

神经体绘制：NeRF及其以外方法

NEURAL VOLUME RENDERING: NERF AND BEYOND

Frank Dellaert
Georgia Institute of Technology
dellaert@cc.gatech.edu

Lin Yen-Chen
Massachusetts Institute of Technology
yenchenl@mit.edu

参考论文总结“NEURAL VOLUME RENDERING: NERF AND BEYOND”，2021年1月，乔治亚理工和麻省理工的合著。

NEURAL VOLUME RENDERING: NERF AND BEYOND

Frank Dellaert
Georgia Institute of Technology
dellaert@cc.gatech.edu

Lin Yen-Chen
Massachusetts Institute of Technology
yenchenl@mit.edu

神经渲染 定义为：

“一种深度图像或视频生成方法，可显式或隐式控制场景属性，如照明、摄像机参数、姿势、几何、外观和语义结构”。

它是一种新数据驱动解决方案，解决了计算机图形学中长期存在的虚拟世界真实感渲染问题。

神经体渲染，是指通过跟踪光线进入场景并沿光线长度进行某种积分生成图像或视频的方法。通常，像多层感知器一样的神经网络将从光线的3D坐标到密度和颜色等的函数进行编码，并集成生成图像。

神经体渲染的直接先驱工作是用神经网络定义隐式曲面表示的方法。许多3D感知图像生成方法使用体素、网格、点云或其他表示，通常基于卷积架构。但在2019年CVPR上，至少有三篇论文介绍了用神经网络作为标量函数近似来定义占有率 (occupancy) 和/或符号距离函数 (SDF) 。

- Occupancy networks引入基于坐标的隐式占有学习。由5个ResNet块组成的网络，输入采用特征向量和3D点，预测二值占用。
- IM-NET用6层MLP解码器，在给定特征向量和3D坐标的情况下预测二值占有。可用于自动编码、形状生成（GAN-样式）和单视图重建。
- DeepSDF直接从一个3D坐标和一个潜代码中提取符号距离函数。它使用了一个8层MPL，其中到第4层有跳连接。
- PIFu表明，将3D点重新投影到像素对齐的特征表示中，可以学习特别详细的隐式模型。这个在PixelNeRF中会重演，效果非常好。

其他采用隐函数的方法有：

- Structured Implicit Functions (2019) 表明，可以组合这些隐式表示，例如，简单地求和。
- CvxNet (2020) 通过采用逐点Max (3D) 组合符号距离函数。
- BSP网络 (2020) 在许多方面类似于CvxNet，但在其核心使用二值空间划分，产生了一种局部输出多边形网格的方法，而不是通过昂贵的网格划分方法。
- Deep Local Shapes (2020) 将深度SDF潜代码存储在体素网格中，表示更大的扩展场景。
- Scene Representation Networks (2019) 或SRN，在架构上与DeepSDF非常相似，但添加可微分光线行进算法找到学习的隐式曲面最近交点，并添加MLP回归颜色，能够从多个姿态的图像中学习。
- Differentiable Volumetric Rendering (2019) 表明，隐式场景表示可以与可微分渲染器相结合，从图像中可以进行训练，类似于SRN。用“体渲染器”一词，但真正的主要贡献是一个巧妙的技巧，使隐式曲面的深度计算可微分：不在体上积分。
- Implicit Differentiable Renderer (2020) 提供了类似的技术，但具有更复杂的表面光场表示，表明它可以在训练期间优化摄像头姿态。

- Neural Articulated Shape Approximation (2020) 或 NASA, 组成隐式函数来表示关节目标, 如人体。
-

并非完全无中生有, 尽管仍然是基于 (warped) 体素的表示, 神经体 (Neural Volumes) 论文引入了用于视图合成的体渲染, 回归了密度和颜色的3D体。潜代码被解码成3D体, 然后通过体渲染获得新图像。

其提出在3D空间中的每个位置使用不透明度和颜色组成的体表示, 其中通过积分投影实现渲染。在优化过程中, 这种半透明的几何表示沿积分射线分散梯度信息, 有效地扩大了收敛范围, 从而能够发现良好的解决方案。

引起大家讨论的论文是神经辐射场NeRF论文 (2020年)。本质上, 用深度SDF架构, 但回归的不是有符号距离函数 (SDF), 而是密度和颜色。然后, 用 (容易微分的) 数值积分方法来近似真实的体渲染步骤。

NeRF模型将体场景表示存储为MLP的权重, 并在许多有已知姿态的图像上进行训练。通过沿每条查看光线以规则间隔集成密度和颜色来渲染新视图。

NeRF非常详细渲染的原因之一是, 使用周期性激活函数 (periodic activation functions), 即傅立叶特征, 对光线上的3D点和相关视图方向进行编码。这一创新后来被推广到周期性激活函数的多层网络, 即SIREN (Sinusoidal-Representation Networks)。这两篇文章都发表在NeurIPS

2020上。

可以说，NeRF论文的影响在于其残酷的简单性：只需一个MLP在5D坐标下输出密度和颜色。有一些华丽点缀（bells and whistles），特别是位置编码和分层抽样方案，但惊讶于如此简单的架构可以产生如此令人印象深刻的结果。尽管如此，原始NeRF留下了许多改进的机会：

- 训练和渲染速度都很慢。
 - 只能表示静态场景。
 - 固定照明。
 - 经过训练的NeRF表示不会推广到其他场景/目标。
-

一些项目/论文旨在改善原始NeRF论文的相当慢的训练和渲染时间。

- JaxNeRF (2020) 用JAX支持 (<https://github.com/google/jax>) 多设备训练，从几天变成几小时，大大加快速度。
- AutoInt (2020) 直接学习体积分，大大加快了渲染速度。
- Learned Initialization (2020) 使用元学习来寻找良好的权重初始化，加快训练速度。
- DeRF (2020) 将场景分解为“软 Voronoi图”，利用加速器内存架构。

- Nerf++ (2020) 建议用单独的NeRF对背景建模，处理无限场景。
- Neural Sparse Voxel Fields (2020) 将场景组织为稀疏体素八叉树，将渲染速度提高10倍。

至少有四项工作侧重于动态场景：

- Nerfies (2020) 及其底层D-NeRF对可变形视频建模，使用第二个MLP对视频的每一帧应用变形。
- D-NeRF (2020) 与Nerfies论文非常相似，甚至使用相同的缩略词，但似乎限制了平移的变形。
- Neural Scene Flow Fields (2020) 采用已知摄像机姿态的单目视频作为输入，深度预测作为先验，并输出场景流进行正则化，用于损失函数。
- Spatial-Temporal Neural Irradiance Field (2020) 仅使用时间作为额外输入。要成功训练这种方法来渲染任意视点视频（来自RGB-D数据！），需要仔细选择损失函数。
- NeRFlow (2020) 使用变形MLP对场景流建模，并跨时域进行积分，获得最终变形。
- NR NeRF (2020) 还用变形MLP对非刚性场景建模。除了摄像机参数之外，它不依赖预计算的场景信息，但与Nerfies相比，生成的输出稍微不那么清晰。

- STaR (2021) 将多视图RGB视频作为输入，并将场景分解为静态和动态体。但是，目前它只支持一个运动目标。

还有两篇论文关注人物的头像/肖像。

- PortraitNeRF (2020) 创建静态NeRF风格的头像 (Avatar)，但仅用一个RGB拍摄图像。为了实现这一点，需要灯光舞台训练数据。
 - DNRF (G2020) 专注于4D Avatar，将变形面部模型纳入流水线，施加强烈的归纳偏差。
-

NeRF风格方法得到增强的另一个方面是如何处理照明，通常是通过可用于重照明场景的潜代码。

- NeRV (2020) 用第二个“可见性”MLP支持任意环境照明和“一次反射”间接照明。
- NeRD (2020) 是另一项工作，其中使用局部反射模型，此外，对于给定场景，去除低分辨率球面谐波照明。
- Neural Reflectance Fields (2020) 通过添加局部反射模型改善NeRF，除密度外。它产生了令人印象深刻的重照明结果，尽管来自单点光源。
- NeRF-W (2020) 是NeRF的首批后续工作之一，优化潜外观代码，以便能够从控制较少的多视图集合中学习神经场景表示。

潜代码也可用于编码形状先验：

- pixelNeRF (2020) 更接近基于图像的渲染，其中在测试时用N个图像。它基于PIFu，创建像素对齐的特征，然后在评估NeRF-样式渲染器时进行插值。
 - GRF (2020) 在设置上非常接近pixelNeRF，但在规范空间而不是视图空间中运行。
 - GRAF (2020) ，即“辐射场的生成模型”，是NeRF的条件变型，添加了外观和形状潜代码，同时通过GAN风格训练获得视点不变性。
 - pi GAN (2020) ，与GRAF类似，但用了一种SIREN式样的NeRF实现，其中每一层都由不同的MLP输出进行调制，该MLP采用潜代码。
-

可以说，所有这些都不会扩展到由许多目标组成的大型场景，因此一个令人兴奋的新领域是如何将目标组成体渲染场景。

- Object-Centric Neural Scene Rendering (2020) 在以目标为核心的坐标系中学习“目标散射函数”，允许使用蒙特卡罗渲染合成场景并真实地照亮场景。
- GIRAFFE (2020) 通过以目标为中心的NeRF模型输出特征向量而不是颜色来支持合成，然后通过平均合成，并以低分辨率渲染为2D特征图，然后在2D中进行上采样。

- Neural Scene Graphs (2020) 支持场景图中的多个以目标为中心的NeRF模型。

最后，至少有一篇论文在（已知）目标姿态估计的背景下使用了NeRF渲染。

- iNeRF (2020) 在姿态估计框架中使用了NeRF MLP，甚至能够通过微调姿态改善标准数据集上的视图合成。然而，它还不能处理照明。

神经体渲染和NeRF风格的论文，最终是否会成功还远未可知。虽然现实世界中确实存在烟雾、烟雾、透明等，但最终大部分光线都从表面散射到眼睛中。由于基于体的方法，NeRF风格的网络可能很容易训练，但已经看到一种趋势，即论文试图在收敛后发现或猜测曲面。事实上，原NeRF论文中的分层抽样方案正是这样。因此，当从NeRF中学习时，可以很容易地看到移回SDF风格的隐式表示或甚至体素，至少在推理时是这样。