

# 透过玻璃看：针对高强度镜面反射的神经表面重建\*

邱佳雄<sup>1</sup> 姜鹏涛<sup>2</sup> 朱一凡<sup>1</sup> 尹泽鑫<sup>1</sup> 程明明<sup>1</sup> 任博<sup>1†</sup>

<sup>1</sup>VCIP, CS, Nankai University <sup>2</sup>Zhejiang University

## 摘要

隐式神经方法已经在弱镜面高光下实现了高质量的 3D 物体表面重建。然而，当我们透过玻璃拍摄目标物体时，高强度镜面反射 (HSR) 经常出现在目标物体的前方。这些场景中复杂的歧义性违反了多视角一致性，从而使最近的方法难以正确重建目标物体。为了解决这个问题，我们提出了一种基于隐式神经渲染的新型表面重建框架 *NeuS-HSR*。在 *NeuS-HSR* 中，物体表面被参数化为隐式符号距离函数 (SDF)。为了减少 HSR 的干扰，我们提出将渲染图像分解为两种外观：目标物体和辅助平面。我们设计了一个全新的辅助平面模块，通过结合物理假设和神经网络来生成辅助平面外观。在合成和真实世界数据集上进行的大量实验表明，*NeuS-HSR* 在针对 HSR 进行精确和鲁棒的目标表面重建方面优于最先进的方法。源代码已开源于 <https://github.com/JiaxiongQ/NeuS-HSR>。

## 1. 引言

使用多视角图像重建三维物体表面是计算机视觉和图形学中一项具有挑战性的任务。近期 NeuS [47] 结合了表面渲染 [3, 12, 37, 55] 和体渲染 [8, 30]，用于重建具有细小结构的物体，并在具有轻微镜面反射的输入上实现了良好的性能。然而，当处理高强度镜面反射 (HSR) 场景时，NeuS 无法恢复目标物体表面，如图1第二行所示。当我们使用相机透过玻璃拍摄目标物体时，高强度镜面反射无

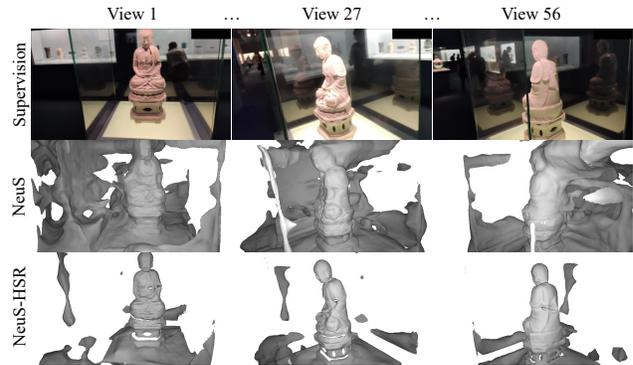


图 1. 高强度镜面反射 (HSR) 下的三维物体表面重建。顶部：手机拍摄的真实世界场景。中间：最先进的方法 NeuS [47] 未能成功重建目标物体 (即佛像)。底部：我们提出的 NeuS-HSR，与 NeuS 相比，它能更精确地重建目标物体表面。

处不在。如图1第一行所示，在拍摄的 HSR 视图中，我们可以看到目标物体前方的虚像。虚像在视觉上引入了物体表面的光度变化，从而降低了多视角一致性并为渲染编码了严重的歧义性，这会误导 NeuS 去重建反射物体而不是目标物体。

为了适应 HSR 场景，一种直观的解决方案是先使用反射去除方法来减少 HSR，然后以增强后的目标物体外观作为监督来重建目标物体。然而，最近的单图像反射去除方法 [4, 9, 23, 24, 27, 42] 需要真实的背景或反射作为监督，而这是很难获得的。此外，这些反射去除方法的测试场景需要存在于训练集中，这限制了它们的泛化。事实表明，显式地使用反射去除方法来增强目标物体的表面是不切实际的。最近一种无监督反射去除方法 NeRFReN [18]，通过隐式表示将渲染图像分解为反射部分和透射部分。然而，它受到视角方向和简单平面反射器的限制。当我们将应用于多视角重建场景时，正如图3所指出的，它会将目标物体作为反射图像中的内容，无法

\*本文为 CVPR' 23 论文 [34] 的中文翻译版。

†任博是通讯作者。

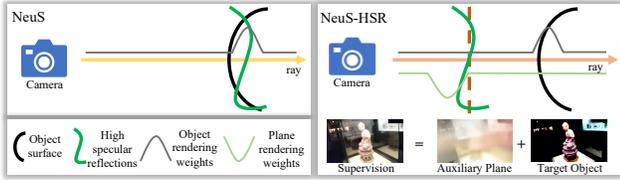


图 2. NeuS-HSR。高强度镜面反射 (HSR) 使 NeuS 倾向于重建在 HSR 中的反射物体。NeuS-HSR 将渲染图像物理解析为目标物体和辅助平面部分，从而使得 NeuS 将注意力集中在目标物体上。

生成正确的透射图像用于目标物体恢复。

如上所述，两阶段的直观解决方案在我们的任务中难以实现。为了解决这个问题，我们考虑了一种比 NeRFReN 更有效的分解策略用以增强目标物体的外观，从而在一个阶段内实现精确的表面重建。为了实现我们的目标，我们构建了以下假设：

**假设 1.** 受 HSR 影响的场景可分解为目标物体和平面反射器两个部分。除目标物体外，HSR 和视角中的大多数其他内容都是通过平面反射器 (即玻璃) 反射和传输。

**假设 2.** 平面反射器与相机视角方向相交，因为所有视角方向矢量一般都指向目标物体并通过平面反射器。

基于上述物理假设，我们提出了 NeuS-HSR——一个全新的物体重建框架，用于从一组 HSR 场景的 RGB 图像中恢复目标物体表面。对于假设 1，如图 2 所示，我们设计了一个辅助平面来表示平面反射器，因为我们的目标是通过它来增强目标物体的外观。借助辅助平面，我们将目标物体和辅助平面部分从监督中准确地分离出来。对于目标物体部分，我们使用 NeuS [47] 中的方法来生成目标物体外观。对于辅助平面部分，我们设计了一个辅助平面模块，以视角方向作为假设 2 的输入，利用神经网络生成依赖于视角的辅助平面的属性 (包括法线和位置)。当辅助平面确定后，我们基于反射变换 [16] 和神经网络获取辅助平面的外观。最后，我们融合两种表面信息，然后得到渲染图像，该图像由拍摄的图片监督并用于单阶段训练。



图 3. NeRFReN 的分解 [18]。NeRFReN 无法分离镜面反射和目标物体在该视角中的外观，从而导致 NeuS 无法成功恢复目标物体表面。

我们进行了一系列实验来评估 NeuS-HSR。实验结果表明，NeuS-HSR 在合成数据集上优于其他 state-of-the-art，并能从真实世界场景中的 HSR 图像中恢复出高质量的目标物体。

综上，我们的主要贡献如下：

- 我们提出通过分离场景中的目标物体和辅助平面部分来恢复受 HSR 影响的目标物体表面。
- 我们设计了一个辅助平面模块来生成辅助平面部分的物理外观，以增强目标物体部分的外观。
- 在合成场景和真实场景上进行的大量实验证明，我们的方法在定量和定性上都比其他 state-of-the-art 方法重建的目标物体更加精确。

## 2. 相关工作

### 2.1. 传统表面重建

经典的多视角曲面重建方法主要包括两类：光度立体重建 [5,6,19,53] 和多视角立体重建 [11,13–15,38–40]。光度立体重建受限于严格的实验环境。对于多视角立体重建，输入图像是在普通场景中围绕目标物体采集的。早期的多视角立体方法 [11,15,38,39] 主要针对具有漫反射材料的物体表面。它们都遵从 Lambertian 假设，即物体表面的同一检测区域在所有视角中变化不大。然而，在现实世界的场景中，物体经常出现明显的镜面反射，例如高光。在有明显镜面反射的真实世界场景中，Lambertian 假设不再成立。广泛使用的运动恢复结构 (SfM) 方法 [36,44,51] 旨在标定摄像机，并在每个视点生成稀疏深度图。然后通过泊松重建 [22] 与深度融合获得物体表面。然而，输出表面的质量很容易受到特征点检测的影响，目标物体上没有丰富纹理的表面区域总是有伪影或

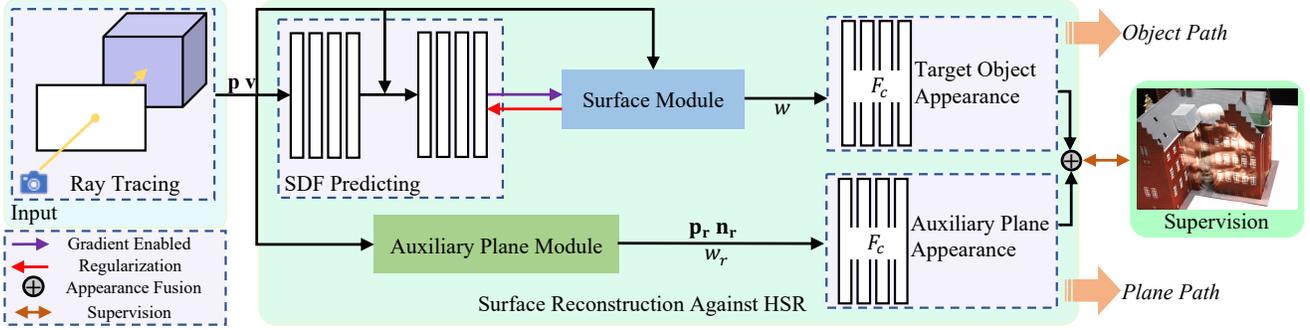


图 4. NeuS-HSR 的流程。采样点  $p$  和视角方向  $v$  分别输入目标物体路径和平面路径。在目标路径中，隐式 SDF  $f$  由位于首部的神经网络生成。表面模块将  $f$ 、 $p$  和  $v$  作为输入，产生渲染权重  $w$ 。在平面路径中，反射模块根据  $p$  和  $v$  生成辅助平面的平面法线  $n_r$ 、三维位置  $p_r$  和渲染权重  $w_r$ 。最后，我们通过外观函数  $F_c$  和体渲染获得两种外观。

空洞。在这项工作中，我们重点研究隐式神经表示方法，以便在更真实的场景中（即 non-Lambertian 表面）获得精确的三维物体表面。

## 2.2. 隐式神经表面渲染

基于神经网络的隐式表示在新视角合成 [26, 28, 30, 41, 49] 和三维重建 [7, 10, 31, 32, 35, 43, 46–48, 50, 54, 55] 方面取得了令人兴奋的成果。它们具有传统方法所不具备的特点，包括灵活的分辨率和自然的全局一致性。基于可微分光线投射的表面渲染应用于不同形式的隐含形状表示的表面重建，如占位函数 [33] 和符号距离函数 (SDF) [55]。IDR [55] 利用 SDF 表示的零水平集提取物体表面的点，并利用神经网络梯度求解可微分渲染公式。UNISURF [32]、VolSDF [54] 和 NeuS [47] 通过引入体渲染方案 [29] 来学习隐式表面函数，以提高从拍摄的图像中重建表面的质量。NeuralWarp [7] 是一种用于完善基本模型（如 VolSDF）的两阶段方法。NeRS [56] 主要通过引入 Phong 模型 [20, 21, 45] 来学习物体表面的外观。它使用一个标准球来表示物体表面，并通过稀疏的图像集学习带有先决掩模的物体纹理，但它主要处理具有反射表面的物体，并生成没有细节的物体表面。与这些工作不同的是，我们提出在一个阶段内将物体表面重建扩展到更具挑战性的 HSR 场景。我们的目标是通过玻璃而不是反射表面来正确恢复物体表面。我们的方法在 HSR 场景中的重建精度和鲁棒性远远优于之前的工作。

## 3. 方法

在这项工作中，我们专注于重建高强度镜面反射 (HSR) 场景中的物体表面。如引言部分所述，HSR 编码非目标物体信息，导致目标物体表面质量较低。针对 HSR 场景，我们引入了一种基于隐式神经渲染的新型物体表面重建方法 NeuS-HSR。NeuS-HSR 的流程如图 4 所示。

具体来说，我们将 HSR 场景分解为两个部分：目标物体和辅助平面。为了渲染目标物体的外观，我们采用了 NeuS 的方案，并将其封装为一个表面模块。为了渲染辅助平面的外观，我们设计了一个基于反射变换 [16] 和多层感知器 (MLPs) 的辅助平面模块。最后，我们应用线性求和来融合两种外观，得到渲染图像，渲染图像接受视角中拍摄图像的监督。下面我们将分三个部分介绍 NeuS-HSR，包括表面模块 (3.1 节)、辅助平面模块 (3.2 节) 和渲染过程 (3.3 节)。

### 3.1. 表面模块

我们使用 NeuS [47] 来渲染目标物体表面。具体来说，NeuS 在每条相机射线  $h_s$  上基于隐式 SDF  $f: \mathbb{R}^3 \rightarrow \mathbb{R}$  建立了一个无偏且可感知遮挡的权重函数  $w$ 。首先， $w$  的定义为：

$$w(t) = T(t)\rho(t), T(t) = \exp\left(-\int_0^t \rho(u)du\right). \quad (1)$$

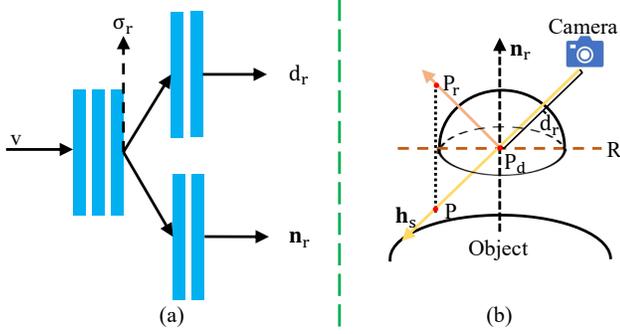


图 5. 辅助平面模块。(a): 我们设计了一种全新的神经网络  $F_r$  用于从输入视角方向  $v$  预测相机射线  $h_s$  上辅助平面  $R$  的体积密度  $\sigma_r$ , 位置  $d_r$  和表面法线  $n_r$ 。(b): 确定辅助平面后, 我们通过反射变换将  $R$  后面的采样点  $P$  投影到其反射点  $P_r$  上 [16]。

其中  $t \in \mathbb{R}$  是沿  $h_s$  的深度值, 接着  $\rho(t)$  被构造为:

$$\rho(t) = \max\left(\frac{-\frac{d\Theta_s}{dt}(f(p(t)))}{\Theta_s(f(p(t)))}, 0\right). \quad (2)$$

其中物体表面  $S$  可以由点  $p$ :  $S = \{p \in \mathbb{R}^3 | f(p) = 0\}$  的零级符号距离集合建模。Logistic 密度分布  $\theta_S(p) = se^{-sp}/(1 + e^{-sp})^2$ , 是 Sigmoid 函数  $\Theta_S(p) = (1 + e^{-sp})^{-1}$  的导数。  $1/s$  是  $\theta_S(p)$  的标准偏差。  $w$  的构建是 NeuS 的关键贡献。它将隐式 SDF 和体渲染正确地连接起来, 以处理复杂的物体结构。相机射线  $h_s$  在点  $p$  处可以表示为:  $h_s(t) = o + tv$ , 其中  $o$  和  $v$  分别表示相机中心和视角方向。我们沿着  $h_s$  采样  $m$  个点, 然后像素颜色值  $C$  由以下公式获取:

$$C = \sum_{i=1}^m w_i c_i. \quad (3)$$

其中  $w_i = \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j)$ ,  $c_i$  是从 MLPs 中学习颜色,  $\alpha_i$  是式 (2) 的离散化表示。

在我们的任务中,  $C$  对 HSR 的内容进行编码, 这会导致  $w$  的歧义。这种歧义性使得 SDF 预测的 MLPs 倾向于模拟 HSR 的内容, 从而在目标物体表面周围产生过多的噪声。因此, 我们提出了一个辅助平面模块, 将 MLPs 的注意力转移到目标物体上, 以处理 HSR 带来的干扰。

### 3.2. 辅助平面模块

在 HSR 场景中, 出现在平面反射器上和目标物体前方的虚像为目标物体的重建提供了具有严重歧义的信息。在没有先验信息的情况下分解渲染图像

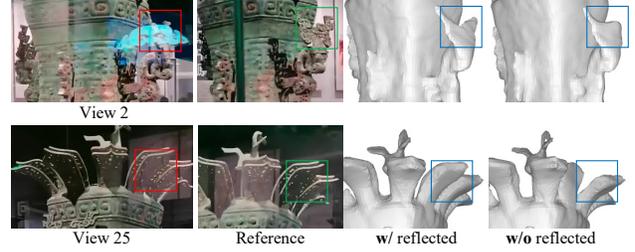


图 6. 将采样点投影到反射点的效果。红色方框展示了 HSR 区域。绿色方框表示与红色方框相关的参考区域。蓝色方框表示 HSR 对物体表面重建的影响。

以增强目标物体的外观是一个难题。NeRFReN [18] 应用深度平滑先验和双向深度一致性约束, 通过隐式表示将渲染图像分成两部分: 透射图像和反射图像。该方案在有限视角方向和简单平面反射体的场景中效果良好。然而, 在 HSR 场景中, 它无法在透射图像中保留目标物体。受 NeRFReN 的启发, 我们提出了一个辅助平面模块来增强目标物体在渲染图像中的外观。

实际上, 我们使用辅助平面  $R$  来表示每条相机射线的实际平面反射器。为了确定  $R$  的物理值, 我们设计了一个新的神经网络  $F_r: \mathbb{S}^2 \rightarrow \mathbb{R} \times \mathbb{R} \times \mathbb{R}^3$ , 如图 5 (a) 所示。  $F_r$  将视角方向  $v$  映射到体积密度  $\sigma_r$  上用于生成渲染权重, 和  $R$  的属性 (包括位置  $d_r$  和平面法线  $n_r$ ), 即

$$\{\sigma_r, d_r, n_r\} = F_r(v). \quad (4)$$

我们假设  $R$  内置在相机坐标系中, 3D 点  $p_d = d_r v$  是  $R$  和  $h_s$  的相互作用。显然  $p_d$  在  $R$  上。给定  $n_r = [A, B, C]$ ,  $R$  可以被定义为:

$$Ax + By + Cz + D = 0. \quad (5)$$

其中  $A^2 + B^2 + C^2 = 1$ 。我们把  $p_d$  代入到式 (5) 中就有了  $D = -d_r n_r \cdot v$ 。

此外, 沿相机射线的采样点也是 MLPs 获取颜色值的部分输入。为了进一步建立 HSR 物理模型, 如图 5 (b) 所示, 对于沿  $h_s$  和  $R$  后方采样的点  $p$ , 我们根据反射变换 [16] 将其投影到沿入射光路径的反射点  $p_r$ 。然后, MLP 可以隐式地跟踪入射光来渲染 HSR。图 6 展示了该操作在减少 HSR 干扰方面的有效性。反射点帮助 MLP 对 HSR 进行物理建模, 从而减少场景的歧义性, 恢复出更精确的目标物体表面。

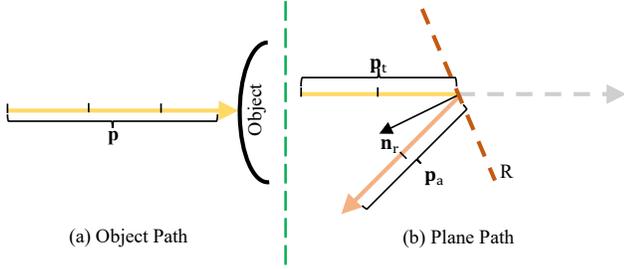


图 7. 沿相机射线的空间采样点用于获取两条路径的颜色值。对于物体路径，我们使用原始采样点  $p$ 。  $p'$  是相机坐标系中的  $p$ 。对于平面路径，我们使用  $p'$  的沿相机射线位于  $R$  前面的部分点  $p_t$  和反射点  $p_a$ 。

投影算法的具体细节请参见补充材料。

### 3.3. 渲染

我们采用神经网络  $F_c$  分别预测物体路径  $c_t$  和平面路径  $c_a$  的颜色值。每条路径的输入都不同。对于物体路径，我们采用 NeuS 方法，将沿相机射线的采样点  $p$ 、目标物体的表面法线  $n$ 、视角方向  $v$  和隐式 SDF 的特征  $f_p$  作为输入。因此，  $c_t = F_c(p, n, v, f_p)$ 。对于平面路径，相机坐标系中的采样点为： $p' = p - o$ 。如图7所示，我们利用  $p'$  在  $R$  前部分点  $p_t$  和反射点  $p_a$  作为输入点  $p_r = p_t \cup p_a$ 。利用平面法线  $n_r$  作为输入法线。对于平面路径，有  $c_r = F_c(p_r, n_r, v, f_p)$ 。

为了生成每个路径的渲染外观，我们还需要构造两种渲染权重。对于物体路径，我们遵循 NeuS 的方案来生成 3.1 中定义的权重  $w$ 。对于辅助平面路径，给定从平面网络  $F_r$  中学习到的体积密度  $\sigma_r$ ，我们采用 NeRFReN 的方案生成权重  $w_r$ ，其计算公式为：

$$w_r^i = \exp\left(-\sum_{j=1}^{i-1} \sigma_r^j \delta_j\right) (1 - \exp(-\sigma_r^i \delta_i)). \quad (6)$$

其中  $\delta_i = t_{i+1} - t_i$ 。最后，可以通过式 (3) 生成目标物体外观  $C_t(w, c_t)$  和辅助平面外观  $C_r(w_r, c_r)$ 。最终渲染出的图像  $C$  由  $C_t$  和  $C_r$  的线性组合获得，其公式为：

$$C = \varphi_1 C_t + \varphi_2 C_r. \quad (7)$$

其中  $\varphi_1 + \varphi_2 = 1$ 。在实验中，我们默认设置  $\varphi_1 = 0.3$  和  $\varphi_2 = 0.7$ 。补充材料中说明了此设置的详细信息。

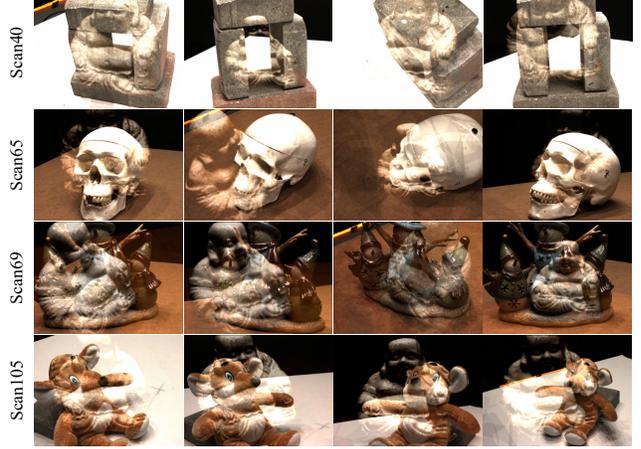


图 8. 合成数据集的示例。我们应用被广泛使用的方法 [58] 在 DTU 数据集 [1] 上合成 HSR 场景。

### 3.4. 损失函数

在 NeuS-HSR 的训练过程中，我们优化了渲染图像  $C$  和拍摄图像  $\tilde{C}$  之间的差异。我们遵循 NeuS 中定义的损失函数，它由三个项组成：颜色损失  $\mathcal{L}_c$  [47, 55]，隐式 SDF 的正则化损失  $\mathcal{L}_r$  [17] 和平面法线的  $\mathcal{L}_n$ 。损失函数公式如下：

$$\begin{cases} \mathcal{L}_c = \frac{1}{b} \sum_i \mathcal{L}_1(C_i, \tilde{C}_i), \\ \mathcal{L}_r = \frac{1}{bm} \sum_{k,i} (|\nabla f(p_k^i)| - 1)^2, \\ \mathcal{L}_n = \frac{1}{b} \sum_i (|n_r^i| - 1)^2, \end{cases} \quad (8)$$

其中  $b$  表示批量大小， $m$  表示沿相机射线采样点的数量。那么最终的损失函数可以定义为：

$$\mathcal{L} = \mathcal{L}_c + \lambda_1 (\mathcal{L}_r + \mathcal{L}_n). \quad (9)$$

其中  $\lambda_1$  是一个常量。实验中，我们默认设置  $\lambda_1 = 0.1$ 。

## 4. 实验

我们进行了大量实验，以表明我们的方法在定量 (表1) 和定性 (图9, 图10) 方面优于其他方法。我们还提供了几个消融实验以揭示我们算法中一些设计选择的必要性 ((图11))。

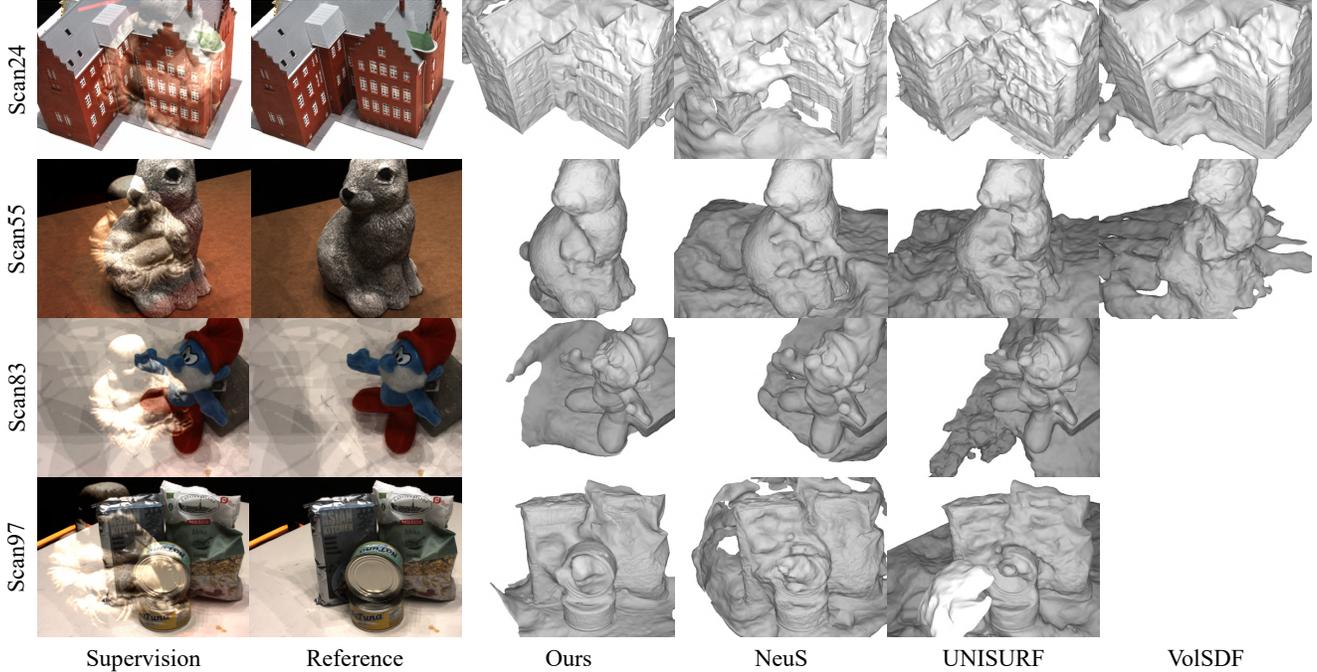


图 9. 对合成数据集的定性比较。“参考”包含目标物体的原始外观。VoISDF [54] 无法恢复“Scan83”和“Scan97”上有意义的物体表面。与其他方法相比，我们的方法实现了更好的物体表面质量。

#### 4.1. 数据集

**合成数据集。**为了定量评估 NeuS-HSR 和其他方法的性能，我们合成了 DTU 数据集集中的 10 个场景 [1]。我们遵循常见的单图像反射合成方法 [58] 来生成合成数据集。给定透射图像  $T$ （即包含目标物体的图像）和反射图像  $R'$ ，具有反射的图像  $I$  可以定义为：

$$I = T + \mathcal{K} \otimes R'. \quad (10)$$

$\mathcal{K}$  是高斯核， $\otimes$  表示卷积运算。我们随机选择一个场景作为反射部分，其他场景设置为透射部分。然后我们采用式 (??) 来获取 HSR 场景。合成数据集的示例如图 8 所示。

**真实世界数据集。**为了验证我们的方法在真实场景中的有效性，我们从互联网上收集了 6 个 HSR 场景。我们利用工具 COLMAP [38] 来估计相机参数。

#### 4.2. 实验设置

**实施细节。**符号距离函数 (SDF)  $f$  由 MLP 参数化，MLP 由 8 个线性层组成。然后通过 marching cubes [29] 从隐式 SDF 生成目标物体表面。辅助平面

函数  $F_r$  由用于预测体积密度的 3 层 MLP 和用于预测平面属性的 2 层 MLP 组成。渲染外观函数  $F_c$  由 4 层 MLP 建模。所有空间点都在单位球体内采样，其中外部场景由 NeRF++ 生成 [57]。采用位置编码 [30] 沿相机射线和视角方向  $v$  采样点  $p$ 。近似 SDF 通过几何初始化 [2] 进行预处理。射线的批大小设置为 512。我们训练 NeuS-HSR 进行了 200k 次迭代，在单个 NVIDIA Tesla V100 GPU 上耗时约 12 小时。

**方法比较。**我们将我们的方法与其他具有公平设置的相关方法进行比较。相关的方法包括 (i) state-of-the-art 隐式神经表面重建方法：NeuS [47]，VoISDF [54] 和 UNISURF [32]，(ii) 经典的多视角立体视觉方法：COLMAP [38]。对于 COLMAP，我们应用屏蔽泊松 [22] 从估计的点云重建其稠密网格。本文中所有基于学习的模型都是在没有真实掩码的情况下训练的。

#### 4.3. 定量比较

对于定量评估，我们在合成数据集上进行了比较。和 [32, 47, 54] 一样，我们使用倒角距离作为评估指标，它表示目标物体的重建质量。我们在表 1 中报

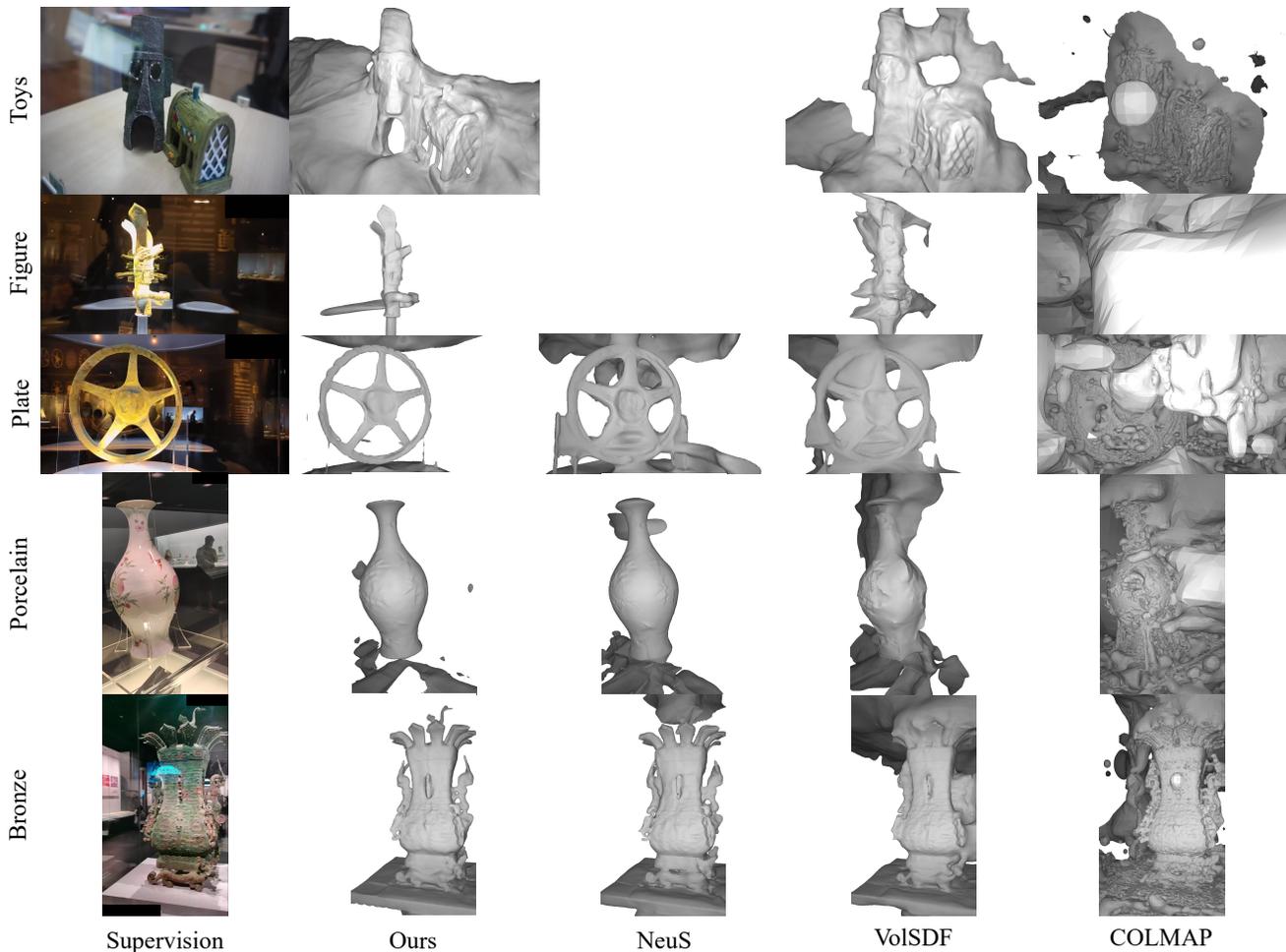


图 10. 真实世界数据集的定性比较。NeuS [47] 无法恢复“Toys ”和“Figure”上有意义的物体表面。我们的方法能针对 HSR 在物理上恢复目标物体的正确表面，而其他方法则会产生带有噪声的结果。

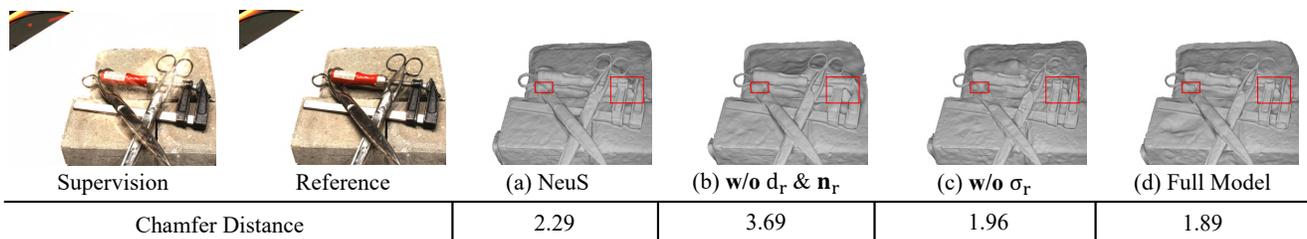


图 11. 对合成场景的消融研究。在没有辅助平面属性 ( $d_r$  和  $n_r$ ) 的情况下分解场景时，性能明显下降。在禁用体积密度  $\sigma_r$  的同时，恢复的目标物体表面失去了精细的细节。

告了得分。可以看到，我们提出的方法在大多数场景下都优于其他方法。因此，NeuS-HSR 在平均倒角距离方面超过了其他方法。

#### 4.4. 定性比较

如图9所示，我们展示了不同方法产生的重建结果。可以看到，其他隐式神经方法会生成带有噪声的

不完整物体表面，并倾向于对附着在目标物体的假镜面反射进行建模。HSR 对这些方法产生了负面影响。相比之下，我们的方法获得了更清晰的结果，并正确重建了物体表面的几何细节。结果表明，我们提出的辅助平面模块可以减少 HSR 带来的影响，重建正确的目标表面。此外，我们进一步评估了每种方法在更具挑战性的真实世界 HSR 场景中的鲁棒性。与

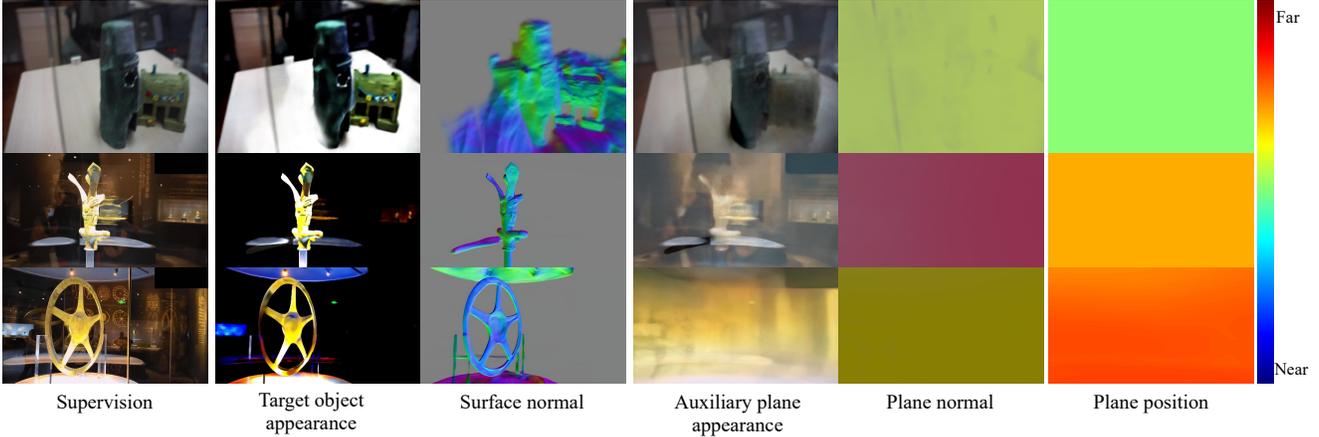


图 12. NeuS-HSR 的组件。NeuS-HSR 的渲染图像被分解为两种外观：目标物体和辅助平面。目标物体外观编码了清晰度较高的完整目标物体，而辅助平面外观则增强了 HSR 的内容。色条代表平面位置的颜色图（“近”表示辅助平面靠近相机）。

表 1. 通过测量倒角距离对合成数据集得出的定量结果。VoISDF [54] 在最后 7 个场景中无法恢复有意义的物体表面。我们的方法的综合表现优于其他方法。最佳的评价结果在表中**加粗**表示。

| ScanID  | UNISURF [32] | VoISDF [54] | NeuS [47]   | Ours        |
|---------|--------------|-------------|-------------|-------------|
| scan24  | 2.92         | 3.89        | 5.30        | <b>2.07</b> |
| scan37  | 4.26         | 2.91        | 2.29        | <b>1.89</b> |
| scan40  | 3.36         | 2.44        | <b>2.02</b> | 2.17        |
| scan55  | 2.11         | 3.95        | 1.73        | <b>1.25</b> |
| scan63  | 2.73         | -           | 2.75        | <b>1.94</b> |
| scan65  | 1.57         | -           | <b>0.93</b> | 1.15        |
| scan69  | 5.00         | -           | 4.15        | <b>3.54</b> |
| scan83  | 1.81         | -           | 2.55        | <b>1.42</b> |
| scan97  | 3.85         | -           | 4.62        | <b>2.82</b> |
| scan105 | 2.01         | -           | 1.53        | <b>1.31</b> |
| mean    | 2.96         | /           | 2.79        | <b>1.96</b> |

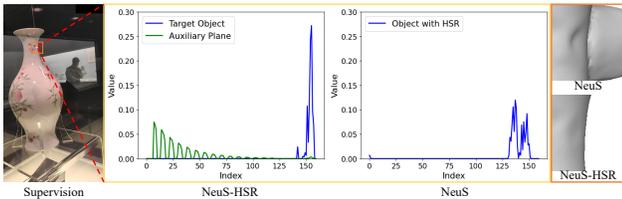


图 13. 真实世界 HSR 场景中的渲染权重。NeuS-HSR 使 NeuS 能够在对抗 HSR 时更加关注目标物体，右侧的裁剪网格就是证明。

合成 HSR 场景相比，真实世界的 HSR 场景编码了更多样化的歧义信息。图 10 说明与 state-of-the-art 神经隐式方法相比，我们的方法在包含目标物体的细小结构时产生了更好的结果。

#### 4.5. 消融研究

我们进行了多个消融实验，以研究不同设置对辅助平面模块的影响，包括体积密度  $\sigma_r$  和平面属性（位置  $d_r$  和平面法线  $n_r$ ）。图 11 展示了每个设置的结果。

**平面属性的作用。**对于视图中的每个相机射线，我们使用 MLP 生成体积密度  $\sigma_r$ ，以及辅助平面的属性（ $d_r$  和  $n_r$ ）。当我们删除辅助平面的属性并仅利用  $\sigma_r$  来生成权重时，与完整模型相比，性能会急剧下降。没有平面属性时，MLP 无法隐式地跟踪入射光并在物理上将目标物体的外观与其他部分分开。

**体积密度的作用。**为了确定体积密度  $\sigma_r$  对于恢复物体表面是否必要，我们禁止 MLP 输出  $\sigma_r$ ，并采用与物体路径相同的权重  $w$  来渲染两种外观。此操作引入了预测 SDF 的两条路径的歧义，然后产生比完整模型更差的结果。然而，由于具有鲁棒性的辅助平面，我们使用此设置的模型仍然比基线 NeuS 获得更好的性能。

## 5. 讨论

**组件。**我们的模型由两部分组成：目标物体和辅助平面。图 12 显示了每个部分的组件。目标物体的外观得到有效的增强，HSR 由辅助平面模块拍摄。辅助平面的表面法线和位置由 MLP 自适应学习。一个视角上所有相机射线的平面法线和位置变得一致，即

在物理上模拟出了一个平面反射器。

**注意力分析。**在 HSR 场景中，为了恢复准确的目标物体，我们的模型应该更加关注目标路径而不是平面路径。如图 13 所示，与辅助平面和 NeuS 的权重相比，目标物体的渲染权重具有更高的峰值和更集中的分布。这说明辅助平面模块可以使 MLP 聚焦于目标物体，进而减少 HSR 的干扰，达到更精确的效果。

**局限性。**所提出的方法仍继承了基于隐式神经表达的多视角重建方法中存在的问题。由于缺乏先验，我们的模型会在不可见区域生成不准确的目标物体几何形状。一个可能的解决方案是引入物体的对称性。

## 6. 总结

在这项工作中，我们提出了在 HSR 干扰下的多视角物体重建任务。为了解决这个任务，我们提出了 NeuS-HSR，一个全新的在 HSR 干扰下准确恢复三维物体表面的框架。我们提出将隔着玻璃拍摄的场景区分为目标物体部分和辅助平面部分，通过辅助平面来增强目标物体。我们设计了一个辅助平面模块，利用 MLP 和反射变换在物理上生成辅助平面外观。在合成场景和真实场景中的实验表明，NeuS-HSR 在定量重建质量和视觉效果方面优于之前的方法。此外，讨论还探讨了我们的分解方法在任务中的有效性。

## Acknowledgments

This work is supported by the National Key Research and Development Program of China Grant (No.2018AAA0100400), NSFC (No.61922046) and NSFC (No.62132012).

## 参考文献

- [1] Henrik Aanæs, Rasmus Ramsbøl Jensen, George Vogiatzis, Engin Tola, and Anders Bjarholm Dahl. Large-scale data for multiple-view stereopsis. *International Journal of Computer Vision*, 120(2):153–168, 2016.
- [2] Matan Atzmon and Yaron Lipman. Sal: Sign agnostic learning of shapes from raw data. In *Proceedings of the*

*IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2565–2574, 2020.

- [3] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.*, 28(3):24, 2009.
- [4] Ya-Chu Chang, Chia-Ni Lu, Chia-Chi Cheng, and Wei-Chen Chiu. Single image reflection removal with edge guidance, reflection classifier, and recurrent decomposition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2033–2042, January 2021.
- [5] Ziang Cheng, Hongdong Li, Yuta Asano, Yinqiang Zheng, and Imari Sato. Multi-view 3d reconstruction of a textureless smooth surface of unknown generic reflectance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16226–16235, 2021.
- [6] Ziang Cheng, Hongdong Li, Richard Hartley, Yinqiang Zheng, and Imari Sato. Diffeomorphic neural surface parameterization for 3d and reflectance acquisition. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–10, 2022.
- [7] François Darmon, Bénédicte Bascle, Jean-Clément Devaux, Pascal Monasse, and Mathieu Aubry. Improving neural implicit surfaces geometry with patch warping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6260–6269, 2022.
- [8] Jeremy S De Bonet and Paul Viola. Poxels: Probabilistic voxelized volume reconstruction. In *Proceedings of International Conference on Computer Vision (ICCV)*, volume 2, 1999.
- [9] Zheng Dong, Ke Xu, Yin Yang, Hujun Bao, Weiwei Xu, and Rynson W.H. Lau. Location-aware single image reflection removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5017–5026, October 2021.
- [10] Qiancheng Fu, Qingshan Xu, Yew-Soon Ong, and Wenbing Tao. Geo-neus: Geometry-consistent neural implicit surfaces learning for multi-view reconstruction. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- [11] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, 32(8):1362–1376, 2009.
- [12] Silvano Galliani, Katrin Lasinger, and Konrad Schindler. Gipuma: Massively parallel multi-view stereo reconstruction. *Publikationen der Deutschen Gesellschaft für Pho-*

- togrammetrie, Fernerkundung und Geoinformation e. V.*, 25(361-369):2, 2016.
- [13] Clement Godard, Peter Hedman, Wenbin Li, and Gabriel J Brostow. Multi-view reconstruction of highly specular surfaces in uncontrolled environments. In *2015 International Conference on 3D Vision*, pages 19–27. IEEE, 2015.
- [14] Shubham Goel, Georgia Gkioxari, and Jitendra Malik. Differentiable stereopsis: Meshes from multiple views using differentiable rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8635–8644, 2022.
- [15] Michael Goesele, Brian Curless, and Steven M Seitz. Multi-view stereo revisited. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, volume 2, pages 2402–2409. IEEE, 2006.
- [16] Ronald Goldman. *Matrices and transformations*. Graphics Gems, 1990.
- [17] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. In *International Conference on Machine Learning*, pages 3789–3799. PMLR, 2020.
- [18] Yuan-Chen Guo, Di Kang, Linchao Bao, Yu He, and Song-Hai Zhang. Nerfren: Neural radiance fields with reflections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18409–18418, 2022.
- [19] Carlos Hernandez, George Vogiatzis, and Roberto Cipolla. Multiview photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):548–554, 2008.
- [20] David S Immel, Michael F Cohen, and Donald P Greenberg. A radiosity method for non-diffuse environments. *Acm Siggraph Computer Graphics*, 20(4):133–142, 1986.
- [21] James T Kajiya. The rendering equation. In *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, pages 143–150, 1986.
- [22] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)*, 32(3):1–13, 2013.
- [23] Chenyang Lei and Qifeng Chen. Robust reflection removal with reflection-free flash-only cues. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14811–14820, 2021.
- [24] Chenyang Lei, Xuhua Huang, Chenyang Qi, Yankun Zhao, Wenxiu Sun, Qiong Yan, and Qifeng Chen. A categorized reflection removal dataset with diverse real-world scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3040–3048, 2022.
- [25] Zhengqin Li, Yu-Ying Yeh, and Manmohan Chandraker. Through the looking glass: Neural 3d reconstruction of transparent shapes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1262–1271, 2020.
- [26] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. Barf: Bundle-adjusting neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5741–5751, 2021.
- [27] Ming Liu, Jianan Pan, Zifei Yan, Wangmeng Zuo, and Lei Zhang. Adaptive network combination for single-image reflection removal: A domain generalization perspective. *arXiv preprint arXiv:2204.01505*, 2022.
- [28] Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. Neural volumes: learning dynamic renderable volumes from images. *ACM Transactions on Graphics (TOG)*, 38(4):1–14, 2019.
- [29] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. *ACM siggraph computer graphics*, 21(4):163–169, 1987.
- [30] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*, pages 405–421. Springer, 2020.
- [31] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3504–3515, 2020.
- [32] Michael Oechsle, Songyou Peng, and Andreas Geiger. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5589–5599, 2021.
- [33] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 523–540. Springer, 2020.
- [34] Jiaxiong Qiu, Peng-Tao Jiang, Yifan Zhu, Ze-Xin Yin, Ming-Ming Cheng, and Bo Ren. Looking through the glass:

- Neural surface reconstruction against high specular reflections. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [35] Sverker Rasmuson, Erik Sintorn, and Ulf Assarsson. Addressing the shape-radiance ambiguity in view-dependent radiance fields. *arXiv preprint arXiv:2203.01553*, 2022.
- [36] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016.
- [37] Johannes L Schönberger, Enliang Zheng, Jan-Michael Frahm, and Marc Pollefeys. Pixelwise view selection for unstructured multi-view stereo. In *European conference on computer vision*, pages 501–518. Springer, 2016.
- [38] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016.
- [39] Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, volume 1, pages 519–528. IEEE, 2006.
- [40] Steven M Seitz and Charles R Dyer. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 35(2):151–173, 1999.
- [41] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *Advances in Neural Information Processing Systems*, 32, 2019.
- [42] Binbin Song, Jiantao Zhou, and Haiwei Wu. Multi-stage curvature-guided network for progressive single image reflection removal. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.
- [43] 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.
- [44] Shimon Ullman. The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 203(1153):405–426, 1979.
- [45] 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. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5481–5490. IEEE, 2022.
- [46] Jiepeng Wang, Peng Wang, Xiaoxiao Long, Christian Theobalt, Taku Komura, Lingjie Liu, and Wenping Wang. Neuris: Neural reconstruction of indoor scenes using normal priors. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII*, pages 139–155. Springer, 2022.
- [47] 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.
- [48] Yiqun Wang, Ivan Skorokhodov, and Peter Wonka. Hf-neus: Improved surface reconstruction using high-frequency details. In *Advances in Neural Information Processing Systems*, 2022.
- [49] Yi Wei, Shaohui Liu, Yongming Rao, Wang Zhao, Jiwen Lu, and Jie Zhou. Nerfingmvs: Guided optimization of neural radiance fields for indoor multi-view stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5610–5619, 2021.
- [50] Markus Worchel, Rodrigo Diaz, Weiwen Hu, Oliver Schreer, Ingo Feldmann, and Peter Eisert. Multi-view mesh reconstruction with neural deferred shading. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6187–6197, 2022.
- [51] Changchang Wu. Visualsfm: A visual structure from motion system. <http://www.cs.washington.edu/homes/ccwu/vsfm>, 2011.
- [52] Jiamin Xu, Zihan Zhu, Hujun Bao, and Wewei Xu. A hybrid mesh-neural representation for 3d transparent object reconstruction. *arXiv preprint arXiv:2203.12613*, 2022.
- [53] Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K Wong. S3-nerf: Neural reflectance field from shading and shadow under a single viewpoint. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- [54] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*, 34:4805–4815, 2021.
- [55] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and ap-

- pearance. *Advances in Neural Information Processing Systems*, 33:2492–2502, 2020.
- [56] 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, 2021.
- [57] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020.
- [58] Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4786–4794, 2018.

## 附录

与旨在重建 3D 透明物体的最新透明物体重建方法 [25,52] 不同，我们的目标是恢复透明物体（即玻璃）后面的物体表面。在这里，我们提供了所提方法和实验的更多细节。具体来说，我们提供了在正文中采用的反射变换 [16] 和投影算法的详细信息 (A节)、线性求和中比率的选择 (B节)、实验的其他细节 (C节)、附加结果 (D节)、附加分析 (E节) 和未来工作 (F节)。

### A. 变换算法

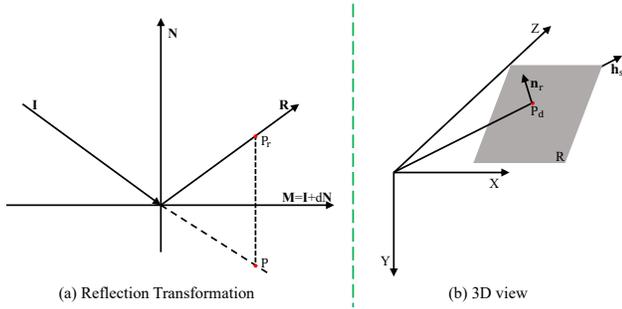


图 14. (a): [16]. (b) 相机坐标系中辅助平面的三维视角。R 是相机射线  $h_s$  的辅助平面， $n_r$  是 R 的单位法向量。

图 14(a) 展示了反射变换的示意图 [16]。假设光线入射到玻璃 M 上，入射方向为 I，反射方向为 R，M 的平面方程定义为：

$$L \cdot P = Ax + By + Cz + D = 0 \quad (11)$$

其中  $L = (A, B, C, D)$  和  $P = (x, y, z, 1)$ 。M 的单位法线向量 N 是  $(A, B, C, 0)$ 。

给定  $P_r$  是反射光上的一点，P 是它的虚像，则  $L \cdot P_r$  是  $P_r$  和 M 之间的垂直距离，可以得到：

$$P = P_r - 2(L \cdot P_r)N = M_r P_r \quad (12)$$

其中  $M_r$  为：

$$M_r = \begin{bmatrix} 1 - 2A^2 & -2AB & -2AC & -2AD \\ -2AB & 1 - 2B^2 & -2BC & -2BD \\ -2AC & -2BC & 1 - 2C^2 & -2CD \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (13)$$

### 算法 1: 沿相机射线 $h_s$ 变换采样点。

**Input:** 平面法线  $n_r = (A, B, C)$ , 平面位置  $d_r$ , 相机中心  $o$ , 深度  $t$ , 视角方向  $v$  和沿着  $h_s$  的采样点  $p$ 。

**Output:** 相机坐标系中平面路径的空间点

$P_r$ 。

- 1  $p' = p_t \cup p'_t = p - o$ ;
- 2  $P_d = d_r v$ ;
- 3  $D = -d_r n_r \cdot v$ ;
- 4  $p_t = \{tv | t \in [0, d_r]\}$ ;
- 5  $p'_t = \{tv | t \in [d_r, 1]\}$ ;
- 6  $\{D, n_r\} \rightarrow M_r$ ;
- 7  $p_a = M_r^{-1} p'_t$ ;
- 8  $p_r = p_t \cup p_a$ .

表 2. 目标物体外观的不同比例对"scan24"的影响。标准偏差为 0.68。

| 比例     | 0.1  | 0.3  | 0.5  | 0.7  | 0.9  |
|--------|------|------|------|------|------|
| 倒角距离 ↓ | 3.29 | 2.07 | 2.43 | 3.99 | 3.17 |

显然式 (12) 是可微函数。在我们的工作中，相机射线沿着 R 的反方向，我们使用反射变换来追踪入射光线。图 14(b) 显示了辅助平面的 3D 视图。辅助平面建立在相机坐标系中。给定笛卡尔观察方向单位向量  $v = (x_v, y_v, z_v)$  和平面位置  $d_r$ ，则我们有：

$$P_d = d_r v = (d_r x_v, d_r y_v, d_r z_v) \quad (14)$$

$P_d$  在辅助平面上 (即式 (1))。给定  $n_r = (A, B, C)$ ，我们可以得到：

$$\begin{aligned} D &= -(A d_r x_v + B d_r y_v + C d_r z_v) \\ &= -d_r (A x_v + B y_v + C z_v) \\ &= -d_r n_r \cdot v \end{aligned} \quad (15)$$

综上，我们提出了算法 1 中获取平面路径输入点  $p_r$  的策略。

### B. 线性求和

线性求和可以通过拍摄的 RGB 图像进行监督从而将两条路径的外观融合为渲染图像。由表 2 所示，

不同的目标物体外观比例对目标物体的重建质量有不同的影响。根据这些结果，我们选择 0.3 作为模型中的默认比例。

## C. 其他实验细节

### C.1. 分层抽样

我们遵循 NeuS [47] 的分层采样来生成输入空间点。具体来说，我们首先沿一条射线均匀采样 64 个点，然后进行 4 次重要性采样。采样点总数为 128。根据 NeRF++ [57]，我们在球体外部额外采样了 32 个点。

### C.2. 神经网络架构

整个网络架构由三部分组成：SDF 预测、辅助平面预测和颜色预测。对于 SDF 预测，我们遵循 NeuS 的神经网络架构，它由 Softplus 激活，其中  $\beta = 100$ 。为了稳定的训练采用了权重归一化。输入通过跳跃连接与第四层的特征连接起来。对于辅助平面预测，体积密度部分由三个带有 ReLU 的线性层组成，位置和法线分支均由两个线性层组成。两个分支的隐藏层由 ReLU 激活，而两个分支的最后层分别由 Sigmoid 和 Tanh 激活。对于颜色预测，隐藏层由 ReLU 激活，最后一层由 Sigmoid 激活。

### C.3. 数据集

表3报告了我们的合成数据集和真实数据集的指标。对于合成数据集，我们设置高斯滤波器的核大小为 11 以产生反射效果。我们随机选择 DTU 数据集 [1] 中的一个场景（“Scan114”）作为高强度镜面反射的来源。我们根据视觉现实从 DTU 数据集 [1] 上总共 15 个场景中选择了 10 个场景。它们是：‘Scan24’，‘Scan37’，‘Scan40’，‘Scan55’，‘Scan63’，‘Scan65’，‘Scan69’，‘Scan83’，‘Scan97’ 和 ‘Scan105’。对于真实世界的数据集，我们拍摄一个场景（“Toys”）并从互联网上收集 5 个场景：‘Buddha’<sup>1</sup>，‘Figure’<sup>2</sup>，

<sup>1</sup><https://www.bilibili.com/video/BV1M44y1z7XX>

<sup>2</sup><https://www.bilibili.com/video/BV1BP4y1Y7bV>

‘Plate’<sup>3</sup>，‘Porcelain’<sup>4</sup> and ‘Bronze’<sup>5</sup>。图18显示了我们在实验中的真实世界场景示例。

### C.4. 推理时间

对于物体表面重建，在分辨率 =64 和阈值 =0.0 的条件下，NeuS-HSR 的推理时间为 36 秒。在单张 NVIDIA Tesla V100 GPU 上渲染分辨率为 800 x 600 的新视角时，NeuS-HSR 在没有真实掩码的情况下需要 96 秒左右。

### C.5. 基线

由于原始的隐式神经基线是在没有 HSR 的数据集上训练和测试的，因此我们在合成和真实世界数据集的每个场景上重新训练所有这些模型。

**NeuS [47].** 为了获得 NeuS 的结果，我们使用了其发布的官方代码<sup>6</sup>，所有场景均采用默认设置。

**UNISURF [32].** 为了与 UNISURF 进行比较，我们采用了其正式发布的代码<sup>7</sup>，并在合成场景中使用了默认设置。

**VolSDF [54].** 为了与 VolSDF 进行比较，我们在所有场景中使用其正式发布的代码<sup>8</sup>和默认设置。

**COLMAP [38].** 为了获得 COLMAP 的结果，我们使用 COLMAP<sup>9</sup> 的官方命令版本，并在所有场景中运行其文档<sup>10</sup>中提供的顺序命令。

### C.6. Q&A

Q1. 在新视角下的渲染图像质量如何？

A1. 我们的工作目标是多视角图像作为监督，针对 HSR 准确的重建目标物体。我们在“Bronze”上对新视角下的合成进行了评估。我们选择连续图像的前 3 幅图像和后 7 幅图像进行测试。测试集的 PSNR 和 SSIM 的平均值见表5。

<sup>3</sup><https://www.bilibili.com/video/BV1BP4y1Y7bV>

<sup>4</sup><https://www.bilibili.com/video/BV1UP4y1h7tW>

<sup>5</sup><https://www.bilibili.com/video/BV1SU4y1E7QR>

<sup>6</sup><https://github.com/Totoro97/NeuS>

<sup>7</sup><https://github.com/autonomousvision/unisurf>

<sup>8</sup><https://github.com/lioryariv/volsdf>

<sup>9</sup><https://github.com/colmap/colmap>

<sup>10</sup><https://colmap.github.io/>

表 3. 实验中使用的数据集指标。

| Scene      | Synthetic   | Buddha      | Toys        | Figure      | Plate       | Porcelain   | Bronze      |
|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Views      | 49/64       | 56          | 23          | 60          | 56          | 60          | 43          |
| Resolution | 1600 × 1200 | 1920 × 1080 | 1372 × 1029 | 1920 × 1080 | 1920 × 1080 | 1080 × 1920 | 1080 × 1920 |

表 4. NeuS-HSR 和基线的模型参数。

| Method  | UNISURF [32] | VolSDF [54] | NeuS [47] | NeuS-HSR |
|---------|--------------|-------------|-----------|----------|
| #Params | 0.8M         | 1.4M        | 1.4M      | 1.5M     |

表 5. ‘Bronze’ 上的新视角合成比较。

| Method | NeRF++ [57] | NeuS [47] | NeuS-HSR |
|--------|-------------|-----------|----------|
| PSNR↑  | 15.92       | 15.51     | 15.93    |
| SSIM↑  | 0.480       | 0.489     | 0.502    |

Q2. 为什么在两条路径中使用相同的外观函数  $F_c$  ?

A2. 首先, 我们使用相同的  $F_c$  来保存模型参数。其次, 由于我们框架中的两条路径是在一个阶段中训练完成的, 而监督训练仅使用拍摄的图像, 因此我们采用相同的  $F_c$  来分离同一域中的两种外观。最后, 我们将  $F_c$  视为一个隐式函数, 它将三维位置、法线、视角方向和神经特征映射为颜色值。我们使用两条路径的不同三维位置和法线作为输入, 通过  $F_c$  获得不同的颜色值。

Q3. 为什么辅助平面建立在相机坐标系中?

A3. 如图14所示, 我们的模型在每次迭代时都由一个视角进行训练, 为了简化训练, 我们在相机坐标系中为每个视角建立了一个辅助平面。我们首先将三维位置变换到摄像机坐标系, 然后我们可以直接应用辅助平面的反射变换。

Q4. NeuS-HSR 在非 HSR 场景中的表现如何?

A4. NeuS-HSR 建立在 HSR 场景下的两个物理假设之上。在非 HSR 场景中, 我们可以将线性求和的比率设置为 1.0, 那么 NeuS-HSR 就退化为 NeuS, 并且可以达到与 NeuS 相同的性能。

## D. 合成数据集的其他结果

### D.1. 符号距离场

我们在图15中将符号距离场可视化。与 NeuS 相比, 我们的模型根据符号距离场的分布提取了更精

确的场景 SDF。具体来说, 我们的符号距离场展示了 Bunny 切平面的几何特征。

### D.2. 组件

图19显示了 NeuS-HSR 在合成数据集上的组成。我们的方法有效地增强了目标物体的外观, 并在没有任何先验的情况下保留了辅助平面外观的 HSR。此外, 在一个视角中, 每个辅助平面在相机射线上的平面法线和位置表明辅助平面倾向于平面反射体。因此, 图19说明我们的模型实现了 HSR 场景的物理分解。

### D.3. 对比

图17显示了 NeuS-HSR 和其他 state-of-the-art 方法在合成数据集上的更多定性比较。所有的隐式神经方法都是在没有真实掩码的情况下训练的。COLMAP [38] 在目标物体表面周围产生了太多的噪声, 因此无法计算我们正文中其结果的度量 (即倒角距离)。

## E. 真实世界数据集上的可训练标准偏差

在 NeuS [47] 中, 优化过程自动降低了标准偏差, 从而使表面变得更加锐利。我们比较了 NeuS-HSR (蓝色曲线) 和 NeuS (橙色曲线) 的可训练标准偏差。与 NeuS 的标准偏差相比, 我们方法的标准偏差收敛到了一个更小的值, 我们的方法在真实世界数据集上取得了比 NeuS 更清晰、更锐利的结果。

## F. 未来工作

未来, 我们计划将方法扩展来处理不同厚度的玻璃上去。在日常生活中, 玻璃越厚, 镜面反射越明显。一种可行的方案是在我们的辅助平面模块中增加厚度。此外, 我们的方法还可以用于处理高反射物体表面 (如汽车)。

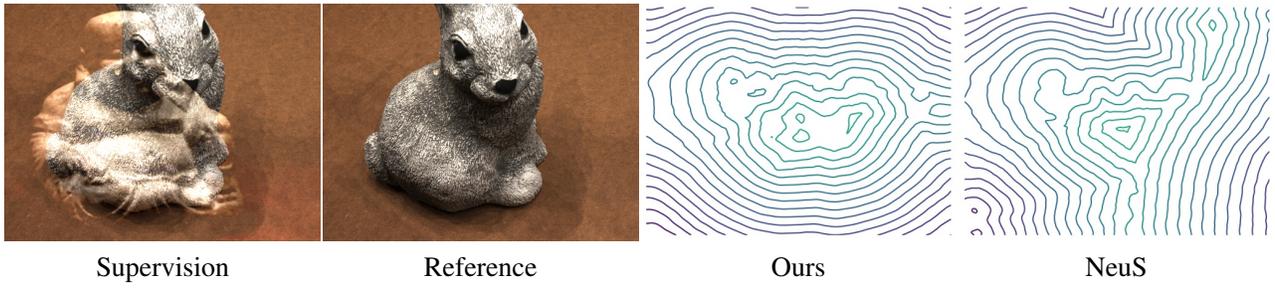


图 15. 符号距离场可视化。

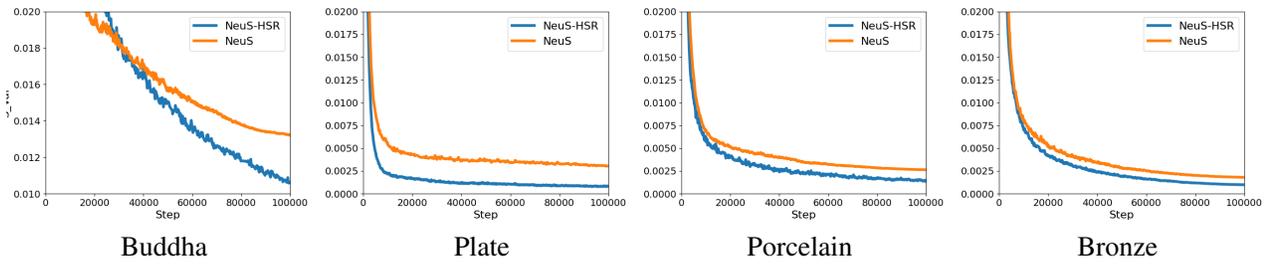


图 16. 可训练标准偏差的比较。

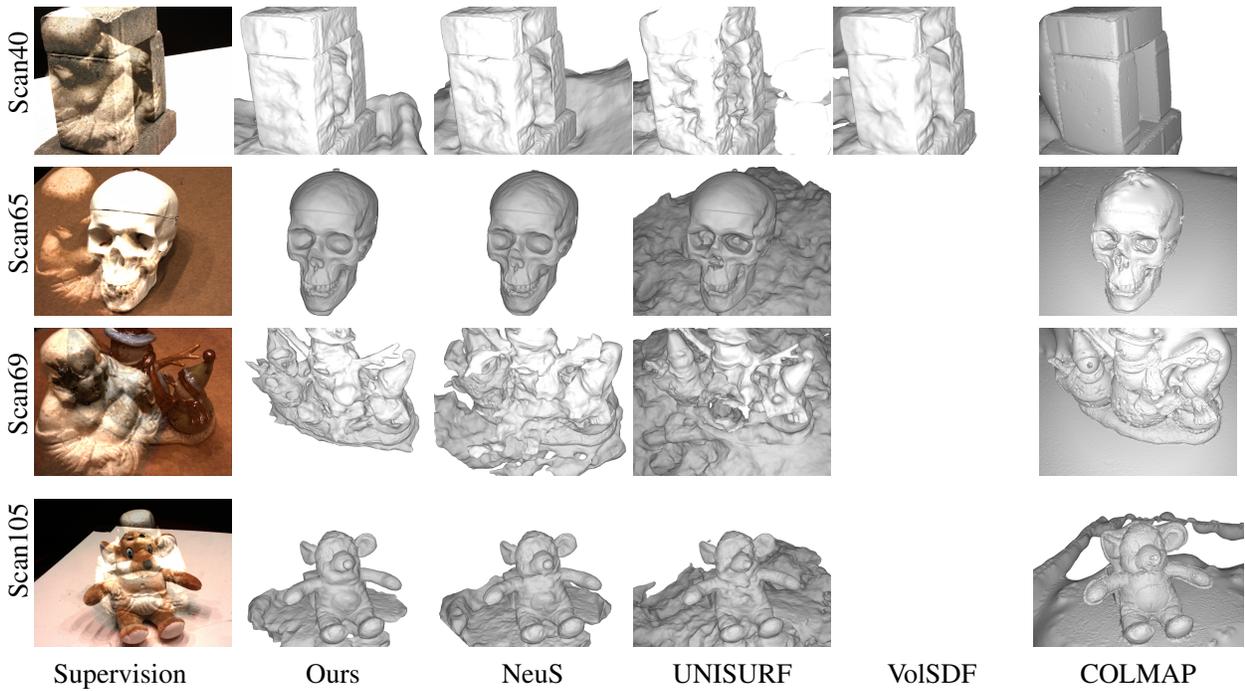


图 17. NeuS-HSR 与基线在合成数据集上的定性比较。

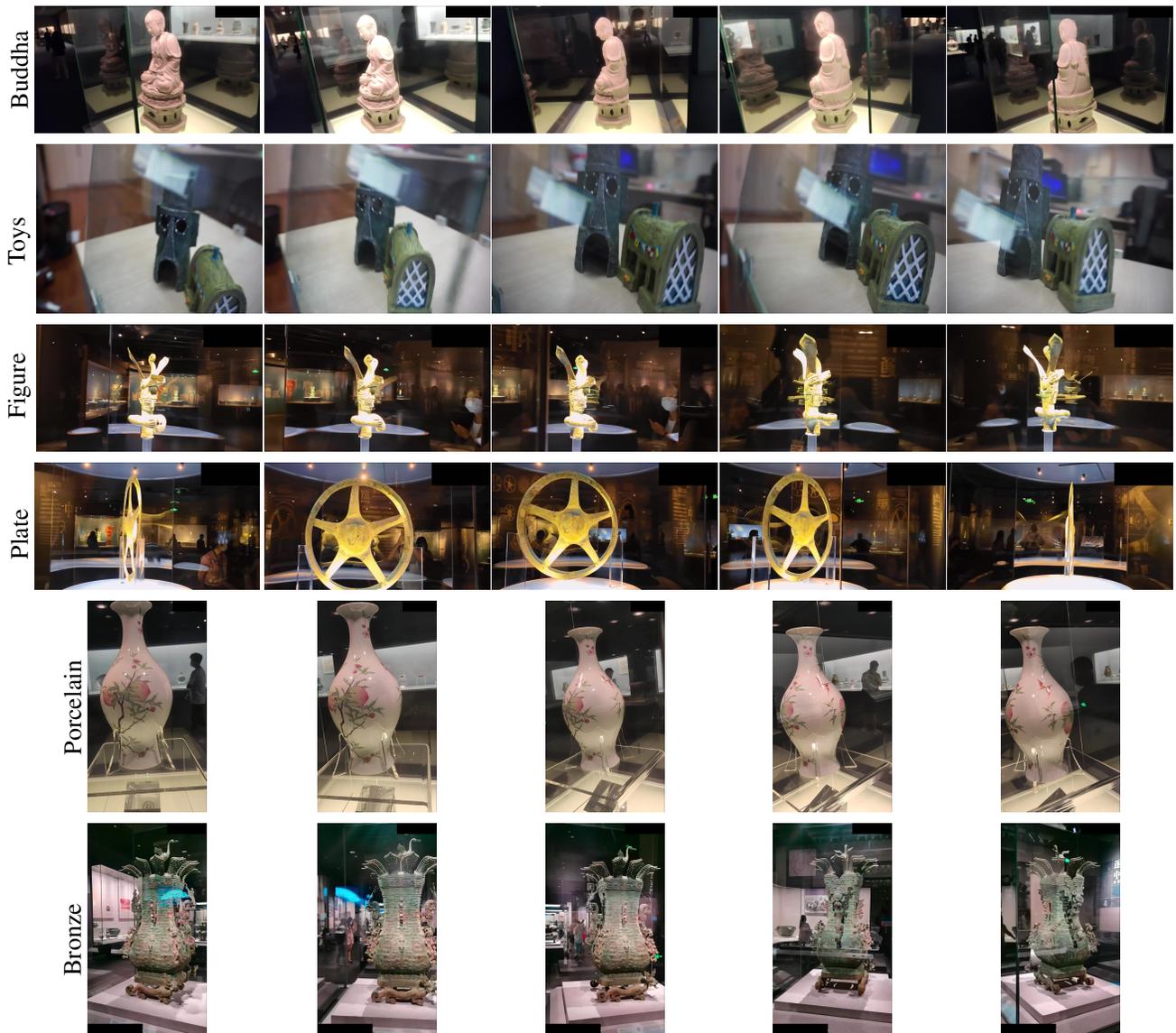


图 18. 真实世界数据集示例。

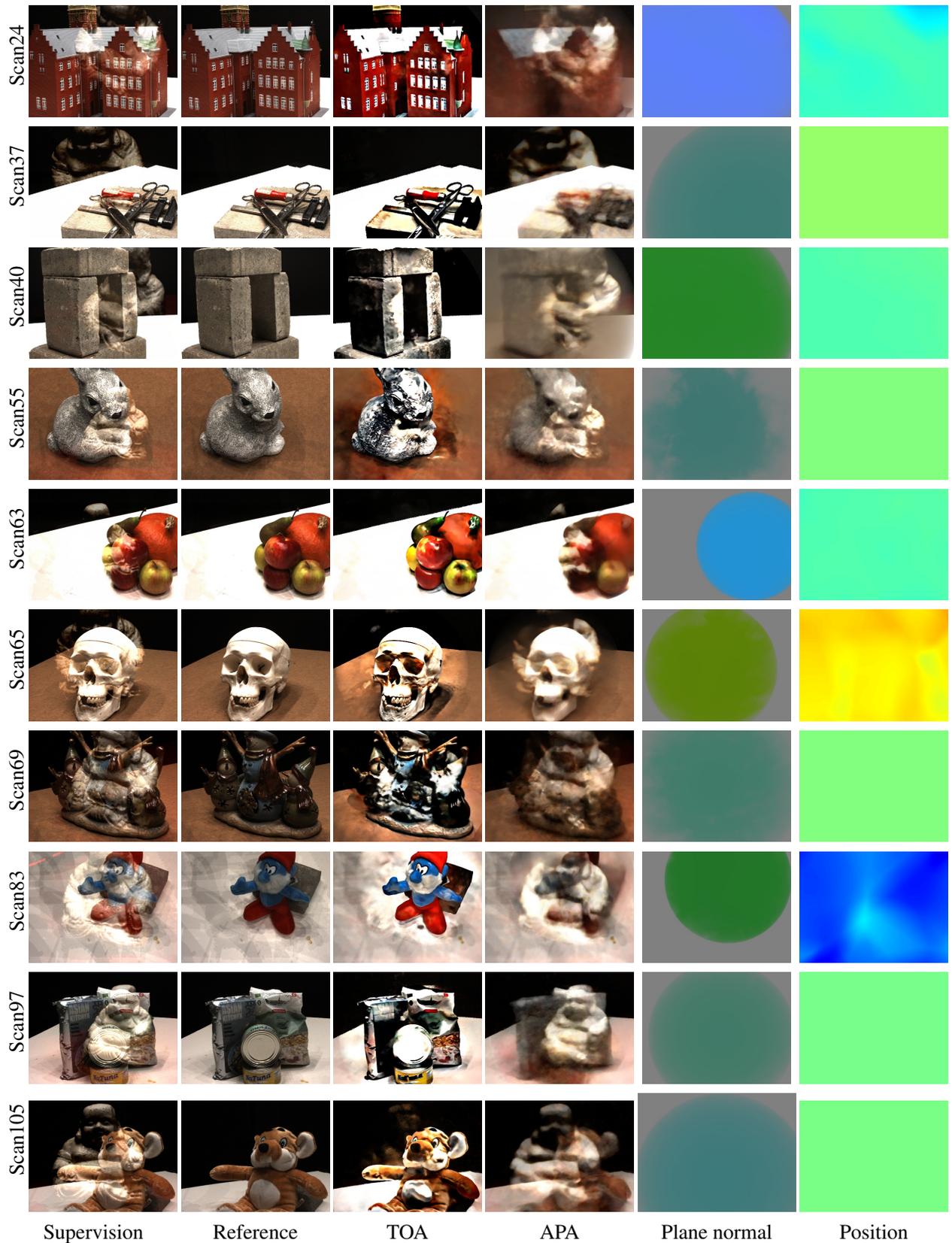


图 19. 合成数据集上的 NeuS-HSR 组件。‘TOA’: 目标物体外观。‘APA’: 辅助平面外观。