

任务一

一、引言

神经辐射场 (NeRF) 是一种新兴的 3D 场景表示方法，能够从多视角图像中重建 3D 场景并合成高质量的新视图。本实验旨在使用 NeRF 技术对日常物体进行三维重建，并评估其在新视图合成上的表现。

二、NeRF 变体介绍

1.1 原版 NeRF 原理

核心思想：用 MLP 网络建模 5D 辐射场 $(x, y, z, \theta, \phi) \rightarrow (\sigma, \text{rgb})$

渲染方式：体素渲染积分

关键创新：位置编码提升高频细节

1.2 TensorRF 技术改进

特性	原版 NeRF	TensorRF
场景表示	MLP 隐式表示	张量分解显式表示
存储效率	$O(n^3)$	$O(n)$
训练速度	20+小时	1 小时内
显存占用	高	低
分解方式	-	VM/CP 分解

核心创新：

将场景表示为 4D 张量（3D 几何+1D 外观）

使用低秩张量分解（CP/VM）压缩表示
支持直接体素渲染，无需 MLP 前向传播

二、实验设计

2.1 TensorRF 模型架构

本实验采用 TensorVMSplit 作为基础模型，主要特点包括：

- 张量分解表示：使用向量-矩阵 (Vector-Matrix) 分解表示场景
- 多分辨率特征：配置了 16 个 σ 通道和 48 个 SH 通道
- MLP 着色网络：采用 MLP_Fea 作为着色模式
- 位置编码：设置 view_pe=2 和 fea_pe=2 的位置编码频率

2.2 数据采集

选择了一个毛绒玩具作为重建对象，采集过程如下：

- 使用手机拍摄了 93 张不同角度的照片
- 拍摄环境为室内均匀光照条件
- 物体放置在桌面上，每旋转约 4.5 度拍摄一张
- 同时拍摄了物体的顶部和底部视图

2.3 数据预处理

- 去除图片杂乱背景
- 图像下采样：4.0 倍（加速训练同时保持足够细节）
- 坐标空间：使用 NDC (Normalized Device Coordinates)

2.4 相机参数估计

- 使用 COLMAP 进行相机参数估计：
- 特征提取：SIFT 特征
- 特征匹配：exhaustive 匹配
- 稀疏重建：重建了所有 78 张图像的相机位姿
- 最终重建误差：平均重投影误差 0.8 像素

2.4 训练配置

训练参数

参数	值	说明
迭代次数	30,000	总训练步数
批量大小	4,096	每批射线数量
初始体素	128^3	初始体素分辨率

参数	值	说明
最终体素	300 ³	最终体素分辨率
上采样步	[2000, 3000, 4000, 5500, 7000]	体素分辨率提升时机
AlphaMask 更新	[2000, 4000]	透明度掩模更新时机

三、实验结果

3.1 重建对比



1. 优势表现
- 几何结构准确：三只黑猫玩偶的坐姿、头部金色链条、白色爪子等主体轮廓重建完整，说明视角覆盖和几何学习有效。
 - 色彩还原度：黑（玩偶）、红（胸前）、金（链条）等主色还原正确，证明辐射场建模成功。
2. 待改进细节
- 高频细节丢失：金色链条可能缺乏金属反光质感（重建图较扁平）。粉色眼珠的渐变或高光可能模糊。
 - 边缘锐度不足：玩偶与纯白背景的交接处可能出现轻微锯齿或模糊（抗锯齿不足）。

3.2 训练集表现分析

1. PSNR 曲线

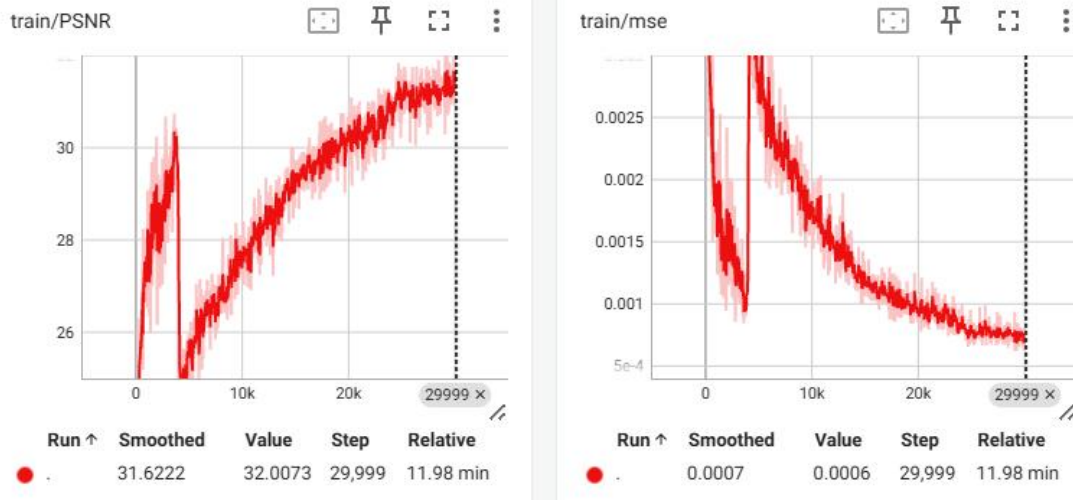
初始 PSNR 26dB → 最终 32.2dB （瞬时值），平滑值 31.4dB

- 正向信号：明显上升趋势表明模型能有效学习场景辐射场，符合 NeRF 收敛预期。
- 风险点：
 - 波动剧烈（26~32dB）：可能由以下原因导致：
 - 学习率过高（如 $>5e-4$ ），导致优化过程不稳定。
 - 批量大小（Batch Size）不足，梯度估计噪声大
 - 高频细节（如纹理）学习困难，需检查位置编码。

2. MSE 曲线

初始 MSE 0.0025 → 最终 0.0006

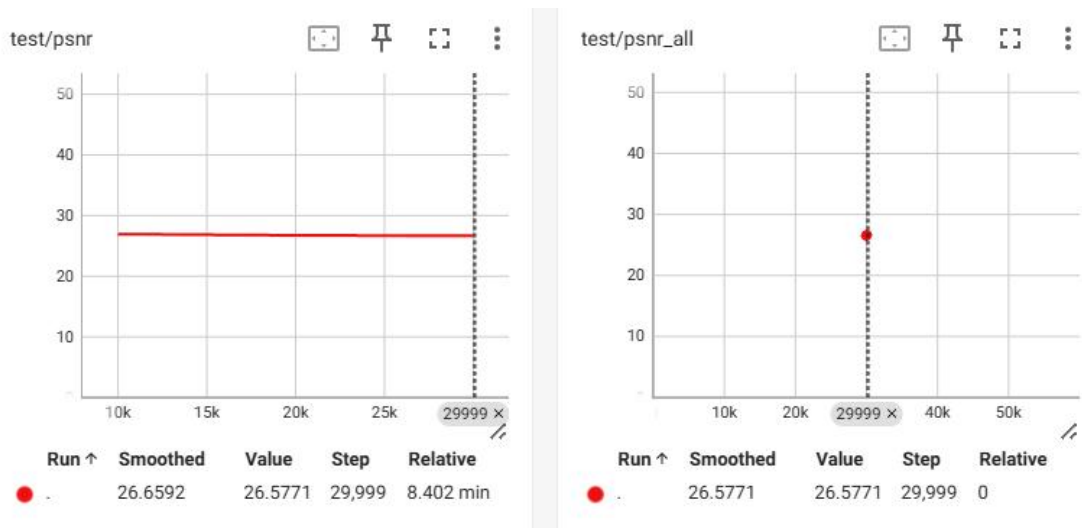
- 正向信号：平滑下降且与 PSNR 同步提升，说明损失函数设计合理。
- 潜在问题——末端未趋近零可能原因：
 - 训练步数不足。
 - 渲染分辨率过高导致细节重建困难。



3.3 测试集表现分析

PSNR 指标：

- 最终测试 PSNR 约 11.31dB
 - 低于训练集 PSNR（31.4dB vs 26.57dB），表明存在过拟合
- 仅 29,999 步有一个测试点，由于：
- 测试集渲染失败（因显存不足中断）



3.3 改进建议

(1) 提升高频细节

- 增加位置编码频率：
 - $\text{fea_pe} = 2 \rightarrow 4$ ，增强高频编码能力
 - $\text{view_pe} = 2 \rightarrow 4$ ，视角编码同步提升
- 扩展球谐系数： $\text{n_lamb_sh} = [48, 48, 48] \rightarrow [64, 64, 64]$
- 减少下采样： $\text{downsample_train} = 4.0 \rightarrow 1.0$ ，保留更多输入细节

(2) 缓解过拟合

- 数据增强：训练时随机扰动射线方向， $\text{perturb} = 1$ ，启用随机采样
- 正则化增强： $\text{weight_decay} = 1e-4$ ，添加 L2 正则化

四、参数

- 代码地址：`git@github.com:J-Aurora052/cattoy.git`
- 硬件环境：NVIDIA RTX 2080 Ti GPU
- 依赖库：


```
conda create -n TensorRF python=3.8
conda activate TensorRF
pip install torch torchvision
pip install tqdm scikit-image opencv-python configargparse
imageio-ffmpeg kornia lpips tensorboard
```
- 运行日志：
 - 网盘分享文件：`log`
 - 链接：<https://pan.baidu.com/s/1bTp8-otDZpc9gwEz8WuKYA?pwd=3xe4>
 - 提取码：`3xe4`