

# Milestone for CV Final Project: NeRF

Xiaole Wang, Gaoxiang Ye (Equal Contribution)

January 17, 2024

Additional results of our ablation studies (including images for each epoch and the PSNR curve), as well as our source code are at [this link](#).

## 1 Introduction

Neural radiance fields (NeRFs) is a novel implicit 3D scene representation and a new approach to novel view synthesis, achieving outstanding visual fidelity. In this work, we attempt to implement one such NeRF and explore possible extensions to it, by generalize it to dynamic scenes and modifying it to target a non-NVS use case.

## 2 Problem Statement

### 2.1 Description

The problems we try to solve in this work are four-fold:

1. Implement positional encoding and two-dimensional neural rendering: two important building blocks for NeRF.
2. Implement a version of the original NeRF.
3. Implement a neural radiance field that supports dynamic scenes. We plan to achieve this by adding time to the input of the neural network or by introducing an additional neural network that models the deformation field.
4. Apply NeRF to a novel use case. The exact use case is still under research.

### 2.2 Dataset

The following is the datasets we may use:

[Nerf-Data](#)

[Tanks and Temples -data](#)

We believe that the combination of these datasets comprehensively reflect NeRF's capacity to handle forward-facing and inward-facing, as well as synthetic and real-life scenes.

### 2.3 Expected result

Firstly, try to implement NeRF to fit one image and multi-view images. In this step, we will try our best to realize the achievements of predecessors. And realize NeRF for dynamic scenes step by step. Based on this, implement our own NeRF to make some innovation.

### 2.4 Evaluation

We plan to qualitatively assess the visual quality of the rendered results, as well as take into account metrics such as PSNR and SSIM.

## 3 Technical Approach

The ideas, algorithms and tricks used in the original NeRF and related works are well-conceived and have been thoroughly tested. Thus, we consider it best to draw inspiration from existing works as to our technical approach.

To be more specific, we will reference three prominent papers: [NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis](#) for vanilla NeRF, [D-NeRF: Neural Radiance Fields for Dynamic Scenes](#) for deformation field-based dynamic NeRF, and [Space-time Neural Irradiance Fields for Free-Viewpoint Video](#) for space-time neural network-based dynamic NeRF.

## 4 Preliminary Results

Now we have completed the task 1(Fit a single image with positional encoding), and examples for our preliminary result about task 1 are as following(The left is our result). We have also performed an ablation study on the effect of positional encoding, the number of frequencies in positional encoding, the width of the network, and its depth.



Figure 1: epoch=50, PSNR=23.37



Figure 2: epoch=50, PSNR=31.08

#### 4.1 Ablation Study on Positional Encoding

With positional encoding



No positional encoding



Figure 3: W/Wo Positional Encoding

The difference between the results with and without positional encoding are apparent. The no-positional-encoding version struggles to capture fine details in the image, only rendering rough geometric shapes. We speculate this is due to the difference in neighboring pixel coordinates sufficiently emphasized in later elements in the positional encoding and the lack thereof in the no-positional-encoding version. Note the lack in representation power manifests as "blocky" artifacts.

## 4.2 Ablation Study on Number of Frequencies in Positional Encoding



Figure 4: Diff frequencies in Positional Encoding

Perhaps unsurprisingly, as the number of frequencies go up, so does the quality of the image. However, we also noted that the difference between  $L = 15$  and  $L = 30$  is not so apparent.

## 4.3 Ablation Study on Depth of Neural Network



Figure 5: Diff depth of NN

Much like with the number of frequencies, the visual quality goes up with the depth of the network, with diminishing returns.

## 4.4 Ablation Study on Width of Neural Network



Figure 6: Diff width of NN

Still, the visual quality goes up with the width of the network, also with diminishing returns. It seems that the width of the model has a larger impact than its depth. At the thinnest width, we can notice "gridy" artifacts.