Neural Radiance Fields (NERF): Synthesizing Novel Views of a Scene from Photographs

Chris Fritz

Background:

Given a visual scene, a set of images captured of that scene (e.g. with a camera), and the directions of those captures, (the 3D coordinates and viewing direction of the camera), ***view synthesis*** the process of generating an image (view) of the scene from an arbitrary viewing direction and location. Traditional, discrete solutions to the visual synthesis problem rely on sampling the scene as a voxel grid [1][2], leading to substantial limitations in synthesis resolution due to the trade-off between resolution and the number of sampled needed to represent a 3D voxelized space. Recent advances in computer vision have leveraged multi-layer perceptrons (fully-connected neural networks) in order to learn a ***Neural Radiance Field (NERF),*** a function mapping 5D position and direction coordinates to an RGB color and volumetric density. [2][3] The novel view is then synthesized through volumetric rendering of rays cast from a camera at the viewpoint. In this project, I wish to implement and train a NERF network in PyTorch. Time-permitting, I would also implement NERF-W, an extension of NERF that can synthesize scenes from not only arbitrary viewpoints, but also arbitrary lighting and background conditions.

Related Work:

NERF is not the first attempt to leverage neural networks as a solution to the view synthesis problem. Such networks have been used to map location coordinates to signed distance functions [5] or implicit representations of a 3D shape [6]. While these methods are theoretically capable of representing complex geometry, they appear limited in practice to simple, over-smoothed geometry [3]. Conversely, NERF and its extensions generate state-of-the-art photo-realistic representations of geometry, as the reader is encouraged to verify through the author's supplemental video. [link to NERF, link to NERF-W] (N.B: at present, only the original NERF authors have released code. The network models are coded in Tensorflow, so implementation in PyTorch is non-trivial)

Project Overview:

The bulk of this project will be a) the implementation of network models specified in research literature as well as b) the ray-traced rendering of images from the output of the network models. Thus the project contains substantial content from both a computational imaging (reconstruction from sparse samples, neural network-based image processing) and computer graphics (ray-tracing, volumetric rendering) perspective. The goal is reproduction of the authored works with as much fidelity as possible, notwithstanding the difference in GPU compute resources. (my NVIDIA RTX 3080 Ti vs their NVIDIA V100)

Milestones:

Timelines are notoriously unreliable. Here are my goals for the project in order of completion. It is unlikely I will accomplish them all by the submission deadline, particularly 6, but will see how far I can get.

1. Develop full understanding of NERF input data, network structure, and volumetric rendering techniques.
2. Implement a simplified NERF model (not featuring Hierarchical Sampling and full-5D positional encoding)
3. Implement a complete NERF model, (including Hierarchical Sampling and full-5D positional

encoding)

4. Develop full understanding of NERF-W extensions (including camera pose estimation and architectural changes to the original NERF work)
5. Implement NERF-W extensions (including variable illumination and uncontrolled / internet-scraped image input)
6. Incorporate a NERF / NERF-W rendering into a game engine (e.g. Unreal Engine 4) to enable scene interactivity, e.g. a player moving around the scene to novel viewpoints, and options to tweak lighting and appearance, etc.

References:

1. Seitz, S. M., & Dyer, C. R. (1999). Photorealistic scene reconstruction by voxel coloring. International Journal of Computer Vision, 35(2), 151-173.
2. Kutulakos, K. N., & Seitz, S. M. (2000). A theory of shape by space carving. International journal of computer vision, 38(3), 199-218.
3. Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020, August). Nerf: Representing scenes as neural radiance fields for view synthesis. In European conference on computer vision (pp. 405-421). Springer, Cham.
4. Martin-Brualla, R., Radwan, N., Sajjadi, M. S., Barron, J. T., Dosovitskiy, A., & Duckworth, D. (2021). Nerf in the wild: Neural radiance fields for unconstrained photo collections. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7210-7219).
5. Park, J. J., Florence, P., Straub, J., Newcombe, R., & Lovegrove, S. (2019). Deepsdf: Learning continuous signed distance functions for shape representation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 165-174).
6. Genova, K., Cole, F., Sud, A., Sarna, A., & Funkhouser, T. (2020). Local deep implicit functions for 3d shape. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4857-4866).