中图法分类号:TP751.1　　**文献标识码**:A　　**文章编号**:1006-8961(2025)02-0559-16

论文引用格式:Zeng Z H, Wang Z J, Zhang Y B, Cai W N, Zhang L L, Guo Y and Liu J Y. 2025. Semantic and consistent neural radiance field reconstruction method based on intrinsic decomposition via classification. Journal of Image and Graphics, 30(02):0559-0574(曾志鸿, 王宗继, 张源奔, 蔡伟南, 张利利, 郭岩, 刘俊义. 2025. 利用本征属性分类的神经辐射场视角及语义一致性重建. 中国图象图形学报, 30(02):0559-0574)[DOI:10.11834/jig.240140]

利用本征属性分类的神经辐射场视角及语义一致性重建

曾志鸿[1,2,3,4],王宗继[1,2,3],张源奔[1,2,3*],蔡伟南[1,2,3,4],张利利[1,2,3],郭岩[1,2,3],刘俊义[1,2,3]

1. 中国科学院空天信息创新研究院,北京　100190; 2. 目标认知与应用技术重点实验室(TCAT),北京　100190; 3. 网络信息体系技术重点实验室(NIST),北京　100190; 4. 中国科学院大学电子学院,北京　100190

摘　要:**目的**　基于神经辐射场(neural radiance field,NeRF)的3D场景重建与新视角生成工作正受到研究者的广泛重视,然而现有的神经辐射场方法通常对给定的场景进行高度专门化的表征,且将场景的几何与外观表征为"混合场",这对场景的几何与外观编辑、场景泛化和3D资源的使用造成了不便。**方法**　提出了一个学习对象本征属性的神经辐射场分类网络,通过图像增强的方式去除高光和阴影,并使用分类的方式实现颜色分解,即从现实场景中提取室内场景语义级目标的本征属性,在此基础上进行神经辐射场的重建。提出了前点优胜模块与颜色分类模块。前点优胜模块在体渲染阶段优化射线代表的本征属性,从而提升神经辐射场的语义一致性;颜色分类模块在辐射场重建阶段,通过全连接网络进行本征属性的分类优化,提高辐射场的语义及视角间一致性。两个主要模块共同作用,使重建的辐射场具备良好的针对外观的泛化能力,可支持场景重上色、重光照以及针对阴影与高光的编辑等任务。**结果**　相比于现有的基于神经辐射场的学习进行本征分解的Intrinsic NeRF方法,在Replica数据集中的充分实验表明,在有限的GPU显存和运行时间下,重建的本征属性神经辐射场具备语义及视角间一致性。针对提升语义一致性的前点优胜模块,本文方法在基线模型Semantic NeRF的基础上提高了4.1%,在未加入该模块的基础上提高了3.9%。针对提升本征分解语义及视角间一致性的颜色分类模块,本文方法在Intrinsic NeRF的本征分解工作基础上提升了10.2%,在未加入颜色分类层的基础上提升了1.7%。**结论**　本文方法构建的本征属性神经辐射场具备语义及视角间一致性,可描述复杂场景几何关系且具备良好外观泛化性。在场景重上色、重光照、阴影与高光的编辑等任务中取得了视角间一致的逼真效果。

关键词:图像处理;场景重建;神经辐射场(NeRF);本征分解;场景编辑

Semantic and consistent neural radiance field reconstruction method based on intrinsic decomposition via classification

Zeng Zhihong[1,2,3,4], Wang Zongji[1,2,3], Zhang Yuanben[1,2,3*], Cai Weinan[1,2,3,4], Zhang Lili[1,2,3], Guo Yan[1,2,3], Liu Junyi[1,2,3]

1. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing　100190, China; 2. Key Laboratory of Target Cognition and Application Technology (TCAT), Beijing　100190, China; 3. Key Laboratory of Network Information System Technology (NIST), Beijing　100190, China; 4. School of Electronic, Chinese Academy of Sciences, Beijing　100190, China

收稿日期:2024-03-09;**修回日期**:2024-04-22;**预印本日期**:2024-04-29

*通信作者:张源奔　zhangyb@aircas.ac.cn

基金项目:基础加强计划技术领域基金项目中科院A类先导专项课题级智能信息系统数字孪生技术(2023-JCJQ-JJ-0760)

Supported by:Foundation Strengthening Program Technology Field Fund Project (2023-JCJQ-JJ-0760)

Abstract: Objective Reconstruction of indoor and outdoor 3D scenes and placement of 3D resources in the real world constitute important development directions in the field of computer vision. Early researchers used voxel, occupancy, grid, and other computer graphics representation methods to achieve good results in terms of storage and rendering efficiency in a variety of mature application areas. However, these methods require time-consuming and laborious manual modeling, experienced modelers, and considerable time and energy. The time-consuming, laborious, and difficult modeling process must be simplified to enhance the application prospects in the 3D scene reconstruction field. By invoking and calculating the implicit field representation, researchers can obtain a realistic scene end-to-end, eliminating the complicated process of traditional modeling. The neural radiance field (NeRF) is the most popular implicit field representation method. Compared with other implicit field methods, the neural implicit field is known for its simplicity and ease of use, but its problems still exist and are rooted in the defects of the implicit field itself. The implicit field is a multidimensional function defined on spatial and directional coordinates, which codifies the geometry of the scene together with the appearance color, resulting in the entangled representation of the independent attributes of the target, causing inconvenience to the application of 3D resources. An important direction regarding implicit fields is "disentanglement" between geometry and appearance. First, the intrinsic decomposition uses some physical priors to avoid the initialization of complex networks. Second, the image is preprocessed into an albedo image independent of the observation direction and a shading image dependent on that. Intrinsic NeRF was the first to apply intrinsic decomposition methods in NeRF, but the decomposition they have used could not produce more reasonable appearance editing results. **Method** In this study, a NeRF classification network is proposed to learn the intrinsic properties of objects and target characteristics. It separates specular factors from 2D images via the image enhancement method, extracts the intrinsic colors (performs intrinsic decomposition) via the classification method, and then presents shading maps and direct illuminations of semantic-level objects in scenes via intrinsic decomposition expression. On this basis, the NeRF is learned, and the semantic consistency is provided with the help of "the front-point dominance module", which is a module from the volume rendering stage that optimizes the albedo by "front points". The consistency between views of the scene is provided with the help of "the color classification layer module", which is a fully connected neural network from the reconstruction stage that fixes the albedo between different perspectives. Finally, a neural radiation field representing the intrinsic properties of the scene is reconstructed. After the rays are obtained by the internal and external parameters of the image, the positions of the sampling points and the directions of the rays are calculated in the neural network, producing the corresponding 3D properties. In the embedding layer, the position and direction are embedded and transformed into high-dimensional embedding features, which are the inputs of the network. After the 8-layer fully connected multilayer perceptron, the network outputs a 1-dimensional volume density, a 256-dimensional feature vector, and an N-dimensional semantic vector (where N is the number of semantic classes). The 256-dimensional feature vectors are then input into each 1-layer fully connected network to obtain color, a shading map, and direct illumination. In the inference stage, the model uses the Monte Carlo integral method to transform properties based on sampling points into properties based on rays (i. e., pixels), resulting in a synthetic result of the novel view. The model can disentangle attributes independent of and dependent on the observer. The resulting albedo output has semantic and multiview consistency independent of the observation direction. The implicit field shows good generalization for appearance and supports scene recoloring, relighting, and editing for shadows and specular factors. **Result** Compared with the existing Intrinsic NeRF method for intrinsic decomposition based on NeRF learning, experiments on the Replica dataset show that under limited GPU memory and running time, this work can obtain intrinsic decomposition results with semantic and multiview consistency. For the "front-point dominance module", which ensures semantic consistency, this work improves the performance by 4. 1% compared with that of the Semantic NeRF. The ablation study revealed an improvement of 3. 9% over the baseline model. For the "color classification layer module", which improves semantic multiview consistency, this work improves the Intrinsic NeRF's intrinsic decomposition method by 10. 2% and the baseline model by 1. 7%. **Conclusion** A novel NeRF classification network that can learn the intrinsic properties of objects and target characteristics is proposed. Experiments show that the work in this study can produce intrinsic decomposition results with semantic and multiview consistency. Moreover, an implicit field of albedo classification is constructed, which can describe the geometric relationship of complex scenes and shows good generalization for appearance. Realistic and multiview-consistent effects are achieved in

tasks of scene recoloring, relighting, and shadow and specular factor editing.

Key words: image processing; scene reconstruction; neural radiance field(NeRF); intrinsic decomposition; scene editing

0 引　言

在现实世界重建室内外三维场景并放置三维资源是计算机视觉领域的重要发展方向。研究者在早期使用占用率网格(Mescheder等,2019)、多边形网格(Wang等,2018)和点云(Insafutdinov和Dosovitskiy,2018)等图形学表示方法,在存储和渲染效率上已经取得了良好的成果,有成熟的应用领域,但这些方法需要耗时费力的人工建模,需要经验丰富的建模师和大量的时间精力。新兴的辐射场表示方法(Sitzmann等,2019)同样具备良好的计算性能和存储性能,更简化了门槛颇高的建模过程,从而大幅提升了三维领域的应用前景。例如,虚拟现实场景浏览、三维动画CG、游戏和仿真模拟等三维应用可以使用范围更大更精细的模型。通过辐射场表示的调用和解算,研究者可以端到端地获得一个逼真场景,省去了一些传统建模的复杂工序。神经辐射场(neural radiance field,NeRF)(Mildenhall等,2022)是辐射场表示方法中最广泛流行的基础工作,该工作利用多视角图像的内外参数计算成像平面和相机在世界坐标系下的分布,通过发射光线并索引神经网络的方式从空间中获得采样点对应的体密度和颜色。在推理阶段,射线行进(Kajiya和von Herzen,1984)方法支持网络通过一维射线上的采样点的蒙特卡罗式(Porter和Duff,1984)计算对应像素的颜色,从而产生新的合成视角图像。通过立方体行进算法(Lorensen和Cline,1987),研究者可以基于辐射场给出的体密度提取目标的三维表面,从而得到单纯的三维重建结果。

神经辐射场仍然存在一些缺陷限制了领域的发展。辐射场是空间坐标和方向坐标的多维函数,它将场景或者目标的几何形状与外观颜色一同编入,造成了目标属性的相互耦合,这对辐射场为基础的三维资源使用造成了不便。关于辐射场一个重要的方向就是几何与外观的“解耦(disentanglement)”问题(Carbonneau等,2024)。类似于本征分解,这也是一个逆渲染问题。一个解决的方式就是几何与外观在辐射场中的分离,IDR(implicit differentiable renderer)(Yariv等,2020)借助符号距离场(signed distance field, SDF)(Oleynikova等,2016)将多维函数分解为几何与外观的函数,从而分别计算它们的几何与外观;NeuS(neural implicit surface)(Wang等,2021)将SDF扩展到NeRF的方法重建几何;NeRD(neural reflectance decomposition)(Boss等,2021)、NeRFactor(Zhang等,2021b)等方法则能单独重建外观,将外观分解为材质和光照等属性。这种方法的问题一方面在于需要合理的初始化,或者添加物理先验的预训练模型;另一方面在于通过这种方式得到的几何与外观并不是相互独立的,存在一定的联系:例如物体的几何与物体的材质均为独立于观察方向的属性。通过基于本征分解的辐射场重建可以避免这些问题。首先,本征分解应用了一定的物理先验,从而避免了复杂网络的初始化工作;其次,本征分解方法将图像预处理为独立于观察方向的albedo与依赖观察方向的shading。Intrinsic NeRF(Ye等,2023)最早尝试在NeRF中应用本征分解方法,然而其分解并未考虑阴影的分离与纹理的提取,无法产生更合理的外观编辑结果。另一方面,该工作将分解过程放在NeRF的训练流程中,降低了辐射场的重建效率。

本文提出了一个学习对象本征属性的神经辐射场分类网络,通过图像增强的方式去除高光和阴影,并使用分类的方式从现实场景中提取室内场景语义级目标的本征属性,在此基础上进行神经辐射场的重建。本文提出了前点优胜模块与颜色分类模块。前点优胜模块在体渲染阶段优化射线代表的本征属性,从而提升神经辐射场的语义一致性;颜色分类模块在辐射场重建阶段,通过全连接网络进行本征属性的分类优化,提高辐射场的语义及视角间一致性。两个主要模块共同作用,使重建的辐射场具备良好的针对外观的泛化能力,可支持场景重上色、重光照以及针对阴影与高光的编辑等任务。实验证明,本文方法构建的本征属性神经辐射场具备语义及视角间一致性,可描述复杂场景几何关系且具备良好外观泛化性。在场景重上色、重光照、阴影与高光的编辑等任务中取得了视角间一致的逼真效果。

1 相关工作

1.1 神经辐射场的解耦

神经辐射场方法利用多视角图像的内外参数计算成像平面与相机在世界坐标系下的分布，通过发射光线并索引神经网络的方式从空间中获得采样点对应的体密度和颜色。相比其他的图形学表示方法辐射场的优势在于方法的简洁、高效，但辐射场的泛化性能极差，通过大量神经网络单元构成的函数在高维空间存在严重的稀疏性，无法找到一个控制辐射场的主成分来显著地变化其特定的几何、外观等属性，这使得对辐射场的编辑成为一个近乎不可能完成的任务。一个解决方案是对几何与外观的解耦，例如Conditional NeRF（Liu等，2021）、NSVF（neural sparse voxel fields）（Liu等，2020b）等工作，但其缺陷在于需要合理的初始化以及大量的物理先验，且解耦得到的外观同时具备依赖视角方向和独立视角方向的属性，缺乏泛化性。其他工作诸如iNeRF（Yen-Chen等，2021）、NeRFactor（Zhang等，2021）、PhySG（physics-based sphere Gaussian）（Zhang等，2021a）等启发于逆渲染方法，将辐射场分解为几何、材质、光照等相互独立的属性，然后通过基于物理的渲染（physics based rendering）（Pharr和Humphreys，2010）的流程实现外观的渲染。本征分解同样是逆渲染的一种，相比于其他的逆渲染方法，本征分解能够产生独立视图和依赖视图的属性，而这更符合NeRF的多阶段输出流程，正是本文方法利用本征分解实现属性解耦的原因。

1.2 神经辐射场的本征分解

对本征分解的研究在1971年开始，Harrys教授根据人类的视觉机制提出了Retinex理论（Land和McCann，1971），该理论认为图像可以分解为梯度变化剧烈的不同色块，以及梯度变化缓慢的照明与阴影效果。后续的研究简化了Harrys教授的模型并将其应用于Lambertian世界，即为Narihira等人（2015）和Zhou等人（2015）提出的本征扩散模型。模型将图像分解为梯度变换剧烈的albedo和梯度变化缓慢的shading。

相比简单的Lambertian世界，一种颜色在图像中的表现，其涉及的参数与物理影响的复杂程度要远远超出这种简单模型的模拟，需要一种更复杂的分解方式，能够同时满足分解的准确性与面对不同环境条件的通用性。Seitz等人（2005）、Dong等人（2015）和Ye等人（2014）假设图像可以被分解为光线的多次反弹结果的加和，也就是本征残差模型。光线的第1次反弹被视为是直接光照，而其余的高次反弹结果被认为是全局照明效应。依据这个设想，研究者在本征分解的分解方程中添加了第3项残差层，以表征直接光照的光学效应。

在神经辐射场诞生后，也有研究者尝试通过将这种模型引入辐射场的重建流程，Intrinsic NeRF（Ye等，2023）将NeRF方法与本征残差模型结合，从而在三维空间中分离了反射层、照射层和残差层属性，但其残差层并未应用光照相关的物理知识先验，而是采用了阶段性的学习策略。本文沿袭Intrinsic NeRF的思路，但并未在重建过程中逐渐获取本征属性，而是在预处理阶段完成后通过网络进行学习。

除了分解数值模型的改进，后续还有研究者基于深度学习网络进行通用的本征分解。例如王玉洁等人（2022）提出了基于深度学习的本征图像分解方法，在合成数据集上获得了分解纹理和光照并进行图像编辑的能力。不过本文方法提出的本征分解进行的是多视角渲染任务，不需要特别精确的通用分解结果，因此仍然采用无监督聚类的方式。

本文采用了本征残差模型，将图像分解为反射率（albedo）、照射率（shading）和高光（specular）。为保证后续用词的一致，此后使用本文方法进行的本征分解，产生的3个相关结果均使用英文单词写法。albedo通过无监督聚类的方式，将颜色空间中混杂的像素点根据颜色分解为不同的色块；shading和specular则是通过Gamma变换的方式，先从复杂场景中提取到高光的准确位置，然后根据本征分解的模型得到shading和specular的模拟结果。

2 方 法

2.1 方法总述

给定一组带位姿的图像，目的是提取二维图像的本征属性，并将这些属性推广到三维空间的场景中，从而形成三维的本征属性辐射场重建和视图渲染结果。针对二维图像，通过Gamma变换进行图像增强，从而优化图像中亮度过高或者过低的区域，将高光与阴影部分从图像中提取出来。在不属于高光

或者阴影的部分，以目标所属的语义类别为输入，通过分类的方式进行颜色的本征分解，从而保证颜色分解在像素级上的分解精度；同步计算所有视角以保证解析结果在语义及不同视角间的一致性（本征图像具有独立于观察方向和位置的一致性）。图像分解具体为

$$\boldsymbol{I}(x,y)=\boldsymbol{A}(x,y)\times\boldsymbol{S}(x,y)+\boldsymbol{Re}(x,y) \tag{1}$$

式中，$\boldsymbol{I}$、$\boldsymbol{A}$、$\boldsymbol{S}$与$\boldsymbol{Re}$分别代表原始图像、albedo、shading和specular。本征属性提取阶段并未涉及对通用数据的分解，所提取的属性是为了后续的辐射场重建阶段获得对应的参考。

在二维本征属性逆渲染三维辐射场的阶段，本文方法提出了一个新的“神经辐射分类场”，在原始的NeRF的基础上，在辐射场输出空间采样点颜色c时，增加一个分支以区分多种albedo颜色的类别，对不同类别的概率值进行监督。神经辐射分类场的“分类”过程等效于Intrinsic NeRF（Ye等，2023）中，通过mean-shift算法对颜色空间中的所有像素颜色进行无监督聚类的过程，本质上是对无监督聚类过程的网络参数化学习，因此可以等价于对颜色空间中所有像素点对应颜色标签的语义分割任务。类似的思路体现在宁小娟等人（2023）提出的方法中，该方法结合语义分割将语义标签聚类并匹配现有模型，从而实现室内场景重建的方法，证明了借助聚类二维信息进行的网络化参数学习在三维领域具有广泛的应用。

但是相比于Intrinsic NeRF，原本的无监督聚类过程需要在每一次通过网络获得射线颜色之后进行一遍，需要消耗大量的时间，而本文方法则是通过一劳永逸的参数化网络表征，将聚类过程抽象为一层分类场网络的学习，在优化方法流程的同时，也大幅缩短了训练和测试的时间，具体参见第3节的实验部分。

与其他基于NeRF的逆渲染方法不同，本文方法的神经辐射分类场根据是否独立于视角来得到对应的输出。独立于视角的输出即albedo，依赖于视角的输出包括shading与specular。这种方式有效地利用了NeRF分阶段引入参数与输出的特性。

体渲染阶段，在射线的积分上，本文方法采用“前点优胜”的策略，为射线所经过的透过率小于一定阈值的第一个采样点分配90%的权重，其他后向的所有点分配10%的权重，从而保证提取的射线颜色分类概率分布的相对稳定。在训练理想的情况下，本文方法能构建一个可分离不同本征属性的辐射场，以供后续的使用者推断阴影的位置与方向，从而间接推导场景的光照信息。本文方法总体流程如图1所示。

2.2 基于对比度增强消除图像阴影与高光

要通过无监督聚类分解本征属性，需要首先将颜色变换到HIS（hue saturation and intensity）颜色空

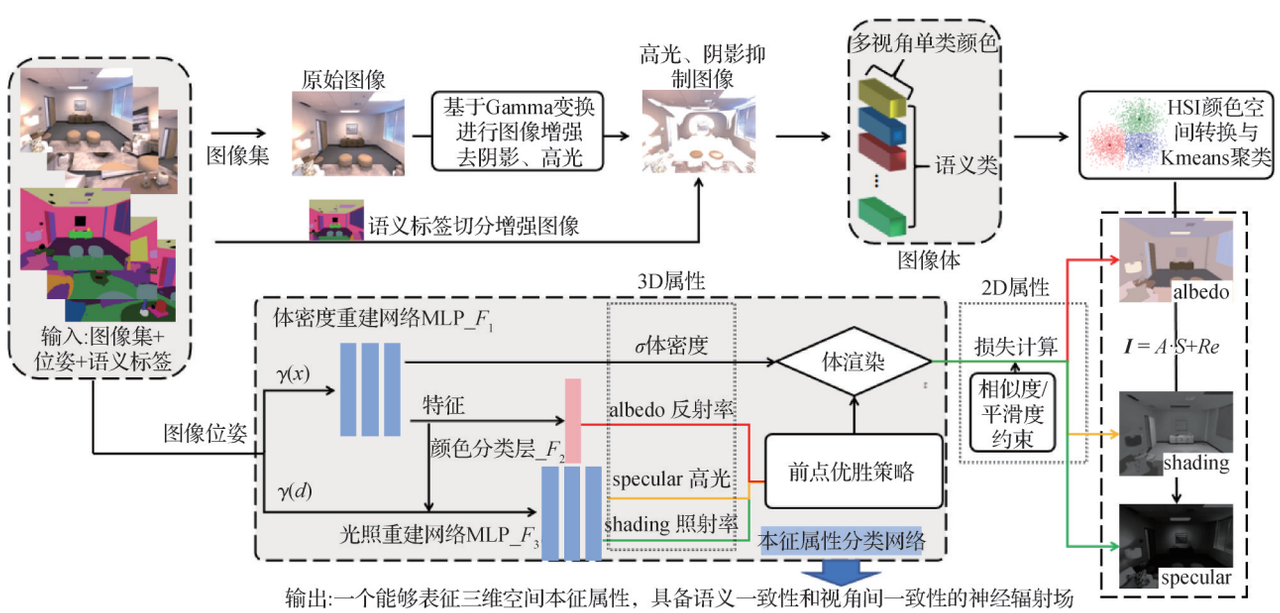


图1 方法总体流程

Fig. 1 Overview of method and process

间。然而在实验中发现，HSI颜色对于受到阴影和高光影响的物体颜色存在色相的偏移，这将导致其分解结果中，受到高光或阴影影响物体的颜色的albedo和shading结果归类到新的颜色中，影响分解的效果。一个可行的做法是在得到HSI颜色前先将阴影和高光部分的颜色去除，在得到聚类结果后，再将相近的albedo和shading赋值给去除的部分。

要得到去掉阴影与高光的本征图像，可以通过图像增强的Gamma变换方法增加图像的对比度，从而将图像的阴影与高光部分从场景中提取出来。Gamma变换提取阴影和高光如图2所示。

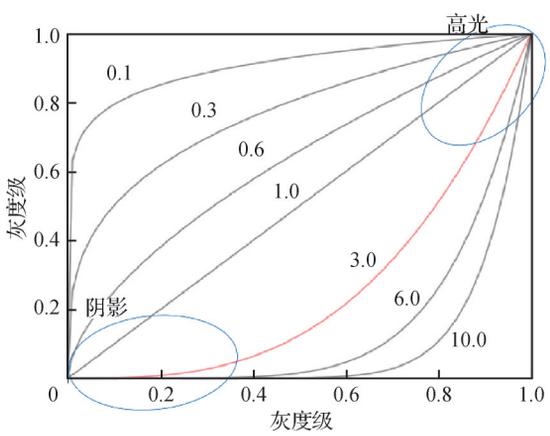


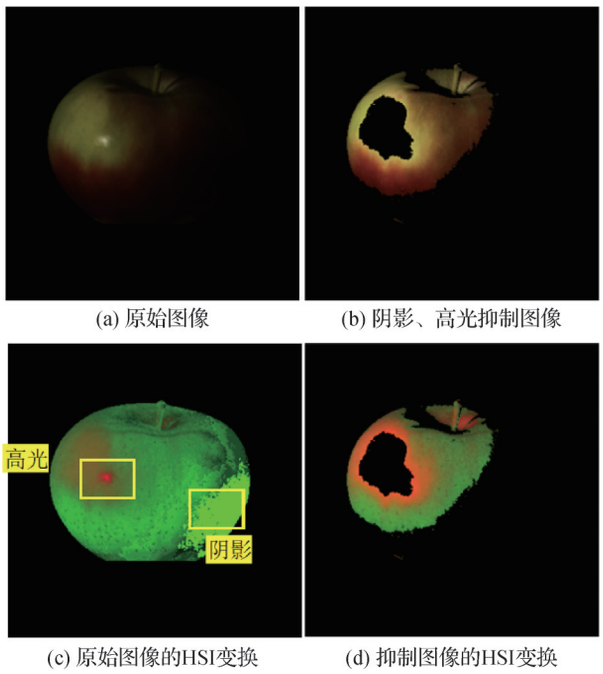
图2　利用Gamma变换提取阴影和高光，图像对比度增强，更加突显非阴影和非高光的颜色

Fig. 2　Using the Gamma transform to extracts shadow and specular factor, the image contrast is enhanced, so that non-shadowed and non-highlight colors are more prominent

图3(a)(b)展示了本文通过Gamma变换去除阴影和高光属性之后，用于颜色分类和本征属性提取的部分在MIT Intrinsic数据集上的简单物体进行展示。

物体颜色受到高光和阴影的影响后，不仅会在亮度上改变，同时也会改变色调。为了在HSI颜色空间中分离本征属性，进行聚类，需要尽可能排除亮度对色调产生的影响，以免将受到光照“调亮”或者“调暗”的同一颜色区域归类到不同的颜色类别中。通过图3(c)(d)展示的HSI变换效果可以看出，在排除了亮度对色调的影响之后，原本的albedo可选范围更加精确了。对于去除的部分，聚类得到albedo和shading后，根据与原像素的向量欧氏距离计算相似性，将聚类结果赋值到去除区域的像素上。

给定图像$\boldsymbol{I}\in\mathbf{R}^{H\times W\times 3}$，对比度增强图像$\boldsymbol{G}$计



(a) 原始图像　(b) 阴影、高光抑制图像

(c) 原始图像的HSI变换　(d) 抑制图像的HSI变换

图3　利用Gamma变换从原始图像提取非阴影、高光颜色及其HSI变换关系

Fig. 3　Using the Gamma transform to extracts color and HSI transformed color without shadow and specular factor from original image ((a) original image; (b) shadow and specular suppressed image; (c) HSI transformation of original image; (d) HSI transformation of suppressed image)

算为

$$\boldsymbol{G}(x,y)=\frac{\boldsymbol{I}(x,y)^{\gamma}}{\max_{(x,y)}\{\boldsymbol{I}(x,y)\}} \tag{2}$$

式中，$\boldsymbol{I}(x,y)$是图像的RGB三通道像素值，其取值归一化到0～1。γ为伽马因子，由于是为了增加对比度，γ值大于1，设定为3。Gamma变换后，图像接近0的部分一般表现为阴影区域，接近1的部分一般表现为高光，方法利用一个预设的权重阈值判断是否属于高光或者阴影的部分，计算为

$$\boldsymbol{diff}_{\text{obj}}(x,y)=\boldsymbol{G}_{\text{obj}}(x,y)-\frac{1}{N}\boldsymbol{I}_{\text{obj}}(x,y) \tag{3}$$

$$\boldsymbol{W}_{\text{obj}}(x,y)=\exp(-\lambda_1(\left|\boldsymbol{diff}_{\text{obj}}(x,y)-w\right|+\left|\boldsymbol{diff}_{\text{obj}}(x,y)+w\right|-2w)) \tag{4}$$

式中，$\boldsymbol{G}$是经过Gamma变换的图像，$\boldsymbol{diff}$是差值图，用于表示经过Gamma变换后的图像和原始图像的差值。根据图2所展示的趋势，显然$\boldsymbol{diff}$在高光和阴影部分的差值相比其他部分是更大的。$\boldsymbol{W}$是绝对值处理和平滑处理后的差值图像，用介于0到1之间的权重表示针对高光和阴影区域的关注，越接近1，表

示越表现为高光或者阴影区域，以支持后续计算。N表示隶属于语义标签obj的图像集合的像素总数乘以3个通道数，λ_1是设定的参数，实验中为10。实验中根据经验设定了阈值w为0.25，从而在Gamma图像中抑制阴影与高光部分，此设定能较好地体现非阴影、非高光的颜色。

本文方法得到对应于图像集合大小的权重后，用阈值$w_0 = e^{-1}$来选取颜色，再将所有权重大于w_0的RGB颜色进行无监督聚类，从而排除阴影与高光的影响。当然，简单地通过Gamma变换去掉的高光相对更加显著，而对应的阴影就有混淆的风险。在实际处理过程中可能存在“黑色区域”和“阴影”混淆的情况，导致部分黑色区域也随着阴影去除被归类到其他的albedo颜色类之中，例如在本文的多视角抽样展示三维场景实验中发现，一些黑色区域被归入到其他颜色类中，详细情况在实验部分说明。这是本文在进行针对阴影的去除过程中的一个不足之处，将在方法局限中对其进行讨论。

2.3 辐射场重建和体渲染

本文方法构建了一种辐射场表示，可以将albedo、shading和specular的解码整合到神经网络中，并通过进行分类任务的解码器确保归属同一本征属性的点的相似性与语义及视角间一致性。如图1的辐射场分类网络所示，方法首先对射线上的点进行采样，通过体密度重建网络“MLP_F_1”后得到体密度σ和对应射线点的256维特征，然后在“颜色分类层_F_2”通过一层全连接网络将特征解码为一个附加在语义标签上的多种颜色的分类结果，颜色标签在预处理阶段得出。体密度重建网络“MLP_F_1”表示为$\boldsymbol{F}_1(\gamma(\boldsymbol{x}))$，“颜色分类层_$\boldsymbol{F}_2$”表示为$\boldsymbol{F}_2(\boldsymbol{f})$。

方法对空间坐标$\boldsymbol{x}$进行编码，以获得高维嵌入表示，具体为

$$\sigma, \boldsymbol{f} = \boldsymbol{F}_1(\gamma(\boldsymbol{x})) \tag{5}$$

$$\boldsymbol{a}_{\mathrm{sl}} = \boldsymbol{F}_2(\boldsymbol{f}) \tag{6}$$

式中，$\boldsymbol{f}$为256维的特征，σ为体密度，$\boldsymbol{a}_{\mathrm{sl}}$为长度为5的向量，对应于所有颜色类别的概率值（这里是针对单独某个语义类给出的颜色），表示albedo的分类。当颜色数量少于5个时，多余的部分概率值设定为0。接下来网络对方向坐标$\boldsymbol{d}$进行编码，然后通过“MLP_F_3”，即光照重建网络，用$\boldsymbol{F}_3(\gamma(\boldsymbol{d}),\boldsymbol{f})$表示，具体为

$$\boldsymbol{s}, \boldsymbol{re} = \boldsymbol{F}_3(\gamma(\boldsymbol{d}), \boldsymbol{f}) \tag{7}$$

式中，$\boldsymbol{s}$表示shading，$\boldsymbol{re}$表示specular。体渲染阶段，在射线上获得采样点的albedo、shading和specular后，以σ为基础进行蒙特卡洛积分（Porter和Duff，1984），具体为

$$\delta_k = t_{k+1} - t_k, \quad \alpha_k = 1 - e^{-\delta_k \sigma(t_k \delta_k)} \tag{8}$$

$$T(\delta_k) = \prod_{k=0}^{k-1}(1 - \alpha(\delta_k)) \tag{9}$$

式中，$T(\delta_k)$表示射线在k点的透过率。此步骤采用一种“前点优胜”的策略，也可以视为某种抑制射线后端受遮挡部分的掩膜。计算机在射线上判定是否经过物体并掩盖物体内部采样点颜色的表达，从而防止部分落在物体内部的采样点由于难以判定颜色类别的归属而对真实的颜色造成影响，本文方法通过深度得到相应的k_0点，从而得到该点的各项属性，“前点”分布如图4所示。

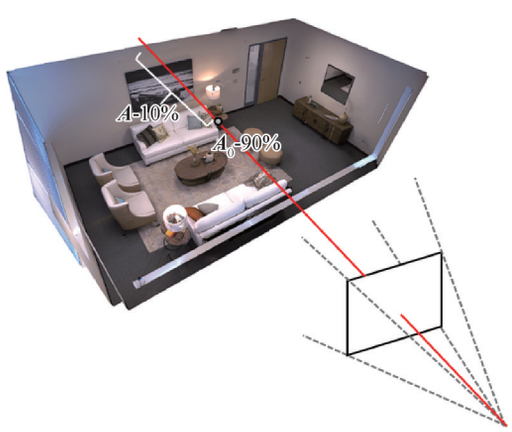


图4 前点在射线上的分布，根据深度计算射线经过首个体密度高于某一阈值的点为前点

Fig. 4 Front point distribution along the rays, based on the depth, the front point are calculated by the density, which is higher than a certain threshold along the ray

射线颜色计算为

$$\boldsymbol{A}_0 = \frac{1}{3}\sum_{k=k_0-1}^{k_0+1} \boldsymbol{a}_{\mathrm{sl}}(t_k) \tag{10}$$

$$\boldsymbol{A} = \lambda_2 \boldsymbol{A}_0 + (1 - \lambda_2 \sum_{k=0}^{K} T(\delta_k)\alpha(\delta_k)\boldsymbol{a}_{\mathrm{sl}}(t_k)) \tag{11}$$

式中，k_0是射线经过物体表面的第1个点，λ_2是“前点优胜”策略的超参数，在训练的开始阶段，λ_2为0，表明此刻正常训练。在50 000个迭代之后，σ的分布趋于稳定，此时将λ_2设定为0.9，表示前点的颜色贡献了射线颜色的90%。

2.4 相似度约束和平滑度约束

根据Retinex理论(Land和McCann, 1971),图像的分解包含许多不同的层次。后续研究者选择了两个最重要的层次,即反射层与照射层,以简化问题。在大多数相关工作(Baslamisli等, 2018; Fan等, 2018; Liu等, 2020a; Luo等, 2020)中,反射层经常描绘图像中的剧烈梯度变化,并且在场景中的照明条件变化时保持不变;阴影层经常描绘图像中的缓慢梯度变化,这些变化发生在目标的阴影边缘或纹理处,并指示目标的照明条件,这就是本征扩散模型。参考模型,在排除了阴影与高光之后,计算机可以通过albedo的相似度约束反射层,通过shading的平滑度约束照射层来获得简化的分解结果。以上所述均是在二维图像上进行的本征分解。但是由于神经辐射场可微渲染的性质,可以通过类似的方法,在已有本征分解结果的情况下将这些方法应用到三维的情况,并实现从任意时空角度对分解结果的观察浏览。这种从二维属性学习三维的思路在万骏辉等人(2024)提出的GANet中也有体现:该方法首先利用目标检测方法在二维图像上提取检测框,然后借助空间变换和网络的学习将二维检测框扩展成为三维点云的网格。因此本节针对辐射场重建的体渲染环节,采用相似度和平滑度约束,将二维图像的分解精确地扩展到三维空间。

对于相似度,在本文方法中通过颜色分类层$F_2(f)$自动地实现了三维空间的albedo属性的类内相似度约束与视角间相似度约束,损失函数为

$$L_a = \mathrm{CE}(\hat{\boldsymbol{A}}(\boldsymbol{r}), \boldsymbol{A}(\boldsymbol{r})) \tag{12}$$

式中,$\boldsymbol{r}$表示射线,CE是交叉熵(cross entropy)损失函数,$\hat{\boldsymbol{A}}(\boldsymbol{r})$是预测值,$\boldsymbol{A}(\boldsymbol{r})$是通过预处理得到的albedo分解结果的标签。

类似地,本文方法也用特殊的策略实现shading属性在三维空间中的平滑度约束,首先是横向约束。光照重建网络$\boldsymbol{F}_3(\gamma(\boldsymbol{d}), \boldsymbol{f})$的输出在三维空间中的分布是连续的,因此计算机可以在反射层梯度变换剧烈的部分分配一个更大的权重来控制$\boldsymbol{F}_3(\gamma(\boldsymbol{d}), \boldsymbol{f})$关注横向的连续性,保证该部分照射层分布的连续性,具体为

$$\omega_{\mathrm{as}}(x, y) = \exp\left(-\alpha_{\mathrm{as}} \left\| \hat{\boldsymbol{a}}_{\mathrm{sl}}(x) - \hat{\boldsymbol{a}}_{\mathrm{sl}}(y) \right\|_2^2 \right) \tag{13}$$

式中,$\hat{\boldsymbol{a}}_{\mathrm{sl}}$是对反射层分类概率值$\boldsymbol{a}_{\mathrm{sl}}$的归一化结果,$\alpha_{\mathrm{as}}$是人为设定的参数,在实验中设定为30。通过权重$\omega_{\mathrm{as}}(x, y)$来控制模型对shading属性训练的关注,具体为

$$L_{s-1} = \sum_{\boldsymbol{r}} \omega_{as}(x, y) \left\| s(x) - s(y) \right\|_2^2 \tag{14}$$

式中,x, y都是对射线$\boldsymbol{r}$采样点的选取方式,用二维坐标的方式表示$s(x)$和$s(y)$是不同射线上的x采样点和y采样点通过网络,预测得到的标量值。其次是纵向约束。方法同上,光照重建网络$\boldsymbol{F}_3(\gamma(\boldsymbol{d}), \boldsymbol{f})$在射线上寻找体密度变化剧烈的部分并分配相应权重,以控制$\boldsymbol{F}_3(\gamma(\boldsymbol{d}), \boldsymbol{f})$关注照射层在三维空间中的纵向连续性,具体为

$$\omega_{\sigma s}(x, y) = \left\| \exp(-\alpha_{\sigma s} \times \sigma(x)) - \exp(-\alpha_{\sigma s} \times \sigma(y)) \right\|_2^2 \tag{15}$$

$\alpha_{\sigma s}$在实验中设定为30,通过权重$\omega_{\sigma s}(x, y)$来控制模型对shading属性训练的关注,具体为

$$L_{s-2} = \frac{1}{N} \sum_{\boldsymbol{r}} \sum_{(x,y)} \omega_{\sigma s}(x, y) \left\| s(x) - s(y) \right\| \tag{16}$$

式中,N是射线的总数,x, y都是对射线$\boldsymbol{r}$采样点的选取。

3 实 验

3.1 数据集

本文实验在Replica数据集(Straub等, 2019)和MIT Intrinsic数据集(Grosse等, 2009)上进行。Replica数据集是大规模室内合成数据集,该数据集包括了酒店、公寓、房间和办公室共90个场景,每个场景包含一个可以浏览的三维合成场景,用于本文方法的室内复杂场景本征分解及辐射场重建。采用Semantic NeRF(Zhi等, 2021)给出的数据,在room_0场景与office_3场景中获取900个不同的视角扫描得到的图像,训练集、验证集的划分方式也与该方法相同。由于神经辐射场"一个场景对应一个神经网络"的性质,基于Replica数据集的各类实验中,训练和测试均在同一场景下进行,只是根据训练的原则分出其中80%的图像做训练,以及20%不重复的图像进行测试。实验过程中表格内数据均在20%不重复图像组成的测试数据范围内得到,并未违反计算机训练和测试原则。MIT Intrinsic数据集是小型单目标通用的本征分解数据集,包括20个受到固定方向光照影响的简单物体,用于验证本文方法采用

的本征分解方法效果。

3.2 实验细节

训练分为3个步骤：第1阶段预热，迭代次数为5 000次，初始学习率为0.001，每500轮学习率衰减一次。Coarse网络的采样数量为64，fine网络的采样数量为128，训练剪裁的图像，与原始图像的宽高比为0.5。第2阶段正式开始，以半分辨率训练完整图像到50 000次迭代，学习率达到0.000 5，此时只通过语义标签与彩色图像训练体密度和颜色。第3阶段，当参数稳定后计算机加入对场景albedo、shading和specular的学习，此时保持体密度重建网络$\boldsymbol{F}_1(\gamma(\boldsymbol{x}))$参数不变，更新颜色分类层$\boldsymbol{F}_2(\boldsymbol{f})$和光照重建网络$\boldsymbol{F}_3(\gamma(\boldsymbol{d}),\boldsymbol{f})$的参数，迭代150 000次。实验在1张RTX4090显卡上完成，平均训练时间为每秒5次迭代。

3.3 本征分解对比效果

为了验证本文方法所提出的基于Gamma图像增强的本征分解、辐射场分类及shading、specular学习网络的有效性，本文在MIT Intrinsic数据集中开展了本征分解的效果对比。表1为一些本征分解通用方法在MIT Intrinsic数据集中本征分解结果的定量对比，实验采用了MSE（mean square error）和LMSE两个指标来判断本文方法使用的本征分解结果的有效性，通过指标判断，本文方法使用的本征分解效果在通用的本征分解方法中效果处于中游，这是由于本文工作重点为研究并设计室内复杂场景三维神经辐射场构建方法，该辐射场的特点是能够对本征属性进行表征，并没有对单个物体的通用的本征分解进行深入研究。

表1 不同方法在MIT Inttinsic数据集上进行本征分解，albedo和原始图像以及shading和原始图像的定量对比

Table 1 Quantitative comparison of different methods among albedo, shading and original image performing intrinsic decomposition on MIT Intrinsic dataset

方法	MSE ↓			LMSE ↓
	albedo	shading	平均	总计
MSCR（Narihira等，2015）	0.020 7	0.012 4	0.016 5	0.023 9
Revisiting（Fan等，2018）	0.013 4	0.008 9	0.011 1	0.020 3
Data Driven（Zhou等，2015）	0.025 2	0.022 9	0.024 0	0.031 9
Intrinsic CNN（Shi等，2017）	0.021 6	0.013 5	0.017 5	0.027 1
本文	0.013 6	0.015 7	0.014 7	0.031 0

注：“↓”表示值越低越好。

图5为MIT Intrinsic通用分解数据集上的本征分解结果展示，用于对比的图像则是MIT Intrinsic的通用本征分解数据集所提供的分解参考以及其他的“通用本征方法”在MIT Intrinsic数据集上的分解结果。MIT Intrinsic数据集提供的参考图像是通过调整真实光照和拍摄得到，其中albedo图像是在环境中提供全局光照，shading图像是通过对拍摄物体刷“灰漆”得到，故而作为许多通用的本征分解方法的Ground Truth使用。“通用本征方法”是对表1中对比方法与其他类似方法的统称，这些方法的共性是利用神经网络，通过深度学习实现图像的本征分解，且方法的主要研究内容就是实现本征分解。

本文应用情境是重建一个辐射场，此辐射场具有本征属性表征的能力，本文方法应用的本征分解方法也不是通过网络学习实现，而是较为简单的无监督聚类。表1显示了本文方法在分解指标上，与其他方法相近，这是由于应用的情境不同。在本文方法的应用情境下，MIT Intrinsic提供的参考结果在通过式（1）实现原始图像恢复的效果并不好，如图5红框中强调的重建图像与原始图像的对比所示。比较图5红框中本文方法产生的重建图像和其他方法产生图像以及参考图像可以看出，MSCR和DataDriven分解方法产生的图像亮度较暗，并且原始图像中的阴影和高光部分没有体现。特别是DataDriven分解的重建结果，更加模糊且缺乏细节。而本文方法在分解前保留了高光和阴影信息，并在重建中使用，使得恢复图像保留了高光和阴影信息，由此进行的外观编辑也可以对高光和阴影部分进行操作；而比较MIT Intrinsic数据集提供的参考图像的重建图像结果可知，本文方法的重建图像在亮度和

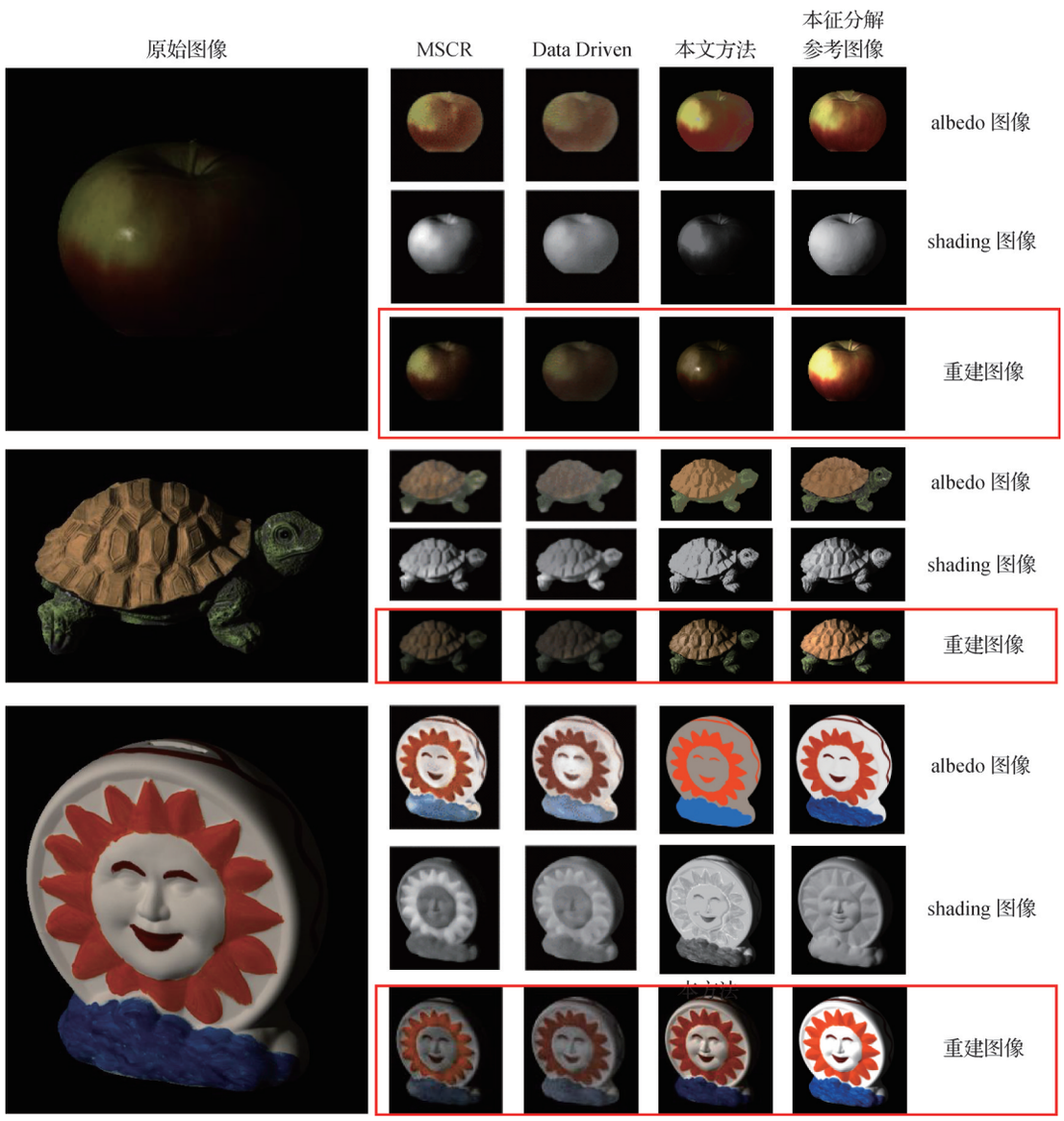


图5　几个典型的MIT Intrinsic通用数据集的本征分解结果(本文方法的重建图像和原始图像更接近)

Fig. 5　A few typical intrinsic decomposition results on MIT Intrinsic dataset(our method performs better on original image recovery)

颜色上具有较好的保持能力。本文方法产生的albedo、shading和specular严格通过式(1)给出,因此恢复的重建图像效果更好。受限于应用情境的不同,MIT Intrinsic数据集提供的通用分解结果仅作为一种展示分解效果的补充。

通过图5可以看出,本文方法产生的分解结果质量与数据集中的分解结果接近程度稍差,但是对原始图像的重建效果更好。这是由于方法和任务目标不同,本文的任务目标是借此重建复杂室内场景的本征属性辐射场,产生albedo图像的方式是通过聚类和语义分割的方式,分解效果与数据集的真实结果并不一致。虽然存在一定的偏差,但一则是本征分解问题本身为病态问题,存在多解,分解结果根据任务需要而定;二则是本文方法为无监督方法,并未使用数据集中的本征属性分解结果进行训练。通过重建结果以及表1中基于albedo的定量评价可以证明本文方法的有效性。

此外,图6表明了本文方法与Intrinsic NeRF方法关于室内复杂场景的本征分解结果的定性对比,红框内体现了本方法所产生分解结果相比Intrinsic NeRF方法的区别。靠图像左边的红框展示了“灯罩”的分解结果,可以看到,Intrinsic NeRF方法给出的albedo图像中仍然具备阴影和色差,未能较好地将阴影从albedo图像中提取出来;靠图像中部的红框展示了“玻璃窗”的分解结果,可以看到,Intrinsic NeRF方法给出的分解结果将玻璃窗的反射当成了

不同的颜色，而本方法通过Gamma变换抑制高光和阴影之后，原本图中被高光照亮的部分被从albedo结果中排除了，而Intrinsic NeRF方法得到的分解结果仍然保留了一部分玻璃窗反射的高光效果。

(a) 原始图像　(b) Intrinsic NeRF　(c) 本文

图6　基于albedo图像的分解结果定性比较(红框内表明本文方法进行本征分解后，场景内物体的高光和阴影被抑制)

Fig. 6　Qualitative comparison of albedo image (our method illustrate that the specular factor and shadow are suppressed in read box region) ((a) original image; (b) Intrinsic NeRF; (c) ours)

图7—图9以3个视角为例，展现了多视角抽样展示三维场景的效果，以及通过本征分解方式控制场景光照渲染的效果，并给出了albedo、shading和specular的分解结果及重上色、重光照的编辑结果。每个视角中，均包括原始图像和分解后网络产生的albedo、shading和specular以及恢复的原始图像的结果。图7(c)、图8(c)和图9(c)展示的shading图像中，发光或高光的部分被抑制；图7(d)、图8(d)和图9(d)中，高光在发光或反光强烈的部分存在强烈表达，同时为了获得较为明显的specular图像，图7(d)、图8(d)和图9(d)展示的specular图像已经在原来的图像基础上乘以3倍，否则图像较暗难以辨认；图7(e)、图8(e)和图9(e)展示的恢复图像损失了一些对光照的渲染效果，这是因为某些光照效果不仅改变了颜色的强度，也改变了颜色的色调。

图7(f)、图8(f)和图9(f)展现了网络能够通过改变albedo图像的三通道颜色来调整场景颜色，同时保证光照条件不变，完成重上色任务，可以看出重上色的图像仍然体现了阴影和高光的效果；图7(g)、图8(g)和图9(g)，以及图7(h)、图8(h)和图9(h)展现了网络能够通过控制shading属性来一定程度地调整场景阴影范围，具体做法是针对方法得到的shading图像进行Gamma变换，当γ值小于1的时候，shading图像对比度降低，阴影部分得到抑制，如图7(g)、图8(g)和图9(g)所示；当γ值大于1的时候，Shading图像对比度提高，阴影部分被放大，如图7(h)、图8(h)和图9(h)所示。图7(i)、图8(i)和图9(i)展现了网络能够通过控制specular属性来调整场景的曝光情况，为场景添加曝光效果。

图7—图9中的recolor和relighting是通过调整albedo图像、shading图像和residual图像的颜色和灰度值实现。根据方法部分的式(1)可知，重建的图像由albedo、shading和residual通过计算给出。由于albedo代表了对象在不受光照影响条件下的颜色，shading代表了对象受到全局光照影响的强度，residual代表了对象受到的直接光照强度，方法可以通过调整3个不同的图像类别来实现场景recolor和relighting的要求。在recolor任务中，可以通过调整albedo颜色分布实现对象颜色的重上色；在relighting任务中，可以调整shading或residual的分布实现对象的重光照，分别是全局光照和直接光照。在实

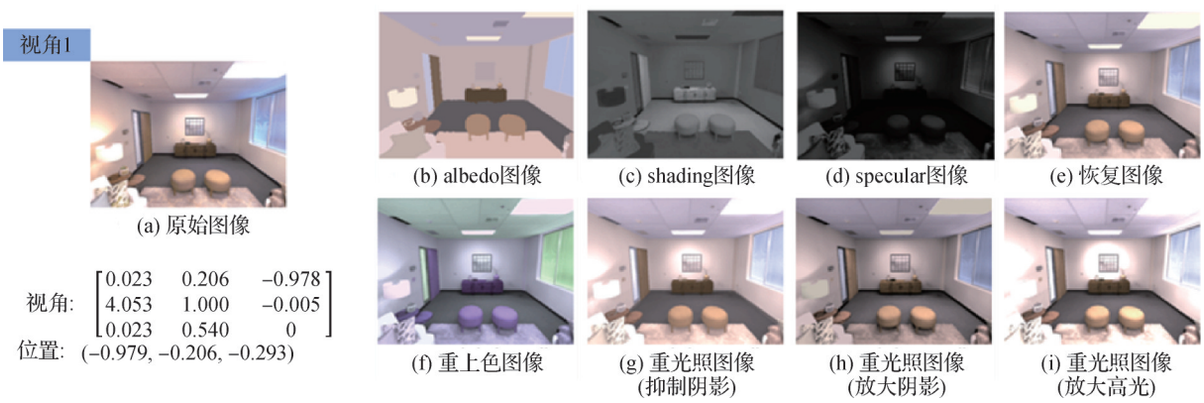


图7　多视角抽样展示三维场景(视角1)

Fig. 7　Multi-view sampling to illustrate 3D scenes(view-01)((a) original image; (b) albedo image; (c) shading image; (d) specular image; (e) recovery image; (f) recoloring image; (g) relighting image(surpass shadow); (h) relighting image(expand shadow); (i) relighting image(expand specular))

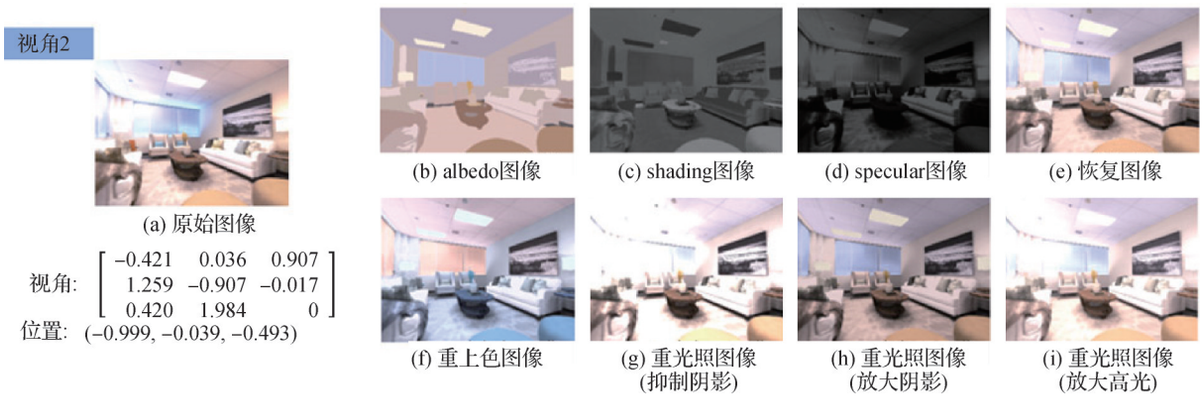


图8　多视角抽样展示三维场景(视角2)

Fig. 8　Multi-view sampling to illustrate 3D scenes(view-02)((a) original image; (b) albedo image; (c) shading image; (d) specular image; (e) recovery image; (f) recoloring image; (g) relighting image(surpass shadow); (h) relighting image(expand shadow); (i) relighting image(expand specular))

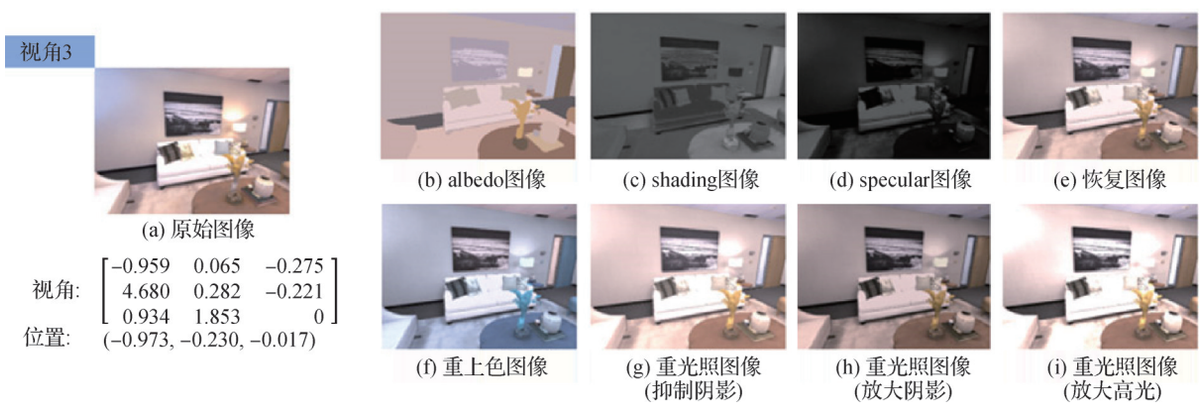


图9　多视角抽样展示三维场景(视角3)

Fig. 9　Multi-view sampling to illustrate 3D scenes(view-03)((a) original image; (b) albedo image; (c) shading image; (d) specular image; (e) recovery image; (f) recoloring image; (g) relighting image(surpass shadow); (h) relighting image(expand shadow); (i) relighting image(expand specular))

验过程中,本文方法在辐射场射线点上采样并根据光照重建网络$\boldsymbol{F}_3(\gamma(\boldsymbol{d}),\boldsymbol{f})$的分支得到相应点的Albedo、Shading和residual属性,例如图7(g)、图8(g)和图9(g),将得到的Shading参数乘以系数1.5,然后通过体渲染和平滑度约束,将三维Shading属性映射到了多视角的图像上,从而产生对应的重光照编辑结果图7(g)、图8(g)和图9(g)。

3.4　渲染质量对比结果

为了验证本文方法所提出的辐射场视图渲染的有效性,本文在Replica的room_0和office_3的两个场景中开展了定量与定性的实验对比。辐射场的视图渲染通过体密度重建网络$\boldsymbol{F}_1(\gamma(\boldsymbol{x}))$和渲染得到的颜色和体密度经过体渲染完成,表2为不同方法在Replica数据集中渲染质量的定量对比,实验采用了PSNR(peak signal to noise ratio)、LPIPS(learned perceptual image patch similarity)和SSIM(structural similarity index)指标,可以看出,本文方法构建的辐射场具备相当的重建和新视图渲染能力,在重建的精度指标上与先前的方法相当,且本方法能够实现

表2　不同方法在Replica数据集上预测的RGB图像和原始图像渲染质量的定量对比

Table 2　Quantitative comparison between predicted RGB image and original image of different methods performing rendering on Replica dataset

方法	PSNR/dB ↑	SSIM ↑	LPIPS ↓
Semantic NeRF	30.977 0	0.895 5	0.106 6
Intrinsic NeRF	30.704 4	0.890 8	0.114 0
本文	28.935 0	0.834 6	0.168 8

优秀的本征分解结果。

此外，表3列出了不同方法在Replica测试集中的显存占用情况以及不同方法在Replica测试集中的运行时间对比（200 k次迭代，在一张RTX4090上运行），可以看出，相比Intrinsic NeRF，本文方法的本征分解过程在预处理阶段完成，因此大幅节省了训练时间和测试时间，同时也尽量减少了在CPU上计算的部分，显存占用率更高。

表3　不同方法在Replica数据集上的训练和测试时间以及显存占用情况对比

Table 3　Comparison of training and test time and GPU memory usage of different methods on Replica dataset

方法	训练时间/h	测试时间/h	显存占用/%
Semantic NeRF	20.66 3	1.750	85.50
Intrinsic NeRF	27.36 4	2.072	80.20
本文	21.04 4	1.750	87.30

3.5　消融实验对比

3.5.1　Gamma变换提取阴影与高光的监督性能验证

一般而言，通过计算机对数据进行监督训练需要监督的目标数据具备良好的性能，例如具备视角间一致性，使得该数据在不同视角下的表现具备连贯的物理特性；或者精确性，要求目标数据在复杂场景中体现场景本身的多目标特性。本文方法在进行辐射场重建时，使用的重建监督目标是通过Gamma变换和本征分解之后得到的albedo、shading和specular结果，Gamma变换的系数是确定的，分解过程是在视角间一致的前提下进行（对整个图像集合的像素进行同步分解），分解也遵循着残差模型，其输入输出确定，从理论上来说其结果相比通用的本征分解模型具备确定性。但是为了侧面验证辐射场训练目标的监督性能，本文在Replica室内场景数据集中进行了辐射场重建的定量实验对比。Intrinsic NeRF的阴影与高光提取并未采用Gamma变换提取高光位置的策略，分解过程与本文方法采取的本征残差模型类似，因此以Intrinsic NeRF进行阴影和高光重建的视角间一致性指标为基准，如表4所示，可以看到，基于Gamma变换进行的本征分解得到的阴影和高光的视角间一致性指标与Intrinsic NeRF方法相比，并未产生明显的变化，证明了通过Gamma变换得到的阴影与高光结果仍然具备相当的监督训练性能。

表4　不同方法在Replica数据集上进行阴影和高光结果提取时，是否引入Gamma变换的视角间一致性指标对比

Table 4　Comparison of multi-view consistency w/o classification layer of different methods on Replica dataset

方法	PSNR/dB ↑	SSIM ↑	LPIPS ↓
Intrinsic NeRF	29.404	0.656 8	0.160 1
本文（with Gamma变换）	29.068	0.650 2	0.161 3

3.5.2　前点优胜策略有效性验证

前点优胜策略的体渲染模块位于体渲染阶段，控制计算机自动地寻找射线经过物体表面的几个采样点，再以分配权重的方式控制颜色分类层$\boldsymbol{F}_2(\boldsymbol{f})$更关注这些采样点的颜色分类以保证分类结果的精准。为了验证前点优胜策略的体渲染模块的有效性，本文在Replica的测试集中进行了定量实验对比，并以语义分割指标准确率、mIoU（mean intersection-over-uniow）对比预测albedo图像与预处理得到的albedo图像的像素分类质量。为了公平对比，本文以Semantic NeRF进行语义分割（分割目标是预处理得到的Albedo）为基准模型，如表5所示。可以看到，相比基准模型，引入前点优胜策略大幅提高了albedo图像的分解准确性，模块的加入使得指标相比基准模型平均提高了4.1%，相比未加入模块提高了3.9%，验证了模块的有效性。

表5　不同方法在Replica数据集上进行语义分割时，引入前点优胜策略的分割指标对比

Table 5　Comparison of semantic segmentation on albedo image w/o front-point dominance of different methods on Replica dataset

方法	mIoU ↑	平均准确率 ↑	总体准确率 ↑
Semantic NeRF	0.887	0.928	0.987
本文w/o前点	0.894	0.929	0.984
本文	**0.948**	**0.975**	**0.991**

注：加粗字体表示各列最优结果。

3.5.3　颜色分类层有效性验证

颜色分类模块即颜色分类层$\boldsymbol{F}_2(\boldsymbol{f})$，支持网络通过分类的方式自动地实现albedo图像的语义及视角间一致性约束。为了验证该模块的有效性，本文

在Replica的测试集中进行了定量与定性实验对比，以语义及视角间一致性为依据对比albedo在3个相近视角间的PSNR、SSIM和LPIPS图像渲染指标。由于是为了验证本征分解效果，实验选定了Intrinsic NeRF方法作为基准模型进行比较，如表6所示。

表6　不同方法在Replica数据集上进行语义分割时，引入颜色分类层的一致性指标对比

Table 6　Comparison of consistency w/o classification layer of different methods on Replica dataset

方法	PSNR/dB ↑	SSIM ↑	LPIPS ↓
Intrinsic NeRF	29.404	0.656 8	0.160 1
本文 w/o颜色分类层	33.470	0.710 9	0.155 8
本文	**33.933**	**0.719 7**	**0.152 0**

注：加粗字体表示各列最优结果。

从表6可以看到，相比基准模型，引入颜色分类层后大幅提高了albedo图像的语义及视角间一致性，该模块加入使得指标相比基准模型平均提高了10.2%，相比未加入模块直接拟合albedo图像结果的方式提高了1.7%，验证了模块的有效性。

4　结　论

本文提出了一种基于神经辐射场学习的本征属性分类网络，可实现语义及视角间一致的本征属性神经辐射场重建。本文方法首先进行本征分解预处理，通过图像对比度增强的形式将原始的二维图像转化为去高光图像，通过自适应Kmeans聚类针对颜色色调产生合理的本征分解结果，在图像集的基础上将分解得到的albedo颜色转化为多视角一致的抽象语义类（在此情况下，一种albedo颜色和一个新的语义类相互等价）。这个将图像颜色转化为多个albedo语义类，然后通过类似语义分割网络进行学习的过程就是方法所使用的“分类”方法，通过分类方法实现本征分解，与之对应的神经网络中包含了学习相应语义类的分支，故命名为“本征属性分类网络”。本文从体现方法特点的角度将题目命名为“利用本征属性分类的神经辐射场视角及语义一致性重建”。通过神经辐射场的三阶段学习，具体包括体密度重建网络$\boldsymbol{F}_1(\gamma(\boldsymbol{x}))$、颜色分类层$\boldsymbol{F}_2(\boldsymbol{f})$和光照重建网络$\boldsymbol{F}_3(\gamma(\boldsymbol{d}),\boldsymbol{f})$，将二维的albedo、shading和specular属性映射到三维辐射场，从而完成场景任意视角重上色、重光照等编辑任务。本方法针对独立于观察方向的albedo属性具备良好的语义及视角间一致性。最后，通过前点优胜的体渲染策略，albedo属性的相似度约束与shading属性的平滑度约束，提升本征分解从二维推广到三维的应用效果。在典型数据集Replica上的充分实验表明，在有限的GPU显存与计算时间下，本文方法能够得到具备语义及视角间一致性的本征属性神经辐射场。对所提出前点优胜模块与颜色分类模块的消融实验表明了本文方法的有效性。

本文工作重点为研究并设计一种三维神经辐射场构建方法，该辐射场能在重建场景体密度和颜色的基础上，也能够表征场景的本征属性，包括albedo、shading和specular等。但是本文方法所使用的本征分解并没有对通用的本征分解进行深入研究。此外，利用Gamma变换去掉高光和阴影再进行分解的思想，在高光情况下表现良好，而在阴影的处理中存在与黑色区域混淆的风险，如图9(b)展示的视角3中分解得到的albedo图像。在未来工作中，计划进一步研究本征分解的物理作用机制，探索场景光照与物体材质相互作用的原理，结合基于物理的渲染方法，解决原有的分解在面对更复杂情况下存在的问题，从而得到更加合理与通用的真实光照渲染，并提高方法的泛化性能，将本文方法扩展到更加复杂的室内室外应用场景中。

参考文献（References）

Baslamisli A S, Groenestege T T, Das P, Le H A, Karaoglu S and Gevers T. 2018. Joint learning of intrinsic images and semantic segmentation//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 286-302 [DOI: 10.1007/978-3-030-01231-1_18]

Boss M, Braun R, Jampani V, Barron J T, Liu C and Lensch H P A. 2021. NeRD: neural reflectance decomposition from image collections//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 12684-12694 [DOI: 10.1109/ICCV48922.2021.01245]

Carbonneau M A, Zaïdi J, Boilard J and Gagnon G. 2024. Measuring disentanglement: a review of metrics. IEEE Transactions on Neural Networks and Learning Systems, 35(7): 8747-8761 [DOI: 10.1109/tnnls.2022.3218982]

Dong B, Dong Y, Tong X and Peers P. 2015. Measurement-based edit-

ing of diffuse Albedo with consistent interreflections. ACM Transactions on Graphics, 34(4): #112 [DOI: 10.1145/2766979]

Fan Q N, Yang J L, Hua G, Chen B Q and Wipf D. 2018. Revisiting deep intrinsic image decompositions//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8944-8952 [DOI: 10.1109/cvpr.2018.00932]

Grosse R, Johnson M K, Adelson E H and Freeman W T. 2009. Ground truth dataset and baseline evaluations for intrinsic image algorithms//Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto, Japan: IEEE: 2335-2342 [DOI: 10.1109/ICCV.2009.5459428]

Insafutdinov E and Dosovitskiy A. 2018. Unsupervised learning of shape and pose with differentiable point clouds//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montreal, Canada: Curran Associates Inc.: 2807-2817

Kajiya J T and von Herzen B P. 1984. Ray tracing volume densities. ACM SIGGRAPH Computer Graphics, 18(3): 165-174 [DOI: 10.1145/964965.808594]

Land E H and McCann J J. 1971. Lightness and retinex theory. Journal of the Optical Society of America, 61(1): 1-11 [DOI: 10.1364/josa.61.000001]

Liu A, Ginosar S, Zhou T H, Efros A A and Snavely N. 2020a. Learning to factorize and relight a city//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 544-561 [DOI: 10.1007/978-3-030-58548-8_32]

Liu L J, Gu J T, Zaw Lin K, Chua T S and Theobalt C. 2020b. Neural sparse voxel fields//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 15651-15663

Liu S, Zhang X M, Zhang Z T, Zhang R, Zhu J Y and Russell B. 2021. Editing conditional radiance fields//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 5773-5783 [DOI: 10.1109/iccv48922.2021.00572]

Lorensen W E and Cline H E. 1987. Marching cubes: a high resolution 3D surface construction algorithm. ACM SIGGRAPH Computer Graphics, 21(4): 163-169 [DOI: 10.1145/37402.37422]

Luo J D, Huang Z Y, Li Y J, Zhou X W, Zhang G F and Bao H J. 2020. NIID-net: adapting surface normal knowledge for intrinsic image decomposition in indoor scenes. IEEE Transactions on Visualization and Computer Graphics, 26(12): 3434-3445 [DOI: 10.1109/TVCG.2020.3023565]

Mescheder L, Oechsle M, Niemeyer M, Nowozin S and Geiger A. 2019. Occupancy networks: learning 3D reconstruction in function space//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 4460-4470 [DOI: 10.1109/cvpr.2019.00459]

Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2022. NeRF: representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1): 99-106 [DOI: 10.1145/3503250]

Narihira T, Maire M and Yu S X. 2015. Direct intrinsics: learning Albedo-Shading decomposition by convolutional regression//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 2992-2992 [DOI: 10.1109/iccv.2015.342]

Ning X J, Gong L, Han Y, Ma T, Shi Z H, Jin H Y and Wang Y H. 2023. Semantic segmentation and model matching-integrated indoor scenario-relevant reconstruction method. Journal of Image and Graphics, 28(10): 3149-3162 (宁小娟, 巩亮, 韩怡, 马婷, 石争浩, 金海燕, 王映辉. 2023. 结合语义分割与模型匹配的室内场景重建方法. 中国图象图形学报, 28(10): 3149-3162) [DOI: 10.11834/jig.220518]

Oleynikova H, Millane A, Taylor Z, Galceran E, Nieto J and Siegwart R. 2016. Signed distance fields: a natural representation for both mapping and planning//Proceedings of 2016 RSS Workshop: Geometry and Beyond—Representations, Physics, and Scene Understanding for Robotics. Ann Arbor, USA: University of Michigan: #134 [DOI: 10.3929/ethz-a-010820134]

Pharr M and Humphreys G. 2010. Physically Based Rendering: From Theory to Implementation. 2nd ed. San Francisco, USA: Morgan Kaufmann Publishers Inc.

Porter T and Duff T. 1984. Compositing digital images//Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques. New York, USA: Association for Computing Machinery: 253-259 [DOI: 10.1145/800031.808606]

Seitz S M, Matsushita Y and Kutulakos K N. 2005. A theory of inverse light transport//Proceedings of the 10th IEEE International Conference on Computer Vision. Beijing, China: IEEE: 1440-1447 [DOI: 10.1109/ICCV.2005.25]

Shi J, Dong Y, Su H and Yu S X. 2017. Learning non-lambertian object intrinsics across shapenet categories//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: 1685-1694 [DOI: 10.1109/CVPR.2017.619]

Sitzmann V, Zollhöfer M and Wetzstein G. 2019. Scene representation networks: continuous 3D-structure-aware neural scene representations//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #101

Straub J, Whelan T, Ma L N, Chen Y F, Wijmans E, Green S, Engel J J, Mur-Artal R, Ren C, Verma S, Clarkson A, Yan M F, Budge B, Yan Y J, Pan X Q, Yon J, Zou Y Y, Leon K, Carter N, Briales J, Gillingham T, Mueggler E, Pesqueira L, Savva M, Batra D, Strasdat H M, De Nardi R, Goesele M, Lovegrove S and Newcombe R. 2019. The replica dataset: a digital replica of indoor spaces [EB/OL]. [2024-03-09]. https://arxiv.org/pdf/1906.05797.pdf

Wan J H, Liu X P, Chen L L, Ao S, Zhang P and Guo Y L. 2024. Geo-

metric attribute-guided 3D semantic instance reconstruction. Journal of Image and Graphics, 29(1): 218-230 (万骏辉, 刘心溥, 陈莉丽, 敖晟, 张鹏, 郭裕兰. 2024. 几何属性引导的三维语义实例重建. 中国图象图形学报, 29(1): 218-230) [DOI: 10.11834/jig.230106]

Wang N Y, Zhang Y D, Li Z W, Fu Y W, Liu W and Jiang Y G. 2018. Pixel2Mesh: generating 3D mesh models from single RGB images// Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 55-71 [DOI: 10.1007/978-3-030-01252-6_4]

Wang P, Liu L J, Liu Y, Theobalt C, Komura T and Wang W P. 2021. NeuS: learning neural implicit surfaces by volume rendering for multi-view reconstruction//Proceedings of the 35th International Conference on Neural Information Processing Systems. [s.l.]: Curran Associates Inc.: #2081

Wang Y J, Fan Q N, Li K, Chen D D, Yang J Y, Lu J Z, Lischinski D and Chen B Q. 2022. High quality rendered dataset and non-local graph convolutional network for intrinsic image decomposition. Journal of Image and Graphics, 27(2): 404-420 (王玉洁, 樊庆楠, 李坤, 陈冬冬, 杨敬钰, 卢健智, Lischinski D, 陈宝权. 2022. 面向本征图像分解的高质量渲染数据集与非局部卷积网络. 中国图象图形学报, 27(2): 404-420) [DOI: 10.11834/jig.210705]

Yariv L, Kasten Y, Moran D, Galun M, Atzmon M, Basri R and Lipman Y. 2020. Multiview neural surface reconstruction by disentangling geometry and appearance//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #210

Ye G Z, Garces E, Liu Y B, Dai Q H and Gutierrez D. 2014. Intrinsic video and applications. ACM Transactions on Graphics, 33(4): #80 [DOI: 10.1145/2601097.2601135]

Ye W C, Chen S, Bao C, Bao H J, Pollefeys M, Cui Z P and Zhang G F. 2023. IntrinsicNeRF: learning intrinsic neural radiance fields for editable novel view synthesis//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 339-351 [DOI: 10.1109/iccv51070.2023.00038]

Yen-Chen L, Florence P, Barron J T, Rodriguez A, Isola P and Lin T Y. 2021. iNeRF: inverting neural radiance fields for pose estimation//Proceedings of 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems. Prague, Czech Republic: IEEE: 1323-1330 [DOI: 10.1109/iros51168.2021.9636708]

Zhang K, Luan F J, Wang Q Q, Bala K and Snavely N. 2021a. PhySG: inverse rendering with spherical Gaussians for physics-based material editing and relighting//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 5453-5462 [DOI: 10.1109/cvpr46437.2021.00541]

Zhang X M, Srinivasan P P, Deng B Y, Debevec P, Freeman W T and Barron J T. 2021b. NeRFactor: neural factorization of shape and reflectance under an unknown illumination. ACM Transactions on Graphics, 40(6): #237 [DOI: 10.1145/3478513.3480496]

Zhi S F, Laidlow T, Leutenegger S and Davison A J. 2021. In-place scene labelling and understanding with implicit scene representation//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 15838-15847 [DOI: 10.1109/iccv48922.2021.01554]

Zhou T H, Krähenbühl P and Efros A A. 2015. Learning data-driven reflectance priors for intrinsic image decomposition//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 3469-3477 [DOI: 10.1109/iccv.2015.396]

作者简介

曾志鸿,男,硕士研究生,主要研究方向为三维场景编辑。
E-mail: zengzhihong17@mails.ucas.ac.cn
张源奔,通信作者,男,副研究员,主要研究方向为时空信息综合处理与应用。E-mail:zhangyb@aircas.ac.cn
王宗继,男,助理研究员,主要研究方向为计算机视觉、计算机图形学、三维场景重建与理解。
E-mail:wangzongji@aircas.ac.cn
蔡伟南,男,博士研究生,主要研究方向为三维场景生成。
E-mail:caiweinan22@mails.ucas.ac.cn
张利利,女,研究员,主要研究方向为时空信息综合处理与应用。E-mail:zhanglili86@126.com
郭岩,男,助理研究员,主要研究方向为时空信息综合处理与应用。E-mail:guoyan@aircas.ac.cn
刘俊义,男,研究员,主要研究方向为信号与信息处理。
E-mail:liujy004735@aircas.ac.cn;