

第二周周报(02.27-03.04)

上周任务总结

1. 阅读 StyleSDF和 NeRF 论文
2. 略读 ENFAF-GAN 和 EG3D
3. NeRF 的代码仍在看, 预计明天可以看完
4. SMPL 阅读了一部分, 理解起来比神经渲染部分的要吃力些

下周任务

1. 下周继续阅读 SMPL 和 EVA3D,特别是 EVA3D 的代码实现, 这两篇可能会放到下下周做汇报, 同时介绍ENFAF-GAN(这个也引入了关于人体姿势的先验)
2. 阅读 StyleSDF 的代码, 如果有需要,会先汇报 StyleSDF 论文, 同时介绍 EG3D;在阅读的过程中感觉这两篇论文有很多相似,值得会比较的地方,可以做一个对比分析;
3. GRAF, NARF, GIRAFFE, 三篇论文做略读, 提取主要内容.
4. 上周老师布置的用 pytorch 编写 CNN 的 tutorial
5. 整理上周论文阅读的简要报告, 在下周周会前会提交

上周论文阅读摘要

1. StyleSDF

StyleSDF它要解决的是在单视角的图像上合成高分辨率的 3D 图像生成问题, 此前针对问题的研究主要是基于3D感知的 GAN方法, 但是这种方法面临两个挑战: 1. 无法生成高分辨率,视角一致的 3D 图像; 2. 3D 表面生成的细节处理不好. 作者认为主要原因是这类基于感知的 3DGAN 方法,要么需要获取成本高昂的多视角图像做训练集, 受限于此, 往往难以渲染高分辨率的图像(感觉这个和原始的NeRF 其实有一些相似,受限于多视角的图像做监督);另一方面, 使用 opacity fields 作为表面重构的代理任务,导致表面的细节处理不好,往往产生从不同视角看来不一致的图像.

作者的方法的主要思想是利用基于 SDF 的 3D 体积渲染和基于 2D 的 styleGAN,是一个两阶段的模型: 第一阶段训练一个SDF的体积渲染器, 以随机噪声做输入,经过渲染映射网络的输出 \mathbf{w} 以及采样点坐标 \mathbf{x} ,采样方向 \mathbf{v} 作为输入, 计算SDF值 $d(\mathbf{x}, \mathbf{z})$, 视角相关的颜色值 $c(\mathbf{x}, \mathbf{v}, \mathbf{z})$ 以及特征向量 $f(\mathbf{x}, \mathbf{v}, \mathbf{z})$,进一步计算3D密度 σ , 经过体积聚合和Marching Cubes输出低分辨率的RGB图片, 2D 特征图和 3Dmesh; 第二阶段将第一阶段的2D特征图和经渲染映射网络得到的 \mathbf{w} 作为生成映射网络输入得到的输出, 一并作为 2DstyleGAN的输入,得到高分辨率的图像.体素渲染有助于学习底层的3D表面信息,将图像的姿势和外观分开,在做推理的时候可以控制相机的位置;StyleGAN2 生成器对低分辨率特征图进行上采样,添加高频细节,并模拟复杂的光传输效果,例如难以使用低分辨率体积渲染器建模的次表面散射和相互反射.

训练体积渲染使用的loss主要是带L1正则的非饱和GAN loss, 除此之外还额外使用了三个loss用于正则: Pose Alignment Loss 用于保证生成的物体是全局对齐的; Eikonal Loss 用于保证学到的SDF在物理上是有效的; Minimal Surface 鼓励 3D 网络用最小量的零交叉来描述场景; 在训练第二阶段 StyleGAN时需要将第一阶段的体积渲染模型冻住,按照styleGAN的设置进行训练.

作者的实验主要是在FFHQ和AFHQ数据集上进行的, 对比的Baseline主要是GIRAFFE, Pi-GAN, GRAF, HoloGAN.定性的实验包括2D新视角图像生成和3D表面比较, StyleSDF生成的新视角图像保持了3D一致性, 并且3D表面质量远高于baseline; 定量比较主要比较了FID和KID分数; 为了比较生成的图像的3D一致性, 和PiGAN比较了depth指标,结果远胜于piGAN,大幅提升了深度一致性.

作者也承认, 在像牙齿这类地方, styleSDF的表现不好(EG3D也存在这个问题, StyleSDF和EG3D在arxiv的发布的时间相近,而且都是做人脸3D的, 不知道这两个一起比较谁的效果更好一些.), 有轻微的混叠;镜面反射或者强光照射引起的深度凹痕,由于styleSDF是在单视角图像上训练的很难消除歧义; 没有处理前景和背景;下一步研究可以从两方面开展: 将原来二阶段的训练过程变为端到端的训练;创建一个与 SDF 表面边界相关的体积采样策略(以减少每次前向传递的查询点数量)并消除对 2D CNN 对特征图进行上采样的需要, 直接将3D几何表面与高分辨率图像联系起来.

2. EG3D

未完待续

原计划周日晚上把这几篇文章的总结写一写的,写CNN那个tutorial时出现一点点小问题,就是自定义实现的CIFAR-10的Dataset在经过归一化变换时,并没有映射到 $[0,1]$ 之间,暂时没有找到问题所在,还在尝试中,所以耽误了晚上写周报和总结的时间.