Lighting Every Darkness with 3DGS: Fast Training and Real-Time Rendering for HDR View Synthesis

Xin Jin^{1,2}* Pengyi Jiao¹* Zheng-Peng Duan¹ Xingchao Yang² Chongyi Li¹ Chun-Le Guo^{1†} Bo Ren^{1†} ¹ VCIP, CS, Nankai University ² MEGVII Technology https://srameo.github.io/projects/le3d



 RawNeRF [28]
 RawGS (Baseline)
 LE3D (Ours)
 (c) Focus on Foreground
 (d) Changing WB, Focus on Background

 Figure 1:
 LE3D 从一组多视角、含噪声的 RAW 图像中重建场景的 3DGS 表示。如左图

 所示,与 RawNeRF [28]
 相比,LE3D 具有快速训练和实时渲染的能力。此外,与我们使用

 RawNeRF 策略训练的 RawGS (一个 3DGS [19])
 相比,LE3D 展现出了更强的抗噪性和表示

 所用
 线性色彩的能力。右图展示了 LE3D 可执行的多种实时下游任务,包括 (a) 曝光变化,(b,d) 白平衡(WB)

Abstract

基于体渲染的方法,如 NeRF,在从 RAW 图像合成 HDR 视图方面表现出 色,尤其适用于夜间场景。然而,由于其依赖于密集采样,这类方法训练 时间长,且无法实现实时渲染。3D 高斯投影(3D Gaussian Splatting,简 称 3DGS)的出现,使得实时渲染和更快速的训练成为可能。然而,直接 基于 RAW 图像实现 3DGS 的视图合成仍面临诸多挑战,主要包括以下几 点:1)在夜间场景中,极低的信噪比(SNR)导致远距离视角的运动恢复 结构(SfM)估计效果较差;2)球面调和函数(SH)的表达能力有限,不 适用于 RAW 线性色彩空间;3)场景结构不准确会影响诸如重聚焦等下 游任务的效果。为了解决上述问题,我们提出了 LE3D(使用 3DGS 照 亮每一片黑暗)。我们的方法引入了圆锥散射初始化,以增强 SfM 估计效 果,并用颜色 MLP 取代球面调和函数(SH),用于表示 RAW 线性色彩 空间。此外,我们还提出了深度畸变和近-远端正则化策略,以提升场景 结构的准确性,从而有利于下游任务。这些设计使得 LE3D 能够执行实时 的新视角合成、HDR 渲染、重聚焦以及色调映射调整。与以往基于体渲 染的方法相比,LE3D 将训练时间缩短至原来的 1%,在2K 分辨率下的 渲染速度(FPS)最高提升至原来的 4,000 倍。代码和可视化工具可访问: https://srameo.github.io/projects/le3d。

*Equal Contribution. This project is done during Xin Jin's internship at MEGVII Technology. [†]C. L. Guo and B. Ren are corresponding authors.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

1 介绍

自从神经辐射场(NeRF) [29] 的出现,新视角合成(Novel View Synthesis, NVS)进入了 蓬勃发展的阶段,从而推动了增强现实和虚拟现实(AR/VR)等相关应用的进步。现有的 新视角合成技术主要使用多张曝光良好且无噪声的低动态范围(LDR)RGB 图像作为输入 来重建三维场景。这在低光照或高对比度的环境中(例如夜间或强光反差区域)显著影响了 获取高质量图像的能力,这类场景通常需要使用高动态范围(HDR)场景重建技术。

现有的 HDR 场景重建技术主要分为两类,且均基于体渲染方法:1)使用多曝光的 LDR RGB 图像进行有监督训练 [17];2)直接在带噪的 RAW 数据上进行训练 [28]。第一类方法 在光照充足的场景中通常非常有效,但在夜间场景中,其重建性能受到 sRGB 数据有限动 态范围的影响而受到限制 [4]。而 RAW 数据在夜景中能够保留更多细节,但也更容易受到 噪声影响。因此,RawNeRF [28] 被提出用于解决 vanila NeRF 缺乏抗噪能力的问题。然 而,RawNeRF 仍面临训练时间长以及无法实时渲染的问题(这是体渲染方法的一个普遍缺 陷)。这极大地限制了当前场景重建技术和 HDR 视图合成的应用。实现 HDR 视图合成的 实时渲染,是将计算摄影引入三维世界的重要一步。

近年来, 3D Gaussian Splatting (3DGS)由于其在实时渲染和高保真重建方面的强大能力 而受到广泛关注。3DGS 使用运动恢复结构 (SfM) [31]进行初始化,并使用一组 3D 高斯 原语表示场景。每个高斯使用球谐函数 (SH)表示方向相关的颜色,并可通过梯度下降优 化颜色、位置、尺度、旋转和不透明度。尽管 3DGS 展现出出色的重建和实时渲染能力,但 它并不适合直接使用夜间 RAW 数据进行训练。这主要由于以下三点:1)基于夜间图像估 计的 SfM 通常不准确,导致远景模糊或可能出现大量漂浮点;2) SH 无法充分表达 RAW 图像中的 HDR 颜色信息,表示能力受限;3) RAW 数据中的噪声会影响 3DGS 的重建质 量,生成额外的漂浮点。最终重建出的结构(例如深度图)表现不佳,导致诸如焦点重建等 下游任务效果不理想。

为使 3DGS 适用于 HDR 场景重建,我们提出了 LE3D,意为使用 **3D**-GS 照亮每一片黑暗, 用以解决上述三类问题。首先,为了解决低光照场景下 SfM 远景估计不准确的问题,我们 提出了圆锥散射初始化用于丰富由 COLMAP 估计的 SfM 结果。该方法在一个由已估计相 机位姿定位的锥形区域内进行随机撒点。其次,我们用一个小型 MLP 替代 SH,在 RAW 线性色彩空间中表示颜色。为了更好地初始化每个高斯的颜色,我们为不同的高斯原语引入 不同的颜色偏置。第三,为了防止噪声在场景中生成额外的漂浮点并增强远景的重建质量, 我们提出了两个正则项:深度畸变损失和远近正则项,以实现从粗到细的结构学习。如图 1 (左)所示,我们的 LE3D 仅需约 1.5 小时 GPU 训练时间(比 RawNeRF [28] 节省 99%), 即可实现 100 FPS 的实时渲染(约比 RawNeRF 快 4000 倍),且在质量上可与之媲美。此 外,如图 1 (右)所示,LE3D 还支持 HDR 渲染、焦点重建、曝光变换等下游任务。

总之,我们的贡献如下:

- 我们提出了 LE3D, 能够从带噪的 RAW 图像中重建 HDR 三维场景并实现实时渲染。相比于基于 NeRF 的方法, LE3D 仅需 1% 的训练时间, 渲染速度提升达 4000 倍。
- 为解决原始 3DGS 在颜色表示方面的缺陷及 SfM 在夜间估计下的不足,我们采用 了具备原语感知偏置的颜色 MLP,并引入圆锥散射初始化来丰富由 COLMAP 初 始化的点云。
- 为增强最终重建结果的结构信息,以提升下游任务的表现(如焦点重建),我们引入 了深度畸变损失和远近正则项来对结构进行监督。

2 相关工作

新视角合成 自从 NeRF [29] 出现以来,新视角合成取得了显著的进展。NeRF 使用一个 MLP 同时表示场景的几何结构和视角相关的颜色。它利用可微分的体积渲染方法 [22],从 而可以通过多视角的二维图像集合进行梯度下降训练。NeRF 的一系列后续变体 [1, 2, 16] 扩展了其抗锯齿能力。为了克服原始 NeRF 在几何结构重建方面的不足,引入了如深度监督 [8, 10] 和失真损失 [2] 等策略。一些方法 [7, 34] 还将 NeRF 应用于低光图像增强等新场 景中。另一些工作 [3, 5, 12, 30] 探索了基于特征网格的方法以提升 NeRF 的训练与渲染速度。尽管这些方法在新视角合成方面取得了较好的效果,但由于体积渲染天然需要密集采样,其训练和渲染速度仍然是一大瓶颈。

近期,三维 Gaussian Splatting (3DGS) [19] 的出现标志着实时新视角合成方法的一次重 大突破。3DGS 使用一组具有不同属性的三维高斯基元表示场景。一些后续工作为高斯表 示增加了抗锯齿能力 [25, 32, 37];还有一些工作通过频域监督增强了高斯的表征能力 [38]。 DNGaussian [23] 提出了一种基于深度正则的框架来优化稀疏视角下的 3DGS,其他一些工 作也采用了深度监督策略 [6, 21]。此外,一些研究 [24, 27, 35] 致力于将 3DGS 应用于动态 场景的表示。然而,这些方法通常只能接收 LDR sRGB 图像作为输入,因此无法重建场景 的 HDR 辐射场,这意味着它们无法完成如 HDR 色调映射、曝光变化等下游任务。相比之 下,LE3D 专为从噪声 RAW 图像中重建场景的 HDR 表示而设计。

HDR 视角合成及其应用 HDR 通常指计算摄影中的一个概念,其核心在于尽可能保留图像的动态范围,以支持更多的后处理操作 [9,11,15,18,26]。现有的 HDR 视角合成技术路 径与传统 2D 图像的 HDR 合成方法有高度相似之处:

1) 直接使用多曝光的 LDR 图像来计算相机响应函数(CRF)并合成 HDR 图像 [9]。该方 法对应 HDR-NeRF [17],它使用 MLP 来学习 CRF;

2) 获取无噪声的欠曝光 RAW 图像,利用 RAW 的线性色彩空间特性,手动模拟多曝光图 像并合成 HDR 图像。这对应 RawNeRF [28],其从带噪 RAW 图像中学习 RAW 色彩空间 下的 NeRF 表示,完成去噪与新视角合成。VR-NeRF [36] 进一步引入感知色彩空间以优化 高动态范围图像的感知质量。

尽管上述方法在视觉效果上表现良好,但由于体积渲染所需的密集采样,它们的训练时长 和渲染效率仍存在明显瓶颈。

LE3D 遵循与 RawNeRF 相同的技术路线,即从噪声 RAW 图像中重建场景表示。这意味着 LE3D 不再强依赖多曝光训练数据,极大地扩展了其适用范围。但其关键区别在于使用 了可微光栅化技术 [19, 20, 33],从而实现了快速训练与实时渲染。

基于重建出的 RAW (线性 RGB) 下的 3D 表征, LE3D 能够实时执行 HDR 视角合成。这也代表着一次将计算摄影引入三维世界的尝试:它使得场景的实时重构与后处理(如改变 曝光、白平衡、HDR 渲染等)成为可能,如图 1 所示。

3 预备知识

3D 高斯点云渲染 3D 高斯点云渲染通过融合多个三维高斯基元来计算像素的颜色和深度, 从而渲染出细致的场景。每个高斯基元由其在三维空间中的中心位置 $\mu_i \in \mathbb{R}^3$,缩放因子 $s_i \in \mathbb{R}^3$,旋转四元数 $q_i \in \mathbb{R}^4$ 以及附加属性如不透明度 o_i 和颜色特征 f_i 所定义。高斯基元的基函数如公式 (1) 所示,其中协方差矩阵由缩放和旋转参数计算得到。

$$G(x) = \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right).$$
 (1)

渲染时,像素颜色由覆盖该像素位置的多个高斯基元贡献的颜色混合得到。该过程包含通 过球面调和函数(SH)将颜色特征 *f_i* 解码为颜色 *c_i*,并通过将不透明度 *o_i* 与其在图像平 面上的二维投影高斯 *G^{2D}* 相乘来计算每个基元的 *α_i*。不同于传统的光线采样策略,3D 高 斯点云渲染采用优化的光栅化器来收集相关的高斯基元。具体地,像素颜色 *C* 通过对覆盖 该像素的 *N* 个有序高斯基元进行混合计算:

$$C = \sum_{i \in N} c_i \cdot \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j), \quad \nexists \psi \quad \alpha_i = G_i^{2D} o_i.$$
⁽²⁾

基于带噪 RAW 图像的 HDR 视图合成 RawNeRF [28] 是 NeRF 的一个强力扩展,专门针对带噪声图像下的高动态范围(HDR)视图合成问题。与低动态范围(LDR)图像不同,HDR 图像中亮暗区域的动态范围可以跨越多个数量级,导致 NeRF 传统的 L2 损失函数无法很好地拟合。为此,RawNeRF 引入了一种加权的 L2 损失,以增强对暗区的监督。RawNeRF 对色调曲线 $\psi = \log(y + \epsilon)$ (其中 $\epsilon = 10^{-3}$)施加梯度监督,并利用加权项 $\psi' = (y + \epsilon)^{-1}$ 对渲染颜色 $\hat{y}i$ 与带噪参考颜色 y_i 之间的 L2 损失进行加权。通过对 y 施加停止梯度操作 $sg(\cdot)$,最终损失函数表达式为:

$$L\psi(\hat{y}, y) = \sum_{i} \left(\frac{\hat{y}_i - y_i}{sg(\hat{y}_i) + \epsilon}\right)^2.$$
(3)



Figure 2: 我们提出的 LE3D 流程。1) 使用 COLMAP 获得初始点云和相机位姿。2) 使用 圆锥散射初始化来丰富远景点云并初始化颜色。3) 进行标准的 3DGS 训练,使用轻量级颜 色 MLP 代替原本用于颜色表示的 SH。4) 使用 RawNeRF 的加权 L2 损失 \mathcal{L} 作为图像级监督,同时使用我们提出的深度畸变损失 \mathcal{R}_{dist} 和远近正则化 \mathcal{R}_{nf} 来约束每条光线上的高斯分布更加集中。在此过程中, f_i 、 b_i 和 c_i 分别表示第 i 个高斯的颜色特征、偏置和最终渲染颜色; o_i 、 r_i 、 s_i 和 p_i 分别表示其不透明度、旋转、尺度和位置。

此外, RawNeRF 采用了多曝光训练策略,利用不同快门速度的图像。该方法通过为每个快门速度 t_i 和每个颜色通道 c 学习缩放因子 $\beta_{t_i}^c$,在线性 RGB 空间对输出颜色进行缩放。具体而言,第 c 个通道的输出颜色 \hat{y}_i^c 会映射为 min($\hat{y}_i^c \cdot t_i \cdot \beta t_i^c$,1) 作为最终渲染输出。

4 提出的方法

我们的 LE3D 流程如图 2 所示。

4.1 对基础 3DGS 表征的改进

为增强 3DGS 在远景和 RAW 线性色彩空间上的表现力,我们提出了圆锥散射初始化和轻量级颜色 MLP。

圆锥散射初始化 为了丰富 COLMAP 初始化的点云 $S = \{s_i\}$ 中的远景点,我们估计所有 相机的位置和朝向,基于此,在预定义的视锥 F 内均匀撒点。视锥 F 由以下参数定义:1) 视点 \mathbf{p} ; 2) 观察方向 $\vec{\mathbf{n}}$; 3) 视场角 Θ ; 4) 近平面和远平面 z_n, z_f 。对于前向场景,观察方 向取所有相机朝向的平均值 $\vec{\mathbf{n}} = \operatorname{avg}\{\vec{\mathbf{n}}_i^c\},$ 视场角取所有相机视场角最大值 $\Theta = \max\{\theta_i^c\}$ 。 视锥需包含所有相机原点 $\{\mathbf{p}_i^c\}$,即包含以中心 $\bar{\mathbf{p}}^c = \operatorname{avg}\{\mathbf{p}_i^c\}$ 和半径 $r = \max\{\|\mathbf{p}_i^c - \bar{\mathbf{p}}^c\|_2\}$, 且垂直于 $\vec{\mathbf{n}}$ 的圆。视锥原点计算如下:

$$\mathbf{p} = \overline{\mathbf{p}}^{c} - \frac{r}{\tan(\Theta/2)} \cdot \frac{\mathbf{n}}{\|\mathbf{n}\|_{2}}, \quad \mathbf{n} = \operatorname{avg}\{\mathbf{n}_{i}^{c}\}, \quad \Theta = \max\{\theta_{i}^{c}\},$$

$$z_{n} = \min\{\|\mathbf{s}_{i} - \mathbf{p}\|_{2}\}, \quad z_{f} = \lambda_{\mathcal{F}} \cdot \max\{\|\mathbf{s}_{i} - \mathbf{p}\|_{2}\}.$$
(4)

其中 z_n 和 z_f 分别取 COLMAP 点云中到 **p** 最近点和最远点距离的 λ_F 倍。然后在视锥 $\mathcal{F} = \{\mathbf{p}, \mathbf{n}, \Theta, z_n, z_f\}$ 内均匀撒点,得到扩充后的点云 $\mathcal{S}' = \mathcal{S} \cup \mathcal{S}^F$, \mathcal{S}^F 为撒点集合。用 \mathcal{S}' 初始化高斯。

带有原始偏置的颜色 MLP 为解决 SH 难以表示 RAW 线性色彩空间问题,使用 tiny color MLP \mathbf{F}_{θ} 替代 SH。每个高斯初始化颜色偏置 b_i ,通过将点云 $\mathbf{s}'_i \in \mathcal{S}'$ 投影到训练视角上,取 投影像素颜色集合 $\{c_{\text{pix}}\}_i$ 的均值作为初始颜色 $c_i^{\mathcal{F}} = \operatorname{avg}(\{c_{\text{pix}}\}_i), \Leftrightarrow b_i^{(0)} = \log(c_i^{\mathcal{F}})$ 。颜色 特征 $f_i^{(0)}$ 从高斯分布 $\mathcal{N}(0, \sigma_f)$ 采样。将 f_i 和相机位姿 v 拼接后输入 \mathbf{F}_{θ} ,采用指数激活函 数模拟无上限 HDR 颜色空间,输出视角依赖颜色,最终颜色为:

$$c_i = \exp\left(\mathbf{F}_{\theta}(f_i, v) + b_i\right). \tag{5}$$

其中 fi 和 bi 是可学习参数, 克隆与分裂时复制给新高斯。

4.2 深度畸变与远近正则化

场景结构对我们框架的下游应用至关重要,尤其是如重聚焦等任务。因此,我们提出了深度畸变和远近正则化,以增强 3D 高斯表示对场景结构的优化能力。借鉴基于 NeRF 的方法 [2],我们利用深度图和权重图对场景结构进行正则化。

深度图和权重图的渲染 近期一些基于 3D 高斯的工作 [23, 6] 采用了对深度的某种监督。 同时,深度图对于重聚焦(见第6节)、机械提取 [14] 和重光照 [13, 39] 等下游任务也十分 重要。深度图 *d* 通过以下方式获得:

$$d = \frac{\sum_{i} z_{i}^{c} \omega_{i}}{\sum_{i} \omega_{i}}, \quad \not \pm \not \oplus \quad [x_{i}^{c}, y_{i}^{c}, z_{i}^{c}]^{T} = W[x_{i}, y_{i}, z_{i}]^{T} + t, \quad \omega_{i} = \alpha_{i} \prod_{j=1}^{i-1} (1 - \alpha_{j}). \tag{6}$$

其中 d 表示深度图, ω_i 为第 i 个高斯的混合权重, $[x_i, y_i, z_i]^T$ 和 $[x_i^c, y_i^c, z_i^c]^T$ 分别为世界坐 标系和相机坐标系下的位置, [W, t] 为相机外参。

权重图中每个像素值描述了经过该像素射线上的分布直方图 \mathcal{H} 。类似于 Mip-NeRF 360 [2], 我们可以通过限制每条射线上高斯基元的集中程度来增强对场景结构的学习。为获得权重 图,首先确定当前相机位置 p^{c} 到最近和最远高斯基元的距离,记为 z_{n}^{c} , z_{f}^{c} 。然后将区间 $[z_{n}^{c}$, $z_{f}^{c})$ 划分为 K 个子区间,第 k 个区间表示为 $[t_{k}, t_{k+1})$ 。因此,直方图第 k 个值 $\mathcal{H}(k)$ 通过渲染如下获得:

$$\mathcal{H}(k) = \sum_{i} \mathbb{1}(z_{i}^{c}, k)\omega_{i}, \quad \mathbb{1}(z_{i}^{c}, k) = \begin{cases} 1 & \text{in} \mathbb{R}z_{i}^{c} \in [t_{k}, t_{k+1}) \\ 0 & \text{and} \end{cases}$$
(7)

渲染 H 十分必要,不仅在正则化中有效,也应用于重聚焦任务。

提出的正则化项 受 Mip-NeRF 360 [2] 启发,我们提出类似的深度畸变正则化 *R*_{dist},用于增强每条射线上高斯基元的集中程度:

$$\mathcal{R}_{dist} = \sum_{u,v}^{K} \mathcal{H}(u) \mathcal{H}(v) \left| \frac{t_u + t_{u+1}}{2} - \frac{t_v + t_{v+1}}{2} \right|.$$
(8)

 \mathcal{R}_{dist} 限制整条射线上的权重要么趋近于零,要么集中在同一个区间。然而,在真实世界的无界场景中,每个区间的距离 $(z_f^c - z_n^c)/K$ 非常大,强行增大 K 来缩短区间长度会增加计算负担。因此, \mathcal{R}_{dist} 只能对每条射线上的高斯基元施加较粗略的监督,即尽可能限制它们集中在同一区间。

为了进一步约束高斯基元的集中程度,我们提出了远近正则化 \mathcal{R}_{nf} 。 \mathcal{R}_{nf} 通过限制每条射 线上最近 M 个和最远 M 个高斯基元的加权深度距离,进一步强化对结构的学习,其中最 远是指在混合权重趋近于 1 时的最后几个高斯。首先提取两个高斯子集 N 和 F,分别包含 每条射线上的最近和最远的 M 个高斯基元。随后渲染它们的深度图 d^{N} , d^{F} 及最终混合权 重图 T^{N} , T^{F} ,其中 $T = \sum_{i} \omega_{i}$ 。定义远近正则化为:

$$\mathcal{R}_{nf} = T^{\mathbf{N}} \cdot T^{\mathbf{F}} \cdot \left| d^{\mathbf{N}} - d^{\mathbf{F}} \right|.$$
(9)

(10)

当两端深度差距较大时, \mathcal{R}_{nf} 通过不透明度的监督抑制射线前后多余的高斯 (依赖于 $T^{\mathbf{N}} \cdot T^{\mathbf{F}}$ 项)。相比 \mathcal{R}_{dist} , \mathcal{R}_{nf} 还能在位置上监督每条射线首尾高斯尽量接近 (依赖于 $|d^{\mathbf{N}} - d^{\mathbf{F}}|$ 项)。因此, \mathcal{R}_{nf} 相较于 \mathcal{R}_{dist} 提供了更细粒度的约束。

除加权 L2 损失 \mathcal{L} 及上述正则化外,我们还对最终混合权重 T 施加约束。鉴于 LE3D 在 真实场景测试,T 应趋近于 1,即所有像素点都应被渲染。因此,我们引入正则项 $\mathcal{R}_T = -\log(T + \epsilon)$,惩罚 T 小于 1 的情况。

5 实验

5.1 实现细节

损失函数与正则化 在我们的实现中,最终的损失函数为:

$$L = \mathcal{L} + \lambda_T \mathcal{R}_T + \lambda_{dist} \mathcal{R}_{dist} + \lambda_{nf} \mathcal{R}_{nf},$$

其中, \mathcal{L} 是加权的 L2 损失, \mathcal{R}_T 、 \mathcal{R}_{dist} 和 \mathcal{R}_{nf} 分别为提出的 T 正则化、深度畸变正则化 和远近正则化。

Table 1: RawNeRF [28] 测试场景的定量结果。最佳结果用粗体表示,次佳用下划线表示。 TM 表示是否支持替换色调映射函数进行 HDR 渲染。支持替换色调映射的方法在 sRGB 上的指标采用 LDR 色调映射以保证公平比较。FPS 测量在 2K 分辨率 (2016×1512)下 进行。Train 表示训练时间,单位为 GPU×小时。LE3D 在保持与先前基于体渲染方法 (RawNeRF [28])相当性能的同时,渲染速度提升了 4000 倍。

方法	тм	FPS↑	训练时间上	RAW		sRGB		
13 14		1101	6480611.1 Y	$\mathrm{PSNR}\uparrow$	$\mathrm{SSIM}\uparrow$	$\mathrm{PSNR}\uparrow$	$\mathrm{SSIM}\uparrow$	$\mathrm{LPIPS}{\downarrow}$
LDR-NeRF [29]	X	0.007	13.66	_	_	20.0391	0.5541	0.5669
LDR-3DGS $[19]$	X	153	0.75	_	_	20.2936	0.5477	0.5344
HDR-3DGS [19]	1	238	0.73	56.4960	0.9926	20.3320	0.5286	0.6563
RawNeRF [28]	1	0.022	129.54	58.6920	0.9969	$\underline{24.0836}$	0.6100	0.4952
RawGS (基线)	1	176	1.05	59.2834	0.9971	23.3485	0.5843	0.5472
LE3D (本方法)	\checkmark	103	1.53	61.0812	0.9983	24.6984	0.6076	0.5071

优化 我们将 λ_F 设置为 10, 以在远景视图中丰富 COLMAP 初始化的点云。损失函数中的 λ_T , λ_{dist} , λ_{nf} 分别设置为 0.01, 0.1, 0.01。对于我们的颜色 MLP **F**_θ, 使用 Adam 优化器, 初 始学习率为 1.0e – 4。每个高斯的颜色特征和偏置的初始学习率分别为 2.0e – 3 和 1.0e – 4。 三者的学习率均按照余弦衰减策略降低至最终学习率 1.0e – 5。除颜色 MLP、基于图元的 颜色偏置和每个高斯的颜色特征外,其他设置与 3DGS [19] 相同。对于包含多重曝光的场 景,我们采用与 RawNeRF [28] 相同的多曝光训练策略。

5.2 数据集与对比

数据集 我们在 RawNeRF 的基准数据集上评估了 LE3D 的性能。该数据集包含 14 个场 景用于定性测试, 3 个测试场景用于定量测试。三个测试场景包含 101 张带噪声的图像和一 张由稳定长曝光合成的干净参考图像。所有训练数据均为短曝光图像,存在曝光不一致现 象,因此我们在测试前采用了与 RawNeRF 相同的仿射对齐操作(详见补充材料)。所有图 像均为 iPhone X 采集的 4032 × 3024 Bayer RAW 图像,保存为 dng 格式。此外,我们还 额外收集了多个真实场景用于定性测试,详情见补充材料。

基线与比较方法 我们比较了两类方法, 3DGS 基方法和 NeRF 基方法。我们对比的基线 为 RawGS,即使用原始 3DGS 作为场景表征,采用 RawNeRF [28] 中提出的加权 L2 损失 和多重曝光训练。此外,我们移除了 SSIM 损失,因为其基于局部邻域的计算会破坏 RAW 数据中的噪声模型。同时,我们对比了 LDR-GS 和 HDR-GS,二者均为原始 3DGS,分别 在后处理的 LDR 图像和未处理的 RAW 图像上训练。NeRF 基方法包括 RawNeRF [28] 和 LDR-NeRF。RawNeRF 是直接在带噪 RAW 图像上用加权 L2 损失和多曝光训练策略训练 的 Mip-NeRF [1]。LDR-NeRF 是基于后处理 LDR 图像,用 L2 损失训练的原始 NeRF [29]。

定量评估 表 1 展示了 RawNeRF [28] 数据集上的定量对比。尽管 NeRF 基方法训练时间 长且渲染慢,但其在 sRGB 上的指标表现良好,这表明其依赖的体渲染对噪声具有较强鲁 棒性(主要由于每条光线的密集采样)。而 3DGS 基方法指标不及 RawNeRF,原因在于其 稀疏的场景表示及较差的抗噪声能力。且高斯基元的分裂依赖梯度强度,使用带噪 RAW 图 像监督会影响其效果,导致结构恢复不完整。而 LE3D 通过结构监督、深度畸变和远近正则 化,实现了更好的结构重建,更适合下游任务,详见第6节。值得注意的是,LE3D 在定量 和定性指标上与 RawNeRF [28] 相当,但训练时间仅为其 1%,渲染速度提升了 3000~6000 倍。

定性评估 图 3 展示了在 RawNeRF [28] 数据集上的定性比较。我们选取了四个场景进行 对比,包括两个室内场景和两个室外场景。前两个场景的数据采集为单一曝光,而后两个场 景包含多重曝光的数据。相比基于 3DGS [19] 的方法,LE3D 表现出更强的抗噪能力,特别 是在前两个场景中。此外,LE3D 在远景重建方面也取得了更好的效果。例如,在第二个场 景中,LE3D 的天空比 RawGS 更加平滑;在第四个场景中,LE3D 恢复的远景细节更加清 断。与 RawNeRF 相比,LE3D 通常能重建出更平滑的效果,同时细节保留也表现良好。最 重要的是,LE3D 拥有更快的训练时间和渲染速度。



Figure 3: LE3D 与其他重建方法的视觉对比(放大观看效果最佳)。训练视角包含线性提亮 后的后处理 RAW 图像(上)和设备直出 JPEG 图像(下)。与 3DGS 基方法相比, LE3D 在 远景恢复中更清晰且抗噪能力更强。与 NeRF 基方法相比, LE3D 在 效果相当的情况下 渲染速度提升 3000~6000 倍。



Training View (a) LE3D (b) w/o CSI (c) w/o Color MLP (d) w/o Regs (e) Results at 7k iter Figure 4: 我们方法的消融实验(最佳观察请赦大)。(b) 中的 CSI 和 (d) 中的 Regs 分别代 表锥形散射初始化和正则化。(e) 展示了训练初期 LE3D 有无颜色 MLP 的渲染结果。

5.3 消融实验

锥形散射初始化(CSI) 在低光环境下,COLMAP 难以获得高质量的稀疏点云。尽管 3DGS 对初始点云质量具有一定鲁棒性,但在初始化区域不足时仍难以实现最佳几何重建。 从图 4 (b) 可见,未使用 CSI 的方法往往在错误深度生成高斯且缺乏细节。而 CSI 扩展了 场景的深度覆盖,使 3DGS 能在较准确的深度位置生成高斯,并表现出更优细节。对比图 4 (a) 和 (b) 表明我们的初始化方法对于实现准确且细致的 3D 重建至关重要。

颜色 MLP 用颜色 MLP 替代 SH 不仅增强了模型的表达能力,还提升了优化过程的稳定 性。图 4 (e) 显示,采用 SH 而非颜色 MLP 的方法在训练初期会出现异常的颜色表现,这 是由于 SH 无法充分表示 RAW 线性颜色空间。虽然渲染结果可能与 LE3D 相似,但该问 题严重影响了最终的结构重建效果,如图 4 (c) 所示。

正则化 优异的三维视觉效果依赖稳健的三维结构重建,这也显著提升了如重聚焦等下游 任务的性能。为此,我们引入深度畸变正则化 *R*_{dist} 和近远正则化 *R*_{nf},约束高斯分布,促 使其聚集于物体表面,从而提升结构重建质量。图 4 (d) 突出了我们提出的正则化方法在场 景三维结构重建上的显著提升。

6 更多应用

重聚焦 结构信息对于重聚焦等任务至关重要。如第 5.3 节所述, LE3D 通过引入深度畸变 和近远正则化,提升了结构细节的学习能力。如图 5 (b, d) 所示, LE3D 凭借更优的结构信





(f) Exposure Variation

(g) Changing Tone-mapping, Viewpoint, and Focus

Figure 5: LE3D 支持多种应用。(d) 中的 RawGS* 表示使用 LE3D 渲染的图像和 RawGS 的结构信息作为重聚焦的输入。(c, e) 分别是 LE3D 和 RawGS 渲染的加权深度图。(f) 展 示了同一场景在不同曝光设置下由 LE3D 渲染的效果。(g) 中,"→"代表全局色调映射,而 "→"表示局部色调映射。

息(在(c)中的深度图得以体现),实现了更加真实的重聚焦效果。而 RawGS 由于结构信 息缺失,在重聚焦时会出现前景与背景的模糊歧义。

详细的重聚焦算法将在补充材料中发布。

曝光变化与 HDR 色调映射 得益于 RAW 数据天然的高动态范围及线性色彩空间的特性, LE3D 能够轻松实现曝光变化,并从过曝数据中恢复细节,如图 5 (f) 所示。图 5 (g) 展示 了 LE3D 可实现的多种色调映射方法,包括色温和曲线调节等全局色调映射,以及我们重 新实现的 HDR+ [15] 局部色调映射。相关实现细节将在补充材料中发布。

综上所述,结合**实时**新视角合成、曝光变化和多种色调映射功能,LE3D 实现了丰富的应用 场景。尽管 RawNeRF [28] 也支持类似应用,但由于其无法实现实时渲染,极大限制了其实 时编辑等应用场景(详见补充材料)。

7 结论

现有的 HDR 视图合成方法大多依赖体积渲染技术,其训练和渲染时间较长,限制了应用范 围。为了解决以往基于体积渲染方法训练时间长且渲染速度慢的问题,我们基于 3DGS 提 出了 LE3D。同时,我们提出了锥形散射初始化 (Cone Scatter Initialization) 和一个用于 线性色彩空间表示颜色的小型 MLP。这解决了使用 COLMAP 初始化时夜间场景远景点缺 失的问题,并且用小型颜色 MLP 替代了球谐函数,更有效地表示了线性色彩空间。最后, 我们引人了深度畸变和近远正则化来强化结构信息,从而更有效、更真实地支持下游的重 聚焦等任务。得益于 LE3D 渲染图像在线性色彩空间,它能够实时实现更加真实的曝光变 化和 HDR 色调映射,丰富了 HDR 视图合成在后续处理中的更多可能性。

致谢

本工作得到国家自然科学基金(编号 62225604, 62272245, 62306153)、中央高校基本科研 业务费(南开大学,编号 63243147)以及天津市科技项目(22JCYBJC01270)的资助。本 工作所用计算设备由南开大学超级计算中心(NKSC)提供支持。

References

[1] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In ICCV, 2021.

- [2] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In CVPR, 2022.
- [3] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In ECCV, 2022.
- [4] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In CVPR, 2018.
- [5] Zhang Chen, Zhong Li, Liangchen Song, Lele Chen, Jingyi Yu, Junsong Yuan, and Yi Xu. Neurbf: A neural fields representation with adaptive radial basis functions. In *ICCV*, 2023.
- [6] Jaeyoung Chung, Jeongtaek Oh, and Kyoung Mu Lee. Depth-regularized optimization for 3d gaussian splatting in few-shot images. arXiv:2311.13398, 2023.
- [7] Ziteng Cui, Lin Gu, Xiao Sun, Xianzheng Ma, Yu Qiao, and Tatsuya Harada. Aleth-nerf: Illumination adaptive nerf with concealing field assumption. In AAAI, 2024.
- [8] David Dadon, Ohad Fried, and Yacov Hel-Or. Ddnerf: Depth distribution neural radiance fields. In *WACV*, 2023.
- [9] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. *Siggraph*, 1997.
- [10] Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. Depth-supervised nerf: Fewer views and faster training for free. In CVPR, 2022.
- [11] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K Mantiuk, and Jonas Unger. Hdr image reconstruction from a single exposure using deep cnns. ACM TOG, 2017.
- [12] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In CVPR, 2022.
- [13] Jian Gao, Chun Gu, Youtian Lin, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3d gaussian: Real-time point cloud relighting with brdf decomposition and ray tracing. arXiv:2311.16043, 2023.
- [14] Antoine Guédon and Vincent Lepetit. Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. *CVPR*, 2024.
- [15] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. ACM TOG, 2016.
- [16] Wenbo Hu, Yuling Wang, Lin Ma, Bangbang Yang, Lin Gao, Xiao Liu, and Yuewen Ma. Tri-miprf: Tri-mip representation for efficient anti-aliasing neural radiance fields. In *ICCV*, 2023.
- [17] Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang. Hdr-nerf: High dynamic range neural radiance fields. In CVPR, 2022.
- [18] Nima Khademi Kalantari, Ravi Ramamoorthi, et al. Deep high dynamic range imaging of dynamic scenes. ACM TOG, 2017.
- [19] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM TOG*, 2023.
- [20] Georgios Kopanas, Julien Philip, Thomas Leimkühler, and George Drettakis. Point-based neural rendering with per-view optimization. In *Computer Graphics Forum*, 2021.
- [21] Pou-Chun Kung, Seth Isaacson, Ram Vasudevan, and Katherine A Skinner. Sad-gs: Shapealigned depth-supervised gaussian splatting. In *ICRA Workshops*, 2024.
- [22] Marc Levoy. Efficient ray tracing of volume data. ACM TOG, 1990.
- [23] Jiahe Li, Jiawei Zhang, Xiao Bai, Jin Zheng, Xin Ning, Jun Zhou, and Lin Gu. Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. arXiv:2403.06912, 2024.

- [24] Zhan Li, Zhang Chen, Zhong Li, and Yi Xu. Spacetime gaussian feature splatting for real-time dynamic view synthesis. CVPR, 2024.
- [25] Zhihao Liang, Qi Zhang, Wenbo Hu, Ying Feng, Lei Zhu, and Kui Jia. Analytic-splatting: Anti-aliased 3d gaussian splatting via analytic integration. arXiv:2403.11056, 2024.
- [26] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-image hdr reconstruction by learning to reverse the camera pipeline. In *CVPR*, 2020.
- [27] Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. arXiv:2308.09713, 2023.
- [28] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In *CVPR*, 2022.
- [29] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In ECCV, 2020.
- [30] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM TOG*, 2022.
- [31] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In CVPR, 2016.
- [32] Xiaowei Song, Jv Zheng, Shiran Yuan, Huan-ang Gao, Jingwei Zhao, Xiang He, Weihao Gu, and Hao Zhao. Sa-gs: Scale-adaptive gaussian splatting for training-free anti-aliasing. arXiv:2403.19615, 2024.
- [33] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *CVPR*, 2022.
- [34] Haoyuan Wang, Xiaogang Xu, Ke Xu, and Rynson WH Lau. Lighting up nerf via unsupervised decomposition and enhancement. In *ICCV*, 2023.
- [35] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. arXiv:2310.08528, 2023.
- [36] Linning Xu, Vasu Agrawal, William Laney, Tony Garcia, Aayush Bansal, Changil Kim, Samuel Rota Bulò, Lorenzo Porzi, Peter Kontschieder, Aljaž Božič, et al. Vr-nerf: High-fidelity virtualized walkable spaces. In Siggraph Asia, 2023.
- [37] Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-splatting: Alias-free 3d gaussian splatting. CVPR, 2024.
- [38] Jiahui Zhang, Fangneng Zhan, Muyu Xu, Shijian Lu, and Eric Xing. Fregs: 3d gaussian splatting with progressive frequency regularization. *CVPR*, 2024.
- [39] Tianyi Zhang, Kaining Huang, Weiming Zhi, and Matthew Johnson-Roberson. Darkgs: Learning neural illumination and 3d gaussians relighting for robotic exploration in the dark. *arXiv:2403.10814*, 2024.