

计算机图形学课程报告

本报告内容为计算机图形学课程上高成英老师主讲的图形学前沿知识的一些回顾。主要介绍了三个部分：三维表征与渲染、神经辐射场的基本原理、神经辐射场的研究内容。让我对图形学三维重建模型和渲染有了新的理解，下面将简要总结在讲座中探讨的一些核心概念。

首先介绍了三维表征与渲染。传统图形学的渲染是基于图像进行三维重建，再基于不同视角进行变换与渲染，该传统渲染理论存在明显缺陷，如二维特征到三维信息的映射存在较大误差，很难得到高质量的三维模型，且三维重建的结果与渲染所需模型存在着差异，因此传统图形学不能很好的处理三维表征与渲染问题。而基于深度学习的渲染流程，将中间部分改为神经表达和可微分渲染，最后进行图像合成，实现从输入到输出的端到端映射，只需对场景信息进行隐式的表示和学习。其中隐式表达能够表达高质量且平滑的表面结构，优于显式表达适应不同的表面拓扑，但高精度的网络同时也需要对三维空间的密集采样。对于三维场景的神经表达有多种方法：RenderNet、Neural Volumes、NeRF 等，在本次讲座中重点介绍了 NeRF 的基本原理和研究内容。

NeRF，即神经辐射场，概念介绍为基于可微体渲染和神经场三维表征的新视角图像合成方法，将三维场景隐式地表示为连续的密度场与颜色场，其中涉及的重要特征有可微体渲染、连续的密度场与颜色场。NeRF 的渲染与优化过程是对多视角图像进行训练，与上述概念相关，学习图像的密度信息与颜色信息，密度信息依赖于位置，颜色信息依赖于方向，是合成图像的必要信息，隐式神经场便是用基于坐标的全连接网络表达颜色场与体密度场，神经场训练不断优化这两项参数，优化基于 NeRF 的表达式，最后利用体素渲染，累计颜色场与体密度场，得到对应视角的渲染图。

神经辐射场的预测具体是通过 $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ 确定光线， \mathbf{d} 为光线方向，沿着光线确定采样点 $\mathbf{r}_i = \mathbf{r}(t_i)$ ，通过隐式神经场 F_Θ 预测 \mathbf{r}_i 对应的密度值 σ_i 与颜色值 c_i ： \mathbf{r}_i 位置预测 σ_i 与几何特征，几何特征与 \mathbf{d} 预测 c_i 。

体素渲染则累计了光线上所有采样点的属性值来计算像素值，将场景看作是由彩色的微小颗粒组成的云，对于穿过场景的光线 $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ ，命中距离为 t 的粒子返回的是颜色值 $\mathbf{c}(t)$ ，根据命中第一个粒子概率： $P[\text{"no hits before"} t] \times P[\text{"hit at"} t] = T(t) \times \sigma(t)$ 将其看作每个采样点颜色值的权重，相乘最后得到 $C(\mathbf{r}) = \int_{t_n}^{t_f} T(t) \sigma(\mathbf{r}(t)) \mathbf{c}(\mathbf{r}(t), \mathbf{d}) dt$ ，由于该积分在实际应用中难以计算，通过在光线远、近端之间采样 N 个点，转换式子得到 $C(\mathbf{r}) = \sum_{i=1}^N T_i \alpha_i c_i$ ，

其中 α_i 为根据采样点及其命中率计算得到的不透明度， $T_i = \prod_{j=1}^{i-1} (1 - \alpha_j)$ 。当提到采样，就难免涉及采样点数量的问题，采样点过多会加大计算开销，过少则导致精度下降，因此提出了分层采样策略，即利用第一次采样点预测的密度值确定第二次采样点的位置，比如第二次采样时在 c_i 的权重 $T_i \alpha_i$ 高的地方侧重采样更多的点，能有效增大采样收益。

通过计算渲染结果与像素点真值间的误差来优化神经辐射场，在训练时，分别监督首次采样预测的像素值与第二次采样预测的像素值；在测试时，则仅利用第二次采样预测的像素值。

以上是 NeRF 的基本工作原理，NeRF 具有连续的表达方式和较强的灵活性，并且采用了端到端的优化方式，使得 NeRF 取得了较好的新视角合成效果。当然 NeRF 也存在一定局限性，如需要三维空间的密集采样，计算效率低，无法重建动态场景和泛化等操作，所以对 NeRF 的许多研究内容也与此挂钩。

NeRF 涉及到许多方面的研究内容,如文本驱动的三维内容生成、三维人脸生成与编辑、大规模城市街景地图重建、机器人定位与导航以及三维数字人重建,在研究内容中就涉及到 NeRF 无法重建动态场景的问题,对此也有相应的解决思路:使用变形场和标准场对动态场景进行建模,利用变形场将所有帧的数据变换到同一种标准表达形式,构造基于标准表达的标准场,在标准场内完成场景的隐式重建。对于动态数字人重建问题,通过隐式重建能够完成新视角的合成,通过隐式重建和驱动能够实现对于指定人体新姿势的图像合成。NeRF 还存在着效率优化问题,其训练时间和渲染时间长,训练需要较长的时间训练拟合 MLP 参数,构造空间内每个点与密度值、颜色值间的映射关系,计算瓶颈在于采样点网络查询时间,可以通过体素化及其分解(利用体素空间存储高维特征或轻量化网络,体素网格分解为低维平面网格表达,空间占用降为平方级),哈希网络轻量化(使用哈希技术压缩高分辨率的体素网格存储),以及稀疏几何表达(跳过无积分贡献的空间仅对重要点进行预测)。在 NeRF 的生成与编辑方面,通过将 NeRF 引入 GAN 模型学习三维空间,可以弥补传统 GAN 模型对三维空间缺少的理解,以及利用 CLIP 模型实现文本对 NeRF 生成场景的编辑,从 text-to-image 进一步到 text-to-3D。

总的来说,NeRF 实现了神经场和图形学组件的有效结合,隐式神经表示扩展了图形学新的发展领域,NeRF 的研究内容就 NeRF 存在的问题和 NeRF 的应用分为了两个大方向,其效果的应用仍需要时间和实验证明,仍具有很大的发展空间。