NeRF

Overview

Neural Radiance Fields (NeRF) 是一种基于学习的三维表示方法，它使用神经网络来表示和渲染三维场景。NeRF 通过对场景进行体绘制（volume rendering），可以生成高质量的、从任意视角看的图像。这种技术特别适用于捕捉复杂的光照效应和细节，如反射、折射、散射和阴影。 NeRF 的工作原理是将场景编码为一个连续的辐射场，其中每个位置都有一个颜色和密度值。神经网络被训练来预测这些值，给定一个三维空间中的位置和观察方向。在渲染过程中，通过沿着从相机位置出发的光线进行积分，可以计算出最终的像素颜色。 以下是 NeRF 的一些关键特点和应用：

1. **高质量渲染**：NeRF 能够生成具有真实感光照和细节的图像，这在传统渲染方法中很难实现。
2. **视图合成**：由于 NeRF 是基于场景的隐式表示，它可以渲染出训练期间未见的视角的图像。
3. **轻量级**：与传统的三维模型表示（如网格或点云）相比，NeRF 可以更紧凑地表示复杂的场景。
4. **灵活性强**：NeRF 可以轻松地与其他基于学习的方法集成，如深度学习模型，用于进一步的任务，如视频合成、风格迁移等。
5. **摄影测量应用**：NeRF 可以用于从一组照片中重建三维场景，这在摄影测量和虚拟现实中有广泛应用。NeRF 的出现为计算机视觉和图形学领域带来了新的研究方向，并在多个应用中展示了其潜力。随着技术的不断进步，NeRF 及其变体可能会在未来的三维内容创建和虚拟现实领域发挥更加重要的作用。

Neural Rendering

> "deep image or video generation approaches that enable explicit or implicit control of scene properties such as illumination, camera parameters, pose, geometry, appearance, and semantic structure."

Neural volume rendering

> "Methods that generate images or video by tracing a ray into the scene and taking an integral of some sort over the length of the ray. Typically a neural network like a multi-layer perceptron encodes a function from the 3D coordinates on the ray to quantities like density and color, which are integrated to yield an image."

Occupancy Nerworks

A network f_θ that predict the probability whether the point in 3D space $p \in \mathbb{R}^3$ is occupied while given the obervation $x \in \chi$

$$f_\theta : \mathbb{R}^3 \times \chi \to [0, 1]$$

Neural Radiance Fields

Reference:

> Ben Mildenhall, et al. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis

Using a neural network to fit a 5D continuous function f that

$$f : (x, y, z, \phi, \theta) \to (c, \sigma)$$

in which x, y, z is the spatial point, ϕ, θ is the angle of the ray, σ indicate the opacity of the point, $c = (r, g, b)$ represents the intensity of the ray.

In this approach the 3D reconstruction is equivalent to training the neural network f by giving the observed images from multiple views. Also, we can gennerate a new image by volume rendering.

Problems: 1. difficult in high resolution. 2. inefficient in few samples. So

> We address these issues by transforming input 5D coordinates with a positional encoding that enables the MLP to represent higher frequency functions, and we propose a hierarchical sampling procedure to reduce the number of queries required to adequately sample this high-frequency scene representation.

Basic frame:

> We encourage the representation to be multiview consistent by restrictingthe network to predict the volume density as a function of only the location x, while allowing the RGB color c to be predicted as a function of both location and viewing direction. To accomplish this, the MLP F rst processes the input 3D coordinate x with 8 fully-connected layers (using ReLU activations and 256 channels per layer), and outputs and a 256-dimensional feature vector. This feature vector is then concatenated with the camera rays viewing direction and passed to one additional fully-connected layer (using a ReLU activation and 128 channels) that output the view-dependent RGB color.

Some tricks refers to the the original refference.