向量-矩阵张量环辐射场新视图合成模型

李莹戈[1],龙 珍[1],苟艺馨[1],林薪雨[2],朱 策[1*]
(1. 电子科技大学信息与通信工程学院,四川成都 611731;2. 重庆大学微电子与通信工程学院,重庆 401331)

摘 要: 基于张量的辐射场方法通过张量回归建立输入(三维空间位置)与输出(体密度、外观特征)之间的映射关系,依托紧凑的场景表示,在保持高质量渲染效果的同时,显著提升了新视图合成效率。然而,现有方法无论是采用传统张量分解还是张量链(Tensor Train,TT)分解,均难以充分挖掘三维场景空间结构信息,对场景深层特征刻画不足。针对这一问题,本文在向量-矩阵(Vector-Matrix,VM)分解框架的基础上,引入张量环(Tensor Ring,TR)分解,提出了向量-矩阵张量环辐射场(Vector-Matrix Tensor Ring Radiance Fields,VMTR-RF)模型用于新视图合成。与现有的张量辐射场方法不同,VMTR分解采用分层建模策略:首先,利用VM分解将场景表示为一系列向量与矩阵因子外积的组合,实现对三维场景的初步紧凑表示;随后,将向量矩阵因子重组为高阶张量,并利用TR分解将其表示为多个核张量构成的张量环网络,从而更充分地捕获三维场景深层特征信息。得益于VMTR分解的优势,VMTR-RF在体密度估计和外观特征学习方面表现出更强的建模能力;最后,利用体渲染技术,结合学习到的体密度与外观特征合成新视图。实验结果表明,VMTR-RF优于现有最先进方法,尤其在保持细节方面表现突出,能够更好地重建锐利边缘、复杂结构和自然纹理,在保持紧凑场景表示的同时实现了更高质量的新视图合成结果。

关键词: 神经辐射场;张量网络;新视图合成;紧凑表示;VMTR分解
基金项目: 四川省自然科学基金(No.2025ZNSFSC0002);国家自然科学基金(No.62401102,No.62401112)
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(XXXX)XX-0001-12
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20260225

Vector-Matrix Tensor Ring Radiance Fields for Novel View Synthesis

LI Yingge[1], LONG Zhen[1], GOU Yixin[1], LIN Xinyu[2], ZHU Ce[1*]
(1. *School of Information and Communication Engineering*, *University of Electronic Science and Technology of China*, *Chengdu*, *Sichuan* 611731, *China*; 2. *School of Microelectronics and Communication Engineering*, *Chongqing University*, *Chongqing* 401331, *China*)

Abstract: Tensor-based radiance field methods established a mapping between inputs (3D spatial positions) and outputs (volume density and appearance features) via tensor regression. These methods relied on compact scene representations and significantly improved the efficiency of novel view synthesis while maintaining high-quality rendering results. However, existing approaches, whether based on conventional tensor decomposition or tensor train (TT) decomposition, are unable to fully exploit structural information in 3D scene space, thereby limiting the representation of deep-level features. To address this issue, we introduced tensor ring (TR) decomposition into the vector-matrix (VM) decomposition framework and proposed a vector-matrix tensor ring radiance fields (VMTR-RF) model for novel view synthesis. Unlike existing tensor radiance field methods, VMTR decomposition adopted a hierarchical modeling strategy: VM decomposition was first used to represent the scene as a combination of outer products of multiple vector and matrix factors, enabling an initial compact representation of the 3D scene. The vector-matrix factors were then reorganized into high-order tensors and further decomposed using TR decomposition, resulting in a tensor ring network composed of multiple core tensors, thereby enabling more effective capture of deep-level features in 3D scenes. Benefiting from the VMTR decomposition, VMTR-RF exhibited stronger modeling capability in volume density estimation and appearance feature learning. Finally, novel view synthesis was performed using volume rendering by combining the learned volume density and appearance features. Experimental results demonstrated that VMTR-RF outperformed existing state-of-the-art methods, particularly in detail preservation, enabling better reconstruction of sharp edges, complex structures, and natural textures, while achieving higher-quality novel view synthesis with a compact scene representation.

Keywords: neural radiance fields; tensor network; novel view synthesis; compact representation; VMTR decomposition
Foundation Item(s): Natural Science Foundation of Sichuan Province (No. 2025ZNSFSC0002); National Natural Science Foundation of China (No.62401102, No.62401112)

收稿日期:2026-04-13;录用日期:2026-05-20;责任编委:李甲
*通信作者:朱策

0 引言

新视图合成旨在根据一组输入图像及其对应摄像机位姿重建三维场景,并生成任意视角下的高质量图像。作为计算机视觉与图形学中的重要研究方向,新视图合成在虚拟现实(Virtual Reality,VR)、增强现实(Augmented Reality,AR)、计算机游戏及视频内容生成等领域[1-3]具有广泛的应用价值,近年来受到持续关注。高质量新视图合成要求模型能够准确重建场景的几何结构,并充分建模场景外观、光照变化和视角相关效应。

为此,传统方法和基于学习的方法先后探索了多种三维场景表示方式,可分为显式场景表示和隐式场景表示两类。显式场景表示主要依赖离散几何元素对场景结构进行刻画。例如,网格[4-5]表示通过顶点、边与三角面片对场景表面进行建模,具有结构清晰、渲染管线成熟等优点,因而被广泛应用于计算机图形学与三维重建任务中。然而,网格表示依赖于高质量的几何表面重建,因此在处理拓扑结构复杂的场景时,其表达能力受到一定限制。点云[6]表示通过离散三维点集直接描述场景几何,具有表示简洁和灵活性高等优势。但由于点云缺乏显式拓扑关系和连续表面约束,其在高保真渲染方面仍存在局限。体素[7-8]表示通过将三维场景划分为规则离散网格,并在每个体素单元中存储对应的场景属性,从而实现对三维场景的结构化表示。但其存储开销往往较大,分辨率受限。相较而言,隐式场景表示则通过连续映射函数对场景进行建模,能够更自然地刻画复杂结构与细粒度细节。神经辐射场(Neural Radiance Fields,NeRF)[9]利用多层感知机(Multi-Layer Perceptron,MLP)网络学习从空间位置坐标及视角方向到体密度与颜色的连续映射函数,实现了对场景几何和外观统一建模,并结合体渲染技术将网络输出的颜色和体密度投影生成图像,在新视图合成任务中获得了高真实感的渲染效果。凭借其出色的表示能力,NeRF及其变体迅速推动了该领域的发展,并被广泛应用于场景重建、三维感知及编辑生成等多个任务[10-15]。

然而,NeRF高度依赖MLP对空间中大量采样点进行逐点查询和计算,带来了较高的训练与渲染开销。针对上述问题,近年来,研究者探索了更加高效的场景建模方式,主要包括基于量化的方法[16-17]以及显式-隐式结合的混合表示方法[18-21]。其中,基于张量的混合神经辐射场方法通过将场景构建为高阶张量,并利用张量分解加速渲染,实现了紧凑的场景表示和高效的新视图合成,因而受到了广泛关注。具体而言,张量辐射场(Tensorial Radiance Fields,TensoRF)[22]将三维场景构建为高阶张量,并利用CP(CANDECOMP/PARAFAC)分解与VM分解对场景紧凑表示,在大幅提高渲染速度的同时,实现了超越NeRF的渲染质量。进一步地,StriVec[23]在TensoRF的基础上进一步将场景划分为多个局部区域,并利用CP分解对每个局部区域进行紧凑表示,增强了局部结构表达能力。BTD-RF[24]则利用块分解(Block Term Decomposition,BTD)与多线性注意力机制,提高了场景表示的紧凑性和渲染速度。尽管上述方法在三维场景紧凑表示与高效重建方面取得了显著进展,但其本质上仍依赖于结构较为简单的传统张量分解模型,这类方法虽然具有较高的计算效率,但往往只能捕捉浅层的场景信息,难以充分刻画场景深层特征,从而在一定程度上限制了视图合成质量。相较经典张量分解模型,张量网络(Tensor Network,TN)近年来被认为是一类更具潜力的高阶数据表示工具[25-26]。TN通过将高阶张量分解为多个相互连接的低阶核张量,能够更有效地捕获张量不同模式之间的相关性。在众多张量网络模型中,张量链(Tensor Train,TT)分解[27]是最简单且应用广泛的方法之一。TT分解已成功应用于压缩辐射场,例如TT-NF[28]、PuTT[29]以及QDLR-NeRF[30],而TT-TSDF[31]则进一步利用TT分解对符号距离场进行压缩表示。

然而,TT分解本身也存在一定局限:一方面,其线性链式结构容易导致秩分布不均衡,中间核张量通常需要承担更多的信息交互与传递,因而往往具有更高的秩,而两端核张量的表示能力相对受限。这种秩不平衡不仅增加了计算负担,也限制了模型的表达能力;另一方面,TT分解对张量维度顺序较为敏感,不同的维度顺序可能显著影响最终分解效果[32-33]。为克服这些不足,研究者进一步提出了张量环(Tensor Ring,TR)分解[32]。TR分解通过将高阶张量表示为一组环状连接的低阶核张量,有效消除了TT分解中的边界限制,使不同模式之间的信息交互更加灵活,也在一定程度上缓解了秩不均衡和维度顺序敏感问题。近期提出的块张量环分解(Block Tensor Ring Decomposition,BTRD)[34],通过将高阶张量表示为多个向量因子与(N-1)阶张量因子外积项的求和形式,并进一步将(N-1)阶张量因子分解为张量环网络,实现了高效的低秩表示。

受此启发,本文提出一种新的张量分解模型即向量-矩阵张量环(Vector-Matrix Tensor Ring,VMTR)分解,并将其应用于新视图合成任务。具体而言,首先,利用向量-矩阵(Vector-Matrix,VM)分解将目标张量表示为一系列向量与矩阵因子外积项的组合,从而获得高效的初步压缩表示;随后,为进一步挖掘因子之间潜在的高阶相关性,将多个向量矩阵因子重组为

高阶张量,并通过TR分解将其表示为由多个核张量构成的张量环网络。基于该分解模型,本文进一步提出向量-矩阵张量环辐射场(Vector-Matrix Tensor Ring Radiance Fields,VMTR-RF)新视图合成方法,分别将三维场景体密度建模为三阶张量,将场景外观特征建模为四阶张量,并采用VMTR分解对其进行紧凑表示。得益于VMTR分解的优势,VMTR-RF能够更充分捕获三维场景深层特征信息,有效提升体密度估计和外观特征学习能力,为新视图合成任务带来更高的渲染质量。

实验结果表明,VMTR-RF相较于对比的方法在3个数据集上取得了一致性能的提升,在峰值信噪比(Peak Signal-to-Noise Ratio,PSNR)、结构相似性(Structural Similarity Index Measure,SSIM)以及感知相似度(Learned Perceptual Image Patch Similarity,LPIPS)等指标上表现出显著优势。同时,在定性的可视化结果方面,所提方法合成的新视图在细节保真度方面表现优越,尤其在自然纹理、复杂结构以及边缘轮廓的重建效果上具有明显改善。

综上所述,本文的主要贡献可以总结如下:

(1)本文提出了一种新的VMTR分解模型:首先,采用VM分解将目标张量分解为一系列向量与矩阵因子外积的组合,从而实现高效地初步压缩表示;进一步地,将多个向量-矩阵因子重组为高阶张量,并利用TR分解将其表示为多个核张量构成的张量环网络,进一步捕捉目标张量不同模式之间的相关信息。

(2)将VMTR分解引入新视图合成任务并进一步提出VMTR-RF方法:分别对体密度张量与外观特征张量进行高效紧凑表示,以进一步挖掘场景深层特征信息。

(3)在多个公开数据集进行实验验证:证明了所提方法在定量指标和定性视觉效果上均优于现有对比方法,尤其在复杂纹理和精细结构恢复方面具有明显优势。

1 符号及定义

1.1 符号

标量、向量、矩阵和张量分别用$x,\boldsymbol{x},\boldsymbol{X}$和$\mathcal{X}$表示,索引通常从1取到其对应的大写字母,例如$i=1,2,\cdots,I$。

定义1 (模式d乘积):给定任意张量$\mathcal{X}\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_D}$与矩阵$\boldsymbol{A}\in\mathbb{R}^{J\times I_d}$,其中,$d=1,2,\cdots,D$,两者之间的模式$d$乘积定义为$\mathcal{Y}=\mathcal{X}\times_d\boldsymbol{A}$。其中,符号$\times_d$表示张量与矩阵的模式$d$乘积,张量$\mathcal{Y}\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_{d-1}\times J\times I_{d+1}\times\cdots\times I_D}$与原始张量$\mathcal{X}$在第$d$个模式上不同,其元素定义为$y_{i_1,i_2,\cdots,i_{d-1},j,i_{d+1},\cdots,i_D}=\sum_{i_d=1}^{I_d}x_{i_1,i_2,\cdots,i_{d-1},i_d,i_{d+1},\cdots,i_D}a_{j,i_d}$。

定义2 (逐元素乘积[35]):设$\mathcal{A},\mathcal{B}\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_D}$为两个同阶同形状的张量,则逐元素乘积表示对它们对应位置的元素分别相乘得到一个与二者相同形状的张量,定义为$\mathcal{C}=\mathcal{A}\odot\mathcal{B}$。其中,$\odot$表示逐元素乘积,张量$\mathcal{C}\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_D}$的元素定义为$c_{i_1,i_2,\cdots,i_D}=a_{i_1,i_2,\cdots,i_D}b_{i_1,i_2,\cdots,i_D}$。

1.2 VMTR分解

定义3 (TR分解[32]):设一个D阶张量$\mathcal{X}\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_D}$,其TR分解可写作如式(1)所示:

$$\mathcal{X}(i_1,i_2,\cdots,i_D)=\mathrm{trace}\left(\mathcal{G}_1(i_1)\mathcal{G}_2(i_2)\cdots\mathcal{G}_D(i_D)\right) \tag{1}$$

其中,$\mathcal{G}_d\in\mathbb{R}^{R_{d-1}\times I_d\times R_d}$为第$d$个三阶核张量,并且满足环状维度约束$R_1=R_{D+1}$;$\mathcal{G}_d(i_d)$为张量$\mathcal{G}_d$的第$i_d$个侧向切片矩阵。图1给出了TR分解的示意图,为简化后续表示与推导,本文将TR分解统一记为$\mathcal{X}=\mathrm{TR}(\mathcal{G}_1,\mathcal{G}_2,\cdots,\mathcal{G}_D)=\mathrm{TR}\left(\{\mathcal{G}^x\}\right)$。

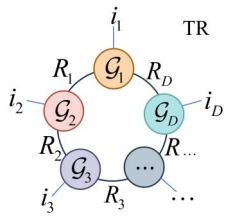

图1 TR分解示意图

Figure 1 The graphical illustration of TR decomposition

定义4 (VMTR分解):如图2所示,对一个三阶张量$\mathcal{F}\in\mathbb{R}^{I\times J\times K}$,VM[22]分解将$\mathcal{F}$表示为3组向量矩阵的外积求和形式,如式(2)所示:

$$\mathcal{F}=\sum_{r=1}^{R_1}\boldsymbol{v}_r^1\circ\boldsymbol{M}_r^{2,3}+\sum_{r=1}^{R_2}\boldsymbol{v}_r^2\circ\boldsymbol{M}_r^{1,3}+\sum_{r=1}^{R_3}\boldsymbol{v}_r^3\circ\boldsymbol{M}_r^{1,2} \tag{2}$$

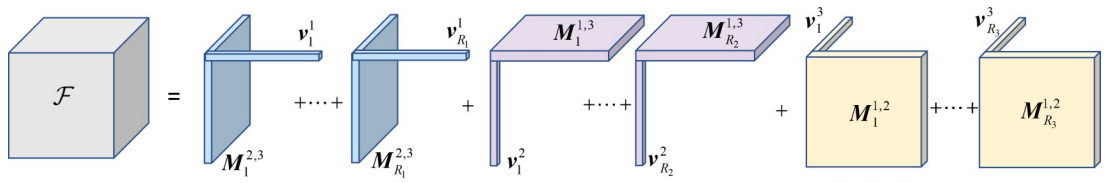

图2 VM分解示意图

Figure 2 The graphical illustration of VM decomposition

其中，$\boldsymbol{v}_r^1\in\mathbb{R}^I$，$\boldsymbol{v}_r^2\in\mathbb{R}^J$，$\boldsymbol{v}_r^3\in\mathbb{R}^K$为3个模式上的向量因子；$\boldsymbol{M}_r^{2,3}\in\mathbb{R}^{J\times K}$，$\boldsymbol{M}_r^{1,3}\in\mathbb{R}^{I\times K}$，$\boldsymbol{M}_r^{1,2}\in\mathbb{R}^{I\times J}$为对应的矩阵因子；$R_1,R_2,R_3$为3个模式对应的VM秩；$\circ$代表外积。

为了便于后续表示，本文将同一模式的因子进行堆叠，即将矩阵因子堆叠为三阶张量$\mathcal{M}^{2,3}\in\mathbb{R}^{J\times K\times R_1}$，其第$r$个切片满足$\mathcal{M}^{2,3}(:,:,r)=\boldsymbol{M}_r^{2,3}$，同理得到$\mathcal{M}^{1,3}\in\mathbb{R}^{I\times K\times R_2}$与$\mathcal{M}^{1,2}\in\mathbb{R}^{I\times J\times R_3}$；将向量因子堆叠为矩阵$\boldsymbol{V}^1\in\mathbb{R}^{I\times R_1}$，其第$r$列满足$\boldsymbol{V}^1(:,r)=\boldsymbol{v}_r^1$，同理得到$\boldsymbol{V}^2\in\mathbb{R}^{J\times R_2}$与$\boldsymbol{V}^3\in\mathbb{R}^{K\times R_3}$。则式(2)可等表示为

$$\mathcal{F}=\mathrm{P}_1\left(\mathcal{M}^{2,3}\times_3\boldsymbol{V}^1\right)+\mathrm{P}_2\left(\mathcal{M}^{1,3}\times_3\boldsymbol{V}^2\right)+\mathrm{P}_3\left(\mathcal{M}^{1,2}\times_3\boldsymbol{V}^3\right) \tag{3}$$

其中，$\times_3$表示模式3乘积；$\mathrm{P}(\cdot)$表示将每一模式结果重排为重构张量$\mathcal{F}$的维度顺序。为统一记号，令$\mathcal{M}^{\bar{n}}$表示第n个模式的矩阵堆叠张量，$\boldsymbol{V}^n$表示第n个模式的向量堆叠矩阵，则有简写形式：

$$\mathcal{F}=\sum_{n=1}^{3}\mathrm{P}_n\left(\mathcal{M}^{\bar{n}}\times_3\boldsymbol{V}^n\right) \tag{4}$$

在VM分解基础上，为了进一步利用张量网络对高阶结构的建模能力，本文通过reshape将$\mathcal{M}^{\bar{n}}$和$\boldsymbol{V}^n$重塑为高阶张量$\mathcal{M}_h^{\bar{n}}$和$\mathcal{V}_h^n$（注：reshape是一种可逆的维度重排操作。以矩阵$\boldsymbol{A}\in\mathbb{R}^{I\times R}$为例，在满足$I=I_1I_2$的条件下，可通过reshape将其重塑为高阶张量$\mathcal{A}_h\in\mathbb{R}^{I_1\times I_2\times R}$）。随后，对$\mathcal{M}_h^{\bar{n}}$和$\mathcal{V}_h^n$施加定义3中的TR分解，于是$\mathcal{M}_h^{\bar{n}}=\mathrm{TR}\left(\{\mathcal{G}^{\mathcal{M}_h^{\bar{n}}}\}\right)$，$\mathcal{V}_h^n=\mathrm{TR}\left(\{\mathcal{G}^{\mathcal{V}_h^n}\}\right)$。进一步本文定义VMTR分解为

$$\mathcal{F}=\sum_{n=1}^{3}\mathrm{P}_n\begin{pmatrix}\mathrm{R}^{-1}\left(\mathrm{TR}\left(\{\mathcal{G}^{\mathcal{M}_h^{\bar{n}}}\}\right)\right)\\ \times_3\mathrm{R}^{-1}\left(\mathrm{TR}\left(\{\mathcal{G}^{\mathcal{V}_h^n}\}\right)\right)\end{pmatrix} \tag{5}$$

其中，$n\in\{1,2,3\}$；$\mathrm{R}^{-1}(\cdot)$代表reshape的逆操作。为了简化表示，本文进一步将式(5)简化记为

$$\mathcal{F}=\sum_{n=1}^{3}\mathrm{T}\left(\{\mathcal{G}^{\mathcal{M}_h^{\bar{n}}}\},\{\mathcal{G}^{\mathcal{V}_h^n}\}\right) \tag{6}$$

本文给出三阶张量VMTR分解的示意图如图3所示。

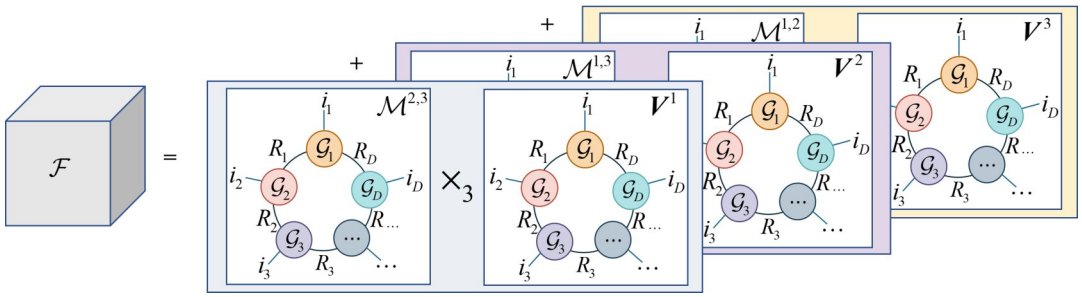

图3 VMTR分解示意图

Figure 3 The graphical illustration of VMTR decomposition

2 方法

2.1 NeRF

NeRF[9]通过MLP将三维位置$\boldsymbol{x}=[x,y,z]$和方向$\boldsymbol{d}=[\theta,\phi]$映射为体密度$\sigma$和辐射颜色$\boldsymbol{c}=[R,G,B]$。在可微体渲染过程中，给定一条起始点为$\boldsymbol{o}$、方向为$\boldsymbol{d}$的射线$r$，通过在$\boldsymbol{x}_q=\boldsymbol{o}+t_q\boldsymbol{d}$处对射线进行顺序采样，可以得到对应采样点密度$\{\sigma_q\}_{q=1}^{Q}$和颜色$\{\boldsymbol{c}_q\}_{q=1}^{Q}$，然后进一步通过式(7)估算出该射线对应的像素颜色$\hat{\boldsymbol{c}}(r)$为

$$\begin{cases}\hat{\boldsymbol{c}}(r)=\sum_{q=1}^{Q}\tau_q\left(1-\exp(-\sigma_q\Delta_q)\right)\boldsymbol{c}_q,\\ \tau_q=\exp\left(-\sum_{p=1}^{q-1}\sigma_p\Delta_p\right)\end{cases} \tag{7}$$

其中，$\Delta_q=t_{q+1}-t_q$表示相邻采样点之间的距离；τ_q表示透射率。为实现场景重建，NeRF通过最小化渲染像素颜色与真实像素颜色$\boldsymbol{c}(r)$之间的均方误差(Mean Square Error，MSE)来更新网络权重，损失函数定义为

$$\mathcal{L}_{\text{render}}=\left\|\boldsymbol{c}(r)-\hat{\boldsymbol{c}}(r)\right\|_2^2 \tag{8}$$

2.2 VMTR辐射场

类似于TensoRF[22]，本文首先使用一个三阶张量$\mathcal{F}_\sigma\in\mathbb{R}^{X\times Y\times Z}$以及一个四阶张量$\mathcal{F}_c\in\mathbb{R}^{X\times Y\times Z\times P}$，将输入位置$\boldsymbol{x}$分别映射为体密度$\sigma$和外观特征$\boldsymbol{w}\in\mathbb{R}^P$。然后，外观特征向量$\boldsymbol{w}$和视角方向$\boldsymbol{d}$通过MLP映射为颜色$\boldsymbol{c}$，上述过程可表示为

$$\begin{cases}\sigma=\mathcal{F}_\sigma(\boldsymbol{x}),\\ \boldsymbol{c}=\mathrm{MLP}(\boldsymbol{w},\boldsymbol{d}),\boldsymbol{w}=\mathcal{F}_c(\boldsymbol{x})\end{cases} \tag{9}$$

为获得更加紧凑的张量表示，本文进一步采用定义4的VMTR分解对$\mathcal{F}_\sigma$和$\mathcal{F}_c$进行参数化。对于三阶体密度张量$\mathcal{F}_\sigma$，有：

$$\mathcal{F}_\sigma=\sum_{n=1}^{3}\mathrm{T}\left(\{\mathcal{G}^{\mathcal{M}_h^{\bar{n}}}\},\{\mathcal{G}^{\mathcal{V}_h^n}\}\right) \tag{10}$$

对于四阶外观特征张量$\mathcal{F}_c$，有：

$$\mathcal{F}_c=\oplus_{n=1}^{3}\mathrm{T}^{*}\left(\{\mathcal{G}^{\mathcal{M}_h^n}\},\{\mathcal{G}^{\mathcal{V}_h^n}\}\right) \tag{11}$$

本文进一步说明四阶外观特征张量(即式(11))的分解形式。需要指出的是,该分解并非由三阶张量的VMTR分解直接进行高阶推广得到,而是基于定义4的式(6)进行两处调整后构建。具体地:

(1)不再对 $\mathcal{M}^n$ 和 $\boldsymbol{V}^n$ 执行模式3乘积($\times_3$),而是通过对它们执行"维度扩张+广播机制[35]+逐元素乘积($\odot$)"替代。因此,式(6)中的 $\mathrm{T}(\cdot)$ 被进一步记作 $\mathrm{T}^{*}(\cdot)$。

(2)不再将3个模式分支的结果相加求和,而是将3项结果沿最后一维进行拼接($\oplus$),从而构造四阶外观特征张量 $\mathcal{F}_c$。因此,式(6)中的 $\sum_{n=1}^{3}(\cdot)$ 被进一步记作 $\oplus_{n=1}^{3}(\cdot)$。

2.3 优化与重建

图4展示了VMTR-RF重建与渲染的整体流程。通过对TensoRF进行VMTR分解,本文将场景体密度张量 $\mathcal{F}_\sigma$ 与场景外观特征张量 $\mathcal{F}_c$ 转化为一系列紧凑的低阶因子张量。给定任意空间坐标 $\boldsymbol{x}$ 与视角方向 $\boldsymbol{d}$,模型通过插值获取对应的体密度值 σ 和外观特征向量 $\boldsymbol{w}$,并通过MLP预测得到体渲染所需的体素颜色 $\boldsymbol{c}$。结合式(9)、式(10)和式(11),VMTR-RF可表示为

$$\begin{cases}\sigma=\sum_{n=1}^{3}\mathrm{T}\left(\{\mathcal{G}^{\mathcal{M}_h^n}\},\{\mathcal{G}^{\mathcal{V}_h^n}\}\right),\\ \boldsymbol{c}=\mathrm{MLP}\left(\oplus_{n=1}^{3}\mathrm{T}^{*}\left(\{\mathcal{G}^{\mathcal{M}_h^n}\},\{\mathcal{G}^{\mathcal{V}_h^n}\}\right),\boldsymbol{d}\right)\end{cases} \tag{12}$$

该模型在实现场景紧凑表征的同时,支持高质量的三维重建。在模型优化阶段,本文采用2.1节的可微体渲染方法合成图像。给定一组具有已知相机位姿的多视角图像,通过Adam优化器对因子张量及MLP参数进行优化。

此外,为防止模型在训练过程中出现过拟合及在梯度下降过程中陷入局部最小化问题,本文在密度因子张量上施加了 $\mathcal{L}_1$ 稀疏正则化约束。最终的损失函数可表示为

$$\min \mathcal{L}_{\mathrm{render}}+\omega\left\|\mathcal{G}_\sigma\right\|_1 \tag{13}$$

其中,$\mathcal{L}_{\mathrm{render}}$ 为式(8)定义的渲染损失;$\mathcal{G}_\sigma$ 表示体密度因子张量;$\|\cdot\|_1$ 为 $\mathcal{L}_1$ 范数;ω 为稀疏正则项权重系数。

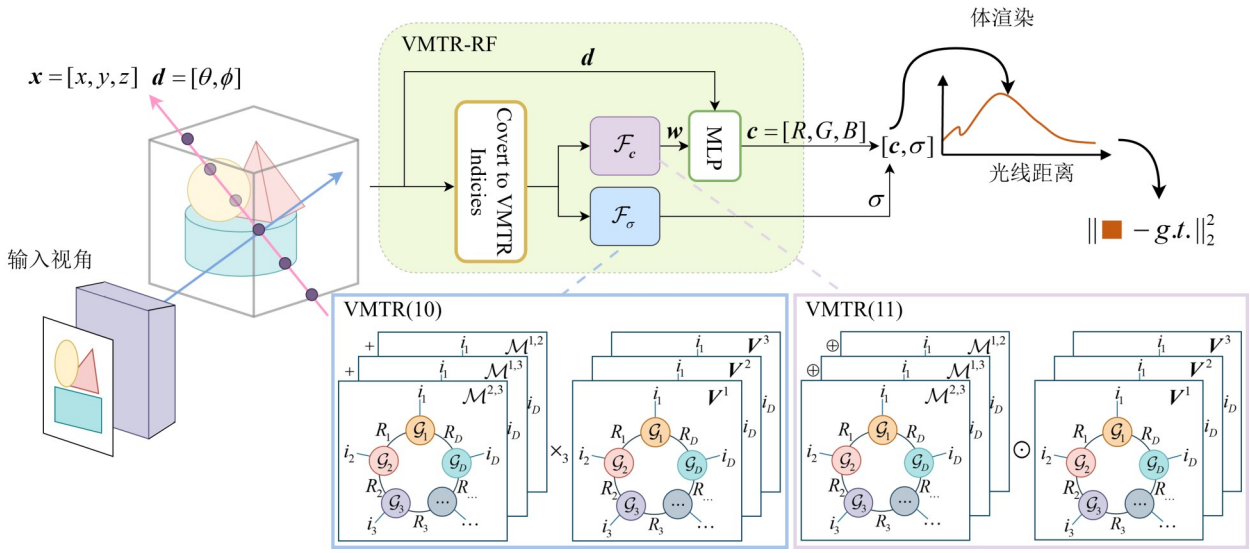

图4 VMTR-RF重建与渲染整体流程图

Figure 4 VMTR-RF reconstruction and rendering pipeline

3 实验

3.1 实验设置

数据集描述:本文使用3个数据集进行实验评估,分别为Synthetic-NeRF[9]、NSVF[18]以及Tanks-Temple[36]。其中,Synthetic-NeRF数据集包含若干结构相对简单的合成三维场景,主要用于模型的初步验证;NSVF数据集包含8个合成场景,图像分辨率为800 × 800,用于进一步评估模型在合成场景中的渲染质量;TanksTemple数据集包含5个场景,图像分辨率为1 920 × 1 080,主要用于检验模型在重建复杂真实场景时的泛化能力与鲁棒性。

对比方法:本文将所提方法与3种先进的基于张量的模型进行对比,分别为TensoRF-CP-384、TensoRF-VM和PuTT-M。所有实验均在同一硬件环境下进行,即使用一张NVIDIA GeForce RTX 3090显

卡(24 GB显存),以确保实验结果的公平性。

实现细节:本文方法采用与TensoRF和PuTT相同的coarse-to-fine的训练策略。模型初始使用100 × 100 × 100分辨率的体素网格,并在第2 000、3 000、4 000、5 500和7 000次迭代时逐步上采样,最终达到400 × 400 × 400的分辨率。TensoRF-CP-384、TensoRF-VM、PuTT-M与VMTR-RF的训练迭代次数分别设置为30 000、30 000、80 000和30 000,对应的初始学习率分别为0.02、0.02、0.000 8和0.001,$\mathcal{L}_1$稀疏正则项的权重则均设置为10^{-5},batch size设置为4 096、8 192、4 096和8 192,其余参数设置均遵循其官方实现。

此外,本文进一步对秩超参数做如下设置与说明:为平衡表达能力与参数规模,VMTR分解在3个模式的VM秩R_1, R_2, R_3统一取$R_1 = R_2 = R_3 = 16$,矩阵因子TR秩依次设置为$[64, 64, 32, 64, 64]$,向量因子的TR秩依次设置为$[64, 8, 64]$,本文所有实验均在上述设置下进行。另一方面,为保证与本文方法在参数规模上的可比性,本文将TensoRF-VM基线3个模式VM秩统一设为5。

3.2 渲染性能分析

表1从PSNR、SSIM以及LPIPS指标对不同方法在Synthetic-NeRF、NSVF和TanksTemple 3个数据集上的性能进行了对比,同时给出了各方法对应的batch size、训练迭代次数以及模型大小。NeRF的实验结果在可获取情况下直接引用自其原始论文。可以看出,本文所提出的VMTR-RF在3个数据集上均取得了稳定且具有竞争力的表现,在PSNR和SSIM指标上整体优于NeRF、TensoRF-CP、TensoRF-VM和PuTT,同时在LPIPS指标上取得了更低的数值,表明其在结构一致性和感知质量方面均具有明显优势。在3个数据集上相较排名第二的方法PSNR平均分别提升了0.31 dB、0.65 dB和0.36 dB,SSIM平均分别提升了0.003、0.003和0.005,LPIPS平均分别降低了0.004、0.002和0.005。在Synthetic-NeRF和NSVF数据集上,尽管TensoRF-VM以及PuTT在部分场景上表现接近或略优,但VMTR-RF仍在PSNR、SSIM和LPIPS 3项指标上保持领先。

表1 本文方法与已有方法在3个数据集上的性能比较

Table 1 Comparison of our method's performance with previous methods across the three datasets

Method	BatchSize	Steps	Size (MB)	Synthetic-NeRF			NSVF			TanksTemple		
				PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
NeRF[9]	4096	100k	5.00	31.01	0.947	0.081	30.81	0.952	0.043	25.78	0.864	0.198
TensoRF-CP[22]	4096	30k	3.8	31.71	0.950	0.075	34.58	0.972	0.044	27.88	0.899	0.177
TensoRF-VM[22]	8192	30k	12.0	<u>32.23</u>	<u>0.955</u>	<u>0.060</u>	35.31	<u>0.976</u>	0.034	<u>28.14</u>	0.909	<u>0.155</u>
PuTT[29]	4096	80k	10.6	31.43	0.950	0.067	<u>35.32</u>	<u>0.976</u>	<u>0.033</u>	<u>28.14</u>	<u>0.910</u>	0.156
VMTR-RF	8192	30k	11.6	**32.54**	**0.958**	**0.056**	**35.97**	**0.979**	**0.031**	**28.50**	**0.915**	**0.150**

注:加粗表示在该项指标中排名第一,下划线表示排名第二。

表2~表4展示了VMTR-RF与对比方法在3个数据集的PSNR对比结果。总体来看,VMTR-RF在不同数据集和场景中PSNR指标大多排名第一,其余部分排名第二。

表2 Synthetic-NeRF数据集各场景的PSNR对比

Table 2 PSNR comparison of each scene on Synthetic-NeRF dataset

Method	Chair	Drums	Ficus	Hotdog	Lego	Materials	Mic	Ship	Avg
NeRF[9]	33.00	25.01	30.13	36.18	32.54	29.62	32.91	28.65	31.01
TensoRF-CP[22]	<u>34.10</u>	25.26	30.78	36.31	34.34	**30.17**	<u>33.95</u>	28.78	31.71
TensoRF-VM[22]	**34.68**	<u>25.34</u>	<u>32.88</u>	<u>36.66</u>	<u>35.37</u>	29.16	33.75	<u>29.97</u>	<u>32.23</u>
PuTT[29]	33.61	25.20	31.70	36.27	33.33	29.24	33.24	28.84	31.43
VMTR-RF	**34.68**	**25.66**	**33.23**	**37.15**	**35.57**	<u>29.80</u>	**34.20**	**30.03**	**32.54**

注:加粗表示在该项指标中排名第一,下划线表示排名第二。

在Synthetic-NeRF和NSVF数据集上,VMTR-RF在Wineholder和Palace场景中较排名第二的方法提升1.00 dB和0.88 dB,而在其他大多数场景中排名第一。在更复杂的TanksTemples数据集上,VMTR-RF在所有场景均取得最高PSNR,显示了其对复杂场景的优秀重建能力。结果表明VMTR-RF在不同数据集与复杂度场景下均能实现稳定高质量重建,验证了其在新视图合成任务中的良好适用性。

表5~表7展示了不同方法在Synthetic-NeRF、NSVF以及TanksTemple数据集各场景下的SSIM对比

表3 NSVF数据集各场景的PSNR对比

Table 3 PSNR comparison of each scene on NSVF dataset

Method	Bike	Lifestyle	Palace	Robot	Spaceship	Steamtrain	Toad	Wineholder	Avg
NeRF	31.77	31.08	31.76	28.69	34.66	30.84	29.42	28.23	30.81
TensoRF-CP	36.83	32.77	36.46	36.09	37.27	35.84	31.45	29.96	34.58
TensoRF-VM	38.01	33.87	36.30	37.16	36.92	**36.70**	33.19	30.35	35.31
PuTT	38.48	34.19	36.16	36.40	**38.44**	34.55	**33.76**	30.56	35.32
VMTR-RF	**38.64**	**34.25**	**37.34**	**37.60**	38.35	36.56	33.47	**31.56**	**35.97**

注:加粗表示在该项指标中排名第一,下划线表示排名第二。

表4 TanksTemple数据集各场景的PSNR对比

Table 4 PSNR comparison of each scene on TanksTemple dataset

Method	Ignatius	Truck	Barn	Caterpillar	Family	Avg
NeRF	25.43	25.36	24.05	23.75	30.29	25.78
TensoRF-CP	28.20	26.30	26.88	25.39	32.62	27.88
TensoRF-VM	28.28	26.81	26.81	25.52	33.27	28.14
PuTT	28.11	26.88	27.07	25.57	33.08	28.14
VMTR-RF	**28.33**	**26.93**	**27.55**	**26.13**	**33.54**	**28.50**

注:加粗表示在该项指标中排名第一,下划线表示排名第二。

表5 Synthetic-NeRF数据集各场景的SSIM对比

Table 5 SSIM comparison of each scene on Synthetic-NeRF dataset

Method	Chair	Drums	Ficus	Hotdog	Lego	Materials	Mic	Ship	Avg
NeRF	0.967	0.925	0.964	0.974	0.961	0.949	0.980	0.856	0.947
TensoRF-CP	0.977	0.921	0.965	0.974	0.972	**0.950**	0.983	0.855	0.950
TensoRF-VM	**0.979**	0.925	0.976	0.978	0.978	0.938	0.984	**0.880**	0.955
PuTT	0.972	0.924	0.972	0.976	0.968	0.942	0.982	0.862	0.950
VMTR-RF	**0.979**	**0.931**	**0.979**	**0.979**	**0.979**	0.948	**0.985**	**0.880**	**0.958**

注:加粗表示在该项指标中排名第一,下划线表示排名第二。

表6 NSVF数据集各场景的SSIM对比

Table 6 SSIM comparison of each scene on NSVF dataset

Method	Bike	Lifestyle	Palace	Robot	Spaceship	Steamtrain	Toad	Wineholder	Avg
NeRF	0.970	0.946	0.950	0.960	0.980	0.966	0.920	0.920	0.952
TensoRF-CP	0.988	0.953	0.973	0.990	0.985	0.986	0.951	0.948	0.972
TensoRF-VM	0.990	0.961	0.972	0.992	0.983	0.988	0.968	0.950	0.976
PuTT	**0.991**	**0.964**	0.970	0.991	**0.988**	0.983	**0.971**	0.952	0.976
VMTR-RF	**0.991**	**0.964**	**0.978**	**0.993**	**0.988**	**0.989**	0.969	**0.960**	**0.979**

注:加粗表示在该项指标中排名第一,下划线表示排名第二。

表7 TanksTemple数据集各场景的SSIM对比

Table 7 SSIM comparison of each scene on TanksTemple dataset

Method	Ignatius	Truck	Barn	Caterpillar	Family	Avg
NeRF	0.920	0.860	0.750	0.860	0.932	0.864
TensoRF-CP	0.937	0.885	0.841	0.884	0.949	0.899
TensoRF-VM	0.943	0.901	0.846	0.898	0.958	0.909
PuTT	0.941	0.901	0.851	0.900	0.956	0.910
VMTR-RF	**0.945**	**0.904**	**0.863**	**0.904**	**0.959**	**0.915**

注:加粗表示在该项指标中排名第一,下划线表示排名第二。

结果。结果显示，VMTR-RF在3个数据集上均取得了最高的平均SSIM值，整体优于NeRF、TensoRF-CP、TensoRF-VM和PuTT。在Synthetic-NeRF与NSVF数据集上，VMTR-RF在多数场景中保持领先，体现了其对场景结构的高效建模能力。在TanksTemple数据集上，尽管整体SSIM水平有所下降，VMTR-RF仍保持稳定优势，表明其在复杂的真实场景中同样能够有效地捕捉场景结构细节。

表8~表10展示了VMTR-RF与对比方法在3个数据集上的LPIPS对比结果（数值越低表示感知质量越好）。结果表明，VMTR-RF在3个数据集上均取得了最低的平均LPIPS，整体优于其他方法。在Synthetic-NeRF数据集上，该方法在Drums、Ficus以及Hotdog等场景中取得最低或并列最低LPIPS值；在NSVF数据集上，在Bike、Palace、Robot以及Wineholder场景中表现最优；在TanksTemple数据集上，在Ignatius、Barn以及Family场景中同样取得最低LPIPS值，体现了其在不同场景下良好的感知建模能力。

表8 Synthetic-NeRF数据集各场景的LPIPS对比

Table 8 LPIPS comparison of each scene on Synthetic-NeRF dataset

Method	Chair	Drums	Ficus	Hotdog	Lego	Materials	Mic	Ship	Avg
NeRF	0.046	0.091	0.044	0.121	0.050	**0.063**	0.028	0.206	0.081
TensoRF-CP	0.041	0.111	0.059	0.051	0.036	0.068	0.033	0.197	0.075
TensoRF-VM	**0.029**	0.096	0.035	**0.039**	0.026	0.079	0.021	**0.156**	0.060
PuTT	0.040	0.092	0.035	0.044	0.042	0.075	0.024	0.184	0.067
VMTR-RF	0.030	**0.084**	**0.027**	**0.039**	**0.024**	0.067	**0.020**	0.157	**0.056**

注：加粗表示在该项指标中排名第一，下划线表示排名第二。

表9 NSVF数据集各场景的LPIPS对比

Table 9 LPIPS comparison of each scene on NSVF dataset

Method	Bike	Lifestyle	Palace	Robot	Spaceship	Steamtrain	Toad	Wineholder	Avg
NeRF	0.019	**0.047**	0.031	0.038	**0.016**	0.031	0.069	0.096	0.043
TensoRF-CP	0.022	0.079	0.029	0.016	0.028	0.031	0.065	0.081	0.044
TensoRF-VM	0.015	0.058	0.028	0.013	0.027	**0.023**	0.046	0.064	0.034
PuTT	0.014	0.056	0.032	0.013	0.020	0.033	**0.039**	0.059	0.033
VMTR-RF	**0.013**	0.057	**0.023**	**0.011**	0.021	**0.023**	0.042	**0.055**	**0.031**

注：加粗表示在该项指标中排名第一，下划线表示排名第二。

表10 TanksTemple数据集各场景的LPIPS对比

Table 10 LPIPS comparison of each scene on TanksTemple dataset

Method	Ignatius	Truck	Barn	Caterpillar	Family	Avg
NeRF	0.111	0.192	0.395	0.196	0.098	0.198
TensoRF-CP	0.098	0.203	0.281	0.220	0.084	0.177
TensoRF-VM	0.084	**0.165**	0.276	0.179	0.072	0.155
PuTT	0.087	0.169	0.270	**0.177**	0.077	0.156
VMTR-RF	**0.081**	0.169	**0.248**	0.182	**0.069**	**0.150**

注：加粗表示在该项指标中排名第一，下划线表示排名第二。

此外，在Barn场景中，VMTR-RF相较于PuTT方法LPIPS下降最为显著，达到了0.022，进一步体现了其在复杂场景中的优势。与此同时，由于Tanks-Temple数据集整体复杂度较高，各方法的LPIPS结果均有所上升。然而，所提出的VMTR-RF方法相较其他方法仍取得最低的平均LPIPS值，说明该方法在复杂真实场景中依然能够稳定合成具有较高感知质量的新视图。以上结果进一步验证了所提方法良好的泛化能力。

3.3 效率分析

本文对比了不同方法在NSVF数据集上Palace场景（分辨率为800 × 800）的训练时间与渲染速度（Frames Per Second，FPS），如表11所示。实验结果表明，VMTR-RF在该场景上的训练时间约为0.76 h，而TensoRF-CP、TensoRF-VM与PuTT分别约为0.37 h、0.34 h和2.65 h。在渲染速度方面，VMTR-RF约为0.27 FPS，TensoRF-CP、TensoRF-VM与PuTT分别约为0.30 FPS、0.36 FPS和0.15 FPS。

表11 Palace场景效率对比

Table 11 Efficiency comparison on the Palace scene

Method	Training Time/h	FPS
TensoRF-CP	~0.37	~0.30
TensoRF-VM	**~0.34**	**~0.36**
PuTT	~2.65	~0.15
VMTR-RF	~0.76	~0.27

注:加粗表示在该项指标中排名第一,下划线表示排名第二。

可以看出,本文方法在训练与渲染阶段相较于TensoRF-VM和TensoRF-CP效率有所下降。造成上述差异的主要原因在于TR分解的引入在一定程度上增加了计算复杂度。尽管本文方法在效率方面有所下降,但在PSNR、SSIM和LPIPS等指标上均取得了稳定提升。综上,本文所提方法在提升模型表达能力的同时,带来了一定的计算开销。

3.4 消融实验

为验证VMTR分解及高阶维度重排(High-order Reshaping,HR)策略的有效性,本节设计了消融实验,对比引入二者对新视图合成质量的影响。实验对比了3种方法:(1)TensoRF-VM:采用VM分解的基线方法;(2)VMTR-RF(w/o HR):采用VMTR分解但未引入高阶维度重排策略;(3)VMTR-RF:采用VMTR分解并引入高阶维度重排策略的完整模型。为保证公平对比,所有方法均采用相同的训练策略、损失函数和评价指标。

在NSVF数据集上的实验结果如表12和表13所示。首先,对比方法(2)与(1),将传统VM分解替换为VMTR分解后,平均PSNR提升0.48 dB,SSIM和LPIPS也均有所改善,单场景PSNR最高提升0.96 dB。同时,模型参数量从12.0 MB降至11.6 MB,验证了VMTR分解在提升新视图合成质量方面的有效性。其次,对比方法(3)与(2),引入高阶维度重排策略后,在保持模型参数量大小相同的条件下,平均PSNR进一步提升0.18 dB,单场景PSNR最高提升0.61 dB。这表明高阶维度重排策略能够在不增加参数开销的前提下,进一步提升新视图合成质量。总体而言,本文方法通过VMTR分解与高阶维度重排策略,在NSVF数据集上较对比的基线方法实现了0.66 dB的平均PSNR提升。二者共同作用,能够更充分地挖掘场景深层特征信息,从而实现更优的新视图合成效果。

表12 NSVF数据集消融实验结果

Table 12 Ablation study results on NSVF dataset

Method	High-order Reshaping	Size/MB	PSNR	SSIM	LPIPS
TensoRF-VM	×	12.0	35.31	0.976	0.034
VMTR-RF(w/o HR)	×	11.6	35.79	0.978	**0.031**
VMTR-RF	√	11.6	**35.97**	**0.979**	**0.031**

注:加粗表示在该项指标中排名第一,下划线表示排名第二。

表13 NSVF数据集消融实验各场景PSNR对比

Table 13 PSNR comparison of ablation study on NSVF dataset across different scenes

Method	Bike	Lifestyle	Palace	Robot	Spaceship	Steamtrain	Toad	Wineholder	Avg
TensoRF-VM	38.01	33.87	36.30	37.16	36.92	**36.70**	33.19	30.35	35.31
VMTR-RF(w/o HR)	38.41	33.99	37.26	**37.64**	37.74	36.66	33.31	31.25	35.79
VMTR-RF	**38.64**	**34.25**	**37.34**	37.60	**38.35**	36.56	**33.47**	**31.56**	**35.97**

注:加粗表示在该项指标中排名第一,下划线表示排名第二。

3.5 可视化分析

为直观验证所提方法在新视图合成任务中的重建效果,本文在合成数据集和真实数据集上的代表性场景中对不同方法进行可视化对比,从复杂结构重建、锐利边缘保持和自然纹理恢复3个维度评估不同方法保持细节的性能。

合成数据集:本文在合成数据集上的4个代表性场景对不同方法进行可视化对比。其中,NSVF数据集的Wineholder场景和Synthetic-NeRF数据集的Lego场景用于评估不同方法对复杂结构与锐利边缘的重建能力;NSVF数据集的Steamtrain场景和Synthetic-NeRF数据集的Drums场景则用于比较不同方法对于自然纹理的保持效果。

图5展示了真实图像(Ground Truth,GT)、VMTR-RF与对比方法在局部放大区域的新视图合成效果。从对比结果可以看出,在Wineholder场景中,本文方法对酒架表面密集的细小镂空结构实现了高精度的重建,边缘更加锐利,结构更为真实;相比之下,PuTT方法合成的视图效果稍差,而TensoRF-VM和TensoRF-CP方法则出现了明显的伪影和过度平滑现象。在Lego场景中,本文方法能够更好地恢复乐高模型中齿轮等小型组件的几何结构,边缘清晰锐利,而其他方法合成结果明显较差。在Steamtrain场景与Drums场景中,本文方法恢复了更清晰的火车玻璃表面和鼓面的纹理细节,相比之下其他方法的视图合成结果存在明显的模糊与伪影。

真实数据集:本文在TanksTemple数据集上的Barn和Caterpillar两个代表性场景中对不同方法进行可视化对比。其中,Barn场景用于评估不同方法对复杂结构与锐利边缘的重建能力,Caterpillar场景则同

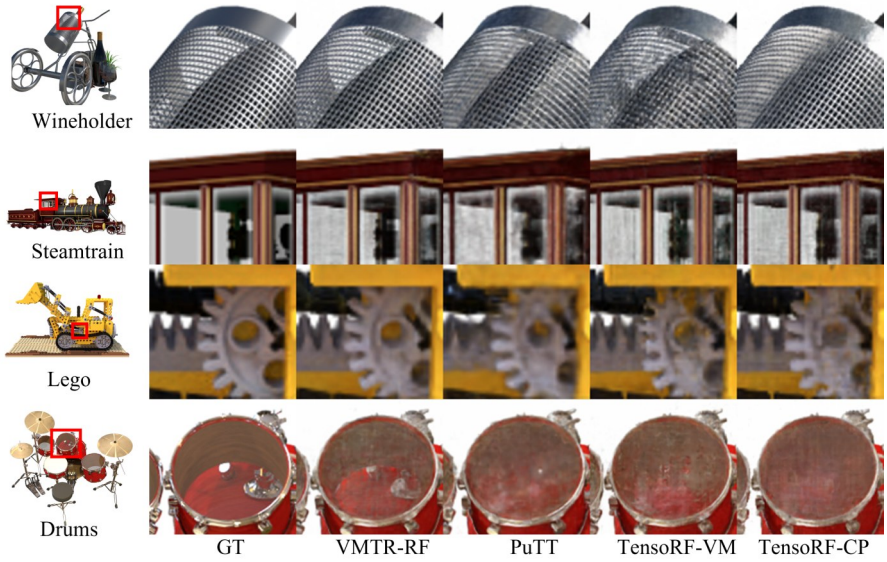

图5 合成数据集可视化结果对比

Figure 5 Comparison of visualization results on synthetic datasets

时用于比较不同方法对复杂结构和自然纹理的保持效果。

图6中展示了GT、VMTR-RF和对比方法合成的新视图放大区域效果图。从结果可以看出，在Barn场景中，本文方法能够更好地重建密集的房顶瓦片的结构，边缘清晰锐利。相比之下，PuTT和TensoRF-VM方法合成的结果存在一定模糊和过度平滑，TensoRF-CP方法的表现最差，出现了明显的伪影。在Caterpillar场景中，本文方法合成的字母纹理更加清晰，散热百叶窗等复杂结构也得以较好保留，与GT更加接近。相比之下，对比方法的合成视图与GT差距较大，视觉质量较低。

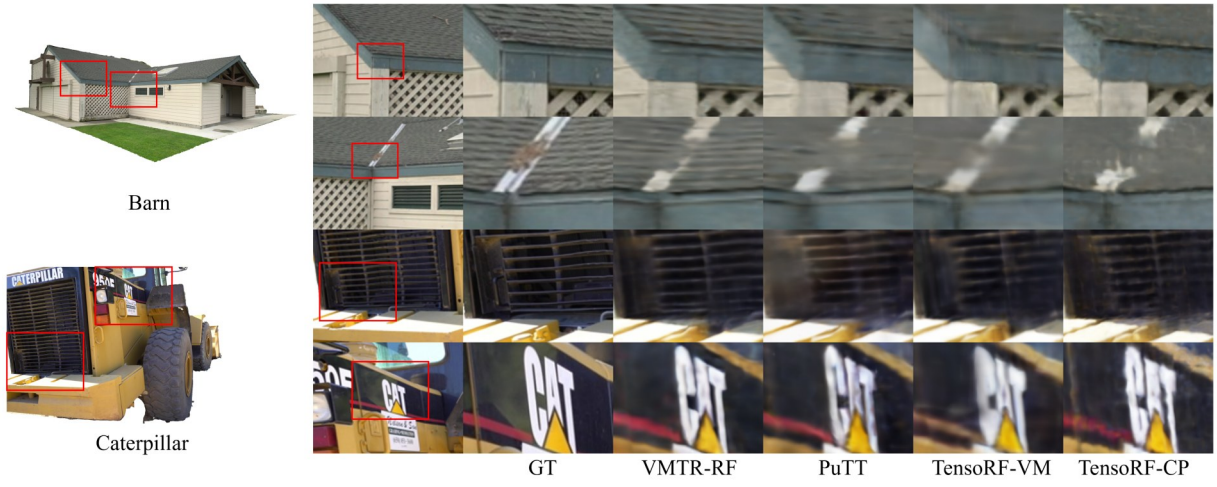

图6 真实数据集可视化结果对比

Figure 6 Comparison of visualization results on real datasets

综上，可视化结果表明，无论是在合成数据集还是真实数据集上，VMTR-RF在细节保持方面均展现出显著优势。相较于对比方法，VMTR-RF能够更准确地重建复杂几何结构、保持锐利边缘并恢复自然纹理，合成的新视图在视觉质量上更接近真实的图像。

4 结论

本文提出了一种张量辐射场表示方法VMTR-RF，通过引入VMTR分解，构建了一种具有分层建模特性的辐射场表示框架。与现有的张量分解方法不同，VMTR分解在VM分解框架下引入TR分解，显著增强了对高阶张量不同模式相关性的建模能力。得

益于VMTR分解因子间多重交互,VMTR-RF能够更加充分地捕获三维场景体密度以及外观特征的深层特征信息,从而合成高质量的渲染图像。实验结果表明,VMTR-RF在多个公开数据集上均取得了一致且稳定的性能提升,在PSNR、SSIM和LPIPS评价指标上优于现有的最先进的方法,验证了VMTR-RF的有效性和泛化能力。同时,定性可视化结果显示,所提方法能够更好地保持边缘结构和纹理细节。综上,VMTR-RF作为一种高效的新视图合成框架,通过紧凑的场景表示实现了出色的视觉重建效果,同时在不同场景的重建中展现出良好的泛化能力。

参考文献

[1] Riegler G, Koltun V. Free view synthesis[C]//Computer Vision - ECCV 2020. Cham: Springer, 2020: 623-640.

[2] Cai Jintong, Lu Huimin. NeRF-based multi-view synthesis techniques: A survey[C]//2024 International Wireless Communications and Mobile Computing. Piscataway: IEEE, 2024: 208-213.

[3] Gao Chen, Saraf A, Kopf J, et al. Dynamic view synthesis from dynamic monocular video[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 5692-5701.

[4] Groueix T, Fisher M, Kim V G, et al. A papier-mache approach to learning 3D surface generation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 216-224.

[5] Wang Nanyang, Zhang Yinda, Li Zhuwen, et al. Pixel2Mesh: Generating 3D mesh models from single RGB images[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 55-71.

[6] Charles R Q, Su Hao, Mo Kaichun, et al. PointNet: Deep learning on point sets for 3D classification and segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 77-85.

[7] Ji Mengqi, Gall J, Zheng Haitian, et al. SurfaceNet: An end-to-end 3D neural network for multiview stereopsis[C]//2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2326-2334.

[8] Qi C R, Su Hao, Nießner M, et al. Volumetric and multi-view CNNs for object classification on 3D data[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 5648-5656.

[9] 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.

[10] Chan E R, Monteiro M, Kellnhofer P, et al. Pi-GAN: Periodic implicit generative adversarial networks for 3D-aware image synthesis[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 5795-5805.

[11] Martin-Brualla R, Radwan N, Sajjadi M S M, et al. NeRF in the wild: Neural radiance fields for unconstrained photo collections[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 7206-7215.

[12] Schwarz K, Liao Yiyi, Niemeyer M, et al. GRAF: Generative radiance fields for 3D-aware image synthesis[PP/OL]. V4. arXiv (2021-03-30) [2026-04-10]. https://doi.org/10.48550/arXiv.2007.02442.

[13] Xiang Fanbo, Xu Zexiang, Hasan M, et al. NeuTex: Neural texture mapping for volumetric neural rendering[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 7115-7124.

[14] Pumarola A, Corona E, Pons-Moll G, et al. D-NeRF: Neural radiance fields for dynamic scenes[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 10313-10322.

[15] Adamkiewicz M, Chen T, Caccavale A, et al. Vision-only robot navigation in a neural radiance world[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 4606-4613.

[16] Gordon C, Chng S F, MacDonald L, et al. On quantizing implicit neural representations[C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2023: 341-350.

[17] Zhong Hongliang, Zhang Jingbo, Liao Jing. VQ-NeRF: Neural reflectance decomposition and editing with vector quantization[J]. IEEE Transactions on Visualization and Computer Graphics, 2024, 30(9): 6247-6260.

[18] Liu Lingjie, Gu Jiatao, Lin K Z, et al. Neural sparse voxel fields[PP/OL]. V2. arXiv (2021-01-06) [2026-04-10]. https://doi.org/10.48550/arXiv.2007.11571.

[19] Yu A, Li Ruilong, Tancik M, et al. PlenOctrees for real-time rendering of neural radiance fields[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 5732-5741.

[20] Fridovich-Keil S, Yu A, Tancik M, et al. Plenoxels: Radiance fields without neural networks[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 5491-5500.

[21] 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.

[22] Chen Anpei, Xu Zexiang, Geiger A, et al. TensoRF: Tensorial radiance fields[C]//Computer Vision - ECCV 2022. Cham: Springer, 2022: 333-350.

[23] Gao Quankai, Xu Qiangeng, Su Hao, et al. Strivec: Sparse tri-vector radiance fields[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2023: 17523-17533.
[24] Kim S B, Kim S, Ahn D, et al. BTD-RF: 3D scene reconstruction using block-term tensor decomposition[J]. Applied Intelligence, 2024, 54(8): 6319-6332.
[25] Cichocki A, Lee N, Oseledets I, et al. Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions[J]. Foundations and Trends in Machine Learning, 2016, 9(4/5): 249-429.
[26] Cichocki A. Era of big data processing: A new approach via tensor networks and tensor decompositions[PP/OL]. V4. arXiv (2014-08-24) [2026-04-10]. https://doi. org/10.48550/arXiv.1403.2048.
[27] Bengua J A, Phien H N, Tuan H D, et al. Efficient tensor completion for color image and video recovery: Low-rank tensor train[J]. IEEE Transactions on Image Processing, 2017, 26(5): 2466-2479.
[28] Obukhov A, Usvyatsov M, Sakaridis C, et al. TT-NF: Tensor train neural fields[J]. IEEE Journal of Selected Topics in Signal Processing, 2024, 18(6): 1024-1035.
[29] Loeschcke S, Wang Dan, Leth-Espensen C, et al. Coarse-to-fine tensor trains for compact visual representations[PP/OL]. V1. arXiv (2024-06-06) [2026-04-10]. https://doi. org/10.48550/arXiv.2406.04332.
[30] Shi Jinglei, Guillemot C. Light field compression *via* compact neural scene representation[C]//ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2023: 1-5.
[31] Boyko A I, Matrosov M P, Oseledets I V, et al. TT-TSDF: Memory-efficient TSDF with low-rank tensor train decomposition[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE, 2020: 10116-10121.
[32] Zhao Qibin, Zhou Guoxu, Xie Shengli, et al. Tensor ring decomposition[PP/OL]. V1. arXiv (2016-06-17)[2026-04-10]. https://doi.org/10.48550/arXiv.1606.05535.
[33] Long Zhen, Zhu Ce, Liu Jiani, et al. Bayesian low rank tensor ring for image recovery[J]. IEEE Transactions on Image Processing, 2021, 30: 3568-3580.
[34] Liu Sheng, Zhao Xile, Zhang Hao. Block tensor ring decomposition: Theory and application[J]. IEEE Transactions on Signal Processing, 2025, 73: 3029-3043.
[35] Matsui Y, Yokota T. Broadcast product: Shape-aligned element-wise multiplication and beyond[PP/OL]. V1. arXiv (2024-09-26)[2026-04-10]. https://doi.org/10.48550/arXiv.2409.17502.
[36] Knapitsch A, Park J, Zhou Qianyi, et al. Tanks and temples: Benchmarking large-scale scene reconstruction[J]. ACM Transactions on Graphics, 2017, 36(4): 1-13.

作者简介

李莹戈 女，2002年12月出生于河南省鹤壁市。现为电子科技大学信息与通信工程学院硕士研究生。主要研究方向为张量信号处理、三维场景重建。
E-mail: 202322011928@std.uestc.edu.cn

龙 珍 女，1993年6月出生于四川省内江市。现为电子科技大学信息与通信工程学院副教授、硕士生导师。主要研究方向为张量信号处理、三维场景重建。
E-mail: zhen.long@uestc.edu.cn

苟艺馨 女，1999年10月出生于甘肃省庆阳市。现为电子科技大学信息与通信工程学院博士研究生。主要研究方向为高维数据表征、三维场景重建。
E-mail: gyx@std.uestc.edu.cn

林薪雨 男，1991年12月出生于四川省德阳市。现为重庆大学微电子与通信工程学院副教授、硕士生导师。主要研究方向为视觉高精度定位、计算机视觉与信号处理、具身智能(无人驾驶与工业机器人)。
E-mail: xinyulin@cqu.edu.cn

朱 策 男，1969年9月出生于四川省自贡市。现为电子科技大学信息与通信工程学院教授、博士生导师。主要研究方向为计算机图像与视频处理。
E-mail: eczhu@uestc.edu.cn