Optimizing a Neural Radiance Field (NeRF)

CVIG
IIT Gandhinagar

Presented By: Aditi Dey (20110007) and Kareena Beniwal (20110095)
Faculty Guide: Prof. Shanmuganathan Raman
PhD Guide: Prajwal Singh
CVIG Lab, Indian Institute of Technology Gandhinagar, India

Introduction

- **Problem Statement:**
  What if there was a way to capture the entire 3D scene just from a sparse set of 2D pictures?
- **Motivation:**
  The ability to synthesize novel views of a scene from a set of sparse images has profound implications in fields such as virtual and augmented reality, film production, and robotics.

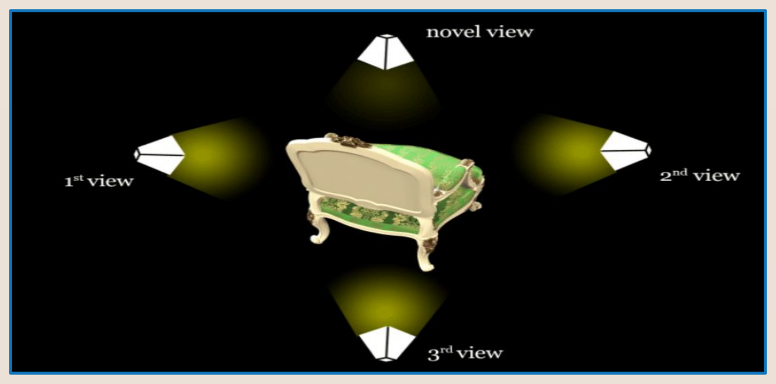


Fig : Novel view Generation

Procedure

In NeRF, a scene is represented as a 5D neural radiance field:

- 3D Coordinates (x, y, z): These determine the position of points in the scene.
- 2D Viewing Directions (θ, φ): These angles define the direction from which the point is observed.

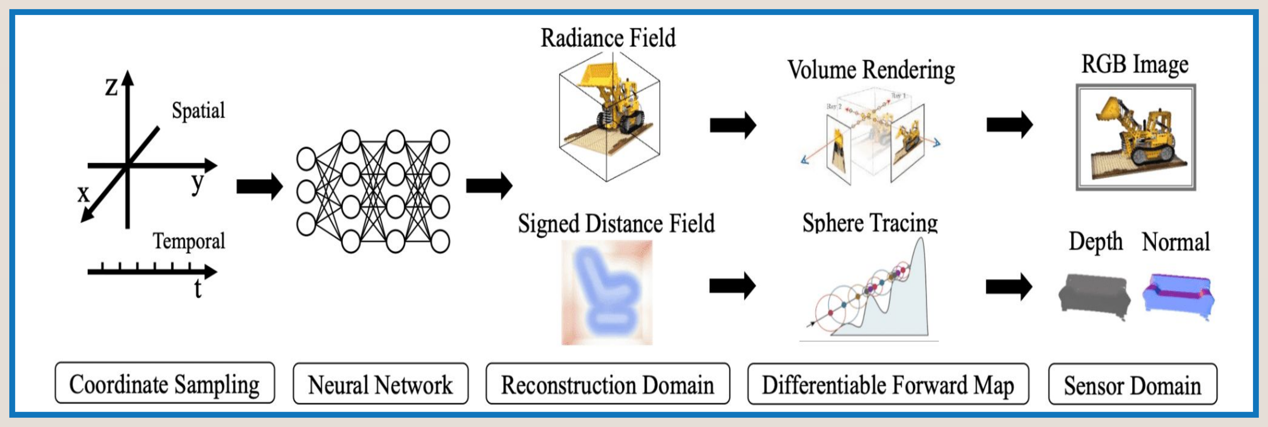


Fig: Feed Forward Neural field algorithm

- **Ray Casting (3D world of the scene ⇨ 2D plane of the image):** For every pixel in the target image, a ray is cast from the camera position through the pixel into the scene.
- **Stratified Sampling:** Initially, a set of points are sampled along the ray in a stratified manner which ensures even coverage and reduces variance.
- **Hierarchical Volume Sampling:** After the initial stratified sampling, the algorithm refines the sampling locations by focusing on regions with higher volume density.
- **Density and Color Estimation:** For each sampled point along a ray, the corresponding position and view direction are input into the neural network. The network outputs a volume density and a color for the point.
- **Volume Rendering:** The colors and densities of the sampled points are combined using volume rendering (using *gradient descent optimization*) to estimate the final color of the ray, corresponding to a pixel in the image.

An Attempt at Quantization

- Reduction of precision of the weights and biases in the network from high-precision floating-point numbers to lower-precision representations.
- Represented with fewer bits, typically 16 or 8.

The difference quantization makes:

- Memory Reduction
- Faster Computation

Trade-offs:

- Potential reduction in model accuracy. Lower precision can lead to information loss, which may affect the model's ability to generalize from its training data

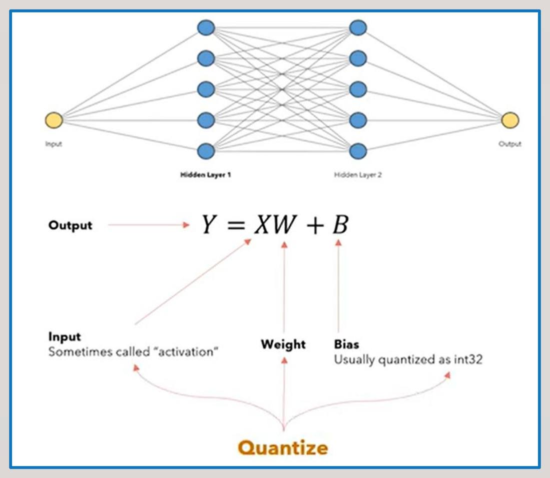
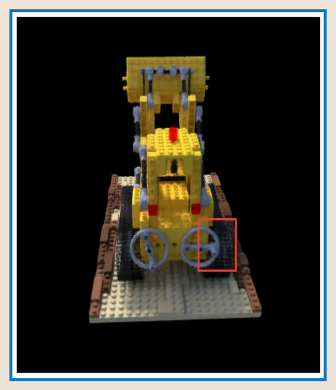


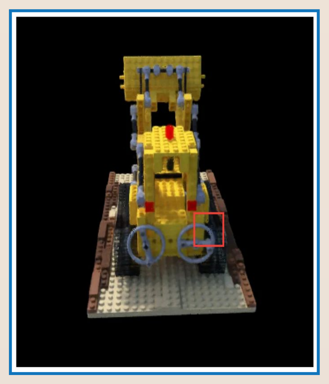
Fig: Quantization

Activation Functions

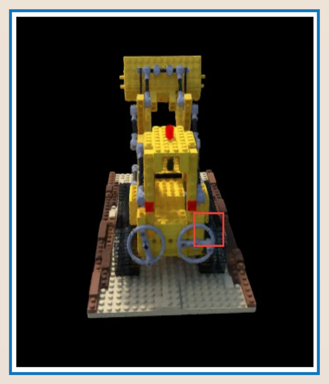
- We trained to model using various activation functions to select the best one for further optimizations.
- The NeRF models gave the best results when trained using GeLU ans SiLU activation on synthetic datasets



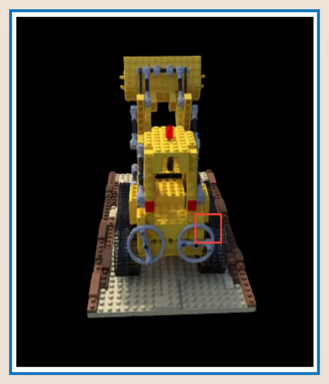
Leaky_ReLU



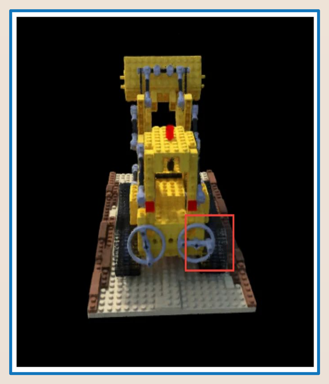
ELU



ReLU



SiLU



GeLU

Activation Function	MMSE	MPSNR	MSSIM	MLPIPS
LeakyReLU	0.0006	32.1577	0.9855	0.0375
ELU	0.0006	32.2008	0.9856	0.0362
ReLU	0.0006	32.2566	0.9859	0.0354
SiLU	0.0006	32.3598	0.9861	0.0348
GeLU	0.0006	32.3903	0.9863	0.0346

Table 1: Performance Metrics for Various Activation Functions

Positional Encodings

Sinusoidal Encoding:

$$\gamma(p) = \left(\sin(2^0 \pi p), \cos(2^0 \pi p), \cdots, \sin(2^{L-1} \pi p), \cos(2^{L-1} \pi p)\right)$$

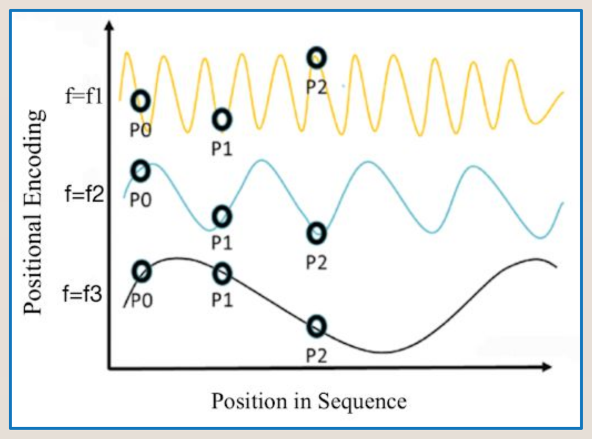


Fig: Sinusoidal Encoding

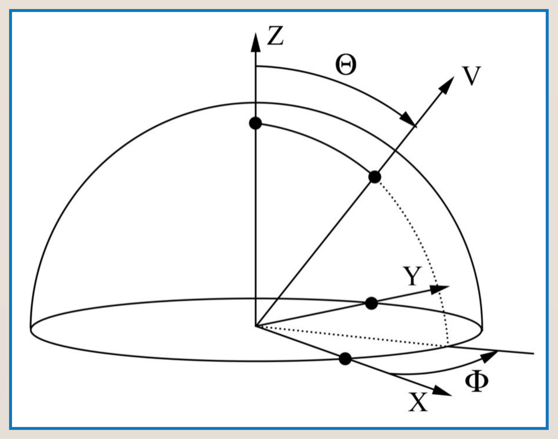


Fig: Spherical Harmonics Encoding

Harmonic Encoding:

- Uses a set of functions known as spherical harmonics, which are solutions to the spherical harmonic differential equation.
- Spherical harmonic function: $Y_\ell^m(\theta, \varphi) = N e^{im\varphi} P_\ell^m(\cos\theta)$

Skip Layer	Net Depth	MMSE	MPSNR	MSSIM	MLPIPS	Time (hrs)
0	2	0.0012	29.4149	0.9668	0.0829	12.09
2	4	0.0008	31.4058	0.9813	0.0482	14.62
3	6	0.0007	31.9831	0.9843	0.0399	18.47
4	8	0.0006	32.3598	0.9861	0.0348	22.36

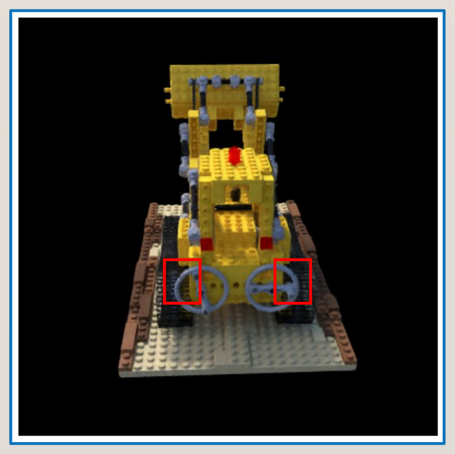
Table 2: SiLU Performance Metrics with Sinosoidal Encoding

Skip Layer	Net Depth	MMSE	MPSNR	MSSIM	MLPIPS	Time (hrs)
0	2	0.0012	29.3488	0.9663	0.0819	11.66
2	4	0.0008	31.2907	0.9807	0.0492	14.58
3	6	0.0007	31.8668	0.9841	0.0399	18.47
4	8	0.0007	32.1549	0.9854	0.0362	22.36

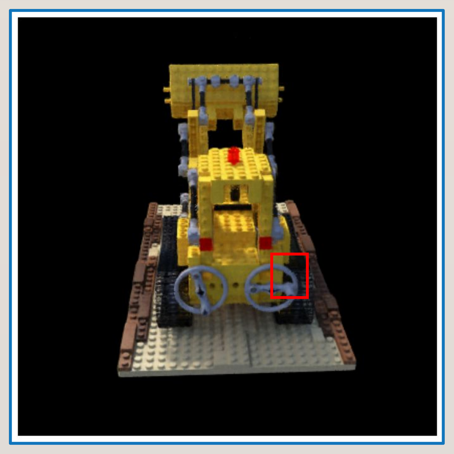
Table 3: SiLU Performance Metrics Using Spherical Harmonics Encoding

Skip Layer	Net Depth	MMSE	MPSNR	MSSIM	MLPIPS	Time (hrs)
0	2	0.0012	29.5260	0.9679	0.0791	11.67
2	4	0.0008	31.4819	0.9819	0.0470	14.58
3	6	0.0007	31.9814	0.9846	0.0389	18.47
4	8	0.0007	32.1551	0.9855	0.0362	22.36

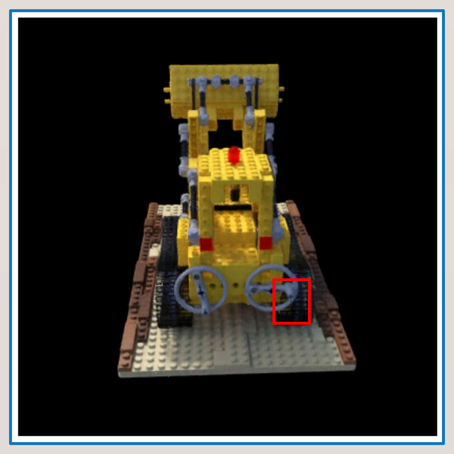
Table 4: GeLU Performance Metrics Using Spherical Harmonics Encoding



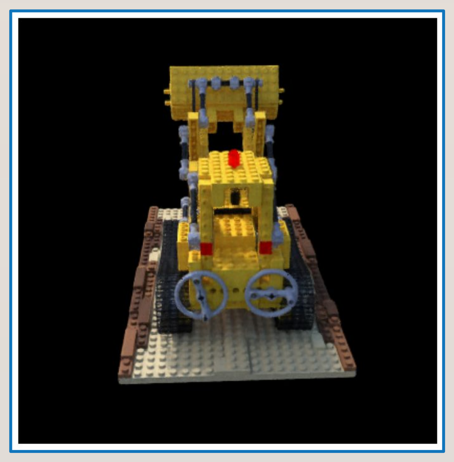
SiLU with sinusoidal encoding



GeLU with sinusoidal encoding



SiLU with spherical encoding



GeLU with spherical encoding

Testing on Real Datasets

- GeLU activation function with spherical encoding gave the best results for synthetic dataset, so we applied it to real dataset.
- We can see that spherical encoding helps increase the performance of models with smaller number of net depth and skip layers. For L2d4 the rendered image is comparable to L8d4.
- Clearly, the MPSNR values are lesser for real datasets compared to synthetic datasets.

Skip Layers: 0
Net Depth: 2

Skip Layers: 2
Net Depth: 4

Skip Layers: 4
Net Depth: 8

Skip Layer	Net Depth	MMSE	MPSNR	MSSIM	MLPIPS	Time (hrs)
0	2	0.0053	22.9144	0.6309	0.0791	10.69
2	4	0.0039	24.3335	0.7240	0.3814	14.58
3	6	-	-	-	-	-
4	8	0.0030	25.3958	0.7733	0.2617	21.38

Table 5: GeLU metrics with Spherical Encoding on Real Dataset

github repo

References

[1] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NERF: Representing scenes as neural radiance fields for view synthesis. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2003.08934

[2] Raha, A. R. G. a. R. (2022, June 15). Computer Graphics and Deep Learning with NeRF using TensorFlow and Keras: Part 2 - PyImageSearch. PyImageSearch. https://pyimagesearch.com/2021/11/17/computer-graphics-and-deep-learning-with-nerf-using-tensorflow-and-keras-part-2/