, YFST18], and point cloud generation [SPK19]. While encoder-decoder-based methods have been widely used for generating information-preserving embeddings over images for image compression and filtering tasks, extending these methods to 3D point cloud data is challenging due to its irregular structure compared to images and 3D voxels. In this work, we design a deep encoder-decoder framework to learn information-rich robust embeddings for several tasks on point clouds, such as clustering, classification, interpolation, completion, and image-based reconstruction.

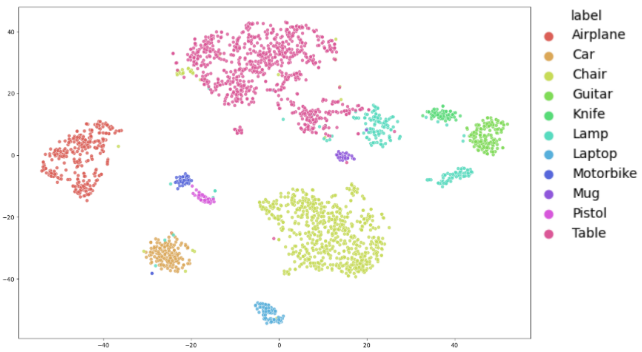

Figure 1: *t-SNE visualization (perplexity value of 40) showing the clusters of learned feature embedding per class.*

2. Method

We extend tree-GAN [SPK19] - a GAN-based (decoder-only) tree-structured framework to an autoencoder framework for efficiently encoding and decoding point clouds. Specifically, instead of using a noise vector $z \in \mathbb{R}^{96}$ sampled from a normal distribution $\mathcal{N}(0, I)$ (as in tree-GAN), we propose to use the embeddings learned from a tree-structured graph convolution-based encoder for generating point clouds. The proposed encoder, when combined with the decoder [SPK19], forms a complete tree-based encoder-decoder framework called TreeGCN-ED. The TreeGCN-ED architecture and the information flow are described in Figure 2. We use Chamfer loss to train the network over the ShapeNetBenchmarkV0 [CFG*15] dataset with 16 object classes. We perform train-validation-test split as per [CFG*15] (more details in the supplementary).

3. Experiments and Results

We compare the performance of TreeGCN-ED with the popular FoldingNet model [YFST18] on the test set of ShapeNetBenchmarkV0 dataset [CFG*15] and evaluate the performance using Chamfer Distance (CD) and Fréchet Point Cloud Distance (FPD). We observed that TreeGCN-ED outperforms FoldingNet by a significant margin. We obtained the CD of 1.21 units (TreeGCN-ED) and 1.48 units (FoldingNet), and FPD of 11.54 units (TreenGCN-ED) and 44.52 units (FoldingNet) averaged over 16 different classes in the dataset. We report the per-class metrics in the supplementary.

Point cloud clustering. Figure 1 shows a t-SNE plot to establish how well our encoder model can generate feature embedding for each class. We observe that the inter-class separation is higher, indicating the discriminative capacity of the proposed encoder model.

Point cloud interpolation. We also perform inter-class and intra-class interpolation from the source and the target point clouds

DOI: 10.2312/pg.20231278

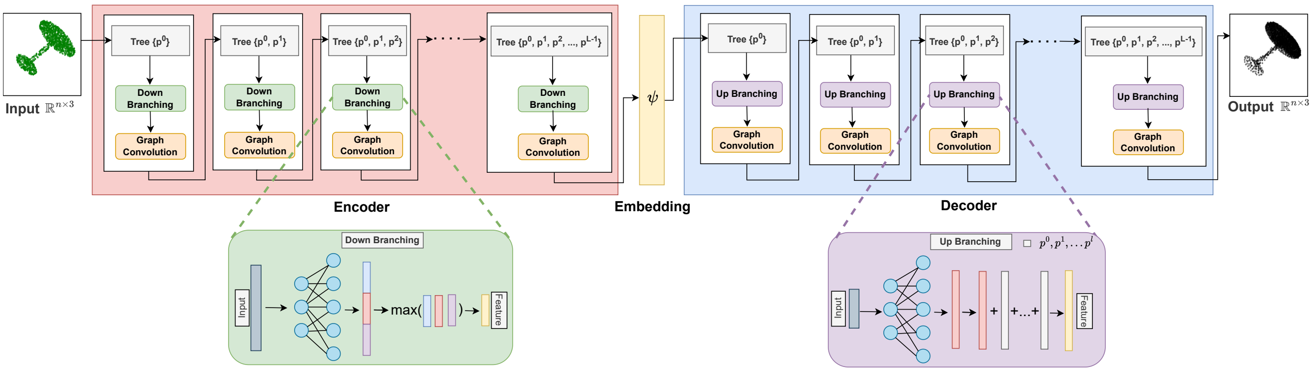

Figure 2: *TreenGCN-ED: The encoder (left) consists of multiple stages of down-branching and graph convolutions for encoding the input 3D point cloud* $\mathbf{p} \in \mathbb{R}^{n\times3}$ *into a feature embedding* $\psi \in \mathbb{R}^K$*. The decoder (right) takes* ψ *as input and reconstructs the 3D point cloud through a set of up-branching and graph convolutions (similar to [SPK19]). The down-branching consists of a fully connected layer followed by max-pooling to accumulate features from ancestors for each node. The output of the fully connected layer is divided into* $\mathbb{C}$ *equal components that are passed to the max-pooling layer. The up-branching is responsible for collecting information from the feature embedding of the ancestors and upsampling, which are then passed to the graph convolution layer for further refinement. Each stage is designed to learn relations between a node and its neighbors, especially when point clouds do not have edge connections.*

in Figure 3. The intra-class interpolation results illustrate the ability of our model to synthesize novel shapes between two given shapes. The generated shapes faithfully represent the object class at each interpolation stage.

Single image-based point cloud reconstruction and point cloud completion. 3D reconstruction from a single image is an ill-posed problem. Figure 4 (a) shows that the learned embeddings can also foster image-based reconstruction. Moreover, we also show how well TreeGCN-ED can fill the missing structures in the point cloud data (see Figure 4 (b)). We observe that TreeGCN-ED also learns inherent semantic information of the point cloud. Specific implementation details and more qualitative results are provided in the supplementary.

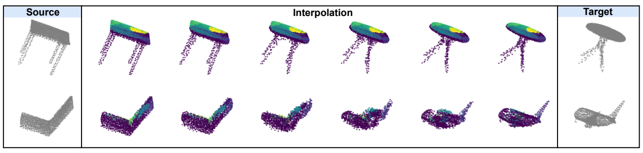

Figure 3: *Intra-class (top) and inter-class (bottom) point cloud interpolation exhibiting a smooth transition.*

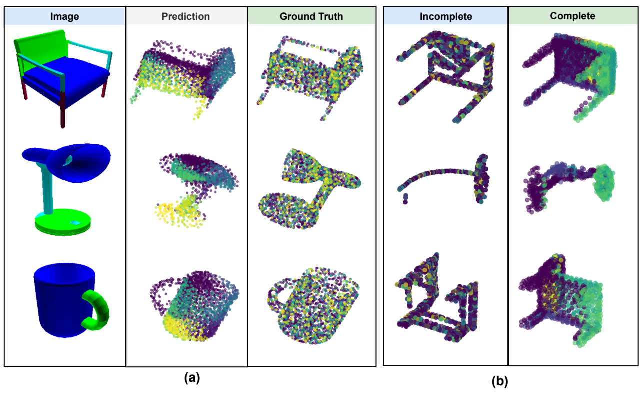

Figure 4: *Results on (a) single image-based point cloud reconstruction and (b) point cloud completion.*

Dataset	No Augmentation		Rotation Augmentation	
	$\psi \in \mathbb{R}^{256}$	$\psi \in \mathbb{R}^{512}$	$\psi \in \mathbb{R}^{256}$	$\psi \in \mathbb{R}^{512}$
ShapeNetCore.v2	10.90	10.07	8.82	7.88
ModelNet10	0.83	0.83	0.85	0.85
ModelNet40	0.71	0.72	0.73	0.73

Table 1: *Ablation study to understand the effect of feature dimension (*$\psi \in \mathbb{R}^{256}$ *vs* $\mathbb{R}^{512}$*) and data augmentation on the Chamfer distance.*

We additionally perform ablation studies to understand the effect of learned feature dimension and data augmentation (like rotation) in Table 1 and obtain the best performance for $\psi \in \mathbb{R}^{512}$ and rotation augmentation.

4. Conclusion

We propose a tree-structured graph convolution-based encoder architecture and combine it with the decoder of tree-GAN to create a complete tree-structured encoder-decoder model for point cloud processing. The results highlight the effectiveness of the encoder model in learning information-rich features for tasks such as point cloud clustering, completion, generation, and single image-based reconstruction, which would be valuable for the graphics community.

References

[CFG*15] Chang A. X., Funkhouser T., Guibas L., Hanrahan P., Huang Q., Li Z., Savarese S., Savva M., Song S., Su H., Xiao J., Yi L., Yu F.: Shapenet: An information-rich 3d model repository, 2015. arXiv:1512.03012. 1

[QSMG17] Qi C., Su H., Mo K., Guibas L. J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. *2017 IEEE (CVPR)* (2017), 77–85. 1

[SPK19] Shu D. W., Park S. W., Kwon J.: 3d point cloud generative adversarial network based on tree structured graph convolutions. *2019 IEEE/CVF (ICCV)* (2019), 3858–3867. 1, 2

[YFST18] Yang Y., Feng C., Shen Y., Tian D.: Foldingnet: Point cloud auto-encoder via deep grid deformation. *2018 IEEE/CVF (CVPR)* (2018), 206–215. 1