利用SVD生成单目视频，然后在SV3D的unet中加入了一个无需训练的自注意力模块，用于让后面一帧多视图图像能够参考前面那一帧的多视图图像，然后用sv3d生成每一帧的多视图图像。

在动态重建的时候，除了原来的变形网络，还会用一个颜色网络改变每个高斯单元的颜色（这个是比较合理的，像点光源下的场景这种确实是因为高斯单元因为光照导致反射的颜色变了，而不是因为原本亮的高斯单元不移动导致的）

然后将渲染出来的图像喂给SDXL-Turbo进行去噪，将原先渲染的和去噪后的进行对比，计算损失，然后去更新mlp