

# 计算机图形学文献报告

21307347 陈欣宇

本次报告所读文献为《CityNeRF: Building NeRF at City Scale》，地图重建是 NeRF 的重要应用分支，这篇文献是来自上次高成英老师讲座介绍 NeRF 中的一篇论文，BungeeNeRF 是香港中文大学团队提出的场景还原模型，也称 CityNeRF。NeRF 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，这种方法在三维目标和受控场景下的建模同样具有优越的性能，应用价值十分广泛，本篇文献介绍的工作首次尝试将 NeRF 引入了城市规模，为传统卫星图像实现地图重建，引入更多尺度下的细节渲染，具有很大的实践意义。

首先概括 CityNeRF 模型需要解决的问题和使用技术，由于捕获城市概况的摄像机到场景的距离很宽，产生了不同细节量级的多尺度图像数据，难以显示建筑复杂的细节变换，这对于普通的 NeRF 也是一个巨大的难题。对此 CityNeRF 模型使用的是一种渐进式学习的方案，用一个渐进的神经辐射场来表示多种尺度之下的场景，根据尺度渐进划分不同网络层的工作，学习图像不同尺度下的细节。且该方式有效激活了位置编码中的高频通道，打开了更复杂的细节，对于视图急剧变化的城市规模场景，CityNeRF 具有优越的适应性。

普通 NeRF 使用多层感知机 MLP，将体积密度和颜色参数化作为输入坐标的函数，通过发出光线  $r(t)$  通过像素，对于光线上查询点  $r(t_k)$ ，取其傅里叶变换的特征得到位置编码，如公式  $(\tau_k, \mathbf{c}_k) = \text{MLP}(\gamma(\mathbf{r}(t_k)))$  最后输出体积密度和颜色。位置编码  $\gamma(\mathbf{x}) = [\sin(2^0 \mathbf{x}), \cos(2^0 \mathbf{x}), \dots, \sin(2^{M-1} \mathbf{x}), \cos(2^{M-1} \mathbf{x})]^T$  通过三维空间内位置的正弦和余弦组合映射到预定频率级  $M$  实现，根据经典体绘制方法进行优化，将估计的体积密度和颜色进行数值正交，得到近似体绘制积分  $\mathbf{C}(\mathbf{r}; \mathbf{t})$ ，最后根据  $\mathbf{C}(\mathbf{r}; \mathbf{t})$  计算损失函数。这种方法会导致一些问题，如果不同距离下的建筑一起训练，会导致近景的纹理模糊，远景的细节不完整，若不同尺度下场景单独训练又会导致连续的尺度变换之间渲染质量不一致，导致几何和纹理突兀。CityNeRF 的优势还体现在激活了位置编码在较深区域的高频通道，对于近距离视图，需要高频傅里叶特征来恢复细节，但这一点对于远程视图来讲是不必需的，普通 NeRF 不能很好地解决不同尺度下傅里叶特征频率通道不同的问题。

CityNeRF 使用多层监督的渐进式学习结构，将训练划分为多个阶段，预设训练阶段的总数 ( $L_{\max}$ )，该总数也代表摄像机和场景之间距离离散后的段数，每个阶段模型都同训练集同步增长，离散尺度的增加使用残差块结构，将新的 residual blocks 附加到网络中，访问地址编码的高频通道，以适应更近视图中出现的细节，残差块的输出头用来预测连续阶段间颜色和密度的残差，能够在距离变近时捕获场景新出现的细节。通过多层监督方式，在每个训练阶段，使用前面阶段更大尺度图像，即最远尺度到当前尺度的样本合并集，联合监督当前阶段残差块的输出头，当前头计算的损失是由所有已训练阶段的输出头聚合而成的，以保持所有尺度下图像渲染质量一致，避免不同尺度训练下细节的不一致问题。同时也将这一阶段的损失回馈到较大尺度的输出头上，为前一阶段的输出添加更多的复杂纹理细节。这种结构的监督也使得更深层次的输出头也保持了细节的复杂性，保证了尺度变化下的渲染视图质量，解决大尺度下出现伪影的问题。

CityNeRF 的测试结果令人满意，在整个测试数据集和每个尺度上都获得了优越的视觉质量，在远程尺度上有显著改善，与联合训练所有图像相比，可以产生最详细的中心目标，并在背景对象中获得良好的视觉质量，渲染视图的外围区域也更清晰、更完整。

总的来说，CityNeRF 通过渐进式学习方式实现了更细致的大尺度场景重建，为实际应用场景地图重建提供了强大可行的解决思路，也进一步完善了 NeRF 这一具有很大潜力的图像合成方法，相信能为未来相关领域的研究和发展打开新的方向。