

DOI:10.3969/j.issn.1003-5060.2026.04.004

结合体素网格和神经辐射场的快速三维重建算法

王安全¹, 李小庆¹, 徐玉华²

(1. 合肥工业大学 电气与自动化工程学院, 安徽 合肥 230009; 2. 奥比中光科技集团股份有限公司, 广东 深圳 518052)

摘要: 三维重建作为计算机视觉的经典任务之一, 在自动驾驶、增强现实、模拟现实等领域有着重要的应用, 相较于传统的三维重建算法, 神经辐射场具有重建精度高、适合复杂物体重建的优点; 但神经辐射场的网络层数深、参数量大, 导致训练时间长。为提高神经辐射场的训练速度, 文章将神经网络与体素网格相结合, 采用体素剪枝等经典算法减少采样点的数量, 最终得到一个速度相较于原神经辐射场有极大提升的网络结构。实验结果表明, 文章算法在 DTU 数据集上重建时间缩减至 1 h 左右, 重建精度比原方法提高 7%。在神经辐射场中使用体素剪枝和光线终止算法, 不仅提高了训练速度, 还提升了重建精度。

关键词: 神经辐射场; 三维重建; 体素剪枝; 体素网格

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1003-5060(2026)04-0456-08

Fast 3D reconstruction algorithm integrating voxel grids and neural radiance fields

WANG Anquan¹, LI Xiaoqing¹, XU Yuhua²

(1. School of Electrical Engineering and Automation, Hefei University of Technology, Hefei 230009, China; 2. Orbbec Inc., Shenzhen 518052, China)

Abstract: 3D reconstruction, as one of the classic tasks in computer vision, has important applications in fields such as autonomous driving, augmented reality, and simulated reality. Compared to traditional 3D reconstruction algorithms, neural radiance fields have the advantages of high reconstruction accuracy and suitability for complex object reconstruction. However, neural radiance fields involve deep network layers and a large number of parameters, resulting in long training times. In order to improve the training speed of neural radiance field, this paper combines neural network with voxel grid, and reduces the number of sampling points by voxel pruning and other classical algorithms. Finally, a network structure with greatly improved speed compared to the original neural radiance field is obtained. Experiments show that the algorithm in this paper reduces the reconstruction time to about 1 hour on the DTU dataset and improves the reconstruction accuracy by 7% compared to the original method. Using voxel pruning and ray termination algorithms in neural radiance fields not only improves the training speed but also improves the reconstruction accuracy.

Key words: neural radiance field; 3D reconstruction; voxel pruning; voxel grid

三维重建是计算机视觉的经典任务之一, 其主要目的是使用二维图片集重建对象的三维模型。三维模型比二维图片更能体现出物体的真实

信息, 更符合普通人的感知。因此三维重建技术在自动驾驶、高精度地图、摄影测量、虚拟现实等领域有着广泛的应用^[1]。传统的三维重建算法主

收稿日期: 2023-12-04; 修回日期: 2024-01-17

基金项目: 国家自然科学基金资助项目(62376083)

作者简介: 王安全(1997—), 男, 河南商丘人, 合肥工业大学硕士生;

李小庆(1990—), 男, 安徽阜阳人, 博士, 合肥工业大学讲师, 硕士生导师, 通信作者, E-mail: lxiaoqing2016@126.com.

要分为彩色(red green blue, RGB)图像重建和深度彩色(red green blue depth, RGBD)图像重建。基于RGB图像的重建方法只适用于特定场景,例如多视图立体视觉法^[2],只适用于表面纹理丰富的物体,否则不能提取有效特征,导致重建失败。基于RGBD图像的重建方法,例如文献^[3-4]方法,需要利用深度相机获取深度信息,重建的场景较小且依赖于深度相机拍摄深度图的效果。近年来,由于体渲染技术的进步^[5],神经网络开始应用于多种计算机视觉任务上,相较于传统方法效果有显著提升,已有研究将其应用于三维重建的任务中提升重建质量。经研究验证,基于神经辐射场的重建效果远优于传统方法。

文献^[6]首次提出神经辐射场概念并应用于新视图合成,使用端到端的全连接网络进行新视图合成,但缺乏对场景几何的约束,重建模型不准确;文献^[7]提出基于神经辐射场的表面重建方法,将符号距离场(signed distance field, SDF)的方法引入到神经辐射场中,将体密度值表示为SDF的函数,提取SDF值为0的集合,可以得到物体的精确模型;文献^[8]提出体素剪枝的方法,对场景中体密度低于阈值的体素不采样,极大地减少了采样次数,相较于文献^[6]的方法,渲染速度提升了10 000倍;文献^[9]引入提议网络,减少了网络的层数和查询网络的次数,在一定程度上缩短了训练时间。在文献^[6]中,使用神经辐射场重建物体时会构建2个网络,即粗网络和细网络,粗网络的结果输出给细网络再次优化,但双网络的优化过程十分漫长,因此需要对采样方式和网络结构进行优化。文献^[10]提出基于体素网格的神经辐射场算法,在每个体素顶点存储体密度值和颜色值,对光线上的采样点插值^[5]得到体密度值和颜色值,因为没有使用神经网络,所以速度很快;但是网格分辨率低,导致重建效果很差,相较于文献^[1]和文献^[9]方法重建出的三维模型有部分缺失,周围有大量的噪点。文献^[11]提出体素网格与神经网络结合的神经辐射场,在体素网格每个顶点存储与体密度相关的特征向量,通过三线性插值法^[12]得到光线上每个采样点的特征向量,然后预测每个采样点的颜色值,减少了训练时间;但是使用体素网格表示三维场景对体素网格的分辨率要求极高,若体素网格分辨率设置得太低,则通过三线性插值得到的体密度信息不准确,导致渲染效果差且模型不准确,若提高体素网格分辨率,则所需的显存容量呈三次方增长,对设备

要求极高。因此,通常基于体素网格的重建比神经网络重建的训练速度快但效果差。

文献^[13]提出多分辨率体素网格,核心思想是构建不同分辨率的体素网格,分辨率由低到高,对输入的采样点在体素中插值得到对应的特征向量;将不同分辨率下的特征向量串联,得到全分辨率下的特征向量;将特征向量输入到神经网络中,预测采样点的体密度值和颜色值,由于神经网络层数只有4层,训练速度较快。文献^[13]方法比文献^[10]方法网格分辨率更高,重建效果更好,但没有对体密度进行约束,导致重建的模型仍有很多噪点,且在重建时仍需要较大的物理显存。

本文将多分辨率网格^[13]应用于神经辐射场表面重建中,极大地减少了网络层数,网络参数从数千万减少到 1.4×10^6 ,加快了收敛速度,但同时面临网格分辨率较高会产生大量无效体素的问题。无效体素增加了大量浮点计算,使网络需要更长的时间收敛,且训练过程中这部分体素占用了大量显存,增加了对显存容量的需求,只能在V100 32G显卡上训练。本文应用体素剪枝的方法^[8],将无效体素从多分辨率网格中删除,不参与计算,在加快收敛速度的同时减少对显存容量的需求,使其能在普通设备上训练,并且使用光线终止算法进一步提高收敛速度。本文算法相较于文献^[13]方法训练速度更快,且重建精度更高;与目前重建效果最优的文献^[7]算法相比,重建模型质量相当;由于减少了对显存的需求,可以在2080Ti-11G显卡上完成训练。本文算法能够在提高训练速度的同时不损失重建精度。

1 基于多分辨率网格的神经辐射场

1.1 神经辐射场基本原理

神经辐射场的原理如图1所示。神经辐射场的输入是一个表示位置信息的三维向量 $\mathbf{x}=(x, y, z)$ 和一个表示方向信息的二维向量 $\mathbf{d}=(\theta, \varphi)$ 。体密度值 σ 是空间位置 \mathbf{x} 的函数 $\sigma(\mathbf{x})$,而颜色值是 \mathbf{x} 和 \mathbf{d} 的函数 $c(\mathbf{x}, \mathbf{d})$ 。这是由于在真实场景中,物体的颜色与光照条件相关,导致从不同的方向观察同一位置时,颜色也会发生变化。

本文采用的神经辐射场使用体素网格与全连接神经网络表示三维场景,在多分辨率网格的顶点存储二维特征向量,用于描述三维场景中的几何信息和颜色信息。仅使用输入的2D图片监督训练神经辐射场,从每张输入图像的像素中心发射一条光线,每条光线上设置128个采样点。

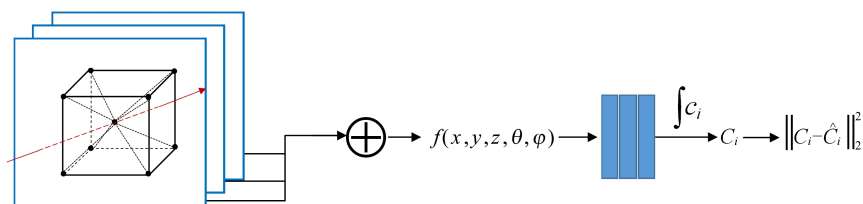


图 1 多分辨率网格神经辐射场原理图

用三线性插值^[5]计算采样点的特征向量,将特征向量输入到全连接网络中解码颜色值和体密度值。因此每个采样点包含一个四维向量,利用体渲染的方法,在输入图像的坐标处将三维场景渲染成与输入图像相同分辨率的二维图片。计算输入图像与渲染图所有像素的颜色误差,将颜色误差作为主要损失值,利用反向传播和梯度下降算法更新神经辐射场中的参数,即

$$\min \frac{1}{n} \sum_{i=1}^n \| \mathbf{I}_i - \mathbf{I}_i' \| \quad (1)$$

其中: n 为图片的数量; \mathbf{I}_i 为输入图像; \mathbf{I}_i' 为渲染图像。逐像素计算每张图像的误差作为网络的损失值,反向传播更新网络的权重 F_θ 。

1.2 网络结构

本文采用显式体素网格与神经网络相结合的方法^[13]重建物体的三维模型,如图 2 所示。

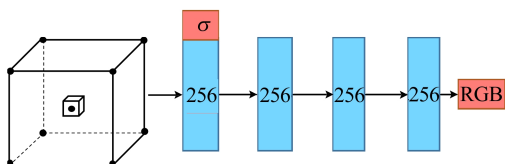


图 2 显式与隐式相结合的网络结构

将场景划分成体素块,每个体素块的各个顶点存储与场景信息相关的特征向量,体素内的采样点通过三线性插值得到该点的特征向量,将该特征向量输入到多层感知机中,解码成体密度值和颜色值,用于渲染二维图像。完全使用神经网络的方法包括 8 层全连接层用于拟合体密度值和 4 层全连接层用于拟合颜色值,这样的网络又分为粗网络和细网络,导致训练过程十分漫长。本文采用的算法将神经网络隐藏层的层数从 24 层减少到 4 层,极大地减少了参数量,相较于仅使用神经网络的方法,重建速度明显加快。

1.2.1 多分辨率体素网格

神经辐射场中使用的神经网络是多层感知机,具有良好的插值性,能够输出平滑的结果,但是难以表达剧烈变化区域的特征^[14]。本文没有

将五维坐标 $\mathbf{p}=(x, y, z, \theta, \varphi)$ 直接作为多层感知机的输入,而是将其编码后再输入到多层感知机中。在文献^[5]中,使用傅里叶编码将欧式空间的样本投射到频率空间,在欧式空间相近的点在频率空间则会相隔很远,使多层感知机能轻易地拟合出这些剧烈的变化。本文采用多分辨率体素网格的方法对输入向量进行编码,将场景划分成多分辨率体素网格^[13],共被排列为 16 个级别的分辨率,每个分辨率下的体素网格的顶点都存储二维特征向量,确定采样点所在的体素网格的级别,通过三线性插值法^[12]得到该采样点在该分辨率下的特征向量,然后将不同分辨率下的特征向量组合,作为多层感知机的输入。

1.2.2 多层感知机

多层感知机由全连接神经网络组成。全连接神经网络原理是每一层的输出经过激活函数后作为下一层的输入,即用全连接神经网络的参数拟合场景中的信息。多层感知机的输入是多分辨率体素网格的输出(见图 2),本文共使用 4 层全连接层作为隐藏层,第 1 个隐藏层输出体密度值和 256 维特征向量,其余隐藏层的输入、输出维数均为 256。第 1 层和最后一层使用 Sigmoid 激活函数,使输出的颜色值与体密度值位于 $[0, 1]$ 之间,其余隐藏层使用 ReLU 作为激活函数,这是由于 ReLU 函数只需要简单的比较运算,计算速度快。

1.3 多分辨率网格的优化

针对多分辨率网格中会产生大量无效体素、增加网络查询次数的问题,本文使用体素剪枝算法对不满足阈值条件的体素剪枝,使其不参与计算,从而减少计算量和显存占用;使用光线终止算法对满足条件的光线不再计算剩余采样点,进一步减少网络查询次数,加快网络收敛速度^[15]。

1.3.1 体素剪枝算法

使用多分辨率体素网格表示场景提高了网格的分辨率,但大部分体素不包含场景信息,无效体素占据的区域对渲染图像几乎没有贡献,而在进行体渲染时仍会对无效体素进行采样,增加了大量的浮点运算。因此需要对这些体素进行限制,

使其不参与计算,从而减少冗余,提高计算效率。本文采用体素剪枝算法对无效体素进行剪枝。

体素剪枝算法原理如下:若一个体素 V_i 内的采样点中体密度的负指数值大于设定阈值,则体素 V_i 被剪去,即

$$\min_{j=1,\dots,G} \{\exp[-\sigma(g_i(\mathbf{p}_j))]\} > \gamma, \quad \mathbf{p}_j \in V_i \quad (2)$$

其中: V_i 为体素网格中的一个体素块; \mathbf{p}_j 为体素块中的第 j 个采样点; g_i 为对采样点进行三线性插值; σ 为计算采样点的体密度值。在训练过程中随着体素分辨率不断提高以及无效体素被剪掉,训练速度得到提升,显存占用同时减少。此外无效体素中存储的信息不会对重建场景造成影响,且能够在一定程度上提高重建精度。

1.3.2 光线终止算法

体渲染的方法可以很好地表示半透明物体和雾状物体,但对于固体表面,使用体渲染的方法沿着光线累积采样点的颜色时,在物体表面重建完成之后仍需要许多不必要的累积步骤才能使累积透明度 $A(\mathbf{p}_0, v)$ 达到 0, 增加网络的无效查询次数。因此,当累积透明度 $A(\mathbf{p}_0, v)$ 下降到某个阈值以下时,提前停止评估这条光线上该采样点之后的采样点^[5]。实验中设置阈值 $\epsilon=0.01$ 能够加快训练速度,且不会导致重建精度下降。

2 表面重建

2.1 符号离场

神经辐射场可以重建深度不连续的场景,但是从预训练的模型中提取高质量的模型很困难,通常情况下重建的物体表面有很多噪点,这是由于缺乏对体密度的几何约束。文献[7]结合体渲染和表面重建的优点,将三维场景约束为符号距离函数,即计算三维空间中的一个点到目标对象表面的距离,并同时距离的符号进行定义,若点在目标物体内部符号为正,在目标物体外部符号为负,则位于目标物体边界上时为 0。重建目标物体的三维模型就是提取零表面集合,即

$$S = \{\mathbf{x} \in \mathbf{R}^3 \mid f(\mathbf{x}) = 0\} \quad (3)$$

其中: \mathbf{x} 为三维空间中的点; $f(\mathbf{x})$ 的值由神经辐射场直接输出。图 2 中输出的体密度值改为输出 SDF 值,再通过概率密度函数 $\phi_x(f(\mathbf{x}))$ 将 SDF 值转换成体密度值,体密度值主要用于渲染图像,概率密度函数为 $\phi_x(\mathbf{x}) = se^{-x}/(1+e^{-x})^2$, $1/s$ 是 $\phi_x(\mathbf{x})$ 的标准差,同时也是可训练参数,用来表示

物体表面的收敛程度。如果采样点的 SDF 值为 0,说明该点是物体的表面,通过提取 SDF 值为 0 的集合,则可以得到物体的精确表面。

2.2 体渲染

神经辐射场将场景表示为空间点的集合,每个点包含体积密度值和颜色值属性,体积渲染算法^[5]将三维场景中的采样点渲染成二维图像,与输入图像计算的像素差作为损失值。

体渲染的原理如下:从输入图像的每个像素中心发射光线,对发射光线的颜色和体密度值进行积分,得到该像素的颜色值,具体过程如式(2)。体密度 $\sigma(x)$ 是从像素中心发射的光线终止于位置 x 处的无穷小粒子的微分概率。 t_n, t_f 表示场景的远、近边界,像素中心发射的光线 $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ 的颜色 $C(\mathbf{r})$ 为:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))c(\mathbf{r}(t), \mathbf{d}) dt, \\ T(t) = \exp\left(-\int_{t_n}^t \sigma(\mathbf{r}(t)) dt\right) \quad (4)$$

其中,函数 $T(t)$ 表示光线从 t_n 到 t 的累计透射率,即光线从 t_n 传播到 t 而不撞击其他粒子的概率。在上述公式中使用积分计算图像每个像素的颜色值,但是在神经辐射场中只能离散地查询采样点的信息,因此上述体渲染的积分公式需改写为离散的形式。若使用离散方式计算像素颜色,则需要采用分层抽样的方法在光线上均匀设置采样点。首先将区间 $[t_n, t_f]$ 划分成 N 个均匀区间,然后从每个区间中随机抽取一个采样点,每个采样点的区间如下:

$$t_i \in \left[t_n + \frac{i-1}{N}(t_f - t_n), t_n + \frac{i}{N}(t_f - t_n) \right].$$

将用积分求像素的颜色值式(4)改写成离散采样点的形式,即

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^N T_i [1 - \exp(-\sigma_i \delta_i)] c_i, \\ T_i = \exp\left(-\sum_{i=1}^{i-1} \sigma_i \delta_i\right) \quad (5)$$

其中: $\delta_i = t_{i+1} - t_i$ 为相邻采样点之间的距离; c_i 为当前采样点的颜色。进一步将式(5)简写为传统的 α 合成形式,即

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^N w_i c_i, \quad w_i = T_i [1 - \exp(-\sigma_i \delta_i)] \quad (6)$$

其中, w_i 为每个采样点的权重。

2.3 损失函数

由于体渲染过程是完全可微的,可计算渲染

图像与一组输入图像所有像素的颜色误差。将颜色误差作为主要损失值,与重建损失和掩码损失加权,然后反向传播训练神经辐射场,而无需任何 3D 监督。神经辐射场的总损失函数如下:

$$L = L_{\text{color}} + \lambda L_{\text{reg}} + \beta L_{\text{mask}} \quad (7)$$

颜色损失 L_{color} 为:

$$L_{\text{color}} = \frac{1}{m} \sum_{\mathbf{r} \in \mathbb{R}^3} \|\hat{C}(\mathbf{r}) - C(\mathbf{r})\|_2^2 \quad (8)$$

其中, m 为输入图像的数量。计算预测图像与输入图像像素的颜色误差并将其作为主要损失值。

重建损失是为了正则化采样点的 SDF 值,即

$$L_{\text{reg}} = \frac{1}{nm} \sum_{k,i} [\|\nabla f(\hat{\mathbf{p}}_{k,i})\|_2 - 1]^2 \quad (9)$$

其中: n 为采样点数量。掩码损失可以在一定程度上提高重建质量,若使用了掩码损失,在训练阶段也需要输入掩码图片集。掩码损失定义为:

$$L_{\text{mask}} = \text{BCE}(M_k, \hat{O}_k) \quad (10)$$

其中, $\hat{O}_k = \sum_{i=1}^n T_{k,i} \alpha_{k,i}$ 为每条光线的权重之和,表示深度信息,与输入的掩码图像计算交叉熵。

3 实验及结果分析

3.1 数据集

1) DTU 是一个大型数据集,包含受控实验室环境中的 128 个场景,其模型使用结构光扫描仪捕获。在 7 种不同的照明条件下,在相同的 49 或 64 个相机位置扫描每个场景,生成分辨率为 $1\,200 \times 1\,600$ 像素的 RGB 图像。数据集涵盖了各种对象和材料,因此非常适合在现实条件下训练和测试深度学习方法。

2) BlendedMVS 数据集包含 113 个场景,涵盖各种不同的场景。每个场景包含 20 到 100 张不等的输入图像,整个数据集共有 17 818 张图像。数据集中的场景包含各种不同的相机轨迹,非结构化的相机轨迹可以更好地建模不同的图像捕捉风格,使网络对真实世界的重建更具实用性。

3.2 实验细节

本文采用 PyTorch 深度学习框架来完成神经辐射场网络的搭建、训练和重建过程,并在 V100 32G GPU 上对训练过程进行加速。

训练过程使用 AdamW 作为优化器,优化器参数 $\beta_1 = 0.9, \beta_2 = 0.99$,采用余弦退火算法^[7]作为学习率更新策略,初始学习率为 0.5。训练神经辐射场的数据集采用 DTU122、DTU106、BlenderMVS-Man,使用全分辨率图像训练,并迭代训

练 100 000 次。

本文选择 3 种算法进行对比,即重建效果最优的神经表面重建算法(NeuS)^[9]、训练速度最优的多分辨率网格算法(instant-ngp)^[13]和传统多视图立体视觉算法(colmap)^[16]。比较本文算法与各种算法的训练速度、重建精度、和渲染效果。

3.3 实验结果

本文算法与 NeuS、instant-ngp 和 colmap 算法完成重建所需训练时间的对比见表 1 所列,数据集为 DTU122,迭代次数 100 000 次。从表 1 可以看出,本文算法的训练速度明显比其他 3 种算法更快,仅在 1 h 之内即可完成重建。

表 1 4 种算法在 DTU122 数据集上训练时间对比

算法	训练时间/min
NeuS	293
colmap	121
instant-ngp	83
本文	52

4 种算法重建的三维模型如图 3 所示。

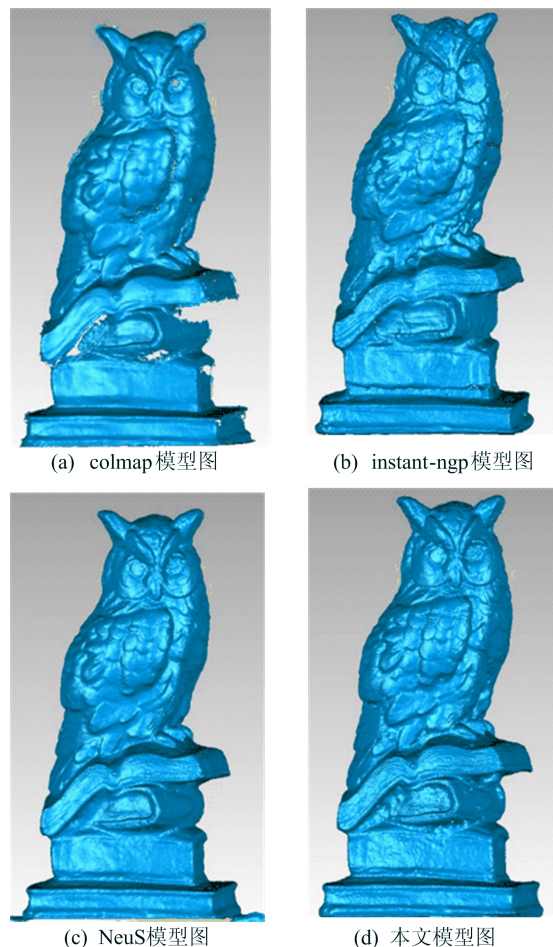


图 3 4 种算法重建的模型比较

从图 4 可以看出: colmap 重建不完整,模型有很多空洞和缺失; instant-ngp 重建的模型顶部明显凹凸不平; NeuS 和本文算法重建效果最优,模型细节重建完整且不包含噪点。

各算法重建结果在 DTU122 数据集上的定量表现见表 2 所列,本文算法的峰值信噪比(PSNR)值、结构相似性(SSIM)值与 NeuS 算法相当,但明显优于另外 2 种算法的重建结果。

表 2 4 种算法在 DTU122 数据集上重建质量定量对比 单位: dB

算法	PSNR	SSIM
colmap	23.82	0.863
NeuS	27.89	0.897
instant-ngp	29.69	0.946
本文算法	29.72	0.969

本文用峰值信噪比 PSNR 和结构相似性 SSIM 作为重建精度的评价指标。PSNR 是到达噪音比率的顶点信号,通常用于最大值信号与背景噪音之间的比率。计算公式如下:

$$M_{SE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2,$$

$$P_{SNR} = 10 \times \lg\left(\frac{I_{max}^2}{M_{SE}}\right) \quad (11)$$

其中: m, n 为图像的宽和高; I 为不包含噪音的纯净图像; K 为噪声图像; I_{max} 为图像像素最大值。PSNR 值 P_{SNR} 越大表示重建效果越好。

结构相似指标可以衡量图片的失真程度,也可以衡量 2 张图片的相似程度。与 PSNR 衡量绝对误差不同, SSIM 是感知模型,它分别从亮度、对比度、结构 3 个方面度量图像相似性。

渲染图像也可以在一定程度上评估算法的好坏。用 DTU106 数据集的渲染图像评估 4 种算法,以 100 000 次迭代为标准,结果如图 4 所示。

从图 4 可以看出: colmap 算法渲染图像有较多的缺失; instant-ngp 渲染图尾部细节丢失; NeuS 算法未完全收敛,羽毛部分过度平滑,细节未重建出来; 而本文算法渲染图没有尾部丢失,且细

节部分也重建较好,效果明显优于其他 3 种算法。

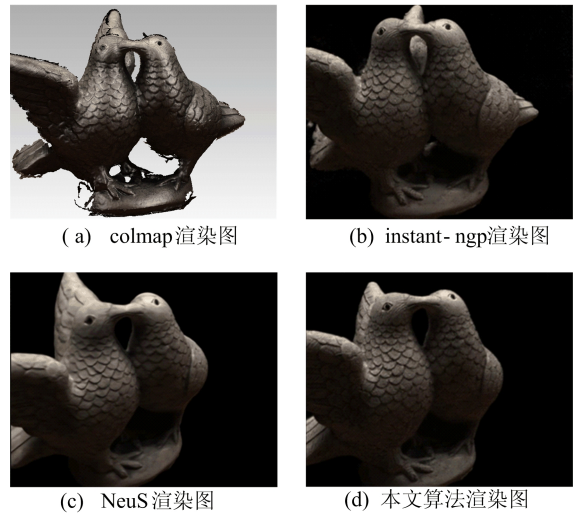


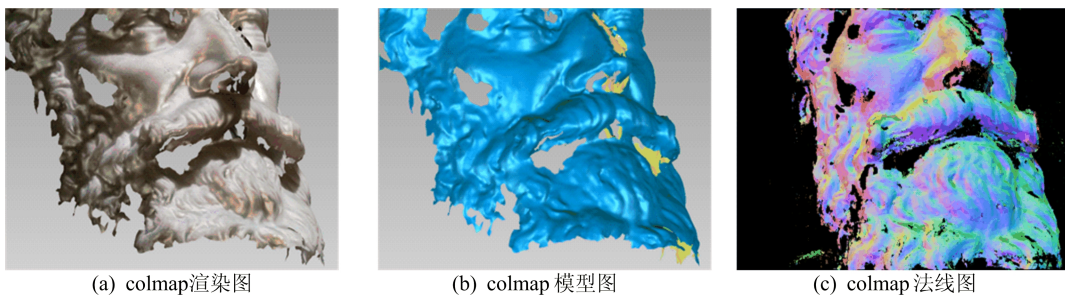
图 4 4 种算法渲染图比较

4 种算法在 DTU106 数据集上训练时间的比较见表 3 所列。从表 3 可以看出,本文重建时间最短 1 h 左右完成重建,相较于其他算法重建时间明显更快。

表 3 4 种算法在 DTU106 数据集上训练时间对比

算法	时间/min
NeuS	322
colmap	196
instant-ngp	86
本文算法	63

4 种算法在 BlenderMVS-Man 数据集上的表现如图 5 所示。从图 5 可以看出: colmap 重建模型有很多缺失,且所需时间最长; instant-ngp 需要 98 min 完成重建,重建结果完整但模型左边有较多噪点; NeuS 需要 283 min 完成训练,重建模型眼部细节表现不好; 本文算法仅需 92 min 完成训练,重建速度比 instant-ngp 算法稍快,重建精度明显优于其他算法。因此,本文算法无论是效果还是速度均最优。



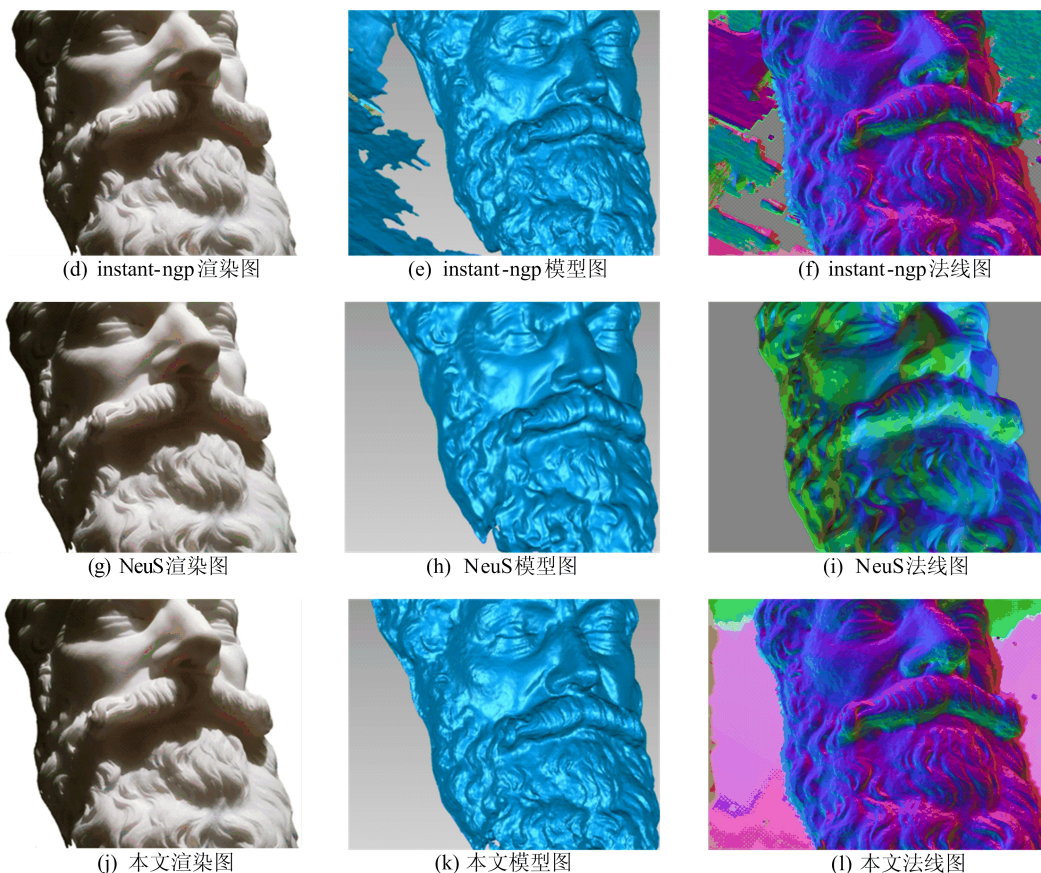


图 5 4 种算法的渲染图、模型图、法线图

4 结 论

本文针对基于深度学习重建算法训练时间长、重建模型精度差等问题,采用体素网格与神经网络相结合的网络结构,极大地减少了神经网络的层数,并使用光线终止和体素剪枝算法^[8]减少无效体素的数量和显存的占用,极大地提高了训练速度。

实验结果表明,本文算法在大部分场景下都能一定程度地提高训练速度,重建精度相较于 NeuS 和 instant-ngp 算法也更优越;在 DTU122 和 DTU106 数据集上的训练时间都在 60 min 左右,而 NeuS 的方法重建需要花费近 300 min;在 BlenderMVS-Man 数据集上,相同的迭代次数下,本文算法重建精度优于 NeuS 和 instantngp 重建算法。综合实验数据表明,本文算法实现了三维重建精度、速度的均衡。

[参 考 文 献]

[1] KANG Z Z, YANG J T, YANG Z, et al. A review of techniques for 3d reconstruction of indoor environments[J]. ISPRS International Journal of Geo-Information, 2020, 9(5):

330-345.

- [2] LEE J Y, DEGOL J, ZOU C H, et al. Patchmatch-rl: deep mvms with pixelwise depth, normal, and visibility[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S. l.]: IEEE, 2021: 6158-6167.
- [3] LZADI S, KIM D, HILLIGES O, et al. Kinectfusion: real-time 3D reconstruction and interaction using a moving depth camera[C]//Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. New York: Association for Computing Machinery, 2011: 559-568.
- [4] DAI A, MATTHIAS N, MICHAEL Z, et al. Bundlerefusion: real-time globally consistent 3D reconstruction using on-the-fly surface reintegration [J]. ACM Transactions on Graphics, 2017, 36(4): 1-18.
- [5] DELL D B, MARTEL J P, WETZSTEIN G, et al. Autoint: automatic integration for fast neural volume rendering [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2021: 14556-14565.
- [6] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. Nerf: representing scenes as neural radiance fields for view synthesis [J]. Communications of the ACM, 2021, 65(1): 99-106.
- [7] BARRON J T, MILDENHALL B, VERBIN D, et al. Mip-nerf 360: unbounded anti-aliased neural radiance fields

- [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE,2022; 5470-5479.
- [8] LIU L, GU J, ZAW L K, et al. Neural spars voxel fields[J]. Advances in Neural Information Processing Systems, 2020, 33:15651-15663.
- [9] WANG P, LIU L, LIU Y, et al. Neus: learning neural implicit surfaces by volume rendering for multi-view reconstruction[J]. Advances in Neural Information Processing Systems, 2021, 34:27171-27183.
- [10] FRIDOVICH S, YU A, TANCIK M, et al. Plenoxels: radiance fields without neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE,2022;5501-5510.
- [11] SUN C, SUN M, CHEN T, et al. Improved direct voxel grid optimization for radiance fields reconstruction[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE,2022;5449-5459.
- [12] GUI P, LI B, TANG T, et al. An improved tri-linear interpolation method for hybrid overset grids and its application [C]//The Proceedings of the 2018 Asia-Pacific International Symposium on Aerospace Technology. Singapore: Springer,2019;601-613.
- [13] MÜLLER T, EVANS A, SCHIED C, et al. Instant neural graphics primitives with a multiresolution hash encoding [J]. ACM Transactions on Graphics, 2022, 41(4):1-15.
- [14] WEGHORST H, HOOPER G, GREENBERG D P. Improved computational methods for ray tracing[J]. ACM Transactions on Graphics (TOG), 1984, 3(1):52-69.
- [15] LIN M, JI R, WANG Y, et al. Hrank: filter pruning using high-rank feature map[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE,2020;1529-1538.
- [16] FISHER A, CANNIZZARO R, COCHRANE M, et al. Colmap: a memory-efficient occupancy grid mapping framework[J]. Robotics and Autonomous Systems, 2021, 142:103755.

(责任编辑 胡亚敏)

(上接第 455 页)

- [12] PROIETTI T, PEDREGAL D J. Seasonality in high frequency time series[J]. Econometrics and Statistics, 2023, 27:62-82.
- [13] BANDARA K, HYNDMAN R J, BERGMEIR C. MSTL: a seasonal-trend decomposition algorithm for time series with multiple seasonal patterns[PP/OL]. arXiv (2021-07-28). <https://doi.org/10.48550/arXiv.2107.13462>.
- [14] WU Q, LIN H. Daily urban air quality index forecasting based on variational mode decomposition, sample entropy and LSTM neural network[J]. Sustainable Cities and Society, 2019, 50:101657.
- [15] GUO Z, YU B, HAO M, et al. A novel hybrid method for flight departure delay prediction using random forest regression and maximal information coefficient [J]. Aerospace Science and Technology, 2021, 116:106822.
- [16] 杨贵军, 李晓霞, 孙玲莉. Prophet-X-12-ARIMA 组合模型及交通运输量预测[J]. 统计与决策, 2023, 39(4):29-34.
- [17] TORRES M E, COLOMINAS M A, SCHLOTTHAUER G, et al. A complete ensemble empirical mode decomposition with adaptive noise[C]//2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). [S. l.]:IEEE,2011;4144-4147.
- [18] 汤俊, 李垠健, 高鑫. 基于 CEEMDAN 的 GNSS 变形监测去噪方法 [J]. 大地测量与地球动力学, 2021, 41(4):408-412.
- [19] YU Y, SI X, HU C, et al. A review of recurrent neural networks; LSTM cells and network architectures[J]. Neural Computation, 2019, 31(7):1235-1270.
- [20] 刘晴晴, 陈华友. 基于赋权 KNN-LSTM 模型的 PM_{2.5} 质量浓度预测 [J]. 合肥工业大学学报(自然科学版), 2021, 44(12):1690-1698.
- [21] ABDOLLAHZADEH B, GHAREHCHOPOGH F S, MIRJALILI S. African vultures optimization algorithm; a new nature-inspired metaheuristic algorithm for global optimization problems[J]. Computers & Industrial Engineering, 2021, 158:107408.
- [22] SIDDIQUI R, AZMAT M, AHMED S, et al. A hybrid demand forecasting model for greater forecasting accuracy: the case of the pharmaceutical industry[J]. Supply Chain Forum; An International Journal, 2022, 23(2):124-134.

(责任编辑 胡亚敏)