DOI:10.20033/j.1003-7241.(2026)08-093-05

一种基于 S3IM 改进的 Aleth-NeRF 网络

何昊宸, 姚 斌, 李 婉

(陕西科技大学 电子信息与人工智能学院,西安 710021)

摘 要:作为一种面向低光照环境的神经辐射场(neural radiance field, NeRF)改进方法,Aleth-NeRF 能够以 RGB 格式低光图像作为输入并获得较好的三维重建与新视角合成效果,但仍存在渲染图像模糊、几何结构细节表达不准确等问题。针对上述不足,提出一种改进的 Aleth-NeRF 网络结构,在渲染分支中引入高斯卷积层对低光照图像及渲染结果进行平滑处理,以抑制噪声并增强特征表达能力;同时设计一种改进型随机结构相似性(stochastic structural similarity, S3IM)损失函数,通过强化结构相似性约束,提升网络对场景几何信息的建模能力。实验结果表明,在多个低光照场景数据集上,所提方法在保持峰值信噪比(peak signal-to-noise ratio, PSNR)基本不变的情况下,有效提升了结构相似性(structural similarity, SSIM)和感知图像相似性(learned perceptual image patch similarity, LPIPS)指标,其中 SSIM 平均提升 0.003,LPIPS 平均降低 0.006。渲染结果在细节表达与几何一致性方面均优于原方法,验证了所提方法的有效性与鲁棒性。

关键词:神经辐射场;低光照成像;三维重建;图像渲染;结构相似性;深度学习

中图分类号: TP391.41;TP391.9 **文献标志码**: A **文章编号**: 1003-7241(2026)08-0093-05

An improved Aleth-NeRF network based on S3IM

He Haochen, Yao Bin, Li Wan

(School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi′an 710021, China)

Abstract: As an improved neural radiance field (NeRF) method for low-light environments, Aleth-NeRF is capable of taking RGB low-light images as input and achieving satisfactory results in 3D reconstruction and novel view synthesis. However, it still suffers from issues such as blurred renderings and inaccurate geometric detail representation. To address these limitations, this paper proposes an improved Aleth-NeRF network architecture. Specifically, a Gaussian convolution layer is introduced into the rendering branch to smooth low-light images and rendered outputs, thereby suppressing noise and enhancing feature representation. In addition, an improved stochastic structural similarity (S3IM) loss function is designed to strengthen structural similarity constraints and improve the network's capability in modeling scene geometry. Experimental results on multiple low-light scene datasets demonstrate that the proposed method effectively improves structural similarity (SSIM) and learned perceptual image patch similarity (LPIPS) while maintaining a comparable peak signal-to-noise ratio (PSNR). Specifically, SSIM is increased by 0.003 on average, and LPIPS is reduced by 0.006. The rendered results exhibit clearer structural details and better geometric consistency compared to the original method, validating the effectiveness and robustness of the proposed approach.

Keywords: neural radiance fields; low-light imaging; 3D reconstruction; novel view synthesis; structural similarity; deep learning

三维图像场景因其能够真实、立体地展示物体和环境,在现代科技和应用中具有重要意义。三维图像场景通过提供详细、真实的视觉信息,广泛应用于医疗、建筑和城市规划、影视和游戏行业、自动驾驶和机器人技术等多个领域。近年来,NeRF[1]因为能够合成高质量、逼真的三维场景图像在计算机图形学和计算机视觉领域引起了极大关注。NeRF 将三维空间坐标和观察方向映射到辐射值,并将场景表示为连续的五维函数。通过在一系列图像和相应的相机姿态上训练神经网络,得到的 NeRF 网络可以渲染场景的新视角,具有逼真的光照、反射和遮挡效果,甚至超过了传统渲染技术(如光栅化或光线追踪)的效果。

然而,NeRF 的构建及其后续的变体网络[2-5]均假设输入的图像处于正常光照条件下,在低光照或过曝光情况下无法正常工作。这是因为 NeRF 是以观察者为中心,它模拟了从重建位置到观察者的光发射量,而没有将照明和材料分离开来。因此,在低光照场景下 NeRF 算法能够得到的三维物体粒子的辐射信息不足,无法对物体材料和几何形状进行估计。在实际应用中,捕获的图像可能存在噪声、低对比度和不足的照明,进一步导致性能下降和重建不准确。

收稿日期:2025-01-06;**录用日期**:2025-02-15

基金项目:国家自然科学基金青年项目(62101312);陕西科技大学科研启动项目(2020BJ-18)

作者简介:何昊宸(1998—),男,硕士研究生,研究方向为计算机图形学与三维重建。

通信作者:姚 斌(1981—),男,博士,副教授,研究方向为计算机视觉与图像特征提取。

引用本文:何昊宸,姚斌,李婉. 一种基于 S3IM 改进的 Aleth-NeRF 网络[J]. 自动化技术与应用,2026,45(8):93-96,179. (He Haochen, Yao Bin, Li Wan. An improved Aleth-NeRF network based on S3IM[J]. Techniques of Automation and Applications,2026,45(8):93-96,179.)

运用图像增强技术[6-7]处理低光图像后再进行 NeRF 建模存在诸多缺点。首先,这些技术会放大图像中的噪声,导致重建结果中的噪声和伪影。其次,色彩失真问题会影响模型对场景的正确理解,导致渲染结果中的色彩偏差。此外,增强过程中可能导致图像细节丢失,使得 NeRF 模型无法捕捉细微特征,影响重建精确性。不同图像片段间的不一致性也会积累,影响场景的整体连贯性和一致性。因此,直接使用这些技术进行 NeRF 建模并不理想,需要专门针对低光照条件进行改进。

RAW-NeRF[8]是一种针对低照度条件下的 NeRF 改进方法。通过在 NeRF 中引入对原始 RAW 格式图像的处理来解决这一问题。RAW 格式图像储存信息丰富却存在难以获得、兼容性差、储存容量高等缺点。针对这一问题,Aleth-NeRF[9]则使用更广泛存在的 RGB 格式图像,通过加入遮蔽场来模拟低光照场景进行 NeRF 网络的训练。在训练结束后去除遮蔽场来渲染得到正常光照下的 NeRF 网络。Aleth-NeRF 也存在一些缺陷,如在一些场景中复原的视角图像存在色差,复原视角的几何边缘存在模糊等。

针对 Aleth-NeRF 存在的上述问题,本文通过改进网络结构,增加高斯卷积网络增强渲染图像的特征表达,减少噪音影响。同时引入由 S3IM 评估模块[10]设计的改进型 S3IM 损失,增强渲染结果。

1 几何增强型 Aleth-NeRF 网络

1.1 原始 Aleth-NeRF 网络结构分析

Aleth-NeRF 与传统 NeRF 方法不同处在于渲染时有无遮蔽场进行参与。Aleth-NeRF 定义遮蔽场为观察视角与观测目标之间的空间存在的阻碍光传播的粒子场,由遮蔽场导致了观测视点的低光照结果。

Aleth-NeRF 建模了两种类型的遮蔽场以减少体积渲染阶段的光传输:局部遮蔽 Ω 表示为在体素级别的遮蔽效果,全局遮蔽场 Ω_θ 表示为在场景级别的遮蔽效果。局部遮蔽场与全局遮蔽场相乘,得到最终的遮蔽场值,从而减少累积透射率。Aleth-NeRF 网络通过输入原始低光照图像作为遮蔽场和低光渲染图像的训练数据。

1.2 改进网络结构

Aleth-NeRF 网络直接将低光照图像作为网络的学习输入,而低光照图像通常包含较多噪声,这些噪声会影响模型的训练过程,使模型难以学习到真实的图像特征,从而在生成图像中出现模糊与几何表达不清。

为了解决上述问题,提出将渲染低光照图像结果与原始低光照图像通过卷积核处理。通过一个 24×24 的高斯卷积层,降低噪音平滑输出结果减少高频噪声,同时平滑低光照图像中的块状伪影,使图像更加干净和平滑。从而能够突显出图像的几何结构信息。经过改进后的网络结构如图 1 所示。

图 1 中,作为输入的空间三维坐标 (x, y, z) 和方位视角 (θ, ϕ) 均提取原始低光照图像,经过位置编码突出高频信息以加强多层感知器(multilayer perceptron, MLP)网络对高频信息的学习。上 MLP 网络为表示网络,输出是一个标量值 σ 表示给定点的体素密度。下 MLP 网络为视角网络,用于预测相机视角下的输出颜色,它接受相机的位置和朝向 (θ, ϕ) 作为输入,并输出一个表示相机视角方向的观测颜色 (r, g, b)。由体素密度 σ 和观测颜色 (r, g, b) 即可组合渲染预测正常光照下的图像。若加入遮蔽场进行渲染,得到的就是低光照渲染图像。

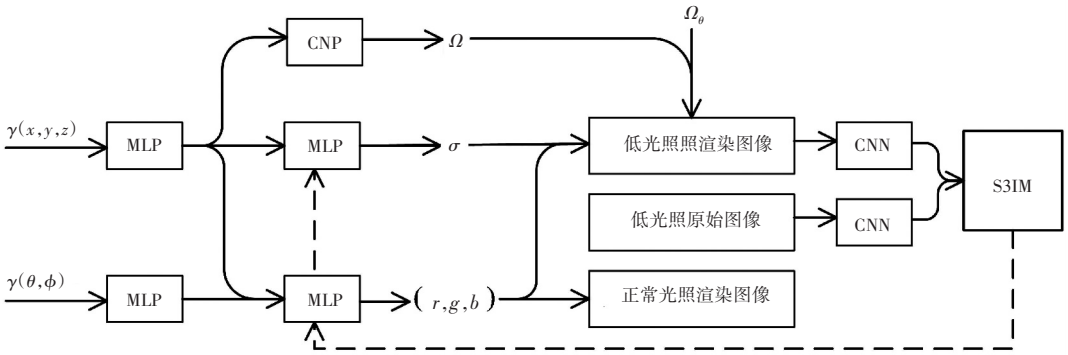


图 1 改进型 Aleth-NeRF 网络结构

Fig. 1 Improved Aleth-NeRF network architecture

低光照渲染图像与低光照原始图像均经过高斯卷积处理过后,作为 S3IM 模块的输入。经过计算 S3IM 损失用于优化表示网络和视角网络对于渲染图像几何信息的表达,而不参与遮蔽场网络的训练。

2 网络损失函数设计

Aleth-NeRF 引入了 3 个无监督损失函数,来促进在低光照条件下改善新视图的合成。分别是:控制总体亮度范围的损失函数 L_{de},控制对比度的损失函数 L_{co} 和色彩损失函数 L_{cc}。

Aleth-NeRF 的总损失函数就由改进过的 MSE 损失和上述 3 个损失函数共同组成如式(1)所示。其中 λ_1、λ_2 和 λ_3 为 3 个非负超参数,取值为 1×10^{-3}、1×10^{-2} 和 1×10^{-8}[9]。

$$L = L_{\text{it-mse}} + \lambda_1 \cdot L_{de} + \lambda_2 \cdot L_{co} + \lambda_3 \cdot L_{cc} \tag{1}$$

分析 Aleth-NeRF 网络的损失函数构成发现,其与

NeRF 网络的损失函数相比增加了 3 个提升色彩表现信息的损失函数。对于图像的几何结构特征没有进行额外的约束,存在优化空间。

2.1 引入 S3IM 损失

S3IM 可显著提升现存 NeRF 类方法的性能指标,其效果已经多个经典 NeRF 方法验证[10]。与传统运用均方误差(mean square error, MSE)的点对点优化相比,S3IM 方法通过随机选取像素点拼接成的像素集进行计算图像随机结构相似性,对重建网络进行全局性结构分析,其计算方法如图 2 所示。通过在原始低光照图像与渲染低光照图像随机的相同位置取像素点,将其拼接为小的像素集。将两个小的像素集计算 SSIM 结构损失作为两幅图像的结构损失。

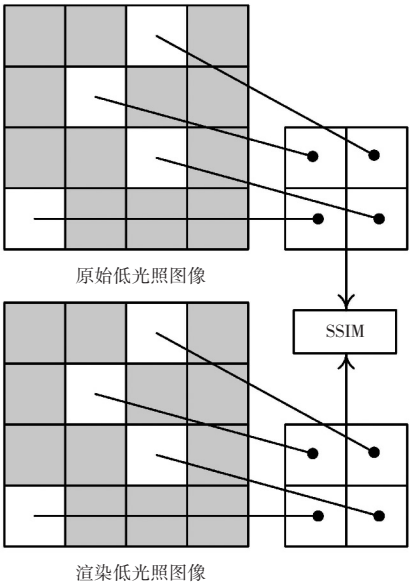


图 2 S3IM 模块演示

Fig. 2 Overview of the S3IM module

但 ALeth-NeRF 网络与传统 NeRF 网络的区别在于,在训练用于合成模拟正常光照图像的 NeRF 网络的同时,也训练了用于模拟低光照环境的遮蔽场网络。得益于 MSE 逐像素计算损失的特性[11],使得 NeRF 网络和遮蔽场网络可以同时通过低光照图像进行训练。而 S3IM 指标注重于全局的结构相似性,在使用低光照图像进行训练时,会导致合成模拟正常光照图像的 NeRF 网络受到过多低光照图像信息的影响[12]。这意味着在 Aleth-NeRF 优化过程中,S3IM 可能无法很好地还原图像的整体对比度信息。具体对比如图 3 所示。



(a) 引入 S3IM (b)原始 Aleth-NeRF

图 3 渲染图像亮度对比

Fig. 3 Comparison of brightness in rendered images.

从渲染结果可以看出,当直接引入 S3IM,渲染图像结果与原方法相比亮度与对比度整体偏低。需要针对 S3IM 损失函数进行改进。

2.2 几何增强型 S3IM 损失

分析 SSIM 评价指标,设有两个离散的非负图像信号 a,b 他们之间的 SSIM 指标计算如式(2)所示。

$$\mathrm{SSIM}(a,b)=\frac{(2\mu_a\mu_b+C_1)(2\sigma_{ab}+C_2)}{(\mu_a^2+\mu_b^2+C_1)(\sigma_a^2+\sigma_b^2+C_2)} \tag{2}$$

可将 SSIM 评价指标分解为亮度,对比度和结构比较度这 3 项组合构成[13]。具体函数表达如下所示。

$$l(a,b)=\frac{2\mu_a\mu_b+C_1}{\mu_a^2+\mu_b^2+C_1} \tag{3}$$

$$c(a,b)=\frac{2\sigma_a\sigma_b+C_2}{\sigma_a^2+\sigma_b^2+C_2} \tag{4}$$

$$s(a,b)=\frac{\sigma_{ab}+C_3}{\sigma_a\sigma_b+C_3} \tag{5}$$

式中,μ_a、μ_b 分别表示图像 a 和 b 的均值,σ_a、σ_b 分别表示图像 a 和 b 的标准差,C_1,C_2 和 C_3 为小常数,通常取 $C_1=(K_1\cdot L)^2$,$C_2=(K_2\cdot L)^2$,$C_3=C_2/2$,$K_1=0.01$,$K_2=0.03$。一般性取 $L=2^B-1$,其中 B 为像素深度。

结合以上原因,为了提高渲染图像的几何结构表达,去除了亮度和对比度分量,从而使模型专注于几何信息的学习,而不是颜色或对比度信息。这有助于在针对低光照图像的 NeRF 建模场景中增强几何信息的学习性能。损失函数定义为 1 减去结构分量的均值,确保其用作最小化目标。同时为了减少由于像素色值带来的亮度与对比度影响,在式中取 $L=1$ 而不是 255。基于 S3IM 修改后的损失函数如式(6)所示。

$$L_{\text{N-S3IM}}=1-\frac{1}{M}\sum_{m=1}^{M}s(\hat{C}^{\text{nor}}(m),C(m))^2 \tag{6}$$

S 函数计算两像素集的结构相似程度,当两图像相似程度越高时,值越趋近于 1。式中,$\hat{C}^{\text{nor}}(m)$,$C(m)$ 分别代表第 m 次由渲染低光照图像和原始低光照图像抽选的像素所拼接的像素集。改进后的 Aleth-NeRF 网络损失函数如式(7)所示,其中 λ 为参数,取值为 1×10^{-3}。

$$L_{\text{imp}}=L+\lambda L_{\text{N-S3IM}} \tag{7}$$

3 实验结果

3.1 实验平台与数据集

本实验环境为 Ubuntu22.04 操作系统,python 版本 3.8,pytorch 版本 1.11.0,cuda 版本 11.3,由一张 Nvidia 4070ti 显卡进行训练。实验数据集为 LOM 数据集,总共包含 5 个场景:“buu”、“chair”、“sofa”、“bike”和“shrub”。每个场景包含 25~65 对低光照与正常光照的图像,其中低光照图像用于网络训练,正常光照图像用于性能评估。各个场景的相机空间位置和视角信息由 COLMAP 计算得出。

3.2 结果分析

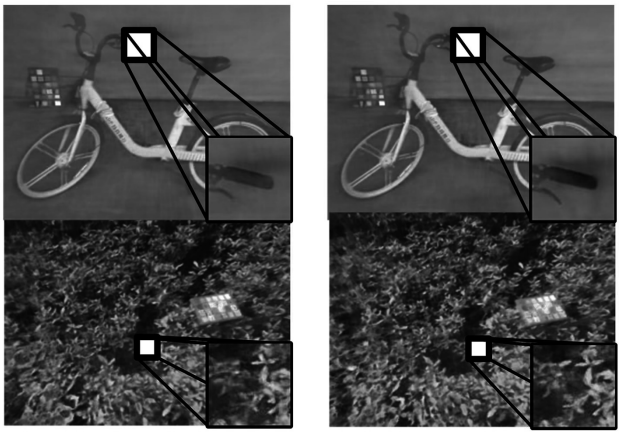
在每个场景对原方法、RetinexNet 增强后的 NeRF 方法和本实验改进型方法对 PSNR、SSIM 和 LPIPS 3 个评价指标进行对比实验。PSNR 为峰值信噪比,主要关注图像的噪声与失真[14];SSIM 考虑了图像的结构信息,能够更好的反应人眼对图像的质量感知;LPIPS 考虑了人眼对于图像感知的敏感性[15],通过计算图像补丁之间的感知距离来评估图像质量。各个方法之间的评价指标结果如表 1 所示。

表 1 各方法渲染正常光照图像与原始图像进行评价的指标对比

Tab. 1 Comparison of evaluation metrics between rendered normal-light images and ground-truth images across different methods

场景	PSNR/SSIM/LPIPS		
	RetinexNet + NeRF	Aleth-NeRF	改进方法
buu	16.521/0.781/0.395	20.303/0.858/0.318	20.249/0.860/0.312
chair	16.946/0.757/0.540	21.102/0.815/0.479	21.109/0.817/0.471
sofa	17.103/0.809/0.576	19.648/0.858/0.357	19.643/0.859/0.355
bike	18.001/0.709/0.480	19.707/0.715/0.547	19.747/0.717/0.543
shrub	14.864/0.286/0.516	18.029/0.489/0.460	18.034/0.497/0.449
平均值	16.687/0.668/0.501	19.757/0.747/0.432	19.756/0.750/0.426

改进网络渲染的正常光照图像与原始 Aleth-NeRF 的渲染结果对比如图 4 所示。



(a) 原始 Aleth-NeRF (b) 改进网络

图 4 渲染图像结构细节对比

Fig. 4 Comparison of structural details in rendered images

观察图 4 渲染结果可以发现,改进方法的渲染结果与原始 Aleth-NeRF 网络相比,对物体的几何表现更加准确,改善了渲染图像中物体模糊与细节表达不清晰的问题。

从表 1 的各指标对比结果可以看出,改进方法在不改变渲染图像的峰值信噪比的情况下,提升了渲染图像的 SSIM 和 LPIPS 指标。图像的结构相似性及感知图像相似性指标上在 5 个训练场景中都有一定程度的提升,尤其是在原 Aleth-NeRF 表现不佳的场景中,改进方法的提升效果更为明显。改进算法将平均 SSIM 指标提升了 0.003,平均 LPIPS 指标降低了 0.006。改进方法在渲染图像的物体几何结构细节表现上要强于原方法,渲染图像的结构表现要更为准确,实验取得了预期效果。

4 结论

本文提出了改进型的 Aleth-NeRF 网络,并引入了改进型 S3IM 损失。在不改变原始渲染图像的峰值信噪比的情况下,有效提高了 Aleth-NeRF 网络渲染图像的结构相似性及感知图像相似性指标,同时显著提升了渲染图像的细节结构表现,在所有场景中都有所提升。但在原始方法已经取得了较好结果的场景中,提升效果较为有限。后续将寻找更优的低光照环境建模,进一步提升图像渲染效果。

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] 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 (ICCV). Montreal, QC, Canada: IEEE, 2021: 5855-5864.

[3] Roessle B, Barron J T, Mildenhall B, et al. Dense depth priors for neural radiance fields from sparse input views[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE, 2022: 12892-12901.

[4] Müller T, Evans A, Schied C, et al. Instant neural graphics primitives with a multiresolution hash encoding[J]. ACM transactions on graphics (TOG), 2022, 41(4): 1-15.

[5] Xu Q, Xu Z, Philip J, et al. Point-NeRF: point-based neural radiance fields[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 5438-5448.

[6] Wu W, Weng J, Zhang P, et al. URetinex-Net: retinex-based deep unfolding network for low-light image enhancement[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 5901-5910.

[7] Guo C, Li C, Guo J, et al. Zero-reference deep curve estimation for low-light image enhancement[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 1780-1789.

[8] Mildenhall B, Hedman P, Martin-Brualla R, et al. NeRF in the dark: High dynamic range view synthesis from noisy raw images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 16190-16199.

(下转第 179 页)