Optimizing 3D Object Reconstruction: A Comparative Study of Differentiable Volumetric Rendering and Signed Distance Functions

Sina MohammadiNiyaki, Ali Nikan
Department of Computing Science
Simon Fraser University
Burnaby, Canada
{sma231, ana106}@sfu.ca

I. Main Goal

A. Computational Problem

3D object reconstruction is a fundamental problem in computer vision with applications in robotics, augmented/virtual reality, gaming, and autonomous systems. Implicit neural representations have emerged as a powerful alternative to explicit 3D models, offering continuous and memory-efficient shape encoding. Differentiable Volumetric Rendering (DVR) reconstructs 3D objects from multiview 2D images by learning an implicit occupancy field and using volumetric rendering for supervision [1]. In contrast, Signed Distance Functions (DeepSDF) learn a continuous shape representation by mapping 3D coordinates to signed distance values, requiring explicit supervision from ground-truth 3D models [2]. Despite their different training signals, both methods leverage neural implicit functions to represent 3D geometry and can extract surfaces via ray sampling or Marching Cubes.

B. Project Goal

This project aims to systematically compare DVR and DeepSDF by evaluating their reconstruction quality, computational efficiency, and scalability. We will benchmark both methods under controlled conditions using standard 3D reconstruction metrics, including Chamfer Distance, IoU, training time, and memory consumption. Additionally, we will investigate optimizations to improve efficiency without compromising accuracy, such as sparse voxel grids for DVR and hash-based encodings for DeepSDF. Finally, we will explore a hybrid approach that combines DVR's ability to infer coarse 3D structure from multiview images with DeepSDF's ability to refine fine geometric details using signed distance supervision.

C. Rendering & Image Generation Component

DVR leverages differentiable volumetric rendering to supervise 3D reconstruction from 2D images, making rendering an integral part of training and optimization [1]. In contrast, DeepSDF employs implicit function learning to represent shapes by mapping 3D points to signed distance values. While DeepSDF does not use rendering for training, it can generate 3D visualizations at inference time using Marching Cubes for mesh extraction or direct raycasting against the learned SDF field [2]. Our experiments will involve generating 3D meshes, volumetric renderings for DVR, and visualizations of DeepSDF reconstructions.

II. Related Work

A. Differentiable Volumetric Rendering (DVR)

DVR [1] introduces a differentiable rendering framework that learns implicit shape and texture representations **directly from 2D images**, enabling 3D reconstruction without explicit 3D supervision. It represents objects as **implicit occupancy fields** and optimizes reconstruction by rendering images and comparing them to input views using **photometric and silhouette losses**. By computing **differentiable depth gradients**, DVR refines the surface estimation and enables **end-to-end training** with only 2D supervision. While DVR produces **high-quality, watertight reconstructions**, it can be **computationally expensive** due to volumetric sampling and **ray-based integration**.

B. DeepSDF

DeepSDF [2] models 3D shapes as a **continuous signed distance function (SDF)**, encoding them in a **latent space** that enables **shape interpolation, reconstruction, and completion**. Unlike DVR, which learns from image supervision, DeepSDF is trained using **precomputed SDF values from 3D datasets**. The method is **memory-efficient** because it does not require explicit voxel grids; instead, it represents geometry via a **multilayer perceptron (MLP)** that maps 3D coordinates to signed distance values. However, it lacks direct 2D supervision and requires **surface extraction techniques** such as **Marching Cubes** or **raycasting** to generate explicit 3D meshes.

C. Relation to Our Work

DVR and DeepSDF both employ **implicit neural representations** to model 3D objects, but they differ

significantly in their **input modalities, supervision, and reconstruction techniques**:

- **DVR learns from multi-view 2D images, using volumetric rendering for supervision.** It is effective for image-based reconstruction but computationally intensive due to volumetric sampling.
- **DeepSDF learns directly from precomputed 3D signed distance values.** It is highly memory-efficient and enables shape generalization but lacks direct image supervision.

Our project will systematically compare these two approaches in terms of **reconstruction accuracy, computational efficiency, and generalization to unseen objects**. Additionally, we will explore novel optimizations to improve runtime performance and scalability. Inspired by Instant-NGP [3], we will investigate how **hash-based encoding techniques** can accelerate training by replacing **dense neural representations with efficient multiresolution hash tables**. This optimization may improve the performance of both DVR (by reducing memory usage in volumetric grids) and DeepSDF (by enabling faster convergence and finer detail reconstruction).

III. Resources

A. Datasets

- **ShapeNet** [4]: A large-scale repository of 3D models suitable for controlled experiments.
- **DTU Multi-View Stereo Dataset** [5]: A collection of real-world multi-view images with corresponding 3D ground truth, ideal for evaluating performance in practical scenarios.

B. Computational Resources

Initial experiments will be conducted using a RTX 3070 GPU. Should further computational power be required, we will utilize the Computing Science Instructional Labs (CSIL) at Simon Fraser University.

IV. Timeline

The planned tasks, dates, and responsibilities for this project are summarized in Table I.

Date	Task	Responsible
Feb 12–18	Set up baseline implementations of DVR and DeepSDF; preprocess datasets	Both
Feb 19–Mar 3	Train baseline models on ShapeNet and DTU datasets	Both
Mar 4–10	Evaluate reconstruction quality (Chamfer Distance, IoU)	Sina
Mar 11–17	Assess computational efficiency (training time, inference speed, memory usage)	Ali
Mar 18–24	Implement optimizations (sparse DVR grid, hash encoding for DeepSDF)	Both
Mar 25–31	Develop a hybrid DVR–SDF model and finalize benchmarking	Both
Apr 1–2	Complete final implementation and result consolidation	Both
Apr 3–6	Prepare and deliver the final project presentation	Both
Apr 7–11	Write and submit the final project report	Both

TABLE I: Work Plan and Responsibilities

V. Expected Outcomes

- **Baseline Comparison:** DVR is expected to yield **higher-quality, watertight reconstructions** by leveraging differentiable rendering with multi-view 2D supervision. However, it will be more **computationally intensive** due to volumetric sampling and ray-based integration [1].
- **DeepSDF Efficiency Trade-off:** DeepSDF should demonstrate **higher memory efficiency** and **generalization ability** across shape categories, as it encodes 3D structures in a compact latent space [2]. However, it lacks direct 2D supervision and requires explicit **surface extraction techniques** like Marching Cubes, which could lead to loss of fine details.
- **Optimized Models:** Inspired by Instant-NGP [3], we anticipate that **introducing hash-based encoding** will reduce training time and improve fine-detail reconstruction for DeepSDF. Additionally, **converting DVR's dense volumetric grid to a sparse representation** should enhance computational efficiency.
- **Hybrid Approach Feasibility:** Combining DVR's **strong multi-view image-based supervision** with DeepSDF's **continuous SDF representation** could enable a hybrid model that **captures coarse structures via DVR while refining details using SDF supervision**. This method, if feasible, could outperform both baselines individually.

References

[1] M. Niemeyer, L. M. Mescheder, M. Oechsle, and A. Geiger, "Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision," *arXiv preprint*, vol. abs/1912.07372, 2019. [Online]. Available: http://arxiv.org/abs/1912.07372

[2] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "Deepsdf: Learning continuous signed distance functions for shape representation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. [Online]. Available: https://arxiv.org/abs/1901.05103

[3] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Transactions on Graphics (TOG)*, 2022. [Online]. Available: https://arxiv.org/abs/2201.05989

[4] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, "Shapenet: An information-rich 3d model repository," *arXiv preprint*, 2015. [Online]. Available: https://arxiv.org/abs/1512.03012

[5] R. Jensen, A. Aanæs, H. W. Jensen, and R. Larsen, "Large-scale multi-view stereo dataset," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 406–413, 2014. [Online]. Available: https://roboimagedata.compute.dtu.dk/?page_id=36