激光与光电子学进展 先进成像

基于改进多层感知机的神经辐射场三维重建方法

侯耀斐[1]，黄海松[1,2*]，范青松[1]，肖婧[1]，韩正功[1]

[1]贵州大学现代制造技术教育部重点实验室，贵州 贵阳 550025；

[2]重庆机电职业技术大学信息工程学院，重庆 402760

摘要 相比传统的三维重建方法，神经辐射场（NeRF）在隐式三维重建方面显示出了优异的性能，然而简单的多层感知机（MLP）模型在采样过程中缺乏局部信息，产生细节模糊的三维重建场景。为解决这一问题，提出一种基于MLP的多特征联合学习方法。首先，在NeRF嵌入层和采样层之间构造多特征联合学习（MFJL）模块，有效解码输入的多视图编码数据，补充MLP模型缺失的局部特征信息。然后，在NeRF采样层和推理层之间建立门控通道变换多层感知机（GCT-MLP）模块，学习高阶特征交互关系，并控制反馈给MLP层的信息流，实现对歧义特征的选择。实验结果表明：所提基于改进MLP的神经辐射场可以避免三维重建中的视图模糊和混叠现象；在Real Forward-Facing数据集部分场景上的平均峰值信噪比（PSNR）、结构相似度（SSIM）、学习感知图像块相似度（LPIPS）分别为28.08 dB、0.887、0.061；在Realistic Synthetic 360°数据集部分场景上的PSNR、SSIM、LPIPS分别为32.75 dB、0.960、0.026；在DTU数据集部分场景上的PSNR、SSIM、LPIPS分别为25.96 dB、0.807、0.208；与NeRF相比，具有更好的视图重建性能，并且在主观视觉效果上得到更加清晰的图像和细节纹理特征。

关键词 神经辐射场；多层感知机；联合学习；三维重建

中图分类号 TN911.73 **文献标志码** A **DOI**：10.3788/LOP223312

3D Reconstruction of Neural Radiation Field Based on Improved Multiple Layer Perceptron

Hou Yaofei[1], Huang Haisong[1,2*], Fan Qingsong[1], Xiao Jing[1], Han Zhenggong[1]

[1]*Key Laboratory of Advanced Manufacturing Technology of the Ministry of Education, Guizhou University, Guiyang* 550025, *Guizhou, China*;

[2]*Information Engineering Institute, Chongqing Vocational and Technical University of Mechatronics, Chongqing* 402760, *China*

Abstract Neural radiation field (NeRF) exhibits excellent performances in implicit 3D reconstruction compared with traditional 3D reconstruction methods. However, the simple multilayer perceptron (MLP) model lacks local information in the sampling process, resulting in a fuzzy 3D reconstruction scene. To solve this issue, a multifeature joint learning (MFJL) method based on MLP is proposed in this study. First, an MFJL module was constructed between the embedding layer and the sampling layer of NeRF to effectively decode the multiview encoded input and supplement the missing local information of MLP model. Then, a gated channel transformation MLP (GCT-MLP) module was built between the sampling layer and the inference layer of NeRF to learn the interaction relations between higher-order features and control the information flow fed back to the MLP layer for the selection of ambiguous features. The experimental results reveal that the NeRF based on the improved MLP can avoid blurred views and aliasing in 3D reconstruction. The average peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and learned perceptual image patch similarity (LPIPS) values on the Real Forward-Facing dataset are 28.08 dB, 0.887, and 0.061; on the Realistic Synthetic 360° dataset are 32.75 dB, 0.960, and 0.026; and on the DTU dataset are 25.96 dB, 0.807, and 0.208, respectively. Overall, the proposed method has a better view reconstruction performance and can obtain clearer images and detailed texture features in

收稿日期：2022-12-13；**修回日期**：2023-01-13；**录用日期**：2023-03-01；**网络首发日期**：2023-03-09

基金项目：贵州省科技计划项目（黔科合平台人才-GCC[2022]006-1，黔科合支撑[2022]一般165，黔科合支撑[2021]一般445，黔科合支撑[2021]一般172，黔科合支撑[2021]一般397，黔科合支撑[2022]一般008）、重庆市自然科学基金面上项目（CSTB2022NSCQ-MSX1600）

通信作者：*hshuang@gzu.edu.cn

subjective visual effects compared with NeRF.

Key words neural radiation field; multiple-layer perceptron; joint learning; 3D reconstruction

1 引 言

三维重建广泛应用于虚拟现实/增强现实(VR/AR)、医疗影像、机器人导航、智慧城市等多个领域与行业,是处理复杂环境下游任务的基础。基于结构光原理的三维扫描仪虽然可以实现较高精度的三维重建,但此类设备的价格昂贵且操作繁琐[1-2]。过去十年,人们将基于物理的多视图几何技术集成到基于深度学习的方法中,直接使用神经网络从二维观测中推断三维场景表示,进而用于三维场景重建任务[3-6]。

目前多视图重建所采用的方法分为几何显式方法和神经隐式方法[7],其中显式方法有基于体素的方法[8]、基于点云的方法[9]、基于曲面网格的方法[10]等。虽然这些方法可以有效地呈现物体的三维特征,但它们通常只对局部区域进行密集采样与重建,除了需要大量的图像输入,查询3D几何先验的方式也增加了对视图重建的内存需求[11]。相比较而言,依赖于隐式神经表征的视图重建方法因场景表征存储小、与视图分辨率无关等优势而应用前景广泛[12]。近年来重建效果优异的隐式方法之一是神经辐射场(NeRF)[13]。NeRF采用由粗到细的采样策略,将一组稀疏视图输入多层感知机(MLP)模型,隐式地将物体的颜色和密度从三维空间位置映射到二维像素点,实现了高质量的新视图重建,但NeRF简单的线性全连接层采样方法会造成局部信息的缺失,导致重建视图的模糊和混叠[14]。针对上述问题,Liu等[15]使用稀疏体素八叉树结构建模局部属性,实现了对重建视图的快速渲染,但由于预定义的体素不能表示无界的三维空间,无法处理真实场景。Zhang等[16]证明了特定的MLP结构可以有效避免像素点隐式映射中可能出现的歧义解的结论,并提出了一种反球面参数化的方法,解决了无界场景渲染困难的问题,整体渲染结果较好,但捕获的局部细节有所欠缺。Trevithick等[17]构造了一个通用辐射场来学习每个像素的局部信息,并引入注意力机制来聚合多个视图的像素特征,隐式地解决了视觉遮挡的问题,但边缘局部特征表达仍有提升的空间。Arandjelović等[18]通过引入尺度较小的MLP建议网络,提出一种联合训练方法,该方法有良好的视图重建性能,但其采样网络与NeRF一致,依然存在渲染模糊和混叠的现象。Yang等[19]在网络的不同阶段递归地应用不同数量线性层的MLP结构,同时采用阈值控制的方法优化采样策略,实现了对不同场景的自适应渲染,但忽略了细节信息,导致重建质量受限。Wang等[20]将整个光线作为网络输入进行信息采样,利用Transformer编码器捕获光线中样本点之间的内在依赖关系,进而促进重建,但Transformer高昂的训练成本增加了实现难度。Fang等[21]放弃粗采样阶段的神经辐射场,构建了一个轻量化的神经样本场,将光线样本分布转换为三维点坐标,提升了采样效率,但未提取全局信息与局部信息之间的依赖关系。

以上文献虽然对NeRF的MLP采样网络做了不同程度的改进,但对于网络训练中多尺度特征信息提取与融合方面的研究仍较少,难以解决局部细节出现频次较少时的特征稀疏或缺失问题。本文在NeRF端到端训练方式的基础上,提出一种基于MLP的多特征联合学习方法,对其简单的全连接层模型进行优化,提高网络提取细粒度特征和抵抗重建退化的能力。主要工作如下:1)在粗采样阶段,设计了多特征联合学习(MFJL)模块,利用归一化MLP瓶颈结构和离散余弦变换(DCT)解码空间全局特征和频域局部特征,并利用无参注意力融合两种特征,为网络提供更丰富的多尺度特征信息,缓解位置编码到定长向量转换造成的特征信息损失的问题;2)在细采样阶段,设计了一个增大网络感受野的标准化中间层(NIL)模块,缓解局部混淆情况以提升模型拟合粗采样特征的能力,纠正粗采样阶段三维特征中错误的部分;3)在特征推理过程中,提出一种门控通道变换多层感知机(GCT-MLP)模块学习高阶特征交互关系,通过优化门控单元,提升网络对像素区域高权重特征的筛选能力。

2 神经辐射场

NeRF结构分为正余弦位置编码网络、特征提取网络和体积渲染网络,整体网络结构如图1所示。其中,特征提取网络包含2个MLP模型,一个为8层全连接的采样MLP模型,另一个为3层全连接的推理MLP模型。

NeRF接收一组同光照条件下某一物体或场景的静态图片。网络输入是由三维位置$X=(x,y,z)$和基于球坐标的二维观看方向$d=(\theta,\varphi)$组成的连续5维坐标。利用MLP网络F_θ,沿采样点提取并推理定向发射的光线在相应像素位置上与视图相关的颜色c和体积密度σ:

$$F_\theta:(X,d)\rightarrow(c,\sigma)。\tag{1}$$

相机成像可以看作光线从像素点出发穿过物体并计算其颜色值来渲染观察到的图像的过程。NeRF通过将相机的光线集成到像素,利用体积渲染方程沿光线独立计算每个采样点的颜色,最终重建图像中像素的颜色$C(r)$:

$$C(r)=\int_0^{+\infty}\sigma[r(t)]\cdot c[r(t),d]\cdot T(t)\mathrm{d}t,\tag{2}$$

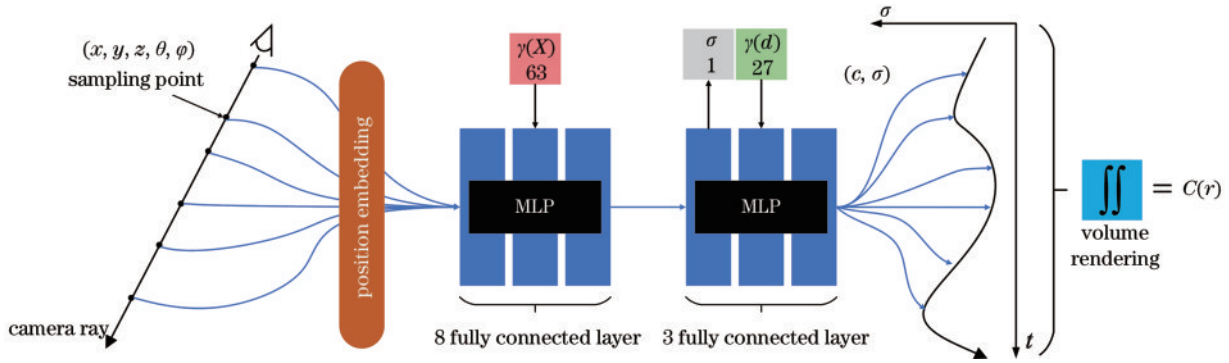

图 1　NeRF 网络的结构

Fig. 1　Structure of NeRF network

式中：$r(t)=o+td$ 表示从相机原点 o 沿二维观看方向 d 发射的光线；t 表示采样长度。$T(t)$ 表示从 0 到 t 的累计透射率，表达式为

$$T(t)=\exp\left\{-\int_0^t\sigma\left[r(s)\right]\mathrm{d}s\right\},\tag{3}$$

式中：s 表示光线长度。

为了使 MLP 更好地拟合高频信息以补偿网络的光谱偏差，NeRF 使用位置编码 γ 在采样前将观测图像中每一个像素点的三维位置 X 和对应相机的姿态 d 都映射到高维空间，使得携带图像信息的光线样本在傅里叶空间转换为高维正弦和余弦信号：

$$\gamma(p)=\left[\sin(2^0\pi p),\cos(2^0\pi p),\cdots,\sin(2^{L-1}\pi p),\cos(2^{L-1}\pi p)\right],\tag{4}$$

式中：p 为三维位置 X 和姿态 d 的形参表示；L 表示位置编码的维度，在 NeRF 中为三维位置 X 设置 $L=10$ 的位置编码 $\gamma(X)$，为相机姿态 d 设置 $L=4$ 的位置编码 $\gamma(d)$。

为了提高采样效率，NeRF 使用粗糙-细腻（coarse-to-fine）双层采样策略将高维信号先后输入到两个特征提取网络中。粗糙表征网络均匀采样 N_c 个样本点来量化光线形成启发式的外形，输出结果可以看作空间中有效采样值的概率密度分布；细腻优化网络在此基础上进行重采样，使用逆变换采样从概率密度分布中抽取 N_f 个与体积相关的样本点，并与均匀采样的样本点进行合并与重排序，使用 N_c+N_f 个样本点来量化光线生成隐式形状表示，然后进行累积渲染步骤，生成最终的 RGB 输出。

从上述描述可以看出，两个特征提取网络的任务并不相同，但在 NeRF 中二者结构相同，这影响了新视图重建的效果。因此，本文分别对两个特征提取网络的结构进行优化，构建基于改进 MLP 的神经辐射场（IP-NeRF）。

3　IP-NeRF 网络构建

IP-NeRF 由改进后的主干、多特征联合学习模块、门控通道变换 MLP 模块组成，网络结构如图 2 所示。为了减少模型的参数量，相较于原始 NeRF 的 8 个全连接层网络结构，IP-NeRF 使用一个深度为 5、通道尺寸为 256 的简化版本作为主干网络，并将细腻优化网络的第二层全连接替换为 NIL，增强网络对粗采样特征的拟合能力。在粗糙表征网络的嵌入层和采样层之间加入 MFJL 模块，通过解码更丰富的细节信息，启发式外形的纹理特征更加明显。此外，在两个特征提取网络的采样层和推理层之间构建门控通道变换 MLP 模块，实现对高权重特征的筛选。

3.1　多特征联合学习模块

MLP 模型擅长于对全局特征的采样与推理，而容易丢失局部特征信息[22]。为了在 NeRF 使用的线性连接 MLP 模型中引入局部性，本文使用离散余弦变换（DCT）[23]采样局部特征；利用 NeRF 的位置编码输入在训练过程中会发生尺度变化的特点，设计了一种归一化 MLP 瓶颈结构采样全局特征；利用无参注意力（SimAM）[24]对两个分支的采样特征进行融合，使 NeRF 的 MLP 模型在捕获全局特征的同时获得更多的局部依赖，从而提高采样性能。综上，构建了 MFJL 模块，结构如图 3 所示。

其中一个分支输入为三维坐标 X 的位置编码 $\gamma(X)$，首先将每个大小为 $C\times H\times W$ 的特征图分割成 n 等份，然后使用 DCT 过滤并补全特征频谱中未提取到的频率分量，聚焦图像中相对重要的局部信息，最后通过在网络中嵌入更多的频率分量信息获得更高的频域特征聚集度。而由于 NeRF 的尺度变化位置编码不利于全局特征的采样，故另一个分支的输入为 $\gamma(X)$ 经层归一化（LN）后重新加权调整在统一范围内的特征向量，首先利用全局平均池化将每个大小为 $C\times H\times W$ 的特征图在通道维度上压缩成一个标量，以捕获通道级的全局依赖，然后采用“先降维，再升维”的策略构建 MLP 瓶颈结构并将其作为特征解码器，使用线性整流单元（ReLU）作为激活函数。此外，基于无参注意力，在通道维度对局部特征与全局特征进行融合，

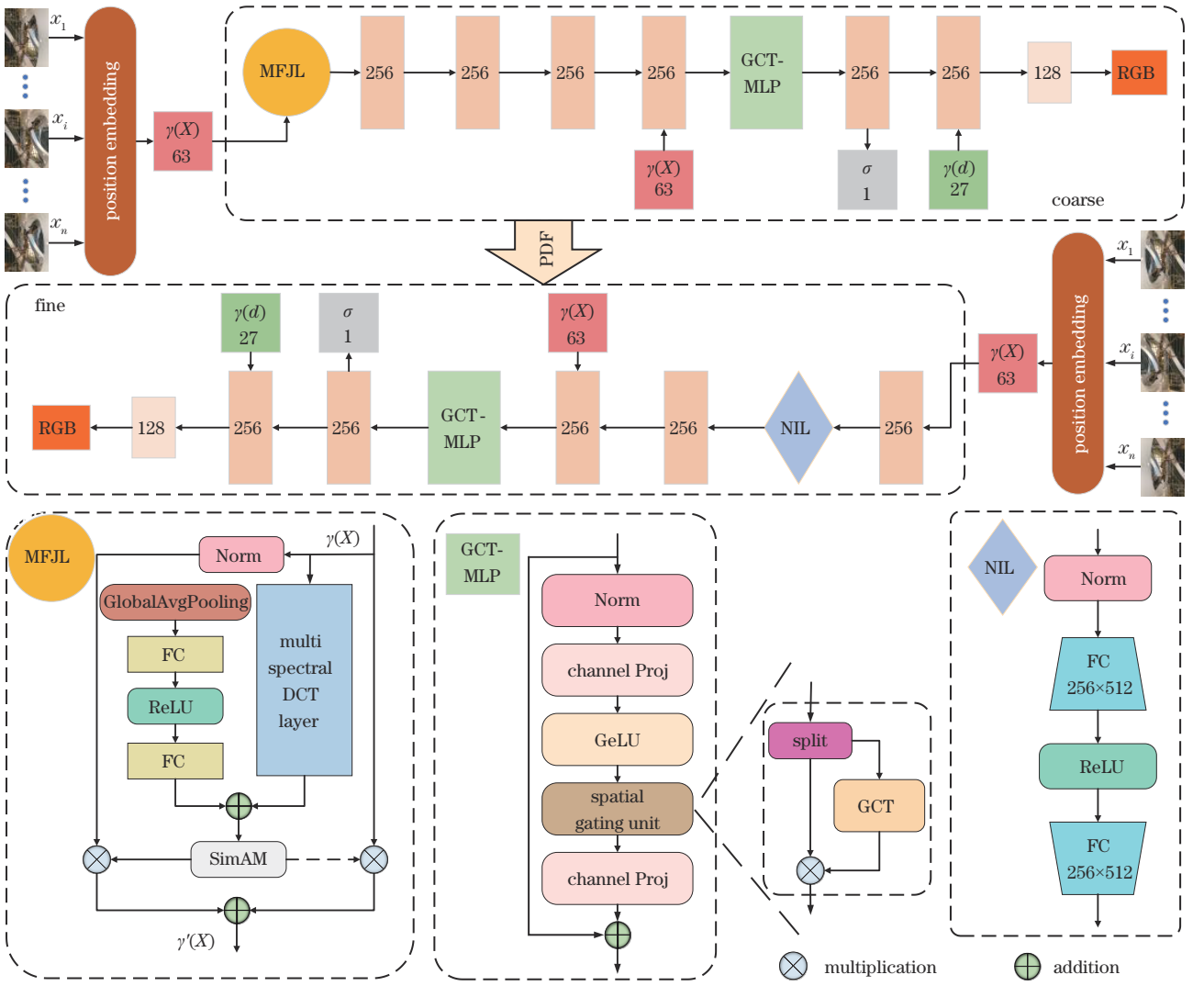

图 2　IP-NeRF 网络的结构

Fig. 2　Structure of IP-NeRF network

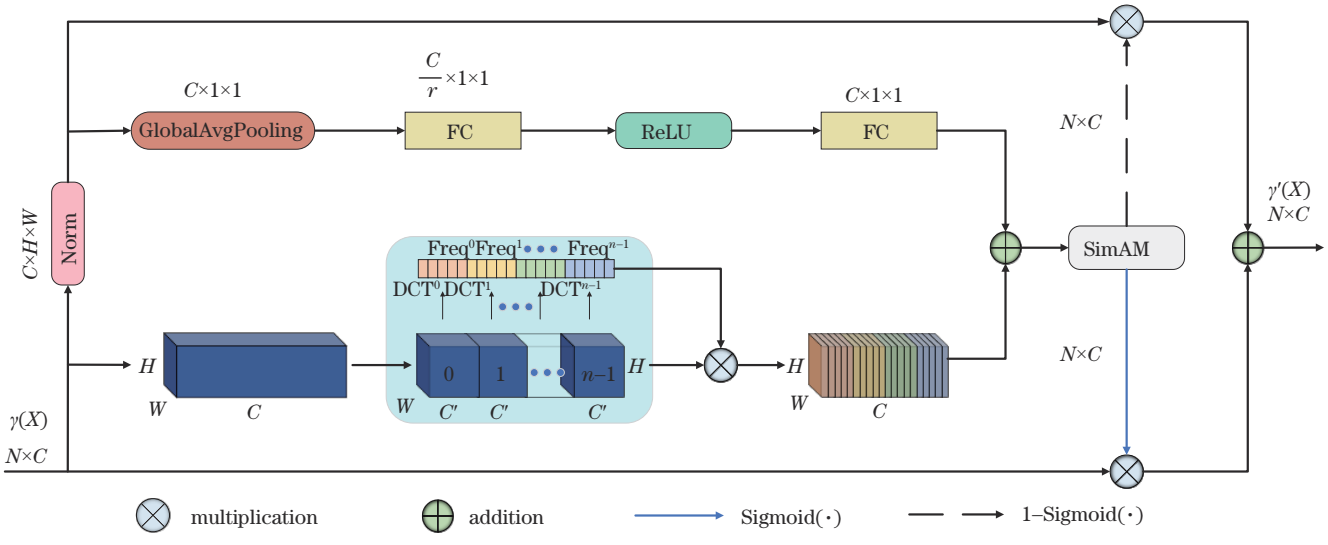

图 3　多特征联合学习模块

Fig. 3　Multi-feature joint learning module

然后通过计算每个神经元的能量函数 e_r，以最小神经元能量之和 e_r^* 为优化目标生成相应的神经元注意力权重，并与对应分支的原始特征图进行加权求和，得到联合学习特征 $\gamma'(X)$：

$$\gamma'(X)=\left\{\left[1-\mathrm{Sigmoid}\left(\frac{1}{\sum_{i=0}^{n-1}e_{\mathrm{r}}^{*}}\right)\right]\left\{\mathrm{DCT}\left[\gamma(X)\right]\oplus\mathrm{nMLP}\left[\gamma(X)\right]\right\}\otimes\gamma(X)\oplus\left[\mathrm{Sigmoid}\left(\frac{1}{\sum_{i=0}^{n-1}e_{\mathrm{r}}^{*}}\right)\right]\left\{\mathrm{DCT}\left[\gamma(X)\right]\oplus\mathrm{nMLP}\left[\gamma(X)\right]\right\}\otimes\mathrm{Norm}\left[\gamma(X)\right]\right\},\tag{5}$$

式中：$\gamma(X)$和$\mathrm{Norm}\left[\gamma(X)\right]$表示原始特征图；Sigmoid(•)表示神经元注意力权重计算；DCT(•)表示局部特征过滤；nMLP(•)表示全局特征采样。

3.2　门控通道变换多层感知机模块

Liu等[25]证明了门控机制与MLP配合的有效性，因此本文在NeRF的采样层与推理层之间加入门控MLP(gMLP)模块以增强对歧义编码特征的空间筛选能力。为使模块更加契合NeRF经傅里叶变换后的高维特征，将gMLP的门控单元由逐点卷积(PW)替换为门控通道变换(GCT)[26]。该方法使NeRF的MLP模型专注于学习权重信息的博弈状态，获得更多与视图高相关性的特征信息进行颜色与密度推理，可以有效减少推理过程中歧义编码特征的干扰，从而提高推理性能，结构如图4所示。

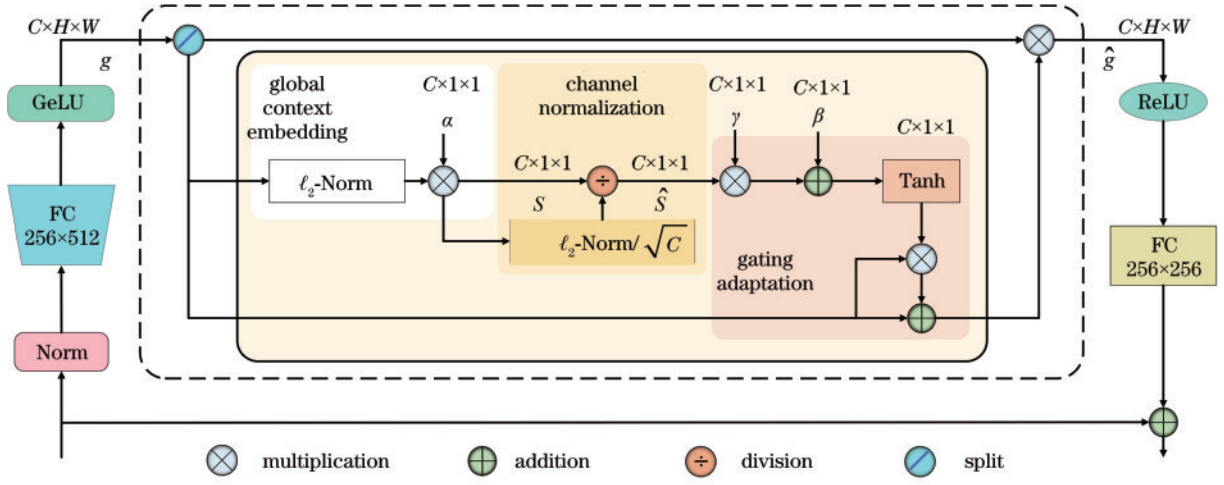

图4　门控通道变换多层感知机模块

Fig. 4　Gated channel tranformation MLP module

门控单元是特征筛选层构成的门控矩阵，通过矩阵点乘来控制特征的通过率，进而控制参数量。GCT采用ℓ_2-Norm归一化方法来建立通道之间的竞争或合作关系，通过门控机制F选择性地传输特征信息，在通道维度实现对特征的重新加权。当一个通道的门控权重被正激活，GCT将促进这个通道的特征和其他通道的特征“竞争”，当一个通道的门控权重被负激活，GCT将促进这个通道的特征和其他通道的特征“合作”，表达式为

$$\hat{g}=F(g|\alpha,\beta,\gamma),\tag{6}$$

式中：α、β、γ是三个可训练的参数，嵌入权重α通过消除局部歧义调整编码输出，门控权重γ和门控偏置β控制通道特征的激活状态。

4　实验结果与分析

4.1　数据集与实验参数设置

使用3个公共数据集进行实验，即Realistic Synthetic 360°数据集、Real Forward-Facing数据集和DTU数据集。Realistic Synthetic 360°数据集是一个真实渲染的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。Real Forward-Facing数据集是手持相机拍摄的真实场景数据集，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，其中1/8用于测试，其余用于训练。DTU数据集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，其中1/8用于测试，其余用于训练。

实验在腾讯云服务器Ubuntu 18.04平台上进行，CPU为Intel® Xeon® Platinum 8255C@2.50 GHz，GPU为NVIDIA® Tesla® T4，采用CUDA11.2加速库进行并行加速，基于Python 3.8的PyTorch深度学习框架实现算法。IP-NeRF采用分层采样策略进行训练，粗糙表征网络的均匀采样点个数为64，细腻优化网络的非均匀采样点个数为128，每个批次的采样光线数量为1024，单次采样数量为32768。使用均方误差(MSE)光度损失，在Adam优化器(β_1=0.9，β_2=0.999，$\varepsilon=10^{-7}$)上以初始值为5×10^{-4}并呈指数衰减

至 5×10^{-5}的学习率进行训练，每个场景训练迭代 20 万次。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）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。

4.2 消融实验

为验证所提方法的有效性，依次设置仅使用改进主干网络的实验组 A、只添加 MFJL 模块的实验组 B、只添加 GCT-MLP 模块的实验组 C、同时引入 MFJL 模块和 GCT-MLP 模块的实验组 D。在 Realistic Synthetic 360° 数据集 Lego 场景和 Real Forward-Facing 数据集的 Trex 场景上验证 4 个实验组的新视图重建性能。表 1 和表 2 展示了消融实验的结果，采用加粗字体表示最优指标，括号中的百分值是相对 NeRF 的提升幅度。

表 1 在 Trex 场景下新视图重建的消融实验

Table 1 Ablation experiment of the new view reconstruction in Trex scene

Ablation experiment	PSNR/dB	SSIM	LPIPS
A	26.90 (+0.37%)	0.881 (+0.11%)	0.057
B	28.23 (+5.33%)	0.925 (+5.11%)	0.051 (−10.53%)
C	28.56 (+6.56%)	0.930 (+5.68%)	0.046 (−19.30%)
D	**28.74** (**+7.24%**)	**0.934** (**+6.13%**)	**0.042** (**−26.31%**)

表 2 在 Lego 场景下新视图重建的消融实验

Table 2 Ablation experiment of the new view reconstruction in Lego scene

Ablation experiment	PSNR/dB	SSIM	LPIPS
A	32.62 (+0.24%)	0.961	0.024
B	33.94 (+4.33%)	0.968 (+0.73%)	0.017 (−29.17%)
C	34.37 (+5.62%)	0.971 (+1.04%)	0.016 (−33.33%)
D	**34.77** (**+6.85%**)	**0.974** (**+1.35%**)	**0.015** (**−37.50%**)

由表 1 和表 2 数据可知：改进后的主干网络可以达到和原始 NeRF 基本一致的性能表现，说明线性全连接层的简单堆叠并不能明显地提升网络的表征能力；相比实验组 A，只添加 MFJL 模块的网络在 Trex 场景下对重建视图的 PSNR、SSIM、LPIPS 平均值提升了 1.33 dB（4.94%）、0.044（4.99%）、−0.006（−10.53%），在 Lego 场景下对重建视图的 PSNR、SSIM、LPIPS 平均值提升了 1.32 dB（4.05%）、0.007（0.73%）、−0.007（−29.17%），说明了局部细节信息对新视图重建的必要性；相比实验组 A，只添加 GCT-MLP 模块的网络在 Trex 场景下对重建视图的 PSNR、SSIM、LPIPS 平均值提升了 1.66 dB（6.17%）、0.049（5.56%）、−0.011（−19.30%），在 Lego 场景下对重建视图的 PSNR、SSIM、LPIPS 平均值提升了 1.75 dB（5.36%）、0.010（1.04%）、−0.008（−33.3%），重建性能并未减弱，说明筛选高阶特征交互关系对重建结果同样重要；相比实验组 A，同时引入 MFJL 模块和 GCT-MLP 模块的网络在 Trex 场景下对重建视图的 PSNR、SSIM、LPIPS 平均值提升了 1.84 dB（6.84%）、0.053（6.01%）、−0.015（−26.31%），在 Lego 场景下对重建视图的 PSNR、SSIM、LPIPS 平均值提升了 2.15 dB（6.59%）、0.013（1.35%）、−0.009（−37.5%），说明所提多特征联合学习方法的有效性。

图 5 展示了两个场景消融实验的定性结果，其中第 1 列为未参与训练的真实图像，第 2～4 列为 4 个消融实验组的新视图重建结果。根据局部放大视图可以定性地观察每个模块对视图重建的贡献：以 Trex 场景为例，实验组 B 处理的前景位置出现了围栏顶部的轮廓，说明 MFJL 模块在防止重建退化方面发挥了重要作用，也表明了对采样网络浅层特征进行处理的有效性；实验组 C 中重建视图的细粒度特征更加清晰，显示了 GCT-MLP 模块捕捉高阶特征进行特征信息筛选的必要性；实验组 D 获得了基本符合真实值的重建视图，表明了所添加的模块可以相互配合，在提高重建性能方面发挥了作用。

4.3 对比实验

为验证所提网络的新视图重建性能，分别在 Real Forward-Facing、Realistic Synthetic 360°、DTU 数据集共 20 个场景上进行测试，并与具有同样实验参数设置的 NeRF、NeRF-ID 进行对比。不同网络的新视图重建定量结果如表 3～5 所示，采用加粗字体表示最优指标，与指标相关的重建视图如图 6 所示。

由表 3～5 数据可知：在 Realistic Synthetic 360° 数据集上，所提网络重建视图的 PSNR、SSIM、LPIPS 平均值相较 NeRF-ID 有 0.41 dB（1.27%）、0.003（0.3%）、−0.003（10.3%）的提升，相较 NeRF 有 1.74 dB（5.6%）、0.013（1.4%）、−0.015（36.6%）的提升，在大部分场景下有较为优异的表现；在 Real Forward-Facing 数据集上，所提网络重建视图的 PSNR、SSIM、LPIPS 平均值相较 NeRF-ID 有 1.32 dB

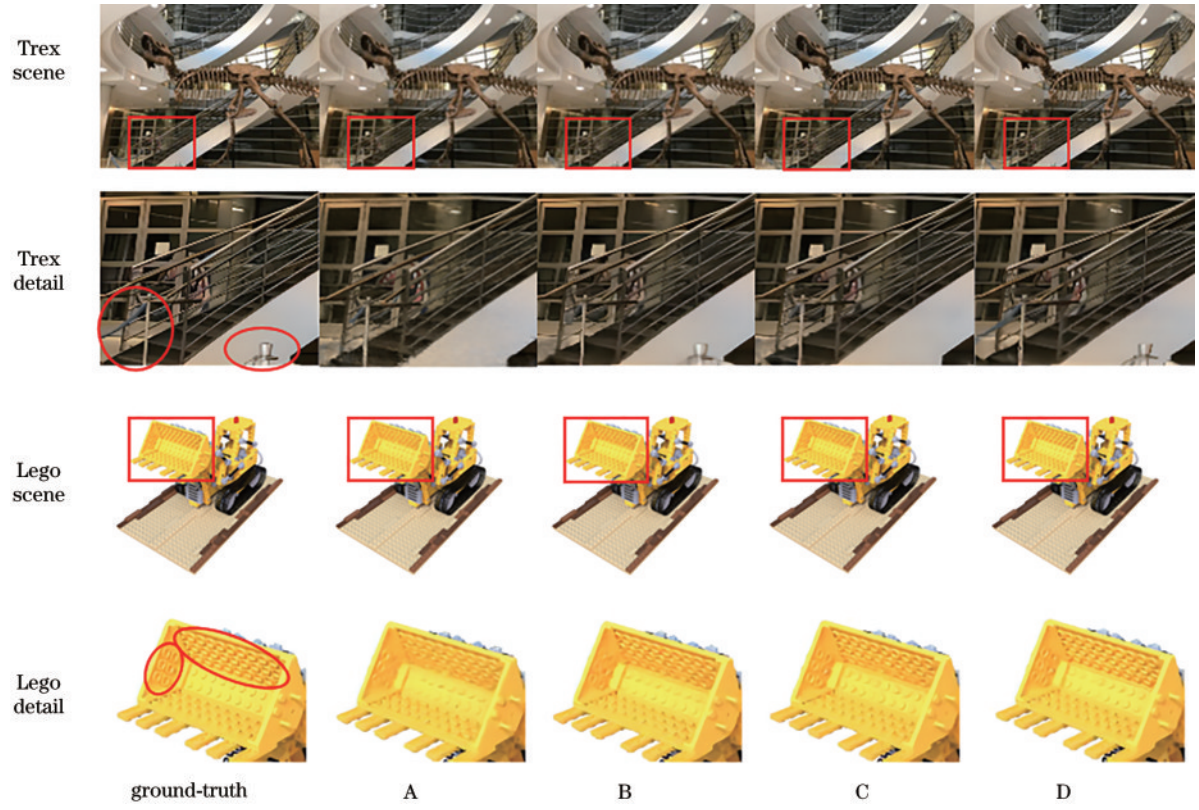

图 5　各模块作用下 Trex 和 Lego 场景的新视图重建可视化结果

Fig. 5　Visualized new view reconstruction results under the modules in Trex and Lego scenes

表 3　不同方法在 Realistic Synthetic 360°数据集上的参数对比

Table 3　Parameter comparison of different methods on Realistic Synthetic 360° dataset

Scene	NeRF			NeRF-ID			IP-NeRF		
	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**
Drums	25.01	0.925	0.058	25.15	0.926	0.057	**25.80**	**0.931**	**0.051**
Ficus	30.13	0.964	0.022	**32.24**	**0.976**	**0.015**	31.86	0.973	0.016
Hotdog	36.18	0.974	0.016	37.26	0.981	0.013	**38.48**	**0.986**	**0.010**
Lego	32.54	0.961	0.024	34.73	**0.974**	**0.015**	**34.77**	**0.974**	**0.015**
Materials	29.62	0.949	0.029	30.37	0.956	0.024	**31.90**	**0.977**	**0.011**
Mic	32.91	0.980	0.023	**34.71**	**0.988**	**0.009**	34.21	0.982	0.018
Ship	28.65	0.856	0.119	29.75	**0.876**	**0.081**	**29.79**	**0.876**	**0.081**
Mean	31.01	0.947	0.041	32.34	0.957	0.029	**32.75**	**0.960**	**0.026**

表 4　不同方法在 Real Forward-Facing 数据集上的参数对比

Table 4　Parameter comparison of different methods on Real Forward-Facing dataset

Scene	NeRF			NeRF-ID			IP-NeRF		
	PSNR /dB	SSIM	LPIPS	PSNR /dB	SSIM	LPIPS	PSNR /dB	SSIM	LPIPS
Fern	25.20	0.792	0.092	25.01	0.800	0.089	**27.08**	**0.868**	**0.079**
Flower	27.40	0.827	0.061	27.85	0.842	0.058	**28.82**	**0.901**	**0.053**
Fortress	31.16	0.881	0.030	31.51	0.888	0.028	**32.94**	**0.933**	**0.024**
Horns	27.45	0.828	0.068	27.88	0.843	0.065	**29.30**	**0.911**	**0.057**
Leaves	20.92	0.690	0.111	21.09	0.708	0.108	**22.53**	**0.825**	**0.100**
Orchids	20.36	0.641	0.121	20.38	0.643	0.120	**21.44**	**0.764**	**0.100**
Room	32.70	0.948	0.041	32.93	0.954	0.039	**33.86**	**0.961**	**0.035**
Trex	26.80	0.880	0.057	27.45	0.897	0.051	**28.74**	**0.934**	**0.042**
Mean	26.50	0.811	0.073	26.76	0.822	0.070	**28.08**	**0.887**	**0.061**

表 5　不同方法在 DTU 数据集部分场景上的参数对比

Table 5　Parameter comparison of different methods on DTU dataset

Scene	NeRF			NeRF-ID			IP-NeRF		
	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**
Scan22	21.55	0.708	0.238	21.98	0.715	0.226	**22.68**	**0.758**	**0.196**
Scan55	26.54	0.794	0.229	26.76	0.800	0.219	**27.23**	**0.812**	**0.206**
Scan109	28.33	0.860	0.236	28.63	0.870	0.226	**29.46**	**0.881**	**0.185**
Mean	24.98	0.779	0.246	25.29	0.787	0.234	**25.96**	**0.807**	**0.208**

(4.9%)、0.065(7.9%)、−0.009(12.9%)的提升，相较 NeRF 有 1.58 dB(5.9%)、0.076(9.3%)、−0.012(16.4%)的提升，在所有场景下的性能指标均优于其他对比实验组；在 DTU 数据集上，所提网络重建视图的 PSNR、SSIM、LPIPS 平均值相较 NeRF-ID 有 0.67 dB(2.6%)、0.020(2.5%)、−0.026(11.1%)的提升，相较 NeRF 有 0.98 dB(3.9%)、0.028(3.6%)、−0.038(15.4%)的提升，在所有随机场景下的性能指标均优于其他对比实验组，这表明所提网络较其他对比网络能有效提升新视图重建的性能。同时，从表 3 可以观察到：所提网络在 Ficus 场景和 Mic 场景下的性能指标低于 NeRF-ID，但高于 NeRF；从数据可以看出所提网络在合成场景的重建性能相较于真实场景波动较大，这是因为局部特征和全局特征联合学习的过程中，无特征背景会对物体表面附近的颜色和密度产生不利影响，这些区域特征点的值会相互干扰，造成特征融合与特征筛选过程中的像素拟合失误，对间隙较多的实体影响较大。

图 6 展示了 Real Forward-Facing 数据集的 Room 场景、Realistic Synthetic 360°数据集的 Materials 场景

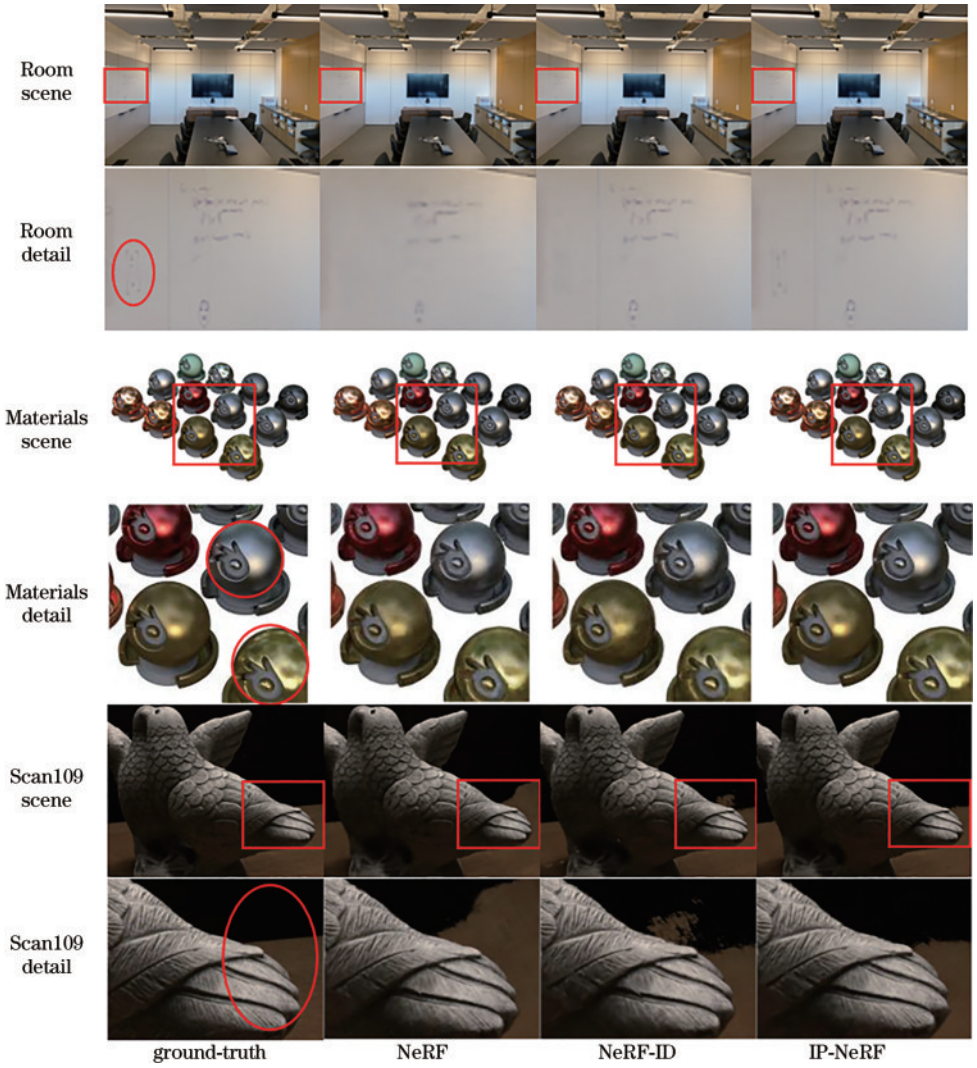

图 6　不同方法在 3 个数据集部分场景的新视图重建可视化结果

Fig. 6　Visualized new view reconstruction results of different methods in the selected scenes of the three datasets

和DTU数据集的Scan109场景下不同方法的新视图重建定性结果。图6第1列为未参与训练的真实图像，第2～4列分别为NeRF、NeRF-ID和IP-NeRF的重建视图。根据第1列圆圈标记的图像真实值，可以观察到不同方法的新视图重建结果的局部变化情况。通过定性结果可以看出，所提方法在避免渲染模糊和混叠的情况下具有优异的视图重建能力，重建细节更加丰富，如板书上的符号、球体的凹凸纹理和鸽子的纹路。

算法运行时间也是衡量算法性能的重要指标，表6列出了各算法训练时间和渲染时间的平均计算耗时。由表6可知，IP-NeRF的计算耗时较高，这是由于NeRF的细腻优化网络的采样点数量为其粗糙表征网络的3倍，门控通道变换MLP模块进行特征筛选的计算量大幅增加，故增加了计算耗时，但相较NeRF，以少量的时间代价换取明显的精度提升，这是可以接受的。s/it中，it表示测试视图，s表示渲染时间，s/it表示每张测试视图所需的渲染时间，对应数值为平均值。

表6 不同方法的计算代价对比

Table 6 Calculation cost comparison of different methods

Dataset	NeRF			NeRF-ID			IP-NeRF		
	PSNR /dB	Train-time /h	Render-time /(s/it)	PSNR /dB	Train-time /h	Render-time /(s/it)	PSNR /dB	Train-time /h	Render-time/(s/it)
Realistic Synthetic 360°	31.01	18.4	21.18	32.34	**14.9**	**17.10**	**32.75**	19.5	22.24
Real Forward-Facing	26.50	16.5	20.10	26.76	**13.3**	**16.17**	**28.08**	17.5	21.20
DTU	24.98	19.4	36.60	25.29	**15.6**	**30.48**	**25.96**	20.5	38.63

通过剔除细腻优化网络的门控通道变换MLP模块，得到IP-NeRF的简化网络，其余参数设置不变，将这种方案命名为“SIP-NeRF”。表7列出了其与NeRF-ID计算代价的对比，可以看出SIP-NeRF相比NeRF-ID在相同时间复杂度的情况下性能较优。针对细腻优化网络计算量大的问题，可以采用一些优化措施，如光线提前终止策略、改进更加轻量高效模块方法。

表7 简化网络的计算代价对比

Table 7 Calculation cost comparison of simplified network

Dataset	NeRF-ID			SIP-NeRF		
	PSNR /dB	Train-time /h	Render-time /(s/it)	PSNR /dB	Train-time /h	Render-time /(s/it)
Realistic Synthetic 360°	32.34	**14.9**	17.10	**32.40**	15.0	**17.08**
Real Forward-Facing	26.76	**13.3**	16.17	**27.68**	13.3	**16.10**
DTU	25.29	**15.6**	30.48	**25.64**	15.6	**30.39**

表8列出了不同方法的PSNR和训练时间，一定程度上可以反映各方法的综合性能。结合表6～8数据可知，所提方法实现了除NeXT外最优的重建质量，而NeXT由于高昂的训练成本也限制了本身的应用，同时这也是所提方法需要改进的方向。从训练时间看，NSVF使用稀疏体素八叉树结构实现了对重建视图的快速渲染，但不适用于真实场景。所提方法的侧重点主要是对像素区域内多尺度特征信息进行联合学习与筛选，可以兼顾重建质量与计算代价。

表8 不同方法的综合性能分析

Table 8 Comprehensive performance analysis of different methods

Dataset	Parameter	NeRF[13]	NSVF[15]	GRF[17]	NeuSample[21]	NeXT[20]	IP-NeRF
Realistic Synthetic 360°	PSNR /dB	31.01	31.75	32.06	31.15	**34.40**	32.75
	Train-time/h	18.4	**1.5**	23.0	14.0	52.7	19.5
Real Forward-Facing	PSNR /dB	26.50	/	26.64	26.83	/	**28.08**
	Train-time /h	16.5	/	20.6	**12.5**	/	17.5

5 结 论

隐式神经表达的有效采样是神经辐射场三维重建的基础，针对NeRF网络表征能力弱的问题，提出一种基于MLP的多特征联合学习方法。所提IP-NeRF网络使用改进后的主干为基本结构，通过在NeRF嵌入层和采样层之间设计归一化MLP瓶颈结构和DCT并行的特征解码器进行全局特征和局部特征提取，结合无参注意力(SimAM)构建了MFJL模块，实现特征融合。此外，通过在NeRF采样层和推理层之间建立门控通道变换MLP，筛选高权重特征，缓解了编码特征歧义值对颜色和密度推理的影响。最后，使用3个公

开数据集对IP-NeRF进行实验训练与测试。实验结果表明，相比NeRF，所提方法的新视图重建各项评价指标均有提高，在Real Forward-Facing数据集上的PSNR、SSIM、LPIPS值提升幅度分别为5.9%、9.3%、16.4%，在Realistic Synthetic 360°数据集上的PSNR、SSIM、LPIPS值提升幅度分别为5.6%、1.4%、36.6%，在DTU数据集部分场景上的PSNR、SSIM、LPIPS值提升幅度分别为3.9%、3.6%、15.4%，且新视图重建定性结果也有较优表现。因此所提方法能为NeRF的MLP模型改进提供一定参考价值，后续可以在编码器网络和推理网络上做进一步的优化。

参 考 文 献

[1] 殷永凯, 于锴, 于春展, 等. 几何光场三维成像综述[J]. 中国激光, 2021, 48(12): 1209001.
Yin Y K, Yu K, Yu C Z, et al. 3D imaging using geometric light field: a review[J]. Chinese Journal of Lasers, 2021, 48(12): 1209001.

[2] 张博霄, 于佳慧, 焦小雪, 等. 线结构光双目融合补缺重建技术[J]. 激光与光电子学进展, 2023, 60(16): 1611001.
Zhang B X, Yu J H, Jiao X X, et al. Line structured light binocular fusion filling and reconstruction technology[J]. Laser & Optoelectronics Progress, 2023, 60(16): 1611001.

[3] 石世锋, 叶南, 张丽艳. 具有远近视距的两目视觉系统标定技术研究[J]. 光学学报, 2021, 41(24): 2415001.
Shi S F, Ye N, Zhang L Y. Calibration of two-camera vision system with far and near sight distance[J]. Acta Optica Sinica, 2021, 41(24): 2415001.

[4] Wang F, Wang C L, Deng C J, et al. Single-pixel imaging using physics enhanced deep learning[J]. Photonics Research, 2022, 10(1): 104-110.

[5] 刘昊鑫, 赵源萌, 张存林, 等. 基于改进U-net的牙齿锥形束CT图像重建研究[J]. 中国激光, 2022, 49(24): 2407207.
Liu H X, Zhao Y M, Zhang C L, et al. Study on tooth cone beam CT image reconstruction based on improved U-net network[J]. Chinese Journal of Lasers, 2022, 49(24): 2407207.

[6] 王明军, 李乐, 易芳, 等. 模拟真实水体环境下目标激光点云数据的三维重建与分析[J]. 中国激光, 2022, 49(3): 0309001.
Wang M J, Li L, Yi F, et al. Three-dimensional reconstruction and analysis of target laser point cloud data under simulated real water environment[J]. Chinese Journal of Lasers, 2022, 49(3): 0309001.

[7] 李明阳, 陈伟, 王珊珊, 等. 视觉深度学习的三维重建方法综述[J]. 计算机科学与探索, 2023, 17(2): 279-302.
Li M Y, Chen W, Wang S S, et al. Survey on 3D reconstruction methods based on visual deep learning[J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(2): 279-302.

[8] Sitzmann V, Thies J, Heide F, et al. DeepVoxels: learning persistent 3D feature embeddings[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 2432-2441.

[9] Bui G, Le T, Morago B, et al. Point-based rendering enhancement via deep learning[J]. The Visual Computer, 2018, 34(6): 829-841.

[10] Riegler G, Koltun V. Free view synthesis[M]//Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12364: 623-640.

[11] Srinivasan P P, Tucker R, Barron J T, et al. Pushing the boundaries of view extrapolation with multiplane images[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 175-184.

[12] Schirmer L, Schardong G, da Silva V, et al. Neural networks for implicit representations of 3D scenes[C]//2021 34th SIBGRAPI Conference on Graphics, Patterns and Images, October 18-22, 2021, Gramado, Rio Grande do Sul, Brazil. New York: IEEE Press, 2021: 17-24.

[13] Mildenhall B, Srinivasan P P, Tancik M, et al. NeRF: representing scenes as neural radiance fields for view synthesis[J]. Communications of the ACM, 2022, 65(1): 99-106.

[14] Zhu Z H, Peng S Y, Larsson V, et al. NICE-SLAM: neural implicit scalable encoding for SLAM[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 12776-12786.

[15] Liu L, Gu J, Zaw L K, et al. Neural sparse voxel fields [J]. Advances in Neural Information Processing Systems, 2020, 33: 15651-15663.

[16] Zhang K, Riegler G, Snavely N, et al. NeRF++: analyzing and improving neural radiance fields[EB/OL]. (2020-10-15)[2022-11-12]. https://arxiv.org/abs/2010.07492.

[17] Trevithick A, Yang B. GRF: learning a general radiance field for 3D representation and rendering[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2022: 15162-15172.

[18] Arandjelović R, Zisserman A. NeRF in detail: learning to sample for view synthesis[EB/OL]. (2021-06-09)[2022-11-02]. https://arxiv.org/abs/2106.05264.

[19] Yang G W, Zhou W Y, Peng H Y, et al. Recursive-NeRF: an efficient and dynamically growing NeRF[J]. IEEE Transactions on Visualization and Computer Graphics, 2022, 14(8): 36194712.

[20] Wang Y X, Li Y J, Liu P D, et al. NeXT: towards high quality neural radiance fields via multi-skip transformer [M]//Avidan S, Brostow G, Cissé M, et al. Computer vision-ECCV 2022. Lecture notes in computer science. Cham: Springer, 2022, 13692: 69-86.

[21] Fang J M, Xie L X, Wang X G, et al. NeuSample: neural sample field for efficient view synthesis[EB/OL]. (2021-11-30)[2022-11-05]. https://arxiv.org/abs/2111.15552.

[22] Ding X H, Xia C L, Zhang X Y, et al. RepMLP: re-

parameterizing convolutions into fully-connected layers for image recognition[EB/OL]. (2021-04-05) [2022-11-09]. https://arxiv.org/abs/2105.01883.

[23] Qin Z Q, Zhang P Y, Wu F, et al. FcaNet: frequency channel attention networks[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2022: 763-772.

[24] Yang L, Zhang R, Li L, et al. Simam: a simple, parameter-free attention module for convolutional neural networks[C]//Proceedings of the 38th International Conference on Machine Learning, July 18-24, 2021, Virtual Event. Cambridge: JMLR, 2021: 11863-11874.

[25] Liu H, Dai Z, So D R, et al. Pay attention to MLPs[C]// Advances in Neural Information Processing Systems 34, December 6-14, 2021, Virtual Event. Cambridge: JMLR, 2021: 9204-9215.

[26] Yang Z X, Zhu L C, Wu Y, et al. Gated channel transformation for visual recognition[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11791-11800.