Neural Networks for Continuous 3D Shape Representation

EE 367 Project Proposal

Youjin Song (ujinsong@stanford.edu)

Problem Statement and Motivation

Given the increasing demand for AR and MR applications in spatial computing, accurate, fast, and memory efficient 3D scene representation from obtainable sources like 2D image inputs is gaining more and more importance. Although 3D imagery entails greater computational demands and observation requirements, there's potential to make the process lighter, by leveraging human intuition or incorporating physical models. This project explores such possibilities using semantic information of a scene, based on previous related works that effectively leverage neural networks for building continuous representation.

Previous and Related Works

Classical methods for representing 3D configurations include voxels, point clouds, and meshes. Voxel representations extend the concept of pixels into 3D space but face limitations due to memory growth, typically restricting resolution. Point clouds offer memory efficiency by only occupying space where objects exist but lack structural connectivity. Mesh representations, formed by deforming a template mesh, are constrained by predefined topologies.

The concept of a continuous **3D Occupancy Function**[1] emerged to address these limitations. This approach involves learning an occupancy function through neural networks, mapping 3D point coordinates to a scalar occupancy probability between 0 and 1. By overcoming previous discretization issues, it enables the representation of realistic, high-resolution meshes within a fixed resolution. Following the emergence of the continuous 3D occupancy function, numerous methods have employed neural networks to learn continuous shape representations. **Deep SDF**[4], for instance, focuses on determining the distance from the surface for each point in 3D space. It achieves this by training a shape-conditioned classifier, where the decision boundary corresponds to the shape's surface, supervised by actual 3D Signed Distance Function (SDF) values. **Scene Representation Networks(SRN)**[5] portray scenes as continuous functions, mapping world coordinates to latent vectors containing local scene properties, solely using 2D images and camera poses(without access to depth or shape information). There also have been efforts to learn implicit shape and texture representations directly from RGB images through approaches like **Differentiable Volumetric Rendering (DVR)**[3]. **NeRF(Neural Radiance Fields)**[2] introduced an innovative approach to represent continuous scenes with intricate geometry and materials using 5D neural radiance fields, parameterized as basic MLP networks. This method leverages volume rendering techniques to achieve photorealistic reconstruction and store semantic features, reducing reliance on depth sensors with their inherent limitations.

Methodology and Experiments

This project will primarily concentrate on implementing the most recent variations of NeRF-like methods and potentially refining the model for improvement using semantic information learned by neural networks, if time allows. Given that this is an individual project and constrained by both time and computational resources, the focus will mainly be on attempting relatively simple tasks, such as representing 3D scenes with straightforward shapes.

Timeline

- **Feb 25 - Mar 2** Choose suitable datasets for the project, implement Occupancy Network
- **Mar 3 - Mar 9** Implement the algorithms and train models, Ablation studies

References

[1] Lars Mescheder et al. "Occupancy networks: Learning 3d reconstruction in function space". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 4460–4470.

[2] Ben Mildenhall et al. "Nerf: Representing scenes as neural radiance fields for view synthesis". In: *Communications of the ACM* 65.1 (2021), pp. 99–106.

[3] Michael Niemeyer et al. "Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 3504–3515.

[4] Jeong Joon Park et al. "Deepsdf: Learning continuous signed distance functions for shape representation". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 165–174.

[5] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. "Scene representation networks: Continuous 3d-structure-aware neural scene representations". In: *Advances in Neural Information Processing Systems* 32 (2019).