Scene Representation Networks:
Continuous 3D-Structure-Aware Neural Scene Representations
Vincent Sitzmann, Michael Zollhöfer, Gordon Wetzstein

Abstract
Unsupervised learning with generative models holds the promise to learn rich representations of 3D scenes. Existing neural scene representations don’t exploit 3D structure. As a result, they’re sample-inefficient, opaque, and don’t generalize to unseen viewpoint transformations.

Scene Representation Networks (SRNs) are a continuous neural scene representation with a 3D inductive bias. Along with a neural renderer, they model both 3D scene geometry and appearance, enforce 3D structure in a multi-view consistent manner, and naturally generalize shape and appearance across scenes.

Problem definition
Data: Tuples of image, camera pose & intrinsics

Train only on data that could be collected by walking around with a camera

Vision: Learn rich representations of 3D scenes by watching videos!

Self-supervised Scene Representation Learning

By using neural renderer, can supervise scene representation with posed images only!

Results
Appearance & geometry from 50 images
Few-shot reconstruction
Latent space interpolation
Non-rigid deformation

Generalize across class of objects with Hypernetworks

Supervise only with posed 2D images via Neural Rendering

Ray marching to estimate ray-geometry intersections

Rendering

World coordinates after n steps of ray marching

Pixel Generator

Feature representation of scene
Persistent across views

Neural Render - Render from different camera perspectives

Neural Scene Representation

By using neural renderer, can supervise scene representation with posed images only!