Fair comparison between recent 3D structure representations
Deep Learning - Project proposal

Benoit Roussel
École normale supérieure Paris-Saclay
4, avenue des Sciences, 91190 Gif-sur-Yvette
benoit.roussel@ens-paris-saclay.fr

Arthur Calvi
École normale supérieure Paris-Saclay
4, avenue des Sciences, 91190 Gif-sur-Yvette
arthur.calvi@ens-paris-saclay.fr

Abstract

An ideal representation of 3D structures is still not unanimous for the 3D reconstruction community. As current explicit reconstructions has not yet satisfied enough requirements, implicit representations using the Signed Distance Function (SDF) or the occupancy function has emerged with better results. However, what is the best choice of implicit representations of the data for learned based methods nowadays ? A fair comparison between the recent 3D structure representations will be led.

1. Motivation and Problem Definition

3D reconstruction is a fundamental problem in computer vision with numerous applications. An ideal representation of 3D geometry should have the following properties: a) encode complex geometries and arbitrary topologies, b) scale to large scenes, c) encapsulate local and global information, and d) be tractable in terms of memory and computation.

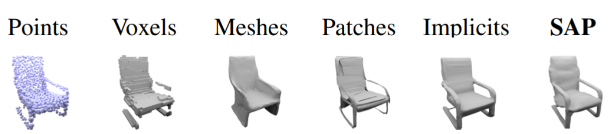

Figure 1. Different representations of the 3D structure

Unfortunately, current representations for 3D reconstruction do not satisfy all of these requirements. Volumetric representations [6] are limited in terms of resolution due to their large memory requirements. Point clouds [3] are lightweight 3D representations but discard topological relations. Mesh-based representations [4] are often hard to predict using neural networks.

On one hand, several works [2, 7, 8, 10] have recently introduced deep implicit representations which represent 3D structures using learned occupancy or signed distance function. In contrast to explicit representations, implicit methods do not discretize 3D space during training, thus resulting in continuous representations of 3D geometry without topology restrictions.

On the other hand, Songyou Peng *et al*. [9] introduced Shape As Points (SAP) that efficiently solves the Poisson Equation during a faster inference by representing the shape as an oriented point cloud.

The aim of this project is to provide a fair study and comparison of these new 3D structure representations with respect to the above properties.

2. Methodology

The purpose of the project is to investigate on current 3D structure representations by evaluating it on the reconstruction of a 3D mesh from a points cloud. For this task, we will mainly focus on two architectures : DeepSDF proposed by J.J. Park *et al*. [8] and Occupancy Network introduced by L.Mescheder *et al*. [7].

Based on the DeepSDF architecture, we will adapt it by switching to the occupancy function. Then we will train both networks on the same dataset $[X, Y]$, where X are points clouds of objects and Y are samples of the 3D meshes of those objects after the Marshing Cube algorithm [5]. Then, both methods will be compared according to appropriate metrics to illustrates pros and cons of both representations for different type of shapes. A third method called Shape as Points (SAP) introduced by S. Peng *et al*. [10] will be eventually compared against the two others. Finally, we will attempt to introduce topological data analysis features such as persistence diagrams into the DeepSDF algorithm to enhance its performance.

The chronological steps are the following :

1. Comprehension of the DeepSDF architecture.
2. Based on the DeepSDF architecture, adaptation with the occupancy function.
3. Based on the DeepSDF architecture, adaptation with the SAP 3D structure representation.
4. Study and comparison of the 3 new architectures on the same dataset [1] and with the same metrics.
5. (Optional) Application : 3D reconstruction of objects captured in 2D images.
6. (Optional) Customization of the DeepSDP by adding vectorized representation of persitence diagram of the points cloud as input of the deep neural network.

3. Evaluation

The different architectures will be assessed on their capacities to a) represent training data and b) represent unseen shapes. First experiment will be mainly focused on the ability of describing geometric details and second experiment will focus on generalization. 3D reconstructions of objects from points clouds will be rated according to the following metrics:

The Chamfer Distance give an indication on the resemblance of two points clouds by returning the closest point's Euclidean distance for each point of both point clouds. S_1 and S_2 are two objects, and x and y are respectively points of those objects. Points are randomly sampled from the obtained meshes.

$$d_C(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} ||x-y||_2^2 + \sum_{y \in S_2} \min_{x \in S_1} ||x-y||_2^2 \quad (1)$$

Earth Mover's Distance (or Wasserstein Metric) give insights on the optimal assignement between two points clouds. Where $\phi : S_1 \longrightarrow S_2$ is the bijection.

$$d_{EMD}(S_1, S_2) = \min_{\phi : S_1 \longrightarrow S_2} \sum_{x \in S_1} ||x - \phi(x)||_2 \quad (2)$$

Volumetric Intersection over Union (IoU) defined as the quotient of the volume of the two meshes' union and the volume of their intersection. Unbiased estimates of the volume of the intersection and the union are obtained by randomly sampling a number of points from the bounding volume and determining if the points lie inside our outside the ground truth / predicted mesh.

Normal consistency score as the mean absolute dot product of the normals in one mesh and the normals at the corresponding nearest neighbors in the other mesh.

As for the dataset, architectures will be evaluated mainly on the ”chair” category of the ShapeNet synthetic dataset. This subset is challenging to represent as it is highly varied, and many models contain high-frequency details.

References

[1] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. Shapenet: An information-rich 3d model repository, 2015. 2

[2] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling, 2019. 1

[3] Haoqiang Fan, Hao Su, and Leonidas Guibas. A point set generation network for 3d object reconstruction from a single image, 2016. 1

[4] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan Russell, and Mathieu Aubry. AtlasNet: A Papier-Mâché Approach to Learning 3D Surface Generation. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1

[5] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '87, page 163–169, New York, NY, USA, 1987. Association for Computing Machinery. 1

[6] Daniel Maturana and Sebastian Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 922–928, 2015. 1

[7] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space, 2019. 1

[8] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation, 2019. 1

[9] Songyou Peng, Chiyu ”Max” Jiang, Yiyi Liao, Michael Niemeyer, Marc Pollefeys, and Andreas Geiger. Shape as points: A differentiable poisson solver, 2021. 1

[10] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks, 2020. 1