

基于外观增强和语义分割的神经辐射场

曹明伟, 黄宝龙, 赵海峰

引用本文

曹明伟, 黄宝龙, 赵海峰. [基于外观增强和语义分割的神经辐射场](#)[J]. 计算机科学, 2025, 52(12): 141-149.

CAO Mingwei, HUANG Baolong, ZHAO Haifeng. [Appearance Enhancement and Semantic Segmentation-based Neural Radiance Fields](#) [J]. Computer Science, 2025, 52(12): 141-149.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[GCP辅助COLMAP框架SFM绝对尺度恢复算法的研究](#)

Optimization and Absolute Scale Recovery of SFM Algorithm in GCP-assisted Colmap Framework
计算机科学, 2025, 52(11A): 250100015-6. <https://doi.org/10.11896/jsjcx.250100015>

[基于下肢骨X光三维重建算法的优化研究](#)

Optimization of 3D Reconstruction Algorithm Based on X-ray of Lower Limb Bone
计算机科学, 2025, 52(11A): 241100152-7. <https://doi.org/10.11896/jsjcx.241100152>

[基于神经辐射场的即时高保真人脸生成算法](#)

Immediate Generation Algorithm of High-fidelity Head Avatars Based on NeRF
计算机科学, 2025, 52(10): 159-167. <https://doi.org/10.11896/jsjcx.241000066>

[基于注意力机制与对比损失的单视图草图三维重建](#)

3D Reconstruction of Single-view Sketches Based on Attention Mechanism and Contrastive Loss
计算机科学, 2025, 52(3): 77-85. <https://doi.org/10.11896/jsjcx.240200102>

[基于区域显著性与空间特征提取的说话人像合成方法](#)

Talking Portrait Synthesis Method Based on Regional Saliency and Spatial Feature Extraction
计算机科学, 2025, 52(3): 58-67. <https://doi.org/10.11896/jsjcx.240300030>

基于外观增强和语义分割的神经辐射场

曹明伟 黄宝龙 赵海峰

安徽大学计算机科学与技术学院 合肥 230601

摘要 神经辐射场(Neural Radiance Fields, NeRFs)因其高效的场景建模和表达能力,已经成为视图合成和渲染领域的重要基础方法。然而,在动态环境中,NeRF在应对复杂光照变化和瞬态对象干扰方面仍存在挑战。由于光照条件发生变化,因此场景中大量不一致外观,进而影响视图合成质量。同时,场景中的动态干扰影响了合成图像的真实感。针对上述问题,提出了一种基于外观增强和语义分割的神经辐射场(Appearance Enhancement and Semantic Segmentation-based Neural Radiance Fields, AS-NeRF)。该方法通过结合锥形体采样与集成位置编码机制,提高外观特征的融合效率,增强模型对光照和相机参数变化的适应能力,从而提升渲染结果的色彩一致性和真实感。此外,采用端到端的轻量级分割网络预测瞬态可视性掩模,有效分离动态对象,避免瞬态元素对合成图像质量的影响。为了验证该方法的有效性,在Photo Tourism数据集上进行了实验,并与多种现有方法进行定性与定量对比分析,实验结果表明所提出的方法在合成精度上优于现有经典方法,并进一步验证了分割掩模在瞬态物体分离中的准确性。

关键词: 神经辐射场; 视图合成; 神经渲染; 三维重建

中图分类号 TP391.41

Appearance Enhancement and Semantic Segmentation-based Neural Radiance Fields

CAO Mingwei, HUANG Baolong and ZHAO Haifeng

School of Computer Science and Technology, Anhui University, Hefei 230601, China

Abstract The accelerated advancement of deep learning has notably propelled 3D reconstruction techniques within the field of computer vision. NeRFs have become an essential methodology due to their adeptness at scene modeling and superior view synthesis. However, challenges persist in dynamic environments, particularly in managing intricate lighting variations and transient object interference. Alterations in imaging conditions may lead to inconsistent scene appearances, thereby degrading the quality of view synthesis. Concurrently, dynamic elements can adversely affect the photorealism of reconstructed scenes. To mitigate these issues, this paper introduces an AS-NeRF. By incorporating frequency regularization and composite positional encoding into the sampling strategy, AS-NeRF enhances the efficiency of appearance feature fusion, thereby augmenting the model's adaptability to variations in lighting and camera parameters. This subsequently improves color consistency and overall rendering realism. Additionally, a lightweight segmentation network is utilized to predict transient visibility masks in an end-to-end manner, effectively isolating dynamic objects and reducing their impact on view synthesis quality. The efficacy of AS-NeRF is verified through experiments conducted on the Photo Tourism datasets, which are compared qualitatively and quantitatively with several existing methods. The experimental results demonstrate that AS-NeRF surpasses existing approaches in terms of synthesis accuracy and further confirms the accuracy of the computed segmentation masks in distinguishing transient objects.

Keywords Neural radiance fields, View synthesis, Neural rendering, 3D reconstruction

1 引言

近年来,隐式神经表示与神经渲染技术的发展,推动了基于神经网络的视图合成方法成为研究热点。在这一背景下,神经辐射场(Neural Radiance Fields, NeRFs)^[1]凭借其高效的场景建模能力和卓越的视图合成效果,成为该领域的代表性方法。NeRF通过多层感知机拟合场景的辐射特性,能够从

有限视角的二维图像数据中生成高质量的新视角图像,在虚拟现实、增强现实、数字文化遗产保护及机器人导航等领域展现出广泛的应用潜力。

尽管NeRF在静态场景的三维重建中表现优异,但其在真实世界的动态环境中仍然面临诸多挑战。例如,在文化遗产保护与虚拟旅游应用中,研究者通常希望利用游客拍摄的图像来重建历史遗迹或者著名地标的三维模型。然而,在处

到稿日期:2025-04-15 返修日期:2025-08-27

基金项目:安徽省高校科研项目(2024AH050045);国家自然科学基金(62372153)

This work was supported by the Anhui Province University Research Project(2024AH050045) and National Natural Science Foundation of China(62372153).

通信作者:曹明伟(caomw@hfut.edu.cn)

理这类动态场景时,NeRF 主要面临两大挑战。一方面,光照条件、天气变化以及相机参数差异导致图像在颜色、亮度和阴影方面存在显著差异,使传统 NeRF 难以实现一致的场景建模,影响视图合成的质量与精度。另一方面,真实场景中常包含大量瞬态对象,如行人、车辆和动物等,这些动态元素时而出现或消失。NeRF 无法区分瞬态与静态内容,重建过程中常混入与场景无关的动态成份,导致生成的三维场景出现重影或虚假结构,降低了视觉真实感。针对上述问题,NeRF-W^[2] 以及后续工作^[3-5] 试图通过引入依赖于图像的外观特征和瞬态掩模来解决这些问题,然而,由于外观特征难以全面适应复杂光照变化,在一些场景下,其渲染结果仍表现出色不一致。此外,瞬态掩模的生成结果往往依赖瞬态向量^[3] 或图像修复方法^[4],难以精准去除动态对象,最终影响场景的重建质量。

针对上述问题,本文提出了一种基于外观增强和语义分割的神经辐射场方法,即 AS-NeRF。该方法通过结合锥形体采样与集成位置编码机制,相比传统 NeRF,能够更高效地将外观特征融入神经场景表示中,特别是在处理像 Photo Tourism^[6] 这样具有不受控视角和光照变化的数据集时,能够生成更加逼真的 3D 视图合成结果。该方法一方面解决了在处理不同分辨率图像和复杂几何尺度时,单射线采样引发的混叠现象;另一方面,其通过在采样过程中建立邻近像素的关联,结合外观特征编码器提取的整体外观特征,实现了多尺度外观特征融合,从而增强了模型对光照和相机参数变化的适应性,提高了渲染结果的色彩一致性和真实感。

此外,AS-NeRF 采用轻量级的分割网络,端到端地精确预测可视性掩模,有效分离瞬态对象,避免了传统 NeRF 方法中动态元素干扰重建效果的问题。实验表明,与现有经典方法相比,AS-NeRF 在处理 Photo Tourism 这种多样化动态场景时,具有更高的精度和鲁棒性,能够有效提升视图合成的精度和视觉效果,实现更真实且连贯的渲染结果。

本文的主要贡献如下:

1) 提出了一种面向动态场景的神经辐射场方法 AS-NeRF,结合外观增强与语义分割策略,针对不受控拍摄条件下的光照变化与动态遮挡问题,实现了更真实一致的 3D 视图合成。

2) 提出了外观增强策略,通过结合锥形体采样与集成位置编码机制,使外观特征更高效地融入神经场景表示中,从而提升了 NeRF 在复杂光照变化和不受控视角数据上的表现。

3) 提出利用轻量级分割网络预测瞬态可视性掩模,通过无监督方式学习图像的语义特征,生成二值可视掩模,实现了静态与瞬态内容的自动分离,增强了其在动态环境中的建模能力。

4) 在 Photo Tourism 数据集上系统地评估了 AS-NeRF 方法,并与现有经典方法进行了全面比较,实验结果验证了本文方法在动态场景中的优越性能。

2 相关工作

2.1 视图合成

传统视图合成方法主要依赖于图像插值^[7-8] 和光场

插值^[9] 技术,通过密集采样来生成高质量的新视角图像。然而,这些方法通常需要大量的输入图像,并且在处理复杂几何结构和遮挡关系时表现不佳,尤其在视角发生较大变化时难以生成自然逼真的图像。为克服这些方法的局限性,基于几何的图像重投影方法^[10-14] 被提出,即先通过二维图像重建三维模型,再基于该模型渲染目标视角图像。这类方法通常依赖于运动恢复结构 (Structure-from-Motion, SfM) 等三维重建技术,并采用点云^[15-16]、网格^[13,17-18] 或体素^[19-20] 等三维表示方式。此外,计算机图形学中常见的光栅化^[21] 和光线追踪^[22-23] 渲染方法在视图合成任务中也发挥了重要作用。其中光栅化方法的计算效率较高,适用于实时渲染;而光线追踪能够模拟光照交互,生成更真实的视觉效果,但计算成本较高。

随着深度学习技术的发展,基于神经网络的视图合成方法取得了显著进展。卷积神经网络^[24-25] 通过端到端学习视角变换和几何关系,提升了自由视角合成能力,但受局部感知特性的影响,在处理复杂几何结构和大视角变化时仍存在一定局限性。生成对抗网络^[26] 在纹理生成、光照模拟和细节恢复方面表现优异,但在稀疏输入视角和复杂光照环境下易出现模糊、细节丢失或光照不一致的问题。近年来,NeRF 作为隐式场景表示方法,在视图合成领域取得了突破。NeRF 结合深度神经网络和体渲染技术,不仅能够建模复杂的光照变化,还能生成高保真的三维结构,避免传统方法中的几何畸变。然而,NeRF 的计算成本较高,训练时间长,并且对输入视角的稠密性要求较高,同时在处理动态场景时易引入伪影。针对上述问题,研究者提出了一系列优化策略,包括提高训练效率^[27-31]、适应稀疏输入^[32-35],以及扩展至动态场景^[2-5] 和动态人体^[36-37] 等。

在面对光照变化和动态遮挡的场景时,NeRF-W^[2] 通过引入外观向量和瞬态向量提升了原始 NeRF 的鲁棒性。其中,外观向量用于标准化不同光照条件下的图像外观,增强场景的一致性;瞬态向量则用于生成掩模,分离短暂出现的动态物体,从而增强模型对数据丢失与噪声的抗干扰能力。后来,Ha-NeRF^[3] 进一步改进了外观建模与瞬态掩模生成策略,增强了模型对复杂光照和遮挡信息的适应性。IE-NeRF^[4] 引入图像修复 (Image Inpainting) 技术,对缺失或损坏区域进行填补,有效提升了动态场景中的视图合成质量。SF-NeRF^[5] 则采用语义感知遮挡滤波方法,缓解遮挡带来的干扰,在真实场景中实现了更高的重建精度和渲染质量。

2.2 隐式神经表示

传统三维数据表示方式包括体素、点云和三角网格,它们在不同场景下具有各自的优势和局限性。体素能够完整表达三维空间信息,但存储成本较高,难以扩展到高分辨率场景。点云表示紧凑,但缺乏拓扑信息,不易直接用于渲染。三角网格适用于精确建模,但在复杂几何结构下受限于拓扑约束。此外,这些离散表示方法在未观测区域容易产生伪影,影响渲染质量。相较之下,隐式神经表示 (Implicit Neural Representation, INR) 通过神经网络学习信号的连续映射,在存储效率、表达能力和渲染质量方面均优于传统方法。

现有的隐式神经表示方法可进一步分为神经隐式表面和

神经体渲染两类。神经隐式表面利用神经网络隐式建模物体几何模型,如 DeepSDF^[38] 采用基于带符号距离函数 (Signed Distance Function, SDF) 进行曲面建模,而 Occupancy Networks^[39] 通过分类边界表示三维表面,后续方法如 PIFu^[40] 和 SIREN^[41] 进一步提升了模型精度和细节表达能力。而神经体渲染则侧重于学习物体的体积属性,并利用体渲染进行射线积分,以合成高质量图像。Neural Volumes^[42] 方法采用编码-解码器网络隐式存储颜色和透明度信息,并结合射线进行算法进行渲染,为 NeRF 的发展奠定了基础。

2.3 基于 NeRF 的多尺度建模

尽管 NeRF 在静态场景下表现出色,但在高分辨率、多尺度和无界场景下存在空间混叠 (Aliasing)、细节丢失以及远景建模不足等问题。为此, Mip-NeRF^[43] 提出了一种基于锥体的分层采样策略,用于替代原始 NeRF 的单个采样方式,并引入集成位置编码 (Integrated Positional Encoding, IPE) 来表达不同尺度下的空间信息。此方法不仅克服了空间混叠问题,还统一了 NeRF 中由粗到细的两阶段 MLP,从而提升了建模效率。在此基础上, Mip-NeRF 360^[44] 针对无界场景提出了非线性空间参数化方法,通过对相机空间的紧致变换增强了远景细节建模能力,进一步拓展了 NeRF 在复杂开放场景中的适用范围。

本文提出的 AS-NeRF 方法在模型结构上延续了 Mip-NeRF 的核心设计理念,采用锥体采样与集成位置编码作为基础框架,并结合图像外观编码和像素关联机制,实现了对动态场景中外观变化的鲁棒建模。相较于 Mip-NeRF, AS-NeRF 进一步引入了可视性分割网络,以无监督方式显式地剥离动态目标,提升了在动态复杂场景下的合成质量。

3 本文方法

本文提出的 AS-NeRF 方法的总体结构如图 1 所示。该方法以 N 幅图像 $\{I_i\}_{i=0}^{N-1} \in \mathbb{R}^{H \times W \times 3}$ 及其相机参数和稀疏点云信息作为输入,面向动态场景的视图合成任务,对传统 NeRF 在处理动态内容方面的局限性进行了改进与扩展。

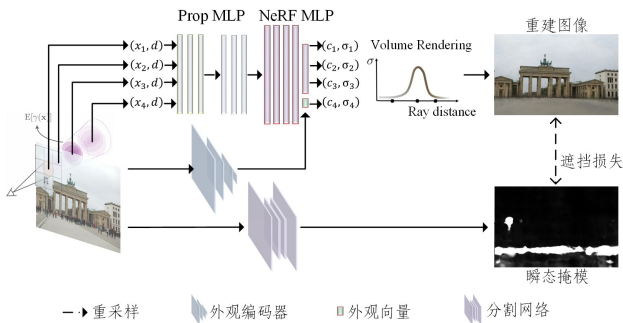


图 1 AS-NeRF 的总体结构

Fig. 1 Overall structure of AS-NeRF

AS-NeRF 在结构设计上继承了 Mip-NeRF^[43] 的锥形体采样与集成位置编码机制,有效缓解了空间混叠问题,并提升了对多尺度空间结构的表达能力。在此基础上, AS-NeRF 在计算像素颜色时,由于锥形体采样自然建立了像素之间的关联,综合考虑了临近像素信息和外观特征编码器提取的图像外观特征,因此能够处理更多细节,并保证了高质量的渲染

效果,尤其在动态场景中的性能更加优越。此外,针对动态场景中瞬态目标对合成质量的干扰, AS-NeRF 采用一个轻量级分割网络^[45],以无监督的方式自动识别场景中的瞬态目标,并生成精确的像素级可视性二值掩模,用于将这些动态对象从场景中剥离出来,有效减少动态目标干扰对渲染质量的影响。

3.1 辐射场优化

如图 2 所示, AS-NeRF 的神经辐射场优化过程分为两个启发式步骤和最终的 NeRF 阶段。具体而言, AS-NeRF 在双启发式阶段通过两个启发式步骤逐步学习场景的粗略表示,即通过小规模参数数量的 MLP 网络预测三维点的体密度,通过优化渲染方程的权重值,逐步确定采样点的精确位置。模型通过双启发式步骤高效构建三维场景的初步几何形状,能够优先关注关键的几何结构特征,而不必立即处理复杂的光照和颜色信息。在最终阶段,通过大规模参数数量的 MLP 网络精准预测体密度和颜色,从而进行图像渲染。总体过程由 3 个独立的 MLP 网络实现,通过对其进行联合优化,以获得场景的隐式表示。

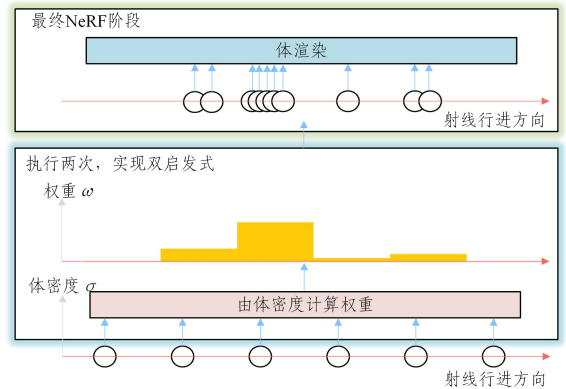


图 2 AS-NeRF 的辐射场优化过程

Fig. 2 Optimization process of the radiance fields in AS-NeRF

同时, AS-NeRF 通过一个外观特征编码器,从输入图像中提取一个紧凑而富有代表性的外观向量。该外观向量全面捕捉了图像的整体色调、光照和纹理信息,有效反映了各幅图像间的外观差异。整体形式化过程如下:

$$\sigma_1(t) = MLP_{\theta_1}[\gamma_x(\mathbf{r}(t))] \quad (1)$$

$$\sigma_2(t) = MLP_{\theta_2}[\gamma_x(\mathbf{r}(t))] \quad (2)$$

$$\sigma(t), \mathbf{c}^c(t) = MLP_{\theta_3}[\gamma_d(\mathbf{d}), \sigma_2(t), \theta_i^c], \theta_i^c = E_\phi(I_i) \quad (3)$$

其中, $\{\theta_i\}_{i=1,2,3}$ 表示 3 个独立 MLP 网络的参数, E_ϕ 表示外观特征编码器, θ_i^c 表示提取的整体外观向量。式 (1) 和式 (2) 表示双启发式阶段, 此时会计算对应采样点的权重值, 用于指导下一阶段采样点位置更新。式 (3) 输出最终体密度值 $\sigma(t)$, 以及依赖于观察视角 \mathbf{d} 和全局外观特征的颜色值 $\mathbf{c}^c(t)$, 随后通过式 (4) 计算渲染射线颜色:

$$\hat{C}(\mathbf{r}) = R(\mathbf{r}, \mathbf{c}, \sigma) = \sum_{k=1}^K T(t_k) \cdot \alpha(\sigma(t_k) \delta_k) \cdot \mathbf{c}^c(t_k) \quad (4)$$

其中, $\alpha(x) = 1 - \exp(-x)$ 表示 Alpha 合成 (Alpha Compositing), $\delta_k = t_{k+1} - t_k$ 表示相邻采样点之间的距离。 $T(t_k)$ 表示从相机起点到射线采样点 t_k 的透射率:

$$T(t_k) = \exp\left(-\sum_{k'=1}^{k-1} \sigma(t_{k'}) \cdot \delta_{k'}\right) \quad (5)$$

3.2 外观增强策略

为了将外观特征融入场景表示中, AS-NeRF 采用了外观

增强策略,以提升不同视角下的颜色一致性和重建质量。具体而言,AS-NeRF通过优化采样策略提取图像外观注入辐射场表示中,以增强场景表示的外观一致性。

首先,AS-NeRF使用圆锥形采样策略^[43-44],自适应地建立当前三维点与邻近射线的空间关联,使得每个采样点在经过多层MLP处理后,能够充分利用周围空间的信息进行补充,生成多尺度颜色特征。其次,AS-NeRF引入基于CNN的外观特征编码器(Appearance Feature Encoder, AFE),从输入图像中提取一个紧凑且具有强表征能力的外观特征向量。该向量与多尺度颜色特征进行拼接,并经过MLP计算出最终的三维点颜色值。通过这一机制,AS-NeRF能够有效地补偿

视角变化导致的外观差异,确保渲染场景在不同视角下呈现出稳定、自然的外观风格。

图3展示了外观特征编码器的网络结构。具体而言,外观特征编码器 E_{ϕ} 接收图像 I_i 作为输入,通过一系列卷积层逐步提取图像的整体外观特征,并最终将其映射为一个48维的外观向量 ℓ_i 。该向量能够捕捉图像整体外观特征,将因光照条件、相机位置等因素带来的大量不一致外观从图像中解耦出来。随后,该外观向量通过与多尺度空间位置的颜色特征相结合生成最终颜色值,该颜色值不仅由几何位置和观察角度决定,还受到输入图像全局外观特征的影响。这种机制有效解决了由外观差异所引发的颜色不一致问题。

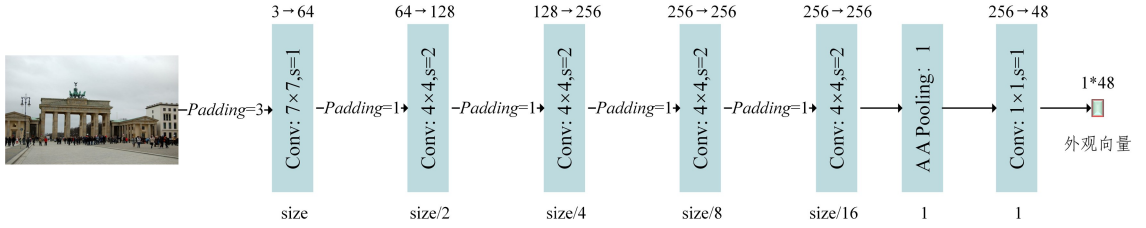


图3 外观特征编码器的网络结构

Fig. 3 Structure of appearance feature module

3.3 分割掩模预测

在视图合成任务中,针对包含瞬态物体的动态场景,需要有效检测和处理的瞬态物体,避免影响最终渲染质量。为此,本文设计了一个基于轻量级分割网络^[45]的分割掩模预测模块,以端到端的方式训练并推理训练图像中的瞬态区域,以无监督学习方式,利用语义分割结果引导渲染优化。

分割网络采用了典型的编码-解码架构,其结构如图4所示。首先,特征提取部分通过多层 3×3 卷积层提取图像的局部纹理信息,并逐步深入捕捉更细致的图像特征。为扩大网

络的感受野,采用了空洞卷积层和普通卷积层相结合的方式,进一步提高特征分辨率,从而捕获更大范围的上下文信息。然后,对得到的特征进行拼接和归一化操作,使用PRReLU进行激活获得联合特征。此外,为了有效减少网络的参数量并增强对全局信息的理解,进一步使用全局自适应池化操作。这一设计使得网络在保持高效计算的同时,能够考虑局部细节和全局上下文,进而提高分割精度。池化后的特征通过两层全连接层生成权重向量。最终,通过与联合部特征进行点乘,精准预测出分割掩模,以区分静态与瞬态物体。

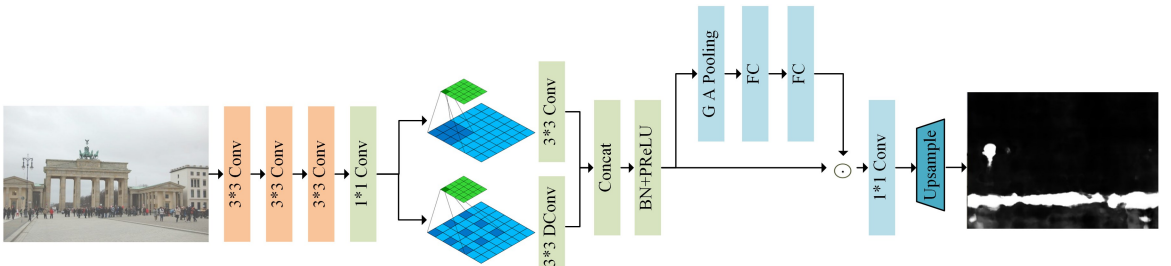


图4 可视性掩模预测模块

Fig. 4 Visible mask prediction module

分割网络和神经辐射场通过联合优化如式(6)所示的损失,以无监督方式进行端到端训练,并解耦静态和瞬态现象。

$$\mathcal{L}_t = M_{ij} \| C(r_{ij}) - \hat{C}(r_{ij}) \|_2^2 + \lambda_t (1 - M_{ij})^2 \quad (6)$$

其中, $C(r_{ij})$ 和 $\hat{C}(r_{ij})$ 分别表示真实颜色和预测颜色; M_{ij} 是由分割网络预测的语义掩模,用于指示像素 (i, j) 是否属于静态场景; \mathcal{L}_t 用于指示图像中的瞬态对象,以避免场景表示受到瞬态对象的干扰,提升最终渲染的质量; λ_t 是平衡两项损失的权重系数。

式(6)中,第一项通过掩模计算静态区域的重建误差,因此期望在静态区域时,掩模 M_{ij} 接近于1,使得颜色重建误差得到有效优化;反之,期望在瞬态区域时,掩模 M_{ij} 接近于0,

减少瞬态对象对重建结果的影响。第一项损失函数会直接促使网络将所有区域划分为瞬态区域,即 M_{ij} 全为0。因此需要添加第二项损失函数,对掩模进行正则化约束,鼓励掩模预测尽可能明确地区分静态和瞬态区域。

3.4 损失函数

为了增强AS-NeRF在动态场景下的外观建模能力,并确保生成视图在外观空间中的多样性,在优化过程中引入了外观模式探索(Appearance Mode Seeking, AMS)损失以及外观正则化(Appearance Regularization, AR)损失。总损失如下:

$$\mathcal{L} = \mathcal{L}_t + \lambda_{AMS} \mathcal{L}_{AMS} + \lambda_{AR} \mathcal{L}_{AR} \quad (7)$$

其中, \mathcal{L}_t 为式(6)所表示的遮挡损失, \mathcal{L}_{AMS} 为外观模式探索损失, \mathcal{L}_{AR} 为外观正则化损失, λ 为对应损失项的权重。

1) 外观模式探索损失

在图像生成任务中,外观特征的多样性是渲染高质量视图的关键因素。然而,在无监督学习或外观编码过程中,可能会出现模式塌缩(Mode Collapse)现象,即模型倾向于生成过于相似的外观特征,导致视图合成过程中出现颜色风格单一、缺乏多样性的问题。为了缓解这一问题,AS-NeRF引入了Mode Seeking损失^[46]。这一损失最初用于生成对抗网络的训练,旨在鼓励生成器覆盖整个数据空间,并提升生成样本的多样性。Mode Seeking损失函数被定义为^[46]:

$$\mathcal{L}_{ms} = \max_G \left(\frac{d_1(G(c, z_1), G(c, z_2))}{d_2(z_1, z_2)} \right) \quad (8)$$

其中, z_1 和 z_2 表示从潜在空间中随机采样的两个不同潜在变量, $G(z)$ 和 $G(z')$ 分别表示生成器 G 以 z_1 和 z_2 为输入生成的图像。

该损失函数鼓励不同潜在向量生成的图像在外观上具有足够大的差异,即潜在变量之间的距离越大,生成图像之间的视觉差异应当越明显,从而提升外观特征的多样性。此外,AS-NeRF除了使用外观向量 ℓ_j ,还随机选取训练集中另一幅图像的外观向量 ℓ_j' ,将这两个外观向量分别输入最终阶段的MLP网络中,并生成对应的渲染图像 $\hat{C}_j^{\ell_j}$ 和 $\hat{C}_j^{\ell_j'}$ 。最终,外观模式探索损失定义为:

$$\mathcal{L}_{AMS} = \frac{|\hat{C}_j^{\ell_j} - \hat{C}_j^{\ell_j'}|_1}{|\ell_j - \ell_j'|_1 + 1 \times 10^{-5}} \quad (9)$$

这一损失的引入确保了AS-NeRF在外观空间中的广泛探索,能够避免陷入模式塌缩,并增强不同视角下的渲染一致性,使得合成图像更加自然、真实。

2) 外观正则化损失

为了进一步确保外观特征编码器在编码过程中能够保持外观的一致性,AS-NeRF还设计了外观正则化损失。该损失的核心目标是保证在“编码—解码—再编码”的循环过程中,外观特征能够保持稳定,从而提升渲染结果的一致性。具体实现方式为:首先,对原始训练图像进行编码,得到对应的外观向量,并将其注入网络中进行渲染,生成合成图像 $\hat{C}_i^{\ell_i}$;随后,对合成图像再次进行编码,获取新的外观向量 ℓ_i' ,并计算其与原始外观向量 ℓ_i 之间的L1损失:

$$\mathcal{L}_{AR} = |\ell_i - \ell_i'|_1 \quad (10)$$

外观正则化损失有效确保了外观特征编码器的稳定性,能避免编码器外观信息的不一致性导致渲染结果出现颜色偏差或风格不稳定的现象。此机制进一步提升了AS-NeRF在多视角合成任务中的外观一致性,使得最终的合成图像更加真实、自然。

4 实验结果与分析

本章展示了AS-NeRF在Photo Tourism数据集中3个著名地标场景上进行的实验结果,验证了其性能。首先,分别介绍了数据集、评估指标和实现细节。接着,将AS-NeRF与现有的最新动态场景视图合成算法进行了对比分析。最后,通过详细的消融实验,进一步分析了AS-NeRF各个模块的贡献和效果。实验结果表明,AS-NeRF在处理动态场景和瞬态物体时,能够显著提高渲染质量和稳定性。

4.1 数据集和评估指标

1) 数据集

Photo Tourism数据集是计算机视觉和三维重建领域的重要基准数据集,由华盛顿大学研究团队开发。其核心特点是利用互联网中的照片集合构建知名地标的三维结构,提供多视角图像及相机参数信息,用于高精度的三维重建和视图合成研究。NeRF-W率先在该数据集上对NeRF方法的视图合成性能进行了评估,并给出了训练集和测试集的划分方案,这一划分方式被后续改进方法广泛沿用,以确保不同模型能够在相同实验条件下进行公平的比较。

Photo Tourism数据集涵盖多个世界知名地标,其中本文使用的Brandenburg Gate(勃兰登堡门)和Sacre Coeur(圣心大教堂)场景分别包含1363和1178幅图像,每幅图像均配备完整的相机内部参数信息,并附带稀疏点云数据。这些数据不仅便于研究人员进行多视角实验,还能用于不同方法的对比评估,以推动视图合成技术的发展。

2) 评估指标

本文实验采用了PSNR,SSIM和LPIPS这3种评估指标来衡量AS-NeRF的性能。PSNR主要用于量化重建图像与原始图像之间的误差,较高的PSNR值通常表明重建图像具有更好的视觉质量。SSIM从亮度、对比度和结构信息等方面评估图像的相似性,能够有效衡量模型对图像结构的保持能力。而LPIPS则是一种感知指标,它基于深度神经网络计算图像之间的感知距离,更贴近人眼的主观视觉感知。通过上述3项指标,可以全面衡量模型在重建质量、结构保持能力和视觉一致性方面的性能。

4.2 实验设置

AS-NeRF的网络结构由3个MLP组成。为了减小模型规模,双启发式阶段仅包含4层全连接层,每层有256个通道,并使用ReLU激活函数,通过Softplus激活函数计算体密度值。最终阶段由8层全连接层组成,每层包含1024个通道,同样采用ReLU激活函数,并使用Softplus计算密度值。在第4层全连接层之后,进行跳跃连接,将初始位置编码后的特征与网络的输出进行拼接,以增加网络的表达能力。此外,最终阶段会先通过一层全连接层将最后的1024个通道压缩到256个通道,然后与编码后的观察视角和外观特征进行拼接,送入一层有128个通道的全连接层,并使用Sigmoid激活函数生成最终的密度值和颜色值。

AS-NeRF是基于PyTorch深度学习框架实现的。在实验过程中,训练集、测试集和验证集的划分与NeRF-W^[2]保持一致,数据预处理在Ha-NeRF^[3]的基础上进行了修改,以实现与Mip-NeRF 360^[44]提出的射线采样方法的无缝衔接。由于AS-NeRF采用了更复杂的网络架构和特征融合机制,每次迭代的计算开销相对较高。为平衡训练效率与收敛效果,实验设置为100万次迭代,相当于7个轮次,而传统方法通常训练20个轮次。此外,优化器选用了Adam,学习率为0.0001,同时为了提高训练的稳定性,采用了指数衰减策略,使学习率在训练过程中逐渐降低。

AS-NeRF的实验过程在NVIDIA RTX 4090显卡上进行,在模型效率与资源开销方面,AS-NeRF具有显著优势。

其训练时间为 23 h, 相比 Ha-NeRF 的 35.1 h 缩短了约 34%。在推理速度上, AS-NeRF 达到 0.18FPS, 是 Ha-NeRF 0.07FPS 的 2.6 倍, 显著提升了实时渲染能力。参数规模方面, AS-NeRF 为 11.9MB, 小于 Ha-NeRF 的 17.4MB, 减少了约 31.6% 的存储开销。综合来看, AS-NeRF 在保持重建质量的同时, 兼具更高的效率和更低的资源消耗, 具备更强的实际部署潜力。

4.3 评估结果

4.3.1 定量实验结果

为了全面评估 AS-NeRF 在视图合成任务中的性能, 本文在 Photo Tourism 数据集上进行了定量和定性分析。表 1 列出了不同方法在 Photo Tourism 数据集两个地标场景上的

表 1 AS-NeRF 等方法在 Photo Tourism 数据集上的定量比较结果

Table 1 Quantitative comparison of AS-NeRF and state-of-the-art methods on the Photo Tourism dataset

方法	Brandenburg Gate			Sacre Coeur		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
NeRF ^[1]	18.90	0.816	0.232	15.60	0.716	0.292
NeRF-W ^[2]	24.17	0.891	0.167	19.20	0.807	0.192
Ha-NeRF ^[3]	24.04	0.877	0.139	20.02	0.801	0.171
SF-NeRF ^[5] (30-fews)	23.23	0.846	0.178	19.64	0.757	0.186
IE-NeRF ^[4]	25.33	0.898	0.158	20.37	0.861	<u>0.169</u>
AS-NeRF(Ours)	25.65	0.880	0.112	20.80	<u>0.816</u>	0.149

4.3.2 视图合成结果

为了更直观地展示和全面分析各方法在动态场景中的表现, 本文选取了包含显著遮挡的场景, 并对比了不同方法在 Photo Tourism 数据集上的视图合成效果。具体结果如图 5 所示。

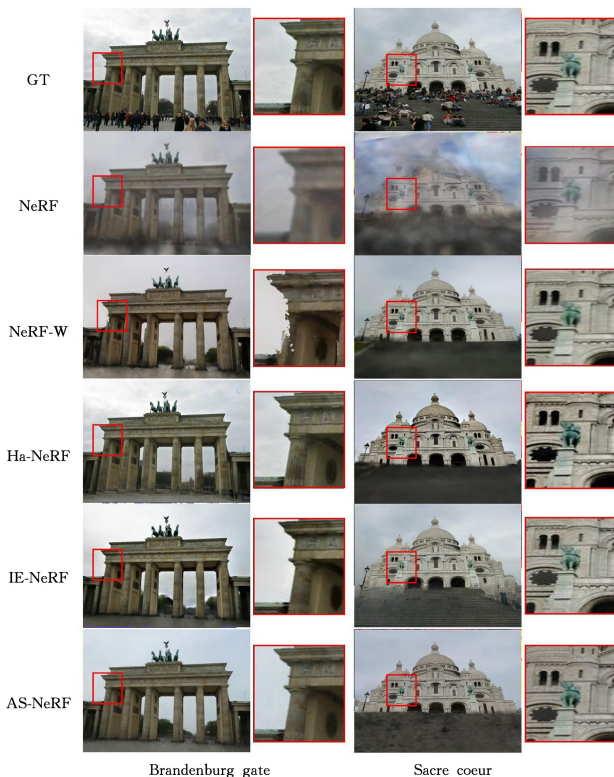


图 5 AS-NeRF 等方法在 Photo Tourism 数据集上的定性分析结果

Fig. 5 Qualitative analysis results of AS-NeRF and state-of-the-art methods on the Photo Tourism dataset

定量比较结果。其中, 加粗表示最佳结果, 下划线表示次佳结果。实验结果表明, AS-NeRF 在 PSNR 和 LPIPS 指标上均取得了最优表现, 同时在 SSIM 指标上也表现出较强的竞争力。这表明 AS-NeRF 在视图合成任务中能够生成高保真的合成图像, 有效保留图像的细节与整体结构。需要说明的是, NeRF-W^[2] 需要额外在测试图像的左侧优化外观嵌入, 然后在右侧计算损失, 这会导致额外的计算开销; 而 SF-NeRF^[5] 仅使用 30 张图像进行训练, 虽然适用于少样本场景, 但在完整场景重建方面存在一定局限性。此外, 为了更直接地比较各方法的合成效果, 选用测试集中的样本均为无遮挡的建筑图像, 以评估各方法在外观建模和瞬态处理方面的能力。

的渲染质量明显优于其他方法。在对比方法中, NeRF 无法有效处理瞬态物体, 导致渲染结果存在显著的色彩偏差和阴影伪影, 严重影响了最终图像的真实性。NeRF-W^[2] 和 Ha-NeRF^[3] 虽然通过引入外观向量建模不同的光照变化, 在一定程度上提升了视觉效果, 但仍然无法完全避免重影和伪影的出现, 尤其是 NeRF-W, 在建筑细节区域, 其伪影问题尤为明显。而 IE-NeRF^[4] 试图通过图像修复过程来消除瞬态影响, 却在重建建筑细节上产生了严重的模糊现象。

相比之下, 本文提出的 AS-NeRF 通过引入 Mip 采样策略, 并结合更精准的分割掩模优化策略, 使得瞬态物体的去除更加精准, 同时有效减少了视图合成中的伪影问题。此外, AS-NeRF 在建筑结构的细节恢复方面也表现优异, 能够准确捕捉复杂纹理, 减少边缘模糊现象, 使得渲染结果更加清晰、逼真。特别是在高动态范围和光照变化较大的场景中, AS-NeRF 依靠增强的全局外观建模能力, 能够更好地保持色彩一致性, 避免局部区域出现色偏或曝光不均的问题, 从而确保整体渲染效果的稳定性和真实性。

4.3.3 掩模划分结果

NeRF-W, Ha-NeRF 和 IE-NeRF 都采用了类似的分割策略, 为了深入评估不同方法在瞬态物体检测和分割上的表现, 图 6 展示了这些方法与本文 AS-NeRF 生成的瞬态掩模的对比结果。可以看出, NeRF-W 和 IE-NeRF 在处理瞬态物体时存在较明显的误识别现象, 很多瞬态区域没有被识别到, 影响了场景重建精度。而 Ha-NeRF 将一些静态区域识别为瞬态区域, 导致重建场景细节丢失。相比之下, AS-NeRF 实现了更加精准的分割掩模预测, 能够更精确地区分静态和瞬态区域, 避免出现瞬态区域残留或过度修复的问题。在复杂场景下, AS-NeRF 的掩模能够精确覆盖瞬态物体, 同时最大程度地保留背景细节, 使得渲染效果更加真实、自然。

从对比结果可以明显看出, AS-NeRF 在处理动态环境时

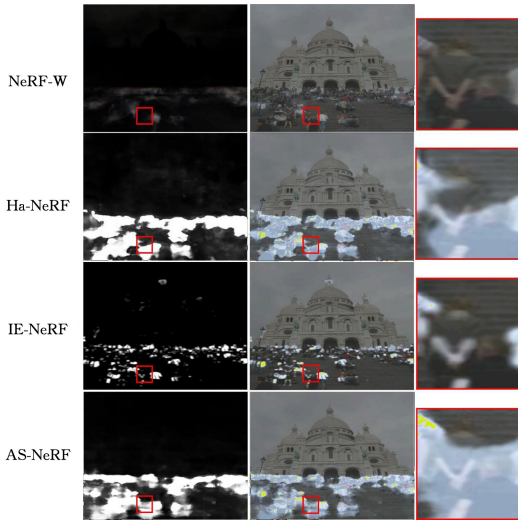


图6 AS-NeRF等方法在Photo Tourism数据集上的掩模划分结果

Fig. 6 Mask segmentation results of AS-NeRF and state-of-the-art methods on the Photo Tourism dataset

4.3.4 外观统一结果

此外,为了验证外观统一编码模块的性能,本文随机选择多张风景图像,并对其进行编码以提取外观向量,从而测试在不同观察视角下外观风格的一致性。这种方式能够渲染出具有随机外观特征的图像,使得在任意视角下都能观察到编码器对外观一致性的影响。为进一步评估外观特征编码器在Brandenburg Gate数据集上的稳定性,本文对不同视角下的合成图像进行可视化分析,可视化结果如图7所示,图中展示了编码器对外观特征的处理效果,包括风格一致性、颜色保持及细节还原能力,直观反映其在复杂场景中的适用性。



图7 在Photo Tourism数据集上的外观可视化结果

Fig. 7 Appearance visualization results of AS-NeRF on the Photo Tourism dataset

4.4 消融实验

本文通过充分的消融实验,验证了AS-NeRF中每个模块的有效性,所有实验的参数均保持一致,以确保结果的可比性和可靠性。通过逐步移除或替换不同的模块,分析其对最终性能的影响。具体而言,AS-NeRF-A表示仅采用本文提出

的外观增强策略,而AS-NeRF-T则表示仅使用语义分割掩模预测模块。

表2列出了消融实验结果,可以明显看出,本文设计的外观增强策略(AS-NeRF-A)与语义分割掩模预测模块(AS-NeRF-T)都显著提升了视图合成的质量。这表明仅采用外观增强策略或仅使用语义分割掩模预测模块均有助于提高图像合成质量。而结合了外观增强与语义分割掩模预测模块的完整模型AS-NeRF在多项指标上都超过了上述单一模块的性能,其中PSNR达到了25.65,SSIM为0.880,LPIPS为0.112。相比基准模型,AS-NeRF在3个评价指标上均取得显著提升,这一结果充分验证了外观增强策略与语义分割掩模预测模块在AS-NeRF框架中的有效性。两者结合的优势在于不仅能够增强图像细节表现和整体一致性,还能更好地适应复杂场景中的动态变化和遮挡,提高视图合成的真实感和稳定性。

表2 在Brandenburg Gate数据集上的消融结果

Table 2 Ablation results on the Brandenburg Gate dataset

方法	PSNR(↑)	SSIM(↑)	LPIPS(↓)
基线	24.04	0.877	0.139
AS-NeRF-A	24.48	0.887	0.116
AS-NeRF-T	24.85	0.888	0.120
AS-NeRF	25.65	0.880	0.112

4.5 超参数消融实验

在计算遮挡损失时,AS-NeRF通过引入第二项损失函数对语义掩模进行正则化约束,从而鼓励分割网络清晰地区分静态和瞬态区域。 λ_r 的设置对模型的性能具有至关重要的影响。具体而言,当 λ_r 偏小,静态区域倾向于被判定为不可见,导致图像细节丢失且重建精度降低;当 λ_r 偏大时,瞬态区域无法有效消除,错误地从这些区域生成重影,影响了最终重建效果。表3列出了不同权重设置下,在Brandenburg Gate场景取得的PSNR值。可以看出,在 λ_r 取0.015时,PSNR值达到了最佳水平,表明在这一权重设置下,模型在静态与瞬态区域的区分和细节上均取得了最优效果。

表3 正则化项权重对PSNR影响的消融结果

Table 3 Ablation results of PSNR with respect to the regularization

weight	
λ_r	Brandenburg Gate
0.006	—
0.008	24.16
0.010	24.25
0.012	25.04
0.015	25.65
0.018	25.12
0.020	24.10
0.050	—

结束语 本文提出了一种基于外观增强和语义掩模的神经辐射场AS-NeRF,旨在解决NeRF在动态环境中的挑战。首先,AS-NeRF引入了锥形体采样与集成位置编码机制,通过改进采样方式提高了空间信息的精确性,同时利用外观特征编码器统一了不同视角下的外观表现,强化了多尺度信息的融合,进而提升了外观一致性与重建效果。其次,AS-NeRF结合轻量级分割网络预测掩模,避免动态物体的影响,

以无监督的方式解决了传统 NeRF 在动态场景中的外观不一致和瞬态物体干扰问题。此外, AS-NeRF 通过 AMS 和 AR 损失优化了外观建模, 确保生成视图在外观空间中具有多样性和一致性。实验表明, 该方法在重建精度和瞬态物体去除上优于现有经典方法, 同时具有较好的外观结果。

参 考 文 献

- [1] 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.
- [2] MARTIN-BRUALLA R, RADWAN N, SAJJADI M S M, et al. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 7210-7219.
- [3] CHEN X, ZHANG Q, LI X, et al. Hallucinated Neural Radiance Fields in the Wild[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 12943-12952.
- [4] WANG S, XU H, LI Y, et al. IE-NeRF: Exploring Transient Mask Inpainting to Enhance Neural Radiance Fields in the Wild[J]. *Neurocomputing*, 2025, 618: 129112.
- [5] LEE J, KIM I, HEO H, et al. Semantic-Aware Occlusion Filtering Neural Radiance Fields in the Wild[J]. *arXiv*: 2303.03966, 2023.
- [6] SNAVELY N, SEITZ S, SZELISKI R. Photo Tourism: Exploring Photo Collections in 3D[J]. *ACM Transactions on Graphics*, 2006, 25(3): 835-846.
- [7] CHAURASIA G, DUCHENE S, SORKINE-HORNUNG O, et al. Depth Synthesis and Local Warps for Plausible Image-Based Navigation[J]. *ACM Transactions on Graphics*, 2013, 32(3): 1-12.
- [8] FITZGIBBON A, WEXLER Y, ZISSERMAN A. Image-Based Rendering Using Image-Based Priors[J]. *International Journal of Computer Vision*, 2005, 63: 141-151.
- [9] LEVOY M, HANRAHAN P. Light Field Rendering [C]// *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, 1996: 31-42.
- [10] CHENG C M, LIN S J, LAI S H, et al. Improved Novel View Synthesis from Depth Image with Large Baseline[C]// *Proceedings of the 19th International Conference on Pattern Recognition*. IEEE, 2008: 1-4.
- [11] DEBEVEC P E, TAYLOR C J, MALIK J. Modeling and Rendering Architecture from Photographs: A Hybrid Geometry-and Image-Based Approach [C]// *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, 1996: 11-20.
- [12] BUEHLER C, BOSSE M, MCMILLAN L, et al. Unstructured Lumigraph Rendering [C]// *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, 2001: 425-432.
- [13] RIEGLER G, KOLTUN V. Free View Synthesis [C]// *Proceedings of the 16th European Conference Computer Vision*, 2020: 623-640.
- [14] FLYNN J, NEULANDER I, PHILBIN J, et al. Deepstereo: Learning to Predict New Views from The World's Imagery [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 5515-5524.
- [15] GROSSMAN J P, DALLY W J. Point Sample Rendering [C]// *Rendering Techniques '98: Proceedings of the Eurographics Workshop in Vienna*, 1998: 181-192.
- [16] ZWICKER M, PFISTER H, VAN BAAR J, et al. Surface Splatting [C]// *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, 2001: 371-378.
- [17] DEBEVEC P, YU Y, BORSHUKOV G. Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping [C]// *Rendering Techniques '98: Proceedings of the Eurographics Workshop in Vienna*, 1998: 105-116.
- [18] SHAN Q, ADAMS R, CURLESS B, et al. The Visual Turing Test for Scene Reconstruction [C]// *Proceedings of the IEEE International Conference on 3D Vision*, 2013: 25-32.
- [19] KUTULAKOS K N, SEITZ S M. A Theory of Shape by Space Carving[J]. *International Journal of Computer Vision*, 2000, 38: 199-218.
- [20] PENNER E, ZHANG L. Soft 3d Reconstruction for View Synthesis[J]. *ACM Transactions on Graphics*, 2017, 36(6): 1-11.
- [21] AKENINE-MOLLER T, HAINES E, HOFFMAN N. *Real-Time Rendering*[M]. AK Peters/crc Press, 2019.
- [22] COOK R L, PORTER T, CARPENTER L. Distributed Ray Tracing [C]// *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*, 1984: 137-145.
- [23] PURCELL T J, BUCK I, MARK W R, et al. Ray Tracing on Programmable Graphics Hardware[J]. *ACM Transactions on Graphics*, 2002, 21(3): 703-712.
- [24] TATARCHENKO M, DOSOVITSKIY A, BROX T. Multi-View 3d Models from Single Images With a Convolutional Network [C]// *Proceedings of the 14th European Conference on Computer Vision*, 2016: 322-337.
- [25] HEDMAN P, PHILIP J, PRICE T, et al. Deep Blending for Free-Viewpoint Image-Based Rendering[J]. *ACM Transactions on Graphics*, 2018, 37(6): 1-15.
- [26] KO J, CHO K, CHOI D, et al. 3d Gan Inversion with Pose Optimization [C]// *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023: 2967-2976.
- [27] LIU L, GU J, LIN K Z, et al. Neural Sparse Voxel Fields [C]// *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020: 15651-15663.
- [28] 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.
- [29] FRIDOVICH-KEIL S, YU A, TANCIK M, et al. Plenoxels: Radiance Fields without Neural Networks [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 5501-5510.

- [30] SUCAR E, LIU S, ORTIZ J, et al. IMAP: Implicit Mapping and Positioning in Real-Time[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021;6229-6238.
- [31] SUN C, SUN M, CHEN H T. Direct Voxel Grid Optimization: Super-Fast Convergence for Radiance Fields Reconstruction [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022;5459-5469.
- [32] CHIBANE J, BANSAL A, LAZOVA V, et al. Stereo Radiance Fields(Srf): Learning View Synthesis for Sparse Views of Novel Scenes[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021;7911-7920.
- [33] CHEN A, XU Z, ZHAO F, et al. MVSNerF: Fast Generalizable Radiance Field Reconstruction from Multi-View Stereo [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021;14124-14133.
- [34] YU A, YE V, TANCIK M, et al. PixelNeRF: Neural Radiance Fields from One or Few Images[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021;4578-4587.
- [35] KIM M, SEO S, HAN B. InfoNeRF: Ray Entropy Minimization for Few-Shot Neural Volume Rendering[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022;12912-12921.
- [36] WANG Z, BAGAUTDINOV T, LOMBARDI S, et al. Learning Compositional Radiance Fields of Dynamic Human Heads[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021;5704-5713.
- [37] WANG X B, ZHANG H, GAO H, et al. Talking Portrait Synthesis Method Based on Regional Saliency and Spatial Feature Extraction [J]. Computer Science, 2025, 52(3): 58-67.
- [38] PARK J J, FLORENCE P, STRAUB J, et al. DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019;165-174.
- [39] MESCHEDER L, OECHSLE M, NIEMEYER M, et al. Occupancy Networks: Learning 3d Reconstruction in Function Space [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019;4460-4470.
- [40] SAITO S, HUANG Z, NATSUME R, et al. Pifu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019;2304-2314.
- [41] SITZMANN V, MARTEL J, BERGMAN A, et al. Implicit Neural Representations with Periodic Activation Functions[C]// Proceedings of the 33th International Conference on Neural Information Processing Systems(NeurIPS). 2020;7462-7473.
- [42] LOMBARDI S, SIMON T, SARAGIH J M, et al. Neural Volumes[J]. ACM Transactions on Graphics, 2019, 38: 1-14.
- [43] BARRON J T, MILDENHALL B, VERBIN D, et al. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022;5470-5479.
- [44] BARRON J T, MILDENHALL B, TANCIK M, et al. Mip-Nerf: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021;5855-5864.
- [45] WU T, TANG S, ZHANG R, et al. CGNet: A Light-Weight Context Guided Network for Semantic Segmentation[J]. IEEE Transactions on Image Processing, 2020, 30: 1169-1179.
- [46] MAO Q, LEE H Y, TSENG H Y, et al. Mode Seeking Generative Adversarial Networks for Diverse Image Synthesis[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019;1429-1437.



CAO Mingwei, born in 1986, Ph.D, associate professor, master supervisor, is a member of CCF (No. 49221M). His main research interests include 3D reconstruction and computer vision.

(责任编辑:柯颖)