DOI:10.19651/j.cnki.emt.2416705

基于神经辐射场的稀疏视角三维重建方法*

张 超[1] 袁 亮[1,2] 肖文东[1] 冉 腾[1] 吕 凯[1]
(1.新疆大学智能制造现代产业学院 乌鲁木齐 830017; 2.上海交通大学文化创意产业学院 上海 200240)

摘 要: 针对神经辐射场在稀疏视角输入条件下渲染结果过于平滑,细节缺失严重等问题,提出一个基于信息关注抑制模块和双阶段损失的网络模型。首先,为解决细节缺失问题,提出一个信息关注抑制模块,该模块在全连接层各层之间采用特征向量归一化模块过滤权重异常值,并以残差网络级联全局信息和局部信息,最后利用通道注意力将将融合后的信息根据重要度进行区分,有效提高了采样点特征向量的准确性。然后,为了解决渲染结果过于平滑导致感知精度低的问题,设计了一种双阶段损失函数,将训练过程划分为两个阶段,粗阶段仅以 RGB 损失和深度损失指导训练,细阶段在此基础上还引入感知损失和全变分损失,通过渐进优化的方式,充分利用图片的高级特征,提升图像感知能力。本文算法与其他经典方法进行对比,在 LLFF 数据集上,定量结果表明,整体性能取得最优值,比次优算法性能提升 1.9%,在 DTU 数据集上,定性结果显示,Scan37、Scan55 和 Scan63 等场景重建的完整性和细节水平具有明显优势。

关键词: 三维重建;稀疏视角;神经辐射场;注意力机制;双阶段损失
中图分类号: TP391.4;TN-9 **文献标识码**: A **国家标准学科分类代码**: 520.6030

Sparse perspective 3D reconstruction method based on neural radiation field

Zhang Chao[1] Yuan Liang[1,2] Xiao Wendong[1] Ran Teng[1] Lyu Kai[1]
(1. College of Intelligent Manufacturing Modern Industry, Xinjiang University, Urumqi 830017, China;
2. School of Cultural and Creative Industries, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: In order to address the issue of the neural radiation field rendering results being overly smooth when sparse viewpoint input conditions are present, resulting in a lack of detail, a network model based on an information attention suppression module and a two-stage loss function has been proposed. The first step is to propose an information attention suppression module, which uses a feature vector normalization module to filter outliers in the weights between layers of MLP. It also uses a residual network to cascade global and local information and employs channel attention to differentiate fused information based on its degree of importance. This process improves the accuracy of the sampling points′ feature vectors. To address the issue of low perceptual accuracy resulting from overly smooth rendering, a two-stage loss function is proposed. This function partitions the training phase into two stages. In the initial coarse stage, training is guided by RGB and depth loss. Subsequently, in the fine stage, perceptual loss and TV loss are incorporated. This approach enables the utilisation of high-level image features, thereby enhancing the image perception ability via gradual optimization. This paper′s algorithm is compared with other classical methods, and on the LLFF dataset, the quantitative results demonstrate that the overall performance reaches its optimal value, which is 1.9% superior to the performance of the sub-optimal algorithm. Furthermore, on the DTU dataset, the qualitative results indicate that the reconstruction′s completeness and detail level, as observed in Scan37, Scan55, and Scan63, are notably enhanced.

Keywords: 3D reconstruction;sparse perspective;neural radiation field;attention mechanism;two-stage loss

0 引 言

从物体的 RGB 图像恢复 3D 形状和纹理是计算机视觉领域的重要课题。多种领域深入研究有赖于获得物体准确三维信息,包括虚拟现实[1]、自动驾驶[2]、定位导航[3]等。

传统的建模方法基于显式表示,使用几何信息对现实

收稿日期:2024-08-21
* 基金项目:国家自然科学基金(52275003)、新疆维吾尔自治区重大科技专项(2023A03001)资助

场景进行表达。常用的几何表示方法有纹理网格[4-6]、体素[7]以及点云[8-11]。虽然传统建模方法能够提供精确的三维结构信息，但由于其计算复杂度高且过分依赖几何模型，从而难以表示复杂、多样化的场景。

隐式神经表示极大克服了显式表示的局限性。基于隐式表示的三维重建算法，通过学习像素值和三维几何之间的非线性关系，能更好地处理透视、遮挡等复杂的几何变换和场景。其中神经辐射场[12](neural radiation field，NeRF)在训练时能够同时学习场景的几何结构和外观信息，在渲染时可以灵活地调整视角和光照条件。依靠这些优势，NeRF 将隐式建模提升到了一个新的水平，使场景的新视图合成产生了令人惊叹的结果。

NeRF 虽然能合成具有复杂几何形状和外观的图像，但只有几张图片用于训练时，其渲染结果会丢失细节。一些方法通过人为设计正则化项对每个场景进行优化，用来消除伪影。FreeNeRF[13]通过限制训练过程中的位置编码频率，避免训练初期发生过拟合，同时对靠近相机的漂浮点利用二进制掩码进行评估惩罚，以消除伪影，但较长频率会影响图像感知精度，同时遮挡正则化会使部分场景的近相机目标拟合过度，从而出现不完整表示。

另外一些算法方法向 NeRF 添加先验信息来扩展 NeRF。PixelNeRF[14]引入图像特征，通过跨多个场景训练学习先验信息，在少量视角下渲染效果良好。但其受数据集可用性限制，无法拓展到多种数据集类型。FlipNeRF[15]使用从输入射线方向和估计的法向量中导出的翻转反射射线作为附加训练射线弥补几何缺失，从而准确估计表面法线并学习三维几何形状，但翻转光线导致计算量上升，对设备要求提高。NerfingMVS[16]用 COLMAP 得到的深度训练专属于当前场景的单目深度网络。后续以网络预测的深度图来指导 NeRF 的学习，并利用滤波器进一步提升渲染质量。DS-NeRF[17]通过 COLMAP 生成的 3D 点云与 2D 图像中点的对应关系来引入深度信息。其通过深度信息联合 RGB 信息来指导光线采样过程，从而利用少量训练图片达到良好渲染效果，尽管如此，其渲染出的图片较为平滑，精度提升受到限制。

针对目前稀疏视角重建存在的过于平滑和细节缺失问题，本文结合深度先验以及正则化方法的优势，提出一种端到端的基于信息关注抑制模块和双阶段损失的 NeRF 网络模型。其中，为提升用于重建的采样点特征向量的准确性，本文提出一个信息关注抑制模块。该模块通过适当的跨通道交互来提取感兴趣区域以获得充分的局部信息。模块中的特征向量归一化组件能过滤掉权重异常点，经过处理的信息再引入通道注意力进行重要度区分，用于重建的点的精度提升明显。为增强细节捕捉能力，本文设计了一种双阶段损失函数。粗阶段在像素空间中进行整体轮廓拟合，细阶段在特征空间进行细节捕捉。通过从像素空间到特征空间进行由粗到细的拟合优化，重建结果的感知精度大幅提升。

1　网络模型

稀疏视角重建的主要目标是通过少量图片生成高质量的三维模型。本文模型遵循神经辐射场的主要结构，首先，根据相机内外参数，得到 RGB 图像中每个像素的射线，并利用 COLMAP 生成的稀疏点云与 RGB 图像中点的对应关系引导每条射线的随机采样。然后，针对位置编码后的采样点，一个信息关注抑制模块被提出，该模块通过向量归一化过滤异常值，筛除错误采样点，然后将局部信息和全局信息进行融合，经过滤以及融合后的信息再使用注意力机制对重要度进行区分，保留重要特征向量。最后，设计了一种双阶段损失函数，用于监督训练过程。其本质上是由粗到细的渐进优化，初始阶段像素空间生成较准确的轮廓信息，第二阶段通过在特征空间里微调关注细节信息。本文网络模型如图 1 所示。本章将详细介绍所提方法。包括信息关注抑制模块的细节和双阶段损失函数的具体实现。

1.1　深度引导采样和位置编码

经典 NeRF 仅以 RGB 信息监督训练过程，在稀疏视角下输入信息减少，会造成细节缺失。为补充信息，本文遵循 DS-NeRF 基线方法，利用 RGB 图像和深度信息共同监督训练过程。其中深度利用 SFM 生成的相机位姿以及稀疏点云与图像的对应关系来引入。在深度信息作用下，光线直接可以跳过空的区域而聚焦在物体表面附近，可以有效提高参与体积渲染过程的采样点比例。

具体地，给定图像 I_j 和相机参数 P_j，通过图像和 3D 点的对应关系，估计可见关键点 $x_i \in X_j$ 的深度，将重投影的 z 值作为关键点的深度 d_{ij}。将光线到达的表面建模为随机变量 D_{ij}，D_{ij} 在深度 d_{ij} 周围服从正态分布，方差为 σ_i。在不受噪声影响的理想状态，光线终止分布服从 δ 分布。而最接近表面深度 d_{ij} 的点的理想射线分布为 $\delta(t-d_{ij})$，其中 t 表示射线距离。通过将 x_i 图像坐标的光线终止分布 $h_{ij}(t)$ 与深度分布 $\delta(t-d_{ij})$ 之间的距离最小化，使采样区域更加贴近物体表面。经深度引导的采样区域，会避免大量空白范围，在采集有限数量的点时，有效点的数量占比会大幅提升。

对于得到的随机采样点，利用全连接网络进行隐式表达。在低维空间中，相邻采样点在 MLP 中表示的过于平滑，为避免该问题，采用正余弦波位置编码方法，将位置信息和观测方向映射到更高维度的空间。如式(1)所示。

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

式中：L 表示维度，q 表示函数输入(位置/方向)。对 3D 位置(x, y, z)，L 设为 10；对观察方向(θ, ϕ)，L 设为 4。

1.2　信息关注抑制模块

经位置编码的采样点信息如果不经过筛选和区分，直接输入到全连接层中会丢失细节信息，并且异常采样点参与训练过程也会降低渲染精度。

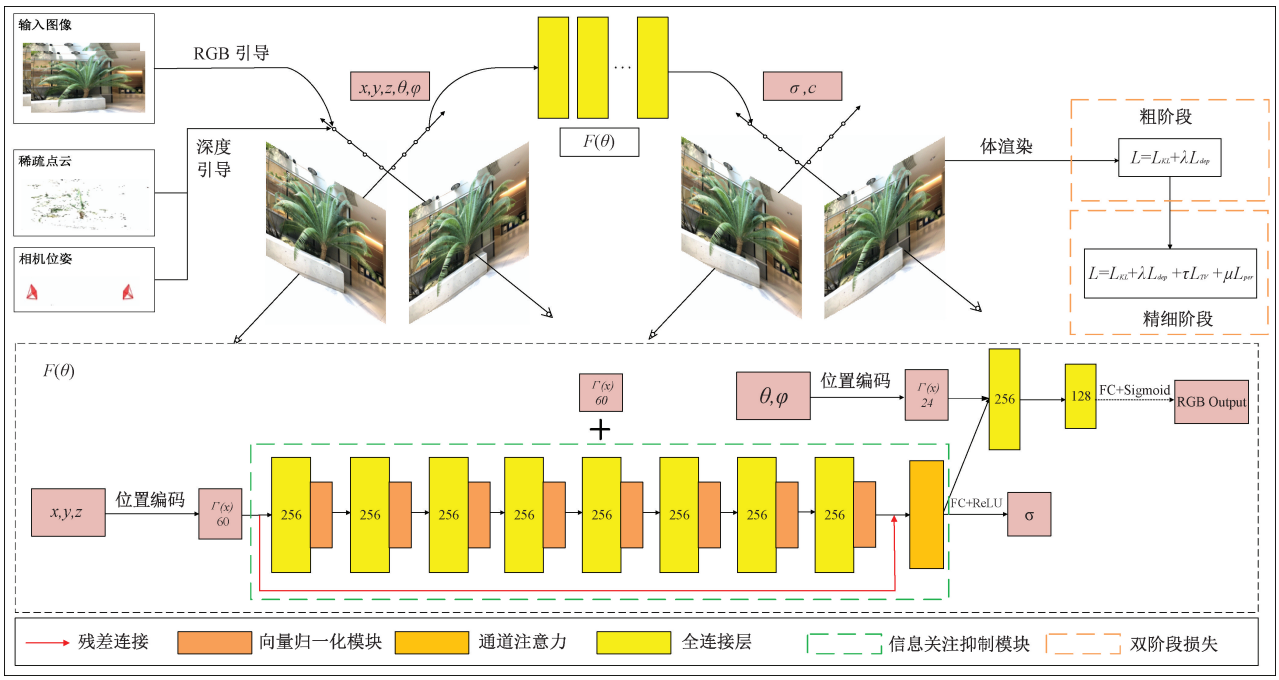

图 1　网络整体结构图

Fig. 1　Overall network structure

区别于目前大多数将采样点的特征向量直接输入全连接层的方法,本文在全连接层的输入输出上对特征向量进行微调和筛选。具体地,在连接层的每层后引入微调模块[18],该模块可以将特征向量的每个元素归一化到单位长度附近,防止由于噪声影响导致的采样点权重随着训练的进行逐步失控。其中,初始特征向量 $\boldsymbol{z}$ 经过微调模块后会转换成线性无关的中间特征向量,从而过滤掉权重值异常的采样点。此外,该模块并未引入训练的参数,其本质上是一个转换模块,不会增加训练负担。

微调模块如式(2)所示。

$$\boldsymbol{b}_{x,y}=\boldsymbol{a}_{x,y}/\sqrt{\frac{1}{N}\sum_{j=0}^{N-1}(\boldsymbol{a}_{x,y}^{j})^{2}+\epsilon} \tag{2}$$

式中:$\boldsymbol{a}$,$\boldsymbol{b}$ 表征两个特征向量,j 表示元素索引,ϵ 是为了防止除数为 0 设置的随机值(本文设置为 10^{-8})。

用于表征场景的全连接层总共 8 层,其中大部分是中间非线性映射,层数相对较多。对于特征向量而言,随着网络层数的加深,全局性信息会逐渐增多,局部信息则被严重削弱。然而,局部信息携带大量的细节信息,对图像的渲染质量存在很大影响。因此,为了充分利用局部信息,模型通过跨通道融合[19],将输入第一层全连接层的信息与最后一层全连接层的输出聚合,让输入信息跨越多个全连接层传播。此时反向传播的梯度也可以跨越多层传播,提高了场景重建的准确性。

如图 2 所示,为进一步提升准确性,对于融合后的信息,利用通道注意力[20]进行重要度区分。对于融合信息,利用全局平均池化(glboal average pooling,GAP)将特征映射转换为低维嵌入,得到 $1\times1\times c$ 特征向量,以此屏蔽空间上的分布信息,更好的利用通道间的相关性。然后,引入以输入为条件的动态特性,以提高特征辨别力。最后以 Sigmoid 函数进行激活。此时完成通道维度上重新标定的信息相较于之前的信息具有更强的表征能力。

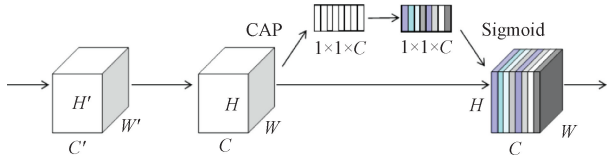

图 2　通道注意力机制

Fig. 2　Channel attention mechanism

1.3　体积渲染

对于已知姿态 P 的 2D 图像,对于图像的每一个体素,从投影中心发射射线 r,从已知角度的方向 d 导出。每个像素的 RGB 值是通过对射线的颜色和密度进行积分得到的。然而实际情况下,这不是一个连续过程。对于每条射线,在近、远平面之间进行随机采样,并使用 MLP 获得采样点的体积密度和颜色。颜色渲染如式(3)所示。

$$C=\sum_{i=1}^{N}T_i(1-\exp(-\sigma_i\theta_i))c_i \tag{3}$$

式中:σ_i 为采样点 i 处体素密度;δ_i 表示采样点 i 与邻近采样点距离,T_i 表示累计透射率,计算过程如式(4)所示。

$$T_i=\exp(-\sum_{j=1}^{i-1}\sigma_j\delta_j) \tag{4}$$

式中:j 表示从第一个采样点到采样点 i 的所有采样点的索引,σ_i 为采样点 i 处的体素密度;δ_i 表示采样点 i 与邻近

采样点的距离。

1.4　双阶段损失函数

本文提出一种双阶段损失函数，旨在提取图片细节。其中，粗阶段通过最小化图像的光线终止分布 $h_{ij}(t)$ 与深度分布的 KL 散度，使得随机变量更加贴近物体表面。通过在像素空间中进行拟合操作，得到较为准确的总体轮廓信息，计算过程如式(5)所示。

$$L_{KL}=E_{D_{ij}}KL[\delta(t-d_{ij})\parallel h_{ij}(t)]=KL[N(D_{ij},\hat{\sigma}_i)\parallel h_{ij}(t)]+const \tag{5}$$

式中：δ 表示 δ 分布，d_{ij} 为重投影深度，D_{ij} 为深度附近的随机变量，t 表示射线距离，$h_{ij}(t)$ 为光线终止分布，N 为正态分布，σ 表示方差。

采用深度监督来训练光线终止分布 $h(t)$，其深度损失如式(6)所示。

$$L_{Depth}=E_{x_i\in X_j}\int\log h(t)\exp\left(-\frac{(t-D_{ij})^2}{2\hat{\sigma}2_i}\right)\mathrm{d}t \tag{6}$$

粗阶段的损失函数为 $L_{COA}=L_{KL}+\lambda_D\times L_{Depth}$，其中 λ_D 是超参数，来平衡颜色和深度监督。

为了更好的提取纹理信息，在精细阶段引入基于梯度的惩罚项和高频感知损失。

在渲染过程中，噪声会对渲染结果产生较大影响，通过添加正则项 TV loss[21] 约束噪声。TV loss 越大表明受噪声影响越大，通过降低 TV loss 来削减图片中相邻像素值的差异。在图像中，像素是离散域，在像素离散域中求和，计算公式如下：

$$L_{TV}=\sum_{i,j}((x_{i,j+1}-x_{ij})^2-(x_{i+1,j}-x_{ij})^2)^{\frac{\beta}{2}} \tag{7}$$

式中：i、j 表示横向、纵向像素角标，β 为可变变量，$\beta<1$，会出现瑕疵小点。$\beta>1$，小点会消除，β 一般取值为 2。

本文引入一种有效的感知损失[22]，对 VGG 网络的卷积输出特征加以限制。如图 3 所示，使用 VGG-16 卷积输出的特征，在特征空间计算 KL 值。

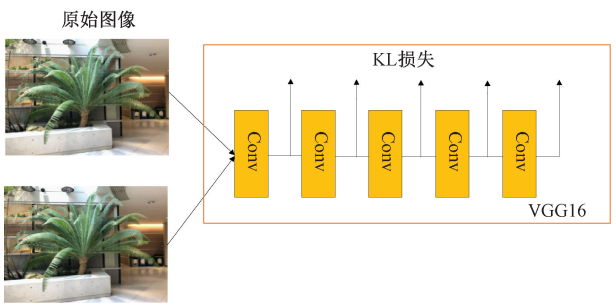

图 3　VGG-16 特征空间计算损失

Fig. 3　Calculating loss in the VGG-16 feature space

感知损失计算公式如下：

$$L_{HA}(G,P)=\frac{1}{CWH}\parallel\Phi(G)-\Phi(P)\parallel_2^2 \tag{8}$$

式中：G 表示地面真值(Ground Truth)，P 表示网络生成结果(Prediction)，Φ 表示预训练 VGG-16 模型的卷积输出，C、W 和 H 是提取特征图的维度。

细阶段的损失函数表示为：$L_{FIN}=L_{KL}+a\times L_{Depth}+b\times L_{TV}+c\times L_{HA}$，其中，$a$，$b$，$c$ 用来平衡颜色和深度。

2　实验结果与分析

2.1　实验设置

1)数据集

LLFF 数据集[23] 由前向视图捕获的 8 个真实世界场景组成。在每个场景中随机选取 2、5 张视图创建训练图像的子集，图片分辨率为 756×1 008。对于每个子集，在其训练子集上运行 COLMAP，以估计相机并收集稀疏关键点用于深度监督。

DTU 数据集[24] 是在具有固定相机轨迹的室内环境中得到的，总共包含 128 个不同的场景，每个场景在 7 种不同的照明条件下拍摄，每个照明条件下都拍摄了 49 张图像。本实验中，采用其测试集的 15 个场景，对于每个场景，使用它们的大小分别为 3、6 的训练视图子集，图片分辨率为 1 600 × 1 200。使用地面实况校准的相机姿势运行 COLMAP 以获得关键点。

2)评价指标

采用峰值信噪比、结构相似性、学习感知图像块相似性作为量化评价指标，峰值信噪比(peak signal-to-noise ratio，PSNR)是一种广泛使用的图像客观评估指标，它基于相应像素之间的误差计算。结构相似性(structural similarity，SSIM)是一种用于测量两幅图像相似性的评估指标，它将失真建模为亮度、对比度和结构的组合。学习感知图像块相似性(learned perceptual image patch similarity，LPIPS)是一种基于深度学习的图像相似性评估方法，通过神经网络提取深度特征进行比较。其中 PSNR 和 SSIM 两项指标数值越高表示实验结果越好，LPIPS 数值越低表示实验结果越好。

3)实施细节

实验设置在每束由相机发出的光线上进行粗采样的参数为 64，进一步进行细采样的参数为 128，每次处理 32 768 根光线，训练批量大小设置为 2 048，学习率设置为从 5×10^{-4} 开始呈指数衰减。进行 100k 次训练迭代至损失函数收敛，其中粗阶段迭代次数为 70k 次，细阶段迭代次数为 30k 次。训练所使用的硬件平台为单张 RTX A5000 显卡，编程语言为 Python，深度学习框架为 Pytorch，软件平台 CUDA 版本为 11.1，PyTorch 版本为 1.10.1。

2.2　特征实验

双阶段损失函数中提到的感知损失，最初是在预训练的深度网络的激活层上，最小化两个激活特征之间的距离。但是激活后的特征非常稀疏，随着网络的加深，特征会更加稀疏，监督效果会越来越差，会严重影响重建质量。

相比之下,卷积层输出的特征则包含更全面的特征,更适合用于监督。

为更直观地展现卷积层输出特征和激活层输出特征的差异,分别输出这两层对单张图像的渲染情况,具体如图 4、5 所示。

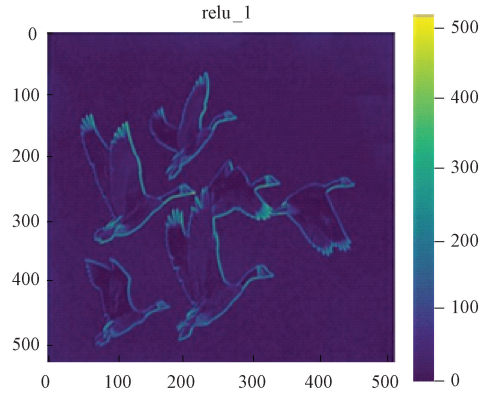

图 4　VGG 激活层输出

Fig. 4　Output of VGG activation layer

实验表明,激活后的特征会产生与真实图像相比不一致的重建亮度,相较于激活层输出的结果,卷积层输出的纹理则更加清晰,为了得到质量更高的渲染图片,使用 VGG-16 卷积输出的特征,在特征空间计算 KL 值。

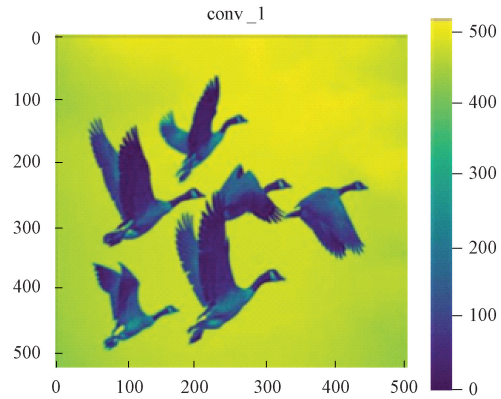

图 5　VGG 卷积层输出

Fig. 5　Output of the VGG convolutional layer

2.3　对比实验

1)定量实验

将提出的模型在 LLFF 数据集上进行评估。对于 LLFF 数据集的 8 个场景,每个场景选择 2 张和 5 张图片作为训练图片,图片分辨率为 756×1 008。

如表 1 所示,使用 LLFF 数据集训练模型时,以 2 张图片作为训练输入,本文模型在 3 个指标上均取得最优值。5 张输入视图时,模型的 LPIPS 指标次优,原因在于 5 张图片输入时,翻转光线引入的补充信息要强于稀疏点云的深度信息。尽管如此,的方法在 PSNR 和 SSIM 上仍取得竞争力的值。

表 1　LLFF 数据集的新视点生成指标

Table 1　View synthesis on LLFF dataset

方法	PSNR ↑		SSIM ↑		LPIPS ↑	
	2-view	5-view	2-view	5-view	2-view	5-view
DS-NeRF	17.08	20.91	0.532	0.603	0.373	0.333
PixelNeRF	16.17	17.03	0.438	0.473	0.461	0.433
FreeNeRF	17.12	21.22	0.476	**0.655**	0.375	0.348
FlipNeRF	16.48	21.48	0.440	0.638	0.351	**0.193**
本文方法	**17.35**	**21.51**	**0.538**	0.653	**0.294**	0.272

使用 DTU 数据集对模型进行评估。对于 DTU 数据集的 15 个测试场景,每个场景选择 3 张和 6 张图片作为训练图片,图片分辨率为 1 600×1 00。

如表 2 所示,使用 DTU 数据集训练模型时,以 3 张图片作为训练输入,PixelNeRF 在 PSNR 和 SSIM 取得最优值,表明其在重建的完整性上存在优势,本文模型在 LPIPS 上取得最优值,表明在重建的细节上表现的更好。6 个输入视图时,本文方法在 PSNR 和 SSIM 上取得最优值,LPIPS 指标取得次优值,实验结果表明了本文方法在稀疏视角下重建的有效性和泛化能力。

2)定性实验

通过可视化方式对重建出的效果进行定性对比,如图 6 和 7 所示。其中 Ground Truth 代表 LLFF 数据集以及 DTU 数据集某一相机视角下的真实图像,其他为本文模型与基线方法同等条件下渲染出的结果。可以看出,在重建完整性方面,DS-NeRF 方法所重建出的物体形状有较大缺失,pixelNeRF 在 DTU 数据集重建完整性较好,但 6 张训练视图,相对于其他算法重建细节较差。原因在于,对于少数视图,当训练和测试域重叠时可以利用数据集先验来补充信息的缺乏。而随着视图增加,信息补充效果受到限制。pixelNeRF 在 LLFF 数据集上重建质量较差,表明其在不同数据上表征能力差异较大,泛化性有待提升。而 FreeNeRF 和 FlipNeRF 方法虽然重建完整性较好,但存在较多的漂浮点,影响重建质量。相较之下,本文提出的模型在重建完整性和结构细节上更接近真实模型。与最近稀疏视角重建方法相比,本文提出的基于信息关注抑

制模块和双阶段损失函数的方法重建的模型更加精确　　完整。

表 2　DTU 数据集的新视点生成指标

Table 2　View synthesis on DTU dataset

方法	PSNR ↑		SSIM ↑		LPIPS ↑	
	3-view	6-view	3-view	6-view	3-view	6-view
DS-NeRF	17.55	21.08	0.577	0.757	0.437	0.339
PixelNeRF	**18.74**	21.02	**0.618**	0.684	0.356	0.293
FreeNeRF	18.56	21.22	0.595	0.734	0.398	0.324
FlipNeRF	18.40	21.37	0.604	0.776	0.369	**0.227**
本文方法	17.93	**21.54**	0.592	**0.785**	**0.348**	0.284

注：加粗数值为该列最优值。

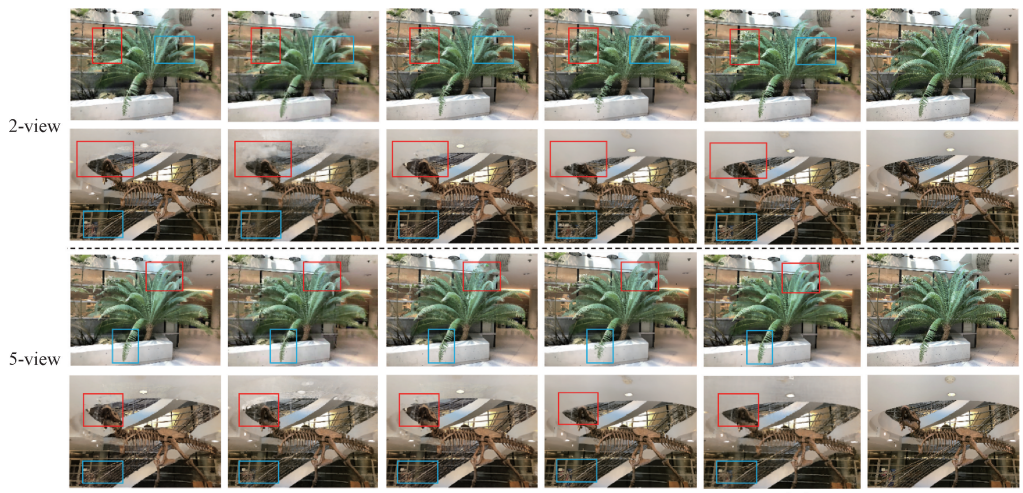

图 6　LLFF 数据集定性比较

Fig. 6　Qualitative comparison of LLFF datasets

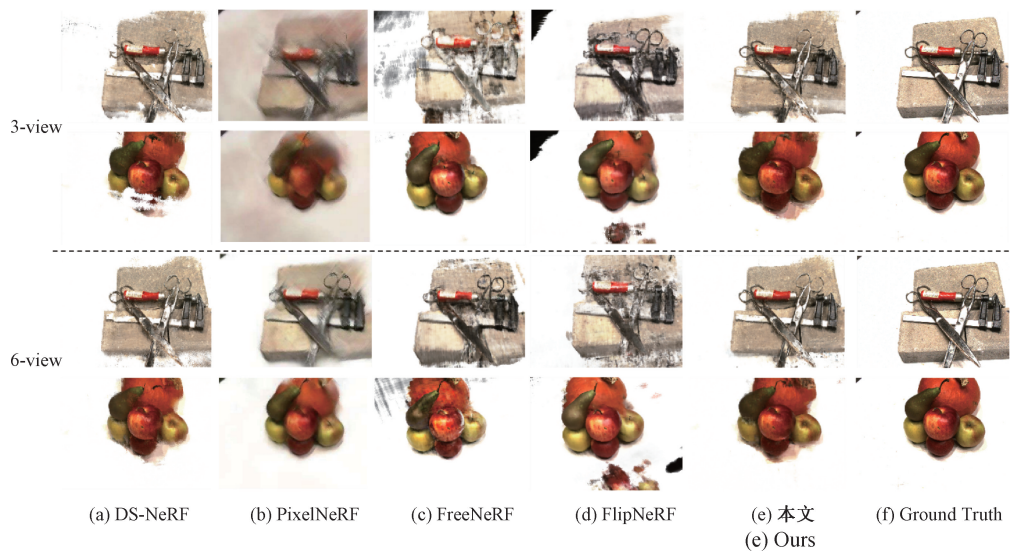

图 7　DTU 数据集定性比较

Fig. 7　Qualitative comparison of DTU datasets

2.4　消融研究

本节在 LLFF 数据集上对模型进行消融研究,以证明所提模块的有效性。表 3 和图 8 分别展示了有无信息关注抑制模块和双阶段损失函数作用的定量和定性结果。

表 3　LLFF 数据集消融实验

Table 3　Ablation experiments on LLFF dataset

关注抑制模块	两阶段损失	PSNR ↑	SSIM ↑	LPIPS ↑
×	×	17.078	0.532	0.373
√	×	**17.379**	**0.538**	0.321
×	√	17.009	0.529	0.317
√	√	17.356	**0.538**	**0.294**

注:加粗数值为该列最优值。

为了验证信息关注抑制模块的性能,在表 3 中展示了有无该模块的定量指标。第 1 行是没有此模块的结果,第 2 行是有此模块的结果。实验结果表明,添加模块后,渲染图片的 3 个指标均提升明显,根本原因在于经信息关注抑制模块处理后的信息去除了包含异常值的采样点特征向量,同时由于残差的作用,用于渲染的信息包含了充分的局部信息,用于细节生成。定性结果如图 8 所示,可以看到应用该模块后伪影大大减少。

对比使用和不使用双阶段损失函数的结果。第 3 行是添加此模块的结果,如表 3 所示。实验表明,双阶段损失函数会使 LPIPS 指标变优,但会令 PSNR 和 SSIM 略有下降。原因在于 PSNR 和 SSIM 是在像素空间中拟合场景,而 LPIPS 则是在特征空间中拟合场景。精细训练阶段本文在特征空间中提取细节时,会使渲染图片中的像素产

图 8　LLFF 数据集消融实验定性对比

Fig. 8　Qualitative comparison of ablation experiments on the LLFF dataset

生小的偏移。

同时使用信息关注抑制器和两阶段损失函数的定性结果如图 8 所示,可以看到相对于单独添加信息关注抑制模块过于平滑,共同作用的渲染图片有更好的感知质量和细节表现能力。

3　结　　论

本文提出一种基于信息关注抑制模块和双阶段损失的 NeRF 网络模型。模型中的信息关注抑制模块利用残差网络跨通道传递信息,兼顾全局信息和局部信息,模块中的归一化微调组件可以过滤异常权重,筛选采样点特征向量,模块还利用通道注意力根据重要度区分关键信息,进一步提升重建精度。此外设计一种双阶段损失对网络进行监督训练,以渐进优化的方式指导训练,充分利用图片的高级特征,提升模型的图像感知以及细节捕捉能力。在 LLFF 数据集上,定量结果表明,整体性能取得最优值,比次优算法性能提升 1.9%,在 DTU 数据集上,定性结果显示,Scan37、Scan55 和 Scan63 等场景重建的完整性和细节水平具有明显优势。然而,模型训练需要较长的时间,因此,未来考虑在提升实时性能和提升内存利用率方面进行改进工作。

参考文献

[1] WU SH CH, WALD J, TATENO K, et al. Scenegraphfusion: Incremental 3D scene graph prediction from RGB-D sequences[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7515-7525.

[2] LI P X, ZHAO H C, LIU P F, et al. RTM3D: Real-time monocular 3D detection from object keypoints for autonomous driving[C]. European Conference on Computer Vision, 2020:644-660.

[3] SARLIN P E, UNAGAR A, LARSSON M, et al. Back to the feature: Learning robust camera

localization from pixels to pose [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021:3247-3257.

[4] LIU L J, XU W P, HABERMANN M, et al. Neural human video rendering by learning dynamic textures and rendering-to-video translation [J]. IEEE Transactions on Visualization and Computer Graphics, 2020, DOI:10.1109/TVCG.2020.2996594.

[5] THIES J, ZOLLHÖFER M, NIEßNER M. Deferred neural rendering: Image synthesis using neural textures[J]. ACM Transactions on Graphics(TOG), 2019, 38(4):1-12.

[6] LOMBARDI S, SIMON T, SARAGIH J, et al. Neural volumes: Learning dynamic renderable volumes from images [J]. ACM Transactions on Graphics(TOG), 2019,38(4):1-14.

[7] SITZMANN V, THIES J, HEIDE F, et al. DeepVoxels: Learning persistent 3D feature embeddings[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019:2437-2446.

[8] 朱代先,孔浩然,秋强,等.注意力机制与神经渲染的多视图三维重建算法[J].电子测量技术,2024,47(5):158-166.
ZHU D X, KONG H R, QIU Q, et al. Attention mechanism and neural rendering algorithm for multi-view 3D reconstruction[J]. Electronic Measurement Technology, 2024,47(5):158-166.

[9] 谢东升,孙滔,史卓鹏,等.基于三维重建的输电线舞动检测方法研究[J].国外电子测量技术,2022,41(3):96-101.
XIE D SH, SUN T, SHI ZH P, et al. Research on transmission line dance detection method based on three-dimensional reconstruction [J]. Foreign Electronic Measurement Technology, 2022, 41(3):96-101.

[10] RÜCKERT D, FRANKE L, STAMMINGER M. ADOP: Approximate differentiable one-pixel point rendering[J]. ACM Transactions on Graphics(TOG), 2022,41(4):1-14.

[11] 高梓皓,张巧芬,王桂棠,等.基于 Resinv-Unet 的图像特征点检测方法[J].国外电子测量技术,2023,42(4):1-7.
GAO Z H, ZHANG Q F, WANG G T, et al. An image feature point detection method based on Resinv-Unet [J]. Foreign Electronic Measurement Technology,2023,42(4):1-7.

[12] 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.

[13] YANG J W, PAVONE M, WANG Y. FreeNeRF: Improving few-shot neural rendering with free frequency regularization[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023:8254-8263.

[14] YU A, YE V, TANCIK M, et al. PixelNeRF: Neural radiance fields from one or few images[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021:4578-4587.

[15] SEO S, CHANG Y, KWAK N. FlipNeRF: Flipped reflection rays for few-shot novel view synthesis[C]. IEEE/CVF International Conference on Computer Vision, 2023: 22883-22893.

[16] WEI Y, LIU SH H, RAO Y M, et al. NerfingMVS: Guided optimization of neural radiance fields for indoor multi-view stereo [C]. IEEE/CVF International Conference on Computer Vision, 2021: 5610-5619.

[17] DENG K L, LIU A, ZHU J Y, et al. Depth supervised NeRF: Fewer views and faster training for free[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022:12882-12891.

[18] KARRAS T, AILA T, LAINE S, et al. Progressive growing of GANs for improved quality, stability, and variation [J]. ArXiv preprint arXiv: 1710.10196, 2017.

[19] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016:770-778.

[20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

[21] RUDIN L I, OSHER S, FATEMI E. Nonlinear total variation based noise removal algorithms[J]. Physica D: Nonlinear phenomena, 1992,60(1-4):259-268.

[22] JOHASON J, ALAHI A, LI F F. Perceptual losses for real-time style transfer and super-resolution[C]. Computer Vision-ECCV 2016: 14th European Conference,2016: 694-711.

[23] MILDENHALL B, SRINIVASAN P P, ORTIZ C R, et al. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines [J]. ACM Transactions on Graphics(TOG), 2019,38(4):1-14.

[24] AANXES H, JENSEN R R, VOGIATZIS G, et al. Large-scale data for multiple-view stereopsis [J]. International Journal of Computer Vision, 2016, 120:153-168.

作者简介

张超,硕士研究生,主要研究方向为计算机视觉、三维重建。
E-mail:13388052309@163.com

袁亮(通信作者),教授,博士生导师,主要研究方向为智能机器人技术、机器视觉与图像处理、数字孪生、工业物联网。
E-mail:yl102439@163.com