计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
2026,35(2):201−208 [doi: 10.15888/j.cnki.csa.010092] [CSTR: 32024.14.csa.010092]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于变分自编码器的神经辐射场三维重建①

徐振宇, 康　睿, 钱　蔚, 曹一鸣, 朱靖恺, 彭　森, 郭　翔

(南京南瑞信息通信科技有限公司, 南京 211106)
通信作者: 徐振宇, E-mail: 1246233565@qq.com

摘　要: 神经辐射场 (neural radiance field, NeRF) 相较于传统的三维重建方法, 能够有效捕获隐式神经表征, 实现高质量的三维重建与新视角合成任务, 然而其需要大量的原始数据进行训练. 为了解决这一问题, 本文借助变分自编码器 (variational autoencoder, VAE) 能够捕捉潜在空间表示的能力, 将其与神经辐射场结合提出一种方法来提高低训练数据量下的三维场景生成效果. 首先, 通过构造变分自编码器的编码器, 选取训练数据中一定比例的原始图片构成向量集, 通过编码器对向量集进行压缩, 从而捕捉其中的潜在特征向量作为输入层数据的全局场景信息补充. 其次, 构造自适应增强采样算法动态调整采样点的分布密度, 增强神经辐射场对场景中细节信息的捕捉能力. 本文选取 3 个公开数据集进行对比实验, 实验结果验证了本方案的有效性, 同时, 所提出的方法在原始训练数据量缺失的情况下能达到与对比网络在完整训练数据量近似的三维重建结果.

关键词: 神经辐射场; 变分自编码器; 自适应采样算法

引用格式: 徐振宇,康睿,钱蔚,曹一鸣,朱靖恺,彭森,郭翔.基于变分自编码器的神经辐射场三维重建.计算机系统应用,2026,35(2):201–208. http://www.c-s-a.org.cn/1003-3254/10092.html

3D Reconstruction of Neural Radiation Field Based on Variational Autoencoder

XU Zhen-Yu, KANG Rui, QIAN Wei, CAO Yi-Ming, ZHU Jing-Kai, PENG Sen, GUO Xiang

(Nanjing NARI Information & Communication Technology Co. Ltd., Nanjing 211106, China)

Abstract: Compared to traditional 3D reconstruction methods, neural radiance fields (NeRFs) can effectively capture implicit neural representations, enabling high-quality 3D reconstruction and novel view synthesis tasks. However, NeRFs typically require a large amount of raw data for training. To this end, this study proposes a method by integrating variational autoencoders (VAEs) with NeRF to improve 3D scene generation performance under limited training data. Firstly, by constructing a VAE encoder, a certain proportion of raw images from the training data are selected to form a vector set. Meanwhile, the encoder compresses this set to capture latent feature vectors, which are then employed to supplement global scene information in the input layer. Secondly, an adaptive-enhanced sampling algorithm is developed to dynamically adjust the distribution density of sampling points, thereby improving NeRF's ability to capture fine scene details. Experiments conducted on three public datasets demonstrate the effectiveness of the proposed method. Additionally, the proposed method achieves 3D reconstruction results comparable to baseline methods trained with full datasets, even under the loss of original training data.

Key words: neural radiance field (NeRF); variational autoencoder (VAE); adaptive sampling algorithm

三维模型重建是计算机视觉和图形学领域中的一个重要研究方向, 其目标是从多视角图像或点云数据中重建出场景的三维结构, 并对模型以及场景的纹理进行映射, 三维重建技术在虚拟现实、增强现实、影

① 收稿时间: 2025-07-01; 修改时间: 2025-08-26, 2025-09-19; 采用时间: 2025-10-14; csa 在线出版时间: 2025-12-19
CNKI 网络首发时间: 2025-12-22

视特效、自动驾驶等领域有广泛应用. 传统的三维重建方法主要包括基于多视觉几何原理 (structure from motion, SFM) 和多视图立体 (multi-view stereo, MVS)[1,2], 这些方法通过估计相机位姿和场景的几何关系来生成三维点云或网格. 然而, 这些方法在处理光照变化、遮挡和复杂材质时, 往往表现不佳. 近年来, 神经辐射场 (neural radiance field, NeRF)[3]作为一种基于深度学习的三维重建方法, 取得了广泛关注. NeRF 通过学习将三维空间中的每一个位置和视角映射到颜色和密度的函数, 使用体渲染技术生成逼真的新视角图像. 与传统方法相比, NeRF 能够捕捉更细腻的光照和反射效果, 并在高质量的图像重建上表现突出.

NeRF 需要进行大量的体积采样与神经网络推理, 导致单幅图像渲染时间通常达到数秒至数分钟, 限制了其实时应用. 针对该问题, Barron 等人[4]提出 Mip-NeRF, 通过多尺度锥体采样替代点采样, 减少反走样效应同时提升采样效率, 实现更快速且高质量的渲染. Yu 等人[5]结合空间数据结构八叉树, 提出 PlenOctrees 方法, 利用空间索引加速神经网络查询, 实现近实时渲染. 此外, Hu 等人[6]提出双阶段高效采样和场景缓存机制, 显著缩短训练与推理时间. Gao 等人[7]提出一套通用的隐式框架, 以 NeDF (neural depth field) 实现光线与隐式表面的快速相交, 加速组合与渲染多个 NeRF 对象. 尽管显著提速, 但仍存在速度-质量权衡的问题, 例如锥体积分在细节/深度不连续处易过度平滑, 且预烘焙的八叉树/体素结构内存占用高、可编辑性与增量更新差, 遇到分布外视角或相机覆盖稀疏时缓存/重要性采样易退化等.

原始 NeRF 针对每个场景需单独训练模型, 训练时间长且难以泛化至未见场景, 限制了其通用性. 为提升泛化能力, Yu 等人[8]提出 PixelNeRF, 结合条件编码器从单张或少量图像提取特征, 实现对新场景的快速推断, 无需重新训练. Trevithick 等人[9]提出 GRF (general radiance field), 将任意场景看作“通用辐射场”, 不仅能还原单个场景, 还能泛化到不同的对象与类别, 只需要一套网络即能描述多种场景. 但前馈泛化在材质跨域等情况下容易失真, 单张或少量视图下尺度歧义与几何漂移常见, 对相机内外参误差、模糊/弱纹理鲁棒性有限.

传统 NeRF 假设场景静止, 难以处理动态或非刚性对象, 并且未对异常点作敏感处理. 为此, Pumarola 等人[10]提出 D-NeRF, 将时间变量引入辐射场输入, 实现动态场景的高质量重建与视角合成. Li 等人[11]提出 NSFF, 将光流场与辐射场结合, 提升动态场景的细节表现和时空一致性. Chen 等人[12]提出“启发式引导分割 (HuGS)”范式, 将 SfM-驱动的启发式与颜色残差启发式相结合, 提升 NeRF 在非静态场景下的三维重建能力. Feng 等人[13]提出 AE-NeRF 解决在非理想事件序列与位姿噪声下的三维重建挑战, 结合联合位姿校正、分层事件蒸馏与事件/时序一致性损失, 将方法推广到更大场景并取得不错的重建表现. 但是动态场景重建在大幅度形变与拓扑变化下的位置仍难精确建模, 长序列优化易漂移且训练与推理成本高, 并且依赖启发式的方案在反射、模糊、弱纹理与遮挡下易失稳等问题.

真实场景中光照变化复杂、多样材质及动态元素对 NeRF 提出挑战. Martin-Brualla 等人[14]提出 NeRF in the wild (NeRF-W), 通过显式建模环境光照变化与非静态物体, 提高了 NeRF 对无约束照片集的适应性和鲁棒性. Barron 等人[15]进一步提出 Mip-NeRF 360, 实现对 360 度全景大范围场景的无缝重建, 提升对远景和复杂结构的表现能力. Bi 等人[16]提出了一种能够同时捕获场景几何、反射属性和光照的神经隐式表示, 采用一种 differentiable ray-marching 框架, 同时沿相机与点光源方向进行光线积分, 支持对阴影、镜面反射等非纹理效果的建模. Cui 等人[17]在体渲染中引入“隐匿场 (concealing field)”对空气透射进行建模, 从而在仅用低照/过曝图像训练时也能适应光照并生成常光条件的新视角. Chen 等人[18]面向存在显著相机位姿异常点的多视角重建, 提出将 NeRF 与场景图联合优化: 通过基于邻域兼容性与渲染一致性的自适应内/外点置信度估计抑制外点影响, 并引入 IoU 几何-位姿联合损失与粗到细训练; 同时发布含外点的新评测数据, 实验在多数据集上比现有方法更稳健、重建质量更高. 但真实世界鲁棒性与大场景外观、光照、几何的解耦仍不充分, 重光照与材质可编辑性受限, 并且超大范围场景在远景细节、尺度与位姿稳定性上仍有退化, 同时存在几何、反射、光照三者联合估计资源消耗严重, 在未知光源与复杂 BRDF 下易退化等问题.

尽管 NeRF 展现了强大的重建能力, 但也依然存在一些问题和挑战. 首先, NeRF 对输入数据的稠密程度要求较高, 当采样点不足或某些区域的图像数据较

稀疏时, 重建的三维场景效果欠佳. 其次, NeRF 的采样策略忽略了图像区域之间复杂度的差异性, 一致性的采样方法会增加训练过程的计算成本. 本文在 NeRF 的训练基础上引入变分自编码器 (variational autoencoder, VAE)[19]对训练数据的潜在信息进行捕捉, 作为训练中全局场景信息的补充, 以应对训练数据稀缺的情况. 同时, 本文提出一种采样策略, 依据不同区域之间复杂度的差异性动态调整不同光线上采样点的数量, 减少训练成本, 增强对复杂区域细节的捕捉.

1 神经辐射场

神经辐射场 (NeRF) 的核心是一种通过多层感知机 (MLP) 构造的神经网络模型, 通过输入光线下采样点的空间坐标$X=(x,y,z)$和视角方向$D=(\theta,\varphi)$, 输出该点在指定视角下颜色和密度值的预测结果, 完整的 NeRF 结构可以分为高维位置编码网络、辐射场网络、体积渲染模块.

位置编码网络是 NeRF 的重要特性, 用于提高模型对高频细节的建模能力. 为了提高辐射场网络对高频信息的拟合, NeRF 对预测光线输入的三维空间坐标$X=(x,y,z)$和视角方向$D=(\theta,\varphi)$构成的位置编码在进行采样前映射到高维空间, 将光线样本的坐标扩展为一系列的高维正弦和余弦函数, 公式如下:

$$o(p)=[\sin(2^0\pi p),\cos(2^0\pi p),\cdots,\sin(2^{L-1}\pi p),\cos(2^{L-1}\pi p)] \tag{1}$$

其中, p为三维位置X和视角方向D的形参表示; L表示位置编码的维度.

辐射场网络由两个 MLP 模型组合而成, 分别为 8 层全连接的采样模型和 3 层全连接的推理模型. 对光线上多个采样点的位置编码进行高维扩展后输入辐射场网络, 经过提取和推理得到该条光线多个采样点的预测结果, 公式如式 (2):

$$F(X,\theta,\varphi)=(c,\sigma) \tag{2}$$

其中, c为颜色, σ为体积密度.

NeRF 通过模拟相机成像的方式, 将光线上的多个像素点的预测结果使用体积渲染公式进行计算, 最终得到光线在相应像素位置上的颜色, 公式如式 (3):

$$C(r)=\int_{t_1}^{t_2}T(t)\cdot\sigma(r(t))\cdot c(r(t),d)\cdot \mathrm{d}t \tag{3}$$

其中, $C(r)$表示重建的图像中像素点颜色, $T(t)$表示光线从t_1传播到t_2而未被拦截的累积透射率.

在 NeRF 模型中, 渲染过程的采样策略决定了渲染效率和视觉质量的平衡. NeRF 采用一种层次化的双层采样策略: 粗采样和细采样. 在保证渲染质量的同时提高渲染效率, 该策略通过初步估计和局部细化相结合的方式, 高效捕捉场景中的光照与体积信息. NeRF 在粗采样阶段对每条光线进行样本点数量为 Ns 的均匀采样, 粗采样采用较大间隔来快速估计物体的整体几何形状和环境的粗略光照分布; 细采样阶段对粗采样阶段中未能精确捕捉的重要细节区域进行样本点数量为 Nf 的重采样, 依赖粗采样结果对光线经过的复杂区域提高采样频率, 保证渲染结果的精度. 使用两次采样的结果进行预测来量化光线的隐式表示, 依据体渲染公式累计生成最终的 RGB 与体密度.

2 VAE-NeRF 神经网络构建

针对 NeRF 在稀疏数据下三维场景重建效果不佳的问题提出 VAE-NeRF 进行改进, 完整的网络结构如图 1 所示. 为了保证在稀疏数据下 NeRF 能够更好地捕捉到场景的全局特征信息, 使用变分自编码器对训练集中的部分原始训练数据进行编码, 将高维数据压缩到低维空间, 获取潜在特征信息作为输入层的补充. 提出一种自适应增强采样算法, 针对双层采样策略中细采样阶段进行优化, 增强细节渲染效果.

2.1 变分自编码器

变分自编码器是一种生成模型, 通过结合自编码器 (autoencoder) 和变分推断 (variational inference) 方法学习数据的潜在表示, 并能够生成新的数据样本. VAE 主要由两个部分构成: 编码器 (encoder) 和解码器 (decoder). 编码器作为推断模型, 通过输入数据X获取其在潜在空间的高斯概率分布, 输出潜在变量Z的概率分布参数: 均值向量μ和方差向量σ; 解码器通过对潜在变量Z进行采样, 从数据空间中恢复或者生成与输入数据具有相似特征的新样本.

神经辐射场依赖于训练数据集的丰富度, 在稀疏数据下无法捕捉到完整的全局信息去实现三维重建. 为了保证神经辐射场在数据稀缺的情况下也能够捕获到较为完整的全局特征信息描述, 从训练集中选取多个不同角度的原始图片数据构成全局信息向量集, 多角度图片的选取可以减少单个视角可能带来的局部偏差问题, 从而减少极端视角或异常图像对全局场景信

息的影响. 接着, 构造编码器对向量集进行潜在特征捕捉, 将输出后的向量作为场景的全局特征潜在表示并作为输入层数据的补充.

首先, 使用编码器对原始训练数据X_i进行潜在表征捕捉, 得到原始图片所表示的概率分布Z_i的分布参数均值μ_i和方差σ_i, 两者共同定义原始图片潜在空间的高斯分布, 其中均值向量提供了图像潜在特征的中心信息, 因此选择均值向量作为输入层数据的补充, 如式 (4) 所示. 接着, 对所有编码后的原始图片的均值向量进行求和计算, 并对其取平均值得到向量$\bar{\mu}$作为全局场景信息, 如式 (5) 所示. 最后, 将进行傅里叶编码后的采样点数据与压缩后的全局场景信息以拼接形式实现对输入层数据的扩充, 输入神经辐射场模型进行训练.

$$Z=\frac{1}{n}\sum_{i=1}^{n}Z_i'(Z_i\sim N(\mu_i,\sigma_i^2)) \tag{4}$$

$$\bar{\mu}=\frac{1}{n}\sum_{i=1}^{n}\mu_i \tag{5}$$

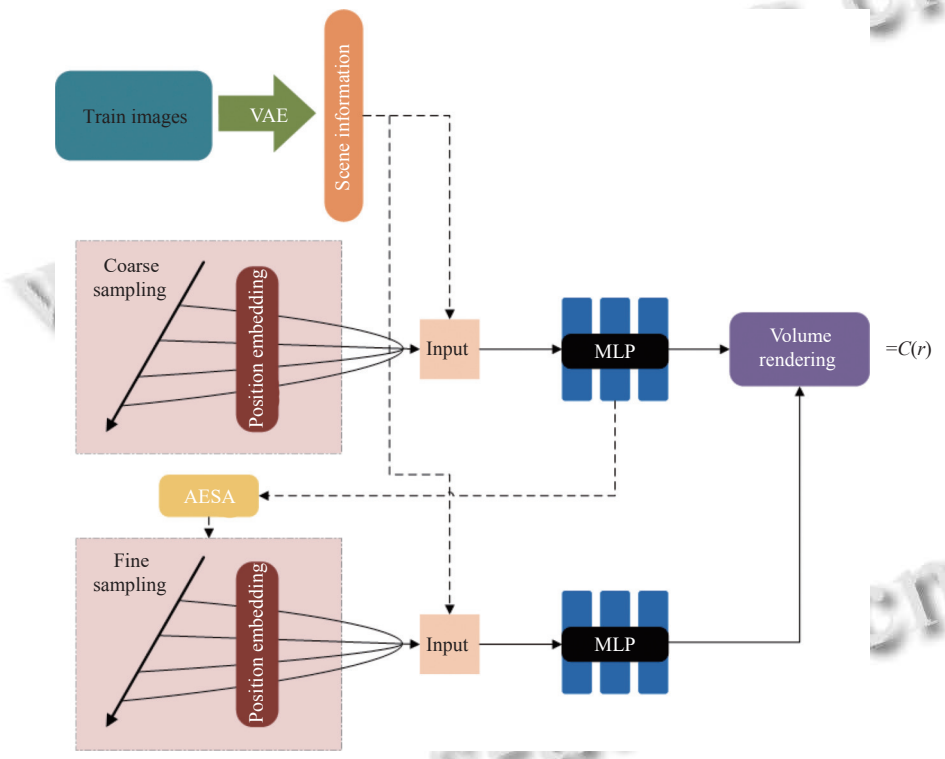

图 1 VAE-NeRF 神经网络结构图

2.2 自适应增强采样算法

NeRF 采用双层采样策略实现对光线上的点进行采样推理, 这种方式在粗采样阶段采用较大间隔来快速估计物体的整体几何形状和环境的粗略光照分布, 并根据粗采样的推理结果, 在细采样阶段对所有采样射线未捕捉到的细节区域进行相同数量点的重采样增强渲染效果. 但是, 这种方式忽视了不同光线之间复杂度的差异性. 因此, 本文提出一种自适应增强采样算法 (adaptive enhanced sampling algorithm, AESA), 对采样点的分布密度进行动态调整, 从而对场景中细节丰富或变化剧烈的区域进行更多的采样, 而在较平坦或变化较少的区域减少采样点数量, 更有效地提升对细节区域的渲染质量.

粗采样阶段与双层采样策略保持同样的方法, 对每条光线以较大间隔平均地进行初始的样点采集, 将采样结果输入到模型中进行推理得到一组结果. 每个粗采样点对应由颜色 RGB 和不透明度 α 构成的四维向量预测结果, 对同一条光线上的相邻采样点进行四维向量的数值差异计算, 对两个向量上每一维数值的差值进行相加求和, d表示一条光线上相邻两点的差异值总和. 设置阈值 γ, 当差异值大于 γ 表示该条光线处于高密度区域, 需要进行细采样, 否则不进行细采样.

$$N_r=N_{\min}+(N_{\max}-N_{\min})\cdot\min\left(1,\left(\frac{d}{d_t}\right)^k\right) \tag{6}$$

其中, N_r 表示这条光线上最终需要细采样点的最少数量, $N_{\min}$ 表示需要细采样点的最少数量, $N_{\max}$ 表示需要细采样点的最多数量, d_t 用于控制差异值的灵敏度. k 是一个调节系数, 用来控制差异值对重采样数量的影响, $k>1$ 时, 差异值对采样数量的影响变缓, 只有在差异值接近 d_t 时, 重采样数量才会快速增加, $k<1$ 时, 差异值对采样数量的影响更为敏感, 较小的差异值也能引发较大的重采样数量变化. 将差异值 d 输入式 (6), 选取 1 和 $\left(\frac{d}{d_t}\right)^k$ 中的最小值来计算, 根据式 (6) 结果设置对应光线上细采样的数量, 进行后续训练.

3 实验结果与分析

3.1 实验数据集与参数设置

本次实验通过 3 个公共数据集进行对比, 分别为 Realistic Synthetic 360°数据集、Real Forward Facing 数据集和 DTU 数据集. Realistic Synthetic 360°数据集[3]是一个真实渲染的 360°合成数据集, 包括 8 个场景、100 张训练视图、100 张验证视图和 200 张测试视图, 分辨率为 800×800, 本文选取其中的 6 个场景进行实验. Real Forward Facing 数据集[3,20]包含 8 个真实场景, 每个场景有 20–62 张分辨率为 1008×756 的视图, 本文选取其中的 6 个场景进行实验. DTU 数据集[21]是搭载可调节亮度灯的工业机器臂拍摄的室内物体数据集, 包含 128 个场景, 本文随机选取 4 个场景进行实验, 每个场景分配 49 张分辨率为 512×640 的视图.

本次实验在本地服务器上运行, CPU 为 Intel(R) Xeon(R) Silver 4214@2.20 GHz, GPU 为 NVIDIA GeForce RTX 3090, 基于 PyTorch 深度学习框架实现算法. 为实现网络对新视图重建结果的定量分析, 使用峰值信噪比 (PSNR)、结构相似性指数 (SSIM) 和学习感知图像块相似度 (LPIPS) 这 3 个性能指标作为评估新视图重建质量的评价标准, 其中 PSNR 和 SSIM 数值越高、LPIPS 数值越低表示新视图重建效果越好.

3.2 消融实验

为了验证本文所提出的方法在三维重建上的有效性, 使用原始的 NeRF 模型和分别只添加变分自编码器的 NeRF 模型 A, 只优化采样算法的 NeRF 模型 B, 添加变分自编码器并同时优化采样算法的 NeRF 模型 C 进行对比实验, 表 1 展示了 4 个模型在 Realistic Synthetic 360°数据集中 Lego 场景下的实验结果对比, 表 2 展示了 4 个模型在 Real Forward Facing 数据集中 Trex 场景下的实验对比, 其中最佳结果使用加粗字体表示.

表 1 在 Lego 场景下新视图重建的消融实验

模型	PSNR (dB)	SSIM	LPIPS
NeRF	32.54	0.961	0.024
A	34.29	0.968	0.017
B	33.13	0.963	0.022
C	**34.81**	**0.975**	**0.013**

表 2 在 Trex 场景下新视图重建的消融实验

模型	PSNR (dB)	SSIM	LPIPS
NeRF	26.80	0.880	0.057
A	27.91	0.910	0.053
B	26.89	0.877	0.056
C	**28.23**	**0.921**	**0.050**

由表 1 和表 2 的数据对比可以发现, 与原始 NeRF 模型相比, 模型 A 在添加变分自编码器加强对全局场景信息的捕捉之后, 在 Lego 场景和在 Trex 场景下均具有更好的表现, 模型 B 在优化采样算法之后, 在整体结果上都有一定的提升, 说明原模型对简单区域实行统一数值细采样的非必要性, 提出的自适应采样算法能够更好地实现对细节的捕捉, 在 Trex 场景下 SSIM 指数略低于原始 NeRF 可能是因为提出的采样算法更擅长边缘与细节的还原但在平缓区域的低采样降低了局部亮度与对比的一致性, 模型 C 在同时进行两种优化之后达到了最好的效果, 说明本文提出方法相较原始 NeRF 模型能有效提升三维重建效果.

3.3 原始训练数据完整情况下的对比实验

本实验在原始训练数据完整的情况下, 将 VAE-NeRF 与具有相同参数的 NeRF、NeRF-ID[22]和 IP-NeRF[23]进行对比实验, 以展示所提方法在 3 个公开数据集下具有广泛的适用性和有效性, 对比结果如表 3–表 5 所示, 其中最优结果以加粗显示, 次优结果以下划线表示. 图 2 展示 Lego 场景下新视图重建可视化结果.

从表 3–表 5 的结果可以看出, 在 Realistic Synthetic 360°数据集上, VAE-NeRF 相较 NeRF 和 NeRF-ID 在大多数情况都有更好的表现效果, 在 Ficus 和 Lego 数据场景下 VAE-NeRF 达到了最好的效果, VAE-NeRF 在 Drums 和 Materials 场景中训练指标次于 NeRF-ID 和 IP-NeRF, 因为这两个场景几何结构相对简单, VAE 的特征学习优势有限提升较少, VAE-NeRF 与 IP-NeRF 的性能指标在其他场景下的表现各具优势. 在 Real Forward Facing 数据集中, IP-NeRF 在整体上具有最优表现, VAE-NeRF 在所有场景下性能指标都优于 NeRF 和

NeRF-ID, 达到次优表现. 在 DTU 数据集的对比实验结果中可以看出, 在 Scan55 场景下 VAE-NeRF 具有最优表现, 其他场景略低于 IP-NeRF, 相较于 NeRF 和 NeRF-ID 整体提升效果显著. 以上实验结果表明, 尽管 VAE-NeRF 是为缺失数据场景设计的三维重建方法, 但在完整数据条件下依然优于 NeRF 与 NeRF-ID, 并在部分场景下超过较新的 IP-NeRF. 总体而言, 与 IP-NeRF 的差距十分有限: PSNR 和 SSIM 仅低约 1%–2%. LPIPS 的绝对差多数在 0.004 左右, 说明所提方法具备与最新方法接近甚至更优的重建性能.

表 3 Realistic Synthetic 360°数据集实验对比

场景	NeRF			NeRF-ID			IP-NeRF			VAE-NeRF		
	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS
Chair	33.00	0.967	0.019	34.54	0.978	0.014	**35.17**	**0.983**	**0.010**	34.89	0.981	0.012
Drums	25.01	0.925	0.058	25.15	0.926	0.057	**25.80**	**0.931**	**0.051**	25.10	0.925	0.058
Ficus	30.13	0.964	0.022	32.24	0.976	0.015	31.86	0.973	0.016	**32.80**	**0.979**	**0.013**
Hotdog	36.18	0.974	0.016	37.26	0.981	0.013	**38.48**	**0.986**	**0.010**	38.12	0.984	0.011
Lego	32.54	0.961	0.024	34.73	0.974	0.015	34.77	0.974	0.015	**34.81**	**0.975**	**0.013**
Materials	29.62	0.949	0.029	30.37	0.956	0.024	**31.90**	**0.977**	**0.011**	30.12	0.951	0.027

表 4 Real Forward Facing 数据集实验对比

场景	NeRF			NeRF-ID			IP-NeRF			VAE-NeRF		
	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS
Fortress	31.16	0.881	0.030	31.51	0.888	0.028	**32.94**	**0.933**	**0.024**	32.10	0.895	0.025
Horns	27.45	0.828	0.068	27.88	0. 843	0.065	**29.30**	**0.911**	**0.057**	28.40	0.856	0.061
Leaves	20.92	0.690	0.111	21.09	0.708	0.108	**22.53**	**0.825**	**0.100**	21.50	0.720	0.105
Orchids	20.36	0.641	0.121	20.38	0.643	0.120	**21.44**	**0.764**	**0.100**	20.80	0.650	0.118
Room	32.70	0.948	0.041	32.93	0.954	0.039	**33.86**	**0.961**	**0.035**	33.60	0.960	**0.035**
Trex	26.80	0.880	0.057	27.45	0.897	0.051	**28.74**	**0.934**	**0.042**	28.23	0.921	0.050

表 5 DTU 数据集实验对比

场景	NeRF			NeRF-ID			IP-NeRF			VAE-NeRF		
	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS
Scan1	23.49	0.754	0.282	23.80	0.765	0.266	**24.47**	**0.778**	**0.248**	24.30	0.775	0.250
Scan22	21.55	0.708	0.238	21.98	0.715	0.226	**22.68**	**0.758**	**0.196**	22.50	0.725	0.212
Scan55	26.54	0.794	0.229	26.76	0.800	0.219	27.23	**0.812**	0.206	**27.30**	0.810	**0.205**
Scan109	28.33	0.860	0.236	28.63	0.870	0.226	**29.46**	**0.881**	**0.185**	29.10	0.880	0.210

图 2 完整训练数据 Lego 场景的新视图重建可视化结果

3.4 原始训练数据缺失下的对比实验

本实验在 NeRF、NeRF-ID 和 IP-NeRF 具有完整训练数据, 而 VAE-NeRF 使用非完整训练数据进行对比, 验证 VAE-NeRF 在缺失数据的情况下依然可以得到近似其他算法具有完整数据的三维重建效果, 其中 VAE-NeRF 使用完整训练数据的 80% 比例进行实验, 实验结果如表 6–表 8 所示, 其中最优结果以加粗表示, 次优结果以下划线表示.

从表 6–表 8 的实验结果可以看出, 在 Realistic Synthetic 360°数据集实验, IP-NeRF 整体表现最好, 尽管 VAE-NeRF 训练数据减少 20%, 但整体表现仍高于 NeRF 在完整训练数据下的重建结果, 在 Drums 和 Materials 这类较为平滑的场景建模结果相对 NeRF 有所下降, 在 Lego 场景达到次优表现. 在 Real Forward Facing 数据集的对比实验中, IP-NeRF 整体上表现最优, VAE-NeRF 在所有场景的表现与 NeRF-ID 非常接

近, 在 Orchids 场景下实现次优结果, 整体表现都优于 NeRF. 在 DTU 数据集实验中, VAE-NeRF 在 Scan55 和 Scan109 场景下总体超过 NeRF-ID, 具有次优的表现, 在 Scan1 和 Scan22 场景下 VAE-NeRF 略逊于具有完整训练数据的 NeRF-ID 达到的重建效果, IP-NeRF 在所有场景表现最好. 实验结果表明, 本文提出的 VAE-NeRF 虽然在训练时较对比的 3 种方法缺少 20% 的数据, 但在 3 个数据集上的重建结果整体依然优于 NeRF, 并与 NeRF-ID 保持高度接近. 尤其是在 Real Forward Facing 和 DTU 数据集上, VAE-NeRF 与 NeRF-ID 的表现几乎一致, 平均差距约为 1%, 部分场景甚至取得更优效果. 同时, VAE-NeRF 在整体重建性能上也与具备完整训练数据的 IP-NeRF 相差不大, 进一步验证了其在数据缺失情况下依旧具备较强竞争力与鲁棒性.

表 6 Realistic Synthetic 360°数据集实验对比

场景	NeRF			NeRF-ID			IP-NeRF			VAE-NeRF		
	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS
Chair	33.00	0.967	0.019	34.54	0.978	0.014	**35.17**	**0.983**	**0.010**	33.70	0.972	0.016
Drums	25.01	0.925	0.058	25.15	0.926	0.057	**25.80**	**0.931**	**0.051**	24.81	0.920	0.060
Ficus	30.13	0.964	0.022	**32.24**	**0.976**	**0.015**	31.86	0.973	0.016	30.82	0.970	0.018
Hotdog	36.18	0.974	0.016	37.26	0.981	0.013	**38.48**	**0.986**	**0.010**	36.25	0.978	0.014
Lego	32.54	0.961	0.024	34.73	**0.974**	**0.015**	**34.77**	**0.974**	**0.015**	33.57	0.968	0.019
Materials	29.62	0.949	0.029	30.37	0.956	0.024	**31.90**	**0.977**	**0.011**	29.13	0.948	0.027

表 7 Real Forward Facing 数据集

场景	NeRF			NeRF-ID			IP-NeRF			VAE-NeRF		
	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS
Fortress	31.16	0.881	0.030	31.51	0.888	0.028	**32.94**	**0.933**	**0.024**	31.40	0.885	0.028
Horns	27.45	0.828	0.068	27.88	0.843	0.065	**29.30**	**0.911**	**0.057**	27.70	0.838	0.064
Leaves	20.92	0.690	0.111	21.09	0.708	0.108	**22.53**	**0.825**	**0.100**	21.00	0.700	0.109
Orchids	20.36	0.641	0.121	20.38	0.643	0.120	**21.44**	**0.764**	**0.100**	20.50	0.645	0.119
Room	32.70	0.948	0.041	32.93	0.954	0.039	**33.86**	**0.961**	**0.035**	32.80	0.951	0.039
Trex	26.80	0.880	0.057	27.45	0.897	0.051	**28.74**	**0.934**	**0.042**	27.30	0.890	0.053

表 8 DTU 数据集实验对比

场景	NeRF			NeRF-ID			IP-NeRF			VAE-NeRF		
	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS	PSNR (dB)	SSIM	LPIPS
Scan1	23.49	0.754	0.282	23.80	0.765	0.266	**24.47**	**0.778**	**0.248**	23.71	0.760	0.272
Scan22	21.55	0.708	0.238	21.98	0.715	0.226	**22.68**	**0.758**	**0.196**	21.77	0.712	0.229
Scan55	26.54	0.794	0.229	26.76	0.800	0.219	**27.23**	**0.812**	**0.206**	26.90	0.798	0.216
Scan109	28.33	0.860	0.236	28.63	0.870	0.226	**29.46**	**0.881**	**0.185**	28.70	0.865	0.223

从第 3.3、3.4 节的实验结果可以看出, VAE-NeRF 虽然是为应对训练数据缺失场景而设计的模型, 但在训练数据完整的情况下, 其整体重建质量依然稳定超越 NeRF、NeRF-ID 等经典方法, 并在性能上与最新的 IP-NeRF 保持高度接近, 充分验证了所提方法在三维重建任务中的有效性与竞争力. 更具意义的是, 当训练数据缺失 20% 时, VAE-NeRF 的重建结果不仅依旧明显优于 NeRF, 并且与 NeRF-ID 几乎持平, 与 IP-NeRF 的差距也有限. 这表明, 即便在数据不完整的现实应用场景中, VAE-NeRF 依然能够逼近甚至超过现有算法在完整数据下的表现, 展现出较强的鲁棒性与泛化能力. 结合在多类场景上的一致优势, 可以进一步证明本文方法的通用性.

4 结论

本文针对训练数据缺失的情况对神经辐射场模型进行优化, 结合变分自编码器提出一种基于变分自编码器的神经辐射场三维重建方法 VAE-NeRF. 引入潜在空间学习场景的隐含结构, 通过变分自编码器对训练数据进行潜在信息的捕捉, 以此来补充神经辐射场训练时对全局场景信息的获取. 同时结合自适应采样, 提升对细节区域的捕捉能力. 最后, 通过在 3 个公开数据集中对 VAE-NeRF 进行实验训练与测试, 实验结果表明, VAE-NeRF 在具有完整训练数据的情况下能够达到更好的三维重建指标, 同时在训练数据缺失的情况下也能够达到近似对比网络具有完整训练数据的场景重建效果, 证明所提方法的有效性.

参考文献

1 Schönberger JL, Frahm JM. Structure-from-motion revisited. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 4104–4113.

2 Furukawa Y, Ponce J. Accurate, dense, and robust multiview stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(8): 1362–1376. [doi: 10.1109/TPAMI.2009.161]

3 Mildenhall B, Srinivasan PP, Tancik M, *et al*. NeRF: Representing scenes as neural radiance fields for view synthesis. Proceedings of the 16th European Conference on Computer Vision (ECCV). Glasgow: Springer, 2020. 405–421.

4 Barron JT, Mildenhall B, Tancik M, *et al*. Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021. 5835–5844.

5 Yu A, Li RL, Tancik M, *et al*. PlenOctrees for real-time rendering of neural radiance fields. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021. 5732–5741.

6 Hu T, Liu S, Chen YL, *et al*. EfficientNeRF-efficient neural radiance fields. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022. 12892–12901.

7 Gao XY, Yang ZY, Zhao YL, *et al*. A general implicit framework for fast NeRF composition and rendering. Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver: ACM, 2024. 1833–1841.

8 Yu A, Ye V, Tancik M, *et al*. pixelNeRF: Neural radiance fields from one or few images. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021. 4576–4585.

9 Trevithick A, Yang B. GRF: Learning a general radiance field for 3D representation and rendering. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021. 15162–15172.

10 Pumarola A, Corona E, Pons-Moll G, *et al*. D-NeRF: Neural radiance fields for dynamic scenes. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021. 10313–10322.

11 Li ZQ, Niklaus S, Snavely N, *et al*. Neural scene flow fields for space-time view synthesis of dynamic scenes. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021. 6494–6504.

12 Chen JH, Qin YP, Liu LJ, *et al*. NeRF-HuGS: Improved neural radiance fields in non-static scenes using heuristics-guided segmentation. Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2024. 19436–19446.

13 Feng CR, Yu WB, Cheng XH, *et al*. AE-NeRF: Augmenting event-based neural radiance fields for non-ideal conditions and larger scenes. Proceedings of the 39th AAAI Conference on Artificial Intelligence. Philadelphia: AAAI Press, 2025. 326.

14 Martin-Brualla R, Radwan N, Sajjadi MSM, *et al*. NeRF in the wild: Neural radiance fields for unconstrained photo collections. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021. 7206–7215.

15 Barron JT, Mildenhall B, Verbin D, *et al*. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022. 5460–5469.

16 Bi S, Xu ZX, Srinivasan P, *et al*. Neural reflectance fields for appearance acquisition. arXiv:2008.03824, 2020.

17 Cui ZT, Gu L, Sun X, *et al*. Aleth-NeRF: Illumination adaptive NeRF with concealing field assumption. Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver: AAAI Press, 2024. 160.

18 Chen YY, Dong SY, Wang XL, *et al*. SG-NeRF: Neural surface reconstruction with scene graph optimization. Proceedings of the 18th European Conference on Computer Vision (ECCV). Milan: Springer, 2024. 188–205.

19 Kingma DP, Welling M. Auto-encoding variational Bayes. Proceedings of the 2nd International Conference on Learning Representations. Banff: OpenReview.net, 2014.

20 Mildenhall B, Srinivasan PP, Ortiz-Cayon R, *et al*. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. ACM Transactions on Graphics, 2019, 38(4): 29.

21 Jensen R, Dahl A, Vogiatzis G, *et al*. Large scale multi-view stereopsis evaluation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus: IEEE, 2014. 406-413.

22 Arandjelović R, Zisserman A. NeRF in detail: Learning to sample for view synthesis. arXiv:2106.05264, 2021.

23 侯耀斐, 黄海松, 范青松, 等. 基于改进多层感知机的神经辐射场三维重建方法. 激光与光电子学进展, 2024, 61(4): 0415004.

(校对责编: 李慧鑫)