

Mip-NeRF 360

论文：《Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields》

地址：<https://arxiv.org/abs/2111.12077>

年份：CVPR 2022 (Oral)

Introduction

任务：新视角合成

技术贡献：

- (1) 对坐标系进行变换，使得能够处理无界场景；
- (2) 原始 NeRF 采用一个 MLP 得到 coarse 的采样结果，另一个得到 fine 的结果，这样会有效率的问题。Mip-NeRF 360 采用一个小 MLP + 大 MLP 的方式，小 MLP 只用于查询点，大 MLP 用于渲染，这样能借助大 MLP 建模大场景的同时保证效率；
- (3) 提出一项正则项，用于解决渲染时出现 floaters 和 background collapse 的问题。

Method

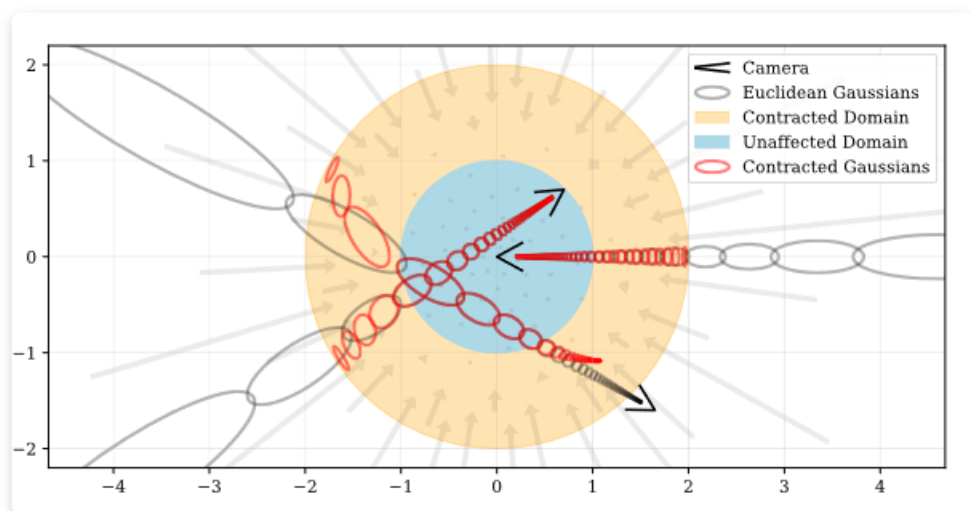
Scene and Ray Parameterization

参考 NeRF++ 或 NDC 坐标，对于无界场景需要将其坐标做变换，使其落在有界区域内。由于 Mip-NeRF 使用高斯来代替光线，所以需要由高斯做变换。

首先定义一个用于坐标变换的函数 $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ，对于 Mip-NeRF 中的一段段的高斯，用 f 将其中心 μ 进行坐标变换，其余点则可以通过 f 在 μ 处的 Jacobian 进行近似。

文章将 f 设计为以下形式：

$$\text{contract}(\mathbf{x}) = \begin{cases} \mathbf{x} & \|\mathbf{x}\| \leq 1 \\ (2 - \frac{1}{\|\mathbf{x}\|})(\frac{\mathbf{x}}{\|\mathbf{x}\|}) & \|\mathbf{x}\| > 1 \end{cases} \quad (1)$$

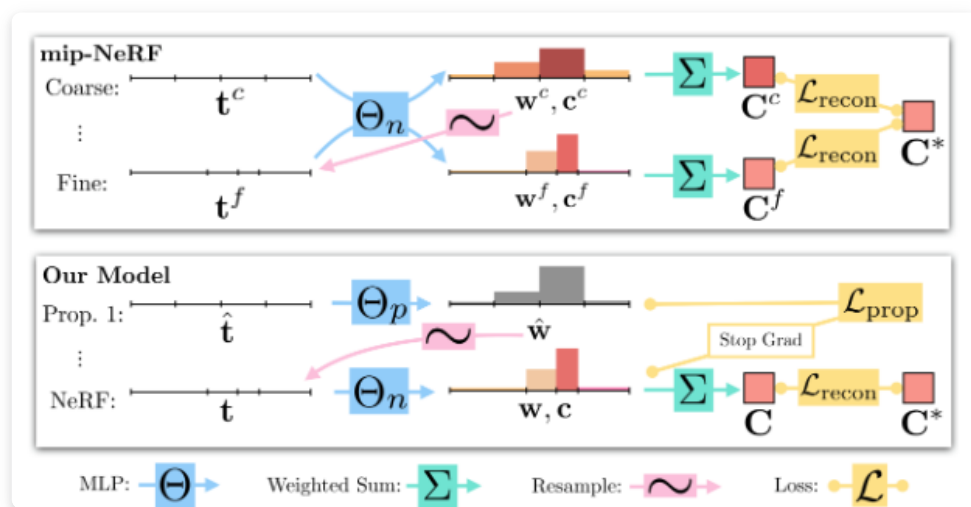


如图所示，在单位球内保持欧式空间，单位球外则压缩到半径为 1 和 2 的球壳中。作者在单位球外的间隔也不采用 t 的等距采样了，而是采用了 s 的等距采样。简单理解 s 与 $1/t$ 相关， s 的等距意味着越到

后面对应的 t 越大，这也符合了近的地方多采样，远的地方少采样的原则。

Coarse-to-Fine Online Distillation

与 Mip-NeRF 不同，这里使用两个 MLP：NeRF MLP 和 proposal MLP，proposal MLP 只用于预测体密度，根据 proposal MLP 预测结果计算权重，再在 NeRF MLP 中采样一次点，渲染图像。这样就能增加 NeRF MLP 参数量提高建模能力的同时保证渲染速度。



proposal MLP 的监督信号不再是 image loss，而是与 NeRF MLP 产生的直方图之间的 loss。文章这里说计算两个直方图之间的 loss 看起来是一件简单的事，但实际上是比较复杂的。因为即使两个直方图对应于同一个分布，由于采样点的数量不同，导致这两个直方图看上去也会有很大的差异。

文章提出的具体做法为：proposal MLP 的直方图的每一段都会有一个明确的上界，我们去约束 NeRF MLP 直方图对应的每一段的值不能超过这个上界。”we only want to penalize the proposal weights for **underestimating** the distribution implied by the NeRF MLP overestimates are to be expected, as the proposal weights will likely be more coarse than the NeRF weights, and will therefore form an upper envelope over it.“ 对于 proposal MLP 预测出来的 \hat{w} 明显低于 NeRF MLP 预测的 w 的部分，说明这部分在 proposal MLP 是 underestimated，因此通过 loss 去拉高 proposal MLP 这部分的值。对于那些反过来的情况，则不需要处理，因为这些区域不是物体所在区域，通过优化第一部分所说的权重就能让这部分的权重变小。

Regularization for Interval-Based Models

NeRF 渲染的图像通常会出现两种 artifact：一种是 floaters，通常是一些不连续的小块，是用于拟合特定视角下的训练图像产生的，在其他视角下看是一团模糊的云；另一种是 background collapse，远处的表面被错误地建模为靠近相机的密集内容的半透明云。对于第一种情况，其在光线上的权重分布会是多峰的，第二种情况的权重分布则是没有显著的峰。

正则项如下所示：

$$\mathcal{L}_{\text{dist}}(\mathbf{s}, \mathbf{w}) = \iint \mathbf{w}_s(\mathbf{u}) \mathbf{w}_s(\mathbf{v}) |u - v| d_u d_v$$

其作用是拔高单峰，压制多峰。因为想要这项 loss 小，对于距离远的 u, v ， $|u - v|$ 前面的权重值则必须要低；对于高权重的 u, v ，则其距离必须要近。

其离散形式如下所示：

$$\mathcal{L}_{\text{dist}}(\mathbf{s}, \mathbf{w}) = \sum_{i,j} w_i w_j \left| \frac{s_i + s_{i+1}}{2} - \frac{s_j + s_{j+1}}{2} \right| + \frac{1}{3} \sum_i w_i^2 (s_{i+1} - s_i)$$

具体推导过程可见 https://www.bilibili.com/video/BV1fw411v7Ey/?spm_id_from=333.1007.top_right_bar_window_default_collection.content.click&vd_source=5a666e99cac2434ec89673bca4c90abc

第一项是两个点不在光线中采样的同一个区间内，所以计算距离再乘上权重。第二项是两个点在同一个区间内的情况，此时直接计算积分结果，就可以得到以上公式。

Experiments

实验在 Mip-NeRF 360 数据集上进行，主要看一下消融实验结果。

	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	Time (hrs)	# Params
A) No $\mathcal{L}_{\text{prop}}$	20.49	0.406	0.573	6.21	9.0M
B) No $\mathcal{L}_{\text{dist}}$	24.41	0.687	0.300	7.08	9.0M
C) No $\mathcal{L}_{\text{dist}}$, w/Noise Injection	24.00	0.655	0.328	7.08	9.0M
D) No Proposal MLP	24.26	0.682	0.307	18.89	8.7M
E) No Prop. MLP w/ [3]'s Training	23.45	0.659	0.328	18.89	8.7M
F) Small NeRF MLP	22.80	0.515	0.480	4.31	1.1M
G) No IPE	23.87	0.664	0.322	7.08	9.0M
H) No Contraction	23.77	0.642	0.347	8.79	10.9M
I) w/DONeRFs Contraction [34]	23.99	0.654	0.334	7.20	9.0M
Our Complete Model	24.37	0.687	0.300	7.09	9.0M

B) 移除 $\mathcal{L}_{\text{dist}}$ 对指标基本没影响，但会影响视觉效果和几何质量，因为没有这项正则项就会出现 artifact; C) 采用原始 NeRF 中的添加扰动方法会有所提升但不如本文的方法; D) 只用一个 MLP 也能有好的结果，但会大幅增加训练时间，因为要在大的 MLP 上做采样; F) 换用小的 NeRF MLP 虽然训练快但会影响性能，大场景还是需要参数量大的 MLP。