

3D无限场景的生成

第一周

图形学最基础知识，包括相机标定，colmap，sfm等方法

[三维重建基础](#)

论文解读

LucidDreamer

In this work, we propose a pipeline called LucidDreamer that utilizes Stable Diffusion and 3D Gaussian splatting to create diverse high-quality 3D scenes from various types of inputs such as text, RGB, and RGBD.

1. **局限**：dataset far from real world（特定领域）
 - a. 利用stable diffusion生成3D模型，难以保证multi-view consistency
 - b. diffusion模型学习：
[diffusion模型基础](#)
2. **Step: dream and alignment**
 - a. 投影点云，利用生成模型。根据深度提升三维空间
 - b. Alignment algorithm: integrate
 - c. 作为3dgs的inital points

简单来说，初始图像投影之后用diffusion模型生成，重新升维到3维空间后，用alignment算法和原先的点云群融合

3. **输入**：初始图像和深度图
4. **输出**：点云
5. **优点**：多场景，配合文字输入，生成更多视角
6. **具体过程**：
 - a. 初始化点云，利用stable diffusion得到img和深度图（深度图的生成根据zero-depth the monocular depth estimation model），把二维图像升维到三维
 - i. $P_0 = \phi_{2 \rightarrow 3}([I_0, D_0], K, P_0)$
 - b. 聚合点云：把生成的点云和原来的点云聚合成更大的点云，需要满足一致性

c. dream过程：投影的image会有没有得到的点，因此利用mask来表示填充和没有填充的点，在没有填充的点继续进行stable diffusion和monocular depth的过程

i. 实际过程中，深度图可能会不一致（因为模型缺陷），所以需要optimal depth

$I_i = \mathcal{S}(\hat{I}, M_i)$, $\hat{D}_i = \mathcal{D}(I_i)$, $D_i = d_i \hat{D}_i$ 新的图像I，是从旧图像和mask利用S来生成的，depth图也可以用来生成，di是系数来控制

$$d_i = \arg \min_d \left(\sum_{M_i=1} \left\| \phi_{2 \rightarrow 3}([I_i, d\hat{D}_i], K, P_i) - P_{i-1} \right\|_1 \right)$$
 这个是用来逼近di

d. alignment：dream的过程是同时生成深度图和新图像，用off the shelf 的深度图生成方法会更好（适配更多的场景，更加的精确）

i. 需要解决consistency的问题，引入alignment算法。因为没有考虑多个depth之间的关系，直接移动点，让之间smooth。（需要计算vector），naive的移动会导致distort，需要利用差值算法解决

图形学基础（有差值算法）

7. 疑问：

a. 相机的内参和外参如何得到：The camera intrinsic matrix and the extrinsic matrix of I0 are denoted as K and P0, respectively. For the case where I0 and D0 are generated from the diffusion model, we set the values of K and P0 by convention regarding the size of the image.

b. 高感知质量是什么：high perceptual quality