LEARNING BASED 3D REPRESENTATION AND RENDERING

INTRODUCTION

Neural network-based approach to 3D scene reconstruction from 2D images.

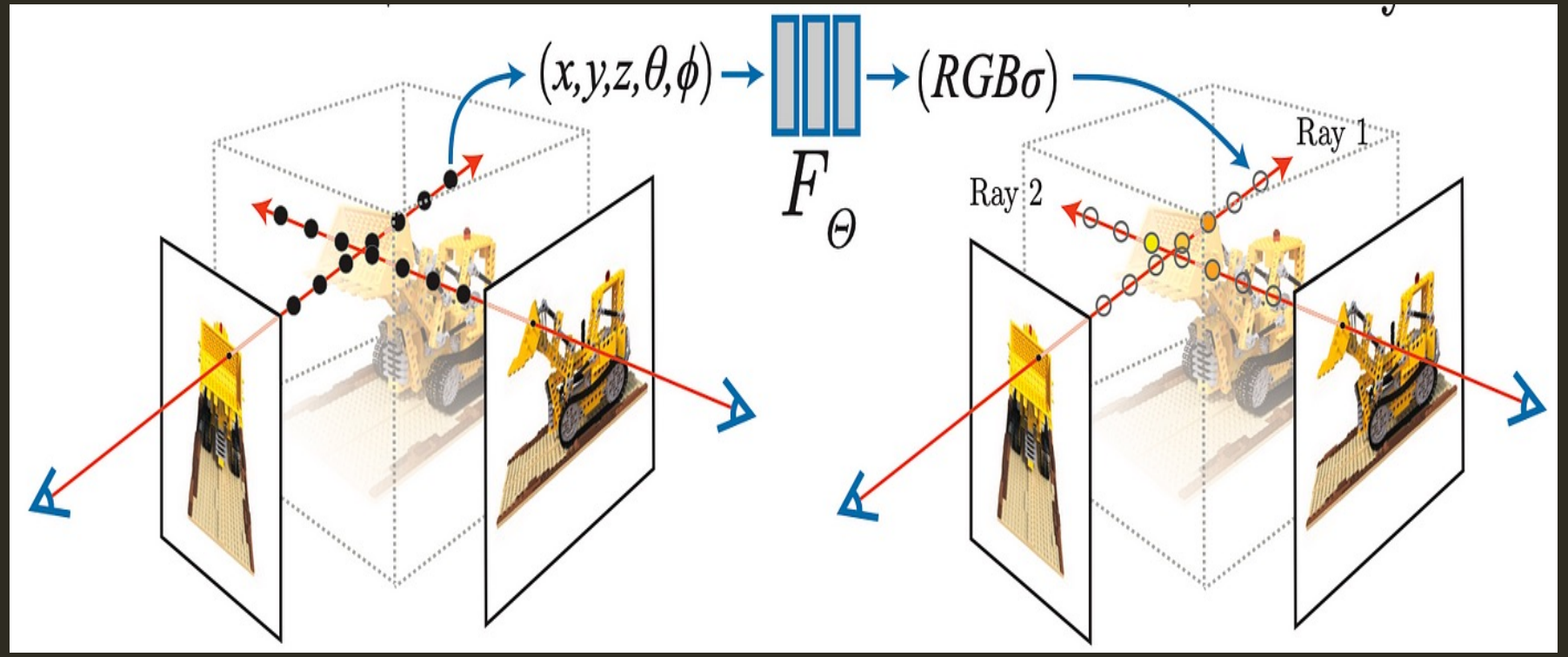
3D representation provides more realistic views of objects and scenes compared to 2D representation.

Aids to capture depth, perspective, and shading of an object or scene.

Various applications such as

- video games, virtual reality, architecture, and medical imaging, among others.

ARCHITECTURE



METHODOLOGY

NeRF (Neural Radiance Fields) is a deep learning-based approach.

A fully connected feedforward network or a convolutional network that takes 5D input vector.

The 5D input vector is comprised of a 3D location (x, y, z) and a 2D viewing direction (θ, φ).

To create a NeRF, we first capture a set of images of the scene from various viewpoints.

Structure-from-motion algorithm is used to estimate the camera poses and create a sparse point cloud of the scene.

Training with differentiable rendering algorithm to backpropagate error between the rendered image and the actual image.

Learns a continuous 3D function that maps a 3D coordinate to the radiance or color and opacity of the corresponding point in the scene.

The radiance field function is represented using a multi-layer perceptron (MLP) neural network.

Outputs the RGB color and density of the scene at that particular point.

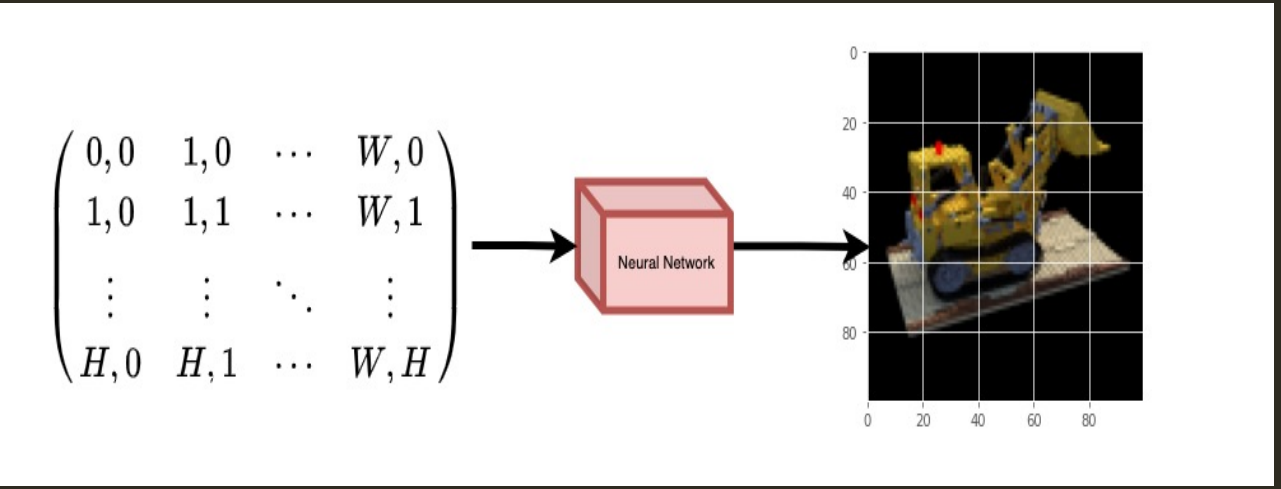
Training - network learns to predict the color and opacity of a voxel given its 3D location and the camera viewpoint.

Rendering - To render an image, the network is queried at each pixel in the output image for the corresponding 3D location in the voxel grid.

Optimization - the network is fine-tuned using an optimization process that minimizes the difference between the predicted and ground truth images.

IMPLEMENTATION DETAILS

1. Learning the 3D scene:
   - A neural network is used to learn a 3D volumetric scene.
   - The network learns to represent the scene as a function that takes in a 3D position and outputs the corresponding color value



2. Discard use of voxels:
   - It requires large number of voxels to represent a complex scene
   - The images of the scene are immediately used by the neural network to train to model.

3. Generating novel views:

- A new image of a 3D scene is generated taken from a camera position
- The network takes in the camera pose as input and produces a set of rays through each pixel of the image
- The network then uses the learned volumetric scene function to estimate the color and opacity of the scene along each ray.
- It combines along each ray to produce final color value for each pixel. It is repeated on each pixel to generate novel view of the scene from new camera pose

DATA SET

106 Legos images

RGB color channel

Camera poses – Camera position

Focal length – Angle of view

DEMO

Video

RESULTS AND CONCLUSION

The model represents the entire 3D scene using 2D training images.

Despite the sparse set of training images, the model performed well.