

MS-NeRF: 多空间神经辐射场

尹泽鑫, 焦鹏屹, 邱佳雄, 程明明, 任博

摘要—现有的神经辐射场 (Neural Radiance Fields, NeRF) 方法在面对反射物体时常常会出现渲染模糊或扭曲的问题。为了解决这一问题, 我们提出了一种多空间神经辐射场 (Multi-Space Neural Radiance Fields, MS-NeRF), 该方法通过在并行子空间中引入一组特征场来表示场景, 而非使用单一的辐射场, 从而可以帮助神经网络更好地理解 and 处理反射和折射物体。我们的多空间方案是一种对现有 NeRF 方法的增强, 仅需极少的额外计算开销即可在训练和推理过程中生成额外的空间输出。我们分别为具有代表性的基于多层感知机 (Multilayer Perceptron, MLP) 和基于网格的 NeRF 方法设计了不同的多空间模块。实验结果显示, 在反射区域, 仅引入约 0.5% 的额外参数, 我们的方法可将 Mip-NeRF 360 的 PSNR 提升 4.15 dB; 同时, 仅引入约 0.046% 的额外参数, 也可将 TensoRF 的 PSNR 提升 2.71 dB, 且不会降低其他区域的渲染质量。此外, 我们构建了一个全新的数据集, 涵盖 33 个合成场景和 7 个真实拍摄场景, 这些场景均具有复杂的反射和折射特性。为了全面评估基于 NeRF 方法在复杂镜面路径下的鲁棒性, 我们设计了具有挑战性的摄像机路径。大量实验表明, 在渲染包含镜面物体和复杂光线路径的高质量场景方面, 我们的方法显著优于现有的单空间 NeRF 方法。源代码、数据集与结果可以通过我们的项目页面获取: <https://zx-yin.github.io/msnerf/>。

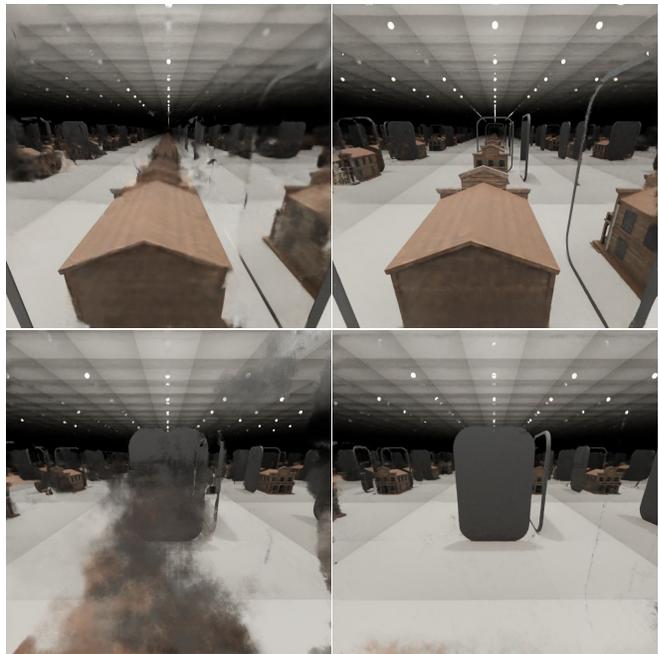
关键词—神经辐射场, 多空间 NeRF, 数据集

1 引言

NEURAL Radiance Fields (NeRF) [2] 及其变体, 正在不断革新神经渲染与三维重建领域, 并且其在更广泛应用中的潜力仍在持续探索之中。NeRF 将场景表示为由简单多层感知器 (MLP) 存储的连续辐射场, 并通过积分密度和辐射度来呈现新颖的视图, 这些密度和辐射度是通过沿从相机到图像平面的射线采样的点由 MLP 映射得到。自 NeRF 首次提出以来 [2], 已有大量研究致力于改进这一方法, 例如: 扩展至无界场景 [3], [4]、处理动态物体 [5]–[7], 或实现对真实世界图像的三维重建 [8]–[11]。

然而, 对于当前最先进的 NeRF 类方法来说, 渲染包含镜面的场景仍然是一项具有挑战性的任务。NeRF 方法的主要假设之一是目标场景的多视图一致性属性 [12]–[15]。当场景中存在镜面时, 若允许视角在场景周围进行 360 度移动, 则镜子正面与背面的视图之间不会保持一致性, 因为镜面本身及其反射的虚像仅在一个很小的视角范围内可见。因此, 通常需要手动标注反射表面, 以避免网络陷入次优的收敛状态 [16]。

在本文中, 我们提出了一种新颖的多空间 NeRF 方法, 可以在无需任何手动标注的情况下, 自动处理镜面物体, 实现场景的 360 度高保真渲染。我们不再将欧几里得场景空间视为单一空间, 而是将其视为由多个虚拟子空间组成, 这些子空间的组成会根据位置和视角方向而变化。



(a) Mip-NeRF 360

(b) MS-Mip-NeRF 360

图 1: 这些是来自新颖的“穿镜路径” (mirror-passing-through path) 的测试视图。第一行位于镜子前方, 而最后一行位于镜子后方。

我们证明了即使在欧氏实空间中, 多视角一致性被严重违反的情况下, 采用多空间分解的方法仍能够成功处理复杂的反射和折射。此外, 我们还证明了, 通过设计一个低成本的多空间模块并用其替换原始输出层, 即可实现上述优势。因此, 我们的多空间方法可以作为对 NeRF 类方法的通用增强方法, 使大多数基于 NeRF 的方法都具有建模复杂反射和折射的能力, 如图 Fig. 1 所示。

- Z.X. Yin, P.Y. Jiao, M.M. Cheng and B. Ren are with the TBI Center & VCIP, Nankai University.
- J. Qiu is with Horizon Robotics, Beijing, China.
- Bo Ren is the corresponding author (rb@nankai.edu.cn).
- A preliminary version of this work appeared at CVPR [1].

现有数据集对包含镜面物体场景的 360 度渲染关注不足, 例如 RFFR [16] 数据集仅包含朝前拍摄的场景, 而 Shiny 数据集 [17] 的视角变化较小, 无法在大角度范围内展现明显的视角依赖效果。因此, 我们构建了一个专门用于评估在包含复杂反射与折射的场景中进行 360 度高保真渲染表现的新数据集。在该数据集中, 我们收集了 33 个合成场景和 7 个真实拍摄的场景。每个合成场景包含 120 张围绕反射或折射物体 360 度环绕拍摄的图像, 其中随机分配 100 张用于训练, 10 张用于验证, 10 张用于评估。此外, 我们为其中 10 个合成场景设计了更具挑战性的路径, 以测试 NeRF 类方法的鲁棒性, 包括 360 度螺旋路径 (其中摄像机从赤道逐渐螺旋上升到极点) 和新颖的镜面穿过路径 (其中摄像机来回穿过镜子)。这些路径各自包含 100 个训练视角和 200 个测试视角, 遵循 NeRF 数据集的标准设置 [2]。每个真实场景均围绕具有反射和折射物体的环境随机拍摄, 图像数量从 62 到 118 张不等, 并按照 LLFF 数据集的组织方式进行整理 [18]。

接着, 我们通过对比实验展示了我们方法的优越性, 使用了三个具有代表性的基线模型, 即 NeRF [2]、Mip-NeRF [19] 和 Mip-NeRF 360 [4], 分别在使用和不使用我们多空间模块的情况下进行对比。此外, 我们还研究了基于网格的加速类 NeRF 方法, 并基于经典方法 (即 TensorRF [20] 和 iNGP [21]) 提出了混合型多空间模块, 以展示我们方案的兼容性。三维重建方法对多视图一致性的依赖更强; 因此, 我们将多空间模块集成进经典的基于 NeRF 的重建方法 NeuS [22] 中, 实验结果表明, 我们的模块同样有助于提升重建类方法的渲染质量。实验结果显示, 我们的方法不仅在包含反射和折射的场景中显著提升了渲染性能, 同时也在非专门针对渲染的模型中展现出良好的鲁棒性。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 度高保真场景渲染中的镜面物体, 在定性比较和定量比较上都比现有的代表性基线方法取得了显著的提升。
- 我们设计了一个轻量级模块, 可以使大多数 NeRF 类方法具备建模反射和折射的能力。
- 我们提出了一种用于 TensorRF 和 iNGP 的混合多空间方案, 展示了我们的方案与基于网格的 NeRF 方法的良好兼容性。
- 我们构建了一个专用于评估包含复杂反射与折射场景的 360 度高保真渲染的数据集, 包括 33 个合成场景和 7 个真实拍摄场景, 涵盖具有挑战性的相机路径。

2 相关工作

基于坐标系的新视图合成。 NeRF [2] 缩小了计算机视觉与计算机图形学之间的鸿沟, 并展示了一种仅依赖带位姿图像即可渲染高质量真实感场景的有前景的方向。这一方案的洞见与良好的泛化能力也推动了多个任务的发展, 如: 三维重建 [22]–[28], 三维感知生成 [29]–[33], 三维感知编辑 [34]–[36], 以及数字人重建与操控 [37]–[40]。研究者们在该方案基础上提出了大量增强方法。Mip-NeRF [19] 通过引入集成位置编码表示 3D 视锥, 从而增强了 NeRF 的抗锯齿能力。[41], [42] 将该方案扩展至 HDR 图像。[3], [4] 将 NeRF 及其变体扩展至无界场景。[22]–[24], [26], [27] 构建了 SDF 与 NeRF 体渲染中的密度之间的关系, 用于三维重建任务。同时, 也有大量工作致力于利用显式或混合表示来加速训练与推理过程 [20], [21], [43]–[47]。

具有高镜面反射的高光材质会对 NeRF 类方法产生显著影响。[48] 受到计算机图形学中基于预计算技术 [49] 的启发, 提出了一种用于表示和渲染视角依赖的高光和反射的方法, 但该方法无法处理镜面反射类型的表面, 因为镜中虚像无法被当作纹理来处理。Guo 等人 [16] 提出将反射表面分解为透射部分与反射部分, 是与我们工作最相关的方法之一。然而, 这种分解方式无法处理包含镜面反射物体的 360 度视角场景, 因为虚像在视角未达到某一特定角度之前与真实物体没有区别。Zeng 等人 [50] 将光线追踪机制引入 NeRF 以建模反射, 但其方法同样缺乏来自镜子背后的视角观测。

另一类与我们方法相似的工作是多重神经辐射场 (multiple neural radiance fields), 但其研究目的与我们不同 [30], [45], [51]–[53]。[30] 使用基于对象级别的神经辐射场进行具备 3D 感知能力的生成与合成。[45], [51] 则通过多个小型 MLP 网络实现高效渲染。[52], [53] 使用多个对象级别的神经辐射场进行三维场景分解与编辑。

常用的数据集。 研究人员已经引入或构建了许多不同的数据集, 以推动 NeRF 方法在各种任务中的发展。Mildenhall 等人 [2] 收集了一个数据集, 包含八组分别渲染的带位姿的物体图像数据, 以及八个真实拍摄的前视场景, 其相机位姿和内参由 COLMAP [54] 估算得出。然而, 这些场景缺乏反射和折射, 而这在现实中非常常见。Wizadwongsa 等人 [17] 提出了一个名为 Shiny 的数据集, 包含八个更具挑战性的场景, 用于测试 NeRF 类方法在渲染视角相关效果方面的表现, 但这些数据仍是以近似前视的方式采集的。Verbin 等人 [48] 构建了一个包含六个光滑物体的数据集, 称为 Shiny Blender, 在与 NeRF 数据集相似的渲染条件下生成, 用于测试方法在建模更复杂材料时的能力。

针对无界场景, Barron 等人 [4] 构建了一个包含 5 个户外场景和 4 个室内场景的数据集, 而 Zhang 等人 [3] 则采用了 Tanks and Temples (T&T) 数据集 [55] 以及 Light

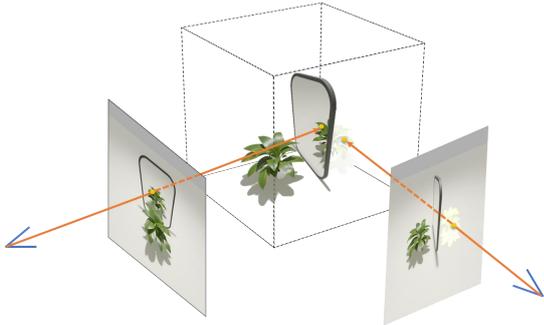


图 2: 镜面产生的虚像只在有限的视角范围内可见, 这违反了多视图一致性。

Field 数据集 [56]。Bemana 等人 [57] 拍摄了一个包含折射物体的数据集, 该数据集由四个场景组成, 摄像机在较大范围内移动。Guo 等人 [16] 收集了六个包含反射和半透明材质的前视场景, 这是迄今为止与我们最相关的数据集, 但我们构建的数据集在挑战性上更高。DTU 数据集 [58] 和 BlendedMVS 数据集 [59] 是常用于三维重建评估的基准数据集。

3 方法

3.1 前置工作: 神经辐射场

Neural Radiance Fields (NeRF) [2] 将场景编码为连续的体积场, 并嵌入到一个多层感知机 (MLP) 的权重中, 同时采用传统体积渲染中的仅吸收模型来合成新视角图像。训练过程只需要一组稀疏的带姿态图像, 朝向场景内场景投射射线 $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$, 其中 $\mathbf{o} \in \mathbb{R}^3$ 是相机中心, $\mathbf{d} \in \mathbb{R}^3$ 是视角方向, 这些射线可通过训练数据中的相机内参和外参计算得到。给定这些射线, NeRF 会在欧几里得空间中按距离相机的距离 t_i 采样一组三维点 $\{\mathbf{p}_i = \mathbf{o} + t_i\mathbf{d}\}$, 并通过以下函数将这些点投影到更高维的空间中:

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

其中, L 是一个超参数, \mathbf{p} 是一个采样点。

给定投影后的特征 $\{\gamma(\mathbf{p}_i)\}$ 和射线方向 \mathbf{d} , MLP 输出密度 $\{\sigma_i\}$ 和颜色 $\{\mathbf{c}_i\}$, 这些值被用于通过 Max [60] 提到的求积法则来估计射线的颜色 $\mathbf{C}(\mathbf{r})$:

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

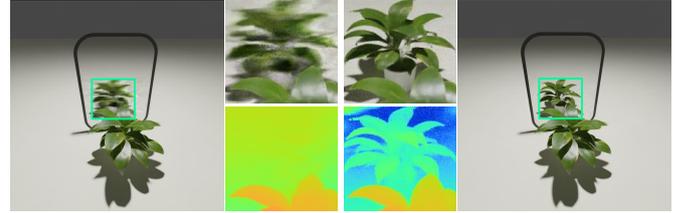
其中 $T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right)$, $\delta_i = t_i - t_{i-1}$ 。由于该公式是可微的, 模型参数可以通过均方误差 (Mean Squared Error, MSE) 损失函数直接优化:

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

其中 \mathcal{R} 表示一个批次的训练射线。此外, NeRF 还采用了一种分层采样策略, 在高权重区域采样更多的点。通过上



(a) 场景 A 的训练视图。 (b) 场景 B 的训练视图。



(c) 场景 A 的渲染视图。 (d) 场景 B 的渲染视图。

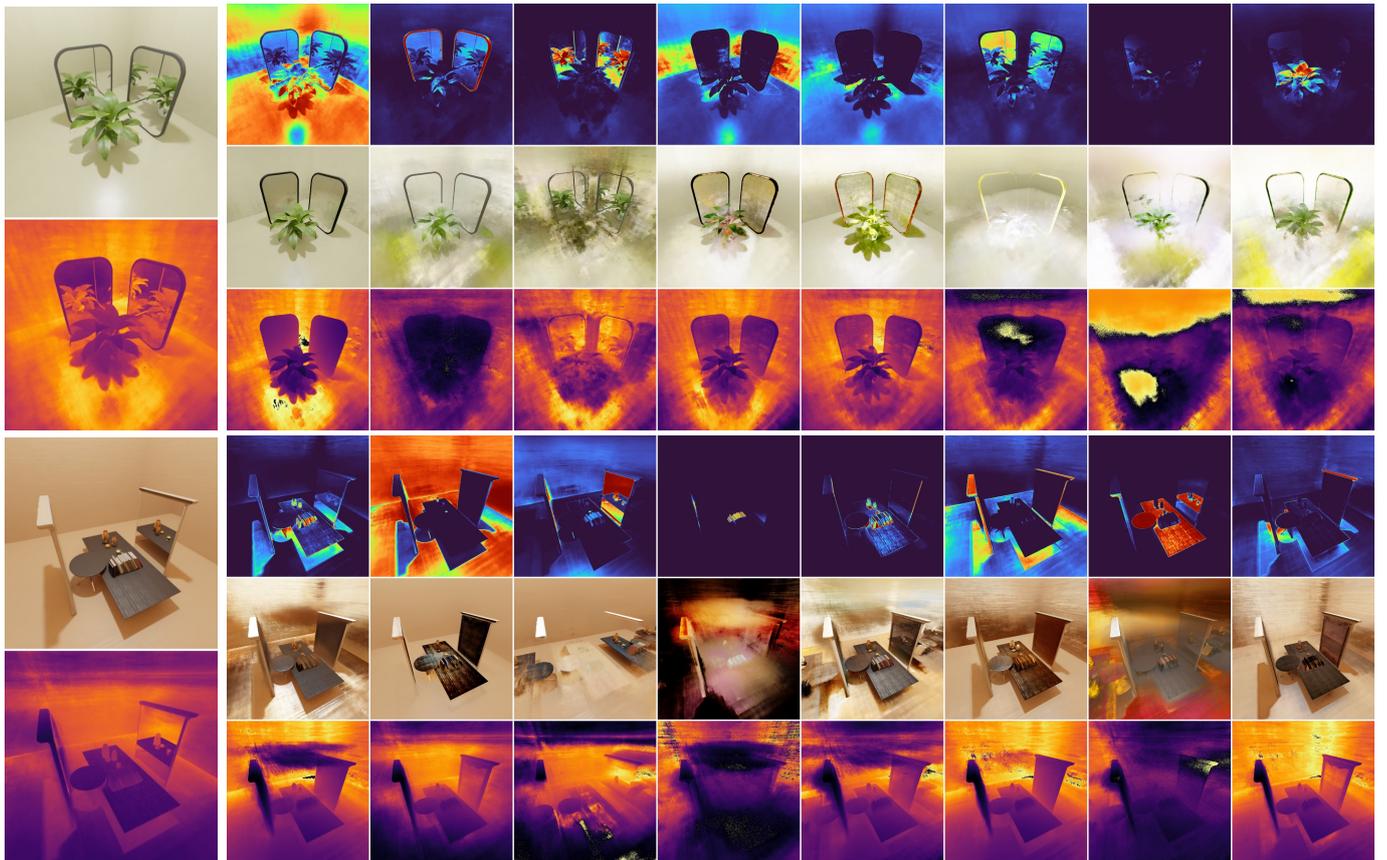
图 3: 第一行为两个场景中的训练视图示例。在场景 A 中, 镜子前面只有一株植物; 而在场景 B 中, 我们精心地放置了另一株植物, 使其位置与虚像的位置完全重合。第二行是测试视角下, 由原始 NeRF 在样例场景中训练后渲染出的深度图。正如所演示的, 当“虚像”满足多视图一致性时, NeRF 可以避免将反射图像视为纹理的陷阱。

述设计, NeRF 在大多数情况下实现了最先进的新视角合成的真实感效果。

3.2 多空间神经辐射场

体积渲染方程以及 MLP 的连续表示能力确实保障了 NeRF 系方法在新视角合成中的成功, 但正如已有研究所指出的那样 [12], [13], [16], 训练过程中还隐藏着一个有助于收敛的不可忽略的属性, 那就是多视图一致性。然而, 多视图一致性在存在反射表面时很容易被破坏。如图 2 所示, 当看向镜子前面时, 人们可以观察到反射的虚像, 就好像它后面有一个物体, 但是当从侧面或向后看时, 镜子后面实际上什么都没有。在实践中, 这意味着将会有完全冲突的训练批次扰乱 MLP 的拟合过程。

为了展示多视图一致性的重要性及其对传统 NeRF 网络结构的影响, 我们使用开源软件 Blender [61] 构建了两个 360 度的样例场景。每个场景包含 100 张训练图像和 10 张测试图像, 其训练视角示例如图 3a 和 Fig. 3b 所示。这两个场景的唯一区别在于, 在后一个场景中, 我们将一个镜像姿势的真实物体放置在镜子后面, 而在前一个场景中则没有。我们在相同设置下分别在这些样例场景上训练原始 NeRF, 并渲染测试集中部分视角, 如图 3c 和 Fig. 3d 所示。结果清楚地表明, 某些视图中消失的虚像 (即违反多视图一致性的部分) 导致模型在与反射相关的区域表现不理想, 并在渲染中产生模糊。有趣的是, 传统的 NeRF 仍在尝试满足多视图一致性假设。从 Fig. 3c



(a) 合成结果

(b) 各子空间的权重图、RGB 图像和深度图。

图 4: 我们可视化了来自 Sec. 6.4 中 MS-NeRF_B 模型在新视角下的合成 RGB 和深度图, 以及解码后的图像, 并附上 MS-NeRF_B 模型中所有子空间的相应权重和深度图。结果表明, 我们的方法成功地将虚像分解到特定的子空间中。

的深度图中, 我们可以很容易地得出结论: 传统的 NeRF 将观察到的虚像视为反射表面上的“纹理”, 在其基本假设与训练数据冲突之间达成了妥协, 尽管这种妥协会导致对真实图像的错误理解和更差的渲染结果。

与传统的 NeRF 相反, 受物理学和计算机图形学中常见观点的启发, 即反射光可以看作是从其镜像对称方向“直接发射”的, 来自可能的“镜子中虚拟空间内的虚拟源”, 我们基于以下假设构建了新颖的多空间 NeRF 方法:

Assumption 1: 在存在反射与折射的情况下, 真实的欧几里得场景空间可以被分解为多个虚拟子空间。每个子空间都满足多视图一致性性质。

基于此我们提出, 子空间的组合权重可以根据空间位置和视角方向动态变化, 所有子空间共同动态地影响最终的渲染结果。通过这种方式, 当存在反射表面时, 真实欧几里得空间中多视图一致性的破坏问题可以被克服——虚像仅被放置在某些子空间中, 而这些子空间仅在特定视角下可见, 如图 Fig. 4 所示。图 Fig. 5 中的深度图进一步验证了这一见解: 我们的多子空间模块赋予了传统 NeRF 理解“镜中虚空间内可能存在的虚拟光源”的能力, 而普通的 NeRF 则甚至无法将复杂反射建模为纹理。

	PSNR↑	SSIM↑	LPIPS↓
Mip-NeRF 360 R	<u>31.35</u>	<u>0.948</u>	0.031
Mip-NeRF 360 F	31.40	0.948	0.031
iNGP R	33.43	0.964	0.015
iNGP F	<u>33.25</u>	<u>0.962</u>	<u>0.017</u>

Table 1: 真实合成 360° 数据集上的结果, “R” 表示神经辐射场, “F” 表示神经特征场。

4 多空间模块

本节中, 我们首先在 Sec. 4.1 对神经特征场进行改进, 然后在 Sec. 4.2 介绍我们基于 MLP 的多空间 (Mutli-Space, MS) 模块, 接着在 Sec. 4.3 中将该 MLP 模块集成到基于 MLP 及基于网格的 NeRF 方法中, 分析其性能表现, 最后在 Sec. 4.4 设计了另一种混合型 MS 模块。

4.1 神经特征场

为了实现我们的多空间神经辐射场, 底层场必须具备以下特性: 重建的场景必须保持三维一致性, 因为子空间是独立的场景, 我们尝试对每个子空间进行场景级建模; 必须能够渲染出二维图像和像素级的权重图, 因为模型仅通过

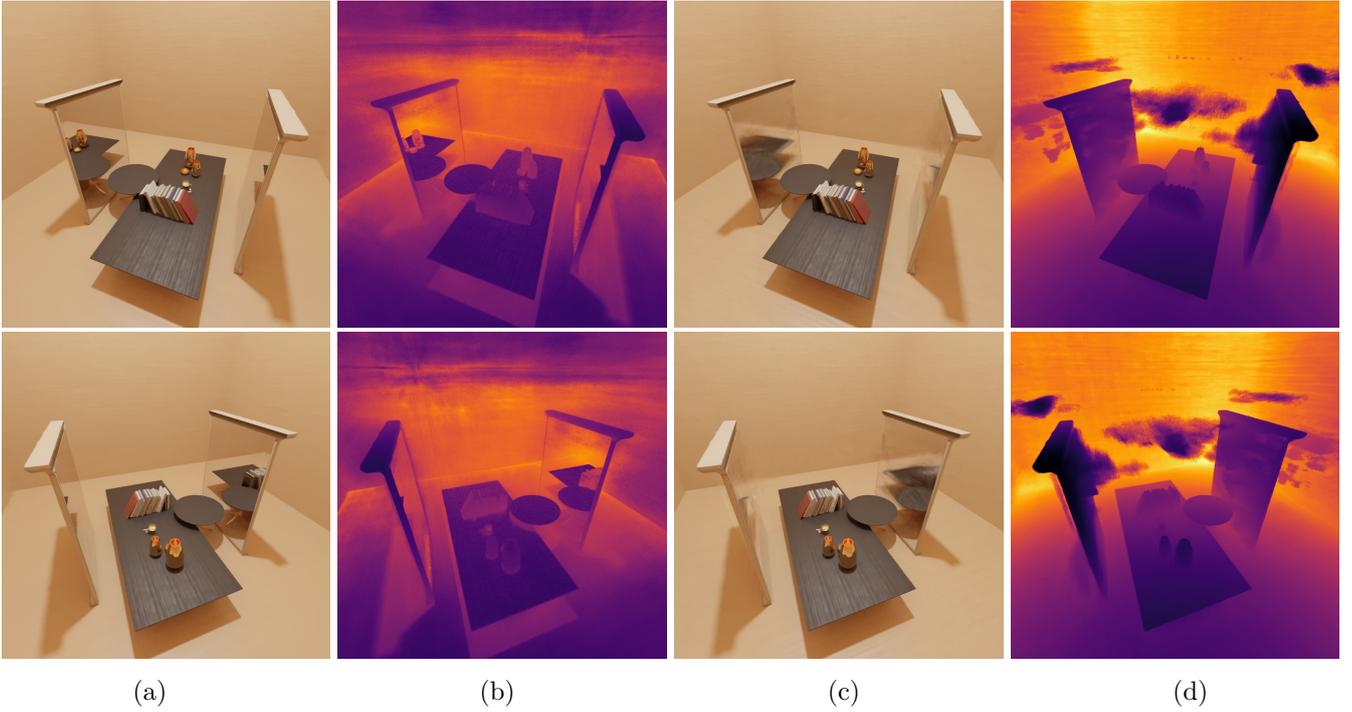


图 5: (a) MS-NeRF_B 的渲染结果。(b) MS-NeRF_B 渲染的深度图可视化。(c) NeRF 的渲染结果。(d) NeRF 渲染的深度图可视化。来自 Sec. 6.4 的可视化结果表明, 我们的 MS 模块能够理解反射发生时的光传输, 且对于普通区域也能正确地进行渲染。

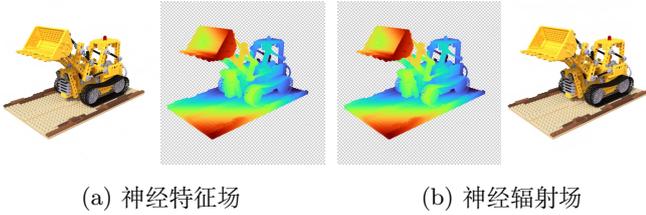


图 6: 基于 Mip-NeRF 360 的神经辐射场和神经特征场渲染的深度图和 RGB 图。

组合后的图像进行监督。虽然我们可以通过给神经辐射场增加一个额外通道来输出组合信息, 但我们在 Sec. 6.5 中通过实验验证, 增加额外通道效果并不理想。我们也可以使用神经特征场来编码 RGB 渲染视图和权重图, 其最初是为了节省内存而设计的 [30], 它用 d 维的特征 $\{\mathbf{f}_i^k\}$ 替代辐射场的输出颜色 \mathbf{c}_i 。渲染时, 神经特征场遵循与 Sec. 3.1 中相同的三维点采样和体积渲染方案, 唯一不同的是渲染得到的是特征图 $\hat{\mathbf{F}}(\mathbf{r})$, 而非颜色图 $\hat{\mathbf{C}}(\mathbf{r})$, 之后通过一个小型的 MLP Θ 将渲染的特征图解码为颜色图:

$$\{\mathcal{F}(\mathbf{r})\} \xrightarrow{\Theta} \{\mathbf{C}(\mathbf{r})\} \quad (4)$$

模型通过真实图像对其进行监督, 监督方式见 (3)。

尽管神经特征场能够对每个视角编码 RGB 图和子空间权重图, 我们仍需进一步验证神经特征场的三维一致性。随着基于 NeRF 方法的发展, 提升渲染质量或渲染速度的

努力主要分为两大分支, 代表性的方法分别是 Mip-NeRF 360 [4] 和 iNGP [21]。前者通过更优的位置参数化、更大的 MLP 以及改进的正则化来提升渲染质量; 后者设计了独特的哈希位置编码、占用网格引导采样策略, 以及高度优化的 CUDA 库, 以提高渲染速度。我们的目标是设计一个统一的框架, 提升两类基于 NeRF 方法在反射下的渲染质量。我们分别基于 Mip-NeRF 360 和 iNGP 构建神经特征场, 并在 NeRF 的真实合成 360° 数据集上进行实验。[2]。如 Tab. 1 和 Fig. 6 所示, 神经特征场在新视角合成任务中的表现与神经辐射场相当。

4.2 具有特征场的多空间模块

基于 Sec. 4.1 中的分析, 我们提出了一个紧凑的多空间模块 (MS 模块), 采用神经特征场方案 [30], 以在标准 NeRF 主干网络结构中充分提取多空间信息, 同时仅带来较小的计算开销。具体来说, MS 模块将替代 NeRF 主干网络的原始输出层。下面我们详细描述该模块的架构。

如图 Fig. 7 所示, 我们的 MS 模块仅修改了原始 NeRF 的输出部分。原始 NeRF 在场景中沿射线上的每个位置计算单个密度 σ_i 和辐射亮度 \mathbf{c}_i , 并使用公式 (2) 进行体积渲染, 得到累积颜色。相反, 我们的多输出层用神经特征场 [30] 替代了神经辐射场。具体来说, 修改后的输出层为沿射线上的每个位置输出 K 组密度 $\{\sigma_i^k\}$ 和 d 维的特征 $\{\mathbf{f}_i^k\}$, 每对密度和特征对应一个子空间, 其中 K 和 d 分别是子空间总数和神经特征场的特征维度的超参数。

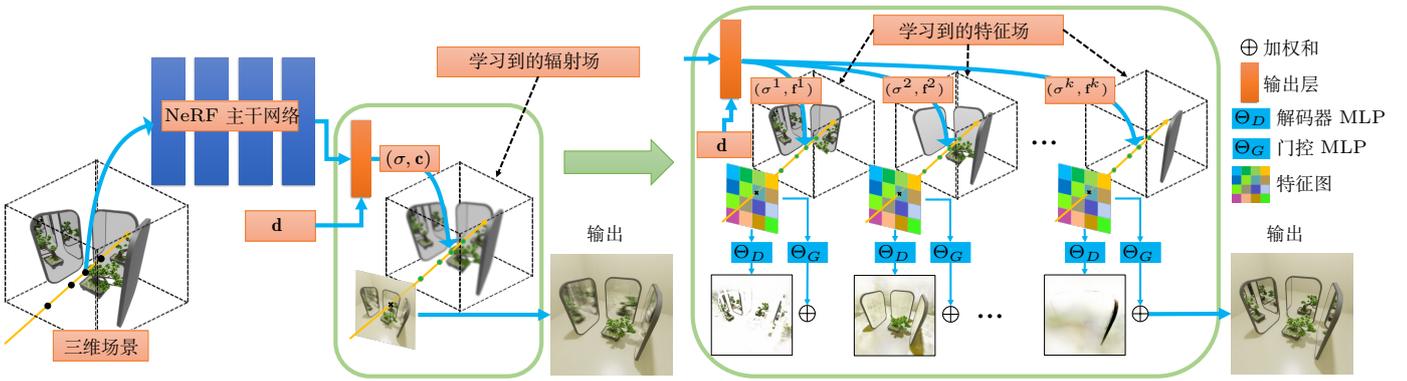


图 7: 我们的多空间模块仅修改了网络的输出部分及体积渲染部分。原始的 NeRF 计算一对密度 σ 和辐射率 \mathbf{c} ，以得到累积颜色。我们的输出层产生多对密度 $\{\sigma^k\}$ 和特征 $\{\mathbf{f}^k\}$ ，它们对应多个并行的特征场。然后，我们使用体积渲染得到多个特征图。通过两个简单的 MLP，即解码器 MLP 和门控 MLP，分别解码为 RGB 图和像素级权重。

然后，我们沿每个子空间中的射线积分特征，以收集 K 张特征图，这些特征图编码了从某个视点看到的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，我们用 $\{\mathcal{F}^k\}$ 简化表示每个像素，并在像素级别描述该操作。每个特征图像素 $\{\mathcal{F}^k\}$ 的计算公式为：

$$\hat{\mathcal{F}}^k(\mathbf{r}) = \sum_{i=1}^N T_i^k (1 - \exp(-\sigma_i^k \delta_i)) \mathbf{f}_i^k, \quad (5)$$

其中上标 k 表示射线穿过的第 k 个子空间。第 k 个密度 σ_i^k 和特征 \mathbf{f}_i^k 对应第 k 个子空间。 $T_i^k = \exp(-\sum_{j=1}^{i-1} \sigma_j^k \delta_j)$ ， $\delta_i = t_i - t_{i-1}$ ，其计算方式与公式 (2) 相同。

然后， $\{\mathcal{F}^k\}$ 会被两个小型的多层感知机 (MLP) 解码，每个 MLP 仅包含一层隐藏层。第一个是解码器 MLP (Decoder MLP)，输入为 $\{\mathcal{F}^k\}$ ，输出 RGB 向量。第二个是门控 MLP (Gate MLP)，输入同样为 $\{\mathcal{F}^k\}$ ，输出权重，用于控制某些子空间的可见性。具体地，我们使用：

$$\{\mathcal{F}^k\} \xrightarrow{\Theta_D} \{\mathbf{C}^k\}, \quad \{\mathcal{F}^k\} \xrightarrow{\Theta_G} \{w^k\}, \quad (6)$$

其中， Θ_D 表示解码器 MLP， Θ_G 表示门控 MLP。最终，MS 模块对 $\{w^k\}$ 应用 softmax 函数，作为每个子空间颜色贡献的权重，以形成最终的渲染结果：

$$\hat{\mathbf{C}}(\mathbf{r}) = \frac{1}{\sum_{i=1}^K \exp(w^i)} \sum_{k=1}^K \exp(w^k) \mathbf{C}^k. \quad (7)$$

与原始的 NeRF 方法相比，(7) 无需额外的损失项。因此，上述轻量级的多空间 (MS) 模块能够作为增强模块，附加到传统的 NeRF 类骨干网络上，我们将在 Sec. 6.4 中展示该方法带来的显著提升。

4.3 具有特征场的基于网格的多空间模块

基于网格的技术 [21] 推动了 NeRF 方法的发展，尤其是在加速领域取得了巨大进步，基于网格的 NeRF 方法能够在几分钟内收敛到相当高的质量，而基于 MLP 的方法则需要数小时到数天的训练时间。这些方法遵循小型 MLP 和

以二维/三维网格组织的可学习显式参数的设计惯例，其中网格显式编码与坐标相关的信息，MLP 更像是解码器。基于网格的 NeRF 方法遵循 (2) 中的渲染方程，但将位置编码函数 (1) 替换为查询函数 $Q(\mathbf{V}, \mathbf{p})$ ，该函数将采样点集合 $\{\mathbf{p}_i\}$ 映射到对应网格位置中的特征值 $\mathbf{V} \in \mathbb{R}^{d_x \times d_y \times d_z \times d_v}$ ，其中 d_x, d_y, d_z 表示网格在 x, y, z 轴上的分辨率， d_v 表示查询特征的维度。由于这些方法仍然像以前的方法一样从单一的辐射场渲染图像，它们也无法渲染高质量的反射和折射效果。

我们选择使用 [62] 中提出的基于采样评估器实现的 iNGP [21] 作为我们的基线，并将我们在 Sec. 4.2 中提出的 MS 模块简单地集成到其中，形成多空间 iNGP。我们在 Sec. 5.2 中的合成数据集的 Scene04 场景 (带环形相机路径) 上进行了实验，因为该场景包含相对复杂的反射和简单的物体。如图 8 所示，尽管我们的 MS 模块避免了模型崩溃，但它仍然难以渲染出高质量的反射效果。

Tab. 1 中的结果也表明，Mip-NeRF 360 从特征场中获得了性能提升，而 iNGP 在集成特征场后性能下降，因此我们怀疑 MS 模块结合特征场的性能与 NeRF 模型中 MLP 的容量有关。为了验证上述推测，我们在合成数据集中 Sec. 5.2 的 Scene05 (带环形相机路径) 上进行了实验，该场景包含无限反射且物体较简单，因此性能提升更为明显。我们将 Sec. 4.1 中的 MS 模块集成到 Mip-NeRF 360，并通过调整网络深度和宽度来缩放基础 MLP 的规模；Fig. 9 中的结果显示，更大的 MLP 能从我们的 MS 模块中获得更明显的性能提升。此外，我们在 Fig. 10 中可视化了多空间 iNGP 和多空间 Mip-NeRF 360 渲染的四个子空间的 RGB 图和通过 PCA 变换得到的特征图，结果表明，较大的 MLP 结合我们的 MS 模块可以重建更干净的子空间，而使用小型 MLP 的 iNGP 直接集成 Sec. 4.2 中的 MS 模块只能重建次优的子空间。



(a) 真实图像 (b) 未用我们的模块 (c) 使用我们的模块
图 8: Sec. 4.2 中带有和不带有我们的 MS 模块的 iNGP 之间的视觉比较。

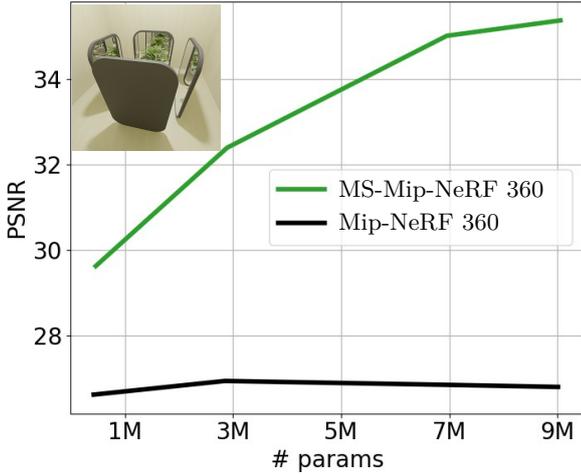
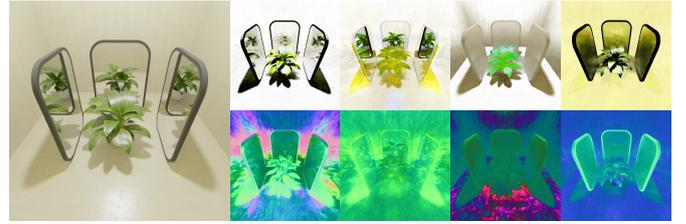


图 9: 我们将包含 6 个子空间的基于 MLP 的多空间模块集成到 Mip-NeRF 360 方法中, 并通过调整模型中 NeRF MLP 的深度和宽度来缩放网络规模; 然后, 在带环形相机路径的场景上, 通过 PSNR 指标评估性能。

4.4 混合多空间模块

虽然带有特征场的 MS 模块在基于 MLP 的 NeRF 方法中表现出良好的性能, 但它与基于网格的方法不兼容。考虑到小型 MLP 的表示能力, 我们使用神经辐射场对基于网格的方法中的子空间进行建模, 如图 Fig. 11 所示。给定沿射线的采样点 $\{\mathbf{p}_i\}$ 和显式网格参数 \mathbf{V} , 我们修改输出层以将 $Q(\mathbf{V}, \mathbf{p})$ 查询的每个特征向量和射线方向 \mathbf{d} 映射到 K 个密度 $\{\sigma_i^k\}$ 和颜色 $\{c_i^k\}$, 它们模拟多个辐射场而不是 \mathbf{d} 特征场, 其中 K 表示子空间的总数。随后执行 (5) 中的多空间积分, 不同之处在于我们积分 K 个颜色 $\{c_i^k\}$ 而不是特征 $\{f_i^k\}$, 以形成 K 个颜色图 $\{C_k\}$ 。

我们从 NeRF 模型中分解出多空间构图信息, 并将其压缩到一个小的 MLP 中, 因为在存在反射和折射的场景中, 虚像的可见性仅与 3D 位置和视线方向相关。具体来说, 我们设计了另一个带有小型 MLP 网络的分支, 该分支将每个 3D 位置 $\{\gamma(\mathbf{p}_i)\}$ (其中 $\gamma(\cdot)$ 为 (1) 中的位置编码) 和视角方向 \mathbf{d} 映射到 d 维的特征 $\{f_i\}$ 。注意, 由于模型容量限制, 我们为每个位置仅记录一个特征向量。然



(a) 多空间 Mip-NeRF 360



(b) 多空间 iNGP

图 10: 使用我们在 Sec. 4.2 中提出的 MS 模块, 对 Mip-NeRF 360 和 iNGP 渲染的子空间可视化。左图为合成图像, 右图为经过 PCA 变换的子空间 RGB 图和子空间特征图。

后, 我们沿射线将这些特征 $\{f_i\}$ 分配到每个子空间, 并结合来自多辐射场分支的 K 个密度 $\{\sigma_i^k\}$, 通过体积渲染得到 K 个特征图 $\{F^k\}$, 与 (5) 中的方法一致。最后, 我们根据 (6) 从 $\{F^k\}$ 解码出逐像素的组合权重图, 并使用来自 NeRF 分支的 K 个颜色图 $\{C^k\}$ 及解码的权重图 $\{w^k\}$, 按照 (7) 组合生成最终渲染结果。

5 数据集

5.1 现有的数据集

我们简要回顾一下与我们的任务最常用或最相关的数据集, 并在 Tab. 2 中列出了它们的属性。借助这些设计良好的数据集, 基于 NeRF 的方法在各种设置下的许多应用中都取得了很大的改进, 例如无界场景中的新型视图合成和 3D 重建。然而, 大多数现有数据集未能涵盖包含复杂光路且相机 360 度移动的场景, 例如镜子前的一杯水, 这在我们的日常生活中很常见。[16] 提出了 RFFR (带反射的真实前向视角) 数据集, 该数据集包含 6 个带反射物体 (如透明玻璃和镜子) 的前向视角场景。但反射物体背后的视角无法被评估, 而这对理解反射非常关键。[57] 提出了一个包含 4 个场景的数据集, 其中摄像机围绕中央的折射物体移动, 视角范围大到 360 度。不过该数据集接近于物体级别, 且折射效果相对简单。

5.2 我们提出的数据集

正如在第 Sec. 5.1 节总结的, 目前缺乏包含复杂反射和折射的 360 度数据集, 制约了相关研究的发展。因此, 我们构建了一个包含 33 个合成场景和 7 个真实拍摄场景的 360 度数据集。

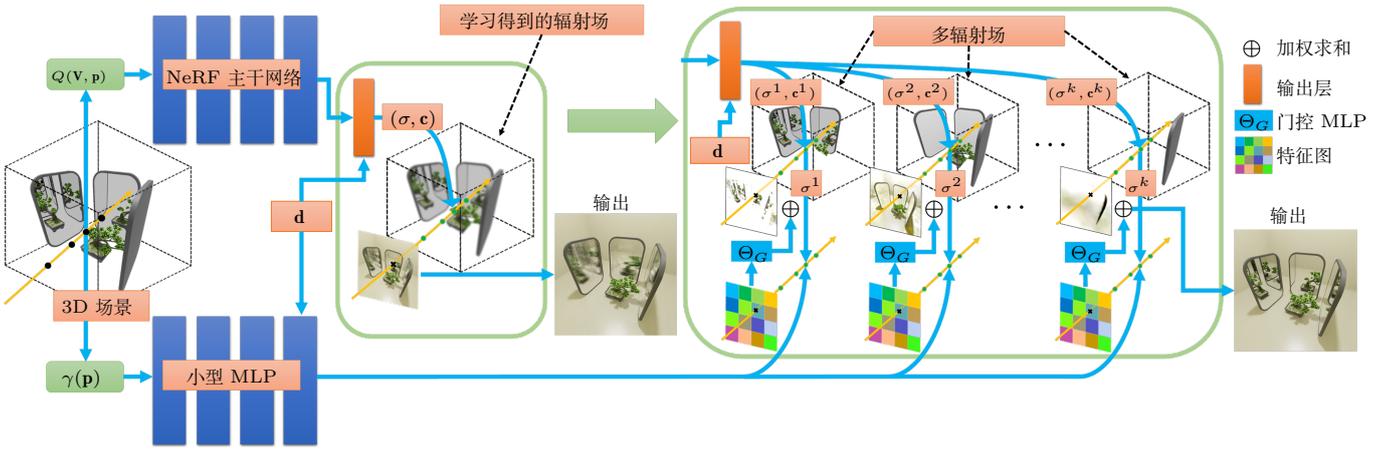


图 11: 所提出的混合 MS 模块示意图。考虑到基于网格的方法架构, 我们将多空间门控信息从辐射场分支中分离, 并用另一个小分支建模多个特征场, 从中解码出门控信息。与多个辐射场一起, 我们进行加权求和以得到最终渲染图像。

对于合成部分 (如 Fig. 12a 所示), 我们使用开源软件 Blender [61], 并利用 BlenderKit 社区提供的 3D 模型设计场景。由于我们的数据集由完整场景组成, 而非单个物体, 我们设计了三种相机路径。其中最简单的是圆形路径, 固定相机高度, 相机始终朝向场景中心, 沿圆周移动以拍摄整个场景。我们对所有场景都渲染该路径, 均匀采样 120 个点, 随机选取 100 张作为训练集, 10 张作为验证集, 10 张作为测试集。此外, 我们设计了一个 360 度螺旋路径, 相机从赤道逐渐螺旋上升至极点, 始终注视场景中心。为了进一步评估未来相关方法的鲁棒性, 我们设计了一种新颖的穿镜路径, 即相机来回穿过场景中的镜面。我们从数据集中选取了 5 个简单场景和 5 个复杂场景来渲染这两条路径, 沿路径均匀采样 300 个点, 随机选取 100 张图像作为训练集, 200 张作为测试集。我们在 Fig. 13 中可视化了这三种相机路径。

构建的数据集中包含了各种反射和折射物体的场景。我们涵盖了不同复杂度的光线路径, 这些复杂度由场景中镜子的数量和布局决定, 镜子数量从一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 [16] 更具挑战性, 因为我们的相机可以从镜子的前方移动到背后。此外, 我们还构建了带镜面墙的房间, 这些房间本质上可以看作是无界场景, 在房间中央添加镜子, 产生无界的虚像。我们进一步制作了更具挑战性的场景, 其中包含反射和折射的组合。如图 Fig. 14 所示, 我们的数据集展现了更为复杂和具有挑战性的特性。另外, 我们还为合成场景中的每个反射和折射物体渲染了实例级别的掩码。我们还包含了 7 个真实拍摄的复杂光照条件场景, 如图 Fig. 12b 所示。这些场景由两个镜子、一颗光滑表面的玻璃球、一颗钻石切割风格的玻璃球、若干玩具和几本书构成。我们以 360 度的视角随机拍摄了这些场景的照片。



(a) 我们合成数据集的一部分。

(b) 我们真实拍摄的数据集的一部分。

图 12: 我们数据集的演示场景 (更多内容见补充)。我们的数据集展现了反射和折射的多样性, 这可以作为验证合成具有复杂光路条件下的新视角合成能力的基准。

6 实验

6.1 基线方法、超参数与基准测试

为了全面评估我们方法的优越性与鲁棒性, 并展示其潜力, 我们基于不同数据集、不同基线方法以及不同规模的模块设计, 进行了多项实验。我们选择了 5 个具有代表性的基于 NeRF 的方法, 并将它们分为两类: a) 基于 MLP 的 NeRF 方法, 包括 NeRF [2]、Mip-NeRF [19] 和 Mip-NeRF 360 [4]; b) 基于网格的 NeRF 方法, 包括 TensorRF [20] 和 iNGP [21]。基于 MLP 的 NeRF 方法包含了新视角合成的代表性方法, 它们通常将底层的辐射场编码进 MLP 的权重中, 目标是渲染高质量的图像。我们基于这些方法, 在我们构建的数据集上, 使用 Sec. 4.2 中设计的基于 MLP 的 MS 模块进行实验, 以验证我们方案的优越性。基于网格的 NeRF 方法采用混合表示, 具体表现为小型 MLP 与

数据集	源头	应用	类型	视角范围	属性	数量
Realistic Synthetic 360°	[2]	新视角合成	S	360 度	非朗伯材质	8
Real Forward-Facing Shiny	[2], [18]	新视角合成	R	前向视角	非朗伯材质	8
Tanks and Temples(T&T)	[17]	新视角合成	R	前向视角	高反射材质, 折射现象	8
Mip-NeRF 360	[55]	新视角合成	R	360 度	无边界场景	4
EikonalFields	[4]	新视角合成	R	360 度	无边界场景	9
RFFR	[57]	新视角合成	R	360 度	折射材质	4
DTU	[16]	新视角合成	R	前向视角	反射现象, 半透明现象	6
BlendedMVS	[58]	重建	R	360 度	非朗伯材质物体	15*
Shiny Blender	[59]	重建	S	360 度	非朗伯材质场景	7*
Ref-NeRF Real captured scenes	[48]	新视角合成	S	360 度	光泽材质	6
	[48], [63]	新视角合成	R	360 度	光泽材质	3

Table 2: NeRF 方法中常用数据集的属性说明。“S”和“R”分别表示合成数据集和真实采集数据集。未命名的数据集我们以方法名称进行标注。“*”表示 NeRF 方法中常用的场景数量，因为原始数据集中包含的场景数量多于标注的数量，因此我们未将多余场景纳入考虑。

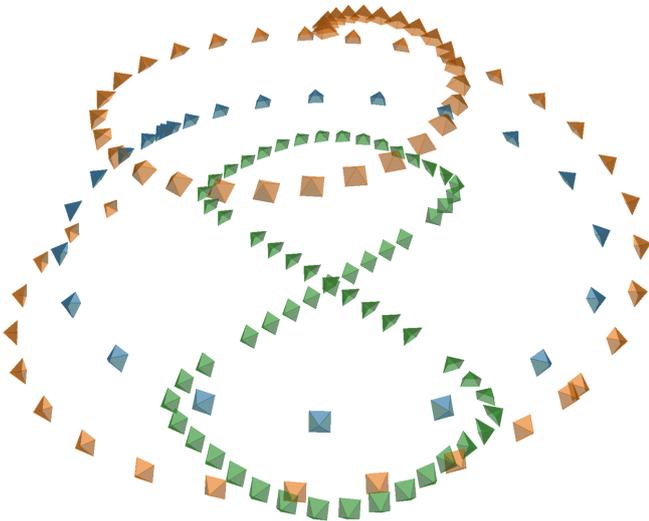


图 13: 我们设计的相机路径示意图。蓝色相机表示圆形路径，绿色相机表示穿镜路径，橙色相机表示螺旋路径。

以 3D/2D 网格组织的离散可学习或固定参数的结合，这大大减少了每次迭代中需要优化的参数量，并能在极短时间内收敛到相对高质量的渲染效果。TensorRF 和 iNGP 分别使用 2D 和 3D 网格中的可学习参数配合小型 MLP 构建网络；因此，我们将 Sec. 4.4 中描述的混合 MS 模块集成到它们中，验证我们的方法。我们基于不同数据集对这些基线方法进行了多项实验，以证明我们方案的优越性和泛化能力。由于我们的基于 MLP 的 MS 模块设计较为简单，我们可以通过三个超参数对模块进行扩展，我们用 K 表示子空间数量， d 表示输出特征的维度以及用 h 表示隐藏的解码器 MLP 和门 MLP 的维度。对于混合 MS 方案，也存在超参数 K 、 d 和 h ，不过这里的 d 和 h 仅控制门 MLP。所有训练细节详见补充材料。我们使用三个常用指标报告结果：PSNR、SSIM [64] 和 LPIPS [65]。

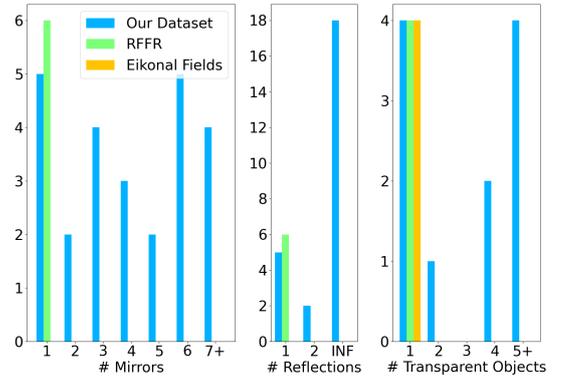


图 14: 我们对比了我们数据集与两个最相关的数据集，即 RFFR [16] 和 EikonalFields [57]，进行了定量比较。我们统计了镜子和透明物体的数量，它们分别与反射和折射直接相关。此外，我们通过场景中发生的最大反射次数来量化光路的复杂性。需要注意的是，在包含两面以上镜子的场景中，必须存在两面相对的镜子；因此，最大反射次数从 2 跃升至无穷大。

6.2 基于 MLP 的 NeRF 方法的实验

我们对 Sec. 4.2 中的模块进行扩展，并将其集成到 NeRF [2]、Mip-NeRF [19] 和 Mip-NeRF 360 [4] 中进行实验。我们大部分采用了 [2], [4], [16], [19], [48] 中的默认设置，唯一不同的是，我们每批次使用 1024 条光线，并对所有场景训练 200k 次迭代。对于基于 NeRF [2] 和 Mip-NeRF [19] 的实验，我们构建了超参数分别为 $\{K = 6, d = 24, h = 24\}$ 的 MS-NeRF_S 和 MS-Mip-NeRF_S。类似地，超参数为 $\{K = 6, d = 48, h = 48\}$ 的 MS-NeRF_M 和 MS-Mip-NeRF_M。以及超参数为 $\{K = 8, d = 64, h = 64\}$

	PSNR↑	SSIM↑	LPIPS↓	# Params
NeRF	23.79/35.24	0.662/0.902	0.240	1.159M
MS-NeRF _S	26.59/35.52	0.737/0.900	0.232	1.201M
MS-NeRF _M	26.71/35.72	0.741/0.901	0.226	1.245M
MS-NeRF _B	26.95/35.87	0.748/0.903	0.226	1.311M
Mip-NeRF	24.47/35.97	0.693/0.906	0.245	0.613M
Ref-NeRF	25.58/36.65	0.716/0.911	0.210	0.713M
MS-Mip-NeRF _S	28.08/36.60	0.779/0.906	0.224	0.634M
MS-Mip-NeRF _M	28.44/36.76	0.788/0.907	0.222	0.656M
MS-Mip-NeRF _B	28.48/36.83	0.789/0.907	0.220	0.689M
Mip-NeRF 360	24.20/37.06	0.733/0.925	0.150	9.007M
MS-Mip-NeRF 360	28.35/37.65	0.822/0.923	0.150	9.007M

(a) 我们在合成数据集的圆形路径上使用了来自 Sec. 4.2 的 MS 模块。

	PSNR↑	SSIM↑	LPIPS↓	# Params
TensorRF	24.25/33.83	0.697/0.937	0.196	17.34 4.00e-2M
MS-TensorRF	26.96/37.74	0.767/0.951	0.147	17.34 4.60e-2M
iNGP	24.04/29.98	0.743/0.915	0.201	14.23 3.28e-2M
MS-iNGP	26.59/33.67	0.800/0.930	0.157	14.23 3.33e-2M

(b) 我们在合成数据集的圆形路径上使用了来自 Sec. 4.4 的 MS 模块。

	PSNR↑	SSIM↑	LPIPS↓	# Params
Mip-NeRF 360	25.21/39.15	0.795/0.969	0.102	9.007M
MS-Mip-NeRF 360	31.31/39.98	0.894/0.966	0.098	9.052M
Mip-NeRF 360	24.87/34.02	0.828/0.915	0.184	9.007M
MS-Mip-NeRF 360	26.74/34.03	0.851/0.910	0.188	9.052M

(c) 我们在合成数据集的螺旋路径和穿镜路径上使用了来自 Sec. 4.2 的 MS 模块。前两行为螺旋路径，后两行为穿镜路径。

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

(d) 在我们真实拍摄的数据集上的比较。

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

(e) 在 RFFR 数据集上的比较。

Table 3: 与现有方法的定量比较。参数中带有“A|B”格式 的表示网格参数和 MLP 参数；其他均为 MLP 参数。对于 我们合成数据集上的 PSNR 和 SSIM 指标，我们根据 反射和折射区域的掩码对渲染图像进行划分，指标以“反 射和折射区域 / 其他区域”的形式报告。

的 MS-NeRF_B 和 MS-Mip-NeRF_B。

对于基于 Mip-NeRF 360 [4] 的实验，我们构建了超参 数为 $\{K = 8, d = 32, h = 64\}$ 的 MS-Mip-NeRF 360。此 外，我们还与 Ref-NeRF [48] 进行了对比，因为它以 Mip- NeRF 为基线，并且在建模光滑材质方面表现优异。由于

	PSNR↑	SSIM↑	LPIPS↓	# Params
Mip-NeRF	30.74	0.942	0.047	1.264M
MS-Mip-NeRF _B	30.81	0.943	0.047	1.295M
Mip-NeRF	25.78	0.775	0.213	1.264M
MS-Mip-NeRF _B	25.59	0.764	0.223	1.295M

Table 4: 真实合成 360° 数据集（前两行）和真实正面拍 摄数据集（后两行）的结果。



(a) Mip-NeRF 360

(b) MS-Mip-NeRF 360

图 15: Mip-NeRF 360 与 MS-Mip-NeRF 360 的视觉对比。 我们的模块能够扩展 Mip-NeRF 360 以建模无界的虚拟场 景。

NeRF、Mip-NeRF 和 Ref-NeRF 均针对有限边界场景设 计，我们在前 25 个带有圆形路径的场景上评估这些方法， 因为后 8 个场景则等价于无限边界场景。对于基于 Mip- NeRF 360 的实验，我们在所有场景的所有路径上进行评 估。

我们还在 RFFR 数据集 [16] 上，将我们的方法与 NeRFReN 进行了比较。NeRFReN 是专门针对前向场景 中镜面材质设计的基于原始 NeRF 的双分支网络，因此我 们基于 NeRF 构建了一个精简版本，称为 MS-NeRF_T，其 超参数为 $\{K = 2, d = 128, h = 128\}$ 。这里使用两个子空 间，是因为 NeRFReN 尝试将反射表面分解为两部分，我 们希望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 NeRFReN，采 用提供的默认设置，唯一不同的是我们将使用的掩码数量 设为 0，因为我们的方法不需要额外的掩码。

6.3 基于网格的 NeRF 方法实验

我们在基于网格的方法上进行了实验，以展示我们方案与 其他应用的兼容性，而不仅仅局限于基于 MLP 的渲染方 法。

iNGP [21] 和 TensorRF [20] 通过结合可学习的 3D/2D 特征网格和非常小的 MLP，实现了快速收敛。我们分别 基于 Sec. 4.4 中提出的混合 MS 模块构建了 MS-iNGP 和 MS-TensorRF，超参数设为 $\{K = 4, d = 8, h = 32\}$ ，并用 一个结构类似但规模更小的 MLP 初始化额外分支，该分 支对应主分支中的 MLP。我们在 MS-TensorRF 实验中遵 循所有默认设置 [20]，并基于 [62] 的实现和其采样评估器 (proposal estimator) 实现了 MS-iNGP。由于 TensorRF 设

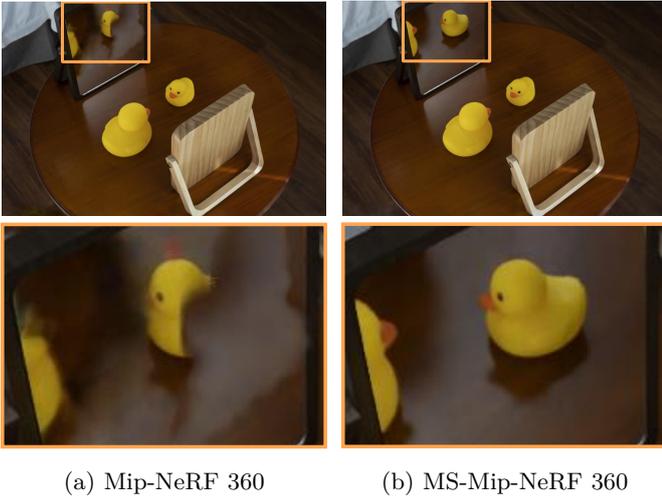


图 16: Mip-NeRF 360 与 MS-Mip-NeRF 360 在我们数据集真实拍摄部分的视觉对比。我们的方法足够鲁棒，能够恢复现实世界中的虚像。

计用于有界场景，我们在前 25 个带有圆形路径的场景上评估 TensorRF。相反，带采样评估器的 iNGP [62] 能够重建无界场景，因此我们在所有带圆形路径的场景上评估与 iNGP 相关的模型。

6.4 对比实验

定量比较。如表 3a 所示，我们基于 MLP 的模块可以集成到大多数类似 NeRF 的模型中，并且在引入极少额外开销的情况下显著提升性能。尤其是在基于 Mip-NeRF 360 的实验中，我们的模块在 PSNR 上提升了 4.15 dB，仅增加了 0.5% 的参数量，且不影响非镜面区域的表现。此外，我们基于 Mip-NeRF 的模型在镜面区域上也大幅优于 Ref-NeRF [48]，后者是基于 Mip-NeRF 的一个变体，具备出色的高光材质建模能力。

含镜面穿越路径和螺旋路径的合成部分及真实拍摄场景更加具有挑战性，因此我们在表 3c 和表 3d 中，将我们的结果与基于 MLP 的方法 Mip-NeRF 360 进行了比较。我们的方法同样展现了显著提升，这表明了我们方案的鲁棒性。具体而言，螺旋相机路径会在虚像上引入更大的视角变化，而镜面穿过相机路径会导致虚像沿光线方向快速消失，结果表明我们的方案能够处理这些问题，而不会降低非镜面区域的性能。

在 Tab. 3b 中的结果表明，我们提出的混合式 MS 模块兼容 TensorRF 和 iNGP。如表中所示，该模块不仅能够提升反射区域的重建性能，同时还能增强其他区域的重建稳定性。

在 RFFR 数据集上，该数据集包含正面视角下的反射表面，我们的方法在训练中不使用任何人工标注掩码的情况下，依然取得了更优的结果，如 Tab. 3e 所示。对于

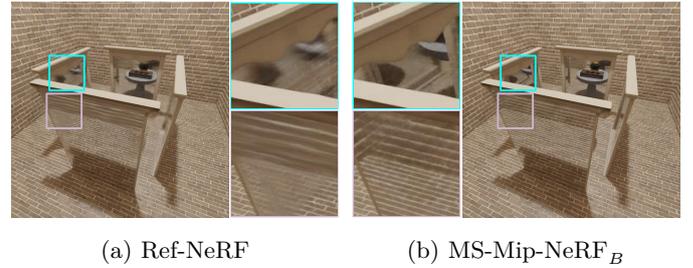


图 17: MS-Mip-NeRF_B 与 Ref-NeRF 的视觉对比。我们的方法在反射表面上显著优于 Ref-NeRF。

NeRFReN，我们使用官方代码在提供的设置下重新训练模型，唯一不同的是我们将用于反射区域的掩码数量设置为 0。

Realistic Synthetic 360° 数据集和 Real Forward-Facing 数据集最早由 [2] 提出，是评估 NeRF 类方法新视角合成能力的常用基准数据集。因此，我们也在这些数据集上训练了 Mip-NeRF 及其扩展版本 MS-Mip-NeRF_B，因为 Mip-NeRF 是 NeRF 方法中被广泛采用的主干网络之一。相关结果展示在 Tab. 4 中，实验表明我们的多空间模块并不会影响 NeRF 方法在普通材质上的表达能力。

综上所述，以上所有实验均展示了我们方法的优越性与良好的兼容性。

定性对比与讨论。除了定量对比外，我们还总结了我们的模块的优势，并通过与相应基线方法的定性或定量对比以支撑我们的结论。

与当前最先进的基于 MLP 的方法 Mip-NeRF 360 的定性对比如图 Fig. 1、Fig. 15 和 Fig. 16 所示。我们的方法能够在合成场景和真实场景中渲染出高保真的虚像，无论是有界的还是无界的。此外，如图 Fig. 1 所示，我们的 MS 模块在相机穿越镜面时发生剧烈变化的场景下也表现出强大的鲁棒性。

我们还与 Ref-NeRF [48] 进行定性比较，该方法基于传统 NeRF 主干构建，因此将虚像理解为纹理进行建模。如图 Fig. 17 所示，由于 Ref-NeRF 同样是以 Mip-NeRF 为基础构建的，我们将我们的 Mip-NeRF 变体与 Ref-NeRF 在相同的主干下进行比较，并在参数量相当的前提下（我们的模型为 0.689M，Ref-NeRF 为 0.713M）进行对比。再次，定性结果显示了我们的显著改进。

我们还与 NeRFReN 模型进行了比较，该模型需要准确标记反射区域的掩模，且仅能处理面向前方的反射。在该对比中，我们在合成数据集上训练其模型，并提供了准确的反射掩码用于训练。如图 Fig. 18 所示，他们的模型在 360 度场景中无法还原出高保真图像，而我们的方法在无需任何额外标注信息的情况下仍能实现高质量渲染。

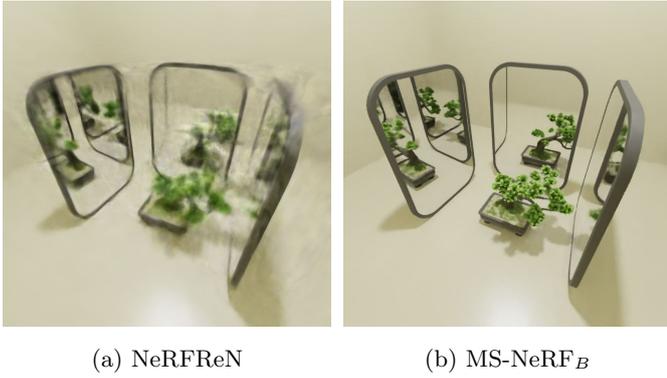


图 18: (a) 使用准确标记的掩码图像进行训练, NeRFReN 甚至无法渲染带有镜子的 360 度场景中的普通部分; (b) 我们的方法无需额外的人工掩码标注, 却能够渲染出高质量图像。

6.5 消融实验

在本节中, 我们评估所提出模块的设计, 并探讨子空间数量与虚像数量之间的关系。

关于使用神经特征场的消融实验。我们实现了一个简单的模块, 该模块输出 K 个标量 $\{\sigma_i^k\}$ 和 K 个四维 RGB-G 向量 $\{\mathbf{c}_i^k\}$, 其中每个向量包含三个 RGB 通道以及一个用于控制多空间组合的门控信息的标量。在渲染时, 我们使用与 NeRF 相同的积分公式对 K 个 RGB-G 向量 $\{\mathbf{c}_i^k\}$ 进行累积, 从而获得每个子空间的 RGB-G 图, 然后将 RGB-G 图拆分为 RGB 图和子空间组合图 (即权重图), 接着在子空间通道维度上对权重图进行 softmax 归一化, 将其作为权重来加权组合各子空间的 RGB 图, 从而形成最终的渲染结果。我们将这一设计集成到 Mip-NeRF 360 中, 记作 MS-Mip NeRF 360_p, 并设定 $K = 8$ 。我们也将第 Sec. 4.4 节所提出的混合模块集成进 Mip-NeRF 360, 记作 MS-Mip NeRF 360_h。实验结果展示在 Tab. 5a 中, 我们还在 Fig. 19 中展示了一些可视化结果, 结果表明, 简单的多空间辐射场假设可以在一定程度上帮助模型应对反射带来的违反多视图一致的问题, 但由于缺乏高效的多空间组合策略, 也会引入过度平滑的问题。

关于子空间数量的消融实验。在欧几里得空间中, 我们可以通过镜子的数量与布局来控制虚拟子空间的数量。例如, 当两面镜子相对摆放时, 可能会产生无限递归的虚像空间; 但当两面镜子背对背放置时, 每面镜子后面仅会有一个虚像。为了对我们模块的设计提供指导, 我们从合成数据集中选择了两个包含不同镜子布局的场景, 在这些场景上训练了若干具有不同子空间数量和特征维度的 NeRF 变体模型。

我们基于 NeRF 构建了多个变体, 输出特征维度设为 $d \in \{24, 48, 64\}$, 子空间数量设为 $K \in \{2, 4, 6, \dots, 16\}$, 然

	PSNR↑	SSIM↑	LPIPS↓	参数数量
Mip-NeRF 360	29.84	0.942	0.100	9.007M
MS-Mip NeRF 360 _h	33.59	0.958	0.081	9.060M
MS-Mip NeRF 360 _p	<u>37.50</u>	<u>0.972</u>	<u>0.064</u>	9.018M
MS-Mip NeRF 360	38.84	0.977	0.054	9.052M

(a) 基于 MLP 的 MS 模块架构的消融研究。

	PSNR↑	SSIM↑	LPIPS↓	参数数量
MS-TensoRF-grid	<u>32.07</u>	<u>0.932</u>	<u>0.130</u>	21.73 5.60e-2M
MS-TensoRF	34.72	0.938	0.113	17.23 4.60e-2M

(b) 混合 MS 模块架构的消融研究。

Table 5: 关于模块架构的消融研究。参数形式为“A|B”的表示网格参数和 MLP 参数; 其余项均为 MLP 参数。

后在这两个场景上训练模型, 并使用 PSNR 作为指标报告实验结果。我们的实验结果如 Fig. 20 所示, 表明子空间数量无需匹配真实虚像空间的数量, 且 6 个子空间可以保证对多空间辐射场的稳定学习。此外, 特征维度为 $d = 24$ 的特征场已经能够编码足够的信息用于空间组合, 但为了获得更稳定的性能, $d = 48$ 是更优的选择。

关于混合型 MS 模块的消融实验。我们进一步设计了一个基于网格的 MS 模块, 为基于网格的方法提供参考。具体来说, 我们将混合型 MS 模块 (详见 Sec. 4.4) 中的位置编码 $\gamma(\mathbf{p})$ 替换为通过 $Q(\mathbf{V}_{vis}, \mathbf{p})$ 查询得到的特征, 其中 \mathbf{V}_{vis} 是一个与主分支中的 \mathbf{V} 类似但分辨率更低的网格。我们将集成了该网格型 MS 模块的模型称为 MS-TensoRF-grid, 并在合成数据集中具有圆形路径的前 10 个场景上, 将其与 Sec. 6.3 中的 MS-TensoRF 模型进行对比实验。如 Tab. 5b 所示, 实验结果表明, 多空间建模更依赖于位置信息, 因此我们的混合型 MS 模块在成本更低的情况下实现了更优的性能。

输入图像数量的消融实验。为了进一步验证我们方法的鲁棒性, 我们将 MS-Mip-NeRF 360 模型与基线方法 Mip-NeRF 360 在合成数据集中的场景 01 至场景 05 上进行比较, 使用包含三条相机路径的训练视角, 分别随机选取 100、75 和 30 张训练图像。详细结果见表 Tab. 7, 结果表明我们的方法在不同输入图像数量下均能稳健提升基线模型在反射区域的性能。我们在补充材料中提供了更多视觉对比, 更好地探索我们方案的行为。

6.6 重建几何

虽然 NeuS [22] 将隐式距离函数 (SDF) 场与体渲染相结合用于三维重建, 但它继承了基于密度的体渲染在镜面重建上的不足。为了推广我们的方案, 解决新视图合成中的多视图不一致性问题, 我们将第 Sec. 4.2 节中提出的 MS 模块集成到 [66] 中实现的 NeuS 中, 构建了 MS-NeuS, 超

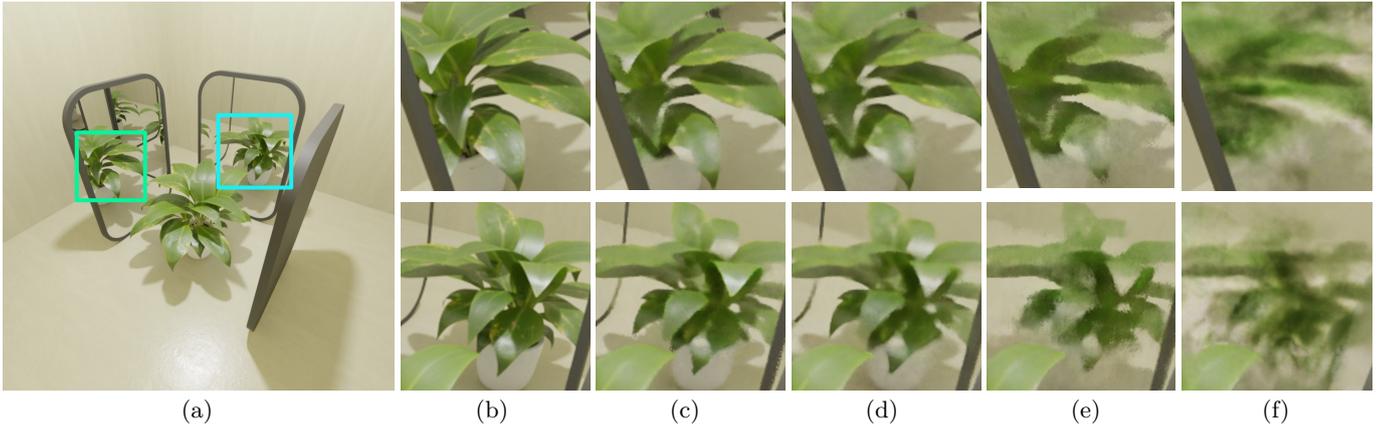


图 19: 渲染图像的详细对比。(a) 场景概览; (b) 真实图像 (Ground-Truth) ; (c) MS-Mip-NeRF 360_p; (d) MS-Mip-NeRF 360_h; (e) MS-Mip-NeRF 360_h; (f) Mip-NeRF 360。

参数设置为 $\{K = 4, d = 16, h = 32\}$ 。我们在合成数据集中圆形路径上进行了实验，渲染视图的定量和定性比较见表 Tab. 6和图 Fig. 21。如图 Fig. 21所示，我们的方案帮助 NeuS 渲染出更好的新视角图像，并对镜面重建了更优的几何结构，但仍难以重建出干净的几何形状，因为我们的 MS 模块设计初衷是新视角合成，且对子空间没有几何约束。

6.7 渲染速度

我们的方案对每个像素执行多次体积渲染操作，因此时间消耗与重要性采样策略和主干网络有关。我们在表 Tab. 8中报告了平均训练和渲染时间。对于基于 MLP 的方法，时间消耗的瓶颈在于网络推理，且每条射线采样的点数是固定的，因此集成我们的方案对速度的影响相对较小。TensorRF 使用基于占据网格的采样策略，而在我们的多空间版本中，我们将所有子空间中占据的位置统一记录到一个占据网格中，导致每条射线采样的点更密集，因此时间消耗显著增加。相反，我们的 iNGP 版本基于 NerfAcc 框架构建，采用采样评估器的采样策略，且采样点数固定，因此时间消耗增加不到一秒。对于所有我们的模块，射线方向上存在冗余采样点，可能的解决方案是对不同子空间进行自适应采样，这将作为未来的研究方向。

7 讨论

结合我们在第 Sec. 3.2节的动机和第 Sec. 6节的大量实验，我们进行了更多实验分析以探讨子空间分解的机制。第 Sec. 7.1节的实验表明，我们的多空间方案赋予密度场处理多视角不一致的能力，且无需图像监督。第 Sec. 7.2节的实验进一步证明，在不同损失的监督下，我们的多空间方案成功地将多视角不一致部分分解到不同的子空间中。以上实验确认，多空间分解仅由虚拟或真实图像决定，无需额外的正则化约束。

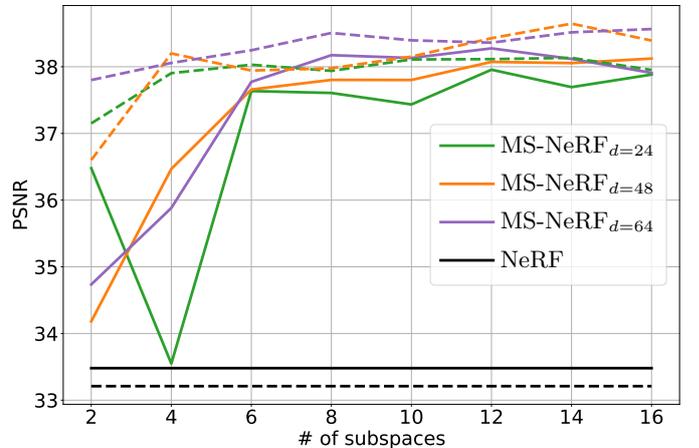


图 20: 我们使用 PSNR 定量评估场景 01 和场景 02 上的消融实验，并分别用实线和虚线绘制结果。

7.1 具有密度场的多空间方案

为了研究多空间方案将多视角不一致空间分解到子空间的机制，我们在仅包含密度输出、没有辐射度输出的密度场上进行了实验，以便展示更清晰的分解效果。具体来说，我们选择了只包含密度输出的原始 NeRF 模型，称为 NeRF-depth，因此该模型通过渲染方程 (2) 监督渲染深度图：

$$\hat{\mathbf{D}}(\mathbf{r}) = \sum_{i=1}^N T_i (1 - \exp(-\sigma_i \delta_i)) t_i \quad (8)$$

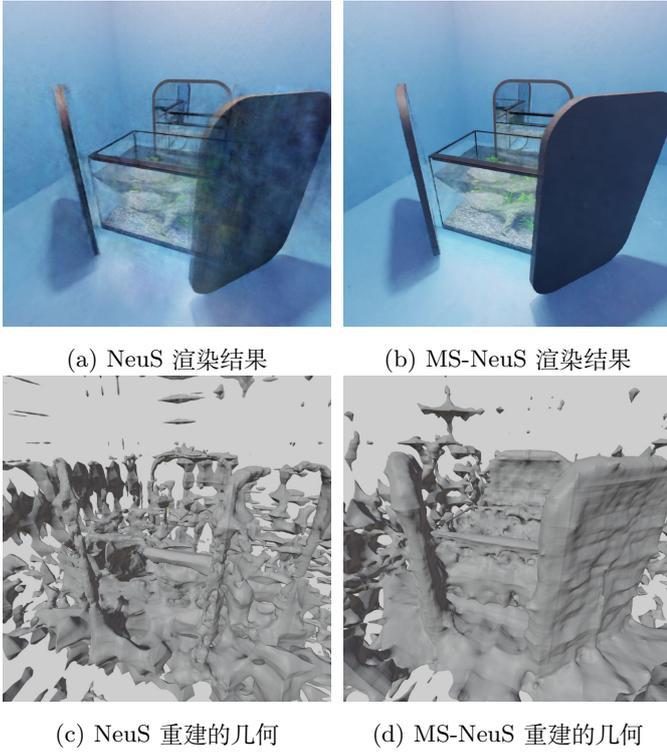


图 21: 我们的 MS 模块帮助 NeuS 渲染高质量的反射新视图, 并改进镜子的重建几何形状。

	PSNR↑	SSIM↑	LPIPS↓	# Params
NeuS	25.86	0.831	0.258	27.96 0.16e-1M
MS-NeuS	28.65	0.859	0.195	27.96 0.37e-1M

Table 6: 基于 NeuS 模型的 MS 模块 (详见第 Sec. 4.2 节) 在我们合成数据集中圆形路径的表现。参数中格式为 “A|B” 表示网格参数和 MLP 参数。

其中, $T_i = \exp(-\sum_{j=1}^{i-1} \sigma_j \delta_j)$, $\delta_i = t_i - t_{i-1}$, $\hat{\mathbf{D}}(\mathbf{r})$ 是渲染得到的深度图。我们通过将第 Sec. 4.2 节中的多空间模块集成到 NeRF-depth 中, 构建了 MS-NeRF-depth 模型, 超参数设置为 $K = 4, d = 8, h = 32$, 同样仅由深度图组成:

$$\hat{\mathbf{D}}(\mathbf{r}) = \frac{1}{\sum_{i=1}^K \exp(w^i)} \sum_{k=1}^K \exp(w^k) \hat{\mathbf{D}}^k. \quad (9)$$

其中, w^k 是如公式 (6) 中所示的子空间组合权重, $\hat{\mathbf{D}}^k$ 是子空间的深度图。为了公平比较, 我们还基于 NeRF-depth 构建了 NeRF-depth-v 模型, 该模型在输出层之前除了位置外还将视角方向作为输入, 因为我们的模块需要视角方向作为输入。所有模型均以深度图进行监督, 且我们直接采用均方误差 (MSE) 损失。我们在渲染的射线长度图上训练这些模型, 图中每个像素包含射线从相机到第一个非反射表面的物理传播长度, 场景配置遵循图 Fig. 22a 中的圆形相机路径。

Model (# input images)	PSNR↑	SSIM↑	LPIPS↓
Mip-NeRF 360 (100)	21.84/40.17	0.660/0.980	0.087
MS-Mip-NeRF 360 (100)	32.11/43.04	0.913/0.984	0.051
Mip-NeRF 360 (75)	21.46/39.37	0.643/0.978	0.088
MS-Mip-NeRF 360 (75)	31.44/42.20	0.890/0.982	0.055
Mip-NeRF 360 (30)	19.08/34.51	0.549/0.966	0.102
MS-Mip-NeRF 360 (30)	28.37/39.84	0.854/0.979	0.060

(a) 圆形相机路径结果

Model (# input images)	PSNR↑	SSIM↑	LPIPS↓
Mip-NeRF 360 (100)	22.26/40.46	0.711/0.982	0.083
MS-Mip-NeRF 360 (100)	32.30/42.39	0.905/0.984	0.049
Mip-NeRF 360 (75)	21.69/39.60	0.692/0.980	0.084
MS-Mip-NeRF 360 (75)	31.12/41.78	0.893/0.983	0.052
Mip-NeRF 360 (30)	18.86/33.40	0.596/0.960	0.116
MS-Mip-NeRF 360 (30)	25.14/36.49	0.792/0.969	0.086

(b) 螺旋相机路径结果。

Model (# input images)	PSNR↑	SSIM↑	LPIPS↓
Mip-NeRF 360 (100)	26.69/35.63	0.857/0.962	0.112
MS-Mip-NeRF 360 (100)	29.39/35.83	0.896/0.962	0.104
Mip-NeRF 360 (75)	25.25/34.93	0.841/0.958	0.118
MS-Mip-NeRF 360 (75)	28.21/34.60	0.883/0.955	0.113
Mip-NeRF 360 (30)	21.60/25.11	0.766/0.889	0.237
MS-Mip-NeRF 360 (30)	21.58/23.04	0.782/0.864	0.287

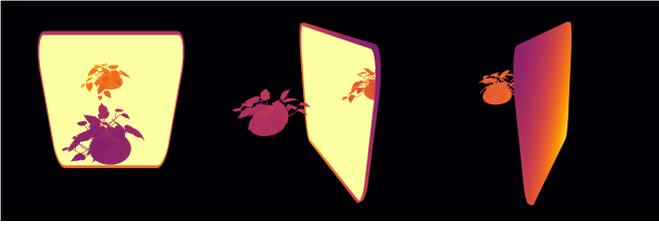
(c) 镜子穿过相机路径的结果。

Table 7: 我们合成数据集中 Scene01 至 Scene05 所有相机路径下, MS-Mip-NeRF 360 与基线模型在不同输入图像数量条件下的定量比较。我们也分别列出了指标, 指标报告方式与 Tab. 3 中一致。

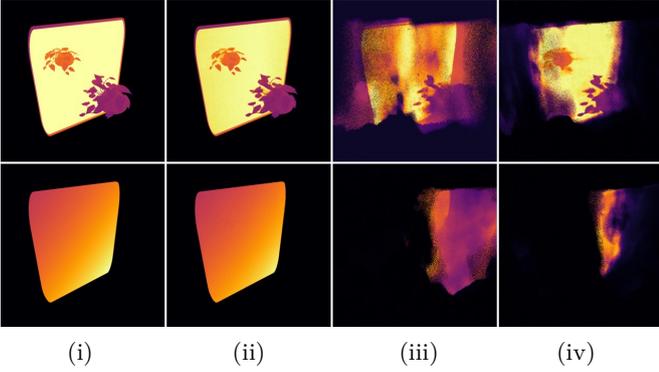
method	training time↓	rendering time(per frame)↓
Mip-NeRF	5.20 h	21 s
MS-Mip-NeRF _B	6.10 h	23 s
Mip-NeRF 360	12.56 h	32 s
MS-Mip-NeRF 360	13.30 h	39 s
TensorRF	0.42 h	3.5 s
MS-TensorRF	1.25 h	11.3 s
iNGP	0.56 h	3.1 s
MS-iNGP	0.67 h	3.6 s

Table 8: 单个 GeForce RTX 3090 GPU 上的平均训练和渲染时间比较。

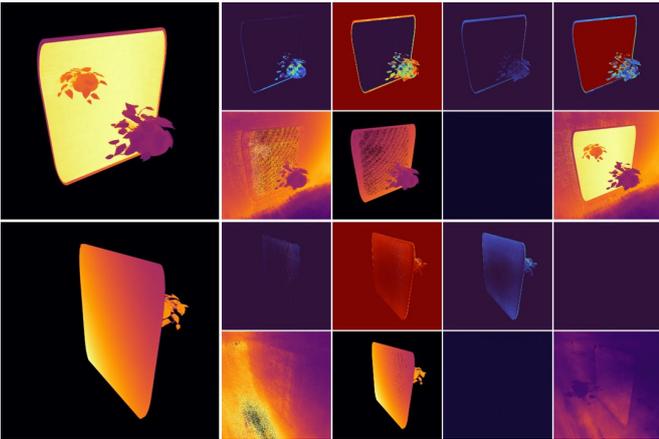
如图 Fig. 22b(iii) 和 Fig. 22b(iv) 所示, 由于反射射线引起的射线传输长度的多视角不一致, NeRF-depth 和 NeRF-depth-v 都无法重建出正确的密度场。相反, 我们的多空间方案成功处理了反射射线, 正如图 Fig. 22b(ii) 所示, 且图 Fig. 22c 的可视化结果证明了这一成功处理是通



(a) 真实值图渲染的光线长度贴图。



(b) 渲染光线长度贴图的视觉比较。(i) 真实值; (ii) MS-NeRF-depth; (iii) NeRF-depth; (iv) NeRF-depth-v.



(i) 完整的渲染结果。(ii) 子空间的权重和射线长度。

(c) 完整渲染结果和分解结果的可视化。

图 22: 仅具有密度场的多空间方案的定性实验。

过自动将虚像与真实图像分离实现的。这些实验验证了我们的方案不仅能够处理外观域中的多视角不一致，还能够处理几何域中的多视角不一致，且子空间分解遵循将多视角不一致部分分离到不同子空间、并在各子空间内保持多视角不一致的原则。

7.2 不同监管的多空间方案

为了探究子空间分解与不同监督损失之间的关系，我们选用了常用的光度损失函数，包括平均绝对误差 (MAE) 损失 \mathcal{L}_{MAE} 、均方误差 (MSE) 损失 \mathcal{L}_{MSE} 、结构相似性指数

	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
NeRF w/ loss (b)	22.00/38.88	0.638/0.972	0.109
MS-NeRF _B w/ loss (a)	24.31/39.24	0.737/ 0.976	0.093
MS-NeRF _B w/ loss (b)	28.09 /39.94	0.831 /0.974	0.075
MS-NeRF _B w/ loss (c)	27.00/ 38.97	0.803/0.971	0.082
MS-NeRF _B w/ loss (d)	26.06/38.35	0.763/0.969	0.096

Table 9: 不同光度损失下的定量评估。训练损失组合分别为 (a) \mathcal{L}_{MAE} ; (b) \mathcal{L}_{MSE} ; (c) $0.5 \times \mathcal{L}_{MSE} + 0.5 \times \mathcal{L}_{SSIM}$; (d) $0.9 \times \mathcal{L}_{MSE} + 0.1 \times \mathcal{L}_{LPIPS}$ 。

(SSIM) 损失 \mathcal{L}_{SSIM} ，以及感知图像补丁相似性 (LPIPS) 损失 \mathcal{L}_{LPIPS} ，对来自我们合成数据集的 Scene01–Scene05 场景中圆形路径下的 MS-NeRF_B 模型进行监督训练。

我们在表 Tab. 9 中提供了定量评估结果，表明我们的多子空间方案兼容不同的监督损失函数。此外，图 Fig. 23 中的分解可视化进一步验证了我们在 Sec. 3.2 中的动机：子空间的分解仅由图像是虚像还是实像决定。

8 结论

在本文中，我们针对基于 NeRF 方法中长期存在的反射表面渲染问题进行了研究。我们提出了一种多子空间 NeRF (MS-NeRF) 方法，将欧几里得空间分解为多个虚拟子空间。所提出的 MS-NeRF 方法相比传统的 NeRF 方法在效果上有显著提升。此外，MS 模块具有轻量化设计，使我们的方法能够作为对传统 NeRF 方法的增强组件使用。我们还构建了一个用于评估此类任务的新颖数据集，希望能够为该领域的后续研究提供帮助。

致谢。 本研究得到了国家自然科学基金 (62132012)、中央高校基本科研业务费专项资金 (南开大学, 编号 63233080) 以及天津市科技计划项目 (22JCYBJC01270) 的资助。计算资源由南开大学超级计算中心 (NKCS) 提供支持。

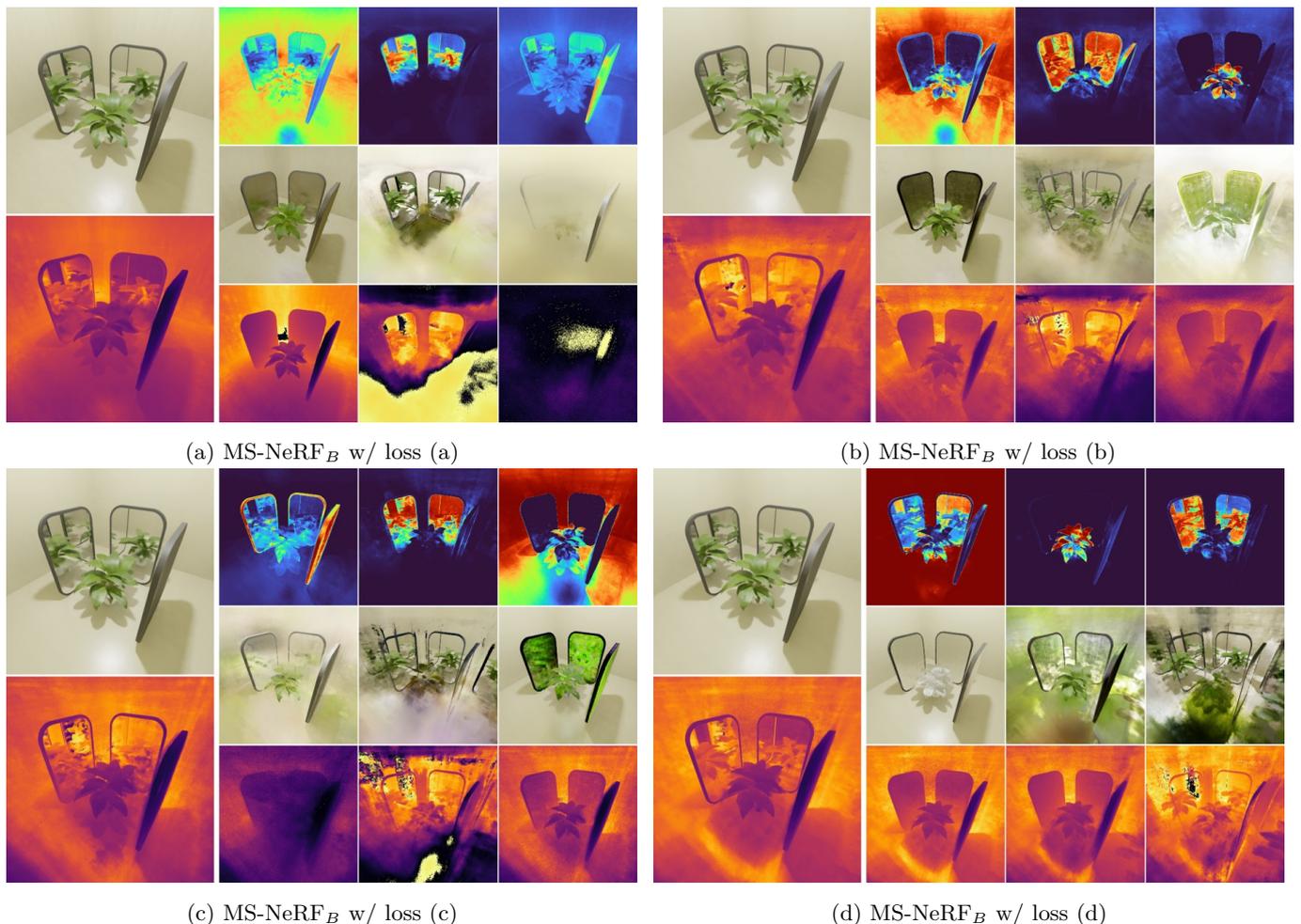


图 23: 不同损失监督下 MS-NeRF_B 子空间分解的可视化。左图展示最终渲染的 RGB 图和深度图, 右图展示各子空间的权重图、RGB 图和深度图。

参考文献

- [1] Z.-X. Yin, J. Qiu, M.-M. Cheng, and B. Ren, "Multi-space neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 12407–12416. 1
- [2] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021. 1, 2, 3, 5, 8, 9, 11
- [3] K. Zhang, G. Riegler, N. Snavely, and V. Koltun, "Nerf++: Analyzing and improving neural radiance fields," *arXiv:2010.07492*, 2020. 1, 2
- [4] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Mip-nerf 360: Unbounded anti-aliased neural radiance fields," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 5470–5479. 1, 2, 5, 8, 9, 10
- [5] A. Pumarola, E. Corona, G. Pons-Moll, and F. Moreno-Noguer, "D-nerf: Neural radiance fields for dynamic scenes," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2021, pp. 10318–10327. 1
- [6] K. Park, U. Sinha, J. T. Barron, S. Bouaziz, D. B. Goldman, S. M. Seitz, and R. Martin-Brualla, "Nerfies: Deformable neural radiance fields," in *Int. Conf. Comput. Vis.*, 2021, pp. 5865–5874. 1
- [7] E. Tretschk, A. Tewari, V. Golyanik, M. Zollhöfer, C. Lassner, and C. Theobalt, "Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video," in *Int. Conf. Comput. Vis.*, 2021, pp. 12959–12970. 1
- [8] R. Martin-Brualla, N. Radwan, M. S. Sajjadi, J. T. Barron, A. Dosovitskiy, and D. Duckworth, "Nerf in the wild: Neural radiance fields for unconstrained photo collections," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2021, pp. 7210–7219. 1
- [9] X. Chen, Q. Zhang, X. Li, Y. Chen, Y. Feng, X. Wang, and J. Wang, "Hallucinated neural radiance fields in the wild," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 12943–12952. 1
- [10] J. Zhang, G. Yang, S. Tulsiani, and D. Ramanan, "Ners: Neural reflectance surfaces for sparse-view 3d reconstruction in the wild," *Advances in Neural Information Processing Systems*, vol. 34, pp. 29835–29847, 2021. 1
- [11] J. Sun, X. Chen, Q. Wang, Z. Li, H. Averbuch-Elor, X. Zhou, and N. Snavely, "Neural 3d reconstruction in the wild," in *ACM SIGGRAPH 2022 Conference Proceedings*, 2022, pp. 1–9. 1
- [12] N. Jain, S. Kumar, and L. Van Gool, "Robustifying the multi-scale representation of neural radiance fields," *arXiv:2210.04233*, 2022. 1, 3
- [13] C.-H. Lin, W.-C. Ma, A. Torralba, and S. Lucey, "Barf: Bundle-adjusting neural radiance fields," in *Int. Conf. Comput. Vis.*, 2021, pp. 5741–5751. 1, 3
- [14] A. Tewari, J. Thies, B. Mildenhall, P. Srinivasan, E. Tretschk,

- W. Yifan, C. Lassner, V. Sitzmann, R. Martin-Brualla, S. Lombardi *et al.*, “Advances in neural rendering,” *Computer Graphics Forum*, vol. 41, no. 2, pp. 703–735, 2022. 1
- [15] W. Bian, Z. Wang, K. Li, J.-W. Bian, and V. A. Prisacariu, “Nope-nerf: Optimising neural radiance field with no pose prior,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 4160–4169. 1
- [16] Y.-C. Guo, D. Kang, L. Bao, Y. He, and S.-H. Zhang, “Nerfren: Neural radiance fields with reflections,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 18 409–18 418. 1, 2, 3, 7, 8, 9, 10
- [17] S. Wizadwongsa, P. Phongthawee, J. Yenphraphai, and S. Suwajanakorn, “Nex: Real-time view synthesis with neural basis expansion,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2021, pp. 8534–8543. 2, 9
- [18] B. Mildenhall, P. P. Srinivasan, R. Ortiz-Cayon, N. K. Kalantari, R. Ramamoorthi, R. Ng, and A. Kar, “Local light field fusion: Practical view synthesis with prescriptive sampling guidelines,” *ACM Trans. Graph.*, vol. 38, no. 4, pp. 1–14, 2019. 2, 9
- [19] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, “Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields,” in *Int. Conf. Comput. Vis.*, 2021, pp. 5855–5864. 2, 8, 9
- [20] A. Chen, Z. Xu, A. Geiger, J. Yu, and H. Su, “Tensorf: Tensorial radiance fields,” in *European Conference on Computer Vision (ECCV)*, 2022. 2, 8, 10
- [21] T. Müller, A. Evans, C. Schied, and A. Keller, “Instant neural graphics primitives with a multiresolution hash encoding,” *ACM Trans. Graph.*, vol. 41, no. 4, jul 2022. 2, 5, 6, 8, 10
- [22] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, “Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 27 171–27 183, 2021. 2, 12
- [23] M. Oechsle, S. Peng, and A. Geiger, “Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction,” in *Int. Conf. Comput. Vis.*, 2021, pp. 5589–5599. 2
- [24] L. Yariv, J. Gu, Y. Kasten, and Y. Lipman, “Volume rendering of neural implicit surfaces,” in *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021. 2
- [25] Z. Yu, A. Chen, B. Antic, S. P. Peng, A. Bhattacharyya, M. Niemeyer, S. Tang, T. Sattler, and A. Geiger, “Sdfstudio: A unified framework for surface reconstruction,” 2022. [Online]. Available: <https://github.com/autonomousvision/sdfstudio> 2
- [26] L. Yariv, P. Hedman, C. Reiser, D. Verbin, P. P. Srinivasan, R. Szeliski, J. T. Barron, and B. Mildenhall, “Baked sdf: Meshing neural sdfs for real-time view synthesis,” *arXiv*, 2023. 2
- [27] Z. Li, T. Müller, A. Evans, R. H. Taylor, M. Unberath, M.-Y. Liu, and C.-H. Lin, “Neuralangelo: High-fidelity neural surface reconstruction,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2
- [28] J. Qiu, P.-T. Jiang, Y. Zhu, Z.-X. Yin, M.-M. Cheng, and B. Ren, “Looking through the glass: Neural surface reconstruction against high specular reflections,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 20 823–20 833. 2
- [29] E. R. Chan, C. Z. Lin, M. A. Chan, K. Nagano, B. Pan, S. De Mello, O. Gallo, L. J. Guibas, J. Tremblay, S. Khamis *et al.*, “Efficient geometry-aware 3d generative adversarial networks,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 16 123–16 133. 2
- [30] M. Niemeyer and A. Geiger, “Giraffe: Representing scenes as compositional generative neural feature fields,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2021, pp. 11 453–11 464. 2, 5
- [31] A. Jain, B. Mildenhall, J. T. Barron, P. Abbeel, and B. Poole, “Zero-shot text-guided object generation with dream fields,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 867–876. 2
- [32] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, “Dreamfusion: Text-to-3d using 2d diffusion,” *arXiv preprint arXiv:2209.14988*, 2022. 2
- [33] Z. Wang, C. Lu, Y. Wang, F. Bao, C. Li, H. Su, and J. Zhu, “Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation,” *arXiv preprint arXiv:2305.16213*, 2023. 2
- [34] C. Wang, M. Chai, M. He, D. Chen, and J. Liao, “Clip-nerf: Text-and-image driven manipulation of neural radiance fields,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 3835–3844. 2
- [35] Y.-J. Yuan, Y.-T. Sun, Y.-K. Lai, Y. Ma, R. Jia, and L. Gao, “Nerf-editing: geometry editing of neural radiance fields,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 18 353–18 364. 2
- [36] A. Haque, M. Tancik, A. Efros, A. Holynski, and A. Kanazawa, “Instruct-nerf2nerf: Editing 3d scenes with instructions,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023. 2
- [37] Y. Zheng, V. F. Abrevaya, M. C. Bühler, X. Chen, M. J. Black, and O. Hilliges, “Im avatar: Implicit morphable head avatars from videos,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 13 545–13 555. 2
- [38] B. Jiang, Y. Hong, H. Bao, and J. Zhang, “Selfrecon: Self reconstruction your digital avatar from monocular video,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 5605–5615. 2
- [39] Y. Feng, J. Yang, M. Pollefeys, M. J. Black, and T. Bolkart, “Capturing and animation of body and clothing from monocular video,” *arXiv:2210.01868*, 2022. 2
- [40] Y. Cao, Y.-P. Cao, K. Han, Y. Shan, and K.-Y. K. Wong, “Dreamavatar: Text-and-shape guided 3d human avatar generation via diffusion models,” 2023. 2
- [41] B. Mildenhall, P. Hedman, R. Martin-Brualla, P. P. Srinivasan, and J. T. Barron, “Nerf in the dark: High dynamic range view synthesis from noisy raw images,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 16 190–16 199. 2
- [42] X. Huang, Q. Zhang, Y. Feng, H. Li, X. Wang, and Q. Wang, “Hdr-nerf: High dynamic range neural radiance fields,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 18 398–18 408. 2
- [43] A. Yu, R. Li, M. Tancik, H. Li, R. Ng, and A. Kanazawa, “Plenotrees for real-time rendering of neural radiance fields,” in *Int. Conf. Comput. Vis.*, 2021, pp. 5752–5761. 2
- [44] S. J. Garbin, M. Kowalski, M. Johnson, J. Shotton, and J. Valentin, “Fastnerf: High-fidelity neural rendering at 200fps,” in *Int. Conf. Comput. Vis.*, 2021, pp. 14 346–14 355. 2
- [45] C. Reiser, S. Peng, Y. Liao, and A. Geiger, “Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps,” in *Int. Conf. Comput. Vis.*, 2021, pp. 14 335–14 345. 2
- [46] C. Sun, M. Sun, and H.-T. Chen, “Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022, pp. 5459–5469. 2

- [47] Z. Chen, T. Funkhouser, P. Hedman, and A. Tagliaschi, "Mobilerf: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures," *arXiv:2208.00277*, 2022. **2**
- [48] D. Verbin, P. Hedman, B. Mildenhall, T. Zickler, J. T. Barron, and P. P. Srinivasan, "Ref-NeRF: Structured view-dependent appearance for neural radiance fields," *CVPR*, 2022. **2, 9, 10, 11**
- [49] R. Ramamoorthi *et al.*, "Precomputation-based rendering," *Foundations and Trends® in Computer Graphics and Vision*, vol. 3, no. 4, pp. 281–369, 2009. **2**
- [50] J. Zeng, C. Bao, R. Chen, Z. Dong, G. Zhang, H. Bao, and Z. Cui, "Mirror-nerf: Learning neural radiance fields for mirrors with whitted-style ray tracing," in *Proceedings of the 31th ACM International Conference on Multimedia*, 2023. **2**
- [51] G.-W. Yang, W.-Y. Zhou, H.-Y. Peng, D. Liang, T.-J. Mu, and S.-M. Hu, "Recursive-nerf: An efficient and dynamically growing nerf," *IEEE Trans. Visual. Comput. Graph.*, 2022. **2**
- [52] B. Yang, Y. Zhang, Y. Xu, Y. Li, H. Zhou, H. Bao, G. Zhang, and Z. Cui, "Learning object-compositional neural radiance field for editable scene rendering," in *Int. Conf. Comput. Vis.*, 2021, pp. 13 779–13 788. **2**
- [53] M. Guo, A. Fathi, J. Wu, and T. Funkhouser, "Object-centric neural scene rendering," *arXiv:2012.08503*, 2020. **2**
- [54] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 4104–4113. **2**
- [55] A. Knapitsch, J. Park, Q.-Y. Zhou, and V. Koltun, "Tanks and temples: Benchmarking large-scale scene reconstruction," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–13, 2017. **2, 9**
- [56] K. Yücer, A. Sorkine-Hornung, O. Wang, and O. Sorkine-Hornung, "Efficient 3d object segmentation from densely sampled light fields with applications to 3d reconstruction," *ACM Trans. Graph.*, vol. 35, no. 3, pp. 1–15, 2016. **3**
- [57] M. Berman, K. Myszkowski, J. Reval Frisvad, H.-P. Seidel, and T. Ritschel, "Eikonal fields for refractive novel-view synthesis," in *ACM SIGGRAPH 2022 Conference Proceedings*, 2022, pp. 1–9. **3, 7, 9**
- [58] R. Jensen, A. Dahl, G. Vogiatzis, E. Tola, and H. Aanæs, "Large scale multi-view stereopsis evaluation," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2014, pp. 406–413. **3, 9**
- [59] Y. Yao, Z. Luo, S. Li, J. Zhang, Y. Ren, L. Zhou, T. Fang, and L. Quan, "Blendedmvs: A large-scale dataset for generalized multi-view stereo networks," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020, pp. 1790–1799. **3, 9**
- [60] N. Max, "Optical models for direct volume rendering," *IEEE Trans. Visual. Comput. Graph.*, vol. 1, no. 2, pp. 99–108, 1995. **3**
- [61] B. O. Community, *Blender - a 3D modelling and rendering package*, Blender Foundation, Stichting Blender Foundation, Amsterdam, 2022. [Online]. Available: <http://www.blender.org> **3, 8**
- [62] R. Li, H. Gao, M. Tancik, and A. Kanazawa, "Nerfacc: Efficient sampling accelerates nerfs," *arXiv preprint arXiv:2305.04966*, 2023. **6, 10, 11**
- [63] P. Hedman, P. P. Srinivasan, B. Mildenhall, J. T. Barron, and P. Debevec, "Baking neural radiance fields for real-time view synthesis," in *Int. Conf. Comput. Vis.*, 2021, pp. 5875–5884. **9**
- [64] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004. **9**
- [65] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 586–595. **9**
- [66] Y.-C. Guo, "Instant neural surface reconstruction," 2022, <https://github.com/bennyguo/instant-nsr-pl>. **12**



尹泽鑫 硕士毕业于南开大学，导师为程明明教授与任博副教授。现为南开大学计算机学院博士研究生，师从谢晋教授。其研究兴趣主要包括神经辐射场与三维计算机视觉。



焦鹏屹 于 2019 年获得北京理工大学学士学位。现为南开大学计算机学院硕士研究生。研究方向包括神经辐射场、计算机图形学与计算机视觉。



邱佳雄 于 2017 年获得大连海事大学学士学位，2020 年获得电子科技大学硕士学位，2024 年获得南开大学计算机学院博士学位。研究方向包括计算机视觉、计算机图形学、机器人与深度学习。



程明明 于 2012 年获得清华大学博士学位，随后在牛津大学担任两年研究员，合作导师为 Philip Torr 教授。现为南开大学教授，领导媒体计算实验室。其研究兴趣包括计算机图形学、计算机视觉与图像处理。他曾获得多项科研奖项，包括国家杰出青年科学基金和 ACM 中国新星奖。他现任 IEEE TPAMI 和 IEEE TIP 的编委。



任博 分别于 2010 年和 2015 年获得清华大学学士与博士学位。现为南开大学计算机学院副教授。研究方向包括计算机图形学、计算机视觉与人工智能。当前研究主要聚焦于基于学习或物理机制的模拟、三维场景几何重建与分析。