课程内容：

1. 三维表征与渲染

传统图形学渲染：(1) 基于图像三维重建 (2) 基于视角的变换与渲染

缺点： (1) 很难得到高质量的三维模型，二维特征到三维信息的映射存在较大误差

(2) 三维重建结果与渲染所需模型存在差异

基于深度学习的渲染流程：

输入图像→神经表达→可微分渲染→图像合成

显式表达与隐式表达：

显式表达：传统渲染，主要分为体素、点云和网络

缺点： (1) 表面表示方法难以适应不同的表面拓扑

隐式能够表达高质量且平滑的表面结构，然而高精度的网络需要密集采样

三维场景的神经表达方法：RenderNet、DeepVoxel、Neural Volumes、SRN、IDR、NeRF

1. 神经辐射场(NeRF)的基本原理

神经辐射场(Neural Radiance Fields, NeRF)：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**

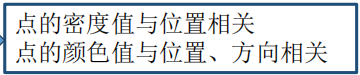
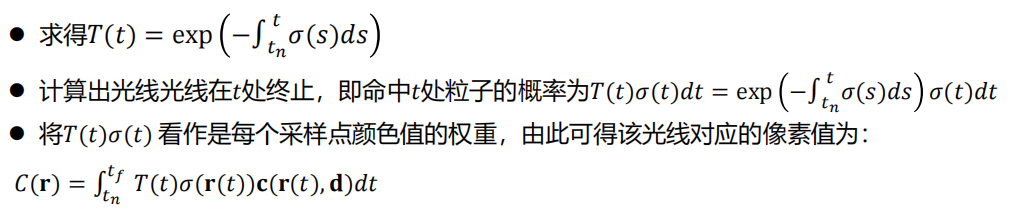
NeRF的渲染与优化过程：

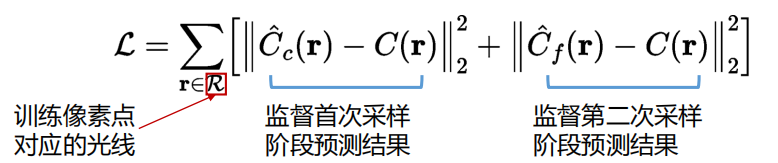
* **多视角图像**(标定好相机参数)进行训练
* 学习 依赖于位置的**密度信息**与依赖于方向的**颜色信息**
* 利用**体素**渲染得到对应视角的渲染图

输入多视角图像→优化基于NeRF的表达式→实现自由视角渲染

隐式神经场：用基于坐标的全连接网络表达颜色场与体密度场

体素渲染：累计颜色场与体密度场渲染为图像

* **r**(t) = **o**+t**d**确定光线 采样点(xi,yi,zi)=**r**i=**r**(ti) (ti是什么)
* 通过隐式神经场 **预测** 采样点ri 对应的**密度值与颜色值**
  + 采样点位置 预测 密度值与几何特征
  + 几何特征与采样点的方向 预测 颜色值
* 体素渲染：**累计光线上**所有采样点的属性值，计算像素值
  + P[“no hits before” t]=T(t) P[“hit at” t]= =𝜎(𝑡)𝑑𝑡
  + 积分难以计算：选择在光线远、近端之间采样N个点
  + 分层采样：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
* 计算渲染结果与像素点真值间的**误差** 优化神经辐射场
  + 损失函数
  + 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，仅利用第二次采样预测的像素值



位置编码：

* 直接使用r(ti)处的坐标进行预测，**难以表现高频细节** (什么是高频细节)
* 位置编码： 坐标映射为高频的形式

总结：NeRF为何取得较好的新视角合成效果？

1. 表达方式**连续**，具有较强的**灵活性**
2. 可以针对渲染结果，进行**端到端的优化**

局限性：

1. 计算效率低
2. 无法重建动态场景
3. 无法进行泛化
4. 重建的曲面表面质量低 等
5. 神经辐射场(NeRF)的研究内容

* 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
* 三维数字人重建
  + 问题：NeRF无法直接针对动态场景进行建模
  + 解决思路：使用**变形场+标准场**对动态场景进行建模
  + 任务：“新视角合成”隐式重建 “新姿势合成”隐式重建+驱动
* 效率优化：NeRF的训练时间长（数十个小时）、渲染时间长（至少30秒一张）
  + 训练耗时：需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
    - 直接存储并输出颜色、密度值？体素化及其分解哈希网络轻量化
    - 跳过无积分贡献的空间仅对重要点进行预测？稀疏几何表达
  + 渲染耗时：计算复杂度≈单个采样点的前传预测时间 X 采样点数量。
* 生成与编辑：传统的 GAN 模型缺少对三维空间的理解，生成的二维图像不能实现多视角一致性
  + 将NeRF引入GAN模型中学习三维空间
  + 研究主题：如何利用文本对生成的物体或场景进行编辑
  + CLIP-NeRF：利用CLIP模型实现文本对NeRF生成场景进行编辑
  + 研究动机：已有的text-to-image 模型取得成功
  + Dreamfusion：结合NeRF与扩散模型实现文本到3D的生成