arXiv:2301.01087v1 [cs.GR] 3 Jan 2023

用于反射新视角合成的神经点焦散

GEORGIOS KOPANAS，Inria 与蔚蓝海岸大学，法国；THOMAS LEIMKÜHLER，马克斯·普朗克信息学研究所，德国；GILLES RAINER，Inria 与蔚蓝海岸大学，法国；CLÉMENT JAMBON，Inria、蔚蓝海岸大学与巴黎综合理工学院，法国；GEORGE DRETTAKIS，Inria 与蔚蓝海岸大学，法国

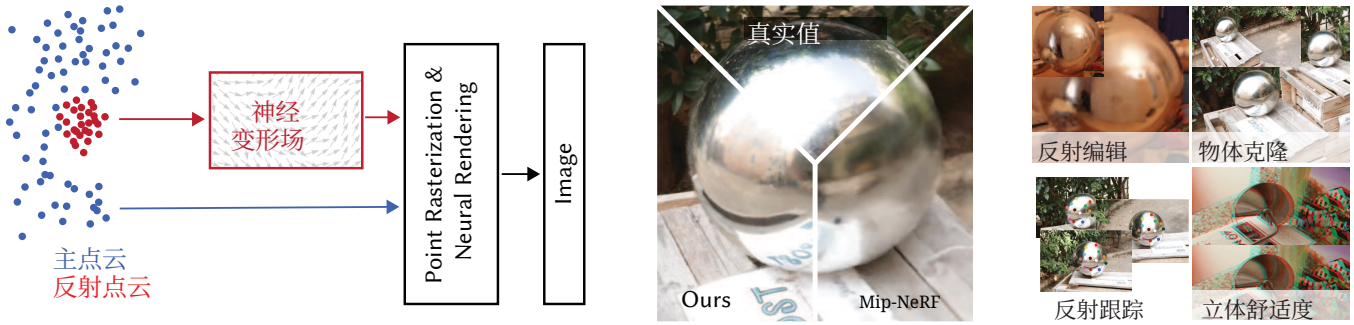

图 1。我们提出了一种对曲面反射体进行新视角合成的方法。我们采用动态点式场景表示，以便对反射焦散轨迹进行建模，从而准确估计反射流。就图像质量而言，我们的方法优于当前最先进的技术，并支持多种额外应用。

反射等视点相关效应对基于图像的渲染和神经渲染算法构成了重大挑战。其中，曲面反射体尤为棘手，因为相机移动时会产生高度非线性的反射流。我们提出了一种新的基于点的表示，用于计算神经点焦散，从而利用一组随手采集的输入照片，对包含曲面反射体的场景进行新视角合成。我们方法的核心是一种神经变形场，用于模拟反射的焦散轨迹，因此可以结合高效的点泼溅与神经渲染器，渲染复杂的镜面效果。我们的主要贡献之一是显式表示反射：使用神经变形场对反射点云进行位移，并优化主点云以表示场景的其余部分。经过简短的手动标注后，我们的方法便可交互式地高质量渲染新视图，并准确呈现反射流。此外，反射流的显式表示支持对采集场景进行多种操控，例如反射编辑、镜面反射物体克隆、跨视图反射跟踪以及舒适立体视觉。我们在 https://repo-sam.inria.fr/fungraph/neural_catacaustics/ 上提供源代码和其他补充材料。

作者地址：Georgios Kopanas，Inria 和蔚蓝海岸大学，法国，georgios. kopanas@inria.fr；Thomas Leimkühler，马克斯·普朗克信息学研究所，德国，thomas.leimkuehler@mpi-inf.mpg.de；Gilles Rainer，Inria 和蔚蓝海岸大学，法国，gilles.rainer.enst@gmail.com；Clément Jambon，Inria、蔚蓝海岸大学和巴黎综合理工学院，法国，clement.jambon@polytechnique.edu；George Dret- takis，Inria 和蔚蓝海岸大学，法国，george.drettakis@inria.fr。

出版权已授权给 ACM。ACM 声明，本稿件由国家政府的雇员、承包商或关联人员撰写或共同撰写。因此，政府保留非独占、免版税的权利，可仅为政府用途出版或复制本文，或允许他人这样做。© 2022年 版权所有归所有者/作者所有。出版权已授权给 ACM。
0730-0301/2022/12-ART201 $15.00
https://doi.org/10.1145/3550454.3555497

CCS 概念：• **计算方法** → **基于点的模型**；**反射率建模**；**光栅化**；• **计算机系统组织** → 神经网络。

附加关键词与短语：基于点的渲染、神经渲染、可微光栅化、反射、焦散

ACM 引用格式：
Georgios Kopanas、Thomas Leimkühler、Gilles Rainer、Clément Jambon 和 George Drettakis。2022年。用于反射新视图合成的神经点焦散。*ACM* 图形学汇刊第41卷第6期，第 201 号文章（2022 年 12 月），共 15 页。https://doi.org/10.1145/3550454.3555497

1 引言

近期的神经渲染方法 [Tewari 等人 2020, 2021] 为采集场景的自由视点渲染带来了出色的视觉质量。这类场景通常包含重要的高频视点相关效果，例如有光泽物体产生的反射；对此，可以采用两种根本不同的方式进行建模（图 2）：一种是采用 欧拉方法，将反射表示为固定形式，并对外观的方向变化进行建模；另一种是采用 拉格朗日方法，在观察者移动时跟踪反射的流动。大多数以往方法都采用前一种方式，根据位置和视线方向来表示静态点上的颜色，使用的要么是计算开销较大的体渲染 [Mildenhall 等人，2020年；Tewari 等人，2021年；Wizadwongsa 等人 2021]，，要么是基于网格的渲染 [Hedman 等人，2018年；Riegler 和 Koltun 2021]。与此不同，我们的方法通过神经变形场直接学习反射流动与视点之间的关系，实质上采用了拉格朗日方法 [Bemana 等人 2020]。我们的基于点的神经渲染

方法能够通过神经场自然地变形反射点，从而实现交互式渲染。

以往的解决方案往往难以兼顾质量与速度，因为为了模拟（相对）高频反射，它们通常采用缓慢的体积光线步进，并进行视图相关查询。快速的近似方案 [Hedman 等人，2021年；Yu 等人 2021] 则牺牲角分辨率，从而降低反射的质量和锐利度。总的来说，这类方法使用多层感知机（MLP）来建模密度和由视线方向参数化的视图相关颜色，并在反射体后方构建反射几何。结合体积光线步进后，通常会产生“雾蒙蒙”的效果，反射中缺乏清晰的细节。近期的一种方案 [Verbin 等人 2021] 提升了这类方法的质量，但仍受制于缓慢的体渲染。此外，使用这类方案难以操控包含反射的场景。

我们采用拉格朗日式的基于点的方法，避免了基于隐式多层感知机的神经辐射场固有的低频偏置；即使使用不同的编码和参数化方式，这种偏置依然存在。我们的方法还有两项优势：推理时开销更低，因此支持交互式渲染；而直接表示则让场景操控变得简单。

我们使用多视图数据集，通过标准的 3D 立体重建提取初始点云；只需进行少量手动操作，在 3–4 张图像上定义反射体掩模，随后便可利用额外的高维特征优化两个独立的点云。在渲染过程中，主点云保持静止，表示场景中以漫反射为主的部分；而第二个反射点云则表示高度依赖视图的反射效果，这些点会受到学习得到的神经变形场的位移（见图 1）。点还带有覆盖范围和不透明度参数，这些参数在训练期间与点的位置一同优化。随后，两个点云的学习特征经光栅化后，由神经渲染器进行解释，从而生成最终图像。

我们的研究受到曲面反射体几何光学理论基础的启发，该理论表明，来自曲面物体的反射会沿焦散曲面 [Hamilton，1828；Lawrence 2013]；这通常会产生高度不规则、快速移动的反射流（图 3）。我们训练一个流场来学习这些轨迹，并将其称为神经点焦散，从而实现交互式自由视点神经渲染。重要的是，基于点的表示具有显式特性，便于对含有反射的场景进行操作，例如反射编辑、反射物体克隆等。

我们首先介绍复杂反射流在曲面反射体上的几何背景，为我们的算法提供指导，然后介绍以下贡献：

- 一种用于神经渲染的新型直接场景表示，其中包含一个独立的反射点云，该点云由反射神经变形场驱动，该变形场学习计算神经点焦散；此外，还包含一个主点云，其参数经过优化，用于表示场景的其余内容。
- 一个神经变形场，用于学习反射点随视点变化的位移 。要稳定训练我们端到端的方法（包括该变形场），需要仔细设计参数化方式和初始化，并采用渐进式训练与点加密。
- 最后，我们提出一种通用的交互式 神经渲染算法，能够高质量地渲染场景中的漫反射辐射和视点相关辐射，从而支持对采集场景进行自由视点导航和交互式渲染。

我们在若干采集场景上展示了我们的方法，并表明，与以往用于渲染曲面物体反射的神经渲染方法相比，我们的方法在定量和定性表现上均更优，同时还能快速渲染和操作此类场景，例如编辑反射、克隆反射物体，或在输入图像中查找反射对应关系。

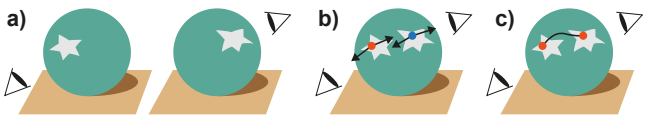

图 2. 视点相关效应的欧拉方法与拉格朗日建模。(*a*) 从两个视点观察一个曲面反射体。星星的反射出现在不同位置，并呈现不同的畸变。(*b*) 在欧拉方法中，每个曲面点都会根据视线方向（黑色箭头）计算外观。红点在左侧视图中可见，在右侧视图中则不可见。必须针对每个曲面点学习反射及其畸变。(*c*) 在拉格朗日方法中，反射物体只需表示一次，并对其运动进行显式建模。许多反射点的运动会自然产生畸变。

2 相关工作

我们仅回顾与本文密切相关的传统反射渲染和基于图像的反射渲染文献，尤其是变形方法；并讨论采用辐射方向表示的神经渲染方法，以及基于点的（神经）渲染方法。

2.1 反射渲染

自计算机图形学（CG）诞生之初，反射的建模与渲染便一直是其重要目标。光线追踪 [Whitted 1979]直接模拟光学现象来渲染反射，极大地提升了计算机图形学渲染的真实感。起初，只有环境映射或反射映射 [Blinn and Newell 1976] 等简化方案具有实用性。

*传统反射渲染。*最初提出了多种利用光栅化渲染反射的方法。对于平面镜，Foley 等人 [1996] 采用新的虚拟相机，并通过多遍渲染虚拟物体 [Diefenbach 和 Badler 1997] 来渲染反射，即将反射场景再次渲染。这一思路后来由 Estalella 等人 [（2005 年）以及 Ofek 和 Rappoport 1998] 扩展至曲面反射体：他们通过变形虚拟网格来渲染反射形状，并使用不同的加速结构。随后，人们提出了多种利用 GPU着色器的解决方案；参见综述 [Szirmay-Kalos 等人 2009]。类似地，我们使用点表示反射几何体，但通过训练神经网络来学习反射点随视点运动产生的流。

可通过对环境图进行预滤波，处理粗糙度各异的光泽表面 [Greene 1986]；随后， [Cabral 等人 1999]通过快速的、基于图像的反射探针变形扩展了这一方法。路径扰动理论 [Chen 和 Arvo 2000] 曾被用于估计镜面路径的位移；

该理论也指导了用于光线追踪的、基于探针的反射流计算 [Rodriguez 等人 2020年a]。Lochmann 等人 [2014] 采用优化方法估计分布式渲染中的反射流。我们的方法与这些方法的出发点相似，因为我们也使用反射流，但应用场景是实拍真实场景，这带来了许多额外挑战，例如几何结构不准确。

利用反射进行反射重建与基于图像的渲染。 三维重建 [Goesele 等人 2007] 指出，含反射场景的基于图像的渲染（IBR）极具挑战性，因为光度一致性会遭到破坏。多种方法采用基于图像的方法采集并渲染反射，通常会将输入图像分离为反射层和透射层 [Szeliski 等人 2000]。反射的图层分离一直是计算机视觉领域的研究课题，既有用户辅助的解决方案 [Levin 和 Weiss 2007]，，也有近期基于偏振的深度学习方法（例如，[Wieschollek 等人 2018]）。更复杂的算法能够估计反射深度 [Sinha 等人 2012]，并支持高效的基于图像的渲染 [Kopf 等人 2013]，，但难以处理曲面反射体。近期有一种专门的方法，可估计汽车车窗的曲面反射体几何结构，并通过临时的图层分离来估计反射流 [Rodriguez 等人 2020年b]。与此不同，我们提出了一种能够学习任意曲面反射体上反射的方法。

理论上，可以使用可微渲染 [Nimier-David 等人 2019] 进行几何结构和材质优化，但尚未证明这类方案能够适用于几何结构不完整的采集场景，尤其是我们这种重建非常不完整的高光反射体。

2.2 神经与基于点的渲染

神经渲染是一个广阔且发展迅速的领域；因此，我们仅讨论与本研究直接相关的工作，即反射渲染/光流和基于点的渲染；近期综述 [Tewari 等人 2020年、2021年；Xie 等人 2021] 对该领域进行了详细介绍。

一些神经渲染方法针对反射/镜面高光提出了专门的处理方式。Thies 等人 [2019] 使用纹理空间中的特征向量来编码镜面效果，但需要对输入相机进行密集的角度采样。X-Fields 方法 [Bemana 等人 2020]学习各种参数（包括镜面反射）的流动，但重点在于小基线的“类光场”相机运动。我们关注的是宽基线随意拍摄，以及完整场景中不受约束的自由视点相机运动。

NeX [Wizadwongsa 等人 2021] 通过学习由基函数线性组合构成的方向性色彩表示，改进了多平面图像方法。神经辐射场（NeRF）[Barron 等，2021；Mildenhall 等人 2020] 使用多层感知机学习密度和视图相关颜色。NeRF 往往会在反射体后方生成与反射物体相对应的密度，并利用视图相关颜色项模拟期望的外观。这在一定程度上能够表示反射流，但通常无法重建清晰的特征，且需要代价高昂的体渲染。针对后一个问题，已有若干变体加以解决 [Garbin 等人 2021；Müller 等人 2022]。与这些表示不同，我们显式的基于点的方法通常能产生更清晰的结果，并支持场景操控。

其他方法将多视图立体视觉重建所提供的网格用作“骨架”。Deep Blending [Hedman 等人 2018] 使用视图相关网格进行输入图像投影，随后通过学习权重进行融合；反射通过简单插值逐渐融入或淡出，往往会产生严重的时间伪影。此外，按输入视图存储数据的方式无法扩展到图像数量庞大的数据集。近期研究针对平面反射显式重建反射图像，并使用超分辨率神经网络进行渲染 [Xu 等人 2021]。神经方法从输入图像中提取视图相关的深度特征，并将其合并到这一共同空间中 [Riegler 和 Koltun 2021]，；而 Philip 等人 [2021] 则通过向神经渲染器提供镜像，显式处理非漫反射。上述所有方法都采用了欧拉方法 [Wu 等人 2012]，，因为它们将方向信息存储在固定的空间引导结构（显式或隐式）中；而我们则采用拉格朗日估计反射流。

最近发表的基于 NeRF 的方法通过重新参数化视图相关颜色来改善反射：它们以反射方向 而非视图作为依据 [Verbin 等人 2021]，，从而改善定向入射光照下的反射；或者学习用于平面反射的独立神经表示 [Guo 等人 2021]。我们的方法处理曲面反射体上的一般反射情形，并支持交互式渲染。

最后，近期方法使用神经变形场学习形变（例如，[Park 等人，2021；Tretschk 等人 2021]），但其目标不同，是处理动态内容；而 Wang 等人 [2021] 则使用相关思路，通过球面镜阵列进行单次拍摄。这些方法可视为拉格朗日方法，但应用背景不同，例如用于运动，或利用反射重建非镜面场景。

基于点的神经渲染。 基于点的渲染是一种灵活的几何结构渲染方法 [Gross 和 Pfister 2007]，，因为它不需要网格连通性。通常会将点泼溅 到屏幕上；必须谨慎选择泼溅点的大小和形状 [Zwicker 等人 2001]。随着可微解的问世 [Wiles 等人，2020；Yifan 等人 2019]，，以及快速近似方法的出现 [Lassner 和 Zollhofer 2021]，基于点的渲染重新受到关注。已有多种神经渲染方法被提出，它们直接渲染场景的基于点的表示 [Aliev 等人 2020][Meshry 等人 2019]，，通过将视图相关特征投影到新视图中 [Kopanas 等人 2021]，；或者采用快速渲染方法，通过优化相机参数来提升质量 [Rückert 等人 2022]。近期研究中，点被用于表示隐式光场 [Ost 等人 2021]。其他方法则将点用作神经基函数 [Xu 等人 2022] 或局部神经场 [Feng 等人 2022]；这两类方法都能在必要时对点云进行上采样或扩展。我们的方法延续了这类研究，利用点的灵活性——点天然适用于我们场景中估计得到的反射流，同时也能灵活表示场景的其余部分，例如在需要时轻松进行点加密。

3 背景与概述

我们的目标是开发一种适用于包含反射体的场景的神经渲染方法，从而支持交互式自由视点渲染和场景

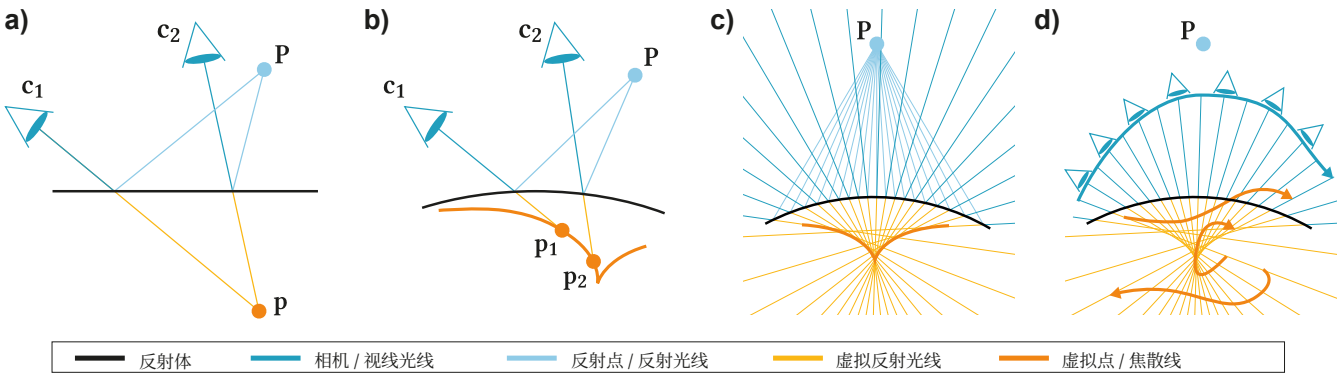

图 3. 焦散的几何结构：*(a)* 对于平面反射体，反射点 P 会产生一个静止的虚拟点 p，其位置不受相机位置 c 的影响。*(b)* 对于曲面反射体（此处以凸面为例），相机运动会使 p 描绘出一个曲面，称为焦散线。*(c)* 单个点 P 的焦散线由虚拟反射光线的包络线定义，如粗橙色曲线所示；包络线上的每一点都与其中一条虚拟反射光线相切。P 的虚拟图像仅在焦散线轨迹上形成。*(d)* 在基于点的渲染设置中，由于不满足相切约束，虚拟点的轨迹并不唯一。无数条轨迹（三个示例如图所示）都会产生相同的表观反射。为清晰起见，省略了反射光线。

操控。接下来，我们将回顾这一背景下的反射几何学，并概述我们的方法。

3.1 曲面物体的反射几何

对于平面反射体（图 3a），其几何结构很简单。考虑点 P；它的反射会在反射体另一侧形成一个静止的虚拟点 p，视点改变时该点不会移动。为在渲染过程中获得正确的反射流，只需将 p 重新投影到新视图 [Foley 等人，1996；Guo 等人，2021；Xu 等人，2021]。

对于曲面反射体（图 3b），情况要复杂得多。在这里，虚拟点的位置取决于视点，也就是说，随着视图变化，该点会在空间中沿轨迹移动。虚拟点 p 描绘出一个曲面，称为焦散线 [Hamilton，1828；Lawrence 2013]。该曲面的形状既取决于反射点 P 的位置，也取决于反射体的形状。焦散面通常高度非线性，其复杂度显著高于反射体几何结构 [Josse 和 Pene 2014]。对于解析反射体几何，存在闭式解。例如，圆形反射体会形成心形线 [Glaeser 1999]；对于更复杂的形状，只要已知精确几何结构，也可以通过优化求得焦散面 [Mitchell 和 Hanrahan 1992]。焦散面可能位于反射体表面的后方或前方，具体取决于 P 的位置以及反射体是凸还是凹；这可能导致大幅缩小或放大，以及其他不规则运动。

在物理学/几何光学中，给定所有几何信息后，焦散线的位置和形状便被唯一确定：只有在焦散面上，P 的（虚拟）光学图像才会出现，因为（虚拟）反射光线仅在这些曲面上相交（图 3c）。对于一组曲线，包络线被定义为与该组中每条曲线相切的曲线。具体而言，包络线由这些切点定义 [Bruce and Giblin 1992]。由于焦散是虚拟反射光线的包络线 [Hamilton 1828]（图 3(c)），虚拟点仅沿焦散线的切线光线可见（另见图 3b）。

3.2 曲面物体反射的神经渲染

对这些运动反射进行自由视点神经渲染十分复杂。我们基于拉格朗日的方法估计曲面反射体上反射点的焦散线轨迹，从而只需存储一次反射，并可重复使用。

我们通过对多视图图像进行训练，学习反射的运动：训练神经变形场 [Sitzmann 等人，2019年；Tewari 等人，2021年；Xie 等人， 2021]。该变形场用于对虚拟点云进行变形，使其与输入视图中采集的反射相匹配；再结合神经渲染网络，即可在具有曲面反射体形状的场景中实现交互式自由视点导航。

然而，使用点会导致结果偏离物理规律：由于在基于点的渲染中，点会向所有方向发光，因此从构造上说，它就是一个（虚拟）光学图像，与其位置无关。这意味着反射轨迹不再唯一（图3d）：任何在适当时刻按正确顺序穿过虚拟反射光线的轨迹，都会产生与运动相机所见相同的表观反射（遮挡情况除外），从而造成深度歧义。对于渲染而言，无需恢复物理上正确的焦散面，因为图3d中的所有不同解都会生成相同的图像。

我们的方法基于焦散曲面理论来学习反射点的轨迹，并将其称为 神经点焦散。我们在第 7.7 节对物理焦散与神经点焦散进行了定性分析。

3.3 方法概述

我们的方法以场景的多张宽基线照片作为输入，对于包含曲面反射体的室内或室外场景，通常需要 200–300 张。我们使用标准的运动恢复结构（SfM）对相机进行标定，并采用标准的多视图立体视觉（MVS）方法提取稠密点云。随后执行一个最少的手动步骤，

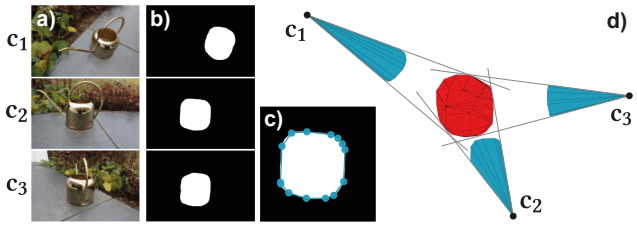

图 4. 反射体体积边界：*(a, b)* 用户需要在少量图像（此处为 3 张）中绘制粗略掩码，标出反射体。*(c)* 根据这些掩码，我们计算简单的二维边界折线（仅显示 c_2）。*(d)* 最后，我们求解满足所有掩码约束的凸三维多面体。

用户在数据集的 3–4 个视图中为反射体创建粗略掩码（此步骤通常不到 1 分钟）；这些掩码用于提取一个包裹反射体的反射体积。

我们的方法如图 5 所示。MVS重建用于初始化主点云，以表示场景的漫反射内容和低频视角相关内容。我们还使用先前创建的反射体积初始化反射点云。训练期间，我们联合优化这两个点云的属性。

为了使用自由视点相机渲染图像，我们首先通过神经变形场在三维空间中移动反射点云，将反射移动到给定相机对应的正确位置。随后，使用 EWA泼溅分别对主点云和反射点云进行光栅化 [Zwicker et al. 2001]。然后，将光栅化后的点云与优化后的环境贴图合成，以补偿缺失的远处几何体。最后，将所得的高维特征输入解码器神经渲染器，合成最终图像。我们的架构采用端到端方式训练，从而实现交互式自由视点渲染器。

4 方法

下面按照图 5（从左到右）介绍我们方法的主要组成部分。我们首先讨论双点云表示，说明反射点如何通过神经变形场发生位移；接着讨论如何对点及其参数进行光栅化；最后介绍我们如何进行神经渲染。

4.1 点云

我们使用三维点表示场景。除了位置和法向量等自然参数外，我们还为这些点添加了其他参数，所有参数均由我们的方法优化。这些属性使主点云能够修正 MVS 重建中错误且不完整的几何结构（见图 7），并且对于学习神经变形场至关重要，从而使反射点能够发生位移以跟随反射流。

我们为两个点云中的所有点优化以下参数：不透明度、覆盖范围和高维特征。对于主点，我们还会优化位置、法向量以及参数 ρ，该参数将用于调制光栅化后的点云。

4.2 神经变形场

我们定义了一个神经变形场，用于随相机运动调整反射点的位置。形式化地说，给定初始反射点位置 $\mathrm{p} \in \mathrm{R}^3$ 和当前相机位置 $\mathrm{c} \in \mathrm{R}^3$,，我们希望通过场 $\mathrm{F} \in (\mathrm{R}^3 \times \mathrm{R}^3) \rightarrow \mathrm{R}^3$，计算神经点焦散上的反射位置 $\mathrm{p}' \in \mathrm{R}^3$：

$$\mathrm{p}' = \mathrm{p} + \mathcal{F}(\mathrm{p}, \mathrm{c}).$$

我们使用带有可训练参数 θwarp的多层感知机来实现 F。初始点位置 p 是固定的，不进行优化（有关我们的初始化流程，请参见第 5.1 节），但每个点都会受到神经变形场的位移。请注意，变形场以初始点位置 p 为输入，因此我们可以对所有反射点使用同一个场，从而紧凑地表示整个反射流。

4.3 光栅化

我们使用可微点泼溅 [Kopanas 等人，2021年；Yifan 等人，2019]对点云进行光栅化。每个点都可以视为一个具有位置 p 和法向量的有向圆盘，并投影到屏幕上，呈现为一个软椭球。每个点还具有一个三维覆盖范围（定义为圆盘半径的平方）、一个 6 通道神经特征以及不透明度 o。

主点云负责表示场景的几何结构。我们为每个主点学习一个额外参数 ρ；该标量值将在最终渲染中控制反射点云的混合程度。

为了对一个点进行光栅化，我们需要确定其二维覆盖范围，同时考虑其与相机之间的距离以及由点的法向量定义的倾斜程度。为此，我们按照以下方式计算每个点的协方差矩阵 [Ren 等人，2002年；Zwicker 等人，2001]:

$$\Sigma = h^2 \mathrm{J}\mathrm{V}\mathrm{J}^T + v\mathrm{I} \tag{1}$$

其中，$\mathrm{J} \in \mathrm{R}^{2\times 2}$ 是从世界空间到视空间的变换的雅可比矩阵，h 是用于补偿像素分辨率的缩放因子 [Zwicker 等人，2001], 而 V 是由该点的三维覆盖范围缩放的单位矩阵。$v\mathrm{I}$ 是低通滤波器；在所有测试中，我们均使用 $v = 0.3$，以在混叠导致的梯度不稳定性与恢复高频特征之间取得平衡。

我们对所有泼溅点执行从前到后合成，以计算像素特征值 c。我们遵循先前的工作 [Kopanas 等人，2021年；Yifan 等人 2019]，并计算落在像素上的所有 N 个点的:

$$c = \sum_{i \in \mathcal{N}} c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j). \tag{2}$$

我们的 α_i 通过计算协方差为 Σ 的二维高斯分布（式 1） [Yifan 等人，并将其与学习得到的逐点不透明度参数 o相乘来获得。不透明度值 o 使优化过程能够让点“消失”，从而修正多视图立体视觉造成的错误过度重建。我们将完整乘积称为:

$$\bar{o} = \prod_{j \in \mathcal{N}} (1 - \alpha_j) \tag{3}$$

累积不透明度，用于第 5.2 节所述的损失，以及下文所述的环境贴图。

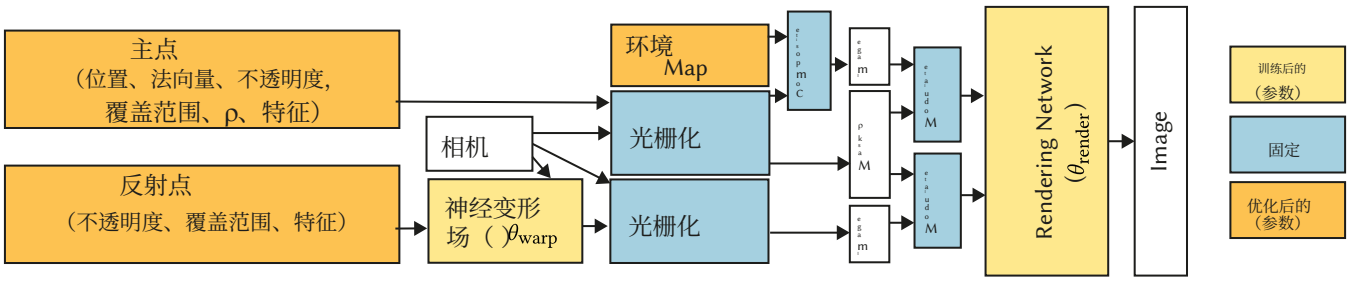

图 5. 我们方法的概览。

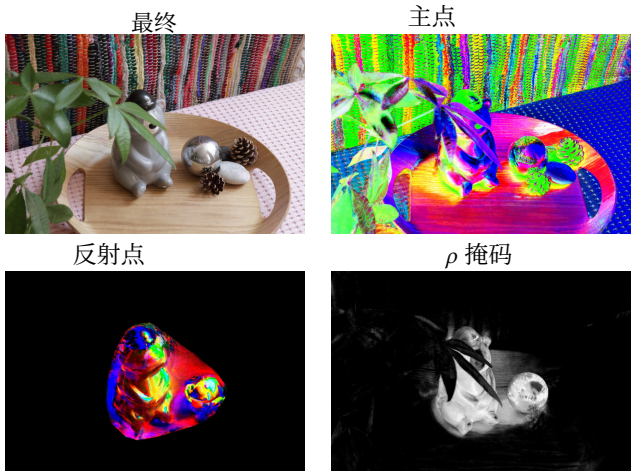

图 6. 提供给神经渲染网络的中间缓冲区。对于主点云和反射点云，我们将前三个通道显示为 RGB。请注意，主点云捕捉了物体的非镜面外观，即瓷熊上的纹理，而反射点云则捕捉了所有镜面细节。掩码 ρ 用于处理镜面反射物体与非镜面物体之间的遮挡。

对于多视图立体视觉无法计算出几何结构的情况，我们使用环境贴图；这通常发生在大型场景中距离较远且仅从少数几个视图观察到的物体上。我们使用极坐标对环境贴图进行参数化，并将其初始化为 6 个值恒为零的特征。渲染时，我们将环境贴图投影到视图中，并使用累积不透明度 $\bar{o}$ 对其进行混合。

4.4 神经渲染

在对所有点及其相应参数（即六通道特征和 ρ）进行光栅化与合成后，我们将这些信息输入一个具有两个分支头的神经渲染器多层感知机。第一个分支头将主点的光栅化特征与视线方向编码为潜在特征。引入视线方向使主点云能够建模低频视角依赖效应（欧拉式），而几乎所有真实世界材质在一定程度上都表现出这种效应。

我们使用掩码 ρ 来调制这些主要的光栅化特征。第二个分支头对反射点特征执行相同的操作，并由 $1 - \rho$进行调制。我们将两个分支的输出拼接起来，并输入解码器网络，以获得最终图像。渲染器使用的所有层均如图 6 所示。我们将该模块的可训练参数记为 θrender。

5 优化与训练

回顾一下，我们有两个点云，一个用于反射，另一个用于其余场景内容；我们联合优化这两个点云和两个网络的参数。这一联合优化过程带来了若干重大挑战，尤其是用于建模复杂运动的反射流容易出现不稳定性，而我们的方法能够解决这些问题。我们将讨论如何使用通过极少用户干预创建的反射体积来初始化反射点，并对其在空间中的运动进行正则化。随后，我们将讨论损失函数，以及使用多分辨率求解器和点加密进行优化的方法。

5.1 反射点的初始化与正则化

在许多场景中，反射表面的空间范围往往有限；我们的每个测试场景都包含一个主要的曲面反射体，输入照片则涵盖了物体周围的视图，以确保反射具有良好的角度覆盖范围，并能看到整个场景，从而实现自由视点渲染。我们考虑包含反射物体的三维体积。这个反射体积表示我们预期神经点焦散在空间中移动的大致区域。我们在反射体积的边界表面上随机放置点，以初始化反射点云。所有实验中我们均使用 400K 个点。在二维曲面上进行初始化是自然的选择，因为我们在反射中看到的场景元素大多由曲面构成（或可近似为曲面）。此外，在优化过程中，我们将使用反射体积对反射点云变形后的位置进行正则化（第 5.2 节）。

我们通过识别输入图像中对应于反射物体的区域来定义反射体积。已有自动检测反射的方法 [Whelan 等人 2018],，但尚未证明这些方法能够可靠地处理曲面反射体。因此，我们选择一种尽量减少用户干预的方法：请用户在少量输入图像中绘制粗略掩码，标出出现反射的物体（图 4a 和 b）。在所有示例中，我们都使用三到四张图像；对于给定的数据集，绘制掩码所需时间不到一分钟。

根据这些掩码以及对应相机的三维标定，我们分三步计算一个包围反射体所占据三维空间的凸多面体。首先，计算每个二维掩码的凸包，并使用道格拉斯-普克算法 [1973] 对其进行简化。由此得到每张图像对应的闭合折线 L_i i（图 4c）。接着，将所有内容提升到三维空间：利用已标定相机 c_i的信息，可以将 L_i 的每个顶点与 c_i的像平面上的对应三维点关联起来。每两个相邻顶点与相机的投影中心共同定义一个平面（图 4d 中的蓝色三角扇），每个平面将三维空间分成两个半空间，其中一个包含反射体，另一个则不包含。最后，我们求解满足所有约束的三维空间边界凸多面体 [Preparata 和 Shamos 1985]（图 4d 中的红色形状）。

5.2 损失

我们的架构针对每个特定场景进行端到端训练——这也是近期神经渲染方法的标准做法 [Kopanas 等人，2021年；Mildenhall 等人，2020年；Tewari 等人， 2021]。

我们的损失由五项组成：

$$\mathcal{L} = \lambda_{\ell_1}\mathcal{L}_{\ell_1} + \lambda_{\text{DSSIM}}\mathcal{L}_{\text{DSSIM}} + \lambda_{\text{P}}\mathcal{L}_{\text{P}} + \lambda_{m}\mathcal{L}_{m} + \lambda_{m_{TV}}\mathcal{L}_{m_{TV}}. \quad (4)$$

Lℓ_1 和 LDSSIM 分别根据 ℓ_1-范数和 DSSIM 指标，惩罚神经渲染器输出与对应输入视图之间的差异 [Loza 等人， 2006]，。LP促使反射点保持在反射体积内。为此，我们使用一个二值掩码 m_{RV}，该掩码由将反射体积投影到当前视图中得到。我们取反射点的（软）累积不透明度 $\bar{o}$（公式 3），并计算

$$\mathcal{L}_{\text{P}} = \|(\bar{o} - m_{\text{RV}})m_{\text{RV}}\|_1. \quad (5)$$

对于位于反射体积投影范围之外的点，此项不会产生影响，因此允许部分点移动到体积之外，但会促使点留在内部，以“填充”反射体表面。Lm 和 Lm_{TV} 对光栅化掩码 ρ进行正则化。Lm 根据 ℓ_1-范数将其与二值掩码 m_{RV} 进行比较，促使掩码在反射体上亮起。最后，Lm_{TV} 通过惩罚 ρ 的总变分来促使其保持平滑。图 6 展示了 ρ 掩码的示例。

对于所有场景的 θwarp，我们都使用 ADAM 优化器和权重衰减，并采用学习率调度。参数的取值为 $\lambda_{\ell_1} = 0.05$，
λ $\lambda_{m_{TV}} = \text{DSSIM} = 0.2$，
$1e-5$
$\lambda_{\text{P}} = \lambda_m = 0.01$，。

5.3 渐进式优化

我们通过渲染输入视图中的图像块，对点参数和神经网络进行端到端优化。每次迭代时，我们从一个随机输入视图中渲染一个随机的 150x150 像素图像块，并根据真实图像计算损失。我们每隔 20 次迭代反向传播一次梯度，这相当于使用批次，以避免内存饱和。

如果采用朴素的方法，优化的初期阶段可能不稳定；整个过程可能会退化，例如将反射点云移到相机视锥之外。为避免这种情况，我们采用多分辨率优化，并逐步增加

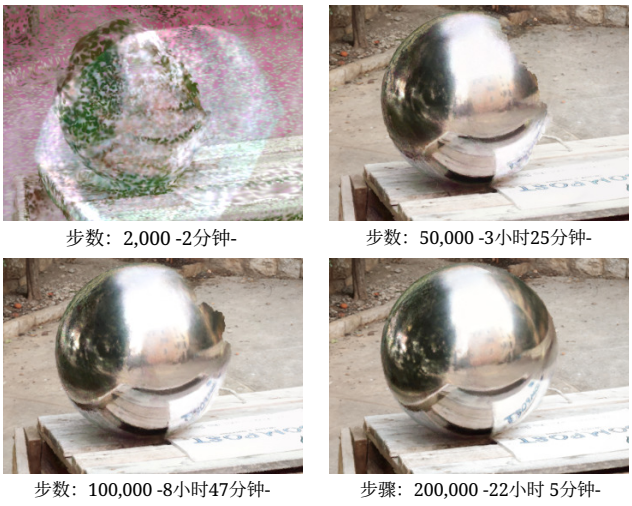

图 7. 优化过程中的最终渲染。左上方展示的是预热结束前的图像。右上方可以看到，质量有了显著提升，但镜面球体仍未完整重建。右下方则展示了完整重建的球体。

主点云的密度，以修正多视图立体视觉产生的大误差，同时保持足够的点密度。

多分辨率方法从低分辨率开始优化，即使用降采样图像，并在开始时每隔 32 个点取一个点，以缩减点云。通过处理较小的图像，也就是更大的图像区域，该方法能够为点的不透明度和六通道特征提供稳定的初始化。在一个简短的预热阶段中，我们进行两次上采样，每次都将分辨率加倍，并逐步恢复更多原始点。预热结束时，我们达到完整图像分辨率和原始点云规模，随后开始优化所有参数。我们在图 7 中展示了优化随时间的变化。

MVS重建可能无法重建场景中的某些区域，从而产生空白区域。这会导致点向这些空白区域移动，形成稀疏点云。为解决这个问题，我们每迭代 2K 次便加密一次点云：对于视空间位置梯度较大的点，我们沿梯度方向生成一个新点。

6 实现

接下来，我们将介绍光栅化、神经网络和交互式渲染器的实现细节。优化过程中的所有渲染以及所有结果均以 1000x666 的分辨率完成。所有源代码和数据均可在此处获取：
https://repo- sam.inria.fr/fungraph/neural_焦散线/

6.1 采集

我们围绕场景中的一个中心反射物体拍摄了数百张照片。随后使用现成的运动恢复结构（SfM）工具对照片进行校准，并生成稠密

点云，再使用多视图立体视觉 [Reality 2018], 初始化主点云。在实验中，我们使用 RealityCapture；在配备 Intel Xeon 5118 Gold 的电脑上处理所有场景时，运动恢复结构耗时 5–8 分钟，多视图立体视觉耗时 30–40 分钟。

6.2 训练用光栅化

我们使用 Pytorch 和自定义 CUDA 内核实现了我们的方法（灵感来自 [Kopanas 等人，2021年；Yifan 等人， 2019]）。虽然点的从前到后合成可直接求导，但若以朴素方式实现，计算光栅化特征相对于点位置和覆盖范围的梯度时，其复杂度会随每个像素中的点数呈平方增长。与近期的先前工作 [Lassner 和 Zollhofer 2021],不同，后者在反向传播中只使用一小部分光栅化点，我们使用所有光栅化点来获得稳定的梯度。为此，我们实现了自定义的两遍递归策略（先从前到后，再从后到前），在降低时间复杂度至线性的同时，得到精确的梯度。

6.3 网络细节

变形场是一个具有 4 层和 256 个特征的 ReLU 激活多层感知机。我们仔细地将相机位置和反射点云归一化到范围 $[-1, 1]$，并将该网络的输出乘以 0.01 进行缩小，以确保优化开始时位移的幅度非常小。

我们的环境贴图分辨率为 1000×400。神经渲染器是一个具有 ReLU 激活的多层感知机：编码器和解码器采用相同的架构，均有 32 个特征和 9 层。每隔 2 层，我们设置一个残差连接。我们使用 FixUp 对其进行初始化。我们选择多层感知机，而非以往工作中使用的卷积神经网络 [Kopanas 等人，2021年；Philip 等人， 2021]。卷积神经网络能够自然地生成高质量渲染结果，但会因屏幕空间滤波而产生时间伪影。此外，初步测试表明，对于我们的端到端优化，卷积神经网络在屏幕空间中修正伪影的能力，会削弱恢复准确三维场景表示的能力。

6.4 交互式渲染

我们发现，处理流程的主要瓶颈是主点云的光栅化，这限制了交互式渲染。为此，我们实现了一个基于 OpenGL 的交互式框架，通过硬件加速的 EWA 表面溅射近似光栅化 [Ren 等人 2002]。我们使用带深度剥离的可见性溅射来生成三个半透明渲染层，并将其合成以形成最终的特征张量。光栅化耗时 100 毫秒；结合处理流程其余部分未经优化的 Python Pytorch 实现，我们实现了 5 fps。我们在第 7.3 节分析由此产生的图像质量。

7 结果与评估

我们首先展示所提方法的结果，随后讨论对反射进行不同操作的应用，最后给出定量和定性评估。

我们采集了 5 个场景用于评估：一个室外场景 Com- post，以及四个室内场景 SilverVase、HallwayLamp、Con- caveBowl 和 CrazyBlade。后两个场景包含凹面反射器。由于我们对反射的运动进行建模，因此通过视频路径最能体现我们的结果；我们为所有场景提供了相应视频，并在补充材料中给出了比较。

7.1 结果

对于每个场景，我们都在单个 RTX8000 GPU 上对端到端模型进行了约 36 小时的优化。图 8 展示了与远离输入视图的路径所对应的真实图像进行对比的结果。我们的方法能够恢复并渲染清晰的反射，且反射运动平滑（请参阅补充视频）；即使面对 CrazyBlade 和 ConcaveBowl 中的凹面反射体等复杂情况，也能取得这样的效果。总体而言，反射的形状、位置和运动都接近真实值。

7.2 应用

除了新视角合成，我们的方法还可自然地应用于更广泛的领域：

7.2.1 反射中的对应关系。 我们的拉格朗日设计带来了一项强大的优势：能够直接支持不同视图中非刚性变形反射的稠密对应关系。在图 9 中，我们展示了如何通过追踪由神经焦散场定义的虚拟反射点的学习投影轨迹，忠实地跨不同视图跟踪反射点。我们设想，镜面表面重建 [Roth 和 Black 2006年；Sankaranarayanan 等人 2010] 以及反射物体的视图一致图像标注 [Caelles 等人 2017] 等下游应用，将能从这些稠密对应关系中显著受益。

7.2.2 场景与反射编辑。除了物理正确性之外，一些渲染应用还需要考虑艺术目标。反射尤其适合进行艺术性修改，因为它们能够传达空间关系，而人类观察者对偏离物理正确性的情况较为宽容 [Ramanarayanan 等人 2007]。针对合成场景的反射编辑工具已有开发 [Ritschel 等人 2009；Schmidt 等人 2013]。尽管我们的处理流程经过训练，能够忠实再现输入视图，但训练完成后，我们的方法允许通过将用于光栅化的相机与输入神经焦散场（负责扭曲反射）的相机解耦，来对反射进行富有表现力的编辑。这样，我们就能以在空间上和跨视图间都连贯的方式移动反射，从而可信地编辑反射透视（图 10 和补充视频）。

此外，我们显式的场景表示支持通用编辑。图 11 展示了一个物体克隆示例。为实现这一效果，我们复制希望克隆的体积中包含的主点，并对其进行变换。随后，我们复制反射点云，并对用于扭曲反射点的相机应用逆变换。请注意，目前我们只能对克隆物体进行平移；旋转可能需要专门的训练或参数化。

7.2.3 舒适立体视觉。反射的立体渲染（例如用于虚拟现实应用）颇具挑战，尤其是对于曲面反射体而言，



图 8. 我们的方法在 Compos 上的结果 [ConcaveBowl, HallwayBump, SpikyVase：左侧是我们的渲染结果，右侧是来自未包含在输入视图中的图像。请注意我们的渲染结果未能..... 真实地捕捉反射。



图 9. 我们的神经形变场公式能够自然地建立不同视图中反射之间的对应关系，即使存在严重形变，也能追踪反射。示例对应关系以彩色圆点标出；另请参阅补充视频。

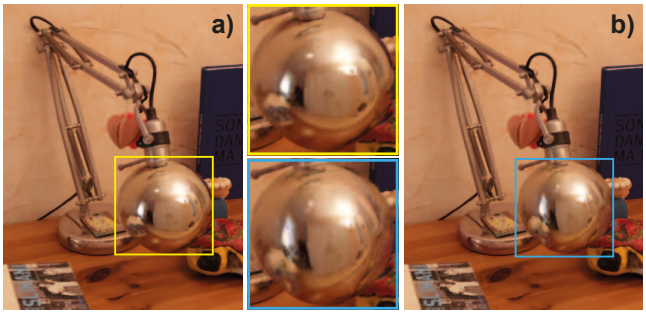

图 10. 反射编辑。（*a*）原始渲染。（*b*）编辑后的反射，放大桌子的右侧部分。编辑后的反射对应于一台位于主视图相机左侧的相机。

因为可能出现过大的水平和垂直视差，从而显著降低观看舒适度 [Lambooij 等，2009；Shibata 等，2011]。虽然已有方法可以将视差范围控制在舒适区间



图 11. 我们显式的场景表示支持多种编辑操作。（*a*）原始场景。（*b*）克隆反射点云以及部分主点云。两个球体上的反射各不相同，因此在新视角下呈现出合理的透视效果（见补充视频）。

对于漫反射外观 [Jones 等人，2001；Lang 等人，2010], 镜面表面对应的问题仅针对能够完全控制图像生成过程的合成场景进行了研究 [Dąbała 等人，2014；Templin 等人，2012]。

我们的方法能够控制随意采集场景中由曲面反射体产生的双目视差。同样，我们利用了这样一点：可以将用于光栅化的立体相机与用于引导形变场的立体相机解耦；当双目视差大到令人不适时，只需减小后者立体相机的瞳距（图 12）。极限情况下，若将单个（独眼）相机作为焦散场的输入，那么双眼的反射都落在学习得到的同一焦散线上，从而呈现漫反射场景元素的立体特性，且与反射体表面存在偏移 [Templin 等人 2012],，避免产生视觉不适。请注意，此修改不会影响处理流程中的漫反射（主）渲染分支。

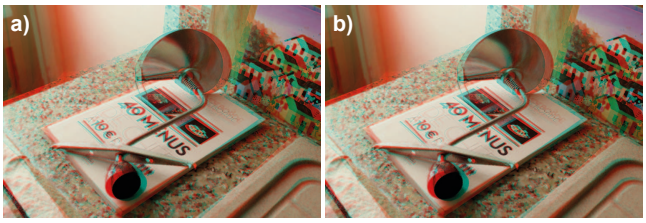

图 12. 我们的方法支持对反射进行舒适立体视觉渲染。（*a*）曲面反射体常会产生令人不适的双眼视差（在叶片左上方尤为明显）。（*b*）我们的方法可以显式调整视差，避免造成视觉不适。请佩戴红蓝眼镜获得立体视觉效果。

7.3 评估

我们从渲染质量方面对方法进行定性和定量评估。首先将其与先前工作进行比较，随后展示消融实验，以评估不同选择之间的权衡。

在定量测试中，除了输入视图外，我们还使用相同的设置采集了一组高速照片序列，帧率约为 10–15 fps。因此，我们获得了与输入视图完全不同的真实值图像序列，用于定量评估。为了评估反射的渲染质量，我们在图像中反射体积投影所覆盖的区域计算所有图像指标。定量评估所用的所有路径和各方法的渲染结果均见补充材料。我们从这些路径中选取了若干子序列，并提供高帧率插值，以便更好地评估反射质量和运动；相关内容也在补充材料中展示。

我们还在图 13 中展示了未优化的交互式渲染器对质量的影响，该渲染器的运行速度为 5 fps。

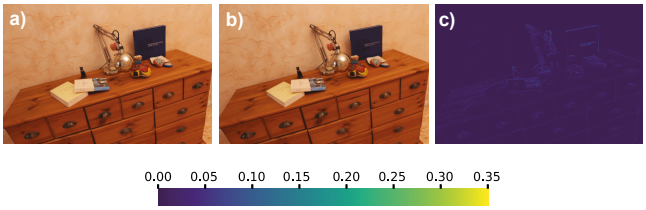

图 13。（*a*）完整渲染器。（*b*）交互式渲染器，采用我们未经优化的 Python 实现，运行速度为 5fps。（*c*）*a*）与*b*）之间的绝对归一化差异。

7.4 比较

我们将 Mip-NeRF [Barron 等人 2021] 与当前最先进且提供代码的 NeRF 方法进行比较 1，并与三种交互式方法进行比较：Deep Blending [Hedman 等人 2018], 提出的基于点的神经渲染（PBNR）方法 Kopanas 等人 [2021], 提出的方法，以及即时神经图形基元 [Müller 等人 2022]提出的方法。所有方法均使用相同分辨率的图像（1000x666）进行训练。我们使用默认批次大小训练 Mip-NeRF，迭代 1M 次，耗时约 48 小时。我们将所有渲染结果与我们完整的未优化方法进行比较；该方法目前的运行速度为 1.5 秒/帧。

[1]提交时，RefNeRF [Verbin 等人 2021] 的代码尚未公开。

图 14 展示了我们的方法与先前工作及真实值之间的定性比较。可以看到，在大多数情况下，我们的方法渲染出的反射更加清晰、准确。令人意外的是，Deep Blending 的表现非常出色；然而，在某些视图下它容易出现灾难性故障（见图 14 中的 ConcaveBowl，其中我们的方法恢复了缺失的几何结构）。Deep Blending 也无法随图像数量扩展：GPU 内存需求会随着图像数量增加而增长——对于当前的 GPU 而言，300 张图像大约就是这种方法所能处理的上限。相比之下，我们的方法在任何场景中使用的 GPU 内存量都大致相同（约 8Gb）。Deep Blending 和基于点的神经渲染在某些场景中会出现严重的闪烁和时间伪影；请参阅补充材料中的 Compost 和 ConcaveBowl。

表 1 展示了所有方法在全部场景上的平均定量结果。误差是在非输入视图上计算的，所用路径也不同于采集时使用的路径。我们计算 D-SSIM、峰值信噪比和学习感知图像块相似度误差；与其他所有方法相比，我们的方法取得了最佳定量结果，但仅略优于 Deep Blending。不过，Deep Blending 存在严重的时间伪影，而这些伪影无法通过这些指标捕捉（请参阅视频）。

表 1. 我们的方法与先前工作的定量评估结果，在五个场景上计算；评估所用路径是专门采集的独立路径，且不同于输入视图。

	$SSIM^{\uparrow}$	$PSNR^{\uparrow}$	$LPIPS^{\downarrow}$
Mip-NeRF [Barron 等人 2021]	0.9772	34.6753	0.0258
InstantNGP [Müller 等人 2022]	0.9769	34.3459	0.0253
基于点的 NR [Kopanas 等人 2021]	0.9790	34.3463	0.0229
Deep Blending [Hedman 等人 2018]	0.9832	35.6316	0.0197
Ours	0.9845	35.8522	0.0179

7.5 消融实验

我们还开展了消融研究，以更好地了解系统的动态特性以及各个组件的影响。我们的实验包括在 Compost 上运行 200k 次迭代（0.5–1 天）。消融实验包含以下不同元素：i) 仅主分量，即仅使用主点云渲染场景；ii) 无稠密化，即不对从 MVS 获得的原始点云进行稠密化；iii) 无LDSSIM 和 无Lm_{TV}，即将这两个损失项置零；iv) 半变形场多层感知机，即将变形场多层感知机的参数数量减少一半。

如表 2 和图 15、16 所示，我们方法的各个组件都能改善结果，其中神经变形场的重要性显而易见，稠密化和多层感知机容量也同样重要。仅主分量和半变形场多层感知机无法恢复反射的运动，因此渲染结果更加模糊。无稠密化配置难以弥补多视图立体视觉过程中几何结构重建中的较大误差。消融实验的另一个有趣发现是，两个损失项 LDSSIM 和

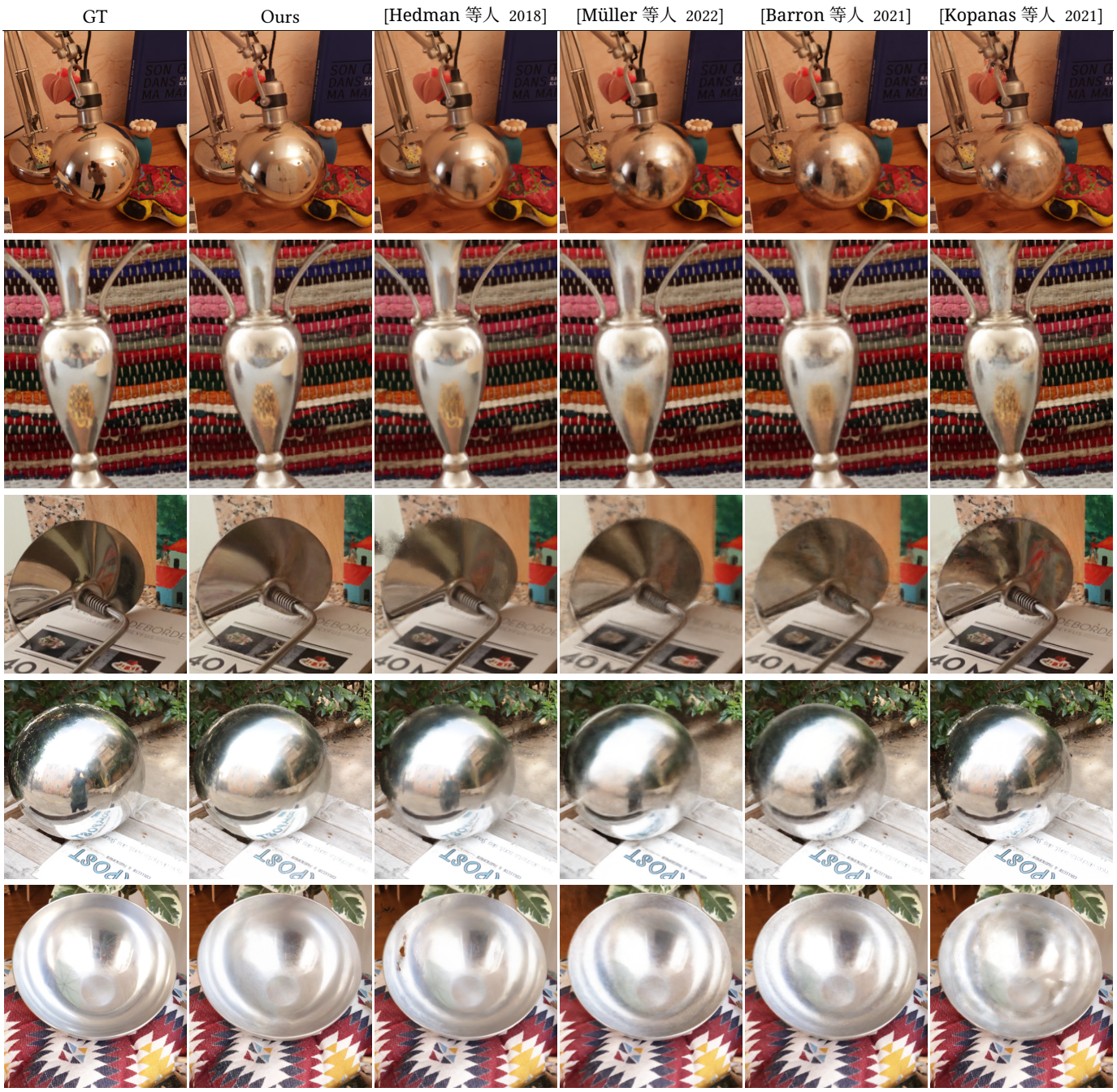

图 14. 我们展示了与先前方法的比较，以及来自输入视图之外路径的对应真实图像。从上到下的场景依次为：HallwayLamp、SilverVase、CrazyBlade、Compost、ConcaveBowl。

$L_{m_{TV}}$ 在很大程度上消除了反射中的摄影师。这也影响了表 2 中的数值结果，因为无 LDSSIM/$L_{m_{TV}}$ 配置实际上较少消除了摄影师，因此相对于真实值产生了偏差。根据具体应用，这可能是可取的，也可能不是。

7.6 遮挡、纹理与多次反射

我们的方法能够处理漫反射物体造成的遮挡、带有纹理的反射体以及多次反射。为说明这一点，我们专门为此采集了两个场景，如图 17 所示。

表 2. 使用与表 1 相同的方法，对算法各个组件进行消融研究的定量结果。

	$SSIM^{\uparrow}$	$PSNR^{\uparrow}$	$LPIPS^{\downarrow}$
仅主分量	0.9689	32.0426	0.0303
无稠密化	0.9727	32.6982	0.0273
无-$L_{DSSIM}/L_{m\ TV}$	0.9704	33.2101	0.0284
半变形场多层感知机	0.9690	32.5671	0.0301
Full	0.9745	33.7041	0.0254

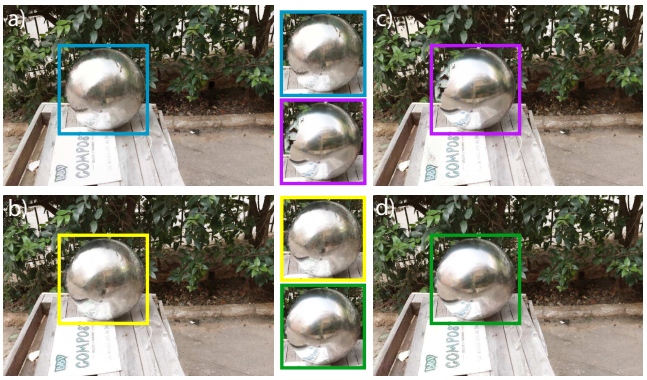

图 15. 我们通过消融实验展示了各个组件对方法结果的影响。a) 仅主分量，b) 半变形场多层感知机，c) 无稠密化，d) 完整方法。

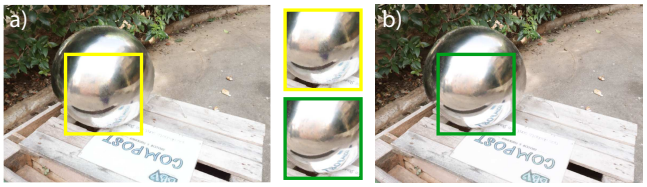

图 16. L_{DSSIM} 和 L^{m}_{TV} 具有令人惊讶的效果：能够从反射中去除摄影师。

当漫反射物体遮挡反射体时，主点云能够推断出遮挡关系，因为其中同时包含反射体和漫反射遮挡物。随后，这些信息会通过处理流程传递到掩码 ρ。具体而言，在漫反射遮挡物位于镜面反射点云前方的区域，ρ 为 0，因此不会渲染镜面反射点云。这一点可见图 6；图 17（第一行及局部放大图）展示了渲染示例。

镜面反射体（例如塑料或瓷器）可能带有漫反射纹理。我们的处理流程可以在不同的点云中分别对二者建模，并根据数值或 ρ 将它们融合，从而正确保留纹理。如图 17（第二行）所示，尽管瓷熊上有彩绘部分，反射仍得到了正确渲染。不过，我们预计高频纹理增多时，效果会下降。

多个具有多次反弹镜面反射效应的镜面反射物体通常会同时出现；当它们彼此接近时，可将其视为一个镜面反射点云



图 17. 两个场景展示了我们的方法如何处理反射体被非镜面场景遮挡的情况（第一行中，浇水壶前方的植被），以及如何处理多重反射（注意第二行中，球体的反射里熊身上的反射）。

（例如走廊中的灯座和灯罩）。对于彼此相距较远的多个反射体，应使用多个镜面反射点云。

由于我们并未显式建模物理光传输，随着场景复杂度增加，高阶反射和全局光照不会变得指数级复杂。这类反射只不过是变形场多层感知机需要

从输入图像中的观测结果提取的一种更复杂的运动。图 17（第二行及插图）展示了这一点：熊身上的反射清晰可见于球体的反射中。

7.7 与物理焦散的比较

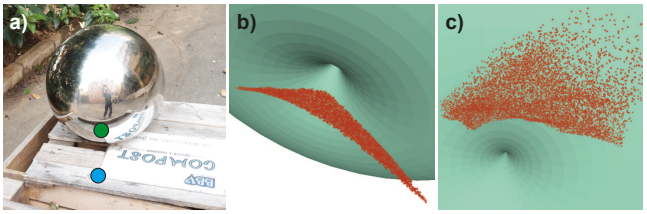

图 18. 神经点焦散与物理焦散的比较。(*a*) 我们考虑一个球面反射器，其焦散面可通过解析方法求得。(*b*,*c*) 几何构型的两个视图。青色网格表示*a*）中蓝色点对应的真实值焦散面。红点是从*a*）中绿色点对应的已学习神经点焦散中采样得到的。另请参阅正文。

由于基于点的表示固有的深度歧义（第 3 节；图 3d），对于特定反射而言，并不存在唯一正确的神经点焦散。我们的系统找到一 种能够实现高质量视角一致渲染的解。尽管如此，我们仍希望对由此产生的焦散几何结构进行定性分析。为此，我们考虑一个已知球面反射器的场景（图 18a），以及一个位置已知的反射点（图 18a 中的蓝点），并计算其解析焦散面（图 18b 和 c 中的青色网格） [Glaeser 1999]。然后，我们

随机采样可见相应反射（图 18a 中的绿点）的相机位置体积，并记录变形场对相应镜面点（图 18b 和 c 中的红点）的输出。尽管这两种解在定性特征上大致相似，但彼此相当不同，尽管它们产生的投影反射流几乎完全相同。

8 讨论与结论

我们的方法能够生成视觉质量更高、定量效果优于所有以往方法的曲面物体渲染结果；对反射点云进行显式表示还带来了额外优势，使我们能够进行多种场景操作，例如反射编辑、反射体克隆、舒适立体视觉渲染，以及跨视图的稠密反射跟踪。

尽管如此，我们的方法仍有局限。目前我们仅在反射体只覆盖场景较小部分的场景中进行了测试；我们预计，只要运动恢复结构/多视图立体视觉能够生成合理的场景初始近似，我们的方法就能取得较好的效果，但很可能仍会出现需要解决的新问题。

我们的工作重点是曲面反射体。平面是一个特殊情况（图 3(a)）；我们只需将镜面点移动到其唯一的静态位置。原则上，我们的方法也能处理这类情况，但需要调整初始化，以避免运动幅度过大。深度歧义（图 3(d)）赋予我们的变形场很大的自由度，使其能够选择点的轨迹而不损害反射质量。这往往会促使其选择较短且简单的轨迹（图 18）。

我们的拉格朗日模型假设，视点相关效应可以通过移动反射来建模。对于具有高频空间变化镜面材料的曲面，这一假设并不成立。尽管我们已证明，该模型能够处理曲面划痕或彩绘瓷器等较小变化，但我们认为，为这一引人入胜且颇具挑战性的情况寻找原则性解决方案，是未来的工作方向。

与大多数近期神经渲染方法一样，我们的方法的一个主要局限是，需要针对每个场景优化／训练模型。解码器多层感知机或许可以通过在一组不同场景上训练来适配；而要跨场景训练通用的神经变形场，则要困难得多。通过加速光栅化步骤，我们可以显著缩短训练时间。可探索多种方案：从训练期间交互式渲染器所用的分层方案变体入手，再考虑其他各种点云光栅化近似方法，不过反向传播时需要谨慎处理。

我们的交互式渲染器还有很大的提速空间。目前，基于点溅射的交互式渲染器实现并不理想，部分原因在于 Pytorch、CUDA 和 OpenGL 之间存在内存复制；我们预计，通过精心优化，可以在真正的实时渲染速度下获得更高质量。

总之，我们提出了一种理论基础扎实的拉格朗日方法，用于渲染采集的多视图场景中曲面物体的反射。我们希望，我们基于几何原理构建的方法以及直接基于点的解决方案，能够启发人们为其他神经渲染问题探索新颖的解决方案。

致谢

本研究由 ERC 高级资助项目 FUNGRAPH No 788065 资助 http://fungraph.inria.fr。作者感谢蔚蓝海岸大学的 OPAL 基础设施以及 GENCI–IDRIS 提供的高性能计算资源（资助项目 2022-AD011013409）。作者感谢匿名审稿人提出的宝贵意见，感谢 P.Hedman 校对早期稿件，感谢 T.Louzi 提供银瓶物体，并感谢 S.Kousoula 帮助剪辑视频，以及 S.Diolatzis 进行富有启发性的讨论。

参考文献

Kara-Ali Aliev、Artem Sevastopolsky、Maria Kolos、Dmitry Ulyanov 和 Victor Lempit- sky。2020年。神经点基图形学。载于*ECCV 2020*。施普林格，696–712。Jonathan T. Barron、Ben Mildenhall、Matthew Tancik、Peter Hedman、Ricardo Martin- Brualla 和 Pratul P. Srinivasan。2021年。Mip-NeRF：用于抗锯齿神经辐射场的多尺度表示。*ICCV*（2021年）。Mojtaba Bemana、Karol Myszkowski、Hans-Peter Seidel 和 Tobias Ritschel。2020年。X-Fields：隐式神经视图、光照与时间图像插值。*ACM*图形学汇刊（*SIGGRAPH Asia 2020*会议论文集）39，6（2020年）。https://doi.org/10.1145/3414685.3417827 James F Blinn 和 Martin E Newell。1976年。计算机生成图像中的纹理与反射。*Commun. ACM*19，10（1976年），542–547。James William Bruce 和 PJ Giblin。1992年。曲线与奇异性：奇异性理论的几何学导论。剑桥大学出版社。Brian Cabral、Marc Olano 和 Philip Nemec。1999年。反射空间图像渲染。载于第*26*届计算机图形学与交互技术年度会议论文集。165–170。Sergi Caelles、Kevis-Kokitsi Maninis、Jordi Pont-Tuset、Laura Leal-Taixé、Daniel Cremers 和 Luc Van Gool。2017年。单样本视频目标分割。载于*IEEE*计算机视觉与模式识别会议论文集。221–230。Min Chen 和 James Arvo。2000年。镜面路径扰动理论及应用。*ACM*图形学汇刊（*TOG*）19，4（2000年），246–278。Łukasz Dąbała、Petr Kellnhofer、Tobias Ritschel、Piotr Didyk、Krzysztof Templin、Karol Myszkowski、Przemyslaw Rokita 和 H-P Seidel。2014年。操控折射与反射双目视差。载于计算机图形学论坛，第33卷。Wiley Online Library，53–62。Paul J Diefenbach 和 Norman I Badler。1997年。多通道流水线渲染：动态环境中的真实感。载于*1997*年交互式三维图形学研讨会论文集。59–ff。David H Douglas 和 Thomas K Peucker。1973年。减少表示数字化线条或其简化图形所需点数的算法。*Cartograph- ica*：国际地理信息与地理可视化期刊10，2（1973年），112–122。Pau Estalella、Ignacio Martin、George Drettakis、Dani Tost、Olivier Devillers 和 Frédéric Cazals。2005年。曲面物体上精确的交互式镜面反射。载于视觉建模与可视化（*VMV 2005*）。柏林：Akademische Verl.-Ges. Aka，2005年，8。Wanquan Feng、Jin Li、Hongrui Cai、Xiaonan Luo 和 Juyong Zhang。2022年。神经点：利用神经场表示点云以实现任意上采样。（2022年）。James D Foley、Foley Dan Van、Andries Van Dam、Steven K Feiner、John F Hughes 和 J Hughes。1996年。计算机图形学：原理与实践。第12110卷。Addison-Wesley Professional。Stephan J Garbin、Marek Kowalski、Matthew Johnson、Jamie Shotton 和 Julien Valentin。2021年。Fastnerf：以200fps实现高保真神经渲染。载于*IEEE/CVF*国际计算机视觉会议论文集。14346–14355。Georg Glaeser。1999年。球体与旋转圆柱面上的反射。几何与图形学期刊3，2（1999年），121–139。Michael Goesele、Noah S navely、Brian Curless、Hugues Hoppe 和 Steven M Seitz。2007年。社区照片集的多视图立体视觉。载于*2007*年*IEEE*第*11*届国际计算机视觉大会。IEEE，1–8。Ned Greene。1986年。环境映射及世界投影的其他应用。*IEEE* 计算机图形与应用6，11（1986年），21–29。Markus Gross 和 Hanspeter Pfister。2007年。点基图形学。爱思唯尔。Yuan-Chen Guo、Di Kang、Linchao Bao、Yu He 和 Song-Hai Zhang。2021年。NeRFReN：带反射的神经辐射场。*arXiv*预印本 *arXiv:2111.15234*（2021年）。William Rowan Hamilton。1828年。光线系统理论。爱尔兰皇家学院学报（1828年），69–174。

Peter Hedman、Julien Philip、True Price、Jan-Michael Frahm、George Drettakis 和 Gabriel Brostow。2018年。用于自由视点基于图像的渲染的Deep Blending。*ACM*图形学汇刊（*SIGGRAPH Asia*会议论文集）37，6（2018年11月）。http://www-sop.inria.fr/reves/Basilic/2018/HPPFDB18 Peter Hedman、Pratul P. Srinivasan、Ben Mildenhall、Jonathan T. Barron 和 Paul Debevec。2021年。烘焙神经辐射场以实现实时视图合成。*ICCV*（2021年）。Graham R Jones、Delman Lee、Nicolas S Holliman 和 David Ezra。2001年。控制立体图像中的感知深度。载于立体显示与虚拟现实系统 *VIII*，第4297卷。SPIE，42–53。Alfrederic Josse 和 Françoise Pene。2014年。关于反射焦散的次数。代数通信42，6（2014年），2442–2475。Georgios Kopanas、Julien Philip、Thomas Leimkühler 和 George Drettakis。2021年。逐视图优化的点基神经渲染。载于计算机图形学论坛，第40卷。Wiley Online Library，29–43。Johannes Kopf、Fabian Langguth、Daniel Scharstein、Richard Szeliski 和 Michael Goesele。2013年。梯度域中的基于图像的渲染。*ACM*图形学汇刊（*TOG*）32，6（2013年），1–9。Marc Lambooij、Marten Fortuin、Ingrid Heynderickx 和 Wijnand IJsselsteijn。2009年。立体显示器的视觉不适与视觉疲劳：综述。*Journal of imaging science and technology* 53，3（2009年），30201–1。Manuel Lang、Alexander Hornung、Oliver Wang、Steven Poulakos、Aljoscha Smolic 和 Markus Gross。2010年。立体三维的非线性视差映射。*ACM*图形学汇刊（*TOG*）29，4（2010年），1–10。Christoph Lassner 和 Michael Zollhofer。2021年。Pulsar：高效的基于球体的神经渲染。载于*IEEE/CVF*计算机视觉与模式识别会议论文集。1440–1449。J Dennis Lawrence。2013年。特殊平面曲线目录。Courier Corporation。Anat Levin 和 Yair Weiss。2007年。利用稀疏先验，通过用户辅助从单幅图像中分离反射。*IEEE*模式分析与机器智能汇刊29，9（2007年），1647–1654。Gerrit Lochmann、Bernhard Reinert、Tobias Ritschel、Stefan Müller 和 Hans-Peter Seidel。2014年。实时反射与折射新视角合成。载于视觉、建模与可视化会议。9–16。A. Loza、L. Mihaylova、N. Canagarajah 和 D. Bull。2006年。基于结构相似性的目标跟踪：视频序列中的应用。载于*2006*年第*9*届国际信息融合会议。1–6。https://doi.org/10.1109/ICIF.2006.301574 Moustafa Meshry、Dan B Goldman、Sameh Khamis、Hugues Hoppe、Rohit Pandey、Noah Snavely 和 Ricardo Martin-Brualla。2019年。真实场景中的神经重渲染。载于*IEEE/CVF*计算机视觉与模式识别会议论文集。6878–6887。Ben Mildenhall、Pratul P Srinivasan、Matthew Tancik、Jonathan T Barron、Ravi Ramamoorthi 和 Ren Ng。2020年。NeRF：将场景表示为神经辐射场以进行视图合成。载于欧洲计算机视觉会议。施普林格，405–421。Don Mitchell 和 Pat Hanrahan。1992年。曲面反射体产生的照明。载于第*19*届计算机图形学与交互技术年度会议论文集。283–291。Thomas Müller、Alex Evans、Christoph Schied 和 Alexander Keller。2022年。采用多分辨率哈希编码的即时神经图形基元。*ACM* 图形学汇刊41，4，文章102（2022年7月），15页。Merlin Nimier-David、Delio Vicini、Tizian Zeltner 和 Wenzel Jakob。2019年。Mitsuba 2：可重定向的正向与逆向渲染器。*ACM*图形学汇刊（*TOG*）38，6（2019年），1–17。Eyal Ofek 和 Ari Rappoport。1998年。曲面物体上的交互式反射。载于第*25*届计算机图形学与交互技术年度会议论文集。333–342。Julian Ost、Issam Laradji、Alejandro Newell、Yuval Bahat 和 Felix Heide。2021年。神经点光场。*arXiv*预印本 *arXiv:2112.01473*（2021年）。Keunhong Park、Utkarsh Sinha、Jonathan T Barron、Sofien Bouaziz、Dan B Goldman、Steven M Seitz 和 Ricardo Martin-Brualla。2021年。Nerfies：可变形神经辐射场。载于*IEEE/CVF*国际计算机视觉会议论文集。5865–5874。Julien Philip、Sébastien Morgenthaler、Michaël Gharbi 和 George Drettakis。2021年。基于多视图立体视觉的自由视点室内神经重光照。*ACM*图形学汇刊（2021年）。http://www-sop.inria.fr/reves/Basilic/2021/PMGD21 Franco P Preparata 和 Michael Ian Shamos。1985年。凸包：基本算法。载于计算几何。施普林格，95–149。Ganesh Ramanarayanan、James Ferwerda、Bruce Walter 和 Kavita Bala。2007年。视觉等价性：迈向图像保真度的新标准。*ACM*图形学汇刊（*TOG*）26，3（2007年），76–es。Capturing Reality。2018年。RealityCapture重建软件。https://www.capturingreality.com/Product。Liu Ren、Hanspeter Pfister 和 Matthias Zwicker。2002年。物体空间EWA表面泼溅：一种实现高质量点渲染的硬件加速方法。载于计算机图形学论坛，第21卷。Wiley Online Library，461–470。

Gernot Riegler 和 Vladlen Koltun。2021年。稳定视图合成。载于*IEEE/CVF*计算机视觉与模式识别会议论文集。12216–12225。Tobias Ritschel、Makoto Okabe、Thorsten Thormählen 和 Hans-Peter Seidel。2009年。交互式反射编辑。*ACM*图形学汇刊（*TOG*）28，5（2009），1–7。Simon Rodriguez、Thomas Leimkühler、Siddhant Prakash、Chris Wyman、Peter Shirley 和 George Drettakis。2020a。用于交互式全局光照的光泽探针重投影。*ACM*图形学汇刊（*SIGGRAPH Asia*会议论文集）39，6（2020年12月）。http://www-sop.inria.fr/reves/Basilic/2020/RLPWSD20 Simon Rodriguez、Siddhant Prakash、Peter Hedman 和 George Drettakis。2020b。利用语义标签和近似反射流对汽车进行基于图像的渲染。*ACM*计算机图形学与交互技术论文集3（2020）。Stefan Roth 和 Michael J Black。2006年。镜面流与表面结构的恢复。载于*2006*年 *IEEE*计算机学会计算机视觉与模式识别会议（*CVPR'06*），第2卷。IEEE，1869–1876。Darius Rückert、Linus Franke 和 Marc Stamminger。2022年。ADOP：近似可微单像素点渲染。*ACM*图形学汇刊41（2022）。Aswin C Sankaranarayanan、Ashok Veeraraghavan、Oncel Tuzel 和 Amit Agrawal。2010年。基于稀疏反射对应关系的镜面表面重建。载于*2010*年 *IEEE*计算机学会计算机视觉与模式识别会议。IEEE，1245–1252。Thorsten-Walther Schmidt、Jan Novak、Johannes Meng、Anton S Kaplanyan、Tim Reiner、Derek Nowrouzezahrai 和 Carsten Dachsbacher。2013年。基于物理的光传输的路径空间操控。*ACM*图形学汇刊（*TOG*）32，4（2013），1–11。Takashi Shibata、Joohwan Kim、David M Hoffman 和 Martin S Banks。2011年。舒适区：预测立体显示器带来的视觉不适。视觉学杂志11，8（2011），11–11。Sudipta N Sinha、Johannes Kopf、Michael Goesele、Daniel Scharstein 和 Richard Szeliski。2012年。含反射场景的基于图像的渲染。*ACM*图形学汇刊（*TOG*）31，4（2012），1–10。Vincent Sitzmann、Michael Zollhöfer 和 Gordon Wetzstein。2019年。场景表示网络：连续的三维结构感知神经场景表示。载于神经信息处理系统进展。Richard Szeliski、Shai Avidan 和 Padmanabhan Anandan。2000年。从包含反射与透明度的多幅图像中提取图层。载于*IEEE*计算机视觉与模式识别会议论文集。*CVPR 2000*（目录号 *PR00662*），第1卷。IEEE，246–253。László Szirmay-Kalos、Tamás Umenhoffer、Gustavo Patow、László Szécsi 和 Mateu Sbert。2009年。图形处理器上的镜面效果：技术现状。载于计算机图形学论坛，第28卷。Wiley Online Library，1586–1617。Krzysztof Templin、Piotr Didyk、Tobias Ritschel、Karol Myszkowski 和 Hans-Peter Seidel。2012年。用于改善光泽呈现的高光微视差。*ACM*图形学汇刊（*TOG*）31，4（2012），1–5。Ayush Tewari、Ohad Fried、Justus Thies、Vincent Sitzmann、Stephen Lombardi、Kalyan Sunkavalli、Ricardo Martin-Brualla、Tomas Simon、Jason Saragih、Matthias Nießner 等。2020年。神经渲染技术现状。载于计算机图形学论坛，第39卷。Wiley Online Library，701–727。Ayush Tewari、Justus Thies、Ben Mildenhall、Pratul Srinivasan、Edgar Tretschk、Yifan Wang、Christoph Lassner、Vincent Sitzmann、Ricardo Martin-Brualla、Stephen Lombardi 等。2021年。神经渲染进展。*arXiv*预印本 *arXiv:2111.05849*（2021年）。Justus Thies、Michael Zollhöfer 和 Matthias Nießner。2019年。延迟神经渲染：使用神经纹理进行图像合成。*ACM*图形学汇刊（*TOG*）38，4（2019），1–12。Edgar Tretschk、Ayush Tewari、Vladislav Golyanik、Michael Zollhöfer、Christoph Lassner 和 Christian Theobalt。2021年。非刚性神经辐射场：从单目视频重建动态场景并进行新视角合成。载于*IEEE*国际计算机视觉会议（*ICCV*）。IEEE。Dor Verbin、Peter Hedman、Ben Mildenhall、Todd Zickler、Jonathan T Barron 和 Pratul P Srinivasan。2021年。Ref-NeRF：神经辐射场的结构化视点相关外观。*arXiv*预印本 *arXiv:2112.03907*（2021年）。Ziyu Wang、Liao Wang、Fuqiang Zhao、Minye Wu、Lan Xu 和 Jingyi Yu。2021年。MirrorNeRF：基于多镜面折反射成像的一次性神经肖像辐射场。载于*2021*年 *IEEE*国际计算摄影会议（*ICCP*）。1–12。https://doi.org/10.1109/ICCP51581.2021.9466270 Thomas Whelan、Michael Goesele、Steven J Lovegrove、Julian Straub、Simon Green、Richard Szeliski、Steven Butterfield、Shobhit Verma、Richard A Newcombe、M Goesele 等。2018年。重建具有镜面与玻璃表面的场景。*ACM*图形学汇刊37，4（2018），102–1。Turner Whitted。1979年。改进的阴影显示光照模型。载于第六届计算机图形学与交互技术年度会议论文集。14。Patrick Wieschollek、Orazio Gallo、Jinwei Gu 和 Jan Kautz。2018年。在真实环境中分离反射与透射图像。载于欧洲计算机视觉会议（*ECCV*）论文集。89–104。

Olivia Wiles、Georgia Gkioxari、Richard Szeliski 和 Justin Johnson。2020年。Synsin：从单幅图像端到端合成视图。载于*IEEE/CVF*计算机视觉与模式识别会议论文集。7467–7477。Suttisak Wizadwongsa、Pakkapon Phongthawee、Jiraphon Yenphraphai 和 Supasorn Suwajanakorn。2021年。NeX：利用神经基展开进行实时视图合成。载于*IEEE*计算机视觉与模式识别会议（*CVPR*）。Hao-Yu Wu、Michael Rubinstein、Eugene Shih、John Guttag、Frédo Durand 和 William Freeman。2012年。通过欧拉视频放大揭示世界中的细微变化。*ACM*图形学汇刊（*TOG*）31，4（2012），1–8。Yiheng Xie、Towaki Takikawa、Shunsuke Saito、Or Litany、Shiqin Yan、Numair Khan、Federico Tombari、James Tompkin、Vincent Sitzmann 和 Srinath Sridhar。2021年。视觉计算及其他领域中的神经场。*arXiv*预印本 *arXiv:2111.11426*（2021年）。

Jiamin Xu、Xiuchao Wu、Zihan Zhu、Qixing Huang、Yin Yang、Hujun Bao 和 Weiwei Xu。2021年。具有反射效果的可扩展的基于图像的室内场景渲染。*ACM*图形学汇刊（*TOG*）40，4（2021年），1–14。Qiangeng Xu、Zexiang Xu、Julien Philip、Sai Bi、Zhixin Shu、Kalyan Sunkavalli 和 Ulrich Neumann。2022年。Point-NeRF：基于点的神经辐射场。（2022年）。Wang Yifan、Felice Serena、Shihao Wu、Cengiz Öztireli 和 Olga Sorkine-Hornung。2019年。用于基于点的几何处理的可微表面泼溅。*ACM*图形学汇刊（*TOG*）38，6（2019年），1–14。Alex Yu、Ruilong Li、Matthew Tancik、Hao Li、Ren Ng 和 Angjoo Kanazawa。2021年。用于神经辐射场实时渲染的 PlenOctrees。载于*ICCV*。Matthias Zwicker、Hanspeter Pfister、JeroenVan Baar 和 MarkusGross。2001年。表面泼溅。载于第*28*届计算机图形学与交互技术年度会议论文集。371–378。