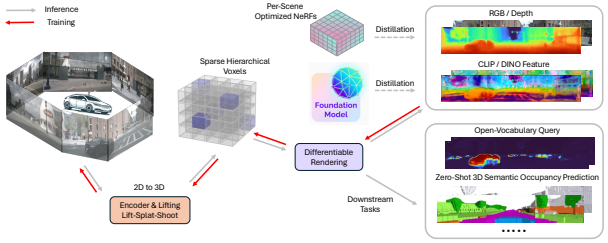
DistillNeRF: Perceiving 3D Scenes from Single-Glance Images by Distilling Neural Fields and Foundation Model Features

Letian Wang, Seung Wook Kim, Jiawei Yang, Cunjun Yu, Boris Ivanovic, Steven Waslander, Yue Wang, Sanja Fidler, Marco Pavone, Peter Karkus

Demo & Code

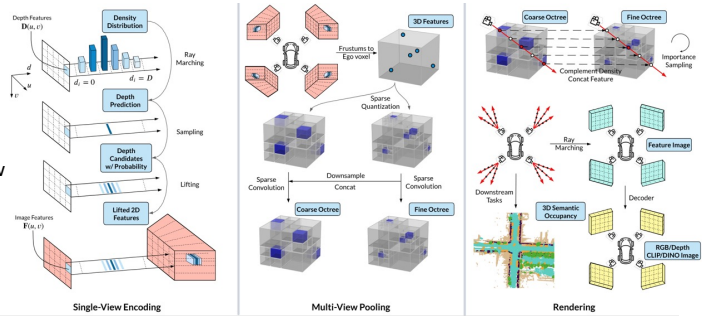
Motivation



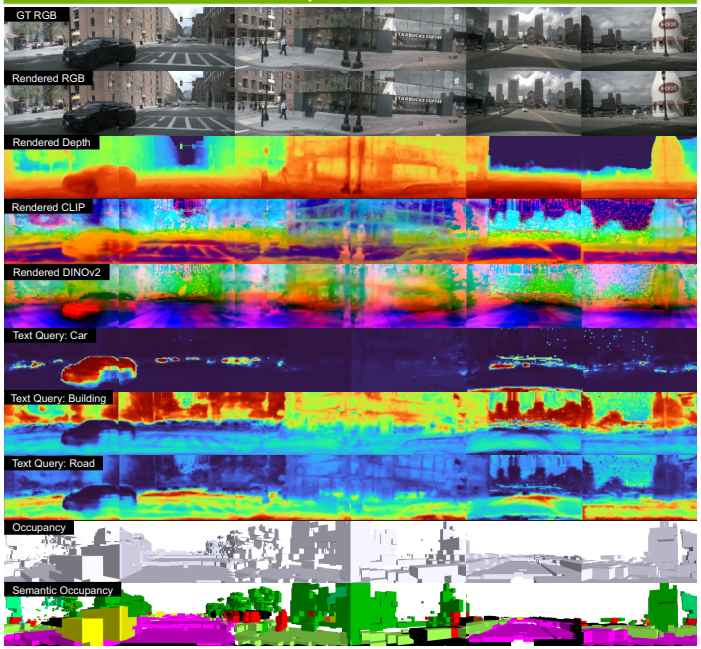
- **Problem:** Perceiving 3D scenes from 2D observations
  - Classic Perception task: un-scalable due to expensive annotation
  - NeRF: not online due to reliance on test-time per-scene training
- **Objective:** bring NeRF to be online and generalizable to new scenes, and enable downstream tasks
- **Insights:** Distillation into an online model
  - Enhancing Geometry: Distill per-scene optimized NeRFs
  - Enriching Semantics: Distill foundation model features

Architecture Details

- **Sparse Hierarchical Voxel**
  - Encoder and lift single-view image via two-stage Lift-Splat-Shoot
  - Fuse multi-view features via sparse quantization and convolution
  - Render from hierarchical sparse voxel and enable downstream tasks
- **Parameterized Neural Field**
  - Downstream: Keep inner voxels at the real scale and high resolution
  - Rendering: Outer voxels captures infinite distance at a lower resolution

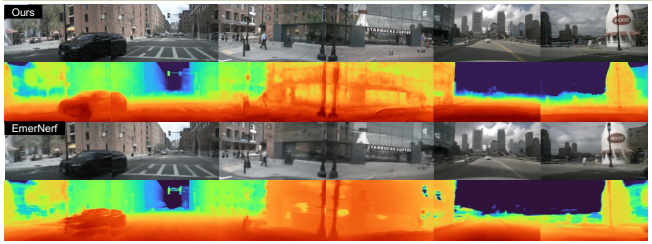


Capabilities

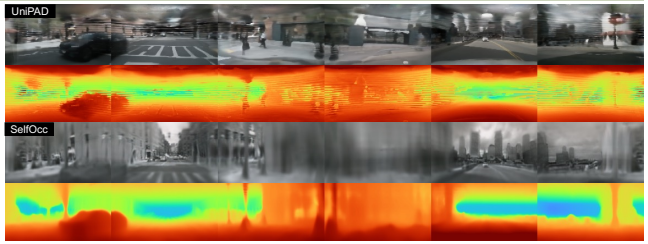


- **Rendering without test-time per-scene optimization**
  - Reconstruction & novel-view synthesis: RGB, Depth, Foundation Feat
- **Downstream Tasks without annotation**
  - Open-vocabulary query
  - 3D semantic occupancy prediction
  - .....

Results & Experiments



On Par with SOTA offline per-scene optimized NeRF



Significantly outperform SOTA online generalizable NeRFs