

学 号:

0122210880110

武汉理工大学

课 程 设 计

题 目 计算机学科前沿讲座

学 院 计算机与人工智能学院

专 业 人工智能

班 级 2201

姓 名 代文博

指导教师 戚欣

2024 年 7 月 6 日

目录

计算机学科前沿课程设计报告—神经辐射场综述	2
1 NeRF 的概念	2
1.2 NeRF 的理论介绍	2
1.2 NeRF 的特点	4
1.2.1 隐式场景表示	4
1.2.2 自监督学习	4
1.2.3 真实感渲染	4
1.2.4 体积渲染	4
1.2.5 多视角一致性	4
1.2.6 可微分体积渲染	4
1.2.7 应用广泛	5
1.3 NeRF 的技术进展	5
1.3.1 速度提升	5
1.3.2 质量改进	5
1.3.3 稀疏视图/少样本训练	5
1.3.4 生成性和条件模型	6
1.3.5 场景编辑和组合	6
1.3.6 姿态估计	6
2 NeRF 目前研究现状	6
2.1 效率优化	7
2.1.1 利用稀疏几何表达	7
2.1.2 体素化	7
2.1.3 体素压缩（哈希表）	8
2.1.4 体素分解	8
2.1.5 体素分解（4D 推广）	9
2.2 人体重建与化身生成	9
2.2.1 动态人体化身	9
2.2.2 人与物体、场景交互	10
2.2.3 数字人生成	10
2.3 3D 生成	11
2.3.1 3D GAN 类别对象生成	11
2.3.2 3D GAN 类别对象生成（三平面改进）	11
2.3.3 3D GAN 类别对象生成（超分辨）	12
2.3.4 通用 3D 对象生成（2D 升维）	13
2.4 3D 编辑	13
2.4.1 物体/场景 NeRF 的编辑	13
2.4.2 基于 GAN 的 NeRF 编辑	14
2.4.3 基于 Diffusion 的 NeRF 编辑	14
2.5 人脸重建与化身生成	15
2.5.1 稀疏视点重建	15
2.6 场景建模	16
3 NeRF 未来发展方向	16
3.1 NeRF 年度发展趋势	16

3.1.1 高质量动态建模	16
3.1.2 与大模型的结合	17
3.1.3 更丰富的信息嵌入	17
3.1.4 应用到其他领域	17
3.2 NeRF 研究的展望	17
3.2.1 设计更高效的表征	17
3.2.2 准确分解几何、光照、材质	17
3.2.3 同时定位与多尺度场景建模	18
3.2.4 联合大模型实现语义联合生成	18
3.3 有关 NeRF 的调研	19
4 体会与总结	20
5 参考文献	21
本科生课程设计成绩评定表	27

计算机学科前沿课程设计报告

—神经辐射场综述

1 NeRF 的概念

1.1 NeRF 的理论介绍

Neural Radiance Field (NeRF)模型是一种新颖的视图合成方法，它使用体渲染和(通常)隐式神经场景表示，通过多层感知器(MLPs)来学习 3D 场景的几何和照明。Mildenhall 等人在 ECCV 2020 上首次介绍了 NeRF[1]，从那时起，它就已经达到了最先进的视觉质量，产生了令人印象深刻的演示，并激发了许多后续作品。

在其基本形式中，NeRF 模型将三维场景表示为由神经网络近似的辐射场。亮度场描述了场景中每个点和每个观看方向的颜色和体积密度。它被写成

$$F(\mathbf{x}, \theta, \varphi) \rightarrow (c, \sigma) \quad (1)$$

其中 $\mathbf{x}=(x,y,z)$ 是场景内坐标， (θ,φ) 表示方位角和极视角， $c=(r, g, b)$ 表示颜色， σ 表示体积密度。该 5D 函数由一个或多个多层预加速器 (MLP) 近似。两个视角 (θ, φ) 通常由 $\mathbf{d}=(dx, dy, dz)$ 表示，这是一个 3D 笛卡尔单位向量。通过将 σ (体积密度 (即场景的内容)的预测限制为与观看方向无关)，该神经网络表示被约束为多视图一致，而允许颜色 c 取决于观看方向和场景内坐标。在基线 NeRF 模型中，这是通过将 MLP 设计为两个阶段来实现的。

- 对于正在合成的图像中的每个像素，通过场景发送相机光线并生成一组采样点。
- 对于每个采样点，使用观看方向和采样位置来提取局部颜色和密度，由 NeRF MLP(s) 计算。
- 使用体绘制从这些颜色和密度产生图像。

更详细地说，给定体积密度和颜色函数，使用体积渲染[2]来获得任何相机射线 $\mathbf{r}(t)=\mathbf{o}+t\mathbf{d}$ 的颜色 $C(\mathbf{r})$ ，相机位置 \mathbf{o} 和观看方向 \mathbf{d} 使用

$$C(\mathbf{r}) = \int_{t_2}^{t_1} T(t) \cdot \sigma(\mathbf{r}(t)) \cdot c(\mathbf{r}(t), \mathbf{d}) \cdot dt \quad (2)$$

其中 $T(t)$ 是累积透射率，表示光线从 t_1 传播到 t 而不被拦截的概率，由

$$T(t) = e^{-\int_t^{t_1} \sigma(\mathbf{r}(u)) \cdot du} \quad (3)$$

$C(\mathbf{r})$ 通过待合成图像的每个像素。这个积分可以用数值计算。最初的实现和大多数后

续方法使用了非确定性分层抽样方法，将射线分成 N 个等间距的仓，并从每个仓中均匀抽取一个样本。然后，等式 (2) 可以近似为

$$\hat{C}(r) = \sum_{i=1}^N \alpha_i c_i T_i, \text{ where } T_i = e^{-\sum_{j=1}^{i-1} \sigma_j \delta_j} \quad (4)$$

δ_i 是从样本 i 到样本 $i+1$ 的距离。 (σ_i, c_i) 是根据 MLPs 计算的在给定射线的采样点 i 上评估的密度和颜色。 α_i 在采样点 i 处合成 alpha 的透明度/不透明度由

$$\alpha_i = 1 - \exp(-\sigma_i \delta_i) \quad (5)$$

可以使用累积的透射率计算射线的预期深度为

$$d(r) = \int_{t_1}^{t_2} T(t) \cdot \sigma(r(t)) \cdot t \cdot dt \quad (6)$$

这可以近似于方程 (4) 近似方程 (2) 和 (3)

$$\hat{D}(r) = \sum_{i=1}^N \alpha_i t_i T_i \quad (7)$$

某些深度正则化方法[3][4][5][6]使用预期的深度来将密度限制为场景表面的类似 delta 的函数，或增强深度平滑度。对于每个像素，使用平方误差光度损失来优化 MLP 参数。在整个图像上，这是由

$$L = \sum_{r \in R} \|\hat{C}(r) - C_{gt}(r)\|_2^2 \quad (8)$$

式中， $C_{gt}(r)$ 为与 r 相关联的训练图像像素的地面真色， r 为与待合成图像相关联的射线批次。NeRF 模型通常采用位置编码，Mildenhall 等[1]表明，这可以极大地改善渲染视图中的精细细节重建。这一点在文献[7]中也得到了更详细的证明，并使用了神经切线核 (Neural Tangent Kernels) 进行佐证。

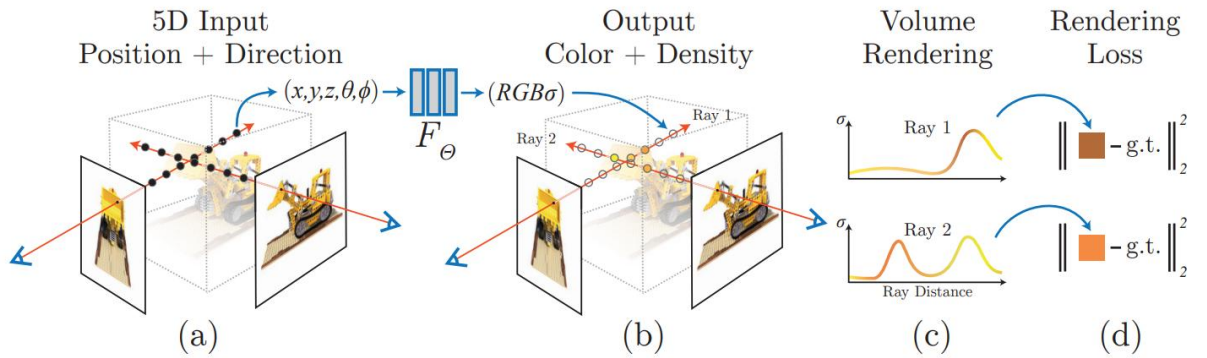


图 1 NeRF 体渲染和训练过程

(a)说明了待合成图像中单个像素的采样点的选择。(b)说明了使用 NeRF MLP 在采样点生成的密度和颜色。(c)和(d)分别说明了通过体绘制沿相关相机光线使用场景内颜色和密度

生成单个像素颜色，以及与地面真实像素颜色的比较。

1.2 NeRF 的特点

1.2.1 隐式场景表示

NeRF（神经辐射场）通过使用多层感知器（MLPs）来学习场景的连续体积密度和颜色信息。这种表示方法不需要显式的几何模型，而是隐式地通过 MLPs 的权重和激活函数来定义场景。这种隐式表示允许 NeRF 在渲染新视角时产生高度逼真的视觉效果，因为模型可以从任意位置和方向对场景进行采样。

1.2.2 自监督学习

NeRF 模型的一个关键优势是它们可以进行自监督学习，这意味着它们可以从多视角图像中学习，而不需要额外的 3D 标注或深度信息。这种学习方式简化了数据需求，因为只需要提供一组从不同角度拍摄的场景图像，模型就可以通过比较不同视角的图像来推断场景的三维结构。

1.2.3 真实感渲染

NeRF 生成的图像具有高度的真实感，这得益于其能够精确模拟光线在三维空间中的散射和遮挡。与传统的三维建模和渲染技术相比，NeRF 能够更自然地处理复杂光照和遮挡效果，从而产生更接近真实世界的视觉效果。

1.2.4 体积渲染

体积渲染是 NeRF 技术的核心，它通过对场景中每个点的颜色和密度进行建模，来合成新视角的图像。这种方法允许 NeRF 在渲染过程中考虑场景的体积属性，如雾化、透明度和散射，从而生成更加细腻和真实的图像。

1.2.5 多视角一致性

NeRF 模型在设计时就考虑了多视角一致性，确保从不同角度观察时，场景内容保持一致。这是通过优化模型参数来实现的，使得从任何给定视角生成的图像都能与实际场景相匹配，无论观察者的位置如何变化。

1.2.6 可微分体积渲染

NeRF 的体积渲染过程是可微分的，这意味着可以通过计算梯度来优化模型参数。这种可微分性是使用梯度下降等优化算法训练 NeRF 模型的基础，使得模型能够通过迭代过程不断改进，以更好地拟合训练数据。

1.2.7 应用广泛

NeRF 模型因其高质量的视觉效果和灵活性，在多个领域内都有应用。例如，在虚拟现实（VR）中，NeRF 可以用来生成逼真的三维场景，为用户提供沉浸式体验。在自动驾驶领域，NeRF 可以用于精确地理解和重建三维场景，这对于车辆的导航和决策至关重要。此外，NeRF 还可以应用于电影制作、游戏开发、建筑可视化等领域，为这些行业带来革命性的视觉效果和设计工具。

1.3 NeRF 的技术进展

1.3.1 速度提升

·烘焙技术

烘焙是一种优化技术，它将训练好的 NeRF 模型的参数和结构固化到一个更高效的数据结构中，如八叉树或哈希表。这样做可以减少在推理时的计算量，因为模型可以直接从优化过的数据结构中检索信息，而不是每次都从头开始计算。

·非烘焙模型

这些模型不依赖于预先烘焙的数据结构，而是在推理时动态地从训练数据中学习场景的特征。通过优化算法和硬件加速，这些模型可以在保持灵活性的同时提高速度。

1.3.2 质量改进

·Mip-NeRF[48]

这是一种改进的 NeRF 技术，它通过在不同尺度上表示场景来提高渲染质量。这种方法可以更好地捕捉场景的细节，从而在渲染时产生更高质量的图像。

·Ref-NeRF[50]

Ref-NeRF 通过引入参考视图来改善学习几何形状的质量。这种方法可以帮助模型更好地理解场景的三维结构，从而在合成视图时产生更准确的结果。

1.3.3 稀疏视图/少样本训练

这种方法通过利用预训练的卷积神经网络（CNN）来减少成功训练 NeRF 模型所需的训练样本数量。CNN 可以作为特征提取器，帮助模型从少量的样本中学习到更多的信息，从而在数据稀疏的情况下也能进行有效的训练。

1.3.4 生成性和条件模型

·基于 GAN 的 NeRF

生成对抗网络（GAN）可以用于生成 NeRF 模型，这些模型能够根据输入的条件（如文本描述或图像）生成相应的三维几何结构。

·基于扩散模型的 NeRF

扩散模型是一种生成模型，它通过逐步去噪来生成数据。在 NeRF 的上下文中，这种模型可以用来生成复杂的三维场景。

1.3.5 场景编辑和组合

NeRF 模型可以被训练来允许用户通过文本或图像输入来编辑场景。例如，用户可以指定要添加或移除的对象，或者改变场景的某些属性，如光照或颜色。

1.3.6 姿态估计

为了使用 NeRF 模型，需要准确的相机姿态信息。研究者们提出了结合同步定位与地图构建（SLAM）和束调整（BA）的方法来进行相机姿态估计。这些方法可以提供精确的姿态估计，从而提高 NeRF 模型的准确性。

2 NeRF 目前研究现状



图2 NeRF 研究现状

2.1 效率优化

·研究动机

朴素的 NeRF 训练时间长，渲染时间长。 其计算瓶颈在于：复杂度=单个采样点网络查询时间 x 采样点数量。

2.1.1 利用稀疏几何表达

·解决思路

利用稀疏几何表达（稀疏体素、八叉树、曲面等）排除对积分无贡献的采样区域，减少采样数。

·SNeRG, Plenoxels, Plenotrees, NSVF[8]

删去无几何区域体素，细化物体表面附近体素，得到稀疏体素或八叉树表达。

·MobileNeRF[9]

将 NeRF 提取到三角网格曲面的稀疏几何上，可利用光栅化在移动端实时渲染。

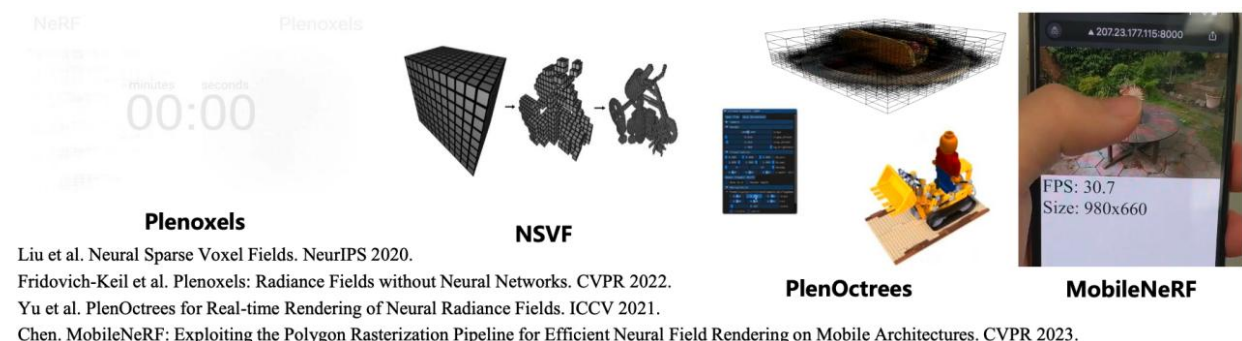


图3 稀疏几何表达

2.1.2 体素化

·解决思路

利用体素空间存储高维特征或轻量化网络，实现低复杂度查询。

·KiloNeRF[10]

空间体素化，每个体素用轻量网络，显著降低运算量并加渲染约数千倍。

·DVGO[11]

通过体素网格低密度初始化、插值后激活等训练策略直接优化体素表达的 NeRF 密度

场与颜色特征场，实现分钟级别的训练收敛。

2.1.3 体素压缩（哈希表）

·解决思路

使用哈希技术压缩高分辨率的体素网格存储。

·InstantNGP[12]

建立多尺度体素网格存储高维特征，将高分辨率网格用哈希压缩，可在低复杂度的条件下实现高分辨率与快速渲染。

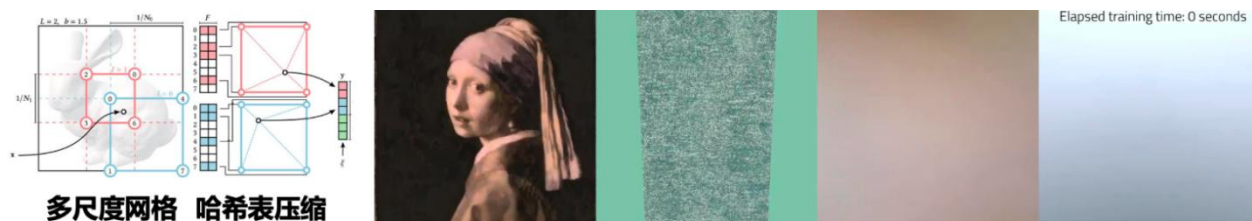


图4 体素压缩

2.1.4 体素分解

·解决思路

体素网格分解为低维平面网格表达，空间占用降为平方级。

·EG3D[13]

将三维坐标对应的体素特征定义为三个正交投影平面的特征。

·TensorRF[14]

将体素网格分解为向量-平面张量积形式的低秩张量之和。

·MeRF[15]

低分辨率体素+高分辨率平面投影。

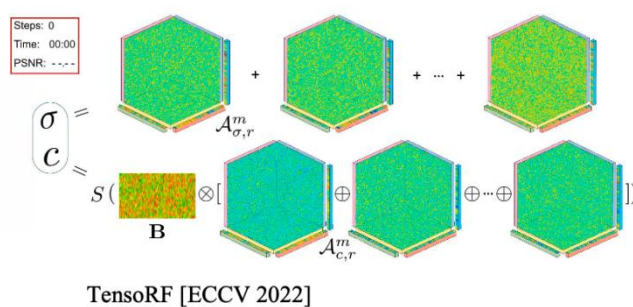


图5 体素分解

2.1.5 体素分解 (4D 推广)

·解决思路

沿用 3D->2D 的分解思路, 进行 4D->2D 的分解。

·Tensor4D[16]

4D 网格 -> 3 个 3D 网格 -> $3 \times 3 = 9$ 个 2D 网格。

·HumanRF[17]

4D 网格 -> 4 个 3D 网格与 1D 网格的张量积, 其中 3D 网格使用哈希压缩。

·HexPlane[18], K-Planes[19]

4D 网格->(x,y,z,t)坐标两两组合得到 6 个 2D 网格。

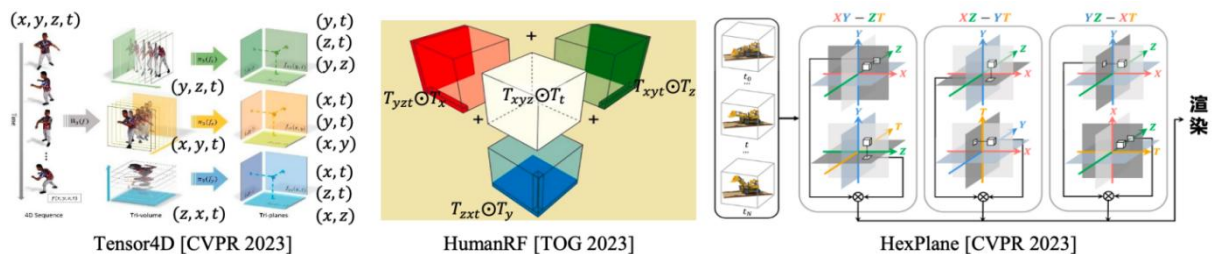


图 6 体素分解 (4D 推广)

2.2 人体重建与化身生成

·研究动机

动态 NERF 建模方法难以适用于人体大范围运动的场景。

·早期工作

以人体参数化模型 SMPL 为先验, 建立帧间大尺度骨架运动联系, 同时优化非刚性变形场与标准姿势下的 NeRF。

2.2.1 动态人体化身

·近期路线

更高质量的可驱动数字人, 关注动态衣服细节的建模。

·AvatarRex[20]

提出局部神经辐射场以及局部特征块以编码细粒度人体衣物细节。

·PoseVocab[21]

提出姿势表征库以编码不同姿态下的人体外观高频变化。

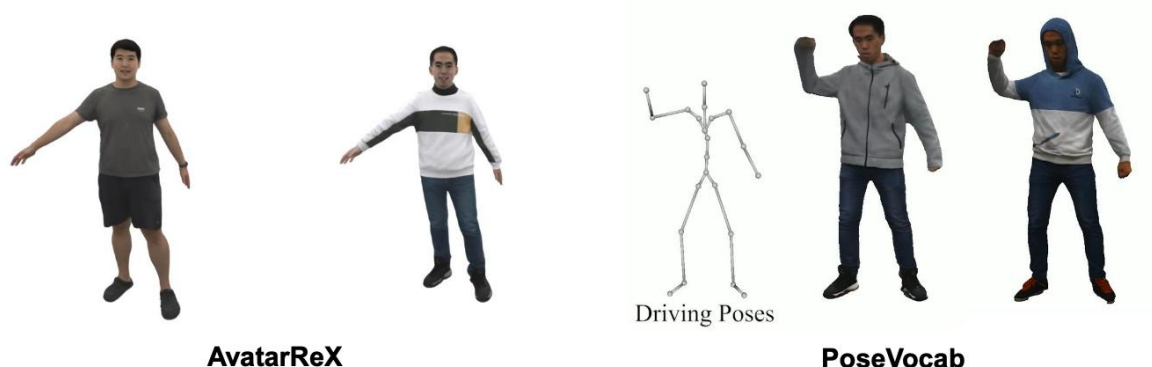


图 7 动态人体化身

2.2.2 人与物体、场景交互

·Instant-NVR[22]

结合非刚性跟踪以及 Instant-NGP 实现人和物体 NeRF 的在线重建。

·HOSNeRF[23]

引入状态隐编码以表征人和物体、场景的不同交互状态。

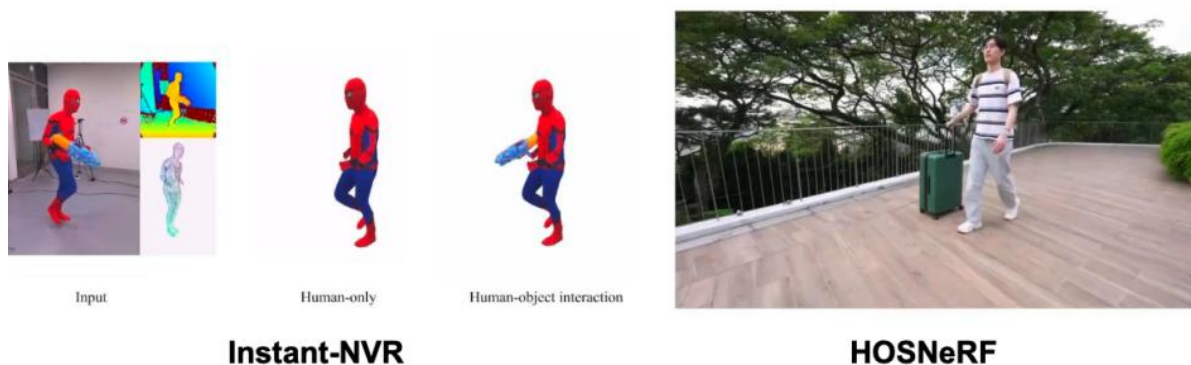


图 8 人景交互

2.2.3 数字人生成

·AvatarCLIP[24]

以 CLIP 为先验，分别生成静态数字人已经运动序列。

·EVA3D[25]

提出组合式人体 NERF，在标准空间中学习三维人体 GAN。

·DreamAvatar[26]

以 Stable Diffusion 为先验，约束基于 NeRF 渲染的图像满足语义输入。



图9 数字人生成

2.3 3D 生成

·研究动机

利用大规模 2D 图像先验，获得对象的生成式先验模型，以支持稀疏视点重建和各类编辑任务。

·近期路线

类别对象 3D 生成-> GAN,通用对象 3D 生成 -> Diffusion。

2.3.1 3D GAN 类别对象生成

·研究动机

NeRF 具有可微渲染的特点，可以从 2D 图片的监督中优化网络参数，因此将 NeRF 与 GAN 结合，构建生成式神经辐射场，学习 3D 内容生成。

·解决方法

基于神经辐射场的 MLP 网络，利用 GAN 的对抗式训练策略从 2D 图片中学习生成式神经辐射场，通过随机噪声产生隐式编码控制其几何与纹理[27]。

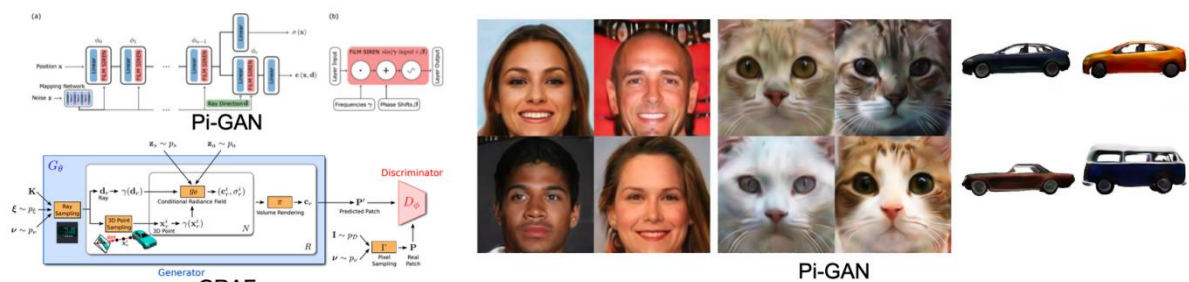


图10 3D GAN 类别对象生成

2.3.2 3D GAN 类别对象生成（三平面改进）

·研究动机

3D GAN 受限于 MLP 的显存消耗和表达能力，生成结果分辨率低。

·解决方案和创新性

提出基于三平面的三维表达，将神经辐射场的高频信号存储在三平面上从而轻量化 MLP 网络，在不损失表达能力的时候，大大降低显存消耗和提升渲染速率，利用高效的 2D styleGAN 生成具有高频细节的 triplane，从而提升生成质量，利用 2D 超分辨率提高渲染分辨率[28][29]。

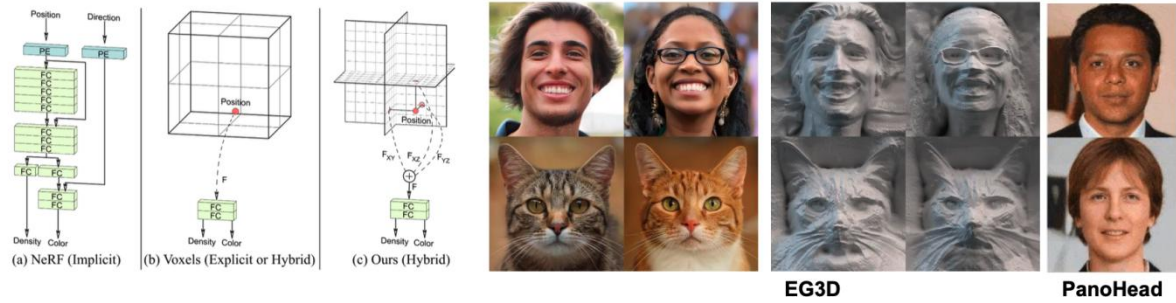


图 11 3D GAN 类别对象生成（三平面改进）

2.3.3 3D GAN 类别对象生成（超分辨）

·研究动机

2D 超分辨率网络将视角信息和图像特征耦合，破坏了三维一致性。

·解决方案和创新性

用 3D 超分辨率代替 2D 超分辨率。

·Gram-hd[30]

在神经辐射场中设置一组隐式曲面流形，并对曲面流形进行超分。

·Mimic3D[31]

通过让生成器的 3D 渲染分支合成的图像模仿其 2D 超分辨率分支生成的图像，使 3DGAN 能够生成高质量的图像，同时保持其严格的 3D 一致性。

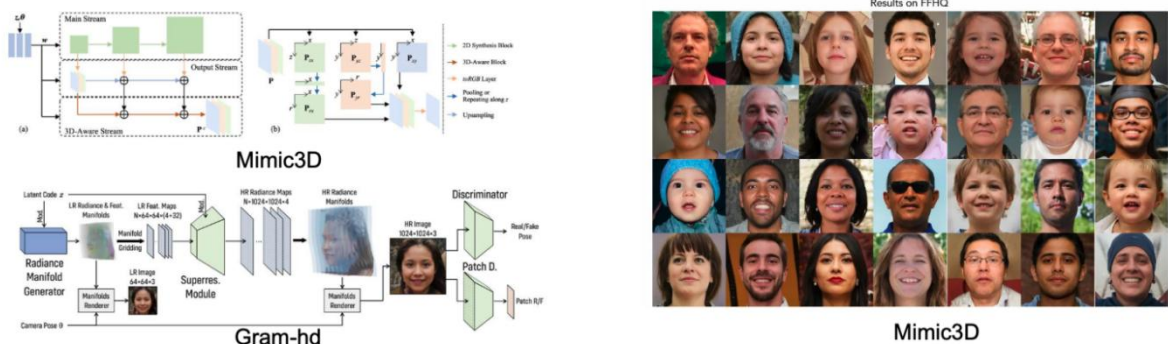


图 12 3D GAN 类别对象生成（超分辨）

2.3.4 通用 3D 对象生成（2D 升维）

·研究动机

2D 生成式大模型具有强大的文本生成图片能力;NeRF 具有表征连续复杂三维对象的能力,并且其渲染方式一种可微逆渲染,因此可通过 2D 监督反向优化辐射场的网络参数,实现通用物体或场景的三维生成。

·解决方案

将预训练 2D 生成式大模型作为先验,利用得分蒸馏采样(SDS)损失,最小化 NeRF 可微渲染图与扩散模型生成图像之间分布的 KL 散度,优化 NeRF 参数,实现文本到三维的生成。

·代表工作

Dreamfusion[32], Magic3D[33], Fantasia3D[34]。



图 13 通用 3D 对象生成（2D 升维）

2.4 3D 编辑

2.4.1 物体/场景 NeRF 的编辑

·研究动机

传统神经辐射场拟合或生成场景或物体,无法对其编辑。

·解决方案

利用不同的网络和隐含向量解耦形状和外观,用户在二维渲染图片上编辑,利用网络和隐含向量进行反向传播优化或前向编辑。

·早期工作

EditNeRF[35],NeRF-Editing[36],NeuMesh[37],ARF[38]。

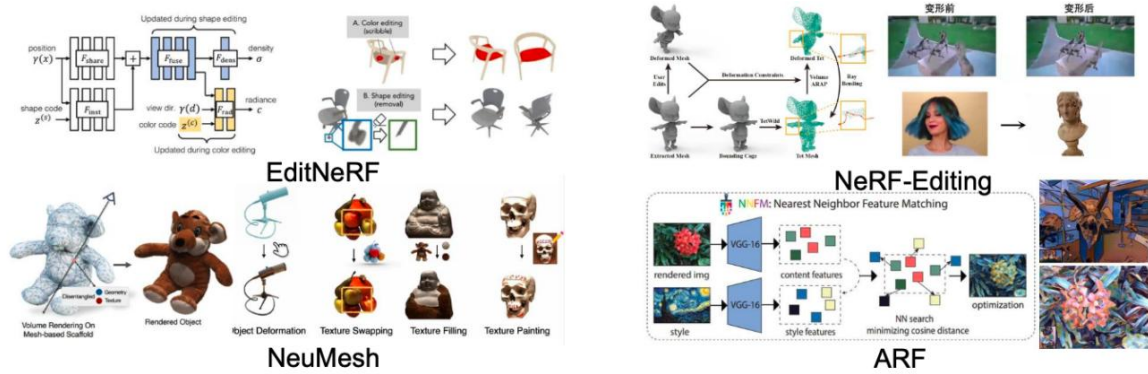


图 14 物体/场景 NeRF 的编辑

2.4.2 基于 GAN 的 NeRF 编辑

·研究动机

PiGAN, GRAF 等 3D GAN 生成丰富的三维人脸，但无法对其进行细粒度编辑。

·解决方案

将外部信号映射到神经辐射场，对其特征进行编辑。

·代表工作

IDE3D[39], NeRFaceEditing[40], AnifaceGAN[41], Next3D[42]。

·IDE3D

提出一个几何和材质解耦的生成式神经语义场，通过在几何分支网络中额外输出语义 mask，对齐三维语义和几何,编辑原理是 2D 语义图编辑映射到语义场，从而编辑三维语义和与其对齐的几何。

·Next3D

提出了一个基于神经纹理贴图的动态三平面表达，驱动表情信号会通过神经纹理光栅化，引起平面特征形变，进而渲染具有相应表情的图像。



图 15 基于 GAN 的 NeRF 编辑

2.4.3 基于 Diffusion 的 NeRF 编辑

·研究动机

基于文生图的扩散模型，利用文本对 NeRF 实现更直观，交互性更好的 3D 或 4D 编辑。

·解决方案

利用扩散模型不断迭代编辑训练集，同时优化神经辐射场参数，使得 NeRF 渲染结果和给定文本生成的编辑图像趋于一致。

·代表工作

InstructNeRF2NeRF[43], Instruct3D-to-3D[44]。

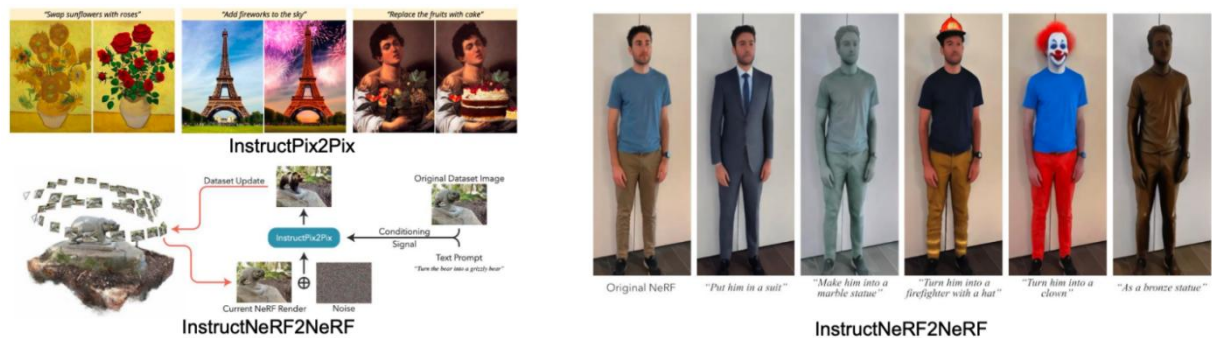


图 16 基于 Diffusion 的 NeRF 编辑

2.5 人脸重建与化身生成

2.5.1 稀疏视点重建

·研究动机

稀疏视点人脸重建，NeRF 容易过拟合到每个视点，新视点合成出现伪影。

·解决方案

引入人脸大数据，关键点和人脸模板等先验，优化 NeRF 重建质量。

·LP3D（静态+实时）

使用 EG3D 生成的人脸数据训练，输入单图像，推理三平面表达的 NeRF。

·HAvatar（动态化身）[45]

采用 3DMM 投影的三平面神经辐射场约束，实现高质量人头动态化身。

·NeRSemble（动态）[46]

引入 3DMM 表情参数，构建带表情语义空间变形场，拟合复杂表情动态。

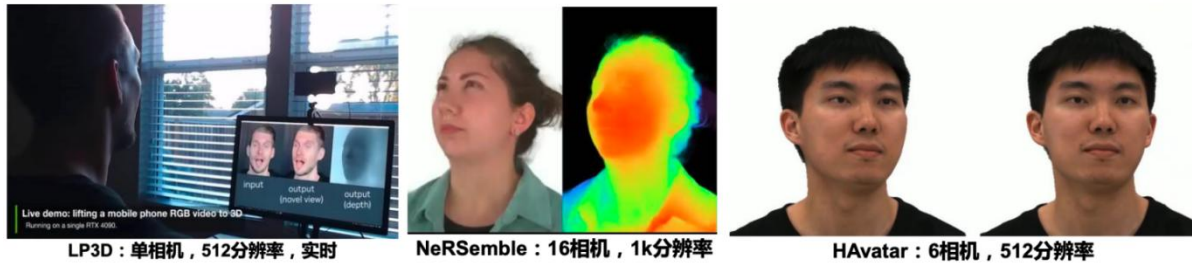


图 17 稀疏视点重建

2.6 场景建模

·研究动机

扩展 NeRF 表征大场景内容，允许对空间跨度大、几何纹理复杂的非结构化图像集合进行准确重建和新视点合成。

·早期工作

NeRF++[47]、Mip-NeRF[48]、Mip-NeRF 360[49]。

·解决思路

通过引入全空间非线性参数化模型，解决无界 3D 场景下 NeRF 建模问题，通过引入考虑采样点高斯区域的集成位置编码，解决 NeRF 在多尺度重建下模糊和混叠问题。

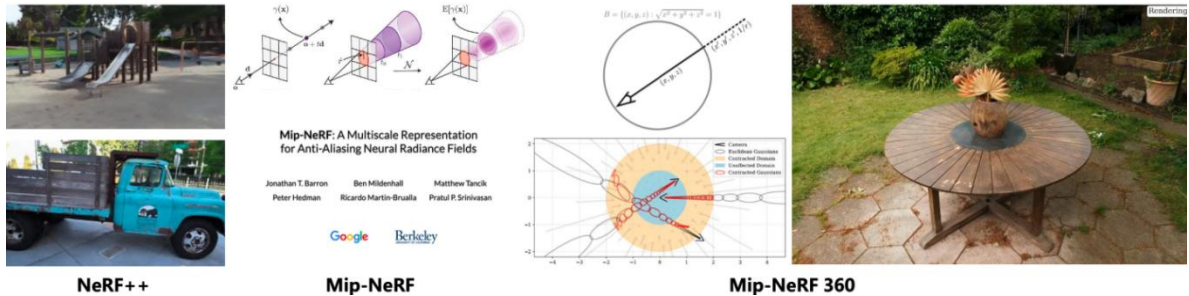


图 18 场景建模

3 NeRF 未来发展方向

3.1 NeRF 年度发展趋势

3.1.1 高质量动态建模

尽管 2022 年以前的 NeRF 方法在静态场景下表现优越，但是对于复杂动态场景的建模效果仍然存在改进空间。2023 年至今的大量工作在此方向作出努力，既包括了对于一般动态场景的 4D 建模改进，也有对于人脸人体的建模改进部分工作甚至在保证实时性的前提下取得了惊艳的效果。

3.1.2 与大模型的结合

大模型的落地应用已然势不可挡。2023 年至今有大量工作致力于将生成式大模型与 NeRF 相结合，从而实现 NeRF 的生成创作。与大模型结合之后，NeRF 不再局限于重建现实物体或场景，而是具备了“无中生有”的创造力。

3.1.3 更丰富的信息嵌入

2022 年以前的 NeRF 工作主要着眼于新视点渲染，因此只考虑了几何和纹理的建模。2023 年至今的工作中，研究者为 NeRF 引入了更多的信息，包括丰富的材质属性，以及更高层次的语义内涵。语义信息的引入则进一步拓宽了 NeRF 的潜在应用场景。

3.1.4 应用到其他领域

在 2022 年，NeRF 仅仅在三维视觉领域受到关注。在 2023 年，NeRF 实现了“破圈”，在机器人、自动驾驶医疗等领域也有了应用，其新视点生成能力能够有效辅助这些领域的数据生成与场景理解。

3.2 NeRF 研究的展望

3.2.1 设计更高效的表征

未来研究可能会探索新的神经网络架构，以提高表征效率和渲染速度。例如，通过自适应采样技术，只在场景的关键区域进行高密度采样，而在其他区域减少采样点，从而减少计算量。

研究者可能会进一步优化网络的权重和激活函数，以实现场景更紧凑、更高效的编码。此外，利用先进的压缩技术，如知识蒸馏和网络剪枝，也有望减少模型大小，提高推理速度。

3.2.2 准确分解几何、光照、材质

为了更准确地模拟现实世界的复杂性，未来的研究可能会专注于如何将几何、光照和材质属性更清晰地分离和建模。这可能涉及到开发新的网络结构，能够独立地学习和表示这些属性。

通过引入物理基础的渲染方程，可以更真实地模拟不同材质的光学特性，以及光照与材质之间的相互作用。此外，利用机器学习技术，如对抗训练，可以进一步提高渲染结果

的真实感。

3.2.3 同时定位与多尺度场景建模

为了实现对大规模场景的有效建模，未来的研究可能会探索结合同步定位与地图构建（SLAM）技术，以实现对场景的精确定位和建模。

多尺度建模方法将被进一步发展，以处理从微观细节到宏观结构的不同层次。这可能包括开发能够处理不同尺度特征的网络，以及能够在全局和局部之间进行有效切换的算法。

3.2.4 联合大模型实现语义联合生成

将 NeRF 与大型语言模型或视觉模型结合，可以实现基于语义输入的场景生成。例如，通过将 NeRF 与自然语言处理（NLP）模型相结合，用户可以描述他们想要的场景，然后由模型生成相应的三维视图。

利用生成对抗网络（GAN）或扩散模型等大模型，可以进一步丰富 NeRF 的生成能力，使其能够创造出前所未有的场景和物体。这些模型可以提供强大的先验知识，帮助 NeRF 在有限的数下也能生成高质量的三维内容。

随着计算资源的日益丰富和算法的不断优化，NeRF 技术有望在未来几年内实现跨越式的发展，为三维视觉领域带来更多创新和突破。通过上述研究方向的深入探索，NeRF 将能够更好地服务于虚拟现实、增强现实、自动驾驶、机器人视觉等多个领域，推动相关技术的快速发展和广泛应用。



图 19 NeRF 研究的展望

3.3 有关 NeRF 的调研

神经辐射场（NeRF）作为一项革命性的三维重建技术，自其诞生以来便在计算机视觉领域引起了巨大关注。它通过深度学习模型模拟光线在三维场景中的传播，生成逼真的三维视觉内容，为多个领域带来了前所未有的机遇。在 2024 年的计算机视觉与模式识别会议（CVPR）上，众多研究者对 NeRF 技术进行了深入的研究和讨论，提出了多项创新性的改进方案，进一步推动了该技术的发展。

一项引人注目的研究是 GenN2N[51]，它是一种适用于多种 NeRF 编辑任务的 NeRF-to-NeRF 通用框架，其核心在于用生成的方式来刻画编辑过程多解性，使其可以借助生成式编辑轻松产生大量符合要求的编辑结果供用户挑选。

GenN2N 的核心部分引入了 3D VAE-GAN 的生成式框架，使用 VAE 表征整个编辑空间，来学习与一组输入的 2D 编辑图像对应的所有可能的 3D NeRF 编辑分布，并用 GAN 为编辑 NeRF 的不同视图提供合理的监督，确保编辑结果的真实性。GenN2N 使用对比学习解耦编辑内容和视角，确保不同视角间的编辑内容一致性。在推理时，用户简单地从条件生成模型中随机地采样出多个编辑码，就可以生成与编辑目标对应的各种 3D 编辑结果。

相比于各种 NeRF 编辑任务的 SOTA 方法（ICCV2023 Oral 等），GenN2N 在编辑质量、多样性、效率等方面均优于已有方法。

CVPR 2024 还展示了多篇关于 NeRF 的研究，这些研究覆盖了场景理解、动态场景捕捉、多视角合成等多个维度。这些研究表明，NeRF 技术不仅在理论上具有广泛的应用前景，而且在实际应用中也展现出了巨大的潜力。通过这些研究，我们可以看到 NeRF 技术在多个领域的应用，如虚拟现实、增强现实、游戏开发、电影制作等，都有望带来创新和突破。

尽管 NeRF 技术取得了显著的进展，但在实际应用中仍存在一些挑战。例如，在处理复杂场景时，模型的训练成本和资源消耗仍然较高；对于动态场景和多视角重建的适应性也需要进一步优化。为了克服这些挑战，研究者可以采取以下策略：

- 优化模型架构和算法流程，以减少计算和存储需求。例如，可以采用更精简的神经网络，或者开发更高效的采样和渲染技术。通过这些优化，NeRF 模型可以在保持高质量输出的同时，减少对计算资源的依赖，使得其在资源受限的环境中也能发挥作用。

- 将 NeRF 与其他计算机视觉技术融合，以增强其在复杂环境中的表现。例如，可以结合注意力机制来提升模型对关键信息的捕捉能力。这种融合不仅可以提高模型的性能，还

可以扩展其应用范围，使其能够更好地适应各种复杂的场景和任务。

推动 NeRF 技术在实际场景中的应用，探索其在自动驾驶、虚拟现实、增强现实等领域的潜力，以期为日常生活带来更多创新和便利。通过将 NeRF 技术与实际应用场景相结合，可以进一步验证和优化其性能，同时也能够为相关领域带来新的解决方案和思路。

通过这些努力，NeRF 技术有望在未来的三维重建和计算机视觉领域发挥更大的作用。随着技术的不断进步和应用的不断拓展，我们有理由相信，NeRF 将会成为推动相关领域发展的重要力量，为我们的生活带来更多的可能性和惊喜。

4 体会与总结

在探索计算机视觉和图形学的广阔天地中，三维重建技术无疑是一颗璀璨的明珠，它将现实世界的复杂性转化为数字世界中的精确模型。然而，长久以来，这一领域的技术进步似乎总是被高昂的成本和复杂的计算过程所束缚。但就在这样的背景下，神经辐射场（NeRF）技术的横空出世，犹如一束照亮黑暗的光，为三维重建带来了前所未有的突破。

NeRF 技术的诞生，标志着我们对三维世界理解的一次质的飞跃。它不仅仅是一种技术，更是一种全新的视角，一种将深度学习的力量与三维空间的几何美学完美结合的创新。通过分析一系列从不同角度拍摄的图像，NeRF 能够构建出一个连续的、能够精确描述三维场景的函数。这就像是赋予了计算机一种全新的视觉能力，让它能够以一种前所未有的方式“看到”并理解我们周围的世界。

NeRF 的实现过程，是一场视觉与智能的盛宴。它首先从多视角图像中提取出关键的采样点，这些点就像是散落在三维空间中的珍珠，每一颗都蕴含着物体表面的几何和颜色信息。然后，这些宝贵的信息被输入到一个精心设计的深度神经网络中，网络中的每一个神经元都像是一个小小的画家，共同绘制出场景中每个点的光学属性。

接下来，体素渲染技术的运用，就像是将这些点连成线、线成面、面成体的过程，它根据颜色和密度信息，渲染出从任意视角观察的场景图像。这个过程涉及到光线追踪和渲染算法，就像是在模拟光线在三维空间中的旅行，最终生成一幅幅符合人类视觉感知的图像。

NeRF 的优势在于其能够生成令人惊叹的新视角图像，这些图像不仅细节丰富，而且质量极高。它在处理复杂场景时的表现，就像是一位技艺高超的艺术家，能够巧妙地避免空洞和纹理模糊等常见问题。更重要的是，NeRF 的神经网络基础为其提供了强大的可扩展

展性，这意味着它可以通过不断的学习和训练，不断提升重建的精度和质量。

在实际应用中，NeRF 技术的潜力已经得到了充分的展现。无论是在虚拟现实、增强现实、游戏开发还是电影制作等领域，NeRF 技术都能够帮助我们创造出更加逼真、更加生动的三维模型。它就像是一把打开新世界的钥匙，为这些领域带来了无限的可能性。

例如，在虚拟现实和游戏设计中，NeRF 技术可以帮助开发者快速构建出逼真的虚拟环境，极大地提升用户的沉浸感。在增强现实领域，NeRF 技术可以用于精确的物体追踪和场景理解，让 AR 应用的真实感和互动性得到质的飞跃。在电影制作中，NeRF 技术则能够帮助制作人员创造出高度逼真的特效和场景，让电影的视觉冲击力达到一个新的高度。

尽管 NeRF 技术为三维重建领域带来了巨大的变革，但它也面临着一些挑战。在动态场景或复杂光照条件下的重建效果可能不尽如人意，这就像是在探索未知领域时遇到的迷雾，需要我们去揭开。此外，NeRF 对训练数据和计算资源的需求较高，这在实时应用中可能会成为性能的瓶颈。但这些挑战，正是我们前进的动力，未来的研究需要针对这些问题进行深入的探讨和改进。

通过深入探索 NeRF，我深刻理解了计算机视觉和机器学习技术在三维场景理解中的关键作用。NeRF 作为一种新兴技术，其快速发展和巨大的应用潜力令人充满期待。在未来的研究中，我希望能够进一步探索 NeRF 在不同领域的应用，并为提高其实用性和普及度做出自己的贡献。我相信，随着技术的不断进步，NeRF 将为我们打开一扇扇通往新世界的大门，让我们能够以一种全新的方式，去探索、去创造、去体验这个多姿多彩的世界。

5 参考文献

- [1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” in European conference on computer vision. Springer, 2020, pp. 405–421.
- [2] J. T. Kajiya and B. P. Von Herzen, “Ray tracing volume densities,” ACM SIGGRAPH computer graphics, vol. 18, no. 3, pp. 165–174, 1984.
- [3] M. Niemeyer, J. T. Barron, B. Mildenhall, M. S. Sajjadi, A. Geiger, and N. Radwan, “Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5480–5490.

- [4] K. Deng, A. Liu, J.-Y. Zhu, and D. Ramanan, “Depth-supervised nerf: Fewer views and faster training for free,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 12 882–12 891.
- [5] Y.-C. Guo, D. Kang, L. Bao, Y. He, and S.-H. Zhang, “Nerfren: Neural radiance fields with reflections,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 18 409–18 418.
- [6] D. Xu, Y. Jiang, P. Wang, Z. Fan, H. Shi, and Z. Wang, “Sinnerf: Training neural radiance fields on complex scenes from a single image,”2022
- [7] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng, “Fourier features let networks learn high frequency functions in low dimensional domains,” Advances in Neural Information Processing Systems, vol. 33, pp. 7537–7547, 2020.
- [8] Liu L, Gu J, Zaw Lin K, et al. Neural sparse voxel fields[J]. Advances in Neural Information Processing Systems, 2020, 33: 15651-15663.
- [9] Chen Z, Funkhouser T, Hedman P, et al. Mobilenerf: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 16569-16578.
- [10] Reiser C, Peng S, Liao Y, et al. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 14335-14345.
- [11] Sun C, Sun M, Chen H T. Improved direct voxel grid optimization for radiance fields reconstruction[J]. arxiv preprint arxiv:2206.05085, 2022.
- [12] Li K, Rolff T, Schmidt S, et al. Bringing instant neural graphics primitives to immersive virtual reality[C]//2023 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW). IEEE, 2023: 739-740.
- [13] Bhattarai A R, Nießner M, Sevastopolsky A. Triplanenet: An encoder for eg3d inversion[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024: 3055-3065.
- [14] Chen A, Xu Z, Geiger A, et al. Tensorf: Tensorial radiance fields[C]//European conference

on computer vision. Cham: Springer Nature Switzerland, 2022: 333-350.

[15] Reiser C, Szeliski R, Verbin D, et al. Merf: Memory-efficient radiance fields for real-time view synthesis in unbounded scenes[J]. ACM Transactions on Graphics (TOG), 2023, 42(4): 1-12.

[16] Shao R, Zheng Z, Tu H, et al. Tensor4d: Efficient neural 4d decomposition for high-fidelity dynamic reconstruction and rendering[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 16632-16642.

[17] Işık M, Rünz M, Georgopoulos M, et al. Humanrf: High-fidelity neural radiance fields for humans in motion[J]. ACM Transactions on Graphics (TOG), 2023, 42(4): 1-12.

[18] Cao A, Johnson J. Hexplane: A fast representation for dynamic scenes[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 130-141.

[19] Fridovich-Keil S, Meanti G, Warburg F R, et al. K-planes: Explicit radiance fields in space, time, and appearance[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 12479-12488.

[20] Zheng Z, Zhao X, Zhang H, et al. Avatarrex: Real-time expressive full-body avatars[J]. ACM Transactions on Graphics (TOG), 2023, 42(4): 1-19.

[21] Li Z, Zheng Z, Liu Y, et al. Posevocab: Learning joint-structured pose embeddings for human avatar modeling[C]//ACM SIGGRAPH 2023 Conference Proceedings. 2023: 1-11.

[22] Jiang Y, Yao K, Su Z, et al. Instant-NVR: Instant Neural Volumetric Rendering for Human-object Interactions from Monocular RGBD Stream[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 595-605.

[23] Liu J W, Cao Y P, Yang T, et al. Hosnerf: Dynamic human-object-scene neural radiance fields from a single video[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 18483-18494.

[24] Hong F, Zhang M, Pan L, et al. Avatareclip: Zero-shot text-driven generation and animation of 3d avatars[J]. arxiv preprint arxiv:2205.08535, 2022.

[25] Hong F, Chen Z, Lan Y, et al. Eva3d: Compositional 3d human generation from 2d image collections[J]. arxiv preprint arxiv:2210.04888, 2022.

[26] Cao Y, Cao Y P, Han K, et al. Dreamavatar: Text-and-shape guided 3d human avatar

generation via diffusion models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 958-968.

[27] Chan E R, Monteiro M, Kellnhofer P, et al. pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 5799-5809.

[28] Bhattarai A R, Nießner M, Sevastopolsky A. Triplanenet: An encoder for eg3d inversion[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024: 3055-3065.

[29] An S, Xu H, Shi Y, et al. Panohead: Geometry-aware 3d full-head synthesis in 360deg[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 20950-20959.

[30] **ang J, Yang J, Deng Y, et al. Gram-hd: 3d-consistent image generation at high resolution with generative radiance manifolds[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 2195-2205.

[31] Chen X, Deng Y, Wang B. Mimic3d: Thriving 3d-aware gans via 3d-to-2d imitation[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE Computer Society, 2023: 2338-2348.

[32] Poole B, Jain A, Barron J T, et al. Dreamfusion: Text-to-3d using 2d diffusion[J]. arxiv preprint arxiv:2209.14988, 2022.

[33] Lin C H, Gao J, Tang L, et al. Magic3d: High-resolution text-to-3d content creation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 300-309.

[34] Chen R, Chen Y, Jiao N, et al. Fantasia3d: Disentangling geometry and appearance for high-quality text-to-3d content creation[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2023: 22246-22256.

[35] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance ffields for view synthesis. In European conference on computer vision, pages 405–421. Springer, 2020. 1, 2, 4

[36] Yuan Y J, Sun Y T, Lai Y K, et al. Nerf-editing: geometry editing of neural radiance

- fields[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 18353-18364.
- [37] Yang B, Bao C, Zeng J, et al. Neumesh: Learning disentangled neural mesh-based implicit field for geometry and texture editing[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 597-614.
- [38] Zhang K, Kolkin N, Bi S, et al. Arf: Artistic radiance fields[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 717-733.
- [39] Jingxiang Sun, Xuan Wang, Yichun Shi, Lizhen Wang, Jue Wang, and Yebin Liu. Ide-3d: Interactive disentangled editing for high-resolution 3d-aware portrait synthesis. ACM Transactions on Graphics (TOG), 41(6):1–10, 2022. 1, 4
- [40] Jiang K, Chen S Y, Liu F L, et al. Nerffaceediting: Disentangled face editing in neural radiance fields[C]//SIGGRAPH Asia 2022 Conference Papers. 2022: 1-9.
- [41] Wu Y, Deng Y, Yang J, et al. Anifacegan: Animatable 3d-aware face image generation for video avatars[J]. Advances in Neural Information Processing Systems, 2022, 35: 36188-36201.
- [42] Sun J, Wang X, Wang L, et al. Next3d: Generative neural texture rasterization for 3d-aware head avatars[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 20991-21002.
- [43] A. Haque, M. Tancik, A. A. Efros, A. Holynski, and A. Kanazawa, “Instruct-nerf2nerf: Editing 3d scenes with instructions,” 2023.
- [44] Kamata H, Sakuma Y, Hayakawa A, et al. Instruct 3d-to-3d: Text instruction guided 3d-to-3d conversion[J]. arxiv preprint arxiv:2303.15780, 2023.
- [45] Zhao X, Wang L, Sun J, et al. Havatar: High-fidelity head avatar via facial model conditioned neural radiance field[J]. ACM Transactions on Graphics, 2023, 43(1): 1-16.
- [46] Kirschstein T, Qian S, Giebenhain S, et al. Nersemble: Multi-view radiance field reconstruction of human heads[J]. ACM Transactions on Graphics (TOG), 2023, 42(4): 1-14.
- [47] Zhang K, Riegler G, Snavely N, et al. Nerf++: Analyzing and improving neural radiance fields[J]. arxiv preprint arxiv:2010.07492, 2020.
- [48] Barron J T, Mildenhall B, Tancik M, et al. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields[C]//Proceedings of the IEEE/CVF international conference on

computer vision. 2021: 5855-5864.

[49] Barron J T, Mildenhall B, Verbin D, et al. Mip-nerf 360: Unbounded anti-aliased neural radiance fields[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 5470-5479.

[50] Verbin D, Hedman P, Mildenhall B, et al. Ref-nerf: Structured view-dependent appearance for neural radiance fields[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 5481-5490.

[51] Liu X, Xue H, Luo K, et al. GenN2N: Generative NeRF2NeRF Translation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 5105-5114.

本科生课程设计成绩评定表

班级：人工智能 2201 姓名：代文博 学号：0122210880110

序号	评分项目	满分	实得分
1	学习态度认真、遵守纪律	20	
2	报告分析合理性	30	
3	报告观点正确性，前沿技术分析全面	30	
4	设计报告的规范性	20	
		总得分/ 等级	
<p>评语：</p>			

注：最终成绩以五级分制记。优（90-100分）、良（80-89分）、中（70-79分）、及格（60-69分）、60分以下为不及格

指导教师签名：

2024 年 7 月 日