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

Vincent Sitzmann, Michael Zollhöfer, Gordon Wetzstein

Stanford University

SCI STANFORD COMPUTATIONAL IMAGING

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.

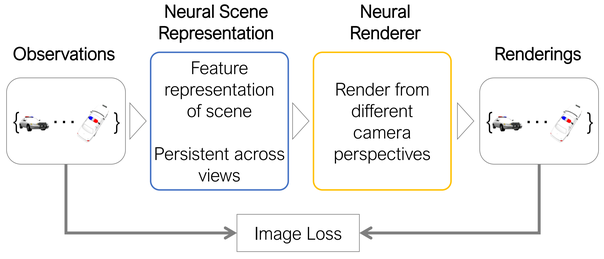
Problem definition

Data: Tuples of image, camera pose & intrinsics

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

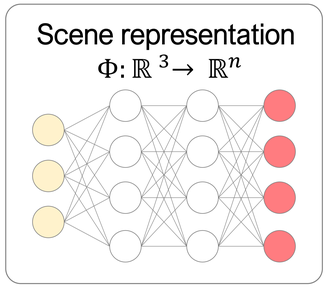
Vision: Learn rich representations of 3D scenes by watching videos!

Self-supervised Scene Representation Learning



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

Scenes as functions that map coordinates to features



Encode scene in weights of Multi-Layer Perceptron (MLP)

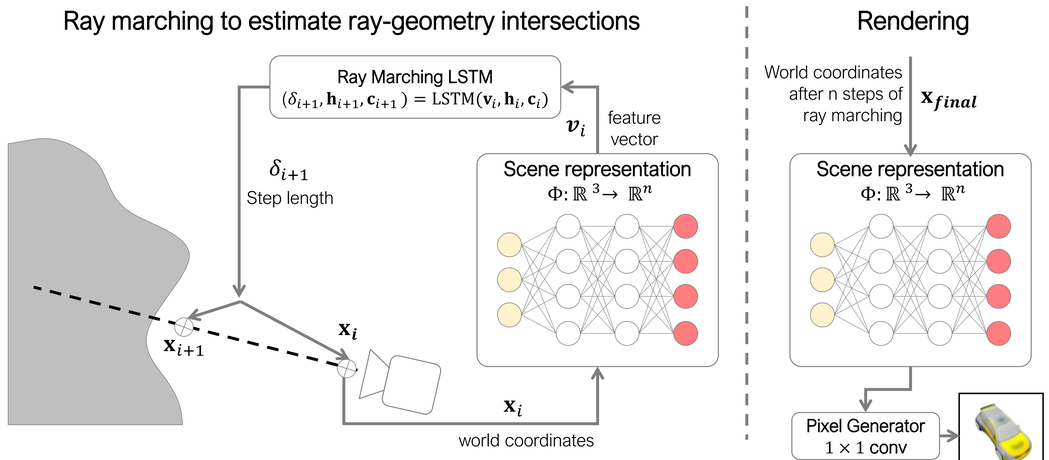
MLP maps every (x,y,z) world coordinate to a feature

Features may represent color, material, signed distance, but also semantic information

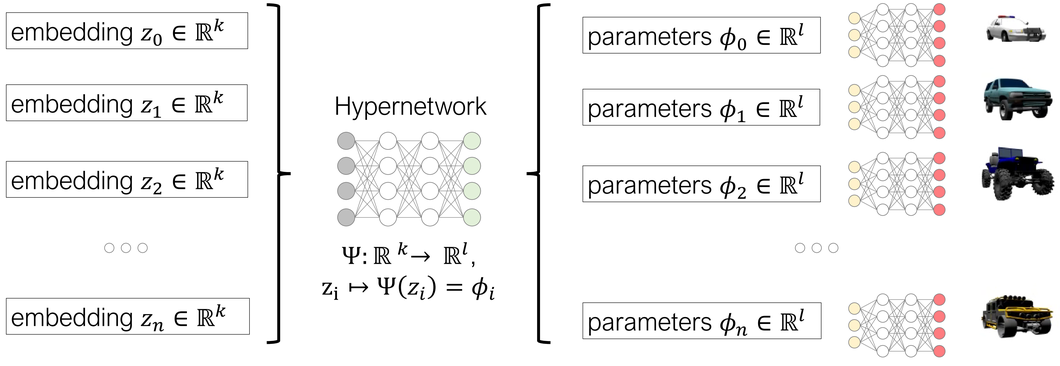
Parameterizes surfaces smoothly

Doesn't scale with resolution, but with scene complexity

Supervise only with posed 2D images via Neural Rendering

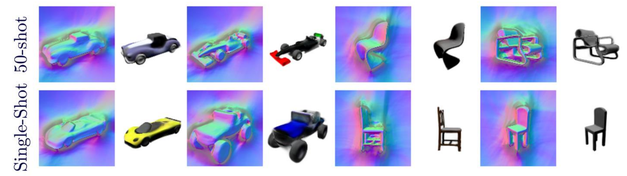


Generalize across class of objects with Hypernetworks



Results

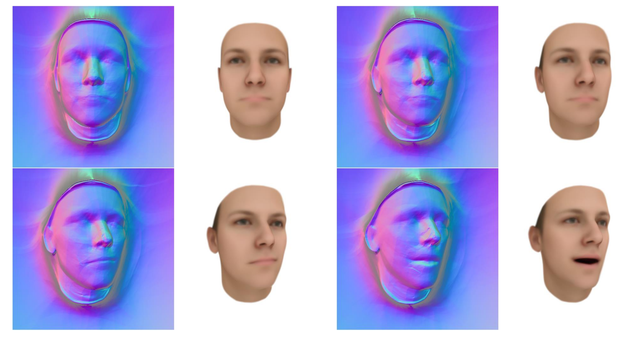
Appearance & geometry from 50 images



Few-shot reconstruction

Latent space interpolation

Non-rigid deformation



Project Page & Contact

vsitzmann.github.io
sitzmann@cs.stanford.edu
@vincesitzmann