Review-46

Scene Representation Networks: Continuous 3D-Structure-Aware Neural Scene Representations

Modelling scene geometry requires explicit 3D supervision in the case of modern geometric deep learning models. The work aims to address this challenge by proposing Scene Representation Networks (SRNs). SRNs make use of continuous 3D structure-aware scene representations by encoding geometry and appearance. These networks represent scenes as continuous functions which map world coordinates to feature representations with local properties. SRNs are trained end-to-end from 2D images and camera poses using a differential ray-marching algorithm. This demonstrates the efficacy of SRNs to complex geometric tasks such as novel view synthesis, few-shot reconstruction, joint shape and appearance interpolation.

The SRN structure relies on the mapping of world coordinates to local feature representations by utilizing a continuous differential function. This is achieved by formulating a scene representation using MLP which maps 3D cartesian coordinates to features at final coordinates. Scene representations obtained from the MLP undergo a nueral rendering algorithm which maps the representation to an image $\mathcal{I}$ based on the camera parameters. The neural rendering algorithm first finds the coordinates of intersections of respective camera rays. Following the determination of world locations, the rendering process maps feature vectors to spatial coordinates of color. The first step is achieved using a novel differentiable ray-marching algorithm. An LSTM maps feature vectors denoting the current estimate of ray intersections to the length of the next marching step. This is called steplength prediction. The second step of the process consists of a Pixel Generator Architecture which employs a per-pixel MLP mapping a feature vector to RGB vector. This is akin to 1x1 convolutions and preserved the global world coordinates of objects in the scene.