

多空间神经辐射场*

尹泽鑫 邱佳雄 程明明 任博†

天津市视觉计算与智能感知重点实验室, 计算机学院, 南开大学

{Zexin.Yin.cn, qiujiaxiong727}@gmail.com, {cmm, rb}@nankai.edu.cn

Abstract

现有的神经辐射场 (Neural Radiance Fields, NeRF) 方法难以渲染反射物体, 通常会产生模糊或畸变的渲染效果。不同于原始的神经辐射场, 我们提出了一种多空间神经辐射场 (Multi-Space Neural Radiance Fields, MS-NeRF), 通过在平行子空间中使用一组特征场来表示场景, 从而更好地让神经网络理解反射和折射物体的存在。我们的多空间方案可以作为现有 NeRF 方法的增强模块, 只需要在训练和推理时花费很小的额外开销来计算多空间输出。我们使用三个有代表性的 NeRF 类模型 (即 NeRF、Mip-NeRF 和 Mip-NeRF 360) 来展示我们方法的优越性和兼容性。我们新构建了一个全新的数据集, 并在其上进行比较实验, 该数据集包括 25 个合成场景和 7 个真实采集场景, 所有场景都包含复杂的反射和折射并且都具有 360 度视角。

大量的实验证明, 在对由镜面产生复杂光路的场景进行高质量渲染的任务上, 我们的方法明显优于现有的单空间 NeRF 方法。

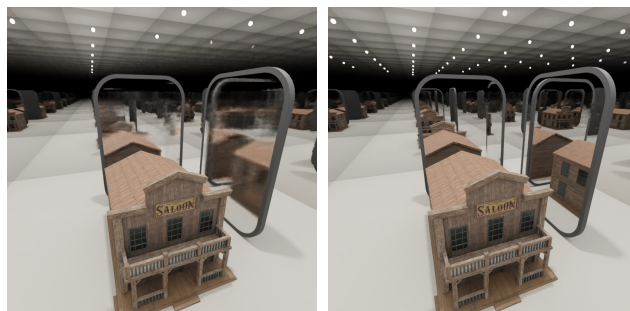
我们的代码和数据集将在以下网址公开发布: <https://zx-yin.github.io/msnerf>。

1. 引言

神经辐射场 (Neural Radiance Fields, NeRF) [25] 及其改进方法正在刷新神经渲染领域, 并且更多有前途的潜在应用仍在不断探索中。NeRF 使用简单的多层

*本文为 CVPR 2023 论文 Multi-Space Neural Radiance Fields 的中译版。

†任博是通讯作者。



(a) MipNeRF 360, SSIM=0.825 (b) 我们的方法, SSIM=0.881

图 1. (a) 尽管 Mip-NeRF 360 [2] 可以处理无边界的场景, 该方法仍然受到反射表面的困扰, 因为虚像违反了多视角一致性约束, 而这对于 NeRF 类方法来说至关重要。(b) 我们的方法可以帮助传统的 NeRF 类方法以较小的额外成本学习虚像。

感知器 (Multi-layer Perceptrons, MLP) 来表示场景的连续辐射场, 并沿着从相机到图像平面的射线方向上采样空间点, 基于采样点从 MLP 中获取密度值和辐照值, 从而使用体渲染方程来渲染得到新的视角。自其首次提出以来 [25], 研究者们已经付出了许多努力来增强该方法, 例如将其扩展到无界场景 [2, 50], 处理移动物体 [29, 30, 37], 或者使用真实环境中的图片进行重建 [6, 21, 35, 49]。

然而, 对于 state-of-the-art 的 NeRF 类方法来说, 渲染包含镜面的场景仍然是一个具有挑战性的任务。NeRF 类方法的一个主要假设是目标场景具有多视图一致性 [16, 20, 36]。当场景中存在镜子时, 如果使视点围绕场景进行 360 度移动, 镜子的前后视图之间就不存在一致性, 因为镜子表面及其反射的虚拟图像只能从一小范围的视角看到。因此, 通常需要手动标记反射表面, 以避免模型收敛到次优解 [12]。

在本文中，我们提出了一种新颖的基于多空间的 NeRF 方法，以便在无需任何手动标记的前提下，自动实现 360 度高清渲染含镜面场景。我们将场景欧氏空间视为由多个虚拟子空间的组合而非单一空间，其组合方式根据位置和视角方向而变化。我们展示了我们的方法通过这种多空间分解，成功处理了欧氏空间中严重违反多视图一致性的复杂反射和折射情况。此外，我们还展示了仅通过设计一个低成本的多空间模块并用其替换原方法中的输出层，就可以实现上述优势。因此，我们的多空间方法能够作为 NeRF 类主干的通用增强模块，赋予了大多数 NeRF 类方法模拟复杂反射和折射的能力，如 Fig. 1 所示。

现有的数据集缺少对于包含镜像物体的 360 度场景的关注，例如 RFFR [12] 只有前向场景，而 [42] 中的 Shiny 数据集视角变化较小，无法展现大角度范围内的一些视角依赖效果。因此，我们构建了一个全新的数据集，专门用于评估包含复杂反射和折射的 360 度高保真场景渲染。在这个数据集中包含 25 个合成场景和 7 个真实世界捕捉场景。每个合成场景包含 360 度围绕镜面或折射物体的 120 张图像，其中 100 张随机用于训练，10 张用于验证，10 张用于评估。每个真实世界场景是随意采集的，围绕着具有镜面和折射物体的场景，包含 62 到 118 张图像，采集到的图像按照 LLFF [24] 的方式进行组织。然后，我们使用了三个代表性基准模型进行对比实验，即 NeRF [25]，Mip-NeRF [1] 和 Mip-NeRF 360 [2]，通过比较这些基准模型在加入和没有加入我们的多空间模块的情况下的表现，展示了我们方法的优越性和兼容性。实验表明，我们的方法在具有反射和折射的场景上，在定量和定性比较上都具有显著提升。我们的主要贡献如下：

- 我们提出了一种多空间 NeRF 方法，可以自动处理 360 度高清场景渲染中的镜像物体，与现有代表性基准模型相比，在定量和定性上都取得了显著的提升。
- 我们设计了一个轻量级模块，可以以较小的计算开销为大多数 NeRF 类方法提供模拟反射和折射的能力。
- 我们构建了一个专用于评估包含复杂反射和折射的 360 度高清场景渲染的数据集，包括 25 个合成场景和 7 个真实采集场景。

2. 相关工作

基于坐标的新视角合成。 NeRF [25] 紧密地连系起了计算机视觉与计算机图形，展现了一种有前景的仅利用姿态图像来渲染照片级高质量场景的方法。这种方法的前景和适应性也为计算机视觉和计算机图形中的各种任务提供了便利，例如 3D 重建 [28,40]、3D 感知生成 [4,15,27]、3D 感知编辑 [39,47] 以及头像重建与编辑 [9,18,52]。此外，研究人员也在大力提升该方法的能力。Mip-NeRF [1] 通过使用集成位置编码表示三维圆锥体，提高了 NeRF 的抗锯齿能力。[14,23] 调整了这种方法，使其适用于 HDR 图像。[2,50] 将 NeRF 及其改进方法的应用范围扩展到无界场景。还有许多工作尝试使用显式表示或混合表示来加速训练和推理速度 [5,7,10,26,32,34,46]。

具有高光反射的光面材料对 NeRF 类方法有很大影响，[38] 受计算机图形中基于预计算的技术 [31] 的启发，提出了 Ref-NeRF 来表示和渲染视角相关的高光和反射，但无法处理镜面反射，因为虚拟图像无法被视为纹理。Guo 等人 [12] 提出将反射表面分解为透射分量和反射分量，这是与我们最相关的工作，但这种分解无法处理含镜像物体场景的 360 度视角，因为在越过一定视角范围之前虚像与实体之间无法被区分。

与我们类似的另一类工作是多重神经辐射场，但该类工作的目的与本文不同 [11,27,32,43,44]。[27] 使用物体级的神经辐射场进行 3D 感知生成和合成。[32,44] 使用多个小型 MLP 进行高效渲染。[11,43] 使用多个物体层级的神经辐射场进行 3D 场景分解和编辑。

常用数据集。 为了促进 NeRF 类方法在各种任务中的发展，研究人员引入或构建了许多不同的数据集。Mildenhall 等人 [25] 收集了一个分别包含八个对象的渲染图像集合，以及八个在真实环境下采集的前向场景，相机姿态和内参是由 COLMAP [33] 估计的。然而，这些场景缺乏在现实场景中非常普遍的反射和折射现象。Wizadwongsa 等人 [42] 提出了一个名为 Shiny 的数据集，其中包含八个更具挑战性的场景，以测试 NeRF 类方法在视角相关效果上的表现，但它们是以大致前向的方式拍摄的。Verbin 等人 [38] 创建了一个包含六个光泽物体的数据集，即 Shiny Blender，它们是在与 NeRF 中相似的条件渲染的，用于测试方法在

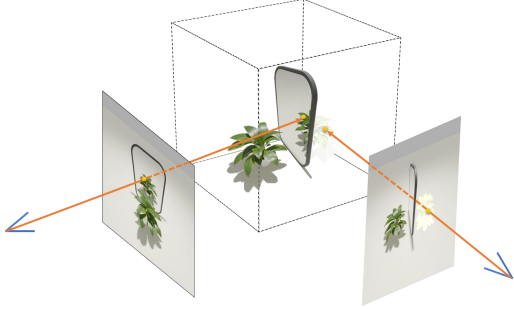


图 2. 镜面创造出的虚像仅在较小的视角范围下可见，这违背了多视图一致性。

建模更复杂材料方面的性能。对于无界场景, Barron 等人 [2] 构建了一个包含 5 个室外场景和 4 个室内场景的数据集; Zhang 等人 [50] 构建了 Tanks and Temples (T&T) 数据集 [19] 和光场数据集 [48]。Bemana1 等人 [3] 采集了一个包含折射物体的数据集, 其中包括四个场景, 其中相机的移动范围较大。Guo 等人 [12] 收集了六个前向场景, 其中包含有反射和半透明材料, 这是迄今为止与我们最相关的数据集, 但本文数据集与其相比要更具挑战性。DTU 数据集 [17] 和 BlendedMVS 数据集 [45] 通常用作 3D 重建评估的基准。

3. 方法

3.1. 预备知识: 神经辐射场

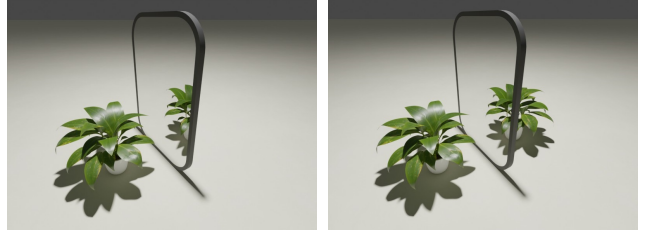
神经辐射场 (NeRF) [25] 将场景以连续体积场的形式编码到多层感知机 (MLP) 的权重中, 并采用传统体积渲染中的吸收模型来合成新的视角。训练过程仅需要一组稀疏的带位姿的图像, 并通过相机中心 $\mathbf{o} \in \mathbb{R}^3$ 和视角方向 $\mathbf{d} \in \mathbb{R}^3$ 来投射出穿过场景的光线 $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$, 这些光线可以通过训练数据中的内参和姿态来计算。给定这些光线后, NeRF 通过在欧氏空间中距离相机 t_i 处采样一组 3D 点 $\{\mathbf{p}_i = \mathbf{o} + t_i\mathbf{d}\}$, 并使用以下函数将这些点投影到更高维空间:

$$\gamma(\mathbf{p}) = [\sin(\mathbf{p}), \cos(\mathbf{p}), \dots, \sin(2^{L-1}\mathbf{p}), \cos(2^{L-1}\mathbf{p})] \quad (1)$$

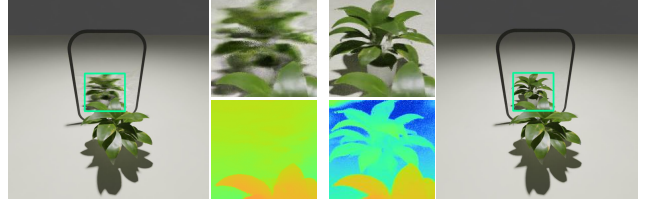
其中 L 是一个超参数, \mathbf{p} 是单个采样点。

给出投影特征向量 $\{\gamma(\mathbf{p}_i)\}$ 和视角方向 \mathbf{d} 后, 多层感知机输出体密度 $\{\sigma_i\}$ 和颜色 $\{c_i\}$ 。将其用于 Max 的 [22] 求积规则中, 从而计算出光线 \mathbf{r} 的颜色 $\mathbf{C}(\mathbf{r})$:

$$\hat{\mathbf{C}}(\mathbf{r}) = \sum_{i=1}^N T_i (1 - \exp(-\sigma_i \delta_i)) c_i \quad (2)$$



(a) 样例场景 A 的一个训练视图。(b) 样例场景 B 的一个训练视图。



(c) 样例场景 A 的一个渲染视图。(d) 样例场景 B 的一个渲染视图。

图 3. 第一行是两个场景中的训练视图示例。在场景 A 中, 镜子前面只有一株植物, 而在场景 B 中, 我们将另一株植物仔细地放置在虚像位置。第二行是用在样例场景上训练的普通 NeRF 渲染出的测试视图和深度。正如图所示, 当“虚像”满足多视角一致性时, NeRF 可以避免将反射图像视为纹理这一误区。

其中 $T_i = \exp(-\sum_{j=1}^{i-1} \sigma_j \delta_j)$, $\delta_i = t_i - t_{i-1}$ 。由于该方程是可微的, 模型参数可以直接通过均方误差 (MSE) 损失进行优化:

$$\mathcal{L} = \frac{1}{|\mathcal{R}|} \sum_{\mathbf{r} \in \mathcal{R}} \|\hat{\mathbf{C}}(\mathbf{r}) - \mathbf{C}(\mathbf{r})\|_2 \quad (3)$$

其中 \mathcal{R} 是光线的一个训练批次。此外, NeRF 还采用了分层采样策略, 在权重累积更高的地方采样更多的点。通过这些设计, NeRF 在大多数情况下都能达到最先进的视角合成效果。

3.2. 多空间神经辐射场

体积渲染方程和 MLP 的连续表示能力确实保证了 NeRF 类方法在新视角合成方面的成功, 但正如先前的研究所指出的 [12, 16, 20], 在训练过程中还存在一个必要的隐藏属性, 它有助于收敛, 这个属性就是多视角一致性。然而, 任何反射表面都很容易违反多视角一致性。如图 Fig. 2 所示, 当我们在镜子前面观察时, 我们可以看到反射的虚像, 就像镜子后面有一个物体, 但是当我们从侧面或背面观察时, 镜子后面实际上什么也没有。这意味着训练批次中将会出现完全冲突的情况, 违反了 MLP 的拟合过程。

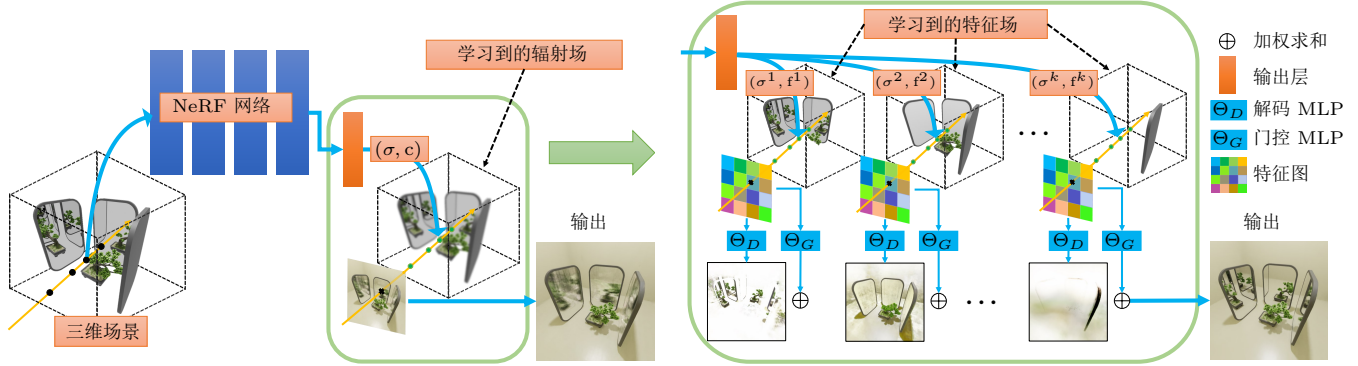


图 4. 我们的多空间模块只修改了网络的输出和体积渲染部分。原始的 NeRF 计算一组体密度 σ 和颜色 c 以得到累积颜色。我们的输出层生成多组体密度 $\{\sigma^k\}$ 和特征 $\{f^k\}$ ，对应于多个并行的特征场。然后，我们使用体积渲染来得到多个特征图。接下来，我们使用两个简单的 MLP，即解码 MLP 和门控 MLP，从这些特征图中解码 RGB 图和像素权重。

为了从实验角度证明多视角一致性的重要性以及其对传统 NeRF 网络结构的影响，我们使用开源软件 Blender [8] 创建了两个 360 度样例场景，每个场景包含 100 个训练图像和 10 个测试图像，其中训练视角示例如图 Fig. 3a 和 Fig. 3b 所示。这两个场景之间唯一的区别在于我们在后一个场景中在镜子后面放置了一个以镜像方式摆放的真实物体，而在前一个场景中则没有这样做。我们在相同的设置下分别在这些玩具场景上训练普通的 NeRF，并从测试集中渲染一些视角。Fig. 3c 和 Fig. 3d 清楚地显示了在一些视角中消失的虚像（即违反了多视角一致性）导致了模型在反射相关区域产生了次优的结果，并在渲染中产生了模糊。有趣的是，传统的 NeRF 在这个过程中仍在努力满足多视角一致性假设。从 Fig. 3c 中的深度图中，我们可以很容易地得出结论，传统的 NeRF 将所观察的虚拟图像视为反射表面上的“纹理”，在其基本原则假设和训练数据冲突之间实现了折衷，尽管这种折衷导致了真实场景的错误理解和糟糕的渲染结果。

与传统的 NeRF 形成对比，我们受到物理学和计算机图形学中常见的观点的启发，即反射光可以被视为从其镜像对称方向“直接发射”出来，从可能存在于镜子内的“虚拟空间内的虚拟源”观察到，我们的新型多空间 NeRF 方法基于以下假设：

假设 1 在存在反射和折射的情况下，真实场景的欧氏空间可以分解为多个虚拟子空间。每个子空间均满足多视角一致性。

进一步地，子空间的组合权重可以根据空间位置

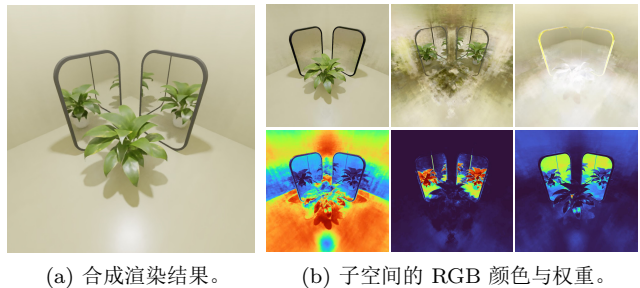


图 5. 在 Sec. 5.2 中，我们可视化了一个新视角，以及一些从 MS-NeRF_B 的子空间中解码得到的图像和对应的权重。结果表明我们的方法成功地将虚像分解到了特定的子空间中。

和视角方向的变化而变化。因此，所有子空间都会动态地对最终的渲染结果产生贡献。通过这种方式，在存在反射表面的真实欧氏空间中违反多视角一致性的问题可以通过将虚拟图像放置在某些子空间中解决，这些子空间只能从特定视角看到，如 Fig. 5 所示。

3.3. 多空间模块

一个朴素的多空间 NeRF 网络的实现方式是使用多个小型的并行 MLP 构建网络，其中每个 MLP 表示一个子空间的信息，然而，这样会大大增加参数量，并且已被证明很难收敛 [32]。此外，我们在 Sec. 3.2 中的实验证明了 NeRF 当前的网络结构具有理解我们的三维场景的潜力。因此，我们提出了一个紧凑的多空间模块（MS 模块），它利用了最初设计用于节省内存的神经特征场方案 [27]，只需很小的计算开销就可以充分地

数据集	来源	应用	类型	视角	特性	数量
Realistic Synthetic 360°	[25]	视角合成	S	360 度	非朗伯	8
Real Forward-Facing	[24, 25]	视角合成	R	前向	非朗伯	8
Shiny	[42]	视角合成	R	前向	高光, 折射	8
Tanks and Temples(T&T)	[19]	视角合成	R	360 度	无界场景	4
Mip-NeRF 360	[2]	视角合成	R	360 度	无界场景	9
EikonalFields	[3]	视角合成	R	大视角	折射	4
RFFR	[12]	视角合成	R	前向	反射, 半透明	6
DTU	[17]	三维重建	R	360 度	非朗伯	15*
BlendedMVS	[45]	三维重建	S	360 度	非朗伯	7*
Shiny Blender	[38]	视角合成	S	360 度	高光	6
Ref-NeRF Real captured scenes	[13, 38]	视角合成	R	360 度	高光	3

表 1. 常用于 NeRF 类方法的数据集的特性。“S” 和“R” 分别表示合成的和真实拍摄的数据。我们用方法的名称来表示那些没有具体命名的数据集。“*” 表示常用于 NeRF 类方法的场景数量，原始数据集中的场景数量多于这个数目，但不在我们考虑当中。

来说，MS 模块将替换 NeRF 主干网络的原始输出层。下面我们将描述模块的详细架构。

如 Fig. 4 所示，我们的 MS 模块只修改了普通 NeRF 的输出部分。普通 NeRF 为穿过场景的每个位置计算一对密度 σ_i 和辐射度 c_i ，并使用 Eq. (2) 进行体积渲染以得到累积颜色。与之相比，我们的多头输出层使用神经特征场 [27] 来替换神经辐射场。具体而言，修改后的输出层为每个射线上的每个位置给出了 K 个密度 $\{\sigma_i^k\}$ 和 K 个特征 $\{f_i^k\}$ ，每个特征是 d 维的，其中 K 和 d 是超参数，分别代表总子空间数和神经特征场的特征维度。

然后我们在每个子空间中沿着射线对特征进行积分，得到 K 个特征图，这些特征图编码了从特定视点观察下每个子空间的颜色信息和可见性。由于所有像素的计算方式相同，为简单起见，我们将每个像素表示为 $\{\mathcal{F}^k\}$ ，并在像素级别描述操作。特征图中的每个像素 $\{\mathcal{F}^k\}$ 的计算方式如下：

$$\hat{\mathcal{F}}^k(\mathbf{r}) = \sum_{i=1}^N T_i^k (1 - \exp(-\sigma_i^k \delta_i)) f_i^k, \quad (4)$$

其中上标 k 代表射线所经过的子空间编号。第 k 个体密度 σ_i^k 和特征 f_i^k 对应第 k 个子空间。 $T_i^k = \exp(-\sum_{j=1}^{i-1} \sigma_j^k \delta_j)$ 与 $\delta_i = t_i - t_{i-1}$ 以与 Eq. (2) 中相似的方式计算。

接下来， $\{\mathcal{F}^k\}$ 通过两个小型 MLP 进行解码，每个 MLP 只有一层隐藏层。第一个是解码 MLP (Decoder

MLP)，它将 $\{\mathcal{F}^k\}$ 作为输入，并输出 RGB 向量。第二个是门控 MLP (Gate MLP)，它将 $\{\mathcal{F}^k\}$ 作为输入，并输出控制某个子空间可见性的权重。具体来说，我们使用以下公式：

$$\{\mathcal{F}^k\} \xrightarrow{\Theta_D} \{C^k\}, \{\mathcal{F}^k\} \xrightarrow{\Theta_G} \{w^k\}, \quad (5)$$

其中 Θ_D 代表解码 MLP， Θ_G 代表门控 MLP。最后，MS 模块将 softmax 函数应用于 $\{w^k\}$ 上，作为每个子空间对整体颜色的贡献，从而形成最终的渲染结果：

$$\hat{C}(\mathbf{r}) = \frac{1}{\sum_{i=1}^K \exp(w^i)} \sum_{k=1}^K \exp(w^k) C^k. \quad (6)$$

与传统的 NeRF 方法相比，Eq. (6) 不需要额外的损失项。因此，上述轻量级的 MS 模块可以作为增强模块添加到传统的 NeRF 类主干网络上，我们将在 Sec. 5.2 中展示我们的方法取得的显著改进效果。

4. 数据集

4.1. 现有数据集

我们简要回顾了常用的或与我们任务最相关的数据集，并在 Tab. 1 中列出了它们的特性。可以看到，现有数据集中缺乏一个具有复杂光线路径的 360 度场景数据集，例如，放在镜子前的一杯水。

4.2. 本文提出的数据集

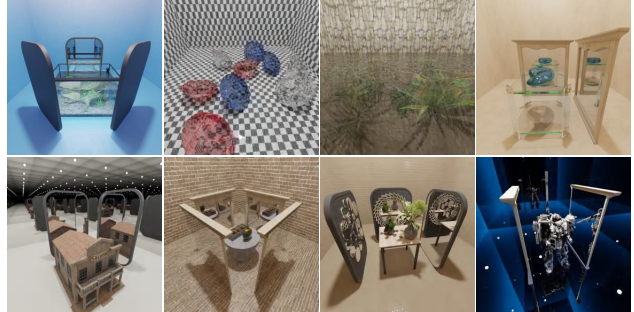
如 Sec. 4.1 中总结的那样，现有数据集中缺乏一个能够促进相关研究的包含复杂反射和折射的 360 度场景的数据集。因此，我们收集了一个由 25 个合成场景和 7 个真实拍摄场景组成的 360 度数据集。对于本文数据集中的合成部分（如 Fig. 6a 所示），我们使用开源软件 Blender [8]，并使用共享三维模型的社区 BlenderKit 中的模型设计我们的场景。由于我们的数据集包含完整的场景而不是单个对象，我们将摄像机的高度固定在俯视场景中心的位置，并围绕环形路径移动摄像机来渲染整个场景。对于每个场景，我们均匀采样了环形路径上的 120 个点，并随机选择了 100 张图片作为训练集，10 张作为验证集，10 张作为测试集。构建的数据集包含各种各样的场景，既有反射物体也有折射物体。我们的数据集有各种复杂的光线路径，由场景中镜子的数量和布局控制，镜子的数量从 1 个到几十个小块不等。需要注意的是，即使我们数据集中只有一个镜子的场景，也比 RFFR [12] 更具挑战性，因为我们的摄像机从镜子的前面移动到后面。此外，我们还构建了带有镜面墙壁的房间，这些房间实际上可以视为无边界的场景，我们在房间中央额外添加了镜子从而创建了无边界的虚像。我们进一步构建了一些包含反射和折射组合的具有挑战性的场景。

我们还包含了 7 个真实场景，这些场景具有复杂的光线条件，如 Fig. 6b 所示。我们在这些场景中使用了两个镜子，一个表面光滑的玻璃球，一个类似于钻石的表面的玻璃球，还有一些玩具和一些书籍。数据集中的图片是随机地以 360 度视点拍摄的。

5. 实验

5.1. 超参数设置与实验基准

我们根据不同的数据集、不同的基准模型以及不同规模的 MS 模块进行了三组实验。由于我们的模块相当简单，我们可以通过三个超参数来调整我们的模块，这三个超参数分别是子空间数量 K 、输出特征维度 d 以及解码 MLP 和门控 MLP 的隐藏层维度 h 。为了公平比较，我们在所有实验中遵循 [1, 2, 12, 25, 38] 中的大多数默认设置，唯一的区别是在所有场景的所有实验中，每批次使用 1024 条光线，并进行了 200,000 次迭代训练。实验细节如下。



(a) 合成数据集的一部分。



(b) 真实采集数据集的一部分。

图 6. 本文的数据集中的一些样例场景（更多详见补充材料）。本文的数据集展示了各种反射和折射的情况，可以作为验证生成具有复杂光线路径的新视角的能力的一个基准。

我们选择了三种代表性的基于 NeRF 的模型作为我们的基准，并将我们的模块与它们集成在一起。对于基于 NeRF [25] 和 Mip-NeRF [1] 的实验，我们构建了 MS-NeRF_S 和 MS-Mip-NeRF_S，其超参数为 $\{K = 6, d = 24, h = 24\}$ ，类似地，MS-NeRF_M 和 MS-Mip-NeRF_M 的超参数为 $\{K = 6, d = 48, h = 48\}$ ，以及 MS-NeRF_B 和 MS-Mip-NeRF_B 的超参数为 $\{K = 8, d = 64, h = 64\}$ 。对于基于 Mip-NeRF 360 [2] 的实验，我们构建了 MS-Mip-NeRF 360，其超参数为 $\{K = 8, d = 32, h = 64\}$ 。此外，我们还与 Ref-NeRF [38] 进行了比较，因为它基于 Mip-NeRF，并具有出色的建模光泽材质的能力。

我们还将我们的方法与 NeRFReN 在 RFFR 数据集 [12] 上进行比较。NeRFReN 是一个基于标准 NeRF 的双分支网络，专门设计用于处理前向场景中的类镜面表面。我们基于 NeRF 以超参数 $\{K = 2, d = 128, h = 128\}$ 构建了我们的方法的一个小版本，称为 MS-NeRF_T。在这里，我们使用两个子空间，因为 NeRFReN 试图将反射表面分解为两个部分，而我们想展示我们的空间分解更有效。为了公平比较，我们使用官方代码在 RFFR 数据集上重新训练了 NeRFReN，并按照提供的参数进行设置，唯一的区别是我们将使用的 mask 数量设置为 0，因为我们的方法不需要额外的 mask。

所有的训练细节可以在补充材料中找到。我们

	PSNR↑	SSIM↑	LPIPS↓	# 参数量
NeRF	30.82	0.865	0.209	1.159M
MS-NeRF _S	32.39	0.872	0.201	1.201M
MS-NeRF _M	32.61	0.875	0.195	1.245M
MS-NeRF _B	32.77	0.876	0.195	1.311M
Mip-NeRF	31.42	0.874	0.215	0.613M
Ref-NeRF	32.37	0.882	0.180	0.713M
MS-Mip-NeRF _S	33.63	0.886	0.195	0.634M
MS-Mip-NeRF _M	33.80	0.887	0.193	0.656M
MS-Mip-NeRF _B	33.90	0.888	0.191	0.689M
Mip-NeRF 360	31.58	0.895	0.145	9.007M
MS-Mip-NeRF 360	35.04	0.906	0.130	9.052M

(a) 在本文的数据集的合成场景部分进行的比较。

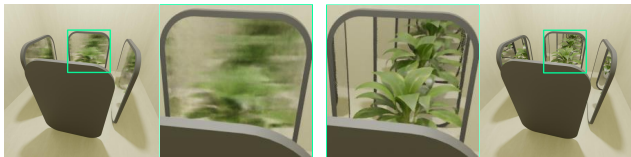
	PSNR↑	SSIM↑	LPIPS↓	# 参数量
Mip-NeRF 360	26.70	0.889	0.113	9.007M
MS-Mip-NeRF 360	28.14	0.891	0.119	9.052M

(b) 在本文的数据集的真实采集部分上进行的比较。

	PSNR↑	SSIM↑	LPIPS↓	# 参数量
NeRFReN*	35.26	0.940	0.081	1.264M
MS-NeRF _T	35.93	0.948	0.066	1.295M

(c) 在 RFFR 数据集上进行比较。“*”代表我们使用官方的代码和配置进行重新训练，但没有使用用于标记反射表面的掩码。

表 2. 和现有方法的定量比较。



(a) Mip-NeRF 360

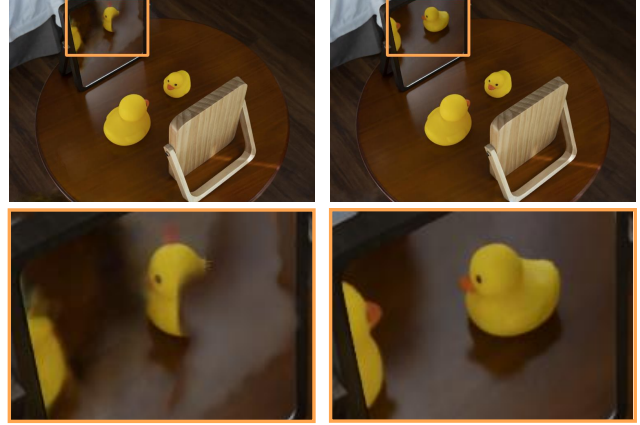
(b) MS-Mip-NeRF 360

图 7. Mip-NeRF 360 与 MS-Mip-NeRF 360 之间的可视化比较。我们的模块可以将 Mip-NeRF 360 的应用范围扩展到含有无边界虚像的场景中。

使用三种常用指标来量化展示我们的结果：PSNR、SSIM [41] 和 LPIPS [51]。

5.2. 对比实验

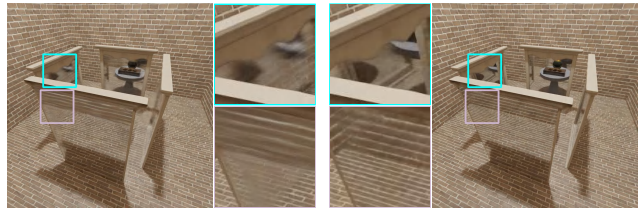
定量对比。如 Tab. 2a所示，我们的模块可以集成到大多数 NeRF 类模型中，并且只需要引入很小的额外成



(a) Mip-NeRF 360

(b) MS-Mip-NeRF 360

图 8. Mip-NeRF 360 与 MS-Mip-NeRF 360 在本文数据集的真实采集部分上的可视化比较。我们的方法足够强大，可以重现现实世界中的虚像。



(a) Ref-NeRF

(b) MS-Mip-NeRF_B

图 9. MS-Mip-NeRF_B 与 Ref-NeRF 之间的可视化比较。我们的方法在反射表面上的表现显著地优于 Ref-NeRF。

本就能大幅提高性能。特别是在基于 Mip-NeRF 360 的实验中，我们的模块在 PSNR 上表现出 3.46dB 的改进，而只增加了 0.5% 的额外参数。此外，我们的基于 Mip-NeRF 的模型也远远优于 Ref-NeRF [38]。Ref-NeRF 是一种基于 Mip-NeRF 的改进模型，具有模拟光泽材质的优秀能力。我们还在 Tab. 2b中展示了我们在本文数据集的真实采集部分与最先进的 Mip-NeRF 360 结果的对比。结果显示，我们的方法有着显著的改进。如 Tab. 2c所示，我们的方法在没有手动标记镜面掩码的情况下，在 RFFR 数据集上表现出更好的结果，该数据集包含前向场景中的反射表面。以上实验证明了我们方法的优越性和兼容性。

定性比较与讨论分析。除了定量比较之外，我们还总结了我们的模块的优势，并通过定性或定量地与相应的基准方法进行比较来验证这些优势。

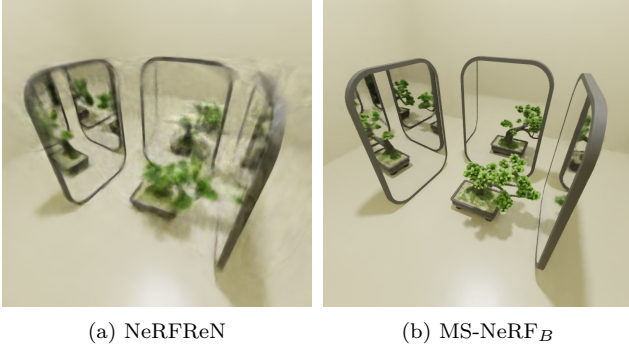


图 10. (a) 即使使用精确标注的掩码进行训练, NeRFReN 甚至无法渲染出 360 度含镜面场景中的实景部分。(b) 我们的方法不需要额外的人工标注, 也能够渲染得到高清图像。

在 Fig. 1、Fig. 7 和 Fig. 8 中, 我们与最先进的方 法 (即 Mip-NeRF 360) 进行了定性比较。我们的方法 在合成和真实场景中均能呈现高清的虚像, 包括有界 和无界的场景。

在 Fig. 9 中, 我们与 Ref-NeRF [38] 进行了定性比 较。Ref-NeRF 使用传统的 NeRF 主干, 并将虚像视 为纹理。由于 Ref-NeRF 也是基于 Mip-NeRF 的, 我 们将我们的基于 Mip-NeRF 的变体与 Ref-NeRF 进行 了比较。在比较中, 我们使用相同的基准和相近的参 数 (具体来说, 我们参数量为 0.689M, Ref-NeRF 的为 0.713M)。定性结果显示了我们在渲染反射表面方面的 显著改进。

我们还与 NeRFReN 模型进行了比较, 该模型在 训练过程中需要准确标记的反射区域的掩码, 并且仅 能处理前向观察的反射表面。在这个比较中, 我们使用 我们的合成数据集来训练他们的模型, 并额外提供了 准确的反射面标注。结果在 Fig. 10 中显示, 他们的模 型无法实现 360 度的高清渲染, 而我们的方法则能够 成功地实现。

5.3. 消融研究

在本节中, 我们评估了我们模块的设计, 并探讨了 子空间数量与虚像数量之间的关系。

神经特征场的消融实验。我们实现了一个模块, 简单地 输出 K 个标量 $\{\sigma_i^k\}$ 和 K 个三维 RGB 向量 $\{c_i^k\}$ 。然 后, 我们使用与 NeRF 相同的积分方程来获取每个子 空间的像素颜色, 并在所有子空间之间进行平均以得 到最终的像素颜色。我们将这个设计集成到原始 NeRF

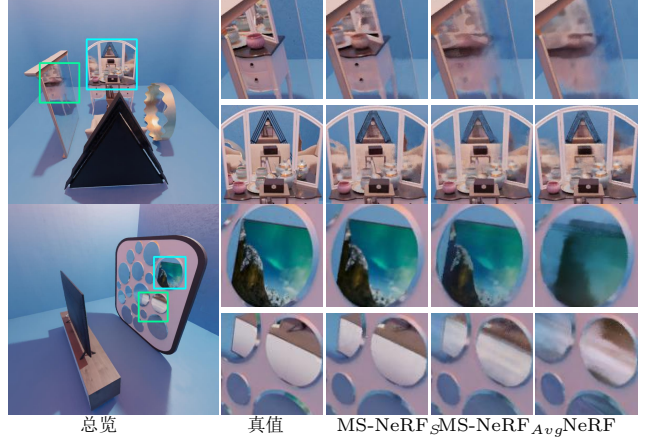


图 11. 将 NeRF, MS-NeRF_{Avg} 与 MS-NeRF_S 进行比较。

	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	# 参数量
NeRF	30.82	0.866	0.209	1.159M
MS-NeRF _{Avg}	31.45	0.868	0.205	1.166M
MS-NeRF _S	32.39	0.872	0.201	1.201M

表 3. 在我们的模块结构方面的消融研究。

中, 标记为 MS-NeRF_{Avg}, 其中我们设置 $K = 6$, 结 果在 Tab. 3 中展示。我们还在 Fig. 11 中展示了一些视 觉结果, 表明简单的多空间辐射场假设可以帮助模型 部分克服反射违反多视图一致性的现象, 但由于缺乏 高效的多空间合成策略, 也会引入过度平滑的问题。

子空间数量的消融实验。在我们的欧氏空间中, 可以通 过镜子的数量和布局来控制虚拟子空间的数量。例如, 当两个镜子彼此相对放置时, 可能会出现无限递归的 虚像空间, 但当两个镜子背靠背放置时, 每个镜子中只 会有一个虚像。为了给我们模块的设计提供指导, 我们 从我们合成的数据集中选择了两个包含不同镜子布局 的场景, 并训练了不同子空间数量和不同特征维度的 NeRF 改进版本。

我们基于 NeRF 构建了多个变种, 其中输出特 征维度分别为 $d \in \{24, 48, 64\}$, 子空间数量分别为 $K \in \{2, 4, 6, \dots, 16\}$, 然后我们在这两个场景上训练模 型, 并使用 PSNR 展示结果。通过分析 Fig. 12 中展示 的结果, 子空间数量并不需要与实际虚像空间的数量 匹配, 且 6 个子空间足以保证多空间辐射场能够稳定 地学习。此外, 特征维度 $d = 24$ 的特征场已经包含了 足够的组合信息, 但为了获得更稳定的性能, $d = 48$ 效

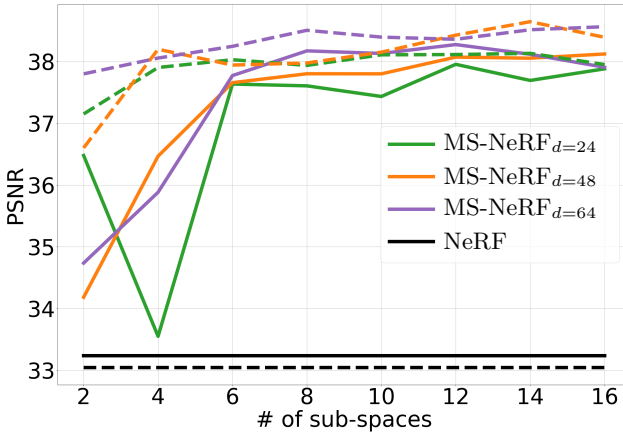


图 12. 我们使用 PSNR 来对场景 01 和场景 02 上的实验进行量化比较，在图表中分别用实线和虚线表示。

果更好。

6. 总结

本文解决了 NeRF 类的方法中渲染反射表面这一长期存在的难题。我们引入了一种多空间 NeRF 方法，将欧氏空间分解成多个虚拟子空间。我们提出的 MS-NeRF 方法相比于传统 NeRF 类方法取得了显著更好的结果。此外，MS 模块的轻量级设计使得我们的方法可以作为传统 NeRF 主干网络的增强模块。我们还构建了一个新的数据集用于评估类似任务，希望能够对未来的研究有所帮助。

致谢：本工作由中国自然科学基金（NO. 62132012）资助。

参考文献

[1] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for

anti-aliasing neural radiance fields. In ICCV, pages 5855–5864, 2021. [2](#), [6](#)

[2] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In IEEE CVPR, pages 5470–5479, 2022. [1](#), [2](#), [3](#), [5](#), [6](#)

[3] Mojtaba Bermana, Karol Myszkowski, Jeppe Revall Frisvad, Hans-Peter Seidel, and Tobias Ritschel. Eikonal fields for refractive novel-view synthesis. In ACM SIGGRAPH 2022 Conference Proceedings, pages 1–9, 2022. [3](#), [5](#)

[4] Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. Efficient geometry-aware 3d generative adversarial networks. In IEEE CVPR, pages 16123–16133, 2022. [2](#)

[5] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In European Conference on Computer Vision (ECCV), 2022. [2](#)

[6] Xingyu Chen, Qi Zhang, Xiaoyu Li, Yue Chen, Ying Feng, Xuan Wang, and Jue Wang. Hallucinated neural radiance fields in the wild. In IEEE CVPR, pages 12943–12952, 2022. [1](#)

[7] Zhiqin Chen, Thomas Funkhouser, Peter Hedman, and Andrea Tagliasacchi. Mobilenerf: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures. arXiv:2208.00277, 2022. [2](#)

[8] Blender Online Community. Blender - a 3D modelling and rendering package. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2022. [4](#), [6](#)

[9] Yao Feng, Jinlong Yang, Marc Pollefeys, Michael J Black, and Timo Bolkart. Capturing and animation of body and clothing from monocular video. arXiv:2210.01868, 2022. [2](#)

[10] Stephan J Garbin, Marek Kowalski, Matthew Johnson, Jamie Shotton, and Julien Valentin. Fastnerf: High-fidelity neural rendering at 200fps. In ICCV, pages 14346–14355, 2021. [2](#)

[11] Michelle Guo, Alireza Fathi, Jiajun Wu, and Thomas Funkhouser. Object-centric neural scene rendering. arXiv:2012.08503, 2020. [2](#)

- [12] Yuan-Chen Guo, Di Kang, Linchao Bao, Yu He, and Song-Hai Zhang. Nerfren: Neural radiance fields with reflections. In IEEE CVPR, pages 18409–18418, 2022. [1](#), [2](#), [3](#), [5](#), [6](#)
- [13] Peter Hedman, Pratul P Srinivasan, Ben Mildenhall, Jonathan T Barron, and Paul Debevec. Baking neural radiance fields for real-time view synthesis. In ICCV, pages 5875–5884, 2021. [5](#)
- [14] Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang. Hdr-nerf: High dynamic range neural radiance fields. In IEEE CVPR, pages 18398–18408, 2022. [2](#)
- [15] Ajay Jain, Ben Mildenhall, Jonathan T Barron, Pieter Abbeel, and Ben Poole. Zero-shot text-guided object generation with dream fields. In IEEE CVPR, pages 867–876, 2022. [2](#)
- [16] Nishant Jain, Suryansh Kumar, and Luc Van Gool. Robustifying the multi-scale representation of neural radiance fields. arXiv:2210.04233, 2022. [1](#), [3](#)
- [17] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engin Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In IEEE CVPR, pages 406–413, 2014. [3](#), [5](#)
- [18] Boyi Jiang, Yang Hong, Hujun Bao, and Juyong Zhang. Selfrecon: Self reconstruction your digital avatar from monocular video. In IEEE CVPR, pages 5605–5615, 2022. [2](#)
- [19] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. ACM TOG, 36(4):1–13, 2017. [3](#), [5](#)
- [20] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. Barf: Bundle-adjusting neural radiance fields. In ICCV, pages 5741–5751, 2021. [1](#), [3](#)
- [21] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In IEEE CVPR, pages 7210–7219, 2021. [1](#)
- [22] Nelson Max. Optical models for direct volume rendering. IEEE TVCG, 1(2):99–108, 1995. [3](#)
- [23] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In IEEE CVPR, pages 16190–16199, 2022. [2](#)
- [24] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. ACM TOG, 38(4):1–14, 2019. [2](#), [5](#)
- [25] Ben Mildenhall, Pratul P Srinivasan, Matthew Tan-cik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1):99–106, 2021. [1](#), [2](#), [3](#), [5](#), [6](#)
- [26] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. ACM TOG, 41(4), jul 2022. [2](#)
- [27] Michael Niemeyer and Andreas Geiger. Giraffe: Representing scenes as compositional generative neural feature fields. In IEEE CVPR, pages 11453–11464, 2021. [2](#), [4](#), [5](#)
- [28] Michael Oechsle, Songyou Peng, and Andreas Geiger. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In ICCV, pages 5589–5599, 2021. [2](#)
- [29] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In ICCV, pages 5865–5874, 2021. [1](#)
- [30] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In IEEE CVPR, pages 10318–10327, 2021. [1](#)
- [31] Ravi Ramamoorthi et al. Precomputation-based rendering. Foundations and Trends® in Computer Graphics and Vision, 3(4):281–369, 2009. [2](#)
- [32] Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In ICCV, pages 14335–14345, 2021. [2](#), [4](#)
- [33] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In IEEE CVPR, pages 4104–4113, 2016. [2](#)
- [34] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for ra-

- diance fields reconstruction. In IEEE CVPR, pages 5459–5469, 2022. 2
- [35] Jiaming Sun, Xi Chen, Qianqian Wang, Zhengqi Li, Hadar Averbuch-Elor, Xiaowei Zhou, and Noah Snavely. Neural 3d reconstruction in the wild. In ACM SIGGRAPH 2022 Conference Proceedings, pages 1–9, 2022. 1
- [36] Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, W Yifan, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, et al. Advances in neural rendering. *Computer Graphics Forum*, 41(2):703–735, 2022. 1
- [37] Edgar Tretschk, Ayush Tewari, Vladislav Golyanik, Michael Zollhöfer, Christoph Lassner, and Christian Theobalt. Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. In ICCV, pages 12959–12970, 2021. 1
- [38] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T. Barron, and Pratul P. Srinivasan. Ref-NeRF: Structured view-dependent appearance for neural radiance fields. *CVPR*, 2022. 2, 5, 6, 7, 8
- [39] Can Wang, Menglei Chai, Mingming He, Dongdong Chen, and Jing Liao. Clip-nerf: Text-and-image driven manipulation of neural radiance fields. In IEEE CVPR, pages 3835–3844, 2022. 2
- [40] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *Advances in Neural Information Processing Systems*, 34:27171–27183, 2021. 2
- [41] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 7
- [42] Suttisak Wizadwongsa, Pakkapon Phongthawee, Jiraphon Yenphraphai, and Supasorn Suwajanakorn. Nex: Real-time view synthesis with neural basis expansion. In IEEE CVPR, pages 8534–8543, 2021. 2, 5
- [43] Bangbang Yang, Yinda Zhang, Yinghao Xu, Yijin Li, Han Zhou, Hujun Bao, Guofeng Zhang, and Zhaopeng Cui. Learning object-compositional neural radiance field for editable scene rendering. In ICCV, pages 13779–13788, 2021. 2
- [44] Guo-Wei Yang, Wen-Yang Zhou, Hao-Yang Peng, Dun Liang, Tai-Jiang Mu, and Shi-Min Hu. Recursive-nerf: An efficient and dynamically growing nerf. *IEEE TVCG*, 2022. 2
- [45] Yao Yao, Zixin Luo, Shiwei Li, Jingyang Zhang, Yufan Ren, Lei Zhou, Tian Fang, and Long Quan. Blended-mvs: A large-scale dataset for generalized multi-view stereo networks. In IEEE CVPR, pages 1790–1799, 2020. 3, 5
- [46] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. Plenotrees for real-time rendering of neural radiance fields. In ICCV, pages 5752–5761, 2021. 2
- [47] Yu-Jie Yuan, Yang-Tian Sun, Yu-Kun Lai, Yuewen Ma, Rongfei Jia, and Lin Gao. Nerf-editing: geometry editing of neural radiance fields. In IEEE CVPR, pages 18353–18364, 2022. 2
- [48] Kaan Yücer, Alexander Sorkine-Hornung, Oliver Wang, and Olga Sorkine-Hornung. Efficient 3d object segmentation from densely sampled light fields with applications to 3d reconstruction. *ACM TOG*, 35(3):1–15, 2016. 3
- [49] Jason Zhang, Gengshan Yang, Shubham Tulsiani, and Deva Ramanan. Ners: Neural reflectance surfaces for sparse-view 3d reconstruction in the wild. *Advances in Neural Information Processing Systems*, 34:29835–29847, 2021. 1
- [50] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv:2010.07492*, 2020. 1, 2, 3
- [51] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In IEEE CVPR, pages 586–595, 2018. 7
- [52] Yufeng Zheng, Victoria Fernández Abrevaya, Marcel C Böhler, Xu Chen, Michael J Black, and Otmar Hilliges. Im avatar: Implicit morphable head avatars from videos. In IEEE CVPR, pages 13545–13555, 2022. 2