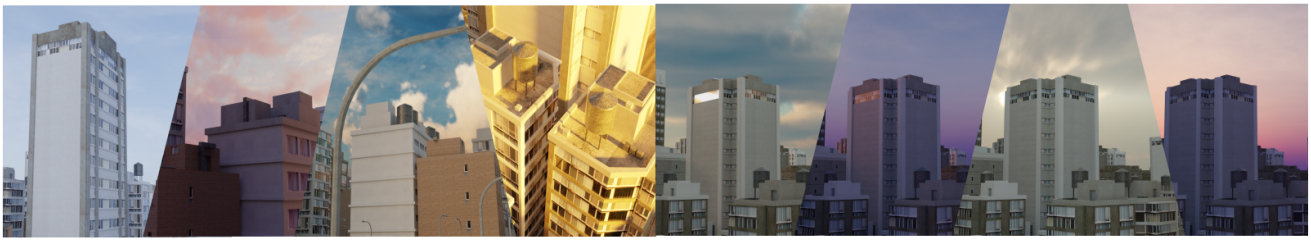
arXiv:2602.01118v1 [cs.CV] 1 Feb 2026

LightCity：面向多光照条件下室外逆向渲染与重建的城市数据集

Jingjing Wang[1] Qirui Hu[1] Chong Bao[1] Yuke Zhu[1] Hujun Bao[1] Zhaopeng Cui[1] Guofeng Zhang [1]浙江大学CAD&CG国家重点实验室



(a) Street & aerial urban views with complicated illumination (b) The urban view with diverse illumination

图 1. 我们推出了一个名为 LightCity 的新型高质量合成城市数据集。该数据集涵盖复杂的城市光照条件，包括多变的光照、逼真的间接照明和阴影效果，以及不同尺度下的街景和航拍图像采集。

摘要

城市场景中的逆向渲染对于自动驾驶和数字孪生等应用至关重要。然而，由于光照条件复杂，包括多光照以及间接光和阴影效果，这一领域面临着重大挑战。由于缺乏合适的数据集，这些挑战对内在分解和三维重建的影响尚未得到探索。本文推出LightCity，这是一个新型高质量合成城市数据集，具有多样的光照条件以及逼真的间接光和阴影效果。LightCity 包含 *300* 多张天空贴图，支持高度可控的光照；涵盖街景视角和航拍视角下不同尺度的 *5* 万多张图像；并提供深度、法线、材质分量、光照和间接光等丰富属性。此外，我们利用LightCity 对城市环境中的三项基础任务进行基准测试，并全面分析这些基准，为相关研究的推进奠定坚实基础。项目主页：*https://23wjj.github.io/LightCity/*。

1. 引言

城市场景中的逆向渲染 [58, 74] 对于自动驾驶、数字孪生和城市规划等各种应用日益重要，但城市场景中复杂的光照条件使这一问题尤其具有挑战性。主要面临两项挑战：1）**多光照**：由于光照不可控且随时间快速变化，因此它是城市场景中一个不受约束的因素。在日常场景中，城市场景会受到各种光照条件的影响，例如太阳位置（日出和日落）、天气变化以及季节变化。2）**间接光与阴影**：城市场景中建筑物之间复杂的空间布局会产生显著的间接光照和阴影效果，从而对场景外观造成重大影响。

应对这些挑战，需要算法能够稳健地适应复杂的光照效果。我们概述了三项关键功能要求：1）城市场景的固有分解能够稳健应对光照变化。例如，对于给定场景，无论光照条件如何变化，分解得到的反射率都应保持一致。2）在城市场景的逆向渲染中，应准确分解复杂的间接光和阴影效果。例如，间接光和阴影应纳入阴影分量，并与反射率分离。3）在复杂的光照挑战下，城市场景重建应得到准确且多视角一致的结果。

然而，这些功能在城市场景分析中在很大程度上仍未被探索，原因在于缺乏合适的

表示共同贡献。表示通讯作者。

数据集。现有具备多重光照、间接光和阴影条件的数据集主要关注单个物体 [18, 20, 23, 24, 36, 56]或小型花园环境 [26]。城市数据集 [13, 29, 33, 34, 57]很少考虑复杂的光照条件。例如，MatrixCity [29] 利用虚幻引擎生成城市场景，其光照仅通过调整光照方向和强度来控制，导致多样性有限，无法反映真实世界的条件。诸如Phototourism [50] 和OMMO [39] 等数据集旨在对单个城市建筑在多重光照条件下进行三维重建，但其简单的空间布局抹除了复杂的间接光和阴影效果。

本文首先提出一种新型合成城市数据集 LightCity，其中包含多样的光照条件，并在更广阔的城市环境中呈现复杂的间接光与阴影效果。其次，我们针对城市场景中的三项基础任务进行基准测试，以探索 LightCity 的三项既定功能。

具体而言，如表 1 所示，LightCity 具有以下主要特点：1）高质量：LightCity 基于 SceneCity [1] 插件构建，并采用 Blender Cycle引擎生成照片级真实图像，呈现逼真的光照与阴影效果。2）丰富的光照多样性：我们纳入了 300 多张天空贴图，涵盖从黎明到夜晚的全天时段，并可通过调整高动态范围成像贴图的旋转角度和强度，灵活控制光照。3）规模：该数据集包含不同尺度的合成城市街区，其中用于内在属性任务的图像超过 3 万个视图，用于重建任务的图像超过 2 万个视图，涵盖街景视角和航拍视角。4）属性全面：该数据集包含多种可支持各类视觉任务的属性，例如用于几何估计的深度与法线图、用于材质估计的漫反射与高光分量等。

为了展示我们的数据集在所述功能上的实用性，我们对三项任务进行基准评测。1）我们评估城市场景中的图像内在分解，以研究不同光照条件下的内在一致性。我们的发现表明，当前的图像内在分解模型在变化的光照条件下缺乏场景的内在连贯性。使用 LightCity 微调的模型往往能学习到不受光照变化影响的、更为一致的内在属性，并预测出更好的阴影属性。2）我们评估单一未知光照下城市场景中的多视图逆向渲染，以研究其内在精度和多视图一致性。我们的发现表明，当前的多视图逆向渲染算法在城市场景材质估计方面存在不足，并且难以分离与视角相关的间接光和阴影效果。3）我们评估多种光照下城市场景中的神经渲染，以研究多光照对新视角 syn- 的影响

数据集	Task	图像数量	级别	来源	内在属性	材质	光照
OMMO [39]	Rec.	14K	Wild	R	✗	✗	R
MatrixCity [29]	Rec.	519K	City	S	A	✓	ID
Phototourism [50]	Rec.	30K	地标	R	✗	✗	R
NeRF-OSR [47]	Rel.	3K	建筑物	R	✗	✗	R
OpenIllum [36]	Rel.	108K	物体		✗	✗	R
MIT Intrinsic [18]	Dec	110	物体	R	ASR	✗	R
MPI Sintel [7]	Dec	2.6K	电影	S	A	✗	✗
CGIntrinsic [31]	Dec	20K	室内	S	AS	✗	✗
IIW [3]	Dec	5K	室内	R	AS	✗	✗
Hypersim [46]	Und	82K	室内	S	ASR	✗	✗
EDEN [26]	Und	439K	花园	S	AS	✗	HID
Ours	重建+12月	50K	City	S	ASR	✓	HID

表1. 我们的 LightCity 数据集与先前数据集之间的属性比较。任务：Rec=重建，Rel=重新照明，12月=分解，未知=理解。来源（Source）：R=真实，S=合成。内在属性：A=反照率，S=阴影，R=残差。光照：R=真实，I=强度，D=方向，H=高动态范围图像。

论文。我们的研究结果表明，与神经辐射场（NeRF）[42]相比，采用三维高斯泼溅（3DGS） [21]的方法能够实现更出色的渲染质量和几何一致性。然而，基于 3DGS 的方法 [14, 25, 54, 60, 68] 会将多重光照引起的外观变化表现为浮动伪影。

我们的贡献总结如下：

- LightCity 提供了 300 多张光照高度可控的天空贴图，包含不同尺度的城市街区，并提供街景视角和航拍视图。此外，它还提供深度、法线、漫反射和光泽材质等丰富属性，以支持多种视觉任务。
- 我们在 LightCity 上对城市场景中的三项基础任务进行基准评测，包括本征图像分解、多视图逆向渲染，以及多重光照条件下的城市场景重建。
- 我们对基准评测结果进行了深入分析。研究结果凸显了多重光照对内在分解一致性的影响、间接光和阴影条件对逆向渲染的影响，以及不同光照条件下城市场景重建的准确性。

2. 相关工作

从神经辐射场（NeRF） [42]到三维高斯泼溅（3DGS） [21, 67],，三维场景表示领域的最新进展已实现高质量重建，包括城市尺度的重建 [37, 57]。然而，这些方法大多假设光照条件理想，而现实世界的数据采集中很少满足这一假设。NeRF-W [40]率先对室外光照变化进行建模，随后出现了基于 NeRF 的 Ha-NeRF [10], 、Neural-Recon [52] 和 CR-NeRF [61], 3, 以及基于 DGS 的 Wild-gs [60],、Wild-gaussian [25], 、Gaussian-wild [68]和 SWAG [14],; 这些方法均在 PhotoTourism 数据集上进行了评估。然而，PhotoTourism 数据集缺乏密集摆放的物体之间多样的光照交互，也缺少全面的地面

真值来进行几何评估，这凸显了构建更复杂、更统一的基准的必要性，以评估多光照约束下的室外重建。

除场景重建之外，理解图像与场景的本征属性（例如反射率 A和阴影S）仍是一个开放性问题。本征图像分解从朗伯假设（$I = A \times S$） [3, 8, 15, 31, 41, 48, 55], 发展到非朗伯模型（$I = A \times S + R$） [9, 49, 64], , 将镜面反射效果分解为残差 R。尽管已得到广泛研究，本征图像分解仍局限于室内数据集 [3, 22],, 而室外基准 [7, 26], 要么质量较低，要么过于简化。鉴于室外多光照效果的复杂性，需要大规模数据集来同时支持训练与基准评测。场景逆向渲染与本征分解目标相近，旨在从多视角图像估计场景反照率，并常常结合基于物理的渲染（PBR）模型以恢复材质属性（例如粗糙度与金属性），用于后续重新照明。现有的NeRF-[47, 56, 63] 和 3DGS- [17, 32] 的逆向渲染方法主要聚焦于物体数据集 [36, 43], , 而城市室外逆向渲染 [35, 45, 58, 66] 因数据集的局限而在很大程度上仍未被探索。更多细节见补充材料第1节。

3. LightCity 数据集

该 LightCity 数据集旨在为本征图像分解、多视图逆向渲染和城市场景重建提供具有挑战性的基准，以应对更加真实的多光照约束。如图 2所示，数据集的构建遵循三阶段流程：收集并标注城市尺度资产和环境贴图（第 3.1节）、采样相机视角（第 3.2节），以及采用基于物理的渲染方法渲染各种属性（第 3.3节）。我们的框架基于开源软件 Blender 构建，可适配各种公开可用的城市资产。

3.1. 数据收集与建模

城市尺度资产。 为展现室外场景中的多光照挑战，我们的数据集满足三个关键标准：具有照片级真实感的高质量城市尺度资产、包含丰富色彩细节的多样化物体类别，以实现全面的场景理解，以及具有复杂物体布局的大规模城市环境，以捕捉长期光照依赖关系。为此，我们采用
SceneCity Blender [1]插件可生成包含 450 多种物体类别和 80 种材质属性的城市地图，从而快速构建场景资产。

区块划分与标注。 我们根据建筑群将基础城市地图划分为五个由大到小的层级， $\{A, B, C, D, E, F\}$。详细信息请参阅补充材料第3节。为了标注城-

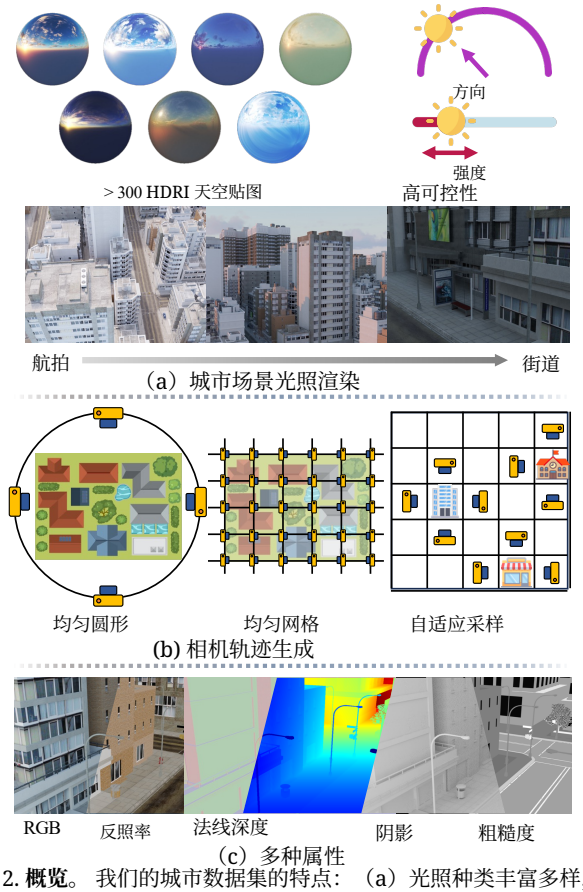


图 2. **概览**。 我们的城市数据集的特点：（a）光照种类丰富多样，且可灵活控制。（b）以不同尺度进行视角采样。（c）多种属性。

市地图中的物体，我们首先将预定义的城市资产拆分为单个物体，因为每个物体都可能属于一个更大的资产组。然后为每个物体分配一个语义标签。

多光照建模。为了模拟室外场景中多样的光照条件，我们选择高动态范围成像天空贴图作为主要表征。与以往仅调整光照强度和方向的合成数据集不同，我们的方法利用天空纹理的变化，这些变化不仅会影响光照，还能指示一天中的时段。我们收集了 300 多张 HDRI贴图，涵盖从黎明到夜晚的时段，并通过旋转 HDRI 和调整环境光强度来控制全局光照。与其他广泛使用的数据集相比，我们的数据集展现出更丰富的动态天空变化，为城市场景任务带来了更复杂的光照挑战。

3.2. 相机轨迹生成

均匀视角采样。圆形采样和均匀网格采样等视角采样方法已广泛应用于以往的数据集 [29, 36, 39]。为进一步消除相机轨迹对室外重建质量的影响，并更好地评估光照

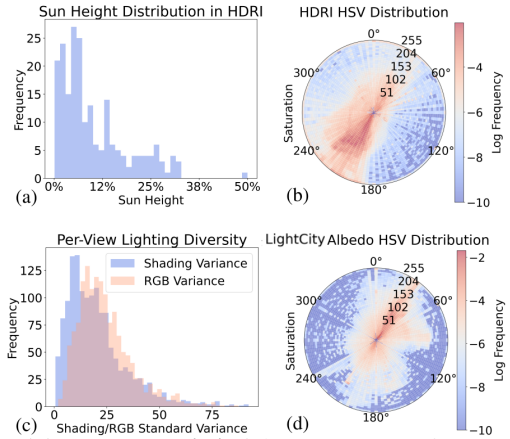


图 3。（a）HDRI 太阳高度分布，其中 0% 代表海平面，覆盖了一天中太阳轨迹的广泛范围。（b、d）HDRI贴图和反照率的 HSV 分布涵盖了广阔的色彩空间，反映出多样的光照和纹理。（c）渲染阴影和 RGB 方差的分布展现了亮度和颜色变化的广泛范围。

因素对其的影响，我们采用了两种均匀视角采样方法：均匀圆形采样和均匀网格采样。

自适应视图采样。为构建更详细、更全面的视点，我们采用自适应视角采样策略，融合街景视图和航拍视图。详细信息请参阅补充材料第 2 节。

标定。大量研究表明，相机视点之间的重叠区域对于高质量三维重建至关重要。相机视角之间重叠不足会显著增加重建失败的可能性。为优化随机生成的相机视角，我们使用 COLMAP 进行粗略的多视图点云重建，并筛除未能建立足够特征对应关系的相机位姿。

3.3. 图像渲染与筛选

渲染。基于城市资产和预先生成的相机视点，我们选择 Blender Cycles（一款基于物理的渲染器）作为渲染引擎。它能够高效模拟全局光照，从而实现照片级真实效果。基于物理的渲染依赖双向散射分布函数（BSDF）着色器来定义材质属性，例如漫反射、光泽反射和透明度等。它利用几何缓冲区分别存储中间渲染结果，包括环境光照、几何、直接与间接光照等。根据 Blender Cycles手册 [4],最终图像 $I(\mathbf{x})$可形式化定义如下：

$$\begin{aligned} I_D(\mathbf{x}) &= g_D(\mathbf{x})(g_{Ddirect}(\mathbf{x}) + g_{Dindirect}(\mathbf{x})), \\ I_G(\mathbf{x}) &= g_G(\mathbf{x})(g_{Gdirect}(\mathbf{x}) + g_{Gindirect}(\mathbf{x})), \\ I(\mathbf{x}) &= I_D(\mathbf{x}) + I_G(\mathbf{x}) + I_T(\mathbf{x}) + I_B(\mathbf{x}) + I_E(\mathbf{x}), \end{aligned} \tag{1}$$

其中 $\mathbf{x}$ 表示像素位置，I_D，I_G，I_T，I_B，I_E表示漫反射双向散射分布函数的渲染贡献，光泽双向散射分布函数、透射双向散射分布函数、背景和发射光源。$g_{D(G)}$、$g_{D(G)direct}$ 和$g_{D(G)indirect}$ 分别表示颜色通道，以及漫反射或光泽双向散射分布函数分量的直接光照通道和间接光照通道。为了从渲染方程中提取图像内在属性，我们将反照率 $A(\mathbf{x})$ 和阴影 $S(\mathbf{x})$定义如下：

$$\begin{aligned} A(\mathbf{x}) &= g_D(\mathbf{x}), \\ S(\mathbf{x}) &= g_{Ddirect}(\mathbf{x}) + g_{Dindirect}(\mathbf{x}), \\ R(\mathbf{x}) &= I_G(\mathbf{x}) + I_T(\mathbf{x}) + I_B(\mathbf{x}) + I_E(\mathbf{x}). \end{aligned} \tag{2}$$

我们搭建了一个自动化多通道渲染管线，为数据集生成多种属性，包括反照率、阴影、材质属性（即粗糙度、金属度和镜面反射）、法线、语义和深度。为了更好地模拟光照效果，我们为每个相机视点随机选择两张 HDRI贴图，并应用四种随机旋转和环境光照强度。为了兼顾渲染效率与视觉保真度，我们将输出分辨率设为 1024×768，最大渲染采样数设为 512，并优先保证照片级真实感，而非渲染速度。

筛选。 在图像筛选的最后阶段，我们会剔除过暗或曝光不足、可能对重建产生负面影响的图像。具体做法是设定强度阈值，移除亮度不足的图像。此外，为了排除信息量极少的图像，例如仅包含一面墙、空旷天空或视野延伸至城市边界之外的图像，我们会分析相应的语义图，并筛除不同物体总数少于两个的图像。通常，对于每张图像 I，我们的筛选标准可表述如下：

$$\begin{cases} Y_I = 0.299R_I + 0.587G_I + 0.114B_I, Y_I < \tau_Y, \\ O_I = |\{l \mid l \in S_I\}|, O_I < \tau_O, \end{cases} \tag{3}$$

其中，Y_I、$R/G/B_I$、S_I 分别表示亮度、单颜色通道和语义图，τ_Y和 τ_O 分别表示用于筛选亮度和语义的阈值集合。

4. 特征分析

本节将分析我们提出的 LightCity 数据集的统计特征。有关该数据集的详细信息见补充材料第3节。

可控多重照明。 与真实世界数据集相比，我们的合成数据集能够精准控制多种户外光照条件。除了调整全局光照的方向和强度外，LightCity 还着重于视角相关天空纹理，并能准确模拟城市环境照明。我们的高动态范围成像贴图涵盖了从清晨到午夜的时段，呈现出丰富的色彩和亮度分布，如图 3 (a)(b)所示。这种全面的控制能力使其更适用于城市重建。

多样化内参。 LightCity 提供自动化的多属性渲染流程，可生成真实标签

LightCity-Outdoor												
方法		D_{train}	反照率					阴影				
			尺度不变峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	尺度不变均方误差↓	尺度不变局部均方误差↓	尺度不变峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	尺度不变均方误差↓	尺度不变局部均方误差↓
DNN	PIE-Net [15]	室外	15.49	0.529	0.449	0.053	0.050	16.15	0.463	0.570	0.114	0.106
	DPF [11]	H	16.20	0.369	0.642	0.020	0.018	17.92	0.445	0.660	0.031	0.028
		H+L	18.94	0.435	0.611	0.013	0.013	21.09	0.496	0.645	0.024	0.022
	CDID [9]	E	15.80	0.370	0.385	0.016	0.015	20.25	0.204	0.542	0.013	0.012
扩散	DMP [27]	H	17.61	0.397	0.448	0.015	0.013	18.38	0.525	0.526	0.023	0.020
		H+L	22.80	0.520	0.363	0.005	0.005	26.58	0.611	0.432	0.003	0.003
	IntrinsicAny [12]	物体	16.82	0.390	0.535	0.049	0.047			/		

表 2. LightCity上的内在分解第一名、第二名和第三名的数值以高亮显示。

EDEN-Outdoor 分布外												
方法		D_{train}	反照率					阴影				
			尺度不变峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	尺度不变均方误差↓	尺度不变局部均方误差↓	尺度不变峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	尺度不变均方误差↓	尺度不变局部均方误差↓
DNN	PIE-Net [15]	室外	6.397	0.094	0.538	0.973	0.956	6.58	0.050	0.499	3.396	3.256
	DPF [11]	H	11.39	0.125	0.592	0.196	0.191	10.56	0.053	0.568	1.602	1.473
		H+L	13.11	0.094	0.580	0.159	0.151	14.16	0.388	0.452	0.552	0.476
	CDID [9]	E+室内.等	17.54	0.234	0.351	0.057	0.055	9.39	0.131	0.550	0.662	0.646
扩散	DMP [27]	H	13.12	0.139	0.382	0.307	0.297	8.09	0.034	0.470	3.760	3.482
		H+L	15.10	0.158	0.347	0.163	0.150	11.54	0.075	0.402	1.766	1.572
	IntrinsicAny [12]	物体	6.293	0.059	0.573	0.866	0.855			/		

表3. 性能 ce对EDEN上内在分解的比较。其中第一，第二，以 被高 d.

用于基于 PBR 的逆向渲染和内在图像分解。它提供分解后的光照分量，以及金属度、粗糙度和镜面反射等基本材质属性。如图3（c）所示，在同一视角下，我们渲染的 RGB 图像和阴影的方差分布范围广泛，表明光照变化丰富。与 MatrixCity（见补充材料第3节）相比，我们漫反射颜色的 HSV 分布覆盖了更广的色彩空间范围（图 3 （d））。丰富多样的内参使我们的数据集能够支持更广泛的未来应用。

5. 实验

本节探讨多光照条件下本征图像分解、逆向渲染和城市场景重建所面临的挑战。我们将最先进的方法适配到 LightCity，并建立了这些任务在多光照约束下的基准，以深入了解它们在不同光照条件下的表现。基线方法详情见补充材料。

5.1. 本征图像分解

实现细节。 由于缺乏室外本征数据集，室外场景的本征图像分解一直鲜有研究。为探究室外环境中本征分解面临的挑战并评估现有方法，我们使用 LightCity 重建数据集开展了两项实验：混合训练（在我们的数据集上进行训练）和直接评估（在我们的数据集上直接测试）。我们采用 DMP [27], a 基于扩散模型的方法和 DPF [11], a 基于深度神经网络的方法作为骨干网络，在仅室内数据集和室内外混合数据集上进行有监督的本征分解训练。此外，为全面评估现有经过充分预训练的模型在室外城市环境中的本征分解质量，我们以三种最先进（SOTA）的方法为基准：IntrinsicAny [12]（基于扩散模型）、CDID [9]（基于深度神经网络）和 PIE-Net [15]（传统的自监督方法）。我们在包括 LightCity 在内的多种室内和户外数据集上对这些方法进行评估，以全面分析其在真实世界本征分解中的有效性。

数据集。 对于室内场景，我们使用广泛采用的合成数据集 Hypersim [46]（简写为 H）作为混合训练和直接评估数据集，该数据集提供各种室内房间的漫反射和阴影信息。对于室外场景，我们使用自己的 LightCity 数据集（L），该数据集提供不同城市景观视角下的多光照本征属性。此外，我们使用 IIW [3]（I）和 EDEN [26]（E）作为室内和室外的域外测试数据集，并使用 BigTime v1 [30] 和 Waymo Open [53] 进行真实场景评估。

指标。 对于反照率和阴影，我们使用尺度不变峰值信噪比（si-PSNR）、结构相似性（SSIM）和学习感知图像块相似度（LPIPS）[71] 来评估分解结果的质量。此外，我们还使用尺度不变均方误差（si-MSE）和尺度不变局部均方误差（si-LMSE）来评估各个分量的统计值误差。除此之外，为与以往研究保持一致，我们还在补充材料第4节中报告了 IIW 上的加权人类判别率（WHDR），作为一项指标。

固有图像评估。 如表2所示，DMP [27]与 LightCity 混合微调后，在 LightCity 上的内在分解任务中表现最佳。相比之下，基于扩散模型的 IntrinsicAny [12] 和基于深度神经网络的方法在城市场景分解任务上均表现不佳，两种内在属性的尺度不变峰值信噪比都低于 20。对于 EDEN（表 3），基于深度神经网络的固有图像模型在反照率分解任务上的平均表现最佳，这归功于 EDEN 被纳入其训练数据集中；而 DPF [11]与 LightCity 混合微调后则在阴影分解任务上展现出更高的

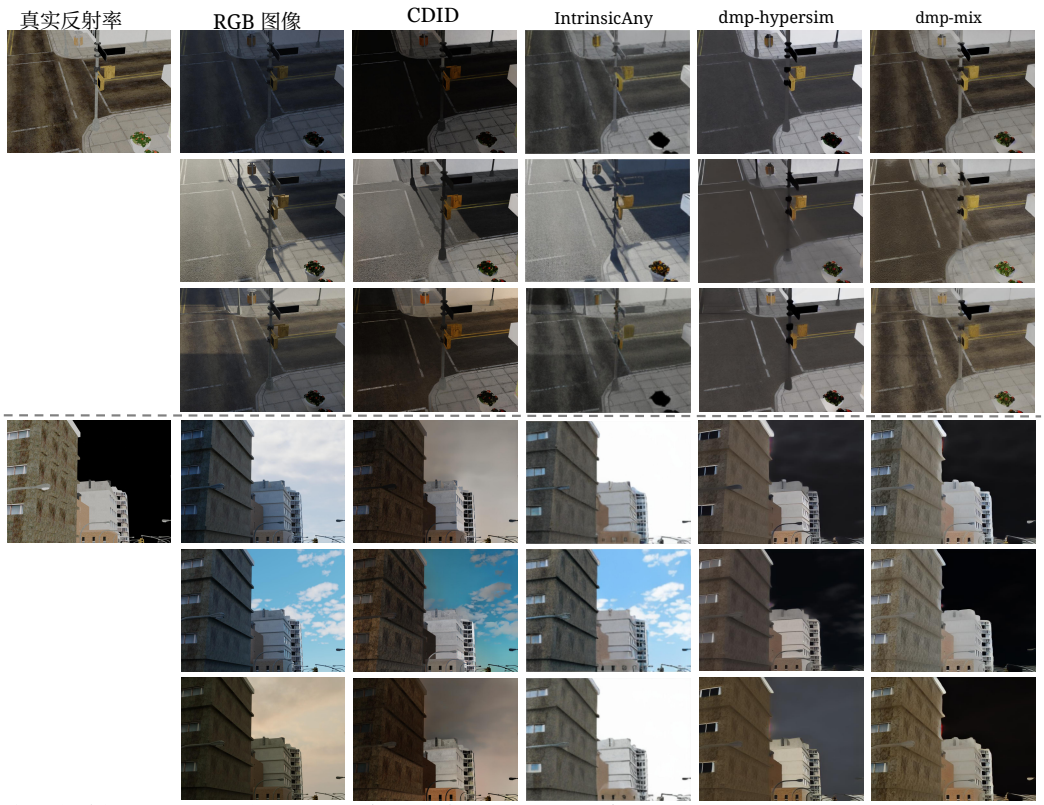


图 4. 图像内在分解中，同一视角在多光照条件下分解得到的反照率可视化。

准确性。值得注意的是，尽管 DPF 和 DMP 未在 EDEN 上进行预训练，却在阴影分解任务上表现最佳。这一优势得益于 LightCity 单视点、多光照的特性：它为城市场景提供了丰富多样的阴影变化，使模型能够学习复杂光照效果的强先验。补充材料第4节中关于 Hypersim 室内数据集的其他结果表明，采用 DMP 主干网络并结合混合训练的 LightCity 数据集，甚至能够提升室内内在分解的效果。这进一步说明，从固定视点学习多样的光照条件能够增强模型的分解能力。为进一步探究多光照对内在图像预测的影响，我们在视点相同但光照条件不同的情况下，将不同方法估计的反照率可视化，如图 4所示。对于光照效果显著的图像，CDID [9] 和 IntrinsicAny [12] 都难以将光照与预测的反照率完全分离，导致街道上仍有阴影和阴影变化，如图 4第3列和第4列所示。此外，对于光照条件不同但场景相同的图像，这些方法预测的反照率存在显著不一致，这可能是由于内在分解模型普遍采用单图像监督范式。相比之下，使用 LightCity 数据集进行混合训练后，DMP 在不同光照条件下生成的反照率预测更加一致。这一高

这凸显了将多光照室外数据集纳入其中的重要性，有助于提升传统内在分解方法在不受约束环境中的泛化能力。BigTime v1 和 Waymo Open 上的更多分解结果见补充材料第4节。

5.2. 多视图逆向渲染

实现细节。 为凸显 LightCity 数据集在室外逆向渲染方面的挑战，我们评估了两种最先进的基于多视图本征分解的逆向渲染方法，即 NeRF-OSR [47] 和 GS-IR [32], , 评估对象为分解后的本征属性和材质属性。

数据集。 我们在均匀环绕视图下，针对四个不同区块 F2、F3、E1 和 E2 评估逆向渲染。测试视图按每隔 8 张图像均匀采样一张。由于 NeRF-OSR 需要多光照输入，而 GS-IR 需要单光照输入，因此我们使用相同的训练视点，但分别以多光照和单光照渲染这些视点，用于 NeRF-OSR 和 GS-IR。为确保公平评估，两种方法均使用相同的新颖固定光照测试视图。

指标。 我们使用三项指标评估渲染质量：峰值信噪比（PSNR）、结构相似性（SSIM）和学习感知图像块相似度 [71]。我们还评估逆向渲染中的内在分解质量。具体而言，GS-IR 和 NeRF-OSR 均恢复出反照率，将其视为与光照无关的内在属性。对于依赖光照的组成

区块	F2		F3	
方法	NeRF-OSR [47]	GS-IR [32]	NeRF-OSR	GS-IR
PSNR	15.11	**15.32**	**15.34**	13.77
SSIM	0.583	**0.630**	0.641	**0.700**
学习感知图像块相似度	0.362	**0.312**	0.343	**0.275**
区块	E1		E2	
方法	NeRF-OSR	GS-IR	NeRF-OSR	GS-IR
PSNR	**15.00**	12.85	**13.92**	13.10
SSIM	0.595	**0.631**	0.555	**0.625**
学习感知图像块相似度	0.385	**0.340**	0.430	**0.349**

表 4. LightCity 上的反照率估计对比

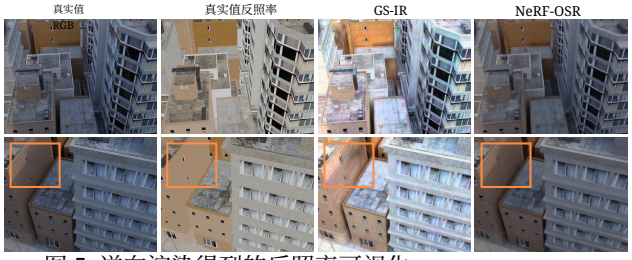


图 5. 逆向渲染得到的反照率可视化。

部分，NeRF-OSR 显式建模光照和阴影，以适应变化的光照条件；而 GS-IR 则将材质属性（即金属度和粗糙度）纳入模型，作为光相互作用的关键因素。我们使用均方误差作为材质属性的评估指标。

逆向渲染。 为研究不同光照解耦策略对内在分解的影响，我们按照表4所示，评估了两种方法的反照率重建质量。在不同场景中，NeRF-OSR 的平均峰值信噪比更高，而 GS-IR 的视觉质量更优；相比之下，NeRF-OSR 通过神经网络进行图像相关的反照率预测，在结构相似性和学习感知图像块相似度方面表现更好。然而，如图 5所示，在城市场景中进行多视图优化时，两种方法都难以实现符合物理规律且具有场景特异性的内在分解，常常将光照效果与漫反射颜色混在一起。此外，GS-IR 基于高斯的体渲染能够合成更逼真的图像，但仍容易出现漏光。相比之下，NeRF-OSR 基于图像的解耦方法往往会在渲染结果中引入表面粗糙度和噪声，这主要是因为阴影与反照率未能充分分离。此外，我们观察到，在全部四个区块中，GS-IR 在新视角合成和几何重建方面的表现均优于 NeRF-OSR。材质估计、几何重建和新视角合成的详细结果见补充材料第 5 节。

5.3. 多光照城市重建

实现细节。 为揭示 LightCity 数据集在多光照条件下进行户外重建所面临的挑战，我们评估了五种最先进的重建模型，即 NeRF- W [40], NeRF-OSR [47], Wild-gaussian [25], Gaussian- wild [68], NexusSplats [54], , 并分析它们在新视角合成和几何恢复方面的表现。

数据集。 为尽量降低极具挑战性的场景对重建质量的影响，并更好地评估各方法处理多光照的能力，我们从 LightCity 数据集中选取了四个相对较小的区块，即 F2、F3、E1 和 E2，并生成均匀圆形相机视角用于重建。每个区块包含约 200 张图像，每 8 张图像中选取 1 张作为测试视角。为弥补在自然光照变化下评估城市重建新视角合成质量方面的不足，我们针对每个区块，在相同的采样视角下构建了两组数据：单光照和多光照。在多光照设置中，每个视角都会从我们的集合中随机选择一张环境贴图。相比之下，单光照设置采用一张单一、光照充足的晴天 HDRI 贴图，该贴图与多光照设置中使用的贴图不同，以确保光照条件受控且明确。我们在两种光照条件下训练所有方法，训练参数严格遵循原论文中的设置。

评估指标。 对于新视角合成，我们采用峰值信噪比、结构相似性和学习感知图像块相似度作为评估指标，并遵循 NeRF-W 的评估设置。对于几何重建，我们使用两个指标评估法线的准确性：平均角误差（MeaAE）和中位角误差（MedAE）。我们在单光照和多光照设置下，分别使用测试视图评估所有方法。

新视角合成。 首先，我们在单光照设置下训练所有方法，将其作为基线，以评估其基础重建质量。我们主要关注在多光照数据集上训练的不同方法的渲染质量。如表5所示，我们使用两个测试集评估新视角合成性能：（1）多光照测试集，其光照分布与训练集相同；（2）单光照测试集，它引入了一种简单却具有挑战性的光照条件，以便进行受控评估。

在横向比较中，所有方法在多光照测试集上的峰值信噪比均持续高于单光照测试集。这表明，当前最先进的多光照重建方法难以泛化到未见过的光照条件。这一局限可能源于它们依赖从训练集中学习的图像依赖的外观先验，而这些先验本质上针对特定分布，在面对不同光照条件时无法有效适应。

在垂直方向上，对于规模相对较小的区块 F2 和 F3，NeRF-W [40] 的峰值信噪比高于基于 GS 的方法，表明其在几何复杂度较低的简单场景中表现出色。然而，对于建筑结构更密集、纹理更丰富的复杂城市街区 E1 和 E2，NeRF-W [40] 的峰值信噪比略低于 Gaussian-wild [68]。与此同时，Gaussian-wild [68]的结构相似性和学习感知图像块相似度得分更优，表明它能更好地保留复杂场景的感

区块	F2						F3					
光照	多光照			单光照			多光照			单光照		
指标	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓
Gaussian-wild [68]	25.72	0.844	0.308	21.54	0.794	0.331	28.62	0.903	0.154	24.88	0.859	0.191
Wild-gaussian [25]	22.20	0.759	0.483	18.80	0.691	0.545	25.93	0.852	0.312	22.06	0.787	0.381
NexusSplats [54]	23.70	0.779	0.476	20.59	0.713	0.517	25.06	0.815	0.393	21.43	0.742	0.458
NeRF-W [40]	26.06	0.832	0.289	24.368	0.797	0.296	29.07	0.874	0.240	25.90	0.819	0.275
NeRF-OSR [47]	25.40	0.778	0.417	21.08	0.697	0.444	26.07	0.802	0.381	20.95	0.703	0.418

区块	E1						E2					
光照	多光照			单光照			多光照			单光照		
指标	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓
Gaussian-wild [68]	27.78	0.855	0.300	22.99	0.778	0.364	27.94	0.861	0.250	23.09	0.803	0.297
Wild-gaussian [25]	24.14	0.763	0.521	19.52	0.655	0.619	24.23	0.750	0.550	19.17	0.645	0.641
NexusSplats [54]	24.91	0.770	0.521	20.13	0.653	0.605	25.36	0.787	0.462	21.00	0.694	0.541
Nerf-W [40]	26.61	0.730	0.465	21.72	0.599	0.507	28.71	0.824	0.325	24.12	0.736	0.368
NeRF-OSR [47]	25.88	0.756	0.476	18.85	0.601	0.525	26.07	00.764	0.473	13.53	0.517	0.558

表 5. 城市场景下多光照条件中的新视角合成对比。 第一和第二名的数值以高亮显示。



图 6. 城市重建中多光照条件下新视角合成的可视化结果。

知与结构保真度。

总体而言，如图 6所示，NeRF-W 能有效捕捉全局光照效果（例如投射在建筑物上的阴影）（第 1、2 列），并且在较简单的场景（例如精细细节重建）中表现稳健（第 3、4 列）。相比之下，Gaussian-wild 更善于适应复杂场景，但代价是对多视图不一致性更加敏感。几何评估和基线方法请参见补充材料第 6 节。

6. 结论

在本文中，我们提出了LightCity，一个具有丰富光照变化的高质量城市场景数据集。它在光照复杂度和属性多样性。借助这些丰富的属性，我们评估并基准测试了复杂城市照明下的三项关键任务：内在分解、逆向渲染和户外重建。研究结果凸显了现有方法在多光照条件下保持结果一致性所面临的挑战。实验表明，我们的数据集提升了内在分解的质量，并支持 NeRF 和 3DGS 用于可变光照城市场景中的逆向渲染和重建评估。我们希望未来的研究能着重提升核心任务在光照变化复杂的城市场景中的鲁棒性。

致谢 本工作部分得到中国国家自然科学基金（编号：62425209 和 62441222）、浙江大学信息技术中心和 CAD&CG 国家重点实验室的资助。

参考文献

[1] Arnold。《Scenecity》。https://www.cgchan.com/，2009年。访问日期：2025-03-07。2, 3[2] Jonathan T Barron 和 Jitendra Malik。通过阴影获取形状、光照和反射率。《IEEE模式分析与机器智能汇刊》，37(8):1670–1687，2014年。12[3] Sean Bell、Kavita Bala 和 Noah Snavely。野外场景中的固有图像。《ACM图形学汇刊》（TOG），33(4): 1–12，2014年。2, 3, 5, 12[4] Blender基金会。渲染通道——Blender手册。https://docs.blender.org/manual/en/2.80/render/layers/passes.html，2019年。访问日期：2025-07-26。4[5] Mark Boss、Raphael Braun、Varun Jampani、Jonathan T Barron、Ce Liu 和 Hendrik Lensch。NeRD：从图像集合中进行神经反射分解。载于 IEEE/CVF 国际计算机视觉会议论文集，第12684–12694页，2021年。12[6] Tim Brooks、Aleksander Holynski 和 Alexei A Efros。InstructPix2Pix：学习遵循图像编辑指令。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第18392–18402页，2023年。16[7] D. J. Butler、J. Wulff、G. B. Stanley 和 M. J. Black。一部用于光流评估的自然风格开源电影。载于欧洲计算机视觉会议（ECCV），第611–625页。施普林格出版社，2012年。2, 3[8] Chris Careaga 和 Yağız Aksoy。通过序数明暗进行本征图像分解。《 ACM 图形学汇刊》，43 (1):1–24，2023年。3, 14[9] Chris Careaga 和 Yağız Aksoy。野外场景中的多彩漫反射固有图像分解。《ACM图形学汇刊》，43 (6)，2024年。3, 5, 6, 12[10] Xingyu Chen、Qi Zhang、Xiaoyu Li、Yue Chen、Ying Feng、Xuan Wang 和 Jue Wang。野外场景中的幻觉神经辐射场。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第12943–12952页，2022年。2, 12[11] Xiaoxue Chen、Yuhang Zheng、Yupeng Zheng、Qiang Zhou、Hao Zhao、Guyue Zhou 和 Ya-Qin Zhang。DPF：通过弱监督学习密集预测场。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第15347–15357页，2023年。5, 13[12] Xi Chen、Sida Peng、Dongchen Yang、Yuan Liu、Bowen Pan、Chengfei Lv 和 Xiaowei Zhou。IntrinsicAnything：学习扩散先验，以在未知光照条件下进行逆向渲染。载于欧洲计算机视觉会议，第450–467页。Springer，2024年。5, 6, 13[13] David J Crandall、Andrew Owens、Noah Snavely 和 Daniel P Huttenlocher。结合马尔可夫随机场的运动恢复结构：用于大规模运动恢复结构的离散-连续优化。IEEE《模式分析与机器智能汇刊》，35(12):2841–2853，2012。2[14] Hiba Dahmani、Moussab Bennehar、Nathan Piasco、Luis Roldao 和 Dzmitry Tsishkou。SWAG：带外观条件高斯的野外图像泼溅。载于欧洲计算机视觉会议，页码325–340。Springer，2024。2, 12[15] Partha Das、Sezer Karaoglu 和 Theo Gevers。PIE-Net：用于本征图像分解的光度不变边缘引导网络。载于 IEEE/CVF计算机视觉与模式识别会议论文集，页码19790–19799，2022。3, 5, 14[16] Sara Fridovich-Keil、Giacomo Meanti、Frederik Rahbæk Warburg、Benjamin Recht 和 Angjoo Kanazawa。K-Planes：空间、时间和外观中的显式辐射场。载于IEEE/CVF计算机视觉与模式识别会议论文集，页码12479–12488，2023。12[17] Ji an Gao、Chun Gu、Youtian Lin、Zhihao Li、Hao Zhu、Xun Cao、Li Zhang 和 Yao Yao。可重光照3D高斯：结合双向反射分布函数分解与光线追踪的逼真点云重光照。载于欧洲计算机视觉会议，页码73–89。Springer，2024。3, 12[18] Roger Grosse、Micah K Johnson、Edward H Adelson 和 William T Freeman。本征图像算法的真值数据集与基线评估。载于2009年IEEE第12届国际计算机视觉会议，页码2335–2342。IEEE，2009。2[19] Rasmus Jensen、Anders Dahl、George Vogiatzis、Engin Tola 和 Henrik Aanæs。大规模多视图立体视觉评估。载于IEEE计算机视觉与模式识别会议论文集，页码406–413，2014。12[20] Haian Jin、Isabella Liu、Peijia Xu、Xiaoshuai Zhang、Song- fang Han、Sai Bi、Xiaowei Zhou、Zexiang Xu 和 Hao Su。TensoIR：张量逆向渲染。载于IEEE/CVF计算机视觉与模式识别会议论文集，页码165–174，2023。2[21] Bernhard Kerbl、Georgios Kopanas、Thomas Leimk¨uhler 和 George Drettakis。用于实时辐射场渲染的三维高斯泼溅。《ACM图形学汇刊》，42(1)，2023：139–2。12[22] Peter Kocsis、Vincent Sitzmann 和 Matthias Nießner。用于室内单视角材质估计的本征图像扩散。在 IEEE/CVF计算机视觉与模式识别会议论文集，页码5198–5208，2024。3[23] Zhengfei Kuang、Kyle Olszewski、Menglei Chai、Zeng Huang、Panos Achlioptas 和 Sergey Tulyakov。Neroic：从在线图像集合中神经渲染物体。《ACM图形学汇刊》（TOG），41(4):1–12，2022。2[24] Zhengfei Kuang、Yunzhi Zhang、Hong-Xing Yu、Samir Agar- wala、Elliott Wu、Jiajun Wu 等。Stanford-ORB：真实世界3D物体逆向渲染基准。神经信息处理系统进展，36:46938–46957，2023。2[25] Jonas Kulhanek、Songyou Peng、Zuzana Kukelova、Marc Pollefeys 和 Torsten Sattler。WildGaussians：野外场景中的三维高斯泼溅。神经信息处理系统大会，2024。2, 7, 8, 12[26] Hoang-An Le、Thomas Mensink、Partha Das、Sezer Karaoglu 和 Theo Gevers。EDEN：多模态合成

封闭式花园场景数据集。载于IEEE/CVF冬季计算机视觉应用会议论文集，页码1579–1589，2021。2, 3, 5[27] Hsin-Ying Lee、Hung-Yu Tseng 和 Ming-Hsuan Yang。利用扩散先验实现可泛化的密集预测。载于IEEE/CVF计算机视觉与模式识别会议论文集，页码7861–7871，2024。5, 13[28] Junyong Lee、Hyeongseok Son、Gunhee Lee、Jonghyeop Lee、Sunghyun Cho 和 Seungyong Lee。利用直方图类比进行深度色彩迁移。《The Visual Computer》，36(10):2129–2143，2020。19[29] Yixuan Li、Lihan Jiang、Linning Xu、Yuanbo Xiangli、Zhenzhi Wang、Dahua Lin 和 Bo Dai。MatrixCity：面向城市尺度神经渲染及其他应用的大规模城市数据集。载于IEEE/CVF国际计算机视觉会议论文集，页码3205–3215，2023。2, 3[30] Zhengqi Li 和 Noah Snavely。通过观察世界学习本征图像分解。载于计算机视觉与模式识别会议（CVPR），2018。5[31] Zhengqi Li 和 Noah Snavely。CGIntrinsics：通过基于物理的渲染改进本征图像分解。载于欧洲计算机视觉会议（ECCV）论文集，页码371–387，2018。2, 3, 12[32] Zhihao Liang、Qi Zhang、Ying Feng、Ying Shan 和 Kui Jia。GS-IR：用于逆向渲染的三维高斯泼溅。载于IEEE/CVF计算机视觉与模式识别会议论文集，页码21644–21653，2024。3, 6, 7, 12[33] Yiyi Liao、Jun Xie 和 Andreas Geiger。KITTI-360：用于二维和三维城市场景理解的新型数据集与基准。IEEE模式分析与机器智能汇刊，45(3):3292–3310，2022。2[34] Liqiang Lin、Yilin Liu、Yue Hu、Xingguang Yan、Ke Xie 和 Hui Huang。采集、重建与模拟：UrbanScene3D数据集。载于欧洲计算机视觉会议，页码93–109。Springer，2022。2[35] Zhi-Hao Lin、Bohan Liu、Yi-Ting Chen、Kuan Sheng Chen、David Forsyth、Jia-Bin Huang、Anand Bhattad 和 Shen- lo ng Wang。UrbanIR：基于单段视频的大规模城市场景逆向渲染。国际三维视觉会议，2025。3[36] Isabella Liu、Linghao Chen、Ziyang Fu、Liwen Wu、Haian Jin、Zhong Li、Chin Ming Ryan Wong、Yi Xu、Ravi Ramamoorthi、Zexiang Xu 等。OpenIllumination：用于真实物体逆向渲染评估的多光照数据集。神经信息处理系统进展，36:36951–36962，2023。2, 3[37] Yang Liu、Chuanchen Luo、Lue Fan、Naiyan Wang、Jun- ran Peng 和 Zhaoxiang Zhang。CityGaussian：利用高斯实现实时、高质量的大规模场景渲染。载于欧洲计算机视觉会议，页码265–282。Springer，2024。2, 12[38] Stephen Lombardi、Jason Saragih、Tomas Simon 和 Yaser Sheikh。用于人脸渲染的深度外观模型。《ACM图形学汇刊》（TOG），37(4):1–13，2018。12

[39] Chongshan Lu、Fukun Yin、Xin Chen、Wen Liu、Tao Chen、Gang Yu 和 Jiayuan Fan。用于新视角合成和隐式场景重建的大规模室外多模态数据集与基准。载于《IEEE/CVF国际计算机视觉会议论文集》，第7557–7567页，2023年。2, 3[40] Ricardo Martin-Brualla、Noha Radwan、Mehdi SM Sajjadi、Jonathan T Barron、Alexey Dosovitskiy 和 Daniel Duckworth。野外NeRF：用于无约束照片集合的神经辐射场。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第7210–7219页，2021年。2，7，8，12[41] Abhimitra Meka、Mohammad Shafiei、Michael Zollh¨ofer、Christian Richardt 和 Christian Theobalt。视频的实时全局光照分解。《ACM图形学汇刊》（TOG），40(3):1–16，2021年。3[42]Ben Mildenhall、Pratul P Srinivasan、Matthew Tancik、Jonathan T Barron、Ravi Ramamoorthi 和 Ren Ng。NeRF：将场景表示为用于视图合成的神经辐射场。《ACM通讯》，65(1):99–106，2021年。2，12[43] Lukas Murmann、Michael Gharbi、Miika Aittala 和 Fredo Durand。野外多光照图像数据集。载于《IEEE/CVF国际计算机视觉会议论文集》，第4080–4089页，2019年。3[44] Gaurav Parmar、Taesung Park、Srinivasa Narasimhan 和 Jun-Yan Zhu。利用文生图模型进行单步图像转换。arXiv预印本 arXiv:2403.12036，2024年。16[45] Ava Pun、Gary Sun、Jingkang Wang、Yun Chen、Ze Yang、Sivabalan Manivasagam、Wei-Chiu Ma 和 Raquel Urtasun。城市场景的神经光照模拟。《神经信息处理系统进展》，36，19326，2023年，19291，3：–。[46] Mike Roberts、Jason Ramapuram、Anurag Ranjan、Atulit Kumar、Miguel Angel Bautista、Nathan Paczan、Russ Webb 和 Joshua M Susskind。Hypersim：用于整体室内场景理解的照片级真实感合成数据集。载于《IEEE/CVF国际计算机视觉会议论文集》，第10912–10922页，2021年。2，5，15[47] Viktor Rudnev、Mohamed Elgharib、William Smith、Lingjie Liu、Vladislav Golyanik 和 Christian Theobalt。用于室外场景重新照明的NeRF。载于《欧洲计算机视觉会议》，第615–631页。Springer，2022年。2，3，6，7，8[48] Li Shen、Ping Tan 和 Stephen Lin。利用非局部纹理线索进行固有图像分解。载于《2008年IEEE计算机视觉与模式识别会议》，第1–7页。IEEE，2008年。3[49] Jian Shi、Yue Dong、Hao Su 和 Stella X Yu。跨ShapeNet类别学习非朗伯物体的固有属性。载于《IEEE计算机视觉与模式识别会议论文集》，第1685–1694页，2017年。3，12[50] Noah Snavely、Steven M Seitz 和 Richard Szeliski。照片旅游：在三维空间中探索照片集合。载于《ACM SIGGRAPH 2006论文集》，第835–846页。计算机协会，2006年。2，12[51] Pratul P Srinivasan、Boyang Deng、Xiuming Zhang、Matthew Tancik、Ben Mildenhall 和 Jonathan T Barron。

NeRV：用于重新照明和视图合成的神经反射与可见性场。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第7495–7504页，2021年。12[52] Jiaming Sun、Xi Chen、Qianqian Wang、Zhengqi Li、Hadar Averbuch-Elor、Xiaowei Zhou 和 Noah Snavely。野外神经三维重建。载于《ACM SIGGRAPH 2022会议论文集》，第1–9页，2022年。2，12[53] Pei Sun、Henrik Kretzschmar、Xerxes Dotiwalla、Aurelien Chouard、Vijaysai Patnaik、Paul Tsui、James Guo、Yin Zhou、Yuning Chai、Benjamin Caine、Vijay Vasudevan、Wei Han、Jiquan Ngiam、Hang Zhao、Aleksei Timofeev、Scott Ettinger、Maxim Krivokon、Amy Gao、Aditya Joshi、Yu Zhang、Jonathon Shlens、Zhifeng Chen 和 Dragomir Anguelov。自动驾驶感知的可扩展性：Waymo开放数据集。载于《IEEE/CVF计算机视觉与模式识别会议论文集》（CVPR），2020年。5[54] Yuzhou Tang、Dejun Xu、Yongjie Hou、Zhenzhong Wang 和 Min Jiang。NexusSplats：野外高效三维高斯泼溅。arXiv预印本 arXiv:2411.14514，2024年。2，7，8[55] Marshall F Tappen、William T Freeman 和 Edward H Adelson。从单张图像中恢复固有图像。《IEEE模式分析与机器智能汇刊》，27(9):1459–1472，2005年。3[56] Marco Toschi、Riccardo De Matteo、Riccardo Spezialetti、Daniele De Gregorio、Luigi Di Stefano 和 Samuele Salti。Relight my nerf：用于真实世界物体新视角合成和重新照明的数据集。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第20762–20772页，2023年。2，3[57] Haithem Turki、Deva Ramanan 和 Mahadev Satyanarayanan。Mega-nerf：面向虚拟飞行穿越的大规模神经辐射场可扩展构建。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第12922–12931页，2022年。2，12[58] Zian Wang、Tianchang Shen、Jun Gao、Shengyu Huang、Jacob Munkberg、Jon Hasselgren、Zan Gojcic、Wenzheng Chen 和 Sanja Fidler。神经场与显式几何表示相结合，用于城市场景的逆向渲染。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第8370–8380页，2023年。1，3[59] Tong Wu、Jiarui Zhang、Xiao Fu、Yuxin Wang、Jiawei Ren、Liang Pan、Wayne Wu、Lei Yang、Jiaqi Wang、Chen Qian 等。OmniObject3D：用于真实感知、重建和生成的大词汇量三维物体数据集。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第803–814页，2023年。12[60] Jiacong Xu、Yiqun Mei 和 Vishal Patel。Wild-gs：从无约束照片集合实时合成新视角。《神经信息处理系统进展》，37:103334–103355，2024年。2，12[61] Yifan Yang、Shuhai Zhang、Zixiong Huang、Yubing Zhang 和 Mingkui Tan。交叉射线神经辐射场：从无约束图像集合合成新视角。载于《IEEE/CVF国际计算机视觉会议论文集》，第15901–15911页，2023年。2，12[62]Chongjie Ye、Lingteng Qiu、Xiaodong Gu、Qi Zuo、Yushuang Wu、Zilong Dong、Liefeng Bo、Yuliang Xiu 和 Xiaoguang Han。StableNormal：降低扩散方差，实现稳定且清晰的法线。《ACM图形学汇刊》（TOG），43(6):1–18，2024年。16[63] Weicai Ye、Shuo Chen、Chong Bao、Hujun Bao、Marc Pollefeys、Zhaopeng Cui 和 Guofeng Zhang。IntrinsicNeRF：学习内禀神经辐射场，实现可编辑的新视角合成。发表于 IEEE/CVF 国际计算机视觉会议论文集，第339–351 页，2023 年。3 – 。[64] Yu-Ying Yeh 、 Koki Nagano、Sameh Khamis、Jan Kautz、Ming-Yu Liu 和 Ting-Chun Wang。通过虚拟灯光舞台和合成到真实域适应，学习重新照明肖像图像。《ACM图形学汇刊》（TOG），41(6):1–21，2022年。3, 12[65] Ye Yu、Abhimitra Meka、Mohamed Elgharib、Hans-Peter Seidel、Christian Theobalt 和 William AP Smith。自监督室外场景重照明。发表于欧洲计算机视觉会议，第 84–101 页。Springer，2020年。19[66] Zheng Zeng、Valentin Deschaintre、Iliyan Georgiev、Yannick Hold-Geoffroy、Yiwei Hu、Fujun Luan、Ling-Qi Yan 和 Miloš Hašan。Rgb↔x：使用材质与光照感知扩散模型进行图像分解与合成。发表于 ACM SIGGRAPH 2024 会议论文。美国计算机协会，2024年。3[67] Hongjia Zhai、Xiyu Zhang、Boming Zhao、Hai Li、Yijia He、Zhaopeng Cui、Hujun Bao 和 Guofeng Zhang。SplatLoc：基于三维高斯泼溅的增强现实视觉定位。31(5):3591–3601，2025年。2[68] Dongbin Zhang、Chuming Wang、Weitao Wang、Peihao Li、Minghan Qin 和 Haoqian Wang。野外高斯：面向无约束图像集的三维高斯泼溅。发表于欧洲计算机视觉会议，第 341–359 页。Springer，2024年。2, 7, 8, 12[69] Kai Zhang、Fujun Luan、Qianqian Wang、Kavita Bala 和 Noah Snavely。PhySG：使用球面高斯进行逆向渲染，实现基于物理的材质编辑与重新照明。发表于 IEEE/CVF 计算机视觉与模式识别会议论文集，第5453 5462, 2021 12 – 。[70] Lvmin Zhang、Anyi Rao 和 Maneesh Agrawala。通过施加一致的光传输，扩展基于扩散模型的野外训练，以实现光照协调与编辑。发表于第十三届国际学习表征会议，2025年。19[71] Richard Zhang、Phillip Isola、Alexei A Efros、Eli Shechtman 和 Oliver Wang。深层特征作为感知度量的非凡有效性。发表于 IEEE 计算机视觉与模式识别会议论文集，第586–595 页，2018年。5, 6[72] Yuanqing Zhang、Jiaming Sun、Xingyi He、Huan Fu、Rongfei Jia 和 Xiaowei Zhou。为逆向渲染建模间接照明。发表于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 18643–18652 页，2022年。12[73] Haonan Zhao、Yiting Wang、Thomas Bashford-Rogers、Valentina Donzella 和 Kurt Debattista。探索生成式 AI 在驾驶数据合成中实现仿真到现实的应用。发表于 2024年

*IEEE*智能车辆研讨会（*IV*），第 3071–3077 页。IEEE，2024年。16[74] Yongjie Zhu、Jiajun Tang、Si Li 和 Boxin Shi。Derendernet：通过形状（不）依赖的阴影渲染，对城市场景进行本征图像分解。发表于 *2021 IEEE* 国际计算摄影会议（*ICCP*），第 1–11 页。IEEE，2021年。1

1. 相关工作补充

1.1. 内在分解

内在分解旨在将图像分离为反射率（反照率）、阴影，有时还包括其他分量。传统内在分解方法依赖不同假设，由此形成三类主要模型：灰度内在模型、RGB内在模型和残差模型。灰度内在模型在早期工作中被广泛使用，基于优化的方法（如 [3]）以及各种数据驱动方法 [31]在单通道假设下估计反射率与阴影。RGB内在模型通过显式估计漫反射颜色和阴影变化，弥补了灰度模型的不足，从而在非均匀照明条件下提高了精度。然而，灰度模型与RGB模型均依赖朗伯假设，难以处理镜面反射。为克服这一问题，残差内在模型 [9, 49, 64], 被提出，将图像分解为反照率 A、阴影 S 和残差项 R，以更好地刻画镜面效应。多项工作已探索这种分解，以改进反射率建模。尽管取得进展，由于数据限制，大多数内在分解与逆向渲染方法仍在简单的室内数据集上进行评估。将这些方法扩展到复杂室外场景仍是一项持续的挑战，尤其是在多样光照条件下。

1.2. 逆向渲染

逆向渲染旨在从图像中恢复反照率、阴影和材质属性等场景内在属性，从而支持重新照明和新视角合成等应用。尽管该领域已取得显著进展，但大多数现有方法和评估仍以物体为中心，对大规模、复杂室外环境的探索有限。

早期的逆向渲染工作依赖基于物理的模型和优化技术，从单张图像中估计反射率和阴影 [2, 38]。随着神经表示的发展，基于 NeRF 的方法得以出现，能够在变化的光照条件下联合学习场景几何与外观。NeRV [51] 和 NeRD [5] 将反射率分解引入 NeRF，但其评估仅限于受控的，以物体为中心的数据集。近期的工作，例如 PhySG [69] 和 InvRender [72], ，将逆向渲染扩展到可处理非朗伯表面和间接光照，但其实验仍主要聚焦于合成物体或小规模真实世界物体。

基于高斯的表示也已应用于逆向渲染。GS-IR [32] 和 Re-lit3DGS [17]通过将场景外观分解为内在组成部分，扩展了三维高斯泼溅在重新照明方面的应用。然而，这些方法仍局限于物体级重建，尚未在大规模室外环境中进行测试。

尽管取得了这些进展，逆向渲染在大型真实场景中的应用仍未得到广泛探索。现有数据集主要以物体为中心（例如 DTU [19],NeRF Synthetic [42], OmniObject3D [59] , ），这限制了这些方法向城市尺度室外环境的泛化。缺乏涵盖复杂户外光照和多样材质的基准，仍是将逆向渲染拓展至物体级场景之外的一大障碍。

1.3. 户外场景重建

户外场景重建已得到广泛研究，神经辐射场（NeRF） [42] 和三维高斯泼溅（3DGS） [21]实现了高质量的场景表示。CityNeRF [57]和CityGaussian [37]等方法进一步提升了大规模城市重建的效果。然而，真实世界的城市级数据采集不可避免地会受到天气、一天中的时段及环境因素的影响，从而产生复杂的光照变化。不一致的光照给户外场景重建带来了巨大挑战。为应对光照变化，近期研究开始融合外观建模。NeRF-W [40]率先引入潜在嵌入，以实现变分光照外观建模。Ha-NeRF [10], 、CR-NeRF [61] 和K-Planes [16] 分别利用基于CNN的方法、跨射线范式和特征网格来建模不同的光照效果。NeuralRecon [52]专注于非受控条件下的几何重建。近期，研究者开始将外观建模扩展至3DGS，相关方法包括wild-gaussians [60], 、Wild-GS [25], 、Gaussian-wild [68], 和SWAG [14]。尽管这些方法提升了模型在Phototourism [50], 等数据集上的鲁棒性，但由于缺乏标准化数据集和统一基准，在极端多光照条件下重建城市场景仍然面临挑战。

2. 相机生成详情

为了生成城市场景的相机视角，我们设计了两种视角采样方法：均匀视图采样和自适应采样。我们还展示了使用给定的相机内参和外参，通过 COLMAP 重建得到的粗略点云，如图7所示。

2.1. 均匀视图采样

对于环绕视图，我们对相机施加两种跟踪约束，并使用帧队列记录其位姿。第一种约束基于用于跟踪的贝塞尔圆形路径。我们在不同区域的中心放置 3 个贝塞尔圆，其半径根据区块的长和宽设定。贝塞尔圆的高度由区域内物体的最大高度决定，以确保从俯视和仰视角度都能获得全面的视图。第二种约束基于标准目标跟踪。我们在场景中心放置一个空物体，使相机在沿贝塞尔曲线移动时保持正确的位姿。曲线的视图密度根据区块的尺度自适应设定。对于网格视图，我们计算每个区块的二维边界框，并根据其尺度层级将该边界框划分为不同分辨率的网格。在每个网格中放置四台相机，俯仰角范围为 20 至 45 度，偏航角分别为 $[0, 90, 180, 270]$ 度。

2.2. 自适应视图采样

对于街景视图，相机沿每个区块内的街道以0.5米的间隔放置。为增强街道及周围建筑物的细节，我们随机生成朝向四个方向的相机，其高度在 [0.5m、0.6m] 以及 [0.9m、1.3m],范围内采样，俯仰角范围为 $[45, 60]$ 度。对于航拍视图，需要注意，上一段所述的均匀视图采样难以充分捕捉密集聚集建筑物内部复杂的遮挡关系。受此启发，我们旨在自适应地将相机布置在密集聚集的建筑物中。具体而言，我们基于区块内所有建筑物的高度和相对位置，以递归方式构建邻接查找表。该查找表使我们能够生成区块的简化空间表示，并高效识别四个方向上的相邻结构。对于位于街道旁的建筑物，我们在建筑物上方可调节的高度处采样面向街道的视图。对于彼此相邻的建筑物，我们基于相对高度关系生成相机位姿，以确保对区块内建筑结构进行更细粒度的覆盖。

3. LightCity 详情

我们的 LightCity 数据集包含两部分，即 LightCity 重建数据集和 LightCity 本征数据集。用于城市场景重建的数据集根据场景簇划分为若干区域，如图9所示。为进一步说明我们数据集多样的漫反射颜色，我们还在图8中可视化了 MatrixCity 数据集的 HSV。

3.1. LightCity 重建数据集

LightCity 重建数据集主要用于多光照条件下的城市场景重建任务。在城市资产的层次划分下，我们通过均匀圆形、均匀网格和自适应采样渲染多视图图像。在相同视点下，我们还构建了单光照条件下的数据集。

3.2. LightCity 内在属性数据集

LightCity 内在属性数据集旨在用于改进和评测户外图像内在属性分解任务。为突出多光照对反照率和阴影预测所带来的挑战，我们随机选择两个天空环境，随机旋转每个视角，并为每个视角随机设置环境光照强度。这种策略能够模拟场景在一天内复杂的光照交互。对于每个视角，我们提供 8 张不同光照条件下的图像。

3.3. LightCity 数据集扩展

由于LightCity 主要关注多样光照对重建和分解的影响，因此我们进一步加入了一个小规模子集，其中包含在雾、雨和雪等极端天气条件下渲染的图像（见图 10）。这一增强旨在支持关于天气感知建模的进一步研究。此外，为丰富场景多样性，我们还提供了如图 11 所示的扩展，该扩展基于由另一个 Blender 插件 City Generator 构建的城市资产，涵盖了更广泛的城市布局。

4. 本征图像分解的更多结果

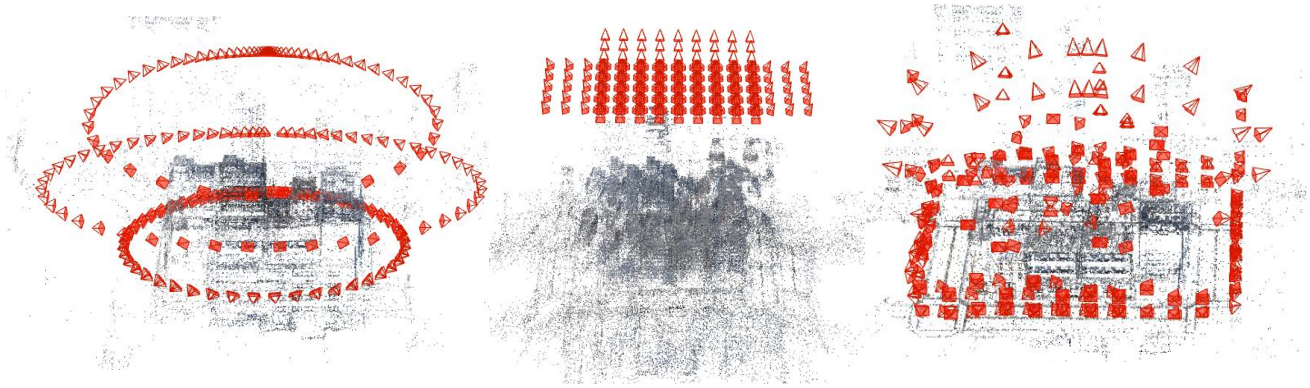
4.1. 基线方法详情

我们简要总结了用于评估本征图像分解的方法。

DPF. [11] DPF（密集预测场）是一种新颖的方法，利用弱点级监督完成密集预测任务。受隐式表示的启发，该方法通过预测查询坐标处的值，利用点级监督进行密集预测。它能够生成高分辨率输出，并在语义解析和本征图像分解方面表现出色。

dmp. [27]DMP 将预训练的文本到图像（T2I）扩散模型用作密集预测任务的先验。它通过插值重新表述扩散过程，从而在输入图像与预测结果之间建立确定性映射。DMP 使用低秩适配进行微调，在 3D 属性估计和本征图像分解等任务中展现出很强的泛化能力。

IntrinsicAny. [12]IntrinsicAnything 旨在解决从已知姿态的图像中恢复物体材质、同时光照未知这一难题。



(a) 均匀圆形 (b) 均匀网格 (c) 自适应采样

图 7. 使用三种相机视角采样方法获取的 F2区块 COLMAP 粗略点云。从左到右依次为均匀圆形、均匀网格和自适应采样。我们的自适应采样生成的点云最为精细且均匀，而另外两种方法的点云则聚集在目标场景的上部。

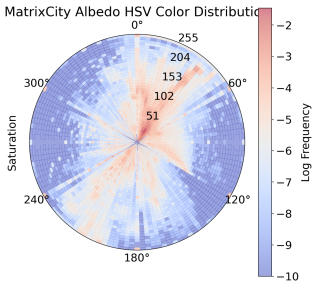


图 8. MatrixCity 反照率图像的 HSV 分布可视化。

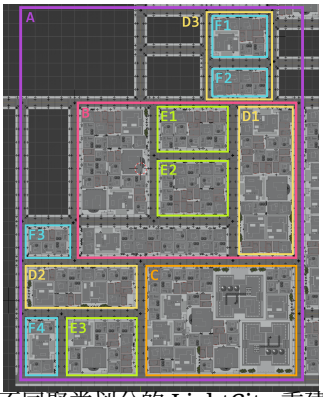


图 9. 按场景的不同聚类划分的 LightCity 重建数据集的不同层级。紫色矩形表示数据集的父节点区块 A，其中包含整个场景的图像。根据不同的区块大小，区块 A 进一步划分为 5 个不同尺度的层级，即 B，C，D，E，F。总计有 13 个区块。我们在倒数第二小的 E，F 层级上进行城市场景重建。

来应对这一问题。它不再仅仅依赖可微渲染，而是引入了一种生成式材质先验

使用扩散模型处理反照率和镜面分量。这有助于消除逆向渲染中的歧义。由粗到细的训练策略进一步确保多视图一致性，从而实现更准确的材质恢复。

CDID。 [8] CDID 通过将图像分解为漫反射反照率、彩色漫反射阴影和镜面残差，解决本征图像分解问题。不同于以往假设光照为单色且世界为朗伯世界的方法，CDID 逐步移除这些限制，从而实现更真实、更灵活的光照感知编辑。

PIENet。 [15]PIE-Net 是一种深度学习方法，通过将三维点云表示为参数曲线（直线、圆、B样条）来检测其中的特征边缘。该方法采用区域提议方法，先识别边缘点和角点，再对其排序以供选择。

4.2. 用于评估的详细数据集

我们使用多个室内和户外数据集，全面评估我们的混合微调机制，并简要总结所用的所有数据集。

Hypersim。 Hypersim 是一个大规模合成数据集，包含具有照片级真实感的室内场景，提供多视角 RGB 图像、深度图、表面法线以及内在分解数据（反照率、阴影）。它可作为室内内在分解、深度估计和逆渲染等任务的基准。

IIW。 IIW 是用于内在图像分解的真实数据集，包含 5,000 多张图像，并附有人类标注的成对反射率比较。它提供了多样化的非受限场景，是评估内在分解方法的重要基准。

EDEN。EDEN 是一个多模态合成数据集，专为面向自然的应用而设计，例如农业和园艺。它包含超过 30 万张来自 100+ 花园模型的图像，并标注了多种视觉模态，包



图10. LightCity 条件变化的更多示例。



图11. 由 City Generator 构建的 LightCity 扩展的更多示例。

括语义分割、深度、表面法线、固有颜色和光流。该数据集可用于语义分割和单目深度预测。

4.3. 室内场景

我们在表 7 和表 6中分别展示了 Hypersim 和 IIW 室内场景图像内在分解的评估结果。对于 Hypersim 数据集，基于深度神经网络的CDID在反照率分解的尺度不变峰值信噪比、尺度不变均方误差和尺度不变局部均方误差方面，平均表现最佳。不过，在反照率的结构相似性高于 0.53、阴影的结构相似性高于 0.62 时，基于扩散的DMP往往具有更好的视觉保真度。这与扩散模型生成图像的高质量相符。此外，经 LightCity 混合微调的DMP在阴影估计方面往往能获得更高的尺度不变峰值信噪比和学习感知图像块相似度。这一发现与此前在室外数据集上的结果一致。对于 IIW 数据集，在 Hypersim 上微调的DMP取得了最佳加权人类差异率得分；经LightCity混合微调的DMP质量略有下降，我们认为这是由于两个数据集之间存在域差异，给扩散模型的学习带来了挑战。然而，经LightCity 混合微调的DPF在加权人类差异率指标上降低了 5%，表现有所提升。

表6. IIW 数据集上反照率估计的性能。其中第一、第二和第三最显示。

IIW-室内			
方法		D_{train}	加权人类判别率 / %
基于深度神经网络	PIE-Net	/	32.77
	DPF	H	43.14
		H+L	38.502
	Intrinsic 2024	物体	21.33
基于扩散	DMP	H	19.08
		H+L	20.37
	IntrinsicAnything	/	27.08

4.4. 仿真到现实讨论

合成数据在计算机视觉和机器人视觉中发挥着至关重要的作用，尤其适用于场景理解和逆向渲染等任务。借助 Blender 或 Unreal 等引擎，可以精确控制光照、材质和几何结构。然而，低质量的合成数据集可能存在较大的仿真到现实差距，从而对模型向真实世界图像的泛化能力产生负面影响。为缓解这一问题，我们采用了最佳的开源渲染引擎 Blender Cycles，以实现照片级真实感。这种高度真实感能够缩小域差异并提升可迁移性，类似于 Hypersim 等数据集 [46]利用基于物理的渲染来提升在真实世界中的表现。

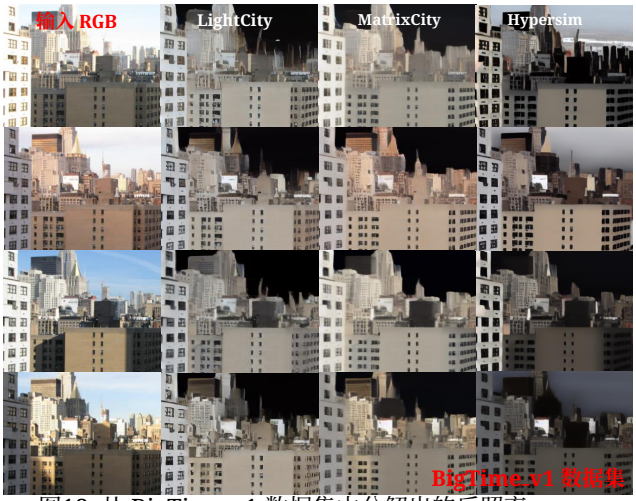


图12. 从 BigTime v1 数据集中分解出的反照率。

为进一步评估真实世界泛化能力，我们利用生成模型强大的泛化能力。近期研究表明，扩散模型展现出

表7。Hypersim室内数据集上的单图像本征分解结果。第一名、第二名和第三名的数

方法		D_{train}	Hypersim-室内 反照率 尺度不变峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	尺度不变均方误差↓	尺度不变局部均方误差↓	阴影 尺度不变峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	尺度不变均方误差↓	尺度不变局部均方误差↓
DNN	PIE-Net	室外	12.55	0.449	0.479	0.101	0.095	14.81	0.513	0.454	0.025	0.024
	DPF	H	15.43	0.445	0.576	0.033	0.031	14.59	0.570	0.531	0.070	0.063
		H+L	13.48	0.403	0.599	0.089	0.082	14.85	0.520	0.570	0.037	0.034
	CDID	E+室内.etc	16.70	0.487	0.372	0.023	0.020	14.85	0.190	0.515	0.011	0.010
扩散	DMP	H	16.48	0.534	0.369	0.036	0.034	15.59	0.624	0.353	0.046	0.043
		H+L	16.56	0.531	0.371	0.033	0.031	15.73	0.621	0.352	0.043	0.040
	IntriAny	物体	12.33	0.407	0.510	0.195	0.182			/		

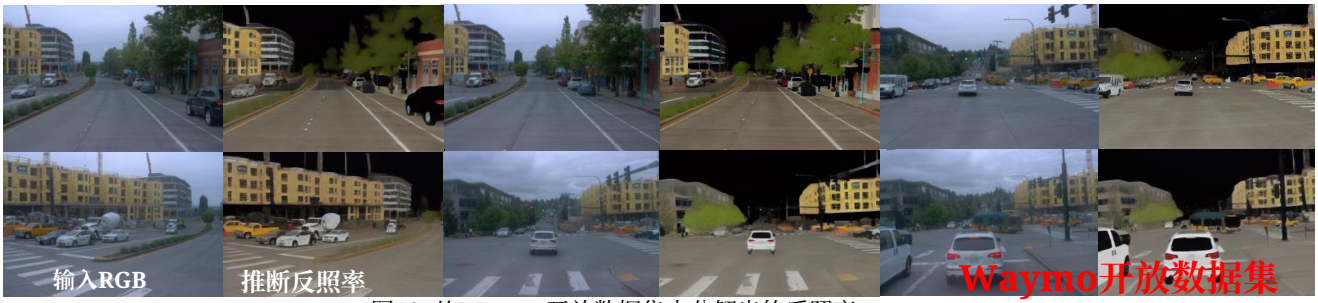


图13. 从Waymo开放数据集中分解出的反照率。

令人瞩目的跨领域泛化能力，包括法线预测等任务（例如 StableNormal [62]）。基于 DMP 这一基于扩散的模型，我们评估了室内和室外真实场景的内在分解性能。对于室内评估，我们报告了 IIW 数据集上的结果（第 4.3 节）。对于室外场景，我们使用 BigTime v1 和 Waymo Open 数据集。BigTime v1 捕捉一天中不同光照条件下的室外环境，而 Waymo Open 数据集则提供了 Waymo 自动驾驶汽车在不同光照和天气条件下采集的多样化城市场景。如图12 所示，在反照率一致性方面，使用 LightCity 进行混合微调的 DMP 平均方差最低，为 0.015；使用 MatrixCity-mix 进行混合微调的模型和纯 Hypersim 的方差则较高，分别为 0.036 和 0.042。使用 LightCity 进行混合微调的 DMP 对真实城市场景也具有良好的泛化能力，如图13所示。

此外，为了进一步缩小仿真到现实差距，生成模型也可以视为用于仿真到现实迁移的领域迁移模型。已有研究已证明，这类方法在弥合合成-真实差异方面有效 [73]。利用 img2img-turbo [44]和 InstructPix2Pix [6] 等基于扩散模型的图像到图像处理流程，为提高合成数据集在真实场景中的适用性提供了一个颇具前景的未来方向。

5. 多图像逆向渲染的更多结果

5.1. 基线方法详情

我们简要介绍用于逆向渲染的基线方法（NeRF-OSR 和 GS-IR）。

NeRF-OSR.是首个能够从多视角、多光照输入图像中学习神经表示，并显式分解场景几何、漫反射反照率和阴影的方法，从而支持更灵活的场景编辑。

GS-IR.首次将 3DGS 扩展至逆向渲染，采用基于物理的渲染框架，利用多视角采集的图像，在物体级和场景级任务中联合重建场景几何、材质属性及未知的自然光照。

5.2. 新视角合成与几何质量

我们还为多视角逆向渲染提供了几何真值，并评估了两种基线方法的几何质量。如表9所示，在新视角合成和几何重建方面，GS-IR 在城市场景逆向渲染中的表现均优于基于 NeRF 的 NeRF-OSR。

5.3. 材质估计

作为基于物理的渲染逆向渲染中的重要组成部分，GS-IR还会优化每个高斯的金属度与粗糙度属性，以产生照片级真实光照效果。因此

我们用真实属性评估分解得到的材质属性，城市逆向渲染中的材质估计精度仍有很大的提升空间。

6. 多光照户外重建的更多结果

6.1. 基线方法详情

我们简要介绍用于多光照户外重建的基线方法（NeRF-W、wild-gaussians 和 Gaussian-wild）。

NeRF-W。 通过引入逐图像学习的低维潜在外观嵌入，并利用 GLO 将其作为共享多层感知机的条件，将隐式 NeRF 扩展至无约束的多光照重建，从而将场景几何与光照不一致性解耦。**wild-gaussians**。 将显式的三维高斯泼溅（3DGS）表示应用于变化的光照条件下的真实场景重建。它采用基于多层感知机的外观建模模块，并通过仿射颜色映射捕捉依赖于图像的高斯颜色，同时保持渲染效率。

Gaussian-wild。 基于 3DGS，将每个高斯的外观分解为固有特征和动态特征，进一步增强场景局部高频变化的表现能力，从而在适应变化的光照条件的同时，更好地捕捉精细场景细节。

NexusSplats. 利用神经网络表示图像特定的全局光照条件，以及高斯对全局光照变化的特定局部响应，从而有效捕捉不同场景中的复杂光照变化。

6.2. 新视角合成

为了为我们的 LightCity 重建数据集提供基线，我们还在单光照数据集上训练了 Gaussian-wild (GS-W)。结果见表 8。与在多光照数据集上训练的模型相比，其性能有所下降，这进一步表明多光照对城市重建性能有显著影响。

在前文中，我们展示了多光照数据集测试集上的可视化评估结果。我们还在图 15中展示了单光照数据集测试集上的结果。与 3DGS 相比，采用外观嵌入建模的方法在质量上有所下降。尽管 NeRF-W 能够还原图像中的阴影（第 1 列），但其在其他未见视角下的表现仍然较差。GS-W 倾向于还原从未见过的输入真实值中更清晰的结构，但仍有

某些部分存在漂浮伪影。这进一步说明了我们的多光照重建数据集所面临的挑战。

我们还深入分析了表现最佳的 GS-W 和 NeRF-W 之间的性能差异，如图所示。 14。第一行展示了 GS-W 的细节模糊，漂浮伪影覆盖了建筑物，导致视觉伪影。第二行展示了 NeRF-W 的细节模糊，它往往会模糊复杂场景的细节。

表8. 在单光照数据集上训练的 Gaussian-Wild 在 F2区块上的新视角合成性能

F2	Gaussian-wild
峰值信噪比	28.74
SSIM	0.878
学习感知图像块相似度	0.174

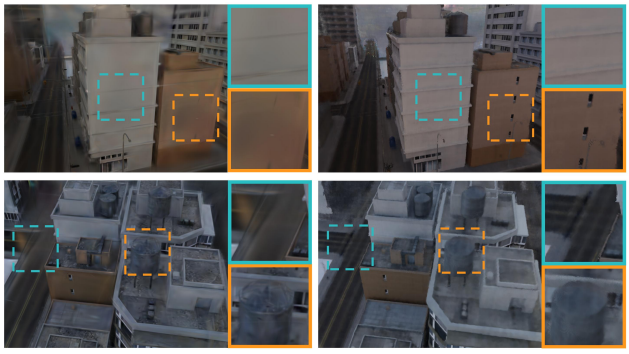


图14. Gaussian-wild 与 NeRF-W 在多光照条件下的新视角合成对比。

6.3. 几何质量

为全面评估LightCity 数据集中场景的重建几何，我们在多光照条件下渲染了所有所用方法的法线图。误差指标列于表11中。在所有区块中，Gaussian-wild 的平均绝对误差和中位数绝对误差均最低，表明与其他方法相比，其几何重建质量具有优势。然而，在

表9. 多视角逆渲染的新视角合成性能。

		数据集			
方法	指标	F2	F3	E1	E2
NeRF-OSR	PSNR	17.35	20.15	21.11	20.95
	SSIM	0.562	0.600	0.622	0.597
	学习感知图像块相似度	0.461	0.413	0.400	0.438
	平均绝对误差	32.81	31.39	30.96	33.98
GS-IR	峰值信噪比	**26.35**	**27.26**	**27.29**	**26.76**
	SSIM	**0.862**	**0.90**	**0.858**	**0.861**
	学习感知图像块相似度	**0.233**	**0.186**	**0.196**	**0.200**
	平均绝对误差	**28.07**	**23.60**	**27.73**	**28.78**

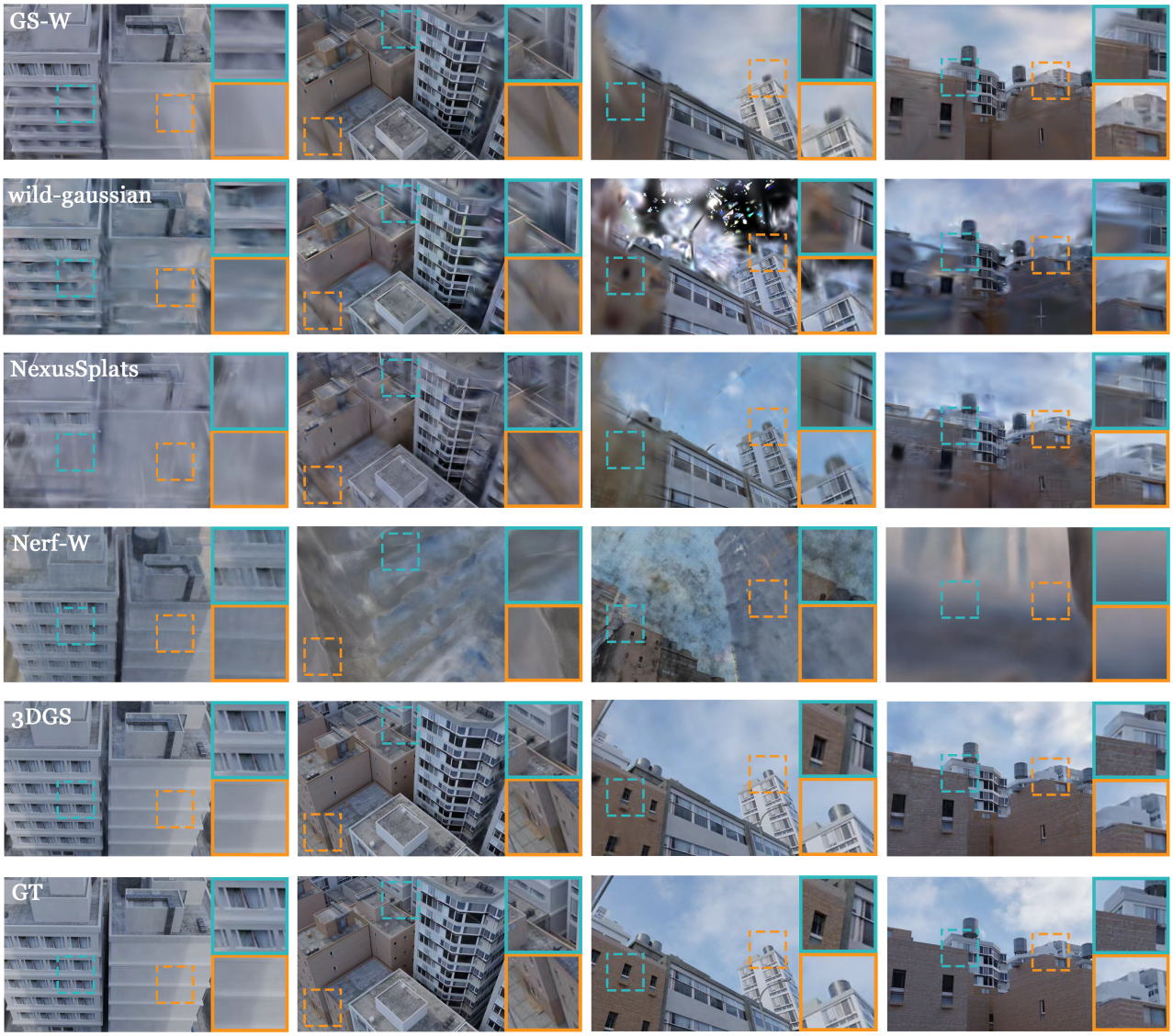


图15。单光照数据集测试集上的新视角渲染结果。

表10。多视角逆向渲染中的材质估计性能。

数据集	指标	GS-IR
A2	金属性均方误差	0.1168
	粗糙度均方误差	0.2643
A3	金属性均方误差	0.1813
	粗糙度均方误差	0.2964
B1	金属性均方误差	0.2888
	粗糙度均方误差	0.2833
B2	金属性均方误差	0.2599
	粗糙度均方误差	0.2466

多光照条件的约束下，这些方法的质量都比原始 3DGS 有所下降。此外，我们在图18中展示了不同方法的法线图。尽管 Gaussian-wild 的法线精度最高，但在一些平坦区域容易出现模糊，这可能是由于多光照输入引入了额外的漂浮伪影。不过，除了一些粗糙度问题外，NeRF-W 的法线相对清晰。这可能归因于其对光线进行离散采样。另两种基于 3DGS 的方法，即 wild-gaussian 和 NexusSplats，几乎无法重建场景法线；大面积的高斯面元覆盖了屏幕空间（第 3 列）。

此外，我们还研究了不同视角之间法线的一致性。如图17所示，NeRF-W

Ta表 11. 多光照条件下城市重建的几何质量性能比较。

数据集	指标	方法 3DGS	Gaussian-wild	wild-gaussian	NexusSplats	Nerf-w
A2	平均绝对误差 21.57	**26.91**	32.23	32.94	27.88	
	中位数绝对误差 21.81	**26.38**	31.30	31.81	27.49	
A3	平均绝对误差 23.50	**25.27**	30.85	32.83	29.81	
	中位数绝对误差 23.33	**26.00**	31.08	32.22	28.45	
B1	平均绝对误差 24.34	**31.15**	37.02	35.76	49.76	
	中位数绝对误差 23.52	**29.65**	36.31	35.45	50.55	
B2	平均绝对误差 23.99	**29.16**	40.63	37.84	37.99	
	中位数绝对误差 23.98	**28.13**	40.34	36.23	36.21	



图 16. 基于三维的重光照视觉效果。

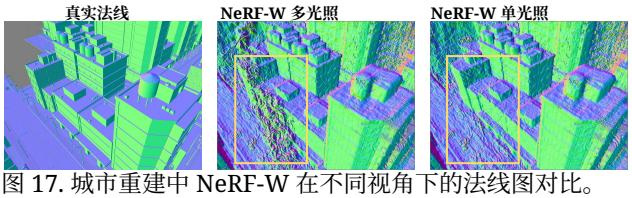


图 17. 城市重建中 NeRF-W 在不同视角下的法线图对比。

倾向于在同一场景的不同视角下重建出不同的法线图，呈现出明显的不一致性。基于高斯泼溅的方法不存在这一问题，因为它们将每个高斯的外观与位置解耦。

7. 更多重新照明结果

重新照明是计算机视觉中的一项重要且具有现实应用价值的任务，可支持内容编辑、光照迁移和场景操控等应用。在我们的多光照重建实验中，我们通过在测试光照条件下优化重建表示来执行基于 3D 的重新照明。可视化示例见图右侧。16。除了 NeRF-OSR 等由 3D 驱动的方法外，我们还评估了基于图像的重新照明技术——这类单图像模型会直接操控输入图像，使其匹配目标光照。我们评估了三种模型：IC-Light [70]，Self-OSR [65]和 ColorTransfer [28]。如图左侧所示，16，IC-Light 难以为复杂的室外场景进行重新照明，而 Self-OSR 和 ColorTransfer 的表现也有限。这些结果表明，当前基于图像的重新照明方法普遍化泛化到室外城市场景时表现不佳。因此，LightCity 有望为未来面向室外环境的基于图像重新照明研究提供支持。

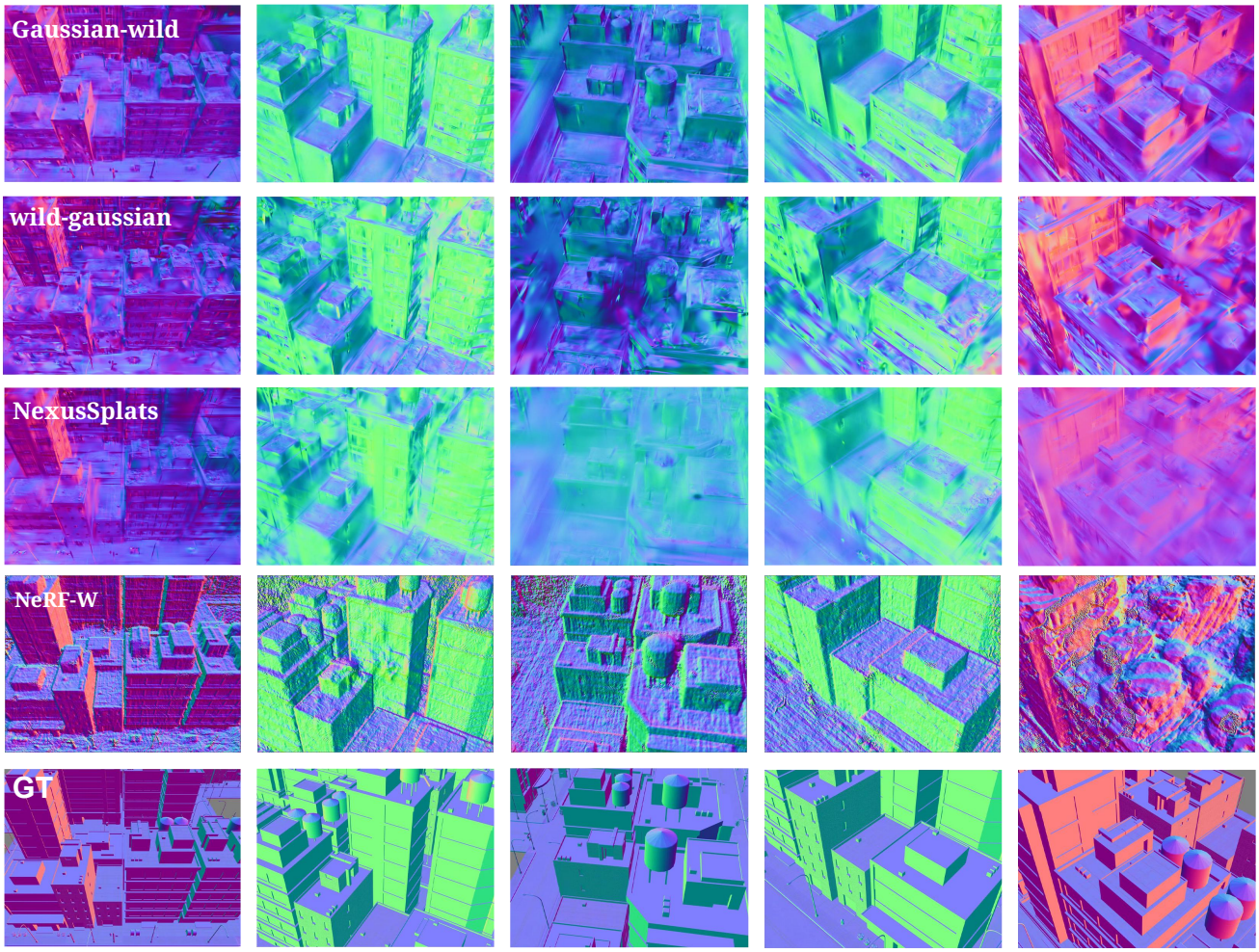


图 18. 多光照条件下城市重建的法线图可视化。