FiT3D

Improving 2D Feature Representations by 3D-Aware Fine-Tuning

Yuanwen Yue Anurag Das Francis Engelmann Siyu Tang Jan Eric Lenssen

ETH zürich Google

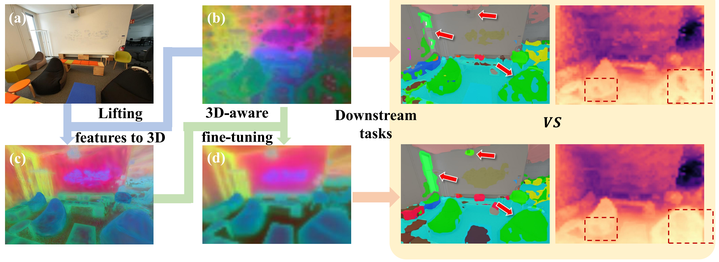
MAX PLANCK INSTITUTE FOR INFORMATICS SIC

Project page

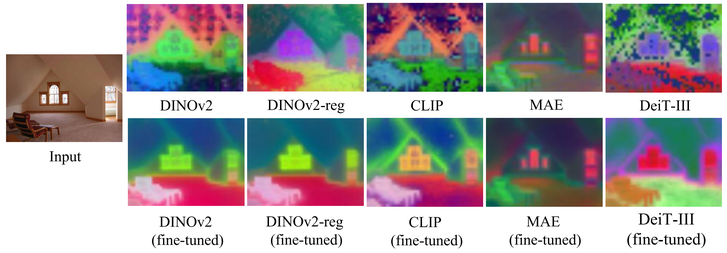
ECCV

Introduction

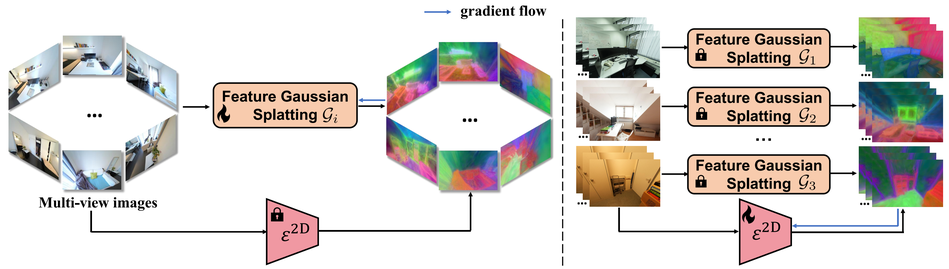
- 2D foundation models are awesome, but we live in a **3D world**.
- How to inject **3D-awareness** into 2D foundation models?
- We propose **FiT3D**: finetuning 2D foundation models with features rendered from 3D Gaussians.
- **Key insight:** 2D features lifted to 3D can in turn improve 2D models.



Universality



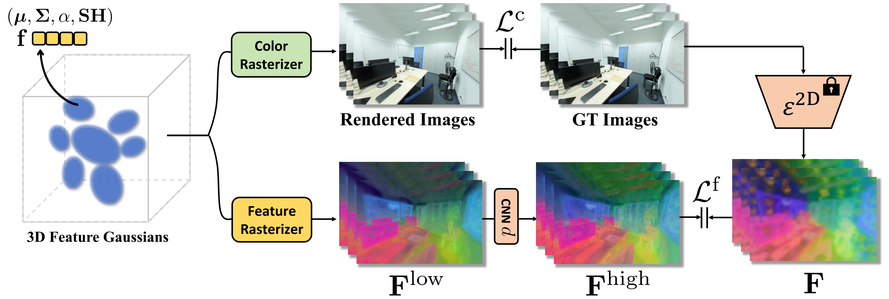
Method



We lift 2D foundation features into 3D-aware features by training 3D Gaussian representation for each scene.

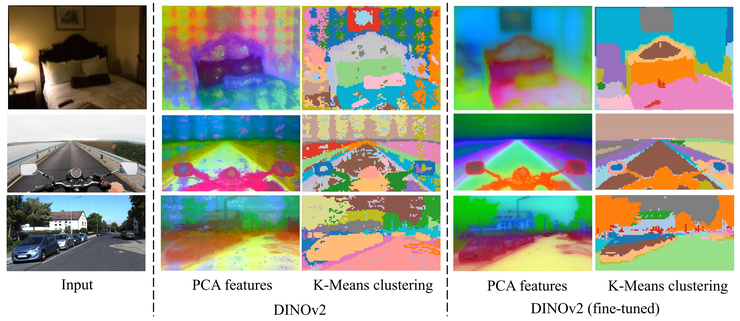
We use the rendered features from pre-trained Gaussians to finetune the 2D foundation model.

Feature Gaussian



$$\hat{\mathcal{G}} = \underset{\{(\boldsymbol{\mu},\mathbf{s},\mathbf{R},\alpha,\mathbf{SH},\mathbf{f})_i\}}{\arg\min} \sum_{i=1}^{N} \mathcal{L}^c(r^{\text{rgb}}(\mathcal{G}, \mathbf{P}_i), \mathbf{I}_i) + \mathcal{L}^f(d(r^{\text{feat}}(\mathcal{G}, \mathbf{P}_i), \mathbf{F}_i)$$

Feature Quality



Linear Probing Results