Computer Science and Application 计算机科学与应用, 2024, 14(5), 265-276
Published Online May 2024 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2024.145135

Hans汉斯

针对高亮物体的高效神经隐式表面重建

何思源，刘兴林

五邑大学电子与信息工程学院，广东 江门

收稿日期：2024年4月28日；录用日期：2024年5月24日；发布日期：2024年5月31日

摘　要

神经隐式表面(Neural Implicit Surface)一直是近年来计算机视觉的热门研究方向。许多工作通过扩展神经辐射场的体渲染管线实现了从二维图像和相机位姿作为输入，在无需三维监督信息下重建高质量三维物体形状。但是，由于利用二维图像进行训练监督，这些方法难以对高亮物体的形状进行合理的推理，其原因是因为物体材质和环境光照所产生的模糊性。本文提出一种针对高亮物体的高效表面重建方法，通过权重插值辐射场(Radiance Field)与反射场(Reflection Field)的方式使得可以更好地表达高亮物体的外观。同时，本文引入了一种渲染损失的方法来缓解高光反射带来的多视角不一致问题，并且引入了两种针对物体法向量的正则化来缓解混合神经场梯度噪声的问题。本工作通过渐进式的训练范式分别对三种数据集进行了实验，实验表明，本方法在多视角合成和高亮物体表面重建任务上都超越了基准模型，并且在训练速度上比基准模型快一个量级。

关键词

神经符号距离场，隐式曲面，表面重建，高亮物体

Efficient Neural Implicit Surface Reconstruction for Glossy Objects

Siyuan He, Xinglin Liu

School of Electronics and Information Engineering, Wuyi University, Jiangmen Guangdong

Received: Apr. 28th, 2024; accepted: May 24th, 2024; published: May 31st, 2024

Abstract

Neural implicit surface has been a popular research direction in computer vision in recent years. Many approaches have extended the volume rendering pipeline of neural radiance field to reconstruct high quality 3D object shapes from 2D images and camera poses as input without any 3D supervision. However, due to the use of two-dimensional images for training supervision, it is dif-

文章引用: 何思源, 刘兴林. 针对高亮物体的高效神经隐式表面重建[J]. 计算机科学与应用, 2024, 14(5): 265-276.
DOI: 10.12677/csa.2024.145135

ficult for these methods to rationally reason about the shape of glossy objects, because of the ambiguity caused by the material and environment lighting. In this paper, we propose an efficient surface reconstruction method specifically designed for glossy objects, which better represents the appearance of such objects through the interpolation of a weighted radiance field and reflection field. Additionally, we introduce a relax rendering loss to alleviate the issue of multi-view inconsistency caused by specular reflections, and two types of regularization for object normal to reduce the gradient noise of the hybrid neural field. Experiments on three datasets using a progressive training paradigm demonstrate that the proposed method outperforms baseline models in novel view synthesis and surface reconstruction tasks, while achieving training speeds approximately one order of magnitude faster than baseline models.

Keywords

Neural Signed Distance Field, Implicit Surface, Surface Reconstruction, Glossy Objects

1. 引言

三维重建是计算机视觉的基础任务。传统的多视角立体视觉方法需要繁琐的处理流程，每一个步骤都会产生误差积累，从而影响到三维物体或场景的形状恢复。近年来，随着神经辐射场的发展，许多工作通过扩展神经辐射场的体渲染管线来拟合隐式表面，从而表达物体的三维形状。

NeuS [1]为扩展辐射场体渲染管线来拟合隐式表面的代表性工作，同期工作有 volSDF [2]、UNISURF [3]等。但是，这类工作使用的纯多层感知机的场景表达所需要的训练时长是难以接受的。一些工作通过利用离散参数化的数据结构进行插值，并与少层数的多层感知机进行结合，从而大幅度缩短训练时长，并且获得的隐式表面有着丰富的几何细节。但是，上述工作对于存在高光反射的场景或物体的拟合效果欠佳，其主要的原因为高光外观与几何存在模糊性，这使得模型陷入局部最小值。同时，高光反射影响了多视角一致性，使得基于二维图像监督的训练管线在训练时引入了极大的偏差。IDR [4]和 Ref-NeRF [5]试图解耦几何与外观来使得模型有更好的拟合效果，或者通过对松绑渲染损失来缓解多视角不一致所带来的影响[5]。但是，这些工作要么无法很好地恢复隐式表面，要么训练时长较长。

本文工作通过提出了一种混合反射场(Reflection Field)与辐射场的插值方法，在缓解反射场训练不稳定的同时，使得整体网络可以更好地表达高光反射的外观。同时，本文引入了一种渲染损失来缓解高光反射所引起的多视角不一致性，并且通过两种法向量的正则化来约束估计的法向量。针对三种不同含高光反射的数据集，实验证明，本工作在新视角合成任务和隐式表面重建任务上超越了基准模型，并且训练时长相比与基准模型缩短了一个量级。

2. 相关工作

2.1. 神经隐式表面

神经隐式表面(Neural Implicit Surface)为一种基于学习的隐式表面表示方法，其使用神经网络对输入的三维坐标点进行映射，得到对应的隐式表面表示，如符号距离场、占据概率等。DeepSDF [6]、Occupancy Network [7]等工作为使用神经网络拟合隐式表面的代表性工作，但这些工作均需要三维监督信息。随着

神经辐射场(Neural Radiance Field, NeRF) [8]的发展，许多工作通过扩展 NeRF 的体渲染管线来拟合对应的隐式表面，整个训练过程仅需要多视角图像与相机内外参数，使得这类方法的实用性大大提高。NeuS 为其中的代表工作，其通过建立符号距离场(Signed Distance Field, SDF)与 NeRF 体渲染中权重函数的联系，使得可以在无需三维监督的情况下拟合出高质量的隐式表面。但是，由于 NeuS 仅采用位置编码与多层数的多层感知机(Multilayer Perceptron, MLP)进行拟合，整体训练速度较慢。一些工作受 Instant-NGP [9]的启发，通过离散参数化的数据结构进行插值，配合少层数的 MLP，使得整体训练速度得到了提升。但这类工作对于具有高亮反射的物体依然难以正确拟合其几何，其原因是物体几何与高亮外观产生的模糊性使得在渲染监督的过程中难以辨别真实的几何形状。

2.2. 针对高亮物体的辐射场

基于神经辐射场(Neural Radiance Field, NeRF)的工作能够表达视角相关(View-Dependent)的物体外观，但对于拟合具有高光反射的外观，这些工作依然难以有很好的拟合效果。IDR [4]提出了通过视角相关方向和法向量作为外观网络的输入以进行几何与物体外观的解耦，其展现出了一定的效果。Ref-NeRF 更进一步，通过反射方向的重参数化，并且提出一种集成方向编码使得外观网络能够很有效地表达具有任意连续粗糙度材质的辐射函数，从而使得对高光物体外观的拟合有着很好的效果。但是，Ref-NeRF 由于是扩展神经辐射场的工作，对输出的体密度进行等值面提取难以得到较好质量的几何。有许多工作尝试对高光物体进行隐式表面的拟合，如 ENVIDR [10]、Ref-NeuS [11]、NeRO [12]等。但是，这类工作仅简单扩展 NeuS 的体渲染管线，其对于隐式表面拟合时间依然较长。本工作引入渐进式多分辨率哈希编码来提升隐式表面的训练速度，通过提出一种混合辐射场与反射场的插值方法，可以很好地提升网络对于高亮物体外观的表达能力。同时，本工作通过引入一种渲染损失来缓解高光反射所带来的多视角不一致性，并且使用两种法向量的正则化来约束估计的法向量，使得可以很好地缓解多分辨率哈希编码这种离散数据结构在求导时所带来的噪声。

3. 方法

本章会先简单回顾神经辐射场(NeRF)的基本原理，并且简要介绍渐进式多分辨率哈希编码。然后，对网络结构进行整体的介绍，之后针对本工作的三个部分进行详细阐述，分别为：(1) 混合辐射场与反射场(Hybrid Radiance–Reflection Field)的插值方法；(2) 缓解多视角不一致的渲染损失函数计算；(3) 正则化项。

3.1. 预备知识

神经辐射场(NeRF) [8]为一种连续的场景表达，其通过输入空间中的三维坐标点和对应的视角方向，输出该点的体密度和辐射值，利用体渲染(Volume Rendering)管线对每条光线的采样点的辐射值进行累积得到对应像素的颜色，从而实现高质量的任意视角的渲染。具体的，给定 N 张多视角图像和对应视角的相机内外参数，在不同视角中以相机原点向图像像素进行光线投射，并对每条光线进行采样，第 i 个采样点记为 x_i，即为空间的三维坐标点。通过对相机内外参数的计算得到对应的视角方向 d，整体的神经辐射场的表达为：

$$F:\left(x_i,d\right)\rightarrow\left(\sigma_i,c_i\right) \tag{1}$$

公式 1 中的 σ_i 为光线上第 i 个采样点的体密度(Volume Density)，c_i 为该采样点估计的辐射值，即该点的颜色。然后，NeRF 通过简单的体渲染对整条光线的所有采样点颜色进行积分，其计算的值为对应像素的颜色值 $\hat{C}$，具体的表示如下：

$$\hat{C}=\sum_{i=1}^{n}T_i\,\alpha_i\,c_i,\quad T_i=\exp\left(-\sum_{j=1}^{i-1}\sigma_j\,\delta_j\right) \tag{2}$$

公式 2 中，T_i 为前 i 个采样点的累积透射率，这里的 α_i 为第 i 段采样间距的不透明度。对于整个训练过程的渲染监督，NeRF 为粗阶段和精细阶段，每个阶段对于像素颜色的损失函数 $\mathcal{L}$ 均为 L2 损失，r 表示对应的光线，$C(r)$ 表示对应光线的真实像素值，具体如下：

$$\mathcal{L}=\left\|\hat{C}(r)-C(r)\right\|_2^2 \tag{3}$$

然而，NeRF 对物体表面缺乏明确的定义，NeuS 等工作通过建立符号距离场(Signed Distance Field)与公式 2 的联系，从而可以通过体渲染管线来优化隐式表面。在 NeuS 中，公式 1 具体表示如下：

$$s_i=\mathrm{MLP}\left(\left[x_i,\mathrm{PE}(x_i)\right]\right) \tag{4}$$

$$c_i=\mathrm{MLP}\left(\left[\mathrm{PE}(\mathrm{d}),enc_g,n\right]\right) \tag{5}$$

NeuS [1]中并不输出体密度 σ，而是符号距离值。这里的 s_i 为对应点的符号距离值，表示该点到物体表面的距离。PE 表示位置编码(Positional Encoding)，enc_g 表示公式 4 输出的几何特征，n 为当前点的法向量，为该点符号距离场的导数。

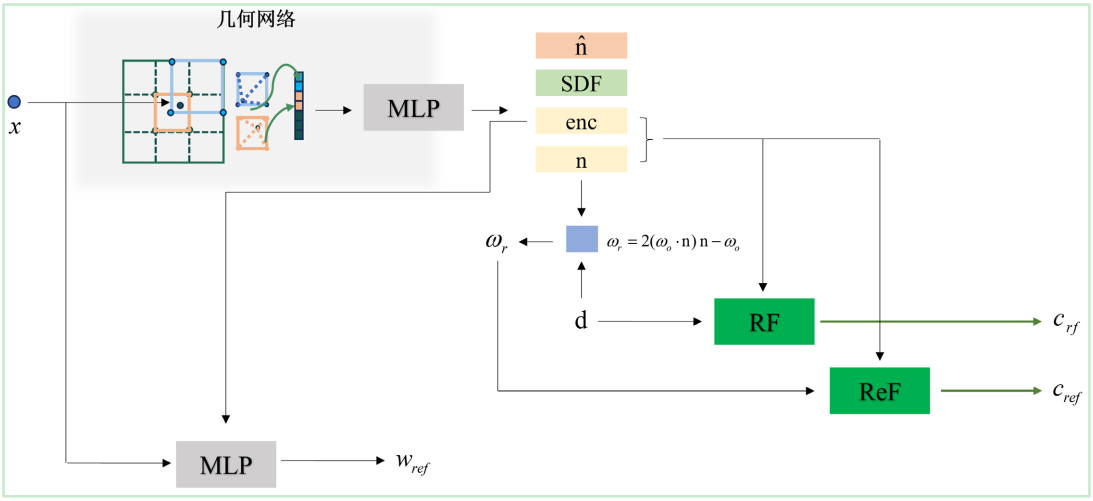

Figure 1. Diagram of the network architecture
图 1. 网络架构图

3.2. 网络结构

本文方法提出了一种能够高效恢复高亮物体隐式表面，同时在新视角合成任务上也有出色性能的方法。如图 1，整体的网络结构分为两个部分：(1) 几何网络(Geometry Network)；(2) 外观网络(Appearance Network)。

对于几何网络部分，本文采用渐进式的多分辨率哈希编码 γ 对三维坐标 x 进行插值，然后将插值好的特征体输入到一层的 MLP 中，通过 Sigmoid 函数输出估计的符号距离值。同时，单层 MLP 输出的几何特征为 enc，用于为外观网络提供几何信息。与先前的工作不同，本方法在几何网络部分针对法向量的估计提供了两种不同计算形式的输出。图 1 中，几何网络估计的法向量 $\hat{\mathrm{n}}$，其网络结构与估计符号距离值相同，但单层 MLP 与估计符号距离值的 MLP 并不共享参数，为独立的 MLP。而图 1 中法向量 n 为符号距离场(Signed Distance Field, SDF)的导数，但这里并不采用 NeuS 的解析导数形式，而是采用一种有限差分的求导方法，其表达式如下：

$$\mathrm{n} = \nabla f\left(x_i\right) = \frac{f\left(\gamma\left(x_i + \epsilon\right)\right) - f\left(\gamma\left(x_i - \epsilon\right)\right)}{2\epsilon} \tag{6}$$

公式 5 中，$f(\cdot)$表示 MLP，这里的层数为 1 层。$\gamma(\cdot)$表示多分辨率哈希编码，与 Neuralangelo [13]类似，采用一种渐进式的编码方式。这里的ϵ为栅格大小，与 Neuralangelo 中采用的占据栅格一致。

虽然，渐进式的编码可以在计算数值梯度(公式 5)时提供有很好的平滑性，但是针对高光物体表面拟合时依然会引入过多的噪声。错误的法向量对于反射的计算和外观网络的对辐射值的拟合会有很大的影响。所以，本文使用上述两种法向量 n 和 $\hat{\mathrm{n}}$ 提供了两种正则化方式，具体见章节 3.5。另外，图 1 中针对符号距离值和法向量 $\hat{\mathrm{n}}$ 的估计，这里 MLP 的初始化与 SAL 的初始化一致，其对 MLP 的权重采用正态分布的初始化方案。

对于外观网络的设计，与先前的工作不同，NeuS [1]、Neuralangelo [13]等工作均采用单一的辐射场表示，即仅对输入的视角方向进行编码，然后使用 MLP 估计出对应点的辐射值，即该点的颜色 c。但是，针对具有高光反射的物体表面，局部区域会有高频的光照反射，而先前工作仅简单采用位置编码或球面谐波等对视角方向进行编码，并采用 MLP 完成外观的表达，这种方式对物体外观的漫反射颜色(Diffuse Color)和光泽颜色(Glossy Color)整合到一起来表达，实际上难以很好地表示出局部强烈变化的外观。本工作提出一种解耦漫反射颜色与光泽颜色的方法，并使用一个估计的权重来衡量该点上两种颜色的贡献程度，更好地实现高光物体表面外观的表达，具体的表示方式见章节 3.3。

3.3. 混合辐射场与反射场

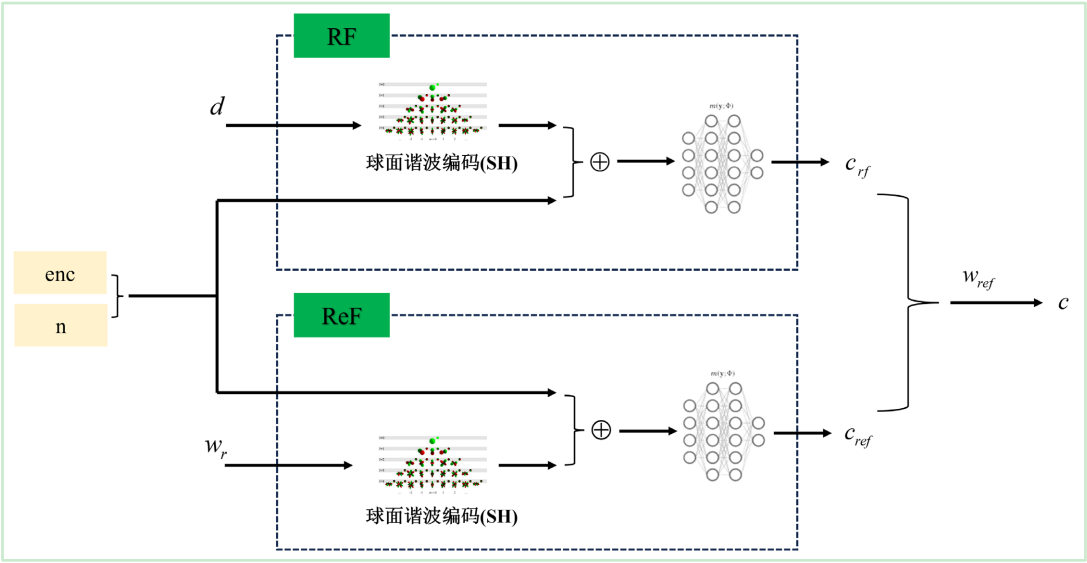

Figure 2. Diagram of the hybrid radiance-reflection field network (appearance network) architecture
图 2. 混合辐射场与反射场网络(外观网络)结构图

本工作对于外观网络采用一种权重衡量辐射场(Radiance Field, RF)和反射场(Reflection Field, ReF)的混合表达方式。对于辐射场，如图 2，实际上与 Instant-NGP [9]等工作的外观表达类似，其采用球面谐波函数对视角方向 d 进行编码，并将编码后的特征与几何特征 enc 和法向量 n 进行联合，然后输入到 2 层的 MLP 中进行漫反射颜色的估计，具体的表示为：

$$c_{rf} = \mathrm{MLP}\left(\left[\mathrm{SH}\left(d\right), enc, n\right]\right) \tag{7}$$

对于反射场，其输入为该采样点的反射方向 ω_r。由图 1 所示，这里使用几何网络估计的法向量 n 和给定的视角方向 d 对反射方向进行计算，表达式为：

$$\omega_r = 2(\omega_o \cdot n)n - \omega_o \tag{8}$$

与 Ref-NeRF [5]等工作类似，通过重参数化外观网络，将反射方向作为输入来表达高频变化的信号。但是，等先前工作[5]不同的是，本方法不采用集成方向编码(Integrated Directional Encoding, IDE)来表达高频信号，这是由于 IDE 估计的材质粗糙度(Roughness)作为输入对高阶球面谐波的编码进行控制，而粗糙度的估计往往与实际有很大的偏差。本工作仅简单采用 4 阶的球面谐波对反射方向的光照信号进行表示。通过实验发现，这种方式能够在拟合隐式表面时产生更加平滑的表面。所以，这里反射场(Reflection Field, ReF)的表示为：

$$c_{ref} = \mathrm{MLP}\left(\left[\mathrm{SH}(\omega_r), enc, n\right]\right) \tag{9}$$

由图 1 所示，本工作通过将球面谐波编码的特征与采样点三维坐标同时输入到 MLP 中进行权重 w 的估计，其用于衡量漫反射颜色 c_{rf} 和光泽颜色 c_{ref} 对当前点颜色的贡献，通过反向传播对权重进行优化。其具体的表达如下：

$$c = w \cdot c_{ref} + (1-w) \cdot c_{rf} \tag{10}$$

3.4. 基于反射得分的渲染损失

实际上，高光反射的物体表面会破坏多视角一致性的假设，使得在计算渲染损失的过程中由于需要最小化与真实像素的相似性而产生伪影。在先前工作中，一般的思路为建立一个反射评分用于衡量该像素点存在反射的可能性，但这种思路仅针对于单一像素，无法很好地联合多个视角对同一个物体表面的像素点进行联合考虑。本工作尝试引入一种类似 Ref-NeuS [10]中的反射得分(Reflection Score)方法。这里并不是建模每个像素存在反射的可能性或不确定性(Uncertainty)，而是通过显式地将物体表面的点反投影回每一个视角，通过比较多个视角的投影坐标上的像素之间的马氏距离(Mahalanobis Distance)来计算出该物体表面点的反射得分。考虑到物体遮挡关系的原因，这里的反射得分通过比较该表面点与每个相机的距离和每个相机到最近表面点的距离来判断遮挡关系，整体的反射得分表示为：

$$\bar{\beta}_i^2(r) = \gamma \frac{1}{\sum_{j=1}^{N} \upsilon_j} \sum_{j=1}^{N} \upsilon_j Mdis \tag{11}$$

$$Mdis = \sqrt{\left(\bar{C}_i(r) - \bar{C}_j(r)\right)^T \sum^{-1} \left(\bar{C}_i(r) - \bar{C}_j(r)\right)} \tag{12}$$

这里的 υ 表示表面点 x^* 相对于每个相机视角的可见性，$\bar{C}$ 表示为表面点 x^* 反投影到对应视角图像空间的像素值，其计算与 Ref-NeuS 一致。但是，表面点 x^* 的计算过程与 Ref-NeuS 略有不同，如图 3 所示，由于本方法采用类似 Neuralanglo [13]的占据栅格来过滤无物体表面相交的空区域，所以在计算光线上每个采样点符号距离值时的计算量更小。表面点 x^* 的计算表达式如下：

$$T_i = \left\{ x_j \,\middle|\, f\left(\gamma\left(x_j\right)\right) \cdot f\left(\gamma\left(x_{j+1}\right)\right) < 0 \right\} \tag{13}$$

$$\hat{S}_i = \left\{ x \,\middle|\, x = \frac{f\left(\gamma\left(x_j\right)\right)x_{j+1} - f\left(\gamma\left(x_{j+1}\right)\right)x_j}{f\left(\gamma\left(x_j\right)\right) - f\left(\gamma\left(x_{j+1}\right)\right)}, x_j \in T_i \right\} \tag{14}$$

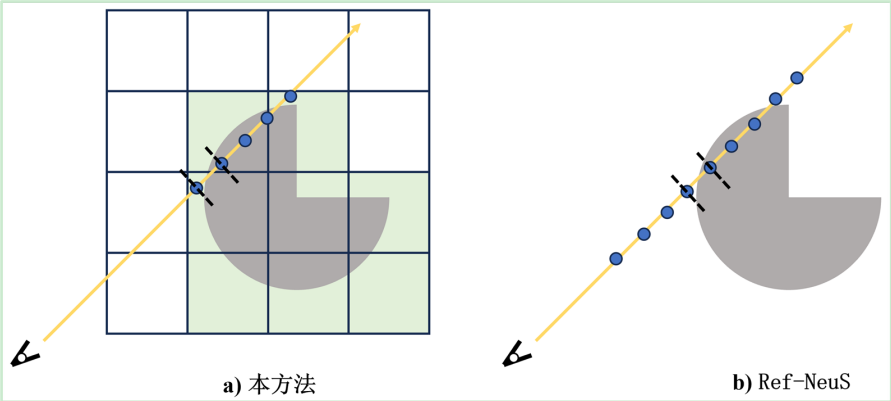

Figure 3. Diagram of surface points calculation comparison
图 3. 表面点集合计算对比图

如上公式，$\hat{S}_i$ 表示第 i 条光线与物体表面相交的点的集合，相交点的计算是采用两侧异号的符号距离值，通过线性插值得到相交点。由于光线与物体会出现多个相交点，这里取距离相机原点最近的相交点，表达式为：

$$x_i^* = \arg\min \mathcal{D}\left(x, o_i\right) \tag{15}$$

这里的 $\mathcal{D}\left(\cdot, o_i\right)$ 为距离计算函数。实际上，本方法会先对可见性 υ 进行计算，然后再对每一个可见的光线进行马氏距离的计算，这样的方式可以减少对无可见光线对图像空间进行反投影的计算开销。如图3，这里将每个视角做光线步进时与物体最近的相交点与视角相机原点 o_i 的距离记为 d_j，而该相机原点到可见性计算的物体表面点 x_i^* 的距离记为 d_j^*，可见性的表示为：

$$\upsilon_j = \mathrm{II}\left(d_j^* \le d_j\right) \tag{16}$$

3.5. 正则化

经过实验观察，直接使用上述几何网络通过有限差分法(Finite Difference Method)求导得到的法向量 n 在拟合高亮物体时会产生很多噪声，整体的法向量并不平滑。为了缓解这些噪声对隐式表面拟合的影响，这里引入了两个对于法向量的损失函数：

$$\mathcal{L}_n = \sum_i w_i \left\| \mathrm{n} - \hat{\mathrm{n}} \right\|^2 \tag{17}$$

这里 $\mathcal{L}_n$ 中的 $\hat{\mathrm{n}}$ 为几何网络估计的法向量，而 n 为符号距离场(SDF)进行有限差分求导得到的法向量。由于几何网络中嵌套了 MLP，其可以很好地为估计的法向量 $\hat{\mathrm{n}}$ 提供一定的平滑性，从而一定程度上影响了有限差分求导的法向量 n 的平滑性。同时，为了保证法向量的朝向正确，这里引入了类似 Ref-NeRF 的法向量惩罚，表示为：

$$\mathcal{L}_o = \sum_i w_i \max\left(0, \mathrm{n} \cdot d\right)^2 \tag{18}$$

上述的 w_i 均为该点在体渲染方程中的权重，其衡量该点在整条光线上积分的贡献程度。除此之外，本方法采用先前工作[1]对符号距离场的约束，即 Eikonal 正则化项 $\mathcal{L}_{eik}$。具体表示如下，其中 P 为所有采样点的总数。

$$\mathcal{L}_{eik} = \frac{1}{P}\sum_{i=1}^{P}\left(\left|\nabla f\left(x_i\right)\right| - 1\right)^2 \tag{19}$$

同时，对于颜色损失 $\mathcal{L}_{color}$ 的考虑，本方法采用 L1 损失，而不是公式 3 的 L2 损失，并且通过上述的公式来缓解优化过程中多视角不一致导致的噪声引入，具体的表述如下：

$$\mathcal{L}_{color}=\sum_{r\in\mathcal{R}}\frac{\left|\hat{C}(r)-C(r)\right|}{\bar{\beta}^2(r)} \tag{20}$$

所以，本方法总体的损失函数可以表示为如下：

$$\mathcal{L}=\mathcal{L}_{color}+\lambda_1\mathcal{L}_{eik}+\lambda_2\mathcal{L}_n+\lambda_3\mathcal{L}_o \tag{21}$$

4. 实验

4.1. 实验设置

4.2.1. 数据集

为了评估本文方法的有效性，我们采用了两种公开的高光物体数据集，分别为 Shiny Blender 数据集和 NeRO [12]中提供的 Glossy Synthetic 数据集。同时，本次实验也对 NeRF Synthetic 数据集中包含高光反射的物体进行了实验。对于 Shiny Blender 数据集，其包含了 6 个物体，每个物体包含多个视角的图像、对应视角的法向量，以及提供了 JSON 文件形式保存的对应视角相机参数。对于 Glossy Synthetic 数据集，其包含 8 个具有高光反射的物体，每个物体文件中包含多视角的图像、对应的深度图和相机参数。原本的数据集并进行划分，本实验对其进行手动划分，以 3:1 的比例，3 份为训练集，1 份为验证测试集。

4.2.2. 基准模型

本文工作选取了多个模型进行新视角合成任务和隐式表面重建任务的对比。对于新视角合成任务，实验选择了在 Shiny Blender 数据集，与三个针对高光物体的模型进行对比，有 Ref-NeRF [5]、ENVIDR [11]和 NeRO [12]。同时，本工作与两种不同渲染方式的模型进行了高光反射的外观拟合的对比，分别为 Nvdiffrec [14]和 NvdiffrecMC [15]，这个模型为均为渲染方程求解范式的工作。对于隐式表面重建任务，本实验选择了两个基准模型，分别为 Ref-NeuS [10]和 NeRO，两个模型均为针对高光反射物体的隐式表面重建工作。

4.2.3. 实现细节

本工作的网络结构主要分为几何网络与外观网络。对于几何网络中的多分辨哈希编码，其编码层级(Level)设置为 16 层，哈希表的大小为 2^{19}，起始分辨率设置为 32，每个层级的哈希特征大小设置为 2。这里的多分辨率哈希编码设置与 Neuralangelo [13]类似，使用渐进式编码的方式从粗到细激活哈希特征，起始层级大小为 4 层，当迭代数进行到 5000 次时开始激活哈希编码，激活的方式与 Neuralangelo 一致。几何网络中的 MLP 为 1 层，隐藏层大小为 64。对于外观网络，辐射场和反射场的球面谐波编码均设置为 4 阶，MLP 层数均为 2。对于整体网络的优化，前 5000 次迭代采用常数因子的学习率衰减方式，之后采用指数衰减，优化器采用 AdamW。几何网络和外观网络的学习率设置为 0.01。对于损失函数的超参数设置，λ_1 为 0.1，λ_2 为 0.0001，λ_3 设置为 0.001。本次实验采用单张 NVIDIA 3090 GPU。

4.2. 实验结果

本节主要对三种上述提到的数据集与基准模型进行对比实验。同时，在章节 4.2.1 中对本文方法进行消融实验。对于新视角合成任务，主要的定量评价标准采用峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和结构性相似度(Structural Similarity, SSIM)。对于隐式表面重建任务，这里采用倒角距离(Chamfer Distance, CD)作为定量评价标准。

4.2.1. 对比实验

对于新视角合成任务，本实验尝试在Shiny Blender数据集和NeRF Synthetic数据集上进行对比实验。从表 1 可以看出，本方法在 Shiny Blender 数据集中的大部分高光物体上都取得较优的结果。虽然，Ref-NeRF [5]和 ENVIDR [11]对于高光物体有的外观拟合有很好的表现，但是这两个工作需要较长时间的训练，均在 6 小时以上。而本方法达到表 1 中的结果仅需要小于 50 分钟的训练时长，并且在各个物体的高光反射拟合都表现稳定，平均值优于对比的方法。表格中加粗的结果表示最优结果，带下划线的结果为次优结果。

Table 1. Comparison with baseline of Shiny Blender dataset
表 1. Shiny Blender 数据集对比实验结果

	Car	Ball	Helmet	Teapot	Toaster	Coffee	平均值
				PSNR ↑			
Nvdiffrec	27.98	21.77	26.97	40.44	24.31	30.74	28.70
NvdiffrecMC	25.93	30.85	26.27	38.44	22.18	29.60	28.88
Ref-NeRF	**30.82**	**47.46**	29.68	47.90	25.70	34.21	35.97
ENVIDR	29.88	41.03	**36.98**	46.14	26.63	**34.45**	35.85
NeRO	26.88	33.66	29.59	40.29	**27.31**	33.76	31.91
本方法	28.95	46.67	34.58	**48.07**	26.96	33.20	**36.41**
				SSIM ↑			
Nvdiffrec	0.963	0.858	0.951	0.996	0.928	0.973	0.945
NvdiffrecMC	0.940	0.940	0.940	0.995	0.886	0.965	0.944
Ref-NeRF	0.955	0.995	0.958	0.998	0.922	0.974	0.967
ENVIDR	**0.972**	**0.997**	0.993	**0.999**	**0.955**	0.984	**0.983**
NeRO	0.949	0.974	0.971	0.995	0.929	0.962	0.963
本方法	0.966	0.995	**0.994**	0.998	0.932	**0.985**	0.978

Table 2. Comparison with baseline of NeRF Synthetic dataset
表 2. NeRF Synthetic 数据集对比实验结果

	Drums	Lego	Materials	Hotdog	Ficus	平均值
			PSNR ↑			
NeRF	25.01	32.54	**29.62**	36.18	30.13	30.70
Ref-NeRF	25.43	**35.10**	27.10	**37.04**	28.74	30.68
ENVIDR	22.99	29.55	29.52	31.44	26.60	28.02
本方法	**26.03**	30.40	29.42	35.93	**32.52**	**30.86**

对于 NeRF Synthetic 数据集，本实验选取了部分包含高光反射的物体进行简单对比。如表 2，本方法在 PSNR 和 SSIM 两项评价标准上均取得较好的结果，并且在平均值上优于其他方法。

对于隐式表面重建任务，本实验针对 Glossy Synthetic 数据集进行了验证。由于 Glossy Synthetic 数据集上的物体存在较多的高光反射，物体表面的材质多为金属材质，这给基于体渲染管线的隐式表面重建方法带来了巨大的挑战。本实验选择了两个针对高光反射物体的先进基准模型，为 Ref-NeuS [10]和 NeRO [12]。对于 NeRO，这里并不进行第二阶段的材质优化，因为材质优化阶段并不会影响到隐式表面的质量，所以这里仅进行第一阶段的形状优化。为了更好地说明本方法的优越性，本实验通过控制模型 Ref-NeuS 和 NeRO 的训练时间在 2 小时内，即迭代方法控制在 30000 次，比较其倒角距离的结果。同时，本实验也与迭代次数为 200000，这里标记为(20w step)的 NeuS [1]、NeRO 进行对比。从表 3 可以看出，本方法

在高光物体上的隐式表面重建质量超越其他对比的方法。同时，本方法达到表 3 结果的训练时长仅需 50 分钟，但与本工作结果接近的 NeRO，其 20 万的迭代次数在 NVIDIA 3090 显卡上使用大约 9 小时。如图 4，本工作在 3 万次迭代的隐式表面重建表现可以很好地恢复出高亮区域几何细节。

Table 3. Experiment results in chamfer distance (CD) on the Glossy Synthetic dataset
表 3. Glossy Synthetic 数据集倒角距离(CD)实验结果

	NeuS (20w step)	Nvdiffrec	NvdiffrecMC	Ref-NeuS (3w step)	NeRO (3w step)	NeRO (20w step)	本方法 (3w step)
Angel	0.0035	0.0056	0.0034	0.0124	0.0058	0.0034	**0.0029**
Horse	0.0053	0.0077	0.0052	0.0067	0.0071	0.0049	**0.0039**
平均值	0.0044	0.0067	0.0043	0.0096	0.0065	0.0042	**0.0034**

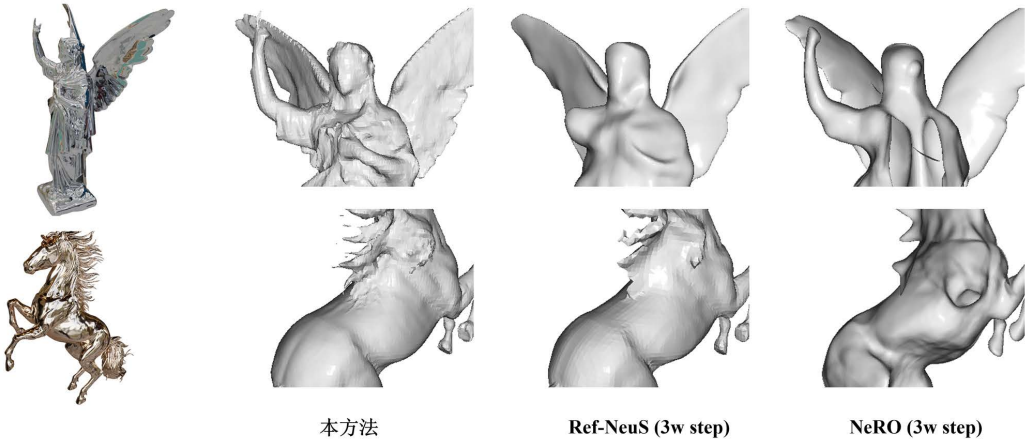

Figure 4. Diagram of reconstruction surface quality comparison on the Glossy Synthetic dataset
图 4. Glossy Synthetic 数据集重建表面质量对比图

4.2.2. 消融实验

为了更好地评估本方法的有效性，该实验针对本文提出了三个部分进行消融实验，分别为：(1) 混合辐射场与反射场(Hybrid Radiance-Reflection Field)，鉴于先前工作均使用辐射场作为外观表达，这里针对反射场进行消融实验，将去掉反射场记为(-ReF)；(2) 这里将去掉法向量正则化，同时去掉反射场的消融实验标记为(-NR & -ReF)；(3) 在消融实验(-NR & -ReF)的基础上，针对本文引入的缓解多视角不一致的渲染损失，这里使用 NeuS 基准模型使用的 L1 颜色损失将其替换，标记为(-NR & -ReF& -RL)。本消融实验选取了 Shiny Blender 数据集和 Glossy Synthetic 数据集中的两个物体进行实验，具体的定量评价如表 4。

Table 4. Results of ablation experiments
表 4. 消融实验结果

	Horse (Glossy Synthetic)	Helmet (Shiny Blender)
	PSNR ↑	PSNR ↑
本方法	24.29	34.57
-ReF	22.90	30.10
-NR & -ReF	22.65	29.77
-NR & -ReF& -RL	22.83	30.60

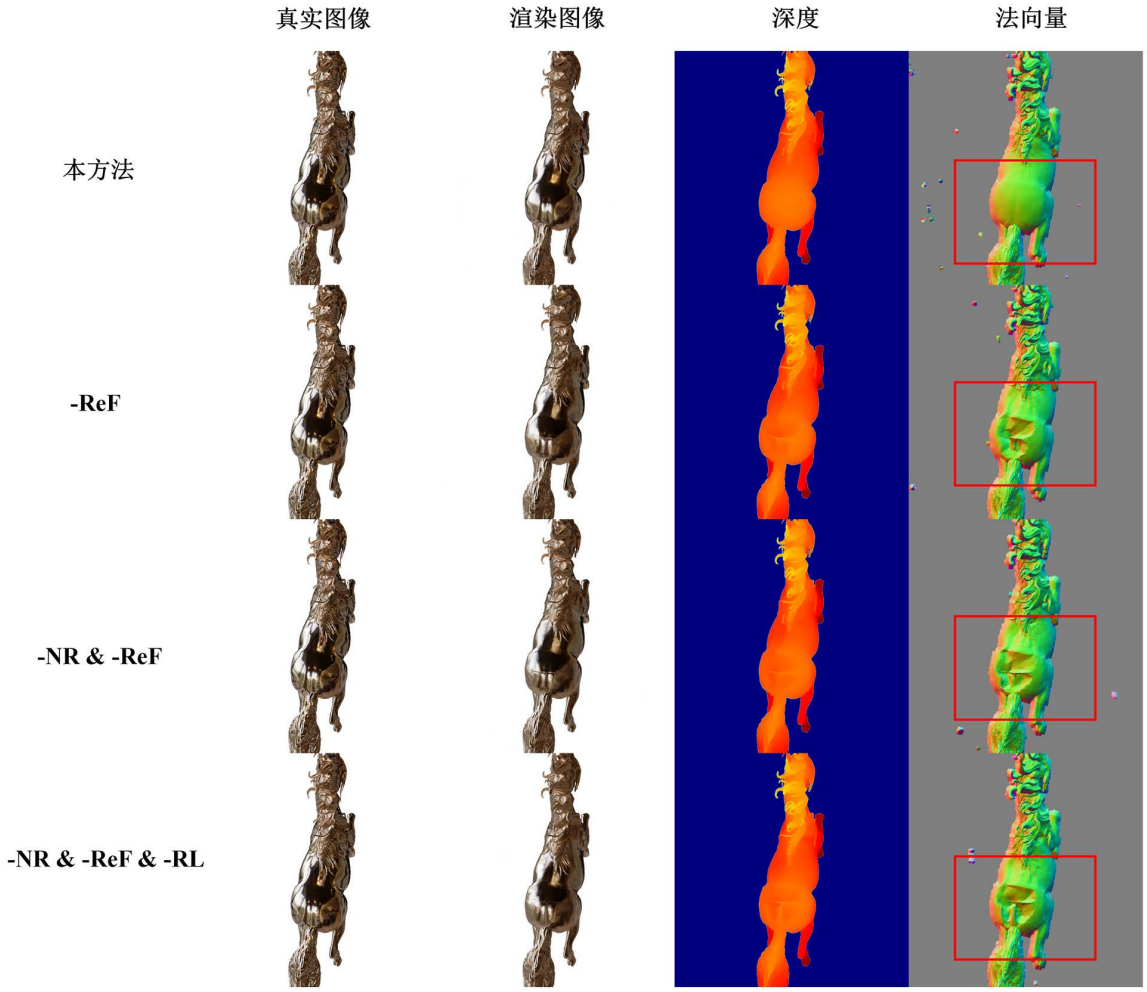

Figure 5. Diagram of geometry quality comparison in ablation experiments
图 5. 消融实验几何质量对比图

从表 4 和图 5 可以看出，本文提出的混合辐射场与反射场的方法可以较好地拟合高光物体的外观。当去掉反射场(-ReF)时，两个数据集的物体在新视角合成任务上的峰值信噪比均有明显的下降，说明反射场能够很好地辅助网络进行高光反射的拟合。对于 Glossy Synthetic 数据集，如图 6 所示，混合方法的外观表示可以很好地辅助隐式表面的拟合，使得在高光反射区域能够拟合出高质量的表面。

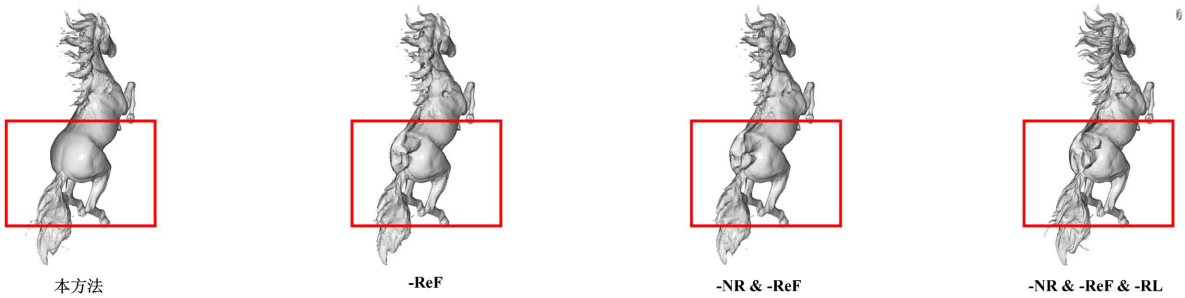

Figure 6. Diagram of 3d mesh visualization of the ablation experiments
图 6. 消融实验三维网格可视化图

5. 结论

本次工作，我们引入了一种混合辐射场与反射场的外观表达方式，并且通过结合一种改进的渲染损失来缓解高光反射带来的多视角不一致性，使得整体网络对于高光物体的隐式表面拟合更加稳定。同时，由于针对高光反射物体的隐式表面拟合对于法向量的质量较为敏感，本工作引入了两种法向量的正则化约束，使得法向量的估计在优化过程中避免陷入局部最小值，同时缓解了混合神经场梯度带来的噪声。该方法在高光物体的隐式表面恢复有着较好的表现，并且训练时长仅为分钟级。同时，其对于高光反射物体的新视角合成任务上也有较好的结果。但是，本文的方法仍然存在不足，比如针对大规模场景中存在的镜面反射或高光反射，该方法难以解决，我们将其作为未来工作的探索。

参考文献

[1] Wang, P., Liu, L., Liu, Y., *et al*. (2021) Neus: Learning Neural Implicit Surfaces by Volume Rendering for Multi-View Reconstruction. *Advances in Neural Information Processing Systems*, **34**, 27171-27183.

[2] Yariv, L., Gu, J., Kasten, Y., *et al*. (2021) Volume Rendering of Neural Implicit Surfaces. *Advances in Neural Information Processing Systems*, **34**, 4805-4815.

[3] Oechsle, M., Peng, S. and Geiger, A. (2021) Unisurf: Unifying Neural Implicit Surfaces and Radiance Fields for Multi-View Reconstruction. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 5589-5599. https://doi.org/10.1109/ICCV48922.2021.00554

[4] Yariv, L., Kasten, Y., Moran, D., *et al*. (2020) Multiview Neural Surface Reconstruction by Disentangling Geometry and Appearance. *Advances in Neural Information Processing Systems*, **33**, 2492-2502.

[5] Verbin, D., Hedman, P., Mildenhall, B., *et al*. (2022) Ref-Nerf: Structured View-Dependent Appearance for Neural Radiance Fields. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 5481-5490. https://doi.org/10.1109/CVPR52688.2022.00541

[6] Müller, T., Evans, A., Schied, C., *et al*. (2022) Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Transactions on Graphics* (*ToG*), **41**, 1-15. https://doi.org/10.1145/3528223.3530127

[7] Park, J.J., Florence, P., Straub, J., *et al*. (2019) Deepsdf: Learning Continuous Signed Distance Functions for Shape Representation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 165-174. https://doi.org/10.1109/CVPR.2019.00025

[8] Mescheder, L., Oechsle, M., Niemeyer, M., *et al*. (2019) Occupancy Networks: Learning 3d Reconstruction in Function Space. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 4460-4470. https://doi.org/10.1109/CVPR.2019.00459

[9] Mildenhall, B., Srinivasan, P.P., Tancik, M., *et al*. (2021) Nerf: Representing Scenes as Neural Radiance Fields for View Synthesis. *Communications of the ACM*, **65**, 99-106. https://doi.org/10.1145/3503250

[10] Ge, W., Hu, T., Zhao, H., *et al*. (2023) Ref-Neus: Ambiguity-Reduced Neural Implicit Surface Learning for Multi-View Reconstruction with Reflection. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Paris, 01-06 October 2023, 4251-4260. https://doi.org/10.1109/ICCV51070.2023.00392

[11] Liang, R., Chen, H., Li, C., *et al*. (2023) Envidr: Implicit Differentiable Renderer with Neural Environment Lighting. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Paris, 01-06 October 2023, 79-89. https://doi.org/10.1109/ICCV51070.2023.00014

[12] Munkberg, J., Hasselgren, J., Shen, T., *et al*. (2022) Extracting Triangular 3d Models, Materials, and Lighting from Images. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June 2022, 8280-8290. https://doi.org/10.1109/CVPR52688.2022.00810

[13] Hasselgren, J., Hofmann, N. and Munkberg, J. (2022) Shape, Light, and Material Decomposition from Images Using Monte Carlo Rendering and Denoising. *Advances in Neural Information Processing Systems*, **35**, 22856-22869.

[14] Liu, Y., Wang, P., Lin, C., *et al*. (2023) Nero: Neural Geometry and Brdf Reconstruction of Reflective Objects from Multiview Images. *ACM Transactions on Graphics* (*TOG*), **42**, 1-22. https://doi.org/10.1145/3592134

[15] Li, Z., Müller, T., Evans, A., *et al*. (2023) Neuralangelo: High-fidelity neural surface reconstruction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, 17-24 June 2023, 8456-8465. https://doi.org/10.1109/CVPR52729.2023.00817