

高效采样的神经隐式表面重建研究

何思源, 刘兴林

五邑大学智能制造学部, 广东 江门

收稿日期: 2024年2月20日; 录用日期: 2024年3月21日; 发布日期: 2024年3月29日

摘要

近几年, 神经隐式表面在三维表达上展现出了巨大的潜力。许多先前工作扩展神经辐射场(Neural Radiance Field, NeRF)的训练管线, 通过在光线上采样点, 然后建模出一种概率分布与隐式曲面建立联系, 从而可以使得隐式曲面拟合。然而, 这些工作要么需要过长的训练时间, 要么需要在特定的加速结构下处理较多的采样点。本文工作, 提出联合占据栅格(Occupancy Grid)和一种非线性函数来指导采样的策略, 使得可以实现在合理的时间内, 处理较少的采样点并达到较好的物体表面重建效果。同时, 这里引入了渐进式训练方式使得模型拟合更加鲁棒。通过在真实数据集和合成数据集的实验表明, 我们的方法在物体表面恢复任务上能取得较好的效果, 并且在新视角合成任务上也能取得出色的表现。

关键词

神经符号距离场, 隐式曲面, 表面重建

Research on Efficient Sampling of Neural Implicit Surface Reconstruction

Siyuan He, Xinglin Liu

Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen Guangdong

Received: Feb. 20th, 2024; accepted: Mar. 21st, 2024; published: Mar. 29th, 2024

Abstract

In recent years, neural implicit surfaces have demonstrated significant potential in 3D representations. Many previous works extend the training pipeline of Neural Radiance Field (NeRF) by modeling a probability distribution to establish a connection with implicit surface using sampled points along rays, thus enabling the good fitting of implicit surface. However, these works either require excessively long training times or deal with a large number of sampled points under specific acceleration structures. In this work, we propose a strategy that combines Occupancy Grid

and a nonlinear function to guide the sampling process, allowing for the achievement of better object surface reconstruction with fewer sampled points in a reasonable amount of time. Additionally, we introduce a progressive training approach to enhance the robustness of the model fitting. Experiments results on real world and synthetic datasets demonstrate the effectiveness of our method in object surface reconstruction tasks and outstanding performance in novel view synthesis tasks.

Keywords

Neural Signed Distance Field, Implicit Surface, Surface Reconstruction

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

从多视角图像中重建三维物体是计算机视觉的一个重要研究热点。传统的运动恢复结构或多视角立体匹配方法做三维重建，需要从多视角图像中恢复出稀疏点云，再到点云配准，恢复稠密点云，异常点处理，点云重建等一系列步骤。整体的流程非常长，每一步都会受到误差的影响。近年来，随着神经辐射场(NeuralRadiance Field, NeRF) [1]一系列工作的扩展，许多方法通过建立神经隐式表面与神经辐射场中的概率密度函数或累积密度函数的联系，利用体渲染管线和反向传播完成了神经隐式表面的拟合。

NeuS [2]通过建立符号距离场(Signed Distance Fields)与神经辐射场(NeRF)中权重函数的关系，实现了不同于传统方法管线的多视角表面重建，同期工作还有 volSDF [3]，UniSurf [4]等。NeuS 的采样策略与神经辐射场的策略类似，都是分层采样策略。当进行光线步进时，先在光线上进行均匀采样，再根据计算得到的概率分布函数进行重要性采样。然而，由于 NeuS 使用了 8 层的多层感知机(Multi-Layer Perception, MLP)，单个采样点的符号距离查询计算消耗较大，所以难以进行非常多的采样。Instant-NGP [5]利用了占据栅格(Occupancy Grid)，通过在每个格子中计算占据概率，以二值化数组表示空间是否有物体表面占据，使得在光线步进时可以跳过空白区域，让采样更集中在有效占据的格子中。虽然，Instant-NGP 采用参数化数据结构结合单层 MLP 的方案使得采样点的符号距离查询计算消耗较小，并且占据网格可以辅助采样使得采样点更有效落在有占据的格子中，但占据的格子中采样策略为均匀采样，分辨率较小，采样数量也少时，会导致整体的拟合质量下降。针对上述问题，本文提出了一种新的混合采样策略。通过结合占据网格让采样更好地在有表面占据的区域进行采样，并采用一种分段指数函数来指导采样点在有效区域更好的靠近表面，使得整体的采样更加有效。

2. 相关工作

2.1. 神经隐式表面

隐式表面(Implicit Surface)是一种传统的三维表示方法。随着深度学习的发展，DeepSDF [6]、Occupancy Network [7]等利用神经网络来表达隐式表面的工作开始出现，但多数工作都需要三维监督，通过不同的方式将点云的三维坐标映射到重建的标量场，而本文的工作主要以多视角图像为输入并作为监督信息。DVR [8]、IDR [9]使用基于表面的渲染方法(Surface-based Rendering Method)，通过多视角图像作为输入，输出高质量的几何与外观。受到神经辐射场(NeRF)的启发，NeuS [2]、volSDF [3]等工作，通过建立符号距离

场(SDF)与体渲染公式中的概率密度函数或累积密度函数的关系, 成功使用基于体渲染的方法进行了隐式表面的估计。为了进一步提高几何估计的质量, PET-NeuS [10]、HF-NeuS [11]等工作相继提出, 提高了几何质量的同时也增加了训练的所需时长, 一般需要将近一到两天的时间。对于训练效率的考虑, NeuS2 [12]从 Instant-NGP [5]中得到了启发, 通过结合多分辨率哈希编码到 NeuS 中, 利用轻量的二阶计算使得输出的几何更加平滑, 并且训练速度非常快, 仅为分钟级。Voxurf [13]的训练同样高效, 通过扩展基于栅格的三维表达, 并引入了多层次的几何特征加强了几何细节的表达。但这些工作的高效都是源于参数化的离散数据结构与少层数的多层次感知机混合的思想, 单个样本基于栅格、基于哈希等结构的查询复杂度仅为 $O(1)$, 配合单层的感知机网络, 其大大降低了对每个样本计算符号距离值的消耗。本文通过探讨采样策略与离散数据结构的配合, 既减小单个样本的计算开销, 又减少不必要采样的计算浪费。

2.2. 高效采样的神经辐射场

由于神经辐射场(NeRF)的训练输入需要从光线步进中进行采样, 而采样的数量和质量一定程度上影响模型训练和推理的速度和质量。所以, 高效的采样策略是非常重要。NeRF [1]和 NeuS [2]采用了从粗到细的分层采样策略, 先对每条光线进行均匀采样, 并计算出累积密度函数, 然后在精细阶段采用逆变换采样方法(Inverse Transform Sampling)来进行重要性采样。但这种采样策略会引入很多无用的采样, 导致计算的浪费。Instant-NGP [5]受到了空间划分的启发, 通过引入占据栅格来划分空间, 以一个二值化数组来标记每个空间是否与物体表面相交, 使得在光线步进的过程中, 可以跳过一些无物体区域, 减少了采样的浪费。NeuS2 [12]、Neuralangelo [14]沿用了这种方法, 使得整体的采样更加有效。但这些工作在占据格子中的采样方法依然为均匀采样, 当使用的占据栅格分辨率较低时, 依然会引入非常多无用样本导致计算的浪费。DoNeRF [15]等方法分析了多种采样策略, 如均匀采样、非线性采样、NDC 空间采样等的有效性, 并引入了一种深度遮挡预测网络来指引采样位置的选择, 从而减少无效采样的计算浪费。但其输入需要 RGBD 图像, 本文工作仅需要 RGB 图像。还有很多工作[16] [17]通过单目深度估计网络来指引采样, 但这既增加了整体模型的参数量, 又受限于深度估计骨架网络的性能。本文工作, 通过利用占据栅格配合一种分段指数函数来指引采样, 在较低的计算开销下使得整体的采样效率增加, 既避免了在无效区域的采样, 又引导了有效区域采样点更加靠近物体表面。

3. 方法

本文工作建立在神经辐射场(NeRF) [1]和 NeuS [2]之上。首先, 本章会先回顾 NeRF 和 NeuS 两个工作。然后, 我们会介绍本工作的整体网络架构, 包括几何网络、外观网络、法向量的计算等。其次, 对于光线上的采样, 我们会分析 NeuS 的分层采样和占据栅格采样, 并提出二阶段采样的方法。最后, 我们会介绍训练过程的正则化项。

3.1. 预备知识

神经辐射场(NeRF)将三维场景表示为一种连续函数, 其利用体渲染可以从任何相机位置渲染出对应视角的图像。给定 N 张不同角度拍摄的图像 $I = \{I_1, I_2, \dots, I_N\}$, 并给定相机的内外参数 $\Pi = \{\pi_1, \pi_2, \dots, \pi_N\}$, 从不同视角图像的像素上投射光线, 然后在光线上进行采样。对于第 i 个采样点 x_i , NeRF 使用神经网络估计出该点的体积密度 σ_i 和采样点颜色 c_i , 即 $F: (x_i, d) \rightarrow (\sigma_i, c_i)$, 这里的 d 为光线的方向, 使用相机内外参数计算得到。然后, 通过经典的体渲染(Volume Rendering) [1]方案对整条光线所有采样点估计的颜色进行积分得到该视角图像的像素颜色 \hat{C} , 其数值积分的近似表达为:

$$\hat{C} = \sum_{i=1}^n T_i \alpha_i c_i, \quad T_i = \exp\left(\sum_{k=1}^{n-1} -\sigma_k \delta_k\right) \quad (1)$$

这里的 T_i 为累积透射率, $\alpha_i = 1 - \exp(-\sigma_i \delta_i)$ 为第 i 段间距的不透明度(Opacity), $\delta_i = t_{i+1} - t_i$ 为相邻采样点的间距, t_i 为从相机原点 o 出发到 x_i 的相对深度值。对于整个工作的优化, 损失函数定义为渲染图像与给定视角图像的颜色损失, 定义为:

$$\mathcal{L} = \sum_x \left\| \hat{C}(x) - C(x) \right\|_2^2 \quad (2)$$

NeuS [2]通过建立符号距离场(SDF)与不透明度 α 的联系, 使得其可以通过 NeRF 的体渲染管线来优化 SDF, 从而实现多视角图像重建隐式表面的任务。SDF 为一种较为常用的隐式表面, 其通过零水平集来表示物体表面 S , 即 $S = \{x \in \mathbb{R}^3 \mid f(x) = 0\}$, 这里的 $f(x)$ 为预测的符号距离值, 使用神经网络来表示。给定采样点 x_i 和符号距离值 $f(x_i)$, 其与不透明度 α 建立的表示为:

$$\alpha_i = \max \left(\frac{\Phi_s(f(x_i)) - \Phi_s(f(x_{i+1}))}{\Phi_s(f(x_i))}, 0 \right) \quad (3)$$

这里的 Φ_s 为 Sigmoid 函数。由于 NeRF 和 NeuS 都采用了较深的多层感知机(MLP)对每一个采样点进行估计, 其计算消耗非常大, 训练速度较慢。Instant-NGP 引入了一种占据栅格(Occupancy Grid)来对空间进行划分, 采用二值化数组来标记每个格子是否与物体表面相交, 从而使得整个过程不需要再无效区域采样, 减少了采样数。同时, 其引入了多分辨率的哈希编码器, 使得对每个采样点的计算消耗大大降低。虽然, Instant-NGP 在新视角合成取得了较好的效果, 但由于其直接对体积密度进行 Marching Cube [18] 而得到三角网格, 使得恢复的几何表面存在很多噪声。

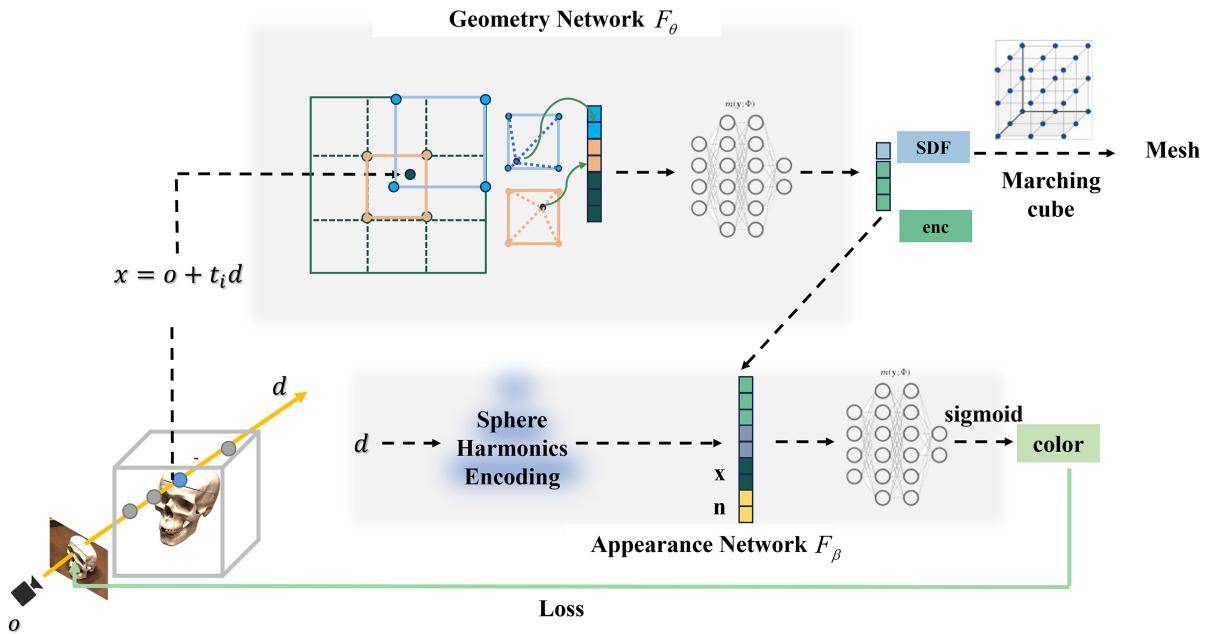


Figure 1. Diagram of the network architecture
图 1. 网络架构图

3.2. 网络结构

我们提出一种结合 NeuS [2]和 Instant-NGP [5]的神经隐式表面重建算法, 通过给定一组环绕物体拍摄的图像和相机的内外参数, 可以高效地重建出静态三维物体表面。整体的方法架构如图 1, 通过从给定

视角的图像像素进行光线投射, 然后从光线中进行采样。与 NeRF [1]、NeuS 和 Instant-NGP 不同, 我们通过结合占据栅格(Occupancy Grid)和一种分段指数函数来指导采样, 具体细节在章节 3.3。对于每一个采样点, 会先通过多分辨率哈希编码 γ 对采样点进行插值, 并对不同尺度插值好的特征进行整合操作, 然后直接送到单层 MLP 中进行符号距离的预测。这里隐式表面的表示为符号距离场(SDF), 表示为:

$$f_{\text{sdf}}(x) = F_{\theta}(\gamma(x)) \quad (4)$$

F_{θ} 为层数为 1 的 MLP。除了输出符号距离值外, 还会输出几何特征, 这里记为 enc 。对于每个采样点颜色的估计, 具体表示如下:

$$c = F_{\beta}(x, n, enc, d) \quad (5)$$

这里的 d 为使用球面谐波(Spherical Harmonics, SH)对视角方向向量 d 编码后的特征, 与 Instant-NGP 的做法一致, 目的是为了使采样点的颜色是视角相关的, 即同一个点, 在不同视角方向上的颜色值不同。 n 为采样点的法向量, 与 NeuS [2]的法向量求解形式不同, NeuS 对采样点的法向量是使用 SDF 的解析梯度(Analytic Gradient), 而本工作使用一种数值方法, 为基于有限差分的 SDF 近似梯度, 主要目的是提供一种更为平滑的法向量。

$$\nabla f_{\text{sdf}}(x) = \frac{f_{\text{sdf}}(x + \epsilon) - f_{\text{sdf}}(x - \epsilon)}{2\epsilon} \quad (6)$$

这里 $n = \nabla f_{\text{sdf}}(x)$, ϵ 为哈希网格的大小。引入 n 作为估计采样点颜色的输入还有一个好处, 就是提供一种几何和外观的解耦[9]。实际上, NeuS2 [12]和 Neuralangelo [14]也是类似的做法, 利用 Instant-NGP 中的多分辨率哈希编码和较少层数的 MLP 对 SDF 进行估计, 由于哈希编码的查询速度快, 加上 MLP 层数较少(1 层), 相对与 NeuS 8 层 MLP 的计算消耗, NeuS2 和 Neuralangelo 的 SDF 计算快了几个量级。然而, NeuS2 和 Neuralangelo 的采样策略与 Instant-NGP 一致, 都为一阶段占据栅格(Occupancy Grid)引导的采样方式, 其在每个占据的格子中依然为均匀采样, 当占据栅格的分辨率较低时, 实际上依然会引入很多无效采样。

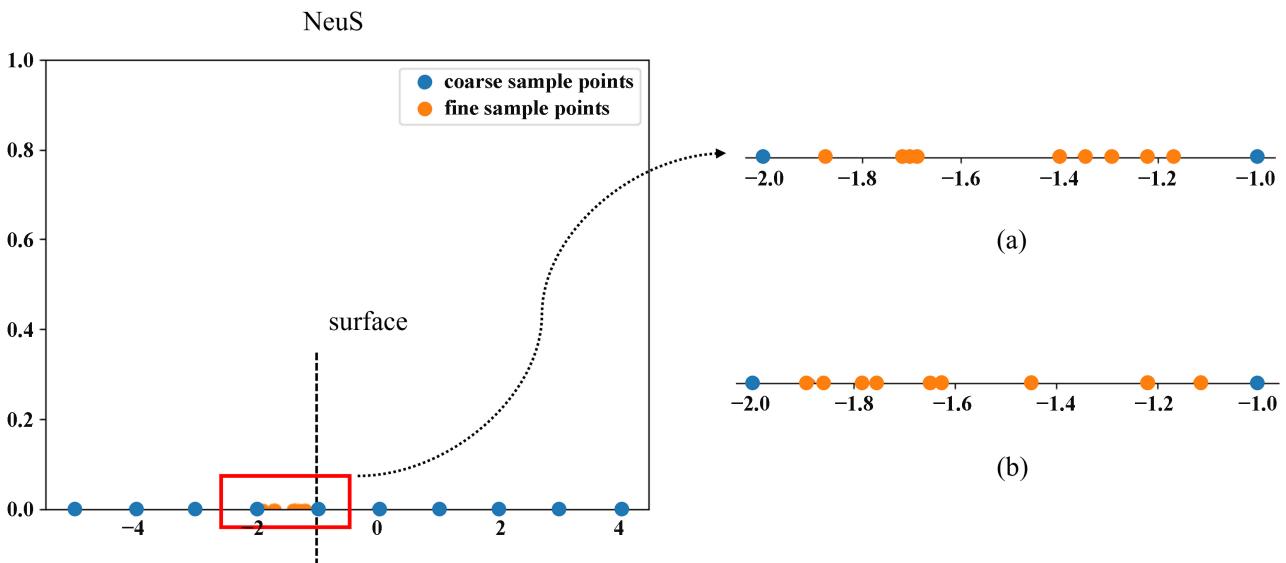


Figure 2. Diagram of the hierarchical sampling strategy of NeuS

图 2. NeuS 分层采样策略图

3.3. 两阶段采样

在公式(1)中, $T_i\alpha_i$ 可以看作整个求和公式中对于采样点颜色 c_i 的权重, 用于衡量光线上每个采样点的贡献程度, 记 $w_i = T_i\alpha_i$ 。对于一个不透明物体, 实际上只有靠近物体表面的采样点才有贡献, 即当采样点靠近物体表面时, 权重 w 的值会趋向于无穷大, 远离物体表面时, 权重 w 的值为 0。

由图 2 所示, 假设对一个单位球(Unit Sphere)进行光线步进, 图的横轴为一维的坐标值, 值为 -1 和值为 1 是单位球的表面边界。以 NeuS 的采样策略, 先对整个空间进行均匀采样, 如蓝色点。然后, 对所有均匀采样点计算 w , 并通过累积权重 w 得到累积密度函数。通过逆向变换采样, NeuS 会做符合 w 分布的随机采样, 具体如图 2(a)。可以看到, NeuS 在精细采样阶段的采样点会随机散落在 w 峰值所在的区间 $[-2.0, -1.0]$, 但并没有对物体表面的偏向性, 即采样点没有偏向于 -1 的位置。并且, 由于随机采样的特性, 每次做逆向变换采样, 精细采样的采样点落实的位置都比较随机, 如图 2(b)。对于 Instant-NGP 的采样策略, 其通过占据栅格(Occupancy Grid)的二值化数组辨别出哪些区域需要跳过, 从而减少像 NeuS 粗阶段均匀采样的计算浪费。但对于一个有占据区域, 依然是均匀采样策略, 无法很好指导采样做得更高效, 如图 3(b-2)。

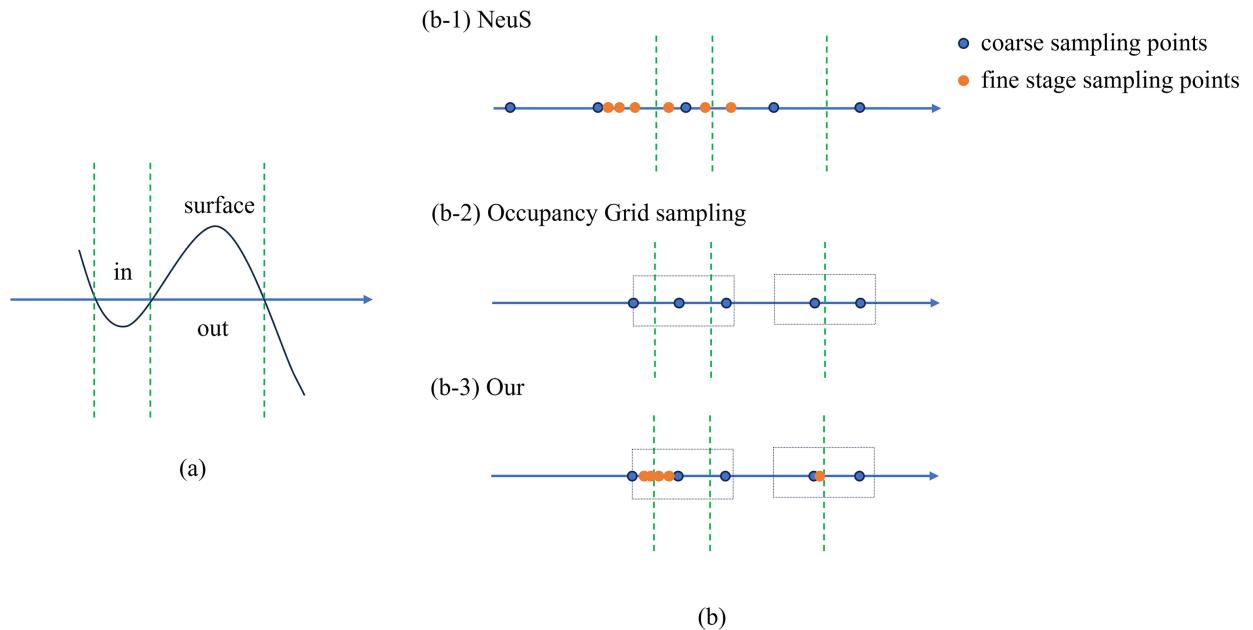


Figure 3. Diagram of the hierarchical sampling strategy comparison
图 3. 采样策略对比图

本文工作采用一种两阶段的采样策略, 如图 3(b-3)。整体步骤为: 1)粗阶段使用基于占据栅格的方式, 使得可以跳过空区域进行采样, 避免采样的无效; 2)精细阶段使用一种非线性函数指引采样的策略, 使得在占据区域的采样点可以更加靠近物体表面。对于精细阶段的非线性函数, 受[19]的启发, 引入一种变换权重函数 \hat{w} , 公式如下:

$$\hat{w}(s) = a \left(\frac{b}{a} \right)^s \quad (7)$$

$$\int_{t_i}^{t_{i+1}} w(t) dt = \int_0^1 \hat{w}(s) ds = \frac{b-a}{\ln b - \ln a} \quad (8)$$

公式(7)中, $a = w(t_i)$ 且 $a \geq 0$, $b = w(t_{i+1})$ 且 $b \leq 1$ 。这里的 w 为粗阶段对采样点计算的权重, 与 NeuS 一致, 即 $w(t_i) = T_i \alpha_i$, a 和 b 为采样间隔端点的权重值。

但与 NeuS 不同的是, 通过引入 \hat{w} , 可以对采样区间进行插值处理, 而不是分段常数。换而言之, 对于一个采样间隔 $[t_i, t_{i+1}]$, NeuS 在间隔内的点的权重均为固定值, 而我们引入的 \hat{w} 可以实现在间隔内插值。对于离散形式的体渲染, 进一步细化整体的概率分布。同时, 公式(7)可以看作通过 a 和 b 的比值自适应地控制区间内点的权重, 当 a 的值远小于 b 值, \hat{w} 随着采样深度 s 的增大, 整体的增大趋势会越来越剧烈, 即越靠近 b 权重值端点的采样点, 其权重越大。当 a 值远大于 b 时, \hat{w} 整体的趋势会往 b 值的端点迅速递减。从而使得, 越靠近大权重值的采样点, 其权重值也相对较大, 越远离越小。进而使得在逆向变换采样的过程中, 可以使精细采样点更靠近表面。

3.4. 正则化

如图 1, 本工作通过从提出的两阶段采样策略从光线上进行采样。每一个采样点输入到网络中预测出符号距离场(SDF)和采样点颜色, 通过损失函数来最小化。总的损失函数 \mathcal{L} 如下:

$$\mathcal{L} = \mathcal{L}_{color} + \lambda_1 \mathcal{L}_{eik} + \lambda_2 \mathcal{L}_{mask} \quad (9)$$

这里的 \mathcal{L}_{color} 为渲染损失(Rendering Loss), 为采样点渲染颜色 $\hat{C}(x)$ 与真实颜色 $C(x)$ 的 L2 损失, 并对光线上全部采样点的 L2 损失进行求和[2]。

$$\mathcal{L}_{color} = \sum_x \left\| \hat{C}(x) - C(x) \right\|_2^2 \quad (10)$$

\mathcal{L}_{eik} 为 Eikonal 正则化项, 用于保证 $f_{sdf}(x)$ 的演化符合符号距离场(SDF)的性质[2]。 $\nabla f_{sdf}(x)$ 为 SDF 的梯度, 使用公式(6)近似求得, 并非使用自动微分求解。

$$\mathcal{L}_{eik} = \sum_x \left(\left\| \nabla f_{sdf}(x) \right\| - 1 \right)^2 \quad (11)$$

除此之外, 我们还使用了 NeuS [2]提供的掩膜损失(Mask Loss), 其定义为每条光线的采样点权重和与数据集提供的掩膜图像的二值交叉熵(Binary Cross-Entropy, BCE), 如公式(12)。

$$\mathcal{L}_{mask} = BCE(M_k, \hat{O}_k) \quad (12)$$

$$\hat{O}_k = \sum_{i=1}^n T_{k,i} \alpha_{k,i} \quad (13)$$

公式(12)中的 M_k 为 0 或 1 值的掩膜, \hat{O}_k 为光线采样点权重和, k 表示第 k 个像素, i 表示第 i 个采样点。最后, 损失函数公式(9)的超参数 λ_1 和 λ_2 在实验中均设置为 0.1。

4. 实验

4.1. 实验设置

4.1.1. 数据集

为了评估本文方法的有效性, 我们采用了两种不同类型的数据集, 分别为合成数据集和真实拍摄的数据集。对于合成数据集, 我们使用 NeRF [1]提供了合成数据集, 其包含 6 个不同的物体, 每个物体里包含不同视角的图像和对应视角的相机参数, 图像分辨率为 800×800 。对于真实数据集, 我们选择了 DTU 数据集。DTU 中包含 15 个场景, 每个场景中有不同视角拍摄的 RGB 图像、对应的物体掩膜(Object Mask)和相机参数, 涵盖了各种不同的材质, 不同拓扑结构的物体。

4.1.2. 基准模型

本文工作选择了三个基准模型进行对比, 分别为 NeuS、Neus-acc 和 Neus-facto。Neus-acc 主要了为体现占据栅格(Occupancy Grid)在表面重建任务中引导采样的作用, 即 NeuS 与 Instant-NGP 的结合。其实现与 NeuS2 [12]类似, 使用多分辨率哈希编码和 1 层 MLP 来估计 SDF。在光线步进时, 会使用占据栅格的二值化数组对无物体相交区域进行跳跃, 从而达到减少无效采样的目的。不使用 NeuS2 的原因是是为了排除过多 GPU 优化技巧的影响, 更专注在比较采样手段上。Neus-facto [20]使用了 Mip-NeRF360 [21]提出了 Proposal Network 来学习采样分布, 其是一种高效的采样手段。

4.1.3. 实现细节

我们使用的多分辨率哈希编码的分辨率一共分 16 级, 从 2^5 到 2^{16} , 每个级别的哈希特征大小设置为 2。几何部分的 MLP 层数为 1 层, 隐含层(Hidden Layer)大小为 64。外观部分的网络 MLP 层数为 2 层, 隐含层大小为 64。我们使用一种混合优化的方法来进行训练, 前 5000 次迭代采用常数因子的学习率衰减, 之后采用指数衰减的方式。优化器使用 AdamW, 对于几何网络(Geometry Network) F_θ 的学习率为 0.001, 外观网络(Appearance Network) F_β 的学习率为 0.001, betas 参数为 0.9 和 0.99。实验所用的 GPU 为单张 V100 32G。

我们采用了一种渐进式的训练方式, 使得模型可以在学习到高频信息的同时, 又减少噪声的产生。通过设定一个更新步数, 这里设定为 1000, 每迭代 1000 次时, 多分辨率哈希编码 γ 的层级会更新, 并激活对应的层级哈希特征, 使得整体的学习为从粗到细的优化, 从而更好地重建出几何表面。

4.2. 实验结果

本文主要在真实数据集(DTU)和合成数据集(NeRF Synthetic)上的检验表面重建的质量和新视角合成的质量。主要的定量评价标准为倒角距离(Chamfer Distance, CD)和峰值信噪比(Peak Singal-to-Noise Ratio, PSNR)。CD 用于评价模型重建的几何质量, PSNR 用于评价模型多视角合成的图像质量。CD 越低, 重建的几何越好。PSNR 越高, 合成的视角质量越好。与先前工作一致[2], 网络估计出的 SDF 会使用 Marching Cube 算法[19]在 $1024 \times 1024 \times 1024$ 分辨率下提取出三角网格进行 CD 的评估。

4.2.1. 真实数据集

对于 DTU 数据集, 倒角距离(CD)的定量评估如表 1, 我们选取了 DTU 数据集中的 14 个场景进行 CD 的评价, 通过控制模型(Neus-acc, Neus-facto, Our)迭代次数为 20,000 次, 可以看出, 我们的方法在大部分物体上取得了较好的结果。从图 4 可以看出, 我们的方法在几何细节上的重建有更好的效果, 这得益于渐进式的训练和两阶段的采样策略。

表 1. Chamfer Distance experiment results of DTU dataset

表 1. DTU 数据集倒角距离(Chamfer Distance, CD)实验结果

Id	24	37	40	55	63	65	69	83	97	105	106	110	114	118
Neus-acc	1.71	1.55	1.35	0.59	1.63	<u>0.86</u>	1.75	<u>0.94</u>	1.69	0.73	0.66	2.27	0.58	2.14
Neus-facto	1.19	1.01	<u>0.97</u>	0.56	1.35	1.04	1.11	1.02	1.53	0.86	0.95	1.96	0.83	0.70
Our	<u>1.08</u>	<u>0.75</u>	1.10	<u>0.38</u>	<u>1.11</u>	0.89	<u>0.63</u>	1.26	<u>1.30</u>	<u>0.72</u>	<u>0.53</u>	<u>1.45</u>	<u>0.39</u>	<u>0.44</u>

对于新视角合成(Novel View Synthesis)任务, 从表 2 可以看出, 我们的方法比 NeuS、volSDF、Neus-acc 有显著的提升, 并在大部分场景中比 Neus-facto 的结果好, 表中最好的结果使用下划线标记。总的来说,

我们的方法对于几何细节的恢复和新视角合成都有一定程度的提高。

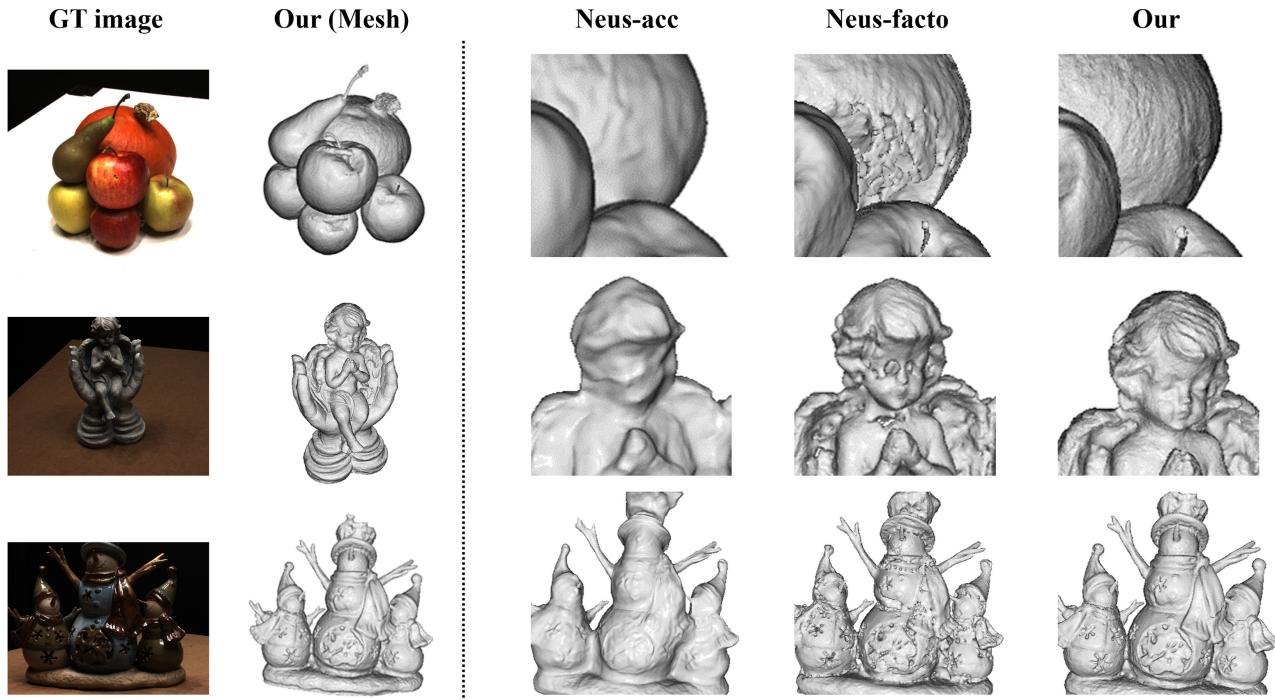


Figure 4. Diagram of the mesh quality comparison on DTU datasets

图 4. DTU 数据集几何质量对比图

Table 2. PSNR experiment results of DTU dataset

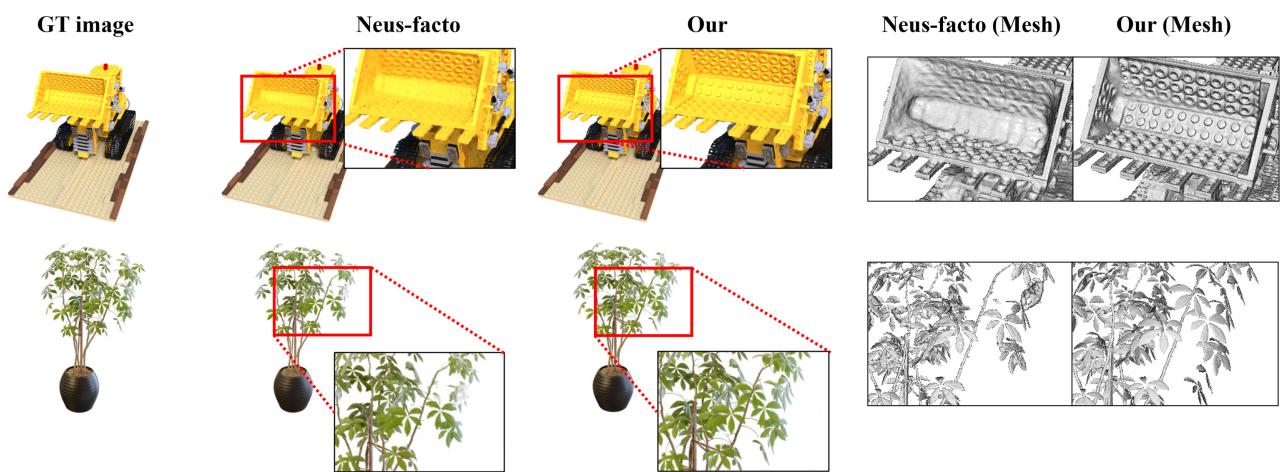
表 2. DTU 数据集峰峰值信噪比(PSNR)实验结果

Id	24	37	40	55	63	65	69
NeuS [2]	28.20	23.64	26.43	25.59	30.61	32.83	29.24
volSDF [3]	26.28	25.61	26.55	26.76	31.57	31.50	29.38
Neus-acc	27.12	24.08	27.56	27.37	30.86	30.74	29.12
Neus-facto	29.97	<u>27.54</u>	29.93	<u>32.27</u>	31.27	29.81	<u>30.51</u>
Our	<u>31.95</u>	26.95	<u>30.78</u>	31.17	<u>35.56</u>	<u>32.79</u>	29.40

4.2.2. 合成数据集

对于合成数据集(NeRF Synthetic), 我们选取了 6 个场景进行几何重建和新视角合成任务的质量评估。从图 5 可以看出, 通过控制迭代次数为 30,000 次, 我们的方法与 Neus-facto 的基于 Proposal Network 采样的方法相比, 无论是生成的三角网格还是新视角合成的质量, 我们的方法都明显优于 Neus-facto。增加迭代次数, Neus-facto 和我们的方法均能在两个任务上得到提升, 但本实验主要为了验证采样策略的有效性, 我们选择了控制在 30,000 次迭代, 训练时长都在 2 小时内。

实际上, 虽然 NeuS 在几何重建任务上取得了一定的效果, 同时也牺牲了其新视角合成任务的性能。如表 3, NeuS 在合成数据集 NeRF Synthetic 上, 新视角合成任务的峰峰值信噪比(PSNR)明显低于 NeRF。但是, 我们的方法不仅在几何质量方面得到了一定的提升, 同时在新视角合成任务中也有很好的效果, 在大部分场景中, 我们的方法都超过了 NeRF。

**Figure 5.** Diagram of the mesh quality comparison on NeRF Synthetic dataset**图 5.** NeRF Synthetic 数据集几何质量对比图**Table 3.** PSNR experiment results of NeRF Synthetic dataset**表 3.** NeRF Synthetic 数据集峰值信噪比(PSNR)实验结果

	Lego	Chair	Ficus	Mic	Ship	Material
NeRF	32.54	33.0	30.15	32.91	28.34	<u>29.62</u>
Neus	29.85	27.95	25.79	29.89	25.46	29.36
Neus-facto	27.56	30.54	25.88	21.67	21.82	-
Our	<u>33.44</u>	<u>34.34</u>	<u>32.92</u>	<u>33.85</u>	<u>29.86</u>	28.78

Table 4. PSNR experiment sampling strategy ablation results of DTU dataset**表 4.** DTU 数据集采样策略消融实验的峰值信噪比(PSNR)实验结果

Id	55	63	65	69	83	97	105	106	110	114	118
Model	25.67	33.94	31.45	27.40	28.29	24.70	29.46	26.19	27.30	28.84	30.73
+ PT	26.14	34.75	31.71	27.30	30.66	25.19	30.36	29.38	27.51	29.34	31.46
+ PTS	<u>31.88</u>	<u>35.94</u>	<u>32.72</u>	<u>29.17</u>	<u>31.81</u>	<u>26.77</u>	<u>32.30</u>	<u>32.25</u>	<u>28.83</u>	<u>31.93</u>	<u>33.59</u>

4.2.3. 消融实验

为了更好地评估我们的方法, 这里对提出的两阶段采样策略和渐进式训练进行消融实验。本次消融实验分别对两种采样数进行实验, 分别为 128 个采样点和 1024 个采样点。

针对 DTU 数据集中选取的 10 个场景, 表 4 为每条光线 128 个采样点进行对比的 PSNR 结果。我们把去掉渐进式训练(Progressive Training, PT)和提出的采样策略标记为 Model。Model 的采样策略为基于占据栅格(Occupancy Grid)的采样。表中的(+PT)表示为 Model 添加了渐进式训练。(+PTS)表示 Model 添加渐进式训练和提出的采样策略, 即本文提出的方法。表中每个方法的训练迭代次数为 30,000 次, 最好的结果均用下划线标记, 可以看出, 当每条光线采样的点数为 128 时, 渐进式的训练方式结合本文提出的两阶段采样策略对于新视角合成任务有显著的提升。从图 7 可以看到, 对于 128 次采样, 结合渐进式训练和我们的采样策略(Model + PTS)对于纹理细节、高光反射等恢复有更好的效果。

如图 6 所示, 当对每条光线进行 1024 次采样, 使用两阶段采样策略的标记为(our sampling), 仅采

用占据栅格的采样标记为(w/o our sampling)。可以看出, 当对光线采样点足够多时, 我们的方法对于几何重建依然有一定程度的提升。

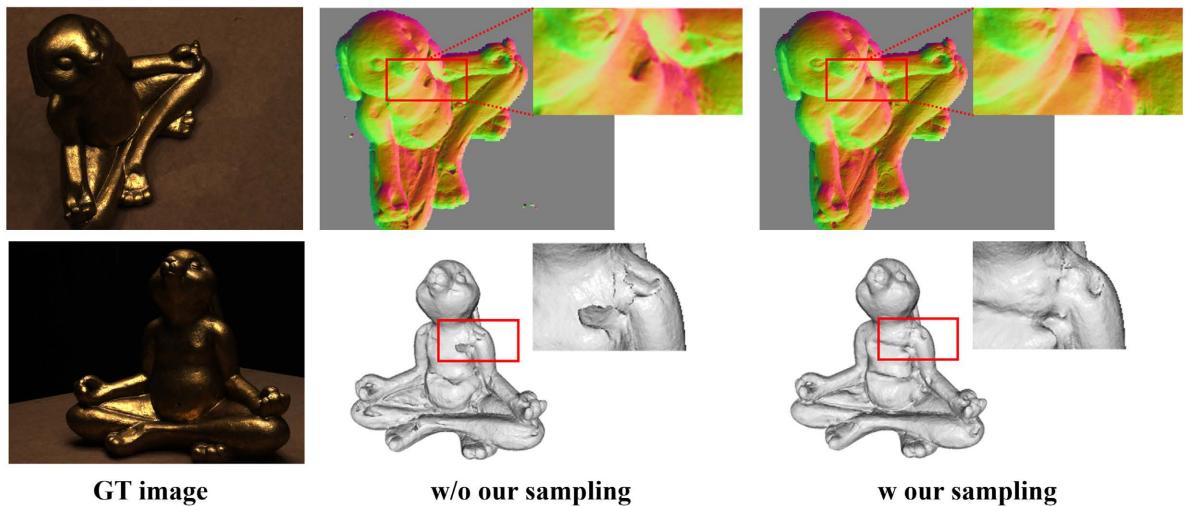


Figure 6. Diagram of the 1024 sampledpoints strategy ablation comparison
图 6. 1024 采样点消融实验对比图

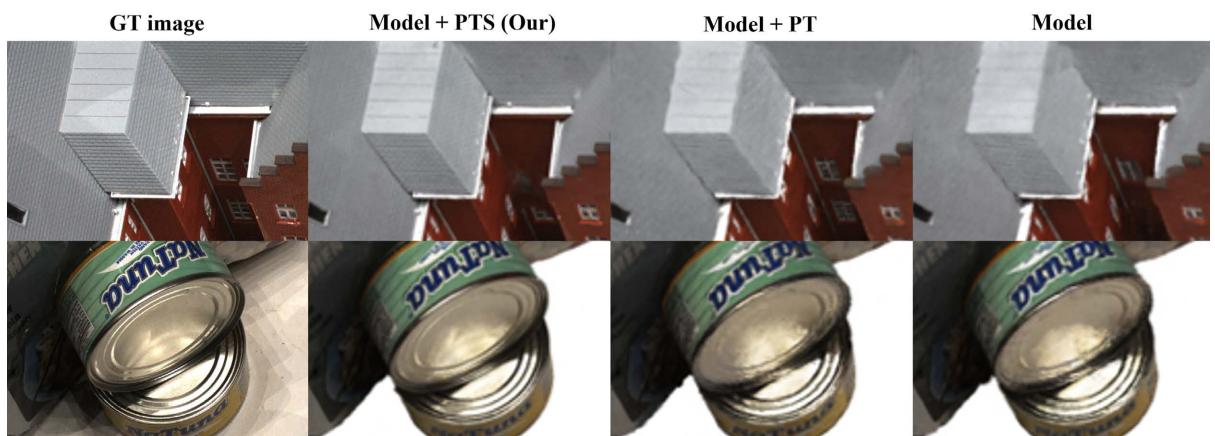


Figure 7. Diagram of the 128 sampledpoints ablation comparison
图 7. 128 采样点消融实验对比图

综合上述, 我们通过两种采样数的消融实验表明, 当对光线进行较少的采样, 我们的二阶段采样策略可以很好地指引采样, 在几何表面重建和新视角合成任务上均取得很好的提升。当对光线进行较多数量的采样, 我们的策略相比与纯占据栅格指引的采样, 在几何表面重建任务上仍有微小的提升。

5. 结论

本次工作, 我们引入了一种两阶段的采样策略, 并结合渐进式训练的方式, 使得可以从多视角图像中重建出高质量的物体表面, 同时保持新视角合成任务的出色性能。我们的方案扩展了 NeuS 分层采样和基于占据栅格的采样策略, 通过引入一种非线性函数指引采样, 使得采样点更好地落在靠近物体表面的位置, 从而能更好地建模光线上概率分布。这种结合既减少了在空白区域采样的浪费, 又在占据的栅格中指引了采样, 是一种更细粒度的采样方式。渐进式的训练可以使得模型在训练初期不过多关注物

体的高频细节从而引入噪声, 通过渐进式激活哈希特征的方式使得符号距离场的拟合更加鲁棒, 并且很好地恢复出几何细节。但是, 本文的方法仍然存在不足, 如针对高光表面拟合不稳定、受 Nyquist 采样定理的限制等, 我们将其作为未来工作的探索。

参考文献

- [1] 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>
- [2] Wang, P., Liu, L., Liu, Y., *et al.* (2021) Neus: Learning Neural Implicit Surfaces by Volume Rendering for Multi-View Reconstruction. arXiv: 2106.10689.
- [3] Yariv, L., Gu, J., Kasten, Y., *et al.* (2021) Volume Rendering of Neural Implicit Surfaces. *Advances in Neural Information Processing Systems*, **34**, 4805-4815.
- [4] 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, 5569-5579. <https://doi.org/10.1109/ICCV48922.2021.00554>
- [5] Müller, T., Evans, A., Schied, C. and Keller, A. (2022) Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Transactions on Graphics*, **41**, 1-15. <https://doi.org/10.1145/3528223.3530127>
- [6] 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>
- [7] 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, 4455-4465. <https://doi.org/10.1109/CVPR.2019.00459>
- [8] Niemeyer, M., Mescheder, L., Oechsle, M., *et al.* (2020) Differentiable Volumetric Rendering: Learning Implicit 3D Representations without 3D Supervision. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 3501-3512. <https://doi.org/10.1109/CVPR42600.2020.00356>
- [9] 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.
- [10] Wang, Y., Skorokhodov, I. and Wonka, P. (2023) PET-NeuS: Positional Encoding Tri-Planes for Neural Surfaces. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, 17-24 June 2023, 12598-12607. <https://doi.org/10.1109/CVPR52729.2023.01212>
- [11] Wang, Y., Skorokhodov, I. and Wonka, P. (2022) Hf-Neus: Improved Surface Reconstruction Using High-Frequency Details. *Advances in Neural Information Processing Systems*, **35**, 1966-1978.
- [12] Wang, Y., Han, Q., Habermann, M., *et al.* (2023) NeuS2: Fast Learning of Neural Implicit Surfaces for Multi-View Reconstruction. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Paris, 1-6 October 2023, 3272-3283. <https://doi.org/10.1109/ICCV51070.2023.00305>
- [13] Wu, T., Wang, J., Pan, X., *et al.* (2022) Voxurf: Voxel-Based Efficient and Accurate Neural Surface Reconstruction. arXiv: 2208.12697.
- [14] 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>
- [15] Neff, T., Stadlbauer, P., Parger, M., *et al.* (2021) DONeRF: Towards Real-Time Rendering of Compact Neural Radiance Fields Using Depth Oracle Networks. *Computer Graphics Forum*, **40**, 45-59. <https://doi.org/10.1111/cgf.14340>
- [16] Prinzler, M., Hilliges, O. and Thies, J. (2023) DINER: Depth-Aware Image-Based NEural Radiance Fields. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, 17-24 June 2023, 12449-12459. <https://doi.org/10.1109/CVPR52729.2023.01198>
- [17] Roessle, B., Barron, J.T., Mildenhall, B., *et al.* (2022) Dense Depth Priors for Neural Radiance Fields from Sparse Input Views. *2022 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June 2022, 12882-12891. <https://doi.org/10.1109/CVPR52688.2022.01255>
- [18] Lorensen, W.E. and Cline, H.E. (1998) Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *Seminal Graphics: Pioneering Efforts That Shaped the Field*, **1**, 347-353. <https://doi.org/10.1145/280811.281026>
- [19] Li, L. and Zhang, J. (2023) L0-Sampler: An L0 Model Guided Volume Sampling for NeRF. arXiv: 2311.07044.

- [20] Yu, Z., Chen, A., Antic, B., *et al.* (2022) Sdfstudio: A Unified Framework for Surface Reconstruction. <https://github.com/autonomousvision/sdfstudio>
- [21] Barron, J.T., Mildenhall, B., Verbin, D., *et al.* (2022) Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June 2022, 5460-5469. <https://doi.org/10.1109/CVPR52688.2022.00539>