
NeRF Project

鲍彦泽
2200012607

孔令珊
2400013075

杨汶帛
2400013559

Abstract

本项目实现了 NeRF 从静态到动态场景的完整技术链，包括位置编码、多视图重建、哈希编码加速、变形场-规范场解耦。我们创新性地提出 Dual-Hash 架构，在提升动态场景的训练效率和时空表达能力方面做出了尝试。

1 引言

NeRF 是近年来计算机视觉领域的重要突破，开创了基于神经网络的新视图合成范式。本项目从静态到动态场景逐步实现 NeRF 技术链，并针对训练效率与时空表达问题提出了创新架构。GitHub 仓库 中包含了本项目的完整代码、详细项目框架说明与实验复现指南。本文附录 A 中提供了几个代表性实验的训练日志，与本报告一同上交的附件中包含了更完整的实验结果，更大的文件（如 checkpoint 等）位于 北大网盘 。

2 简介

2.1 主要原理与方法

2.1.1 神经辐射场基础

本项目遵循 Mildenhall 等人提出的经典神经辐射场（NeRF）框架 [1]。NeRF 使用多层感知机 $F_\Theta : (\mathbf{x}, \mathbf{d}) \rightarrow (\sigma, \mathbf{c})$ 隐式表示场景。像素颜色通过沿光线的体渲染积分计算：

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t) \sigma(\mathbf{r}(t)) \mathbf{c}(\mathbf{r}(t), \mathbf{d}) dt \quad (1)$$

其中 $T(t) = \exp(-\int_{t_n}^t \sigma(\mathbf{r}(s)) ds)$ 为累积透射率。

2.1.2 位置编码

为缓解 MLP 的低频偏置，NeRF 使用位置编码 $\gamma(\cdot)$ 将输入映射至高维频域：

$$\gamma(p) = (\sin(2^0 \pi p), \cos(2^0 \pi p), \dots, \sin(2^{L-1} \pi p), \cos(2^{L-1} \pi p)) \quad (2)$$

2.1.3 加速机制：多分辨率哈希编码与占据网格

针对原版 NeRF 训练缓慢的问题，我们借鉴了 Instant-NGP 的加速思想 [3]。Instant-NeRF 采用**多分辨率哈希编码**替代 Fourier 编码，将 3D 空间划分为 16 个分辨率层级，每个空间点通过哈希映射索引至 2^{19} 大小的特征表，并经由三线性插值获取可学习特征，极大降低了计算开销并实现分钟级训练。

引入 128^3 分辨率的**占据网格 (Density Grid)**，维护粗粒度的场景二值密度状态 ($\sigma > \tau$)。在光线行进过程中，利用占据网格进行空域跳跃，仅在有效区域进行精细采样，并动态更新网格，显著提升了渲染效率。

2.2 数据

2.2.1 数据集

1、**NeRF Synthetic** 包含 Lego、Drums、Ficus、Hotdog、Ship 等静态合成场景，每个场景包含 100 张训练图像和 200 张测试图像

2、**D-NeRF** 包含 Standup、Trex、Jumpingjacks、Hook、Mutant 等动态场景

2.2.2 数据增强

为了提升模型的泛化能力和鲁棒性，我们在训练过程中引入了多种数据增强策略：

1、**光线采样扰动**：在粗采样阶段，对分层采样的每个区间内进行随机扰动 ($t_i \sim \mathcal{U}[t_{n,i}, t_{f,i}]$)，而非固定取中点。这有效地将离散采样转化为对连续光线的积分近似，防止网络过拟合到固定的深度采样点。

2、**随机背景**：为了防止模型在透明区域或物体边缘的渲染出伪影，我们在训练过程中实施随机背景增强。具体而言，对于每个训练批次，我们从均匀分布 $\mathbf{c}_{bg} \sim \mathcal{U}[0, 1]^3$ 中采样一个随机背景颜色，并基于训练图像的 Alpha 通道动态合成目标颜色：

$$C_{target} = C_{gt} \cdot \alpha_{gt} + \mathbf{c}_{bg} \cdot (1 - \alpha_{gt}) \quad (3)$$

同时，渲染方程的背景项也替换为该随机颜色。这种强约束迫使模型正确分离前景和背景，从而显著提高了复杂轮廓的边缘质量和几何重建精度。

3、**坐标与时间噪声注入 (D-NeRF 专用)**：针对动态场景稀疏视角易过拟合的问题，我们在变形网络的输入端引入高斯噪声：

$$\mathbf{x}' = \mathbf{x} + \mathcal{N}(0, \sigma_x^2), \quad t' = t + \mathcal{N}(0, \sigma_t^2) \quad (4)$$

其中 $\sigma_x = 0.005, \sigma_t = 0.04$ 。这种扰动强制变形场在局部邻域内保持平滑 (Lipschitz 连续)，显著减少了新视角下的几何抖动。

3 实验

3.1 实验一：单图像拟合与位置编码

3.1.1 架构与方法

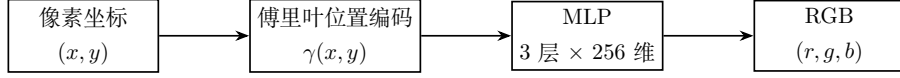


图 1: 单图像拟合架构流程图

3.1.2 实验结果

训练 10 min 取得 PSNR 为 53.35 dB 的结果图 (见附件), 消融实验结果 (见附件) 表明: 1、无位置编码时, 网络只能拟合低频平滑区域, 细节模糊, PSNR 显著下降 2、频率数量要适中, 太低会模糊, 太高有噪声 3、模型加深加宽, 效果会变好, 但是也要综合考虑存储、训练时间和过拟合问题。

3.2 实验二：多视图 NeRF 实现

3.2.1 架构与方法

1. 场景表示与位置编码 采用标准 NeRF 表达场景, 输入为位置 \mathbf{x} 与视角 \mathbf{d} ; 位置编码 $L = 10$, 方向编码 $L = 4$, 由相机位姿生成光线采样点。

2. 网络结构 主干为 8 层 256 维 MLP, 第 4 层 skip; 密度头输出 σ , 颜色支路拼接视角编码后 1 层 MLP 输出 RGB (Sigmoid), 结构如图 1。

3. 体渲染 在 $[t_n, t_f]$ 上分层采样 $N = 64$ (训练扰动), 体渲染加权求和得到颜色, 训练采用 MSE, 评价指标为 PSNR。

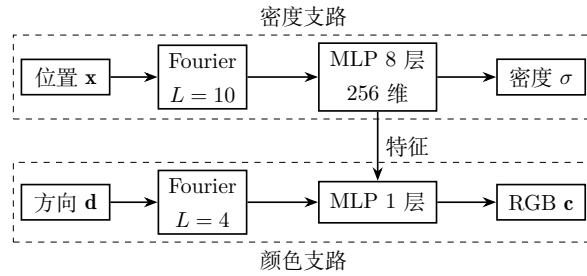


图 2: 标准 NeRF 架构流程图

3.2.2 实验结果

在 NeRF Synthetic 的 Lego 场景上, 训练 20k steps (batch 4096, 5×10^{-4} , $N = 64$, downscale=2), 测试集 PSNR 为 **24.75 dB**。渲染结果主结构与光照一致, 但细节与边缘仍偏软; 已输出渲染帧与视频作为定性结果。

3.3 优化改进：Instant-NeRF

基于 Instant-NGP 实现加速：(1) 多分辨率哈希编码 (16 层, 2^{19}) 替代 Fourier 编码；(2) Tiny MLP (密度 1 层 + 颜色 2 层)；(3) 128^3 占据网格 实现空域跳跃。在笔记本 GPU 上训练 lego, 1 分钟细节可见, 5 分钟达到 26+ dB, 渲染速度 10+ FPS。

3.3.1 性能对比

3.4 实验三：动态场景扩展

3.4.1 架构与方法

架构 1 添加时间编码至 NeRF 输入：将原始三维位置 \mathbf{x} 、时间戳 t 和视角方向 \mathbf{d} 分别进行 Fourier 编码后直接拼接，作为单一 MLP 的输入，直接输出 RGB 颜色和体密度 (σ, \mathbf{c}) 。此方法结构简单，但网络需同时学习复杂的时空映射，导致训练收敛较慢且难以解耦运动与几何。

架构 2 通过解耦变形场和规范场建模动态场景 [2]：

变形网络 Ψ_t ：将观察空间 (\mathbf{x}, t) 映射到规范空间

$$\mathbf{x}_{\text{canon}} = \mathbf{x} + \Psi_t(\mathbf{x}, t) \quad (5)$$

采用 4 层 MLP (128 维)，输入为位置和时间的 Fourier 编码。

规范网络 Ψ_x ：在规范空间表示静态场景

$$\Psi_x(\mathbf{x}_{\text{canon}}, \mathbf{d}, t) \rightarrow (\mathbf{c}, \sigma) \quad (6)$$

采用 8 层 NeRF MLP，额外拼接时间特征。

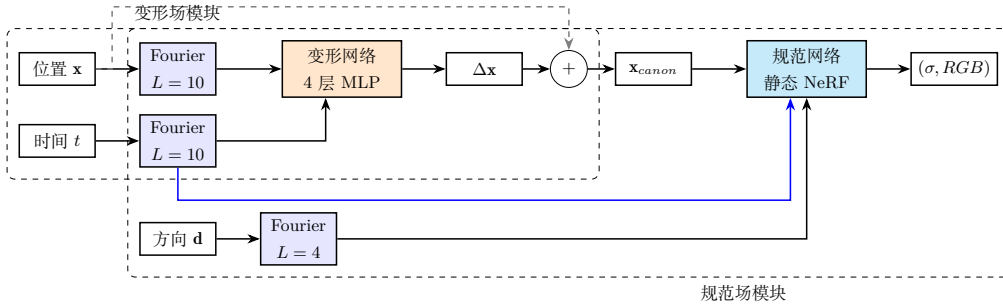


图 3: D-NeRF 架构 2

3.4.2 探索改进：D-NeRF (Instant 混合架构)

1、混合架构：变形场保留 Fourier MLP (4 层 \times 128 维)，规范场替换为哈希网格 + Tiny MLP，兼顾变形解耦能力与 $O(1)$ 查询效率。

2、约束机制：形变 L2 正则 + 形变 TV 正则 + 时序平滑正则 消除哈希网格噪声与时空抖动；坐标/时间噪声注入 强制平滑；随机背景 与 体积守恒约束 增强泛化。

3、动态占据网格：借鉴 TiNeuVox [4] 的时间感知体素策略，采用多时刻联合采样更新 128^3 占据网格，实现时空跳跃加速。通过冻结冷启动、分段调整更新速率、后期再冻结的策略，充分兼顾训练活性与稳定性。

性能提升：通过这些改进，在我们的笔记本显卡上，D-NeRF 第一次有望实现收敛

3.4.3 实验结果

架构	训练步数	训练时间	test PSNR	收敛性
架构 1（直接时间编码）	50000	~16 h	~23 dB	难以收敛
架构 2（变形场解耦）	40000	~13 h	~25 dB	较难收敛，收敛平台低
改进架构（Instant 混合）	20000	~2 h	~30 dB	较快收敛

在 8GB Nvidia GPU Laptop 4060 上进行实验，结果如表：(1) 架构 1 由于时空耦合严重，MLP 庞大干扰严重，预测图像一直非常模糊；(2) 架构 2 通过解耦一定程度上改善了效果，但全 MLP 结构仍然训练缓慢；(3) 改进架构将规范场替换为 Instant-NeRF，使得图像质量显著提升，训练时间大幅缩短。

3.5 实验四：NeRF 创新 – Dual-Hash Dynamic NeRF

在 D-NeRF 训练中，将规范网络替换为 Instant-NeRF 后仍面临训练慢、高频震荡和动静耦合问题。分析表明形变网络的 MLP 巨量参数、难解耦是主因。受 NDVG [5] 显式体素位移场启发，我们创新性地设计了 **Dual-Hash 架构**，将 Instant-NGP [3] 的多分辨率哈希编码应用于变形场建模，用 3 个哈希形变网格替代 MLP，利用网络的局部支持性实现动静解耦，显著提升收敛速度并消除高频震荡。

3.5.1 架构与方法

针对上述问题，我们创新性地提出 Dual-Hash 架构，细节如下：

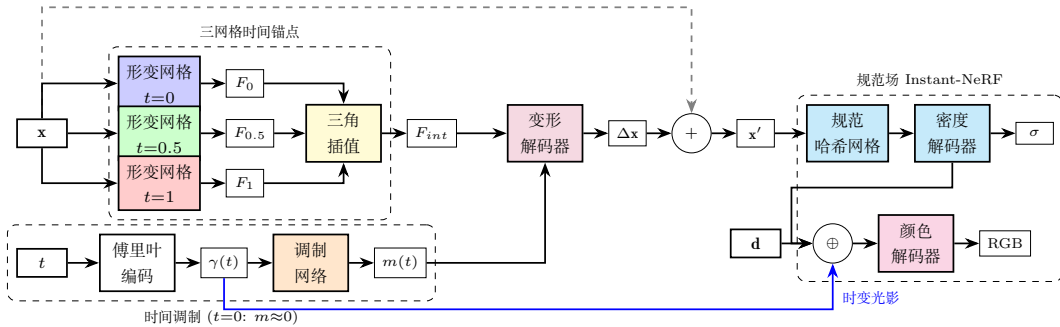


图 4: Dual-Hash 架构

1. 三网格时间锚点插值

针对时变运动的复杂非线性，借鉴 K-Planes [6] 显式时空数据结构思想，我们建立三个独立的时间锚点网格 ($t = 0, 0.5, 1$)，通过三角形加权实现时间连续插值。

$$w_i = \max\left(0, 1 - \frac{|t - t_i|}{0.5}\right), \quad F_{interp} = \sum_i \frac{w_i}{\sum_j w_j} F_i \quad (7)$$

该设计保证了特征 C^1 连续性，且由三个锚点分担不同时间段的运动信息，避免了将时间直接输入哈希表导致的抖动。

2. 零偏置锚定

时间调制网络采用零偏置锚定策略初始化偏置参数，从物理上强制 $t = 0$ 时刻变形场输出为零，确保初始状态与规范空间严格对齐。

3. 时变光影

除了几何变形外，我们将时间特征直接拼接到规范场哈希特征中（图中蓝色连线），增强了模型对随时间动态变化的光照效果（如高光移动）的建模能力。

4. 多重正则化体系

- (1) **形变 L2 正则**: $\lambda_1 \|\Delta \mathbf{x}\|^2$ 约束位移幅度，防止过大形变导致几何畸变。
- (2) **形变 TV 正则**: $\lambda_2 \|\nabla_x \Delta \mathbf{x}\|_1$ 位移场全变分正则化，消除空间抖动 [7]。
- (3) **时序平滑正则**: $\lambda_3 \|\Delta \mathbf{x}(t) - \Delta \mathbf{x}(t + \epsilon)\|^2$ 约束相邻时刻位移变化，保证时间连续性。
- (4) **静态锚点约束**: $\lambda_4 \|\Delta \mathbf{x}(t = 0)\|^2$ 强制初始时刻无变形，确保规范空间稳定无漂移。
- (5) **无监督一致性**: 体积守恒约束，增强物理合理性并提高泛化能力 [7]。

3.5.2 实验结果

在 Standup 数据集上的一次训练结果如下，通过观察发现，Dual-Hash Dynamic NeRF 显著提升了训练速度，且成功消除了 MLP 在尝试拟合高频边缘时会产生过的拟合闪动（可以详细看渲染结果的帽檐，并和使用 MLP 形变网络的 part 3 对比）

训练时间	训练步数	训练集 PSNR	验证集 PSNR	测试集 PSNR
8.759 分钟	5000	32.41 dB	29.56 dB	30.58 dB

但是，由于显卡资源有限，我们仍未能在所有数据集上实现完全收敛，渲染质量还存在瓶颈，未来工作将继续优化训练流程和超参数设置，以期达到更优性能。

4 结论

本项目实现了 NeRF 从静态到动态场景的完整技术链，包括位置编码、多视图重建、哈希加速、变形场解耦。针对传统 MLP 变形网络的不足，创新性地提出 Dual-Hash 架构，用三个时间锚点哈希网格替代全连接网络，在 Standup 数据集上 5000 步达到 30.58 dB PSNR，显著提升训练效率并消除高频边缘闪动，为动态 NeRF 提供了新思路。

References

- [1] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. *NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis*. In *European Conference on Computer Vision (ECCV)*, 2020.
- [2] Pumarola, A., Corona, E., Pons-Moll, G., and Moreno-Noguer, F. *D-NeRF: Neural Radiance Fields for Dynamic Scenes*. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [3] Müller, T., Evans, A., Schied, C., and Keller, A. *Instant Neural Graphics Primitives with a Multiresolution Hash Encoding*. *ACM Transactions on Graphics (TOG)*, 41(4):1–15, 2022.
- [4] Fang, J., Yi, T., Wang, X., Xie, L., Zhang, X., Liu, W., NieSSner, M., and Tian, Q. *Fast Dynamic Radiance Fields with Time-Aware Neural Voxels*. In *SIGGRAPH Asia*, 2022.
- [5] Wang, L., Wang, L., Lin, Y., Zhang, S., and Cai, Y. *Neural Deformable Voxel Grid for Fast Optimization of Dynamic View Synthesis*. In *Asian Conference on Computer Vision (ACCV)*, 2022.
- [6] Fridovich-Keil, S., Meanti, G., Warburg, F. R., Recht, B., and Kanazawa, A. *K-Planes: Explicit Radiance Fields in Space, Time, and Appearance*. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [7] Fang, J., Yi, T., Wang, X., Xie, L., Zhang, X., Liu, W., NieSSner, M., and Tian, Q. *DeVRf: Fast Deformable Voxel Radiance Fields for Dynamic Scenes*. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

A 训练过程可视化

本节展示了不同实验的训练过程 TensorBoard 日志，包括损失曲线和 PSNR 指标变化。

A.1 Part 2: Instant-NeRF 训练日志

Instant-NeRF 在 Lego 数据集上的训练过程如图 5 所示，可以看到模型在前 1000 步快速收敛，5 分钟内即可达到较高质量。



图 5: Part 2 Instant-NeRF 训练日志 (Lego 数据集)

A.2 Part 3: D-NeRF 混合架构训练日志

D-NeRF 混合架构（变形场 MLP + 规范场 Instant-NeRF）在 Standup 数据集上的训练过程如图 6 所示，相比纯 MLP 架构显著加速。

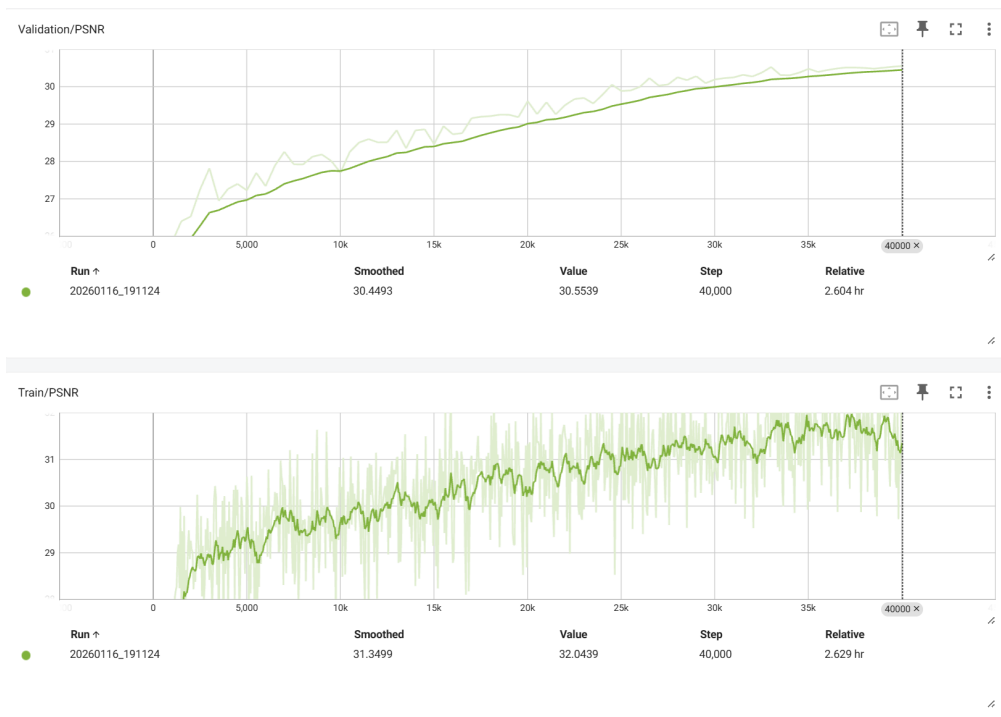


图 6: Part 3 D-NeRF 混合架构训练日志 (Standup 数据集)

A.3 Part 4: Dual-Hash 架构训练日志

Dual-Hash 架构在 Standup 数据集上的训练过程如图 7 所示，全哈希架构实现了更快的收敛速度和更稳定的训练曲线。



图 7: Part 4 Dual-Hash 架构训练日志 (Standup 数据集)