A “Splatacular” Year of 3D Reconstruction

Songyou Peng

Google DeepMind

Stanford University

May 1, 2025

Intelligent systems interact with **3D environments**

3D Reconstruction
Create digital twins from real scenes

3D Scene Understanding
Analyze the scene digitally

Video Credit: YouTube - Real time archviz apartment

My Research During PhD

Learn to **Reconstruct** and **Understand** 3D World

ConvOccNet
ECCV 2020 (Spotlight)

MonoSDF
NeurIPS 2022

Shape As Points
NeurIPS 2021 (Oral)

KiloNeRF
ICCV 2021

NICE-SLAM
CVPR 2022

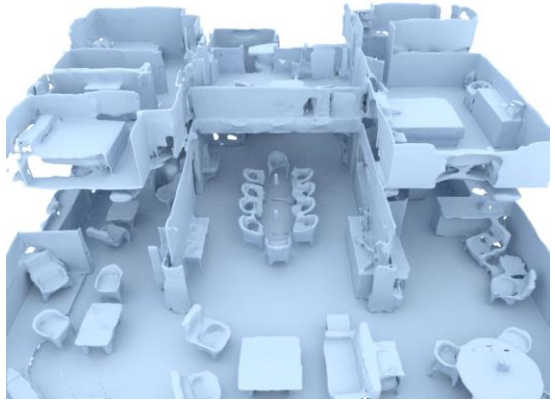
NICER-SLAM
3DV 2024 (Oral)

UNISURF
ICCV 2021 (Oral)

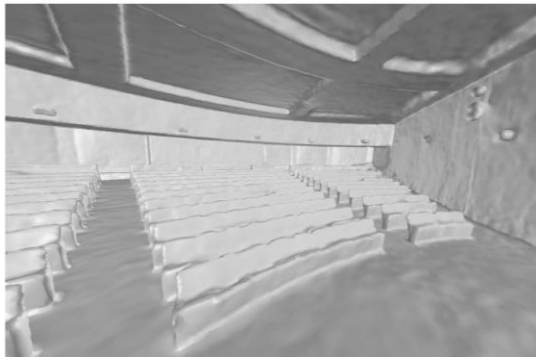
OpenScene
CVPR 2023

My Research During PhD

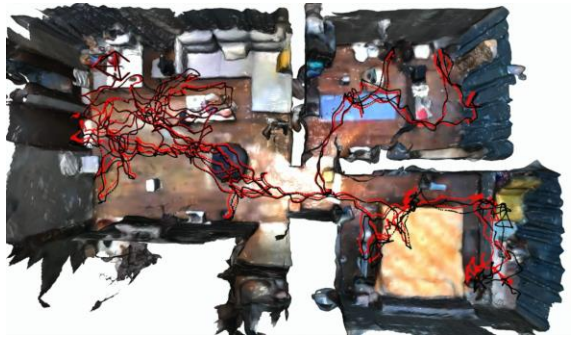
Learn to Reconstruct and Understand 3D World



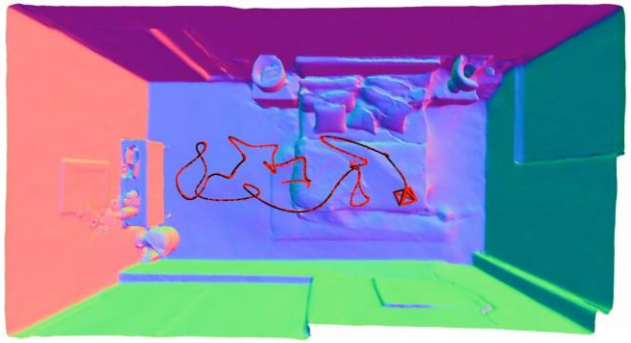
ConvOccNet
ECCV 2020 (Spotlight)



MonoSDF
NeurIPS 2022



NICE-SLAM
CVPR 2022



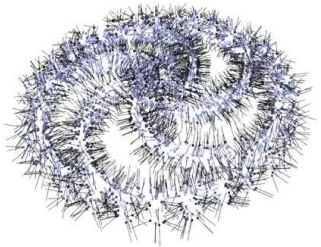
NICER-SLAM
3DV 2024 (Best Honor. Men.)

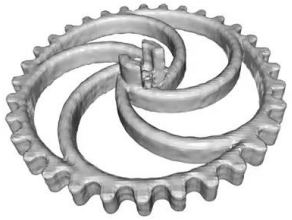
Topic #1:
Reconstruct Complex Scenes

My Research During PhD

Learn to Reconstruct and Understand 3D World

Topic #2:
Fast Inference



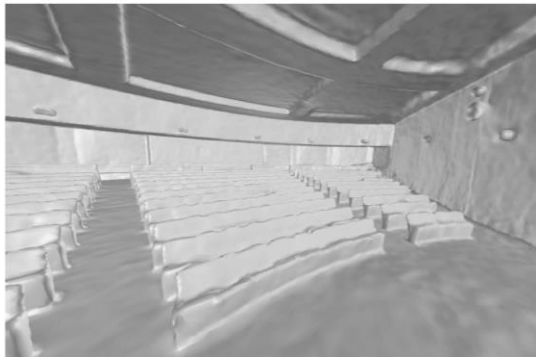


Shape As Points
NeurIPS 2021 (Oral)

KiloNeRF
ICCV 2021

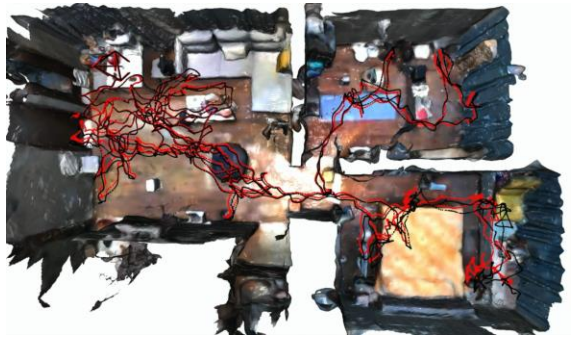
My Research During PhD

Learn to Reconstruct and Understand 3D World

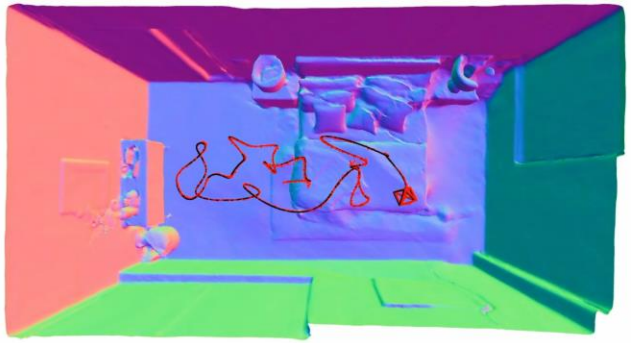


MonoSDF
NeurIPS 2022

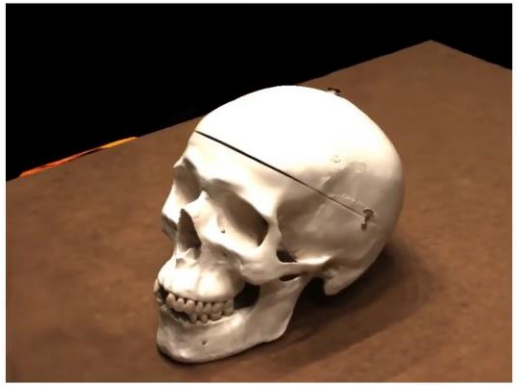
Topic #3:
Reconstruct from 2D Observations



NICE-SLAM
CVPR 2022



NICER-SLAM
3DV 2024 (Best Paper Honorable)



UNISURF
ICCV 2021 (Oral)

My Research During PhD

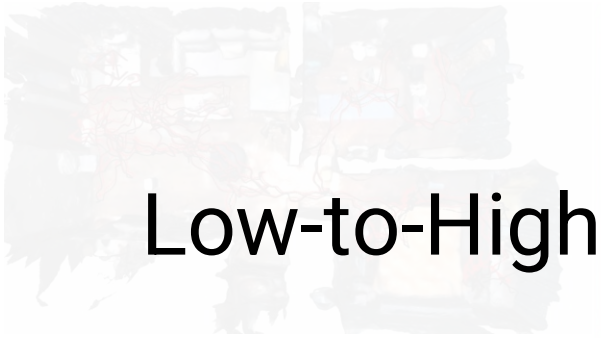
Learn to Reconstruct and Understand 3D World

ConvOccNet
ECCV 2020 (Spotlight)

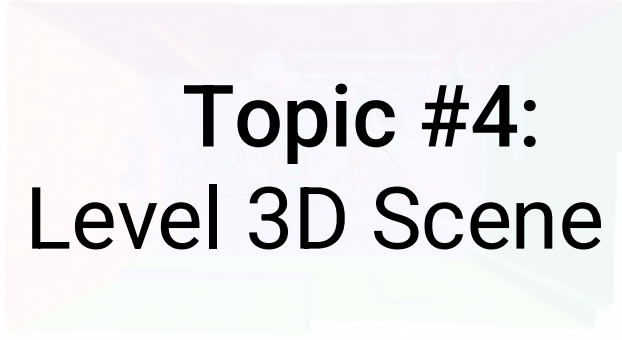
MonoSDF
NeurIPS 2022

Shape As Points
NeurIPS 2021 (Oral)

KiloNeRF
ICCV 2021



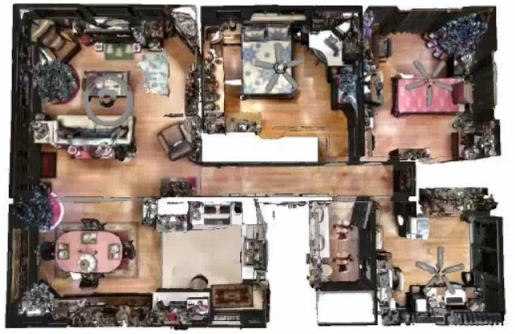
NICE-SLAM
CVPR 2022



NICER-SLAM
3DV 2024 (Oral)

UNISURF
ICCV 2021 (Oral)

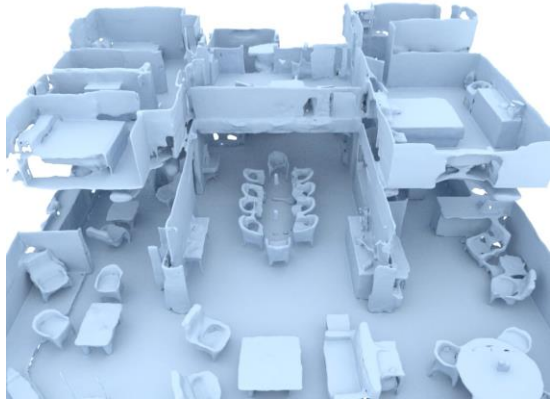
Topic #4:
Low-to-High Level 3D Scene Understanding



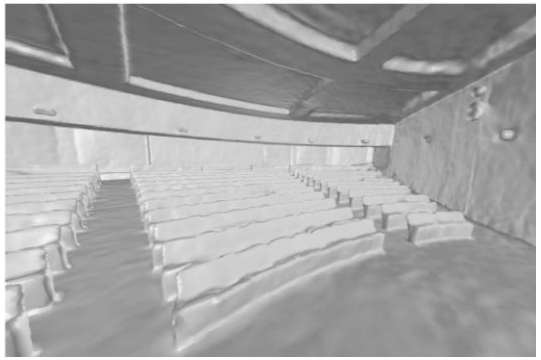
OpenScene
CVPR 2023

My Research During PhD

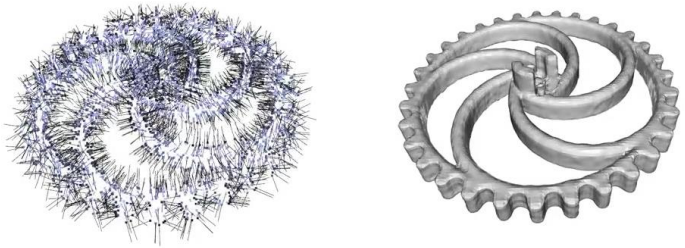
Learn to Reconstruct and Understand 3D World



ConvOccNet
ECCV 2020 (Spotlight)

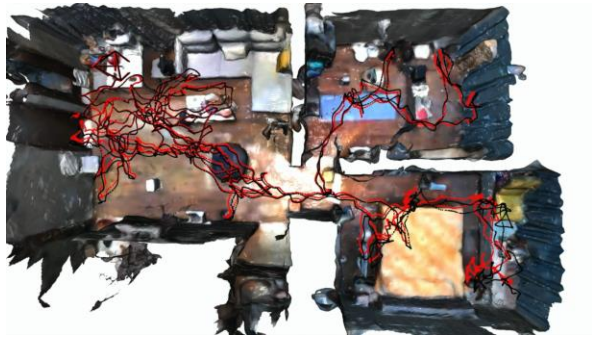


MonoSDF
NeurIPS 2022

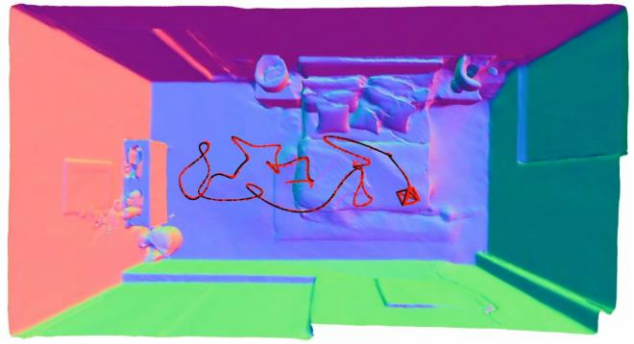


Shape As Points
NeurIPS 2021 (Oral)

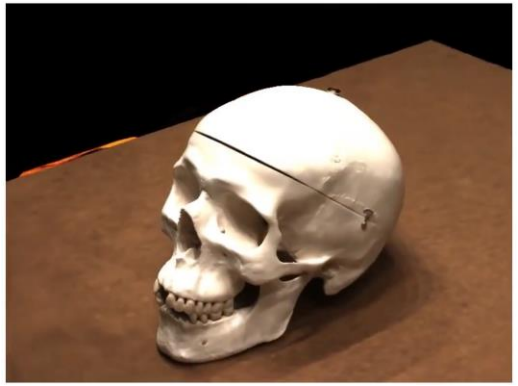
KiloNeRF
ICCV 2021



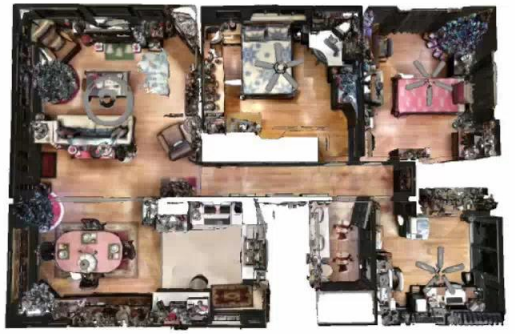
NICE-SLAM
CVPR 2022



NICER-SLAM
3DV 2024 (Best Paper Honorable)



UNISURF
ICCV 2021 (Oral)



OpenScene
CVPR 2023

My PhD Thesis

Already Tackling Some Challenges in 3D Reconstruction

- Reconstruct **at scale**
- Reconstruct **at speed**
- Reconstruct **from 2D observations**

An Ideal 3D Reconstruction Pipeline

Instant, Pose-Free, Real-World 3D Everywhere

Feedforward

Pose-Agnostic

Dynamic

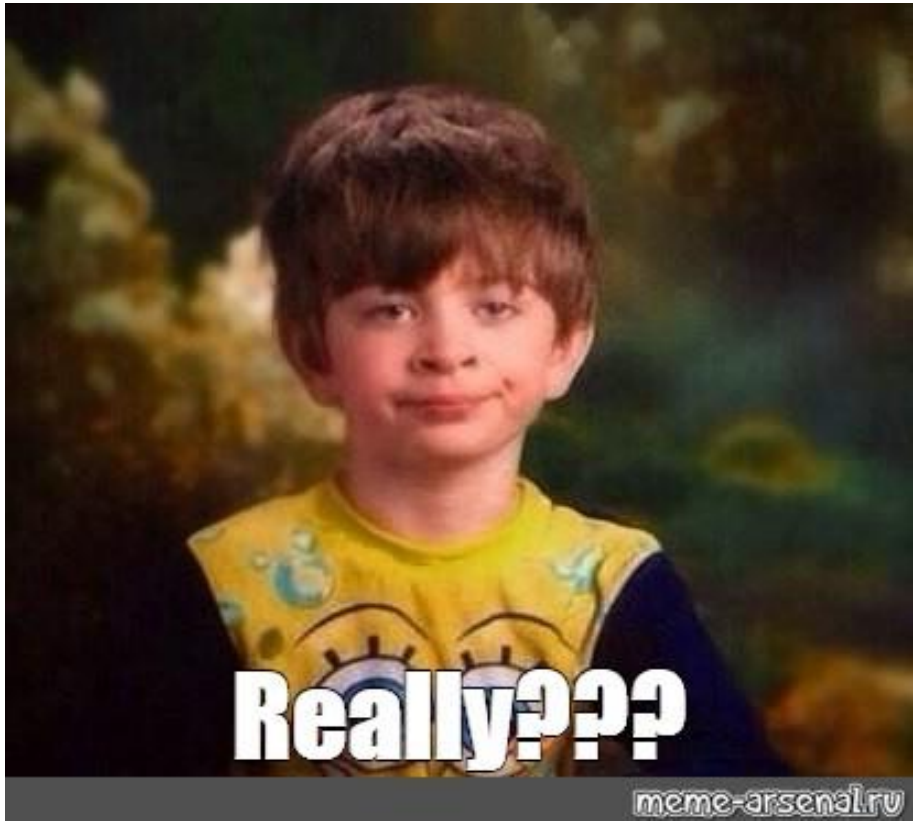
Fast Rendering

Arbitrary Lengths

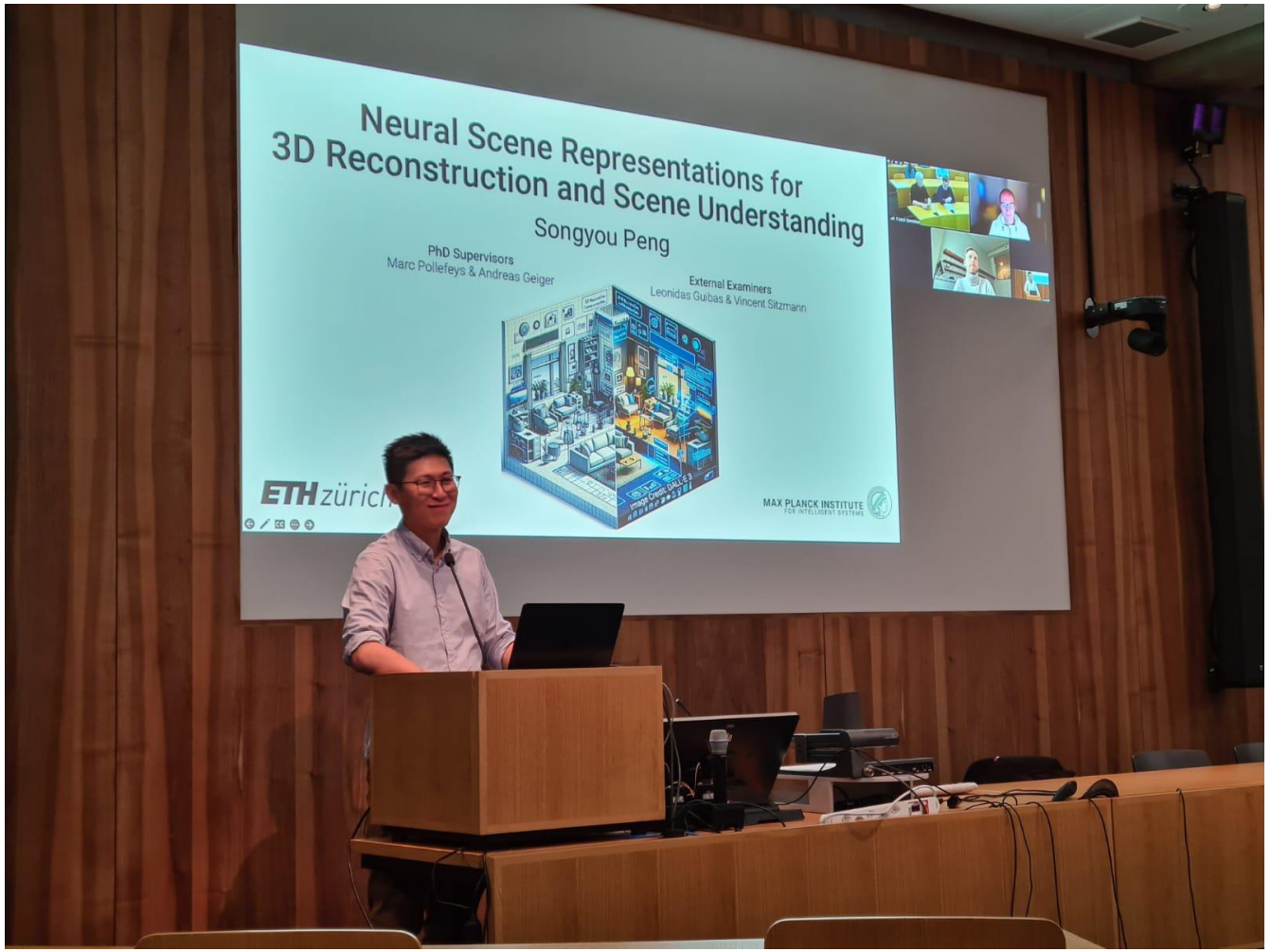
Lighting-Robust

“People overestimate what they can do in **one year**, and underestimate what they can do in **ten years**.”

--- *Bill Gates*



I Defended on Nov 2023



How much can one push forward until **Nov 2024**?

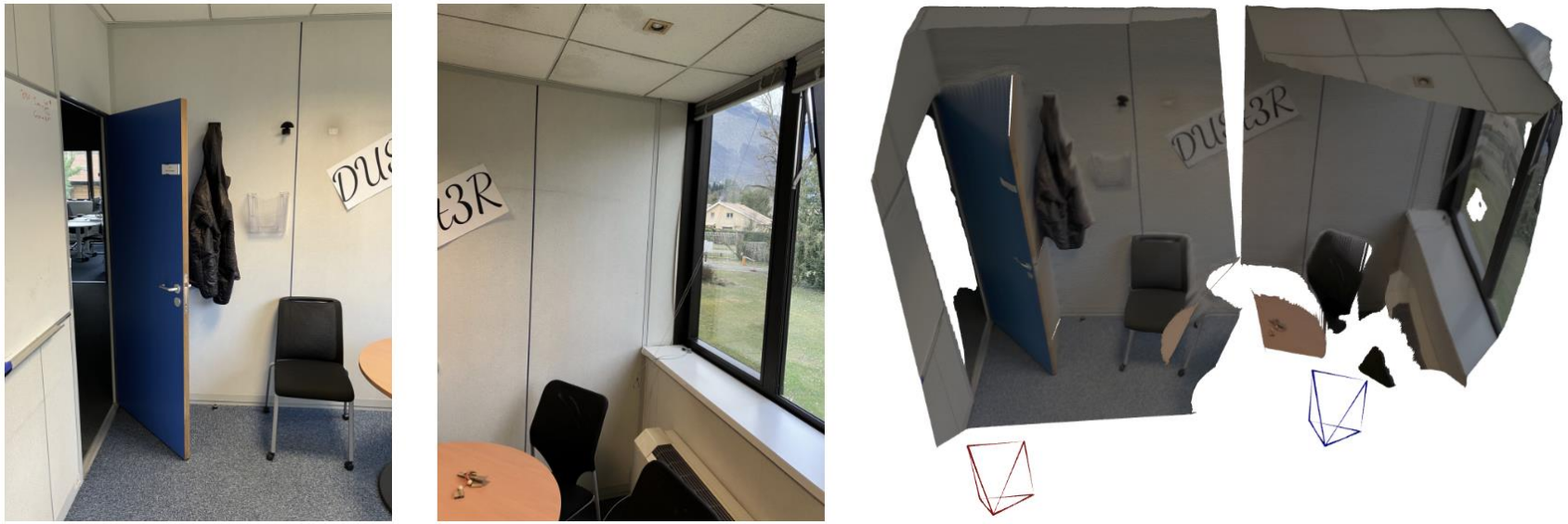
What Came Up in 2023?

Input Posed Images

3DGS

Kerbl*, Kopanas*, Leimkühler, Drettakis: 3D Gaussian Splatting for Real-Time Radiance Field Rendering. SIGGRAPH 2023

What Came Up in 2023?



Wang, Leroy, Cabon, Chidlovskii, Revaud: DUSt3R: Geometric 3D Vision Made Easy. CVPR 24 (*appeared on arXiv at Dec 2023*)

An Ideal 3D Reconstruction Pipeline

Instant, Pose-Free, Real-World 3D Everywhere

Feedforward

Pose-Agnostic

Dynamic

Fast Rendering

Arbitrary Lengths

Lighting-Robust

An Ideal 3D Reconstruction Pipeline

Instant, Pose-Free, Real-World 3D Everywhere

Feedforward

Pose-Agnostic

Dynamic

Fast Rendering

Arbitrary Lengths

Lighting-Robust

Feedforward 3D Gaussian Splatting

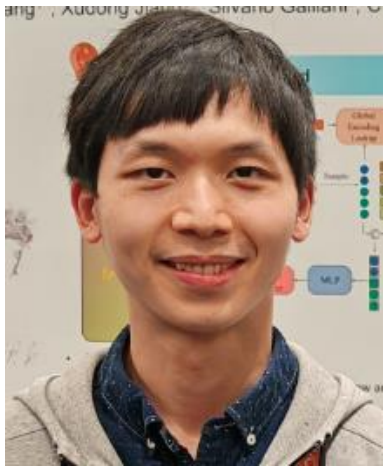
DepthSplat

Connecting Gaussian Splatting and Depth

haofeixu.github.io/depthsplat/

CVPR 2025







Haofei Xu | **Songyou Peng** | Fangjinhua Wang | Hermann Blum | Daniel Barath | Andreas Geiger | Marc Pollefeys

Motivation

pixelSplat / MVSplat

Depth Anything v2

✚ Multi-view Consistent

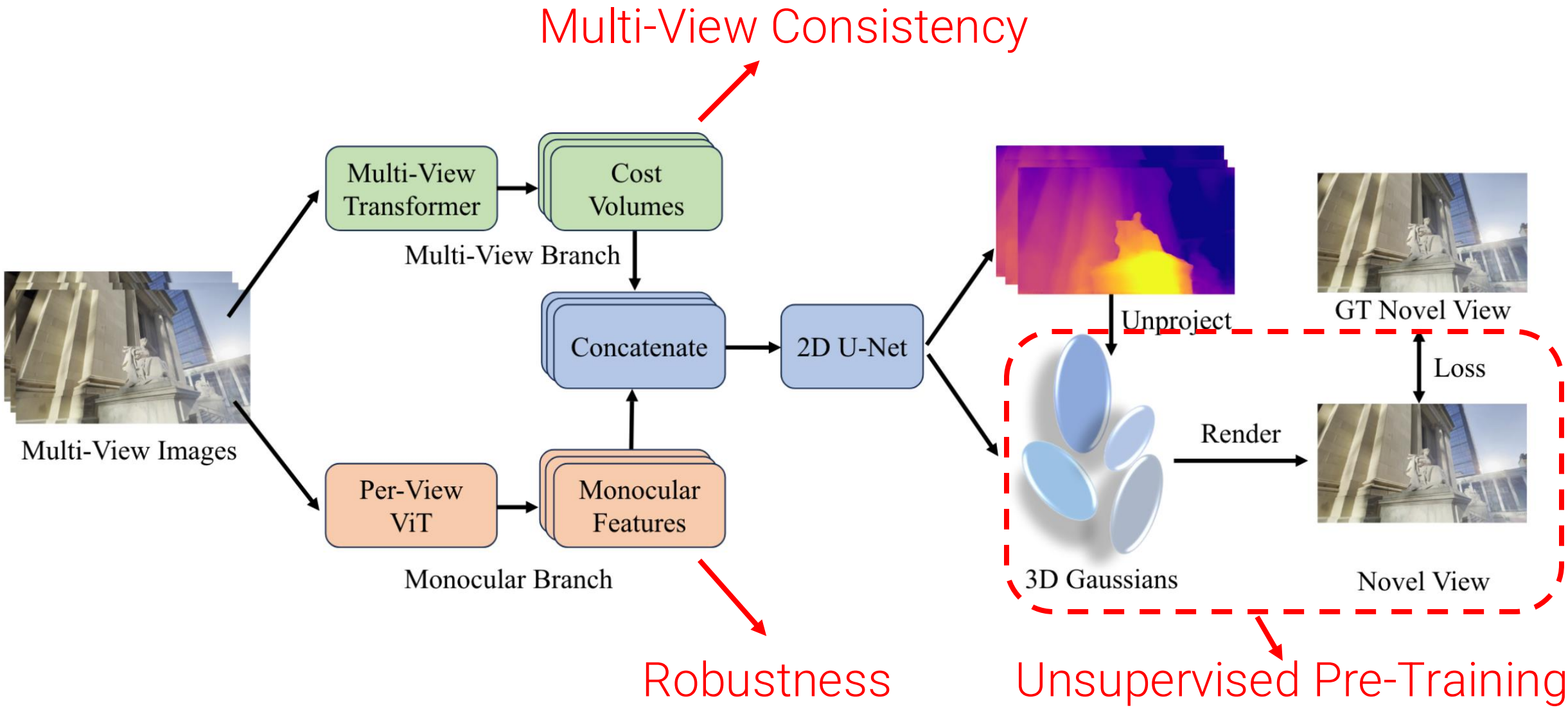
− Lack robustness

✚ Robust

− Unknown scale & shift

Can they benefit each other?

Pipeline


Multi-View Consistency
Multi-View Transformer
Cost Volumes
Multi-View Branch
Concatenate
2D U-Net
Unproject
GT Novel View
Loss
Multi-View Images
Per-View ViT
Monocular Features
Monocular Branch
Render
3D Gaussians
Novel View
Robustness
Unsupervised Pre-Training

Feedforward View Synthesis

6 Input Views

DepthSplat

Rendering

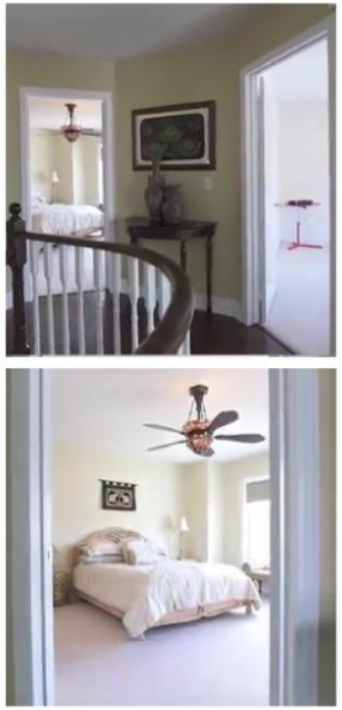
Feedforward View Synthesis

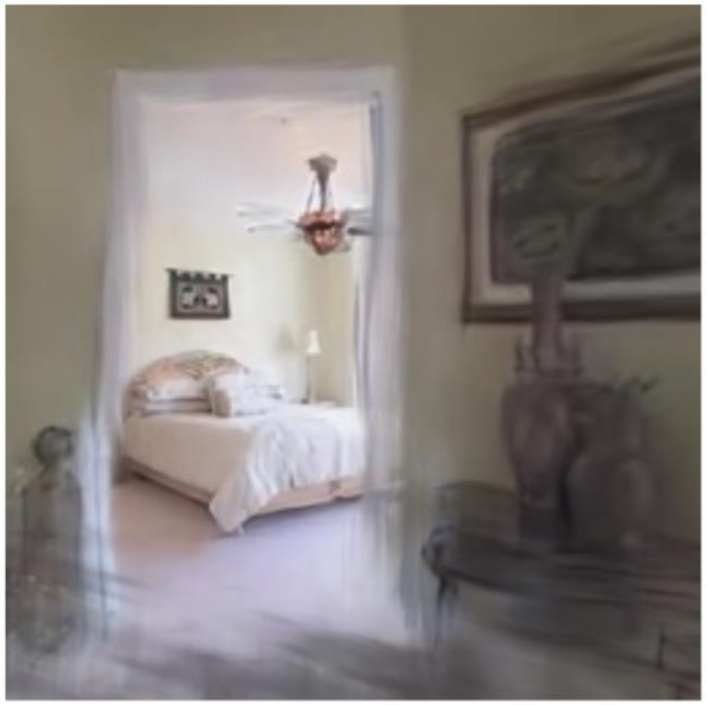
12 Input Views

DepthSplat

Rendering

Comparison

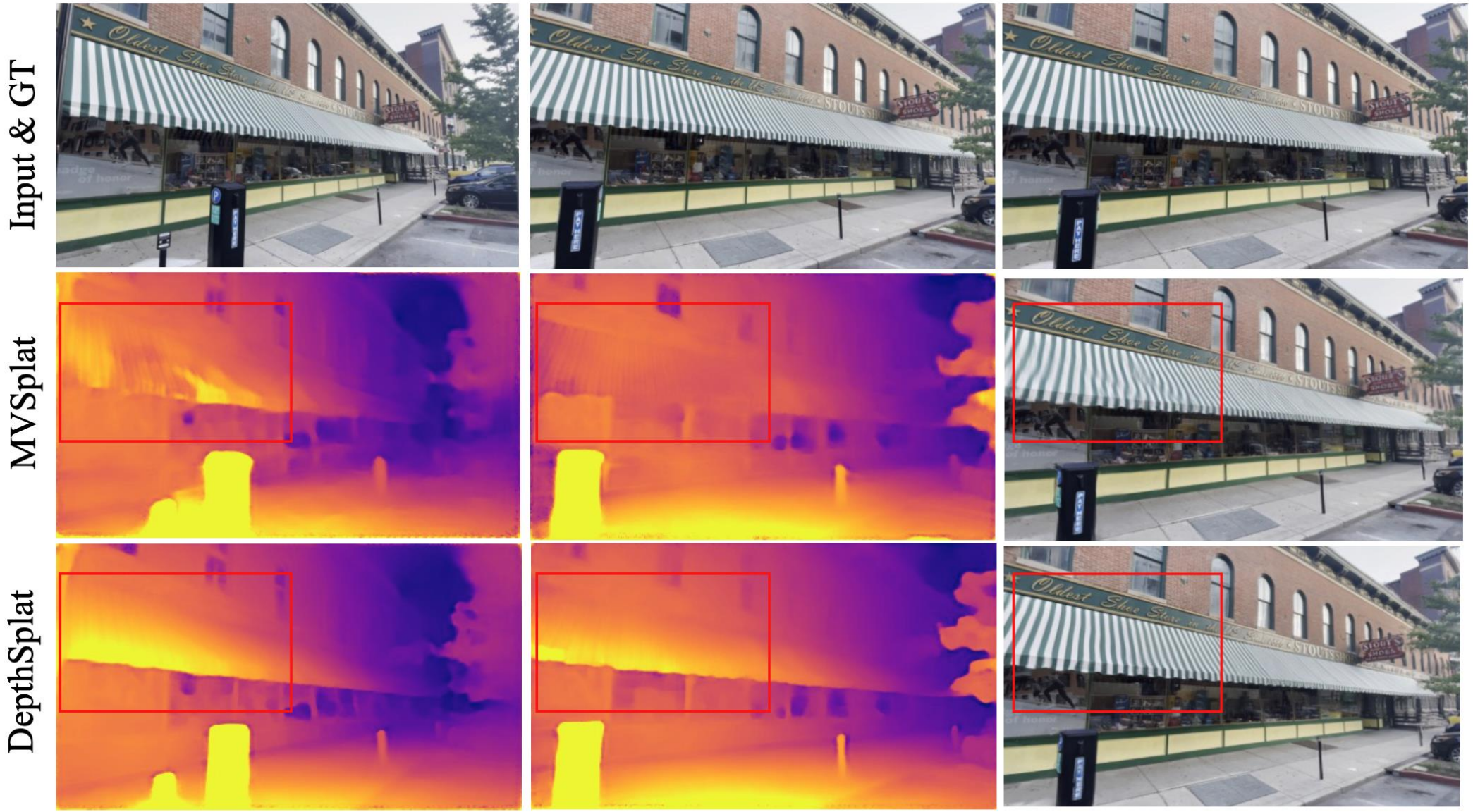




Input pixelSplat MVSplat DepthSplat

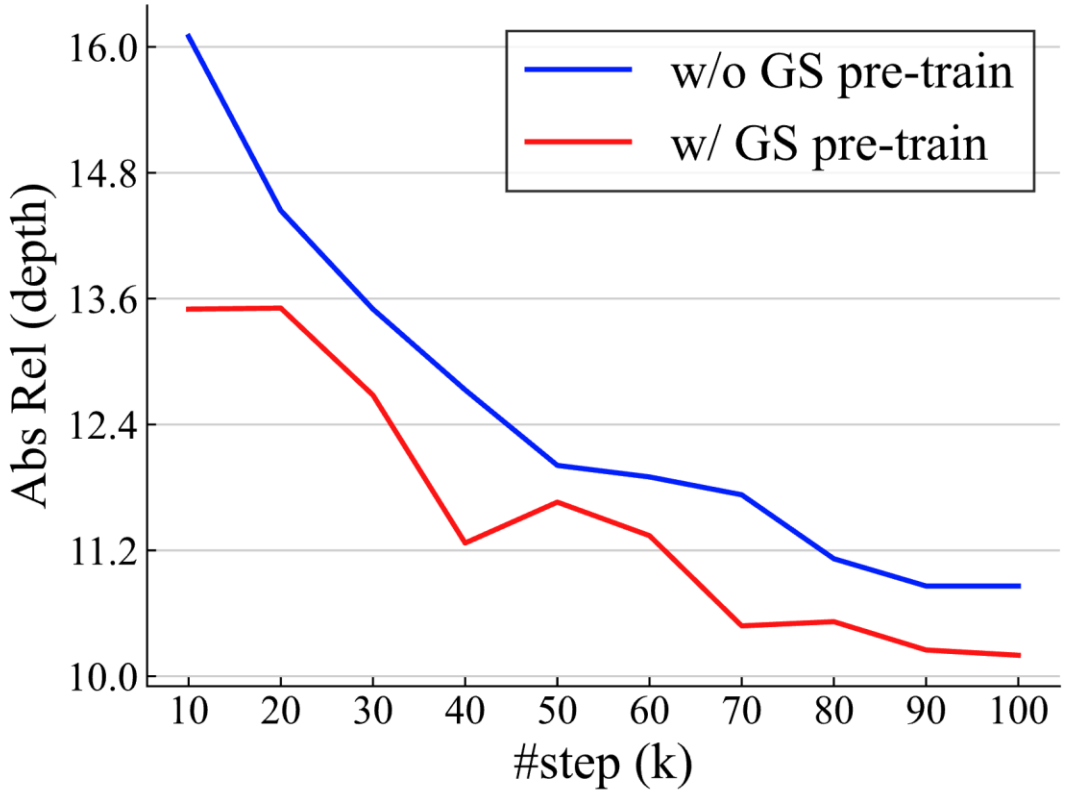
DepthSplat is significantly **more robust!**

Depth → Gaussian Splatting



Gaussian Splatting → Depth

- Unsupervised depth pre-training on RealEstate10K
- Supervised depth fine-tuning on TartanAir & VKITTI2



Validation curves of depth prediction error

Take-home Messages

- Depth and Gaussian splatting are helping each other!
- Feed-forward Gaussian splatting for large-scale scenes
- A few posed images as input… **This is not a practical setting!**

Image 1

Image 2

Non-trivial to get camera poses!

An Ideal 3D Reconstruction Pipeline

Instant, Pose-Free, Real-World 3D Everywhere

Feedforward

Pose-Agnostic

Dynamic

Fast Rendering

Arbitrary Lengths

Lighting-Robust

An Ideal 3D Reconstruction Pipeline

Instant, Pose-Free, Real-World 3D Everywhere

Feedforward

Pose-Agnostic

Dynamic

Fast Rendering

Arbitrary Lengths

Lighting-Robust

Goal: Unposed Feedforward 3DGS

Input Images **w/o poses**

3D Gaussians

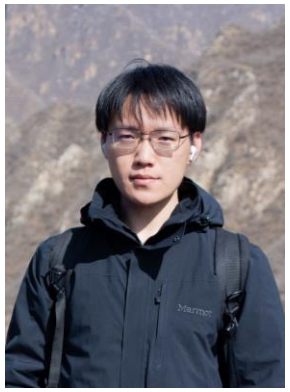
Novel Views

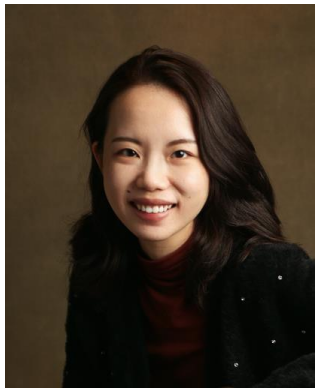
No Pose, No Problem

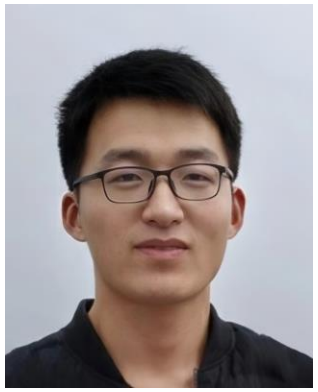
Surprisingly Simple 3D Gaussian Splats from Sparse Unposed Images

(a.k.a **NoPoSplat**)

ICLR 2025 (Oral, top 1.8%)


Botao Ye


Sifei Liu


Haofei Xu

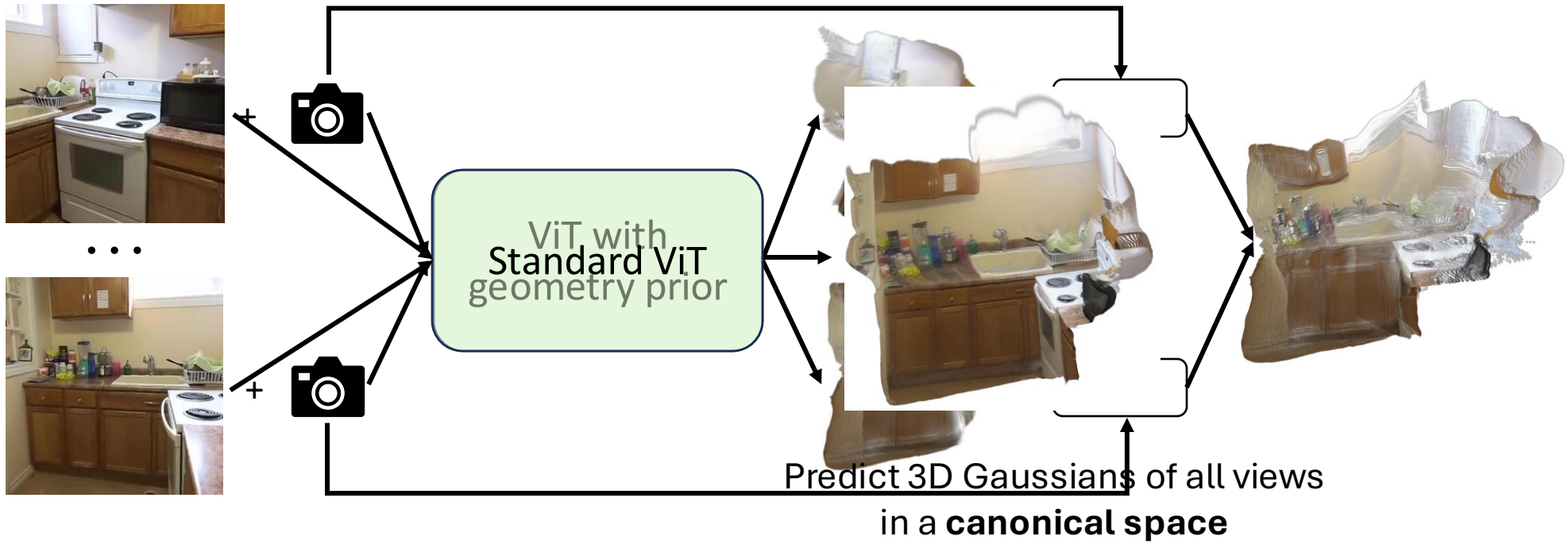

Xueting Li


Marc Pollefeys


Ming-Hsuan Yang

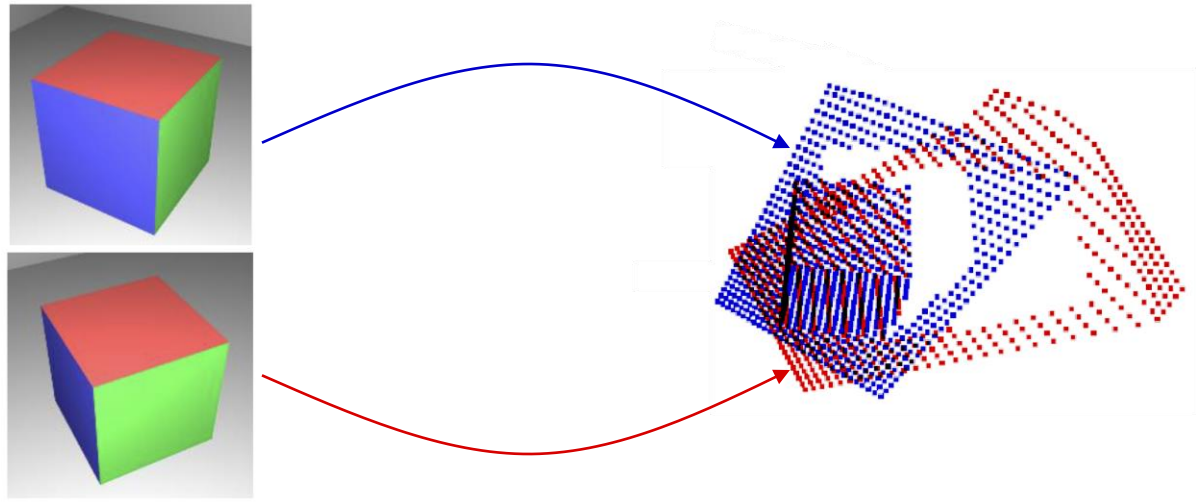

Songyou Peng

Previous Feed-forward 3DGS



Charatan, Li, Tagliasacchi, Sitzmann: pixelSplat: 3D Gaussian Splats from Image Pairs for Scalable Generalizable 3D Reconstruction. CVPR 2024

Canonical Prediction



DUSt3R [CVPR'24]

🤔Does a similar philosophy apply to Gaussians?

Point Maps

- ➖ Discrete representation
- ➖ Ground truth depth needed

3DGS

- ➕ Novel view synthesis
- ➕ Video data for training

NoPoSplat

...

Unposed Images

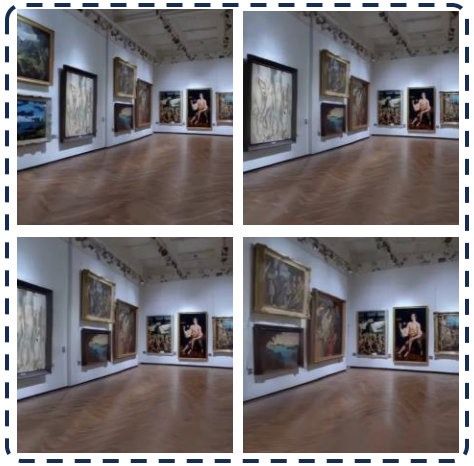
NoPoSplat

~ 15 ms

3D Gaussians
in **canonical space**

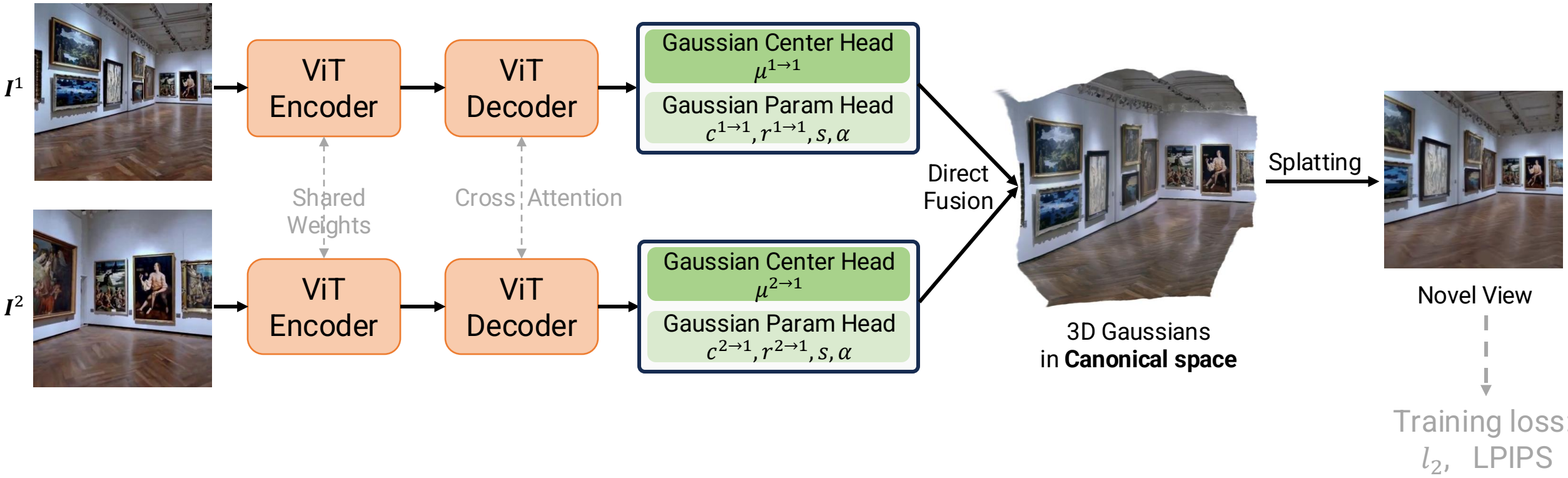
Downstream
Tasks

Pose Estimation

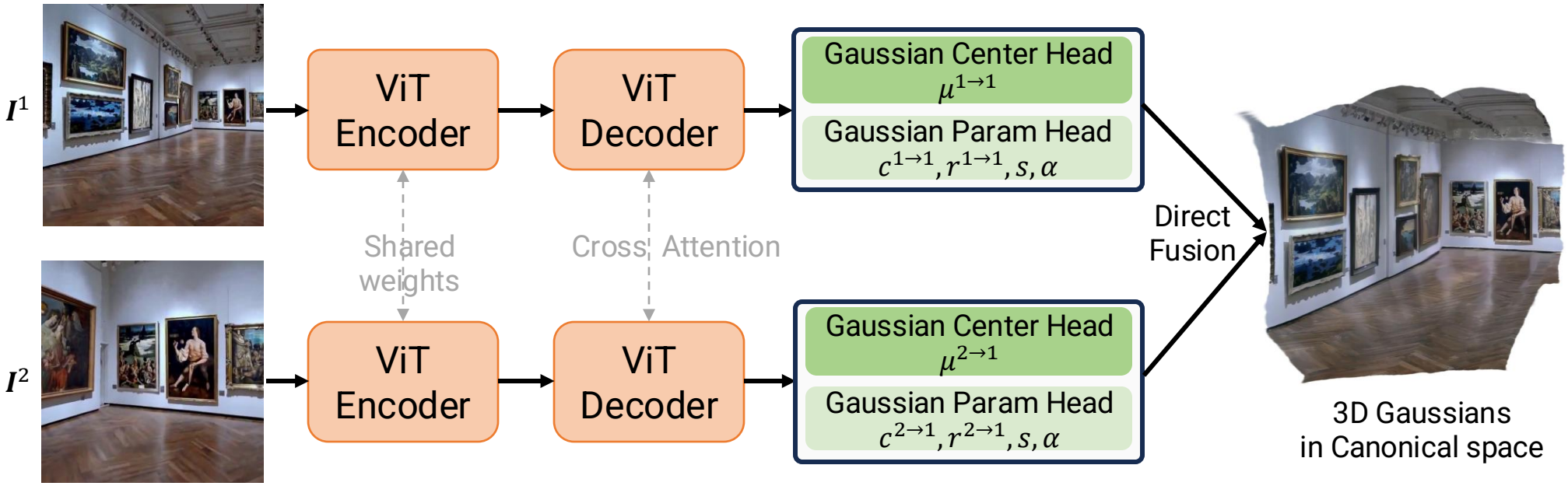


Novel View Synthesis

Architecture

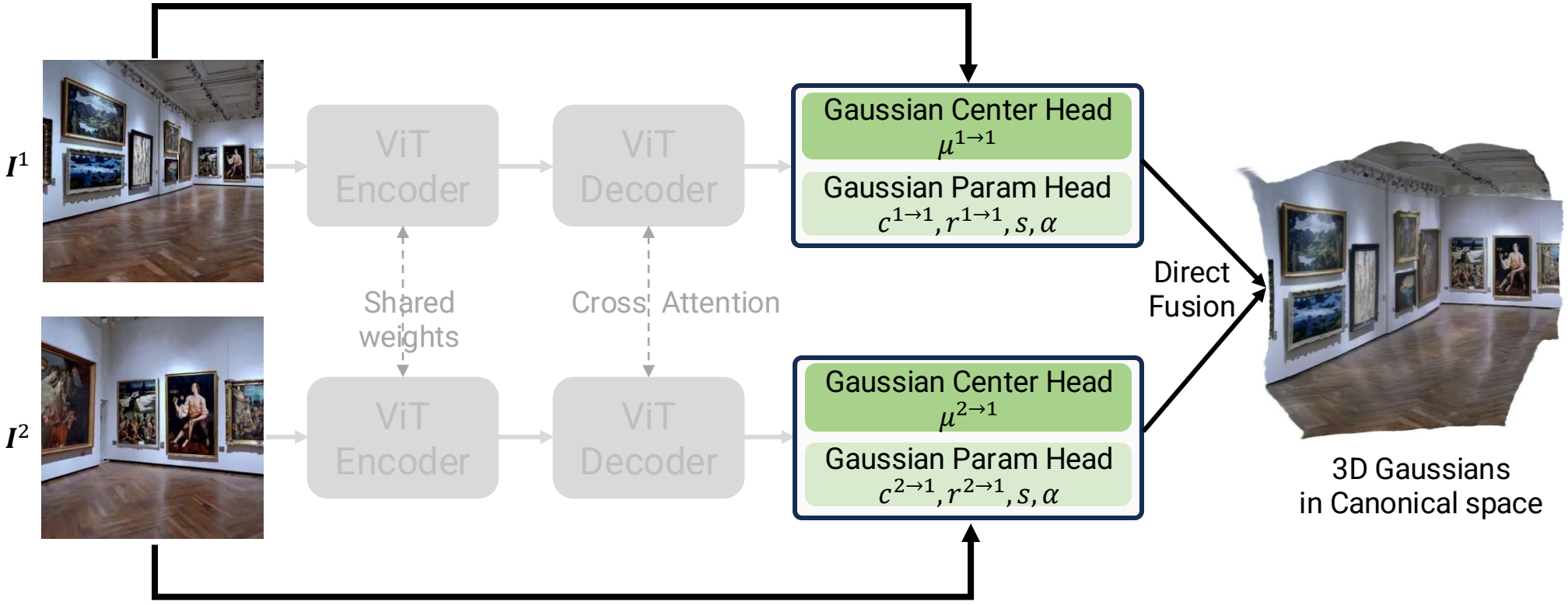


Issue 1: Blurry Rendering



Issue 1: Blurry Rendering

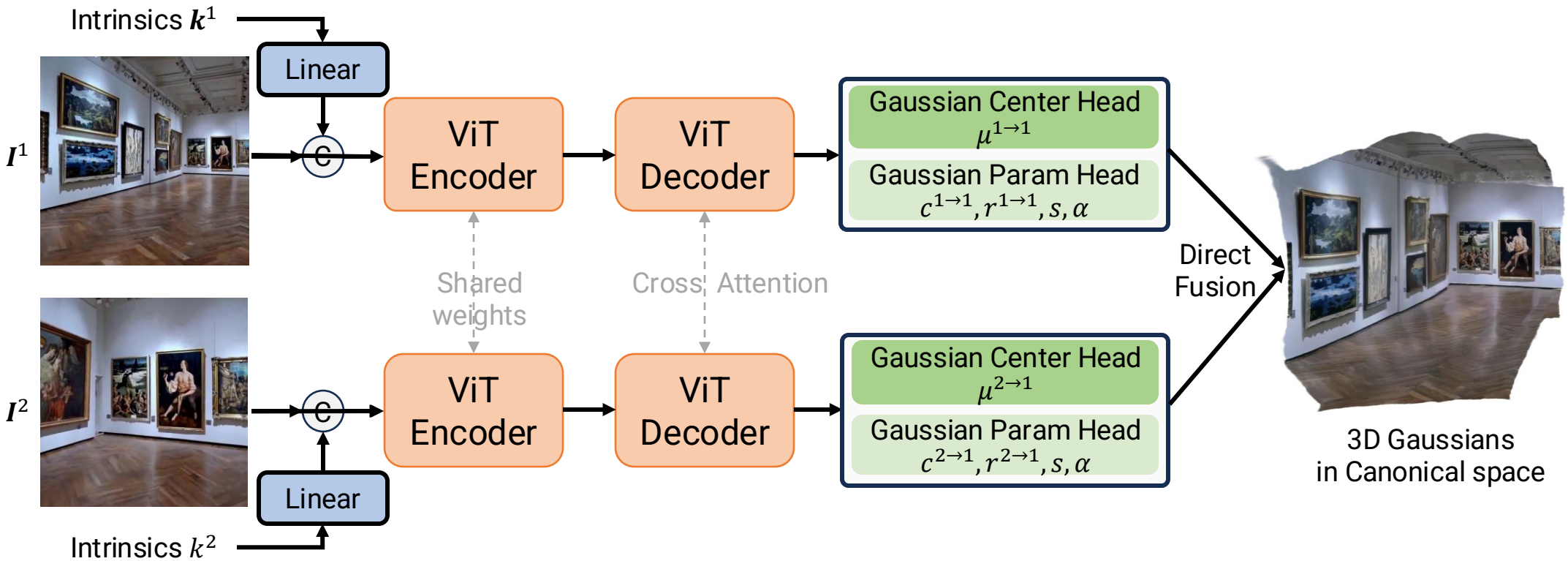
Solution: Add a shortcut!



Issue 2: Scale Ambiguity

Solution: Add the intrinsic embeddings!

$$p = K(RP + t)$$



Issue 3: Inaccurate Pose Estimation

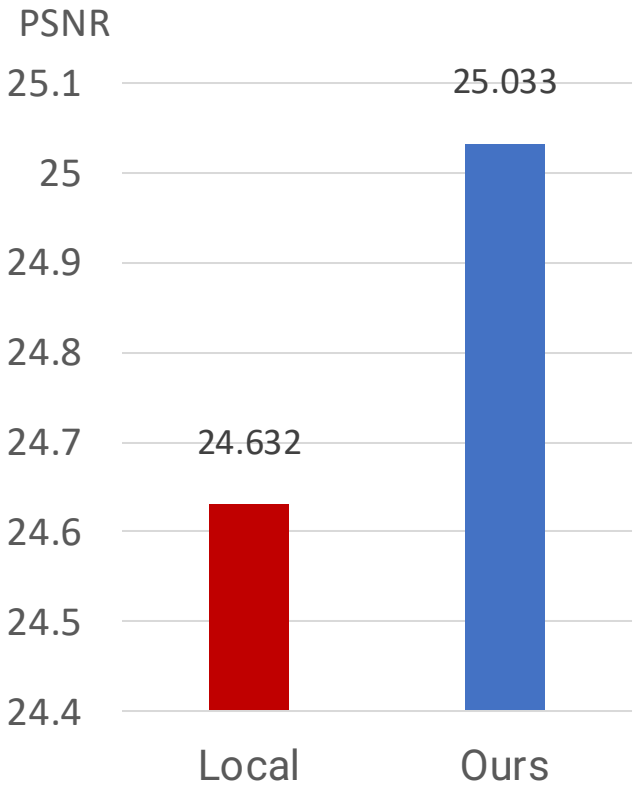
Solution: coarse-to-fine estimation

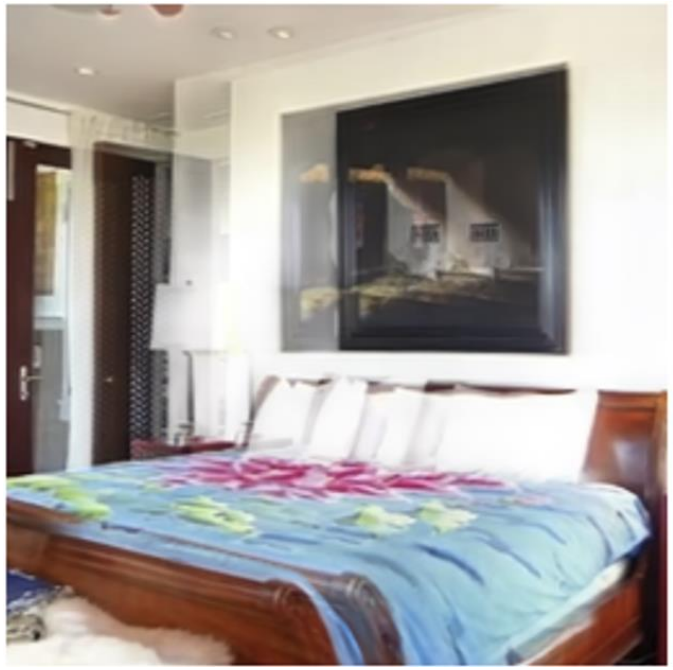
- Coarse stage: run RANSAC-PnP on Gaussian centers
- Refine stage: optimize with photometric loss

PnP	Photometric	5°	10°	20°
✓	✓	**0.318**	**0.538**	**0.717**
✓		0.287	0.506	0.692
	✓	0.017	0.027	0.051

Ablation

Canonical Gaussian prediction



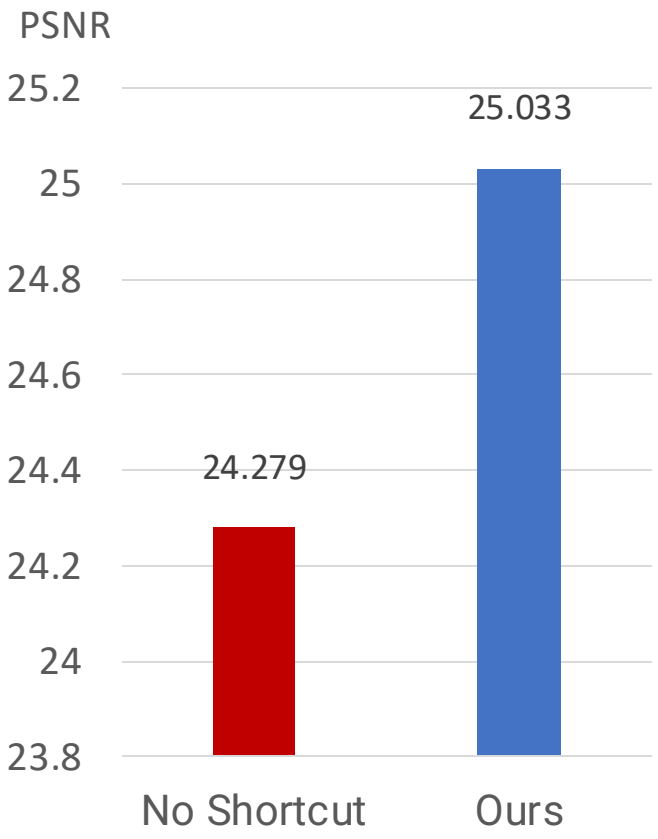


Local

Canonical

Ablation

Image shortcut leads to sharper details

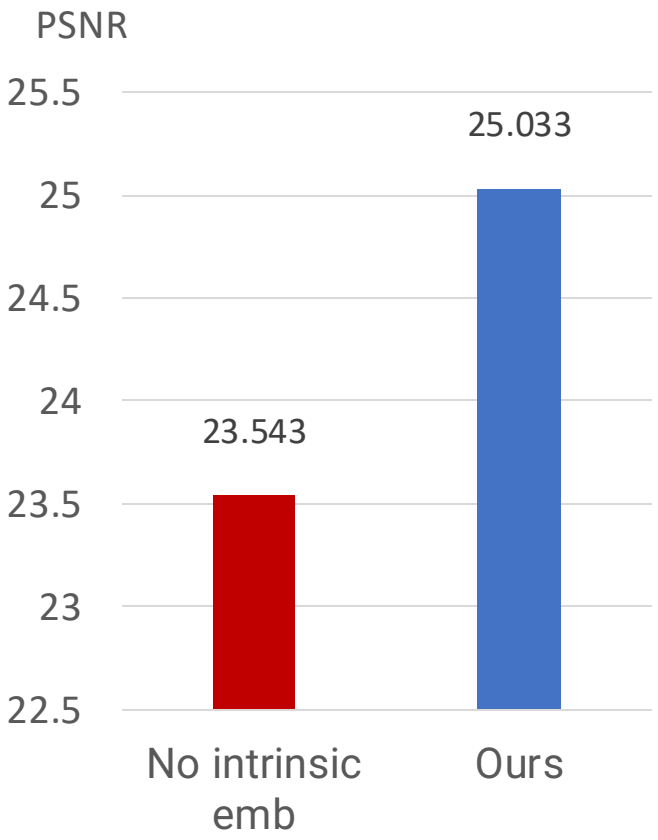


No Shortcut

Ours

Ablation

Intrinsic embedding



No Intrinsic Emb

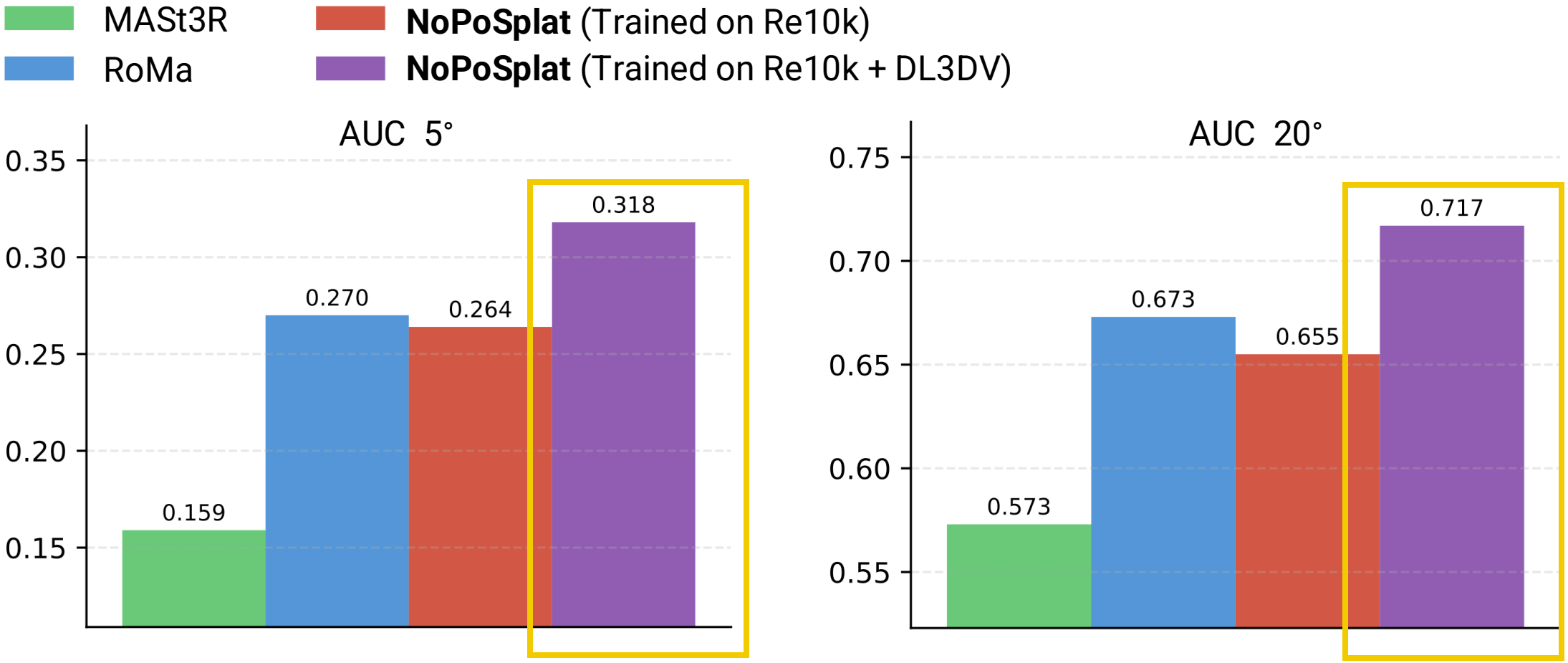
Ours

GT

What is More…

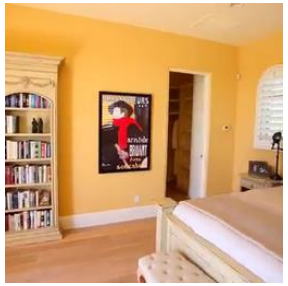
Accurate Pose Estimation

Evaluation on ScanNet

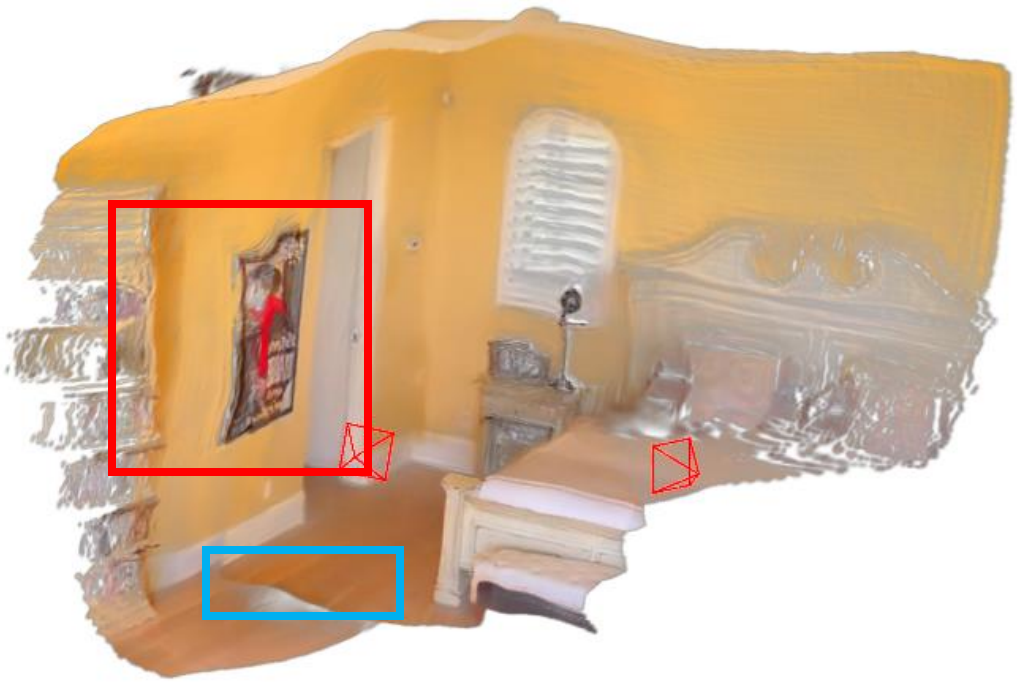


High Quality Geometry

Input Images



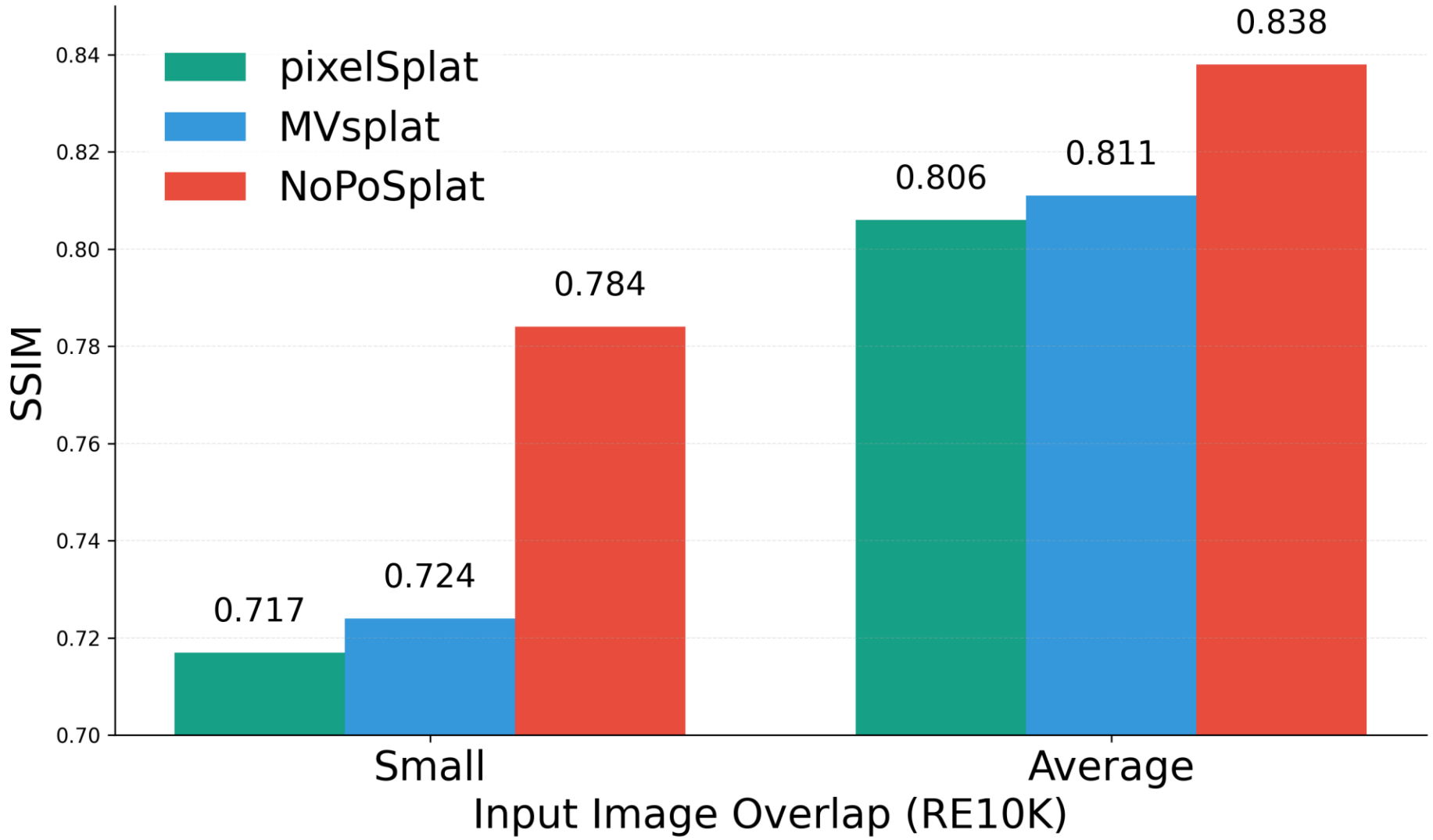
NoPoSplat (pose-free)



MVSplat (pose-required)

Appearance Quality

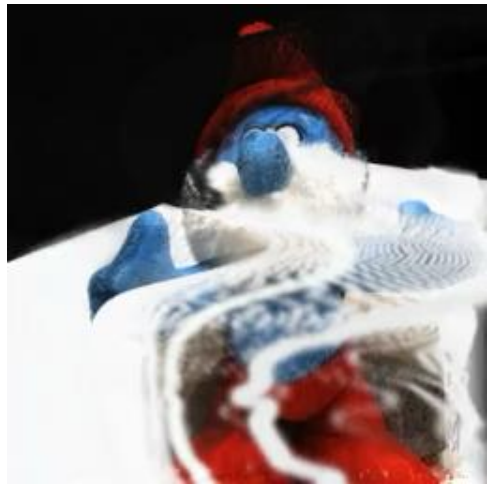
Better even than pose-required methods!



Appearance Quality

Input Views | MVSplat | **NoPoSplat**

Cross-Dataset Generalization

	Input Views		MVSplat	**NoPoSplat**
RE10K → DTU				
RE10K → ScanNet++				

In-the-Wild Data

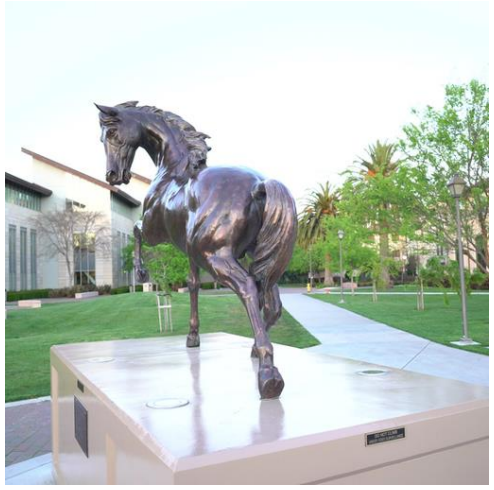
Images extracted from OpenAI Sora

Input Images

Novel Views

In-the-Wild Data

Images from Tanks & Temples

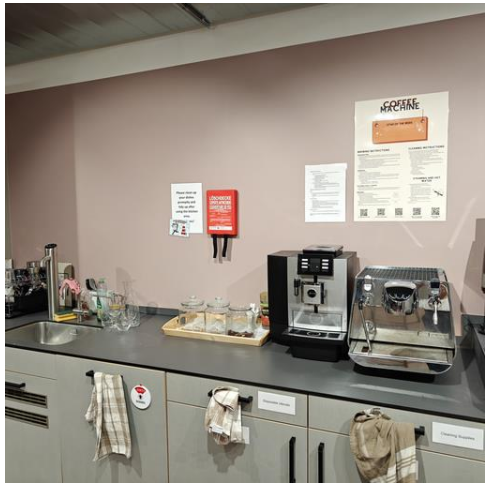


Input Images

Novel Views

In-the-Wild Data

iPhone images



Input Images

Novel Views

Take-home Messages

- Feedforward NVS can be surprisingly simple!
- Side product: SoTA relative pose estimation
- Works well with any static scenes **Not practical enough!**


THE PARISIAN DEPARTMENT STORE
TOUR EIFFEL · CHAMP-DE-MARS · MUSÉE DU LOUVRE · NOTRE-DAME · MUSÉE D'ORSAY · OPERA GARNIER · CHAMPS-ELYSÉES · GRAND PALAIS · TROCADÉRO
BIGBUS PARIS · LESCARSROUGES
BIG BUS
HOP-ON HOP-OF

Motivation

How to obtain **distractor-free 3D reconstruction**
from **casually captured & long** image sequences **in the wild**?

An Ideal 3D Reconstruction Pipeline

Instant, Pose-Free, Real-World 3D Everywhere

Feedforward

Pose-Agnostic

Dynamic

Fast Rendering

Arbitrary Lengths

Lighting-Robust

An Ideal 3D Reconstruction Pipeline

Instant, Pose-Free, Real-World 3D Everywhere

Feedforward

Pose-Agnostic

Dynamic

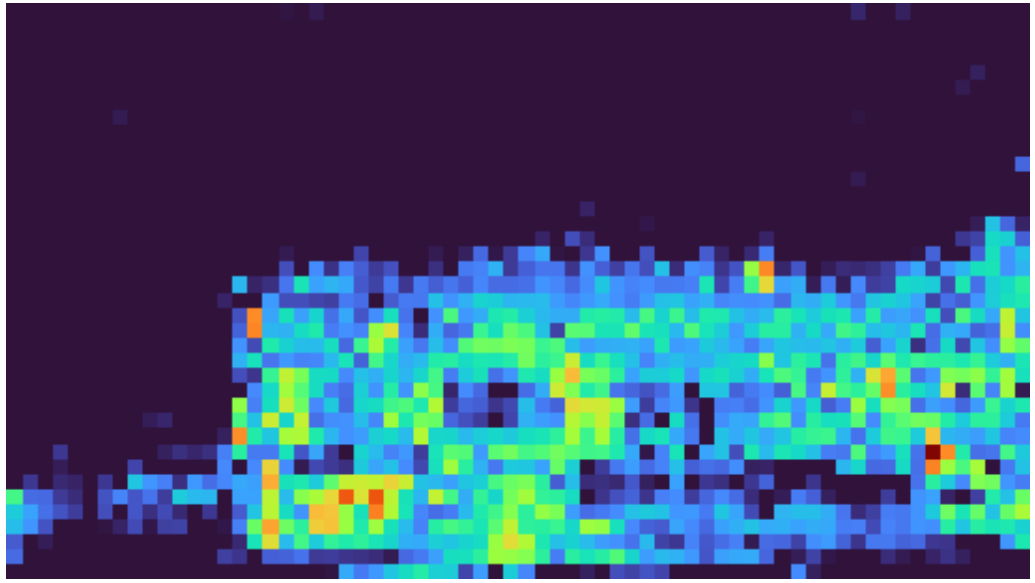
Fast Rendering

Arbitrary Lengths

Lighting-Robust

Uncertainty

Input RGB



Uncertainty Map

How to learn a good uncertainty map?

DINO v2

- A 2D foundation model producing **universal features**
- Preserve temporal-spatial consistency

Oquab et al.: DINOv2: Learning Robust Visual Features without Supervision. arXiv, 2023

How to leverage the DINO v2 to model uncertainty for 3D reconstruction?




NeRF *On-the-go*

Exploiting Uncertainty for Distractor-free NeRFs in the Wild

CVPR 2024

Weining Ren*

Zihan Zhu*

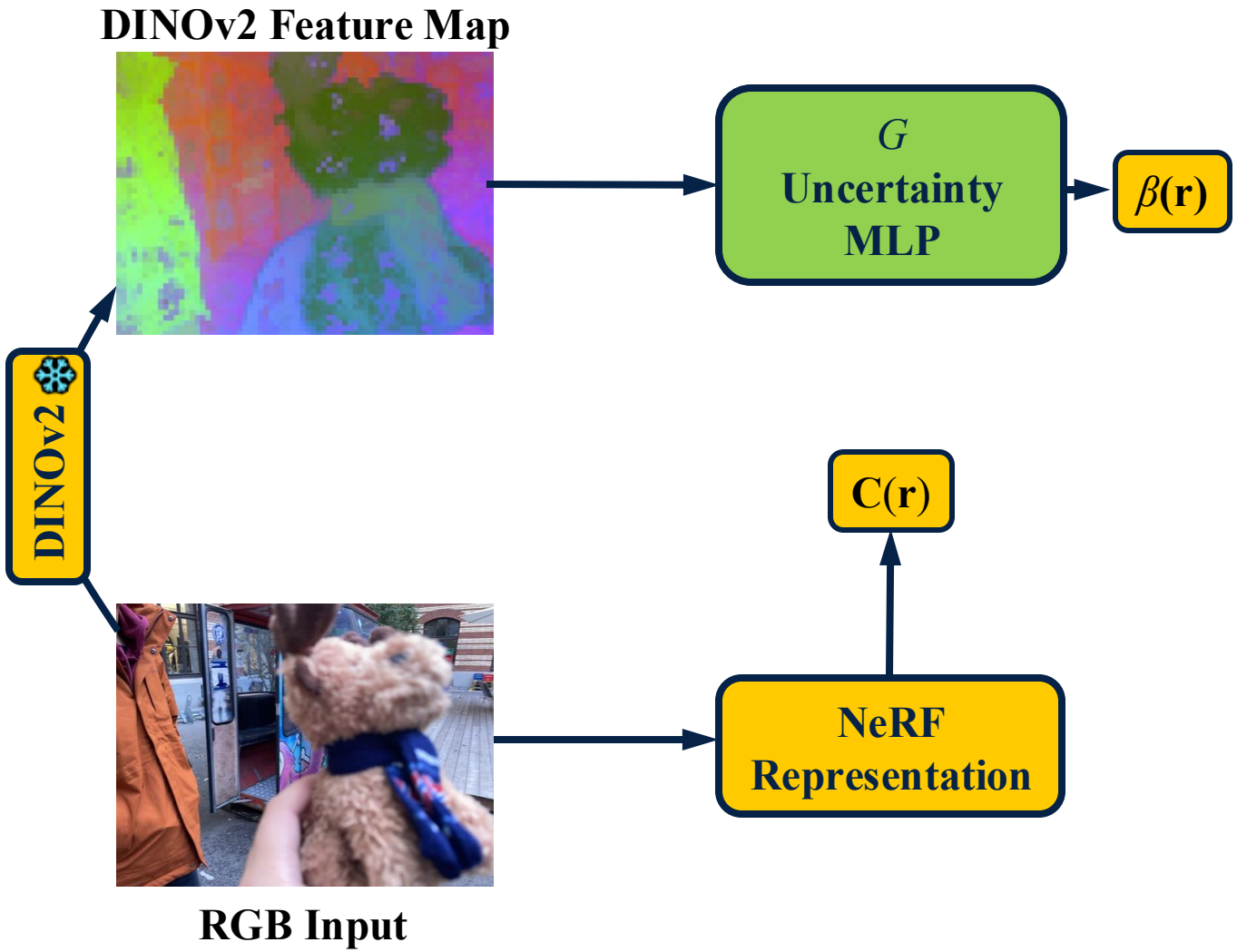
Boyang Sun

Julia Chen

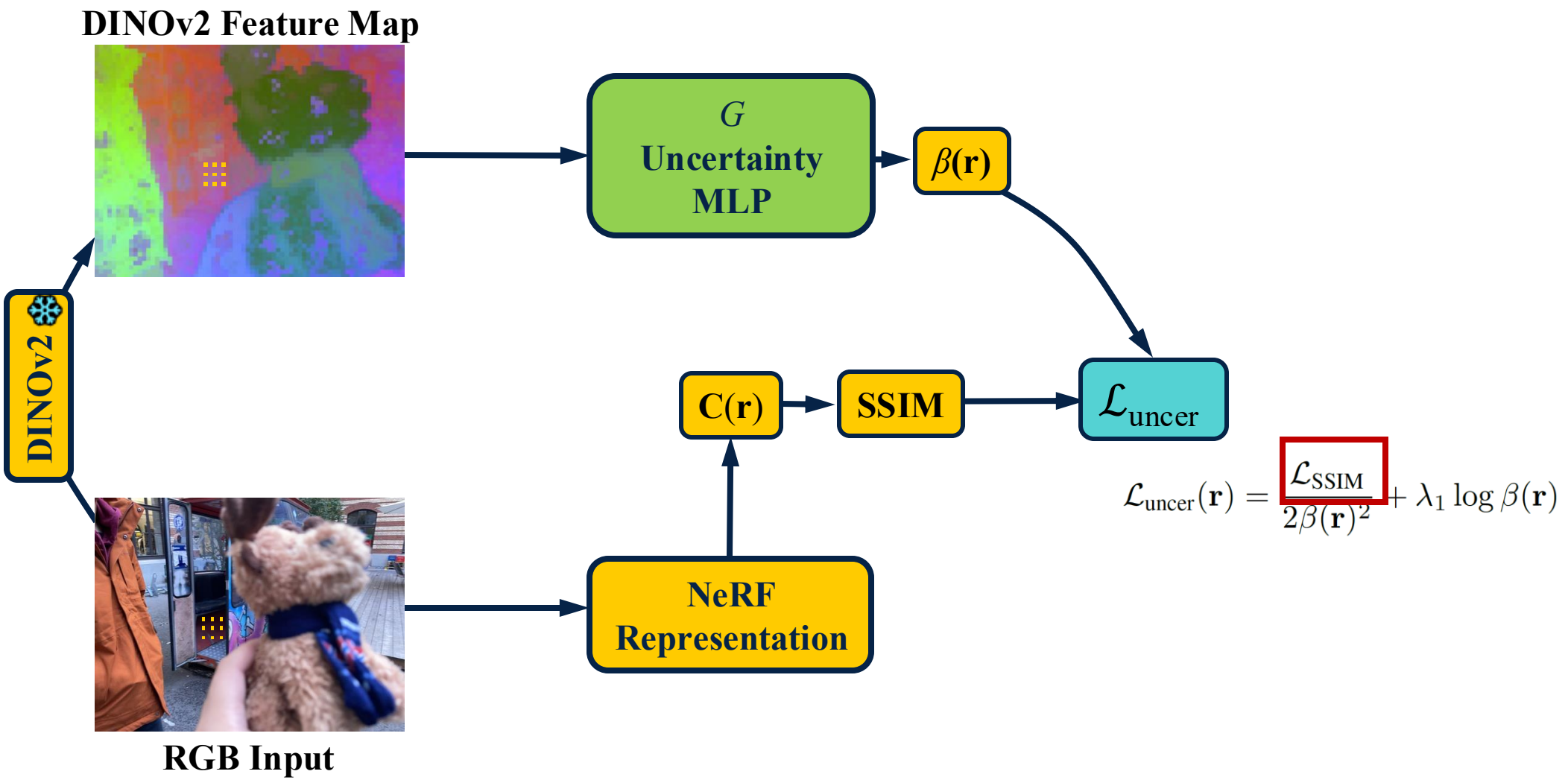
Marc Pollefeys

Songyou Peng

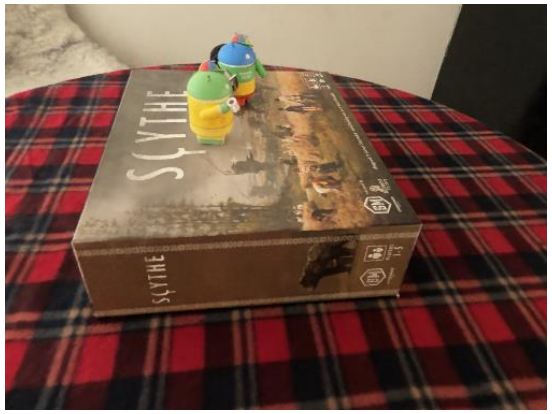
Pipeline


DINOv2 Feature Map
G
Uncertainty
MLP
β(r)
DINOv2
C(r)
NeRF
Representation
RGB Input

Pipeline


DINOv2 Feature Map
G
Uncertainty
MLP
β(r)
DINOv2
C(r)
SSIM
$\mathcal{L}_{\text{uncer}}$
$\mathcal{L}_{\text{uncer}}(\mathbf{r}) = \frac{\mathcal{L}_{\text{SSIM}}}{2\beta(\mathbf{r})^2} + \lambda_1 \log \beta(\mathbf{r})$
NeRF
Representation
RGB Input

To Learn the Uncertainty MLP…

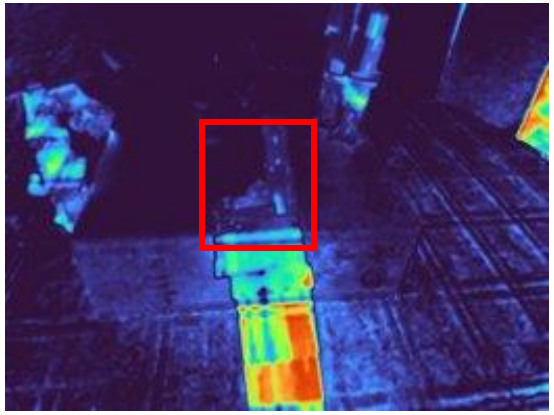


Rendered RGB

Train RGB

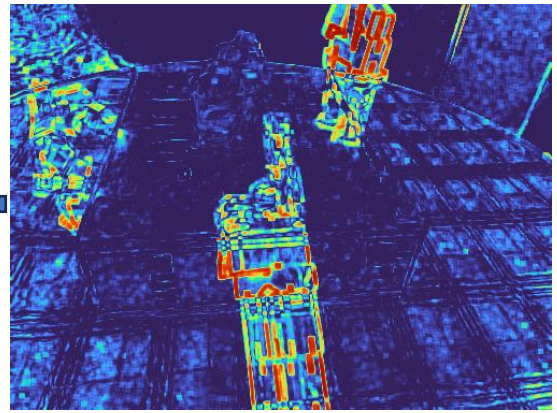
Why SSIM?

Leverage structure information when RGB is similar!



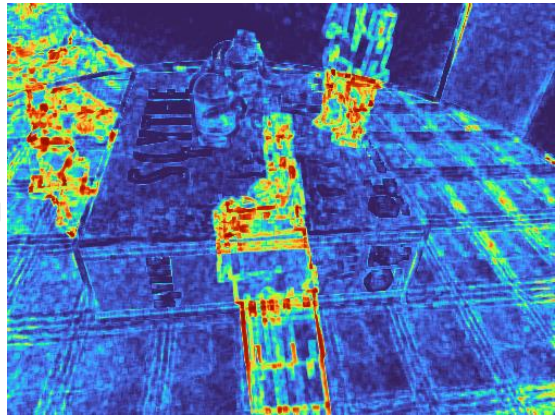
Luminance

+



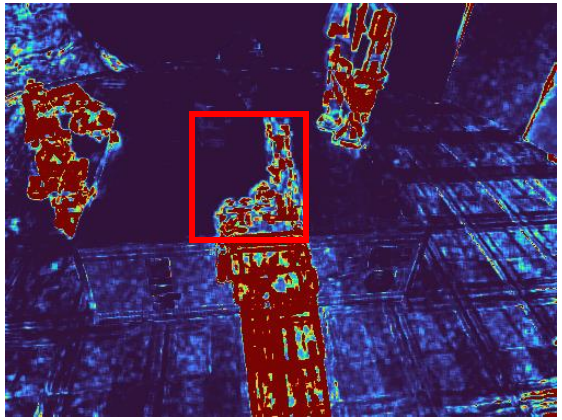
Contrast

+



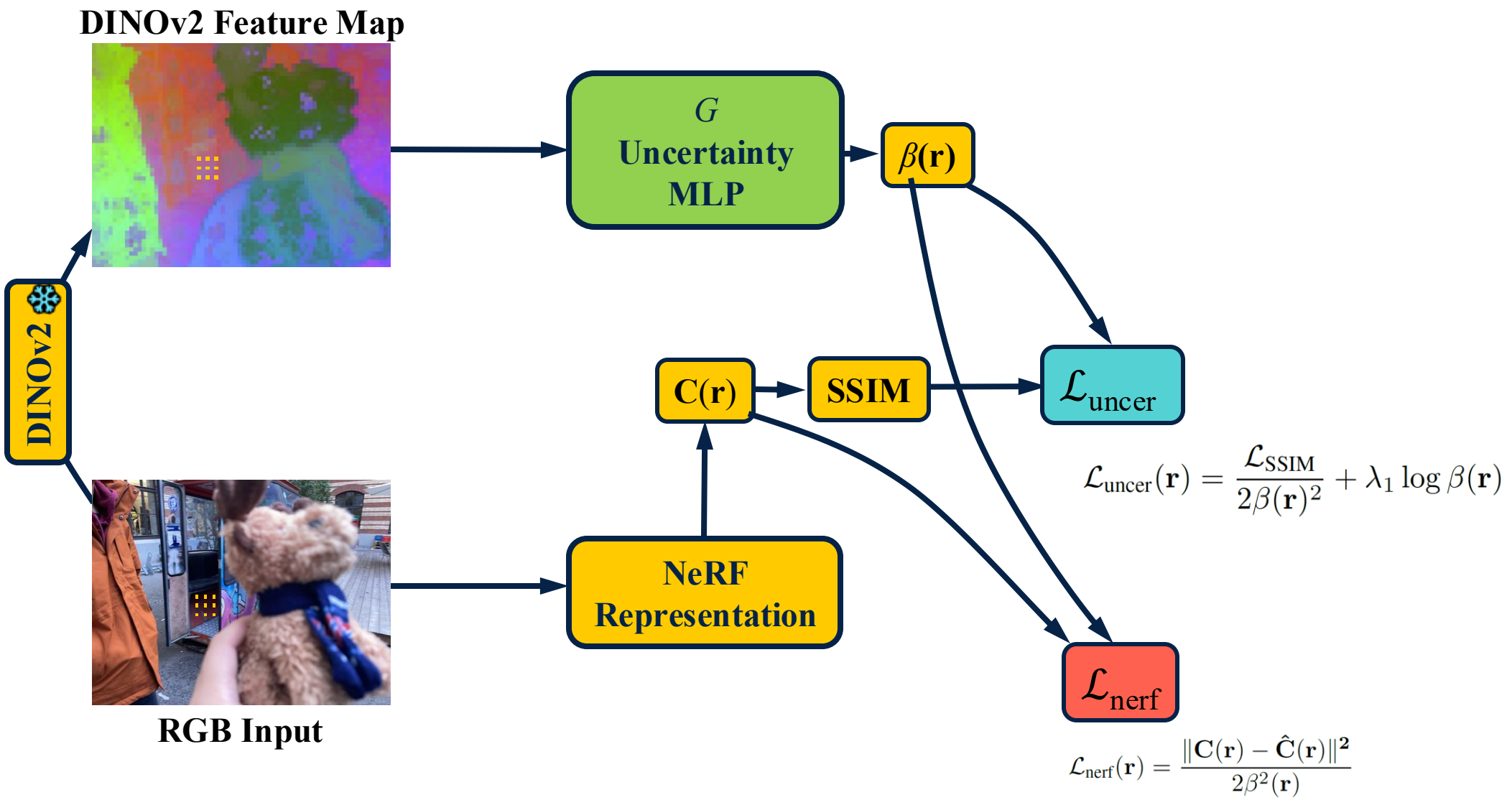
Structure

=



SSIM Error

Pipeline


DINOv2 Feature Map
G
Uncertainty
MLP
β(r)
DINOv2
C(r)
SSIM
$\mathcal{L}_{\text{uncer}}$
$\mathcal{L}_{\text{uncer}}(\mathbf{r}) = \frac{\mathcal{L}_{\text{SSIM}}}{2\beta(\mathbf{r})^2} + \lambda_1 \log \beta(\mathbf{r})$
NeRF
Representation
$\mathcal{L}_{\text{nerf}}$
$\mathcal{L}_{\text{nerf}}(\mathbf{r}) = \frac{\|\mathbf{C}(\mathbf{r}) - \hat{\mathbf{C}}(\mathbf{r})\|^2}{2\beta^2(\mathbf{r})}$
RGB Input

Results

`Train Station` - **Input Images**

`Train Station` - **Rendering Comparisons**

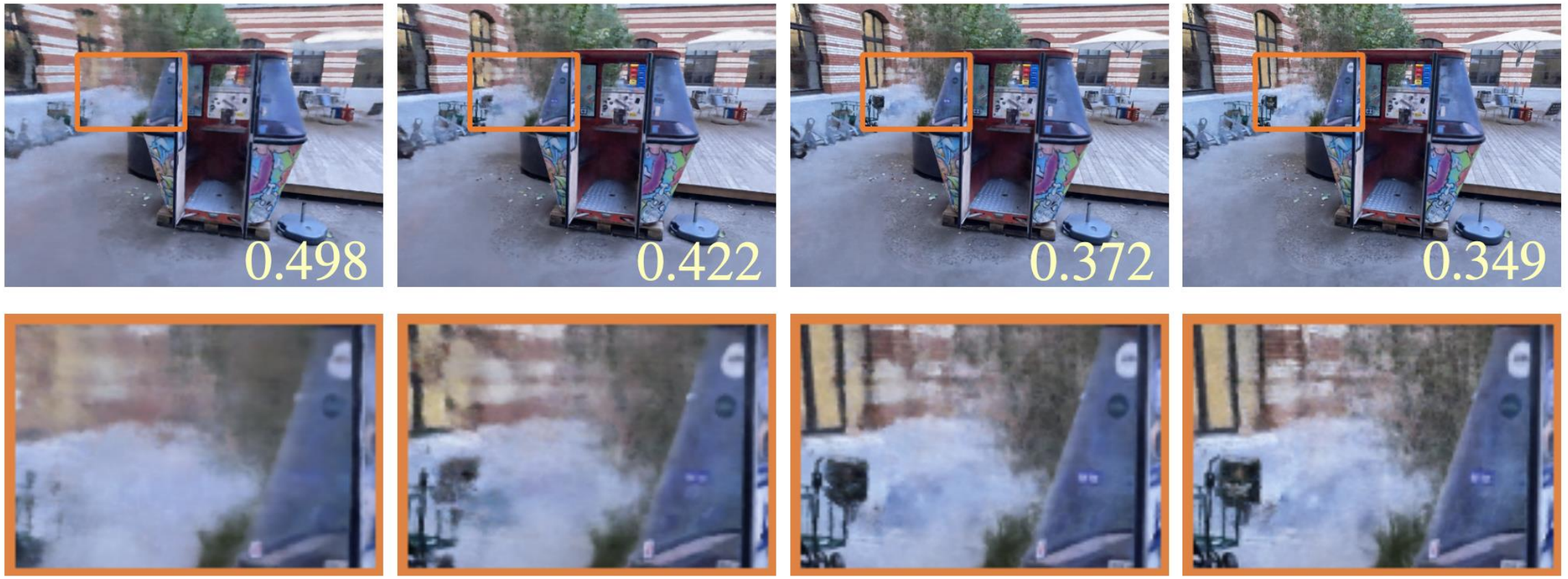
Occlusion Ratio: **High**

`Patio-High` - **Input**

`Patio-High` - **Rendering Comparisons**

Analysis

Analysis - Efficiency



RobustNeRF

Sabour et al.: RobustNeRF: Ignoring Distractors with Robust Losses. CVPR, 2023 (Highlight)

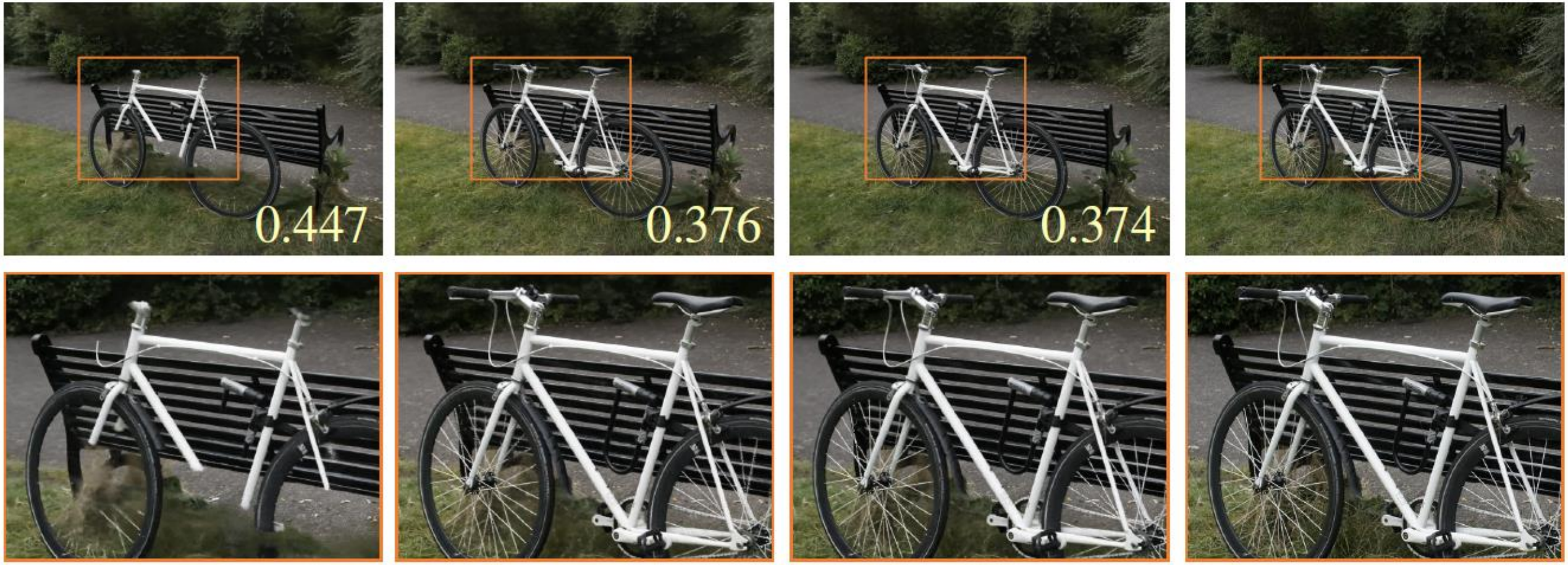
Analysis - Efficiency

25K 50K 100K 250K

NeRF *On-the-go*

(Ours)

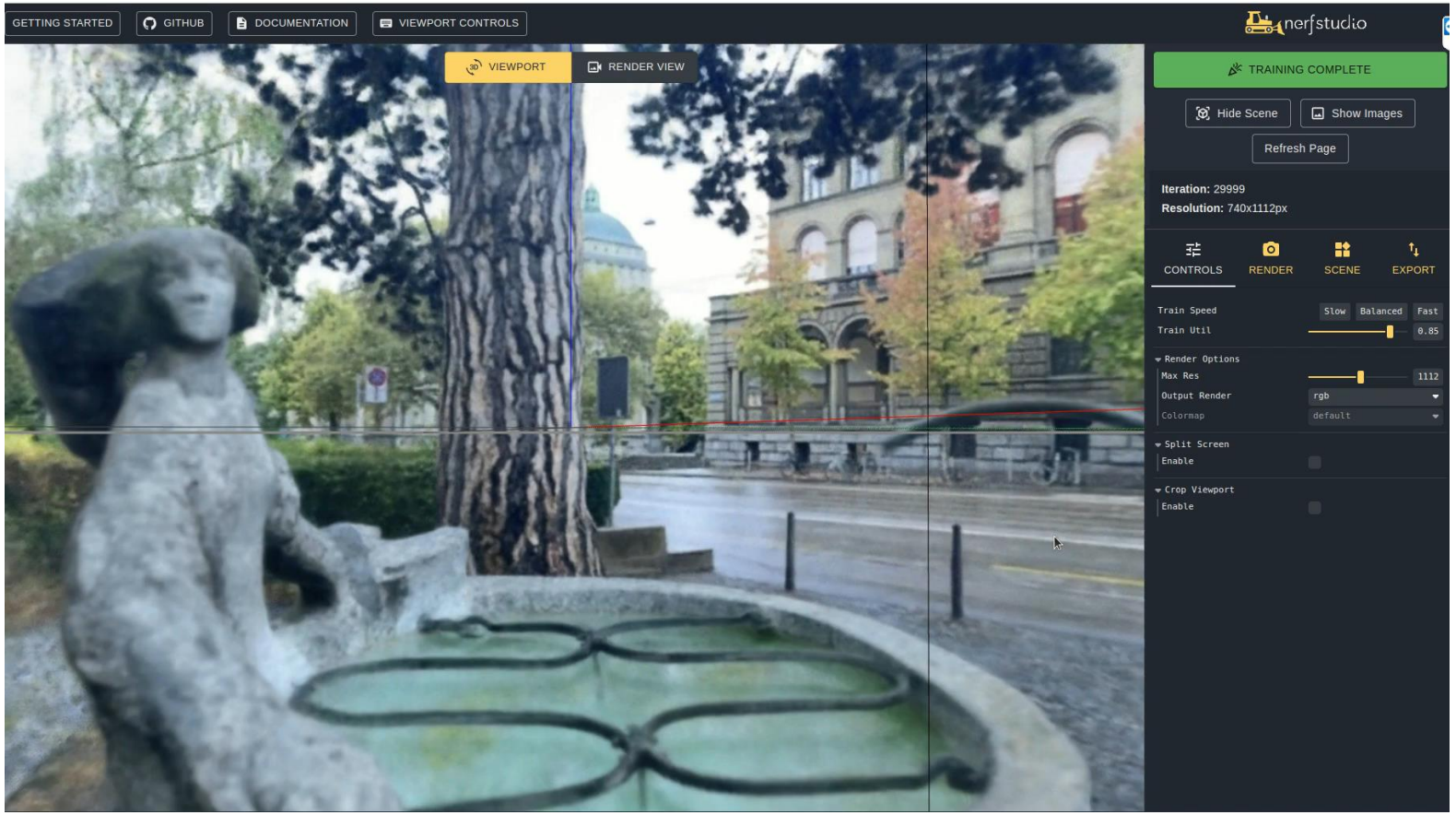
Analysis – Static Scene



RobustNeRF **Ours** MipNeRF 360 GT

Take-home Messages

- ***On-the-go*** module is plug-and-play for all NeRF methods
  - Integrated into NeRFStudio



Take-home Messages

- ***On-the-go*** module is plug-and-play for all NeRF methods
  - Integrated into NeRFStudio

- **2D foundation model** (DINOv2) rocks!

However, it is VERY SLOW

In-the-Wild

An Ideal 3D Reconstruction Pipeline

Instant, Pose-Free, Real-World 3D Everywhere

Feedforward

Pose-Agnostic

Dynamic

Fast Rendering

Arbitrary Lengths

Lighting-Robust

An Ideal 3D Reconstruction Pipeline

Instant, Pose-Free, Real-World 3D Everywhere

Feedforward

Pose-Agnostic

Dynamic

Fast Rendering

Arbitrary Lengths

Lighting-Robust

3DGS

from **casually captured** images **in the wild**

- 😃 Robustly handle **arbitrary occlusions**
- 😃 Model any **illumination changes**
- 😃 **Real-time** rendering!

WildGaussians

3D Gaussian Splatting In the Wild

NeurIPS 2024

Jonas Kulhanek

Songyou Peng

Zuzana Kukelova

Marc Pollefeys

Torsten Sattler

Pipeline

Local appearance modelling

Pipeline

Optimizing with uncertainty

GT image

DINO features

Uncertainty predictor (Affine transform)

SSIM + L

DINO cossim loss weighted by uncertainty

s weighted by uncertainty

Rendering

DINO features

Phototourism Dataset


K-Planes
Wild Gaussians
FPS 0.016
FPS 31
Rendering Speed

An Ideal 3D Reconstruction Pipeline

Instant, Pose-Free, Real-World 3D Everywhere

Feedforward

Pose-Agnostic

Dynamic

Fast Rendering

Arbitrary Lengths

Lighting-Robust

An Ideal 3D Reconstruction Pipeline

Instant, Pose-Free, Real-World 3D Everywhere

Feedforward

Pose-Agnostic

Dynamic

Fast Rendering

Arbitrary Lengths

Lighting-Robust

Stanford University

ETH zürich

WildGS-SLAM

Monocular Gaussian Splatting SLAM in Dynamic Environments

CVPR 2025

Jianhao Zheng*

Zihan Zhu*

Valentin Bieri

Marc Pollefeys

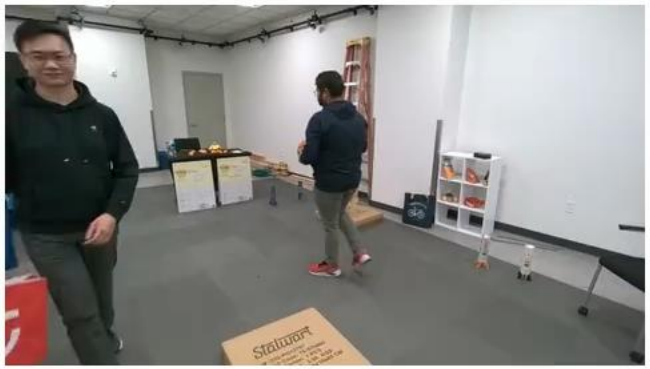
Songyou Peng

Iro Armeni

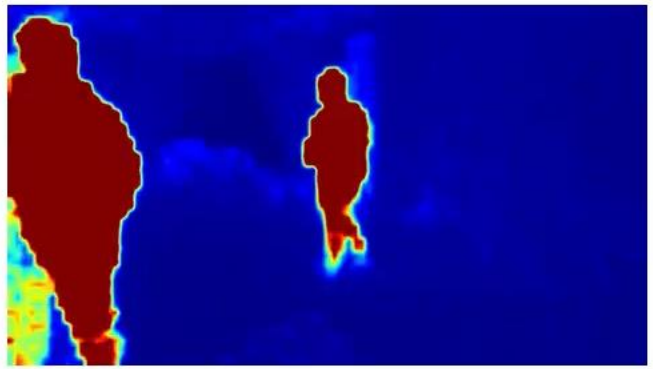
Online mapping

— GT traj — Est. traj

Input: RGB frames



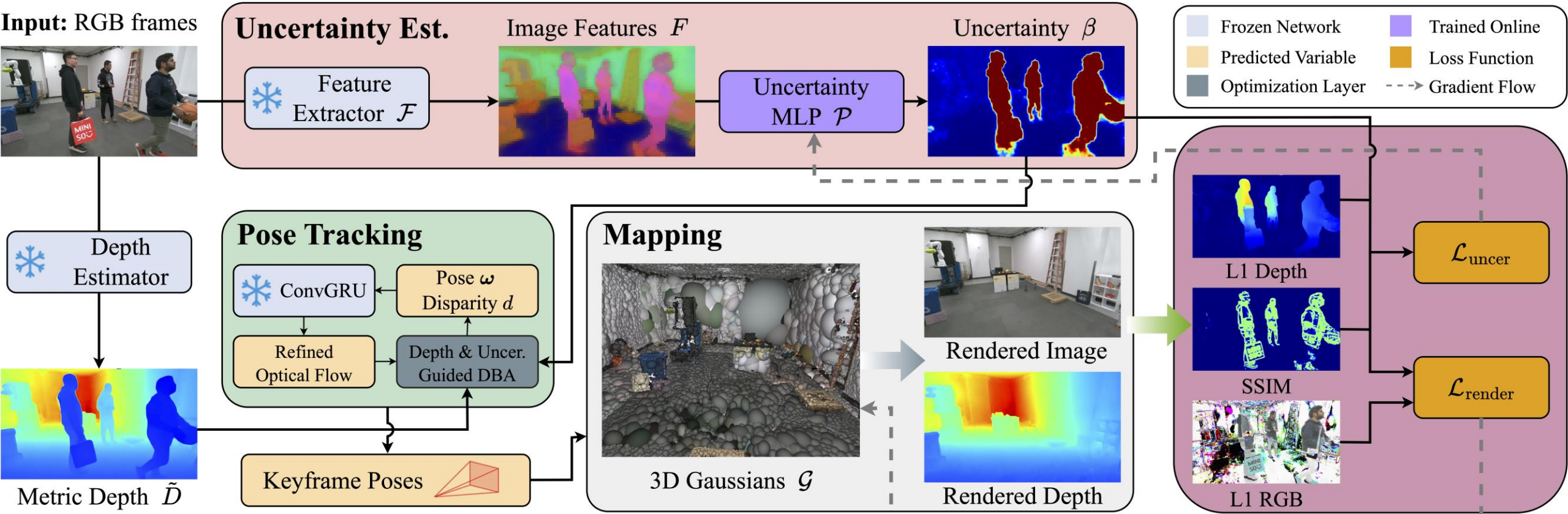
Uncertainty



Rendered (online)

Rendered (final)

WildGS-SLAM

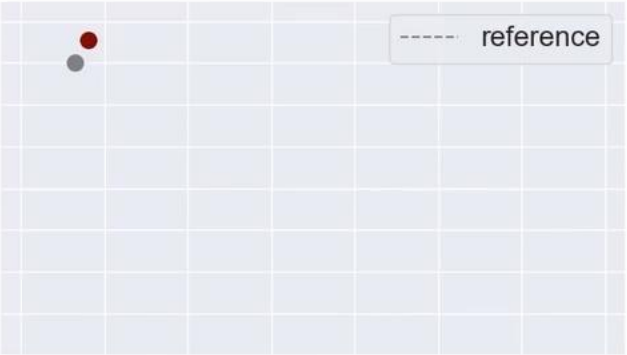

Input: RGB frames
Uncertainty Est.
Image Features F
Uncertainty β
Feature Extractor $\mathcal{F}$
Uncertainty MLP $\mathcal{P}$
Frozen Network
Trained Online
Predicted Variable
Loss Function
Optimization Layer
Gradient Flow
Depth Estimator
Pose Tracking
ConvGRU
Pose ω
Disparity d
Refined Optical Flow
Depth & Uncer. Guided DBA
Keyframe Poses
Metric Depth $\tilde{D}$
Mapping
3D Gaussians $\mathcal{G}$
Rendered Image
Rendered Depth
L1 Depth
SSIM
L1 RGB
$\mathcal{L}_{\text{uncer}}$
$\mathcal{L}_{\text{render}}$

Results

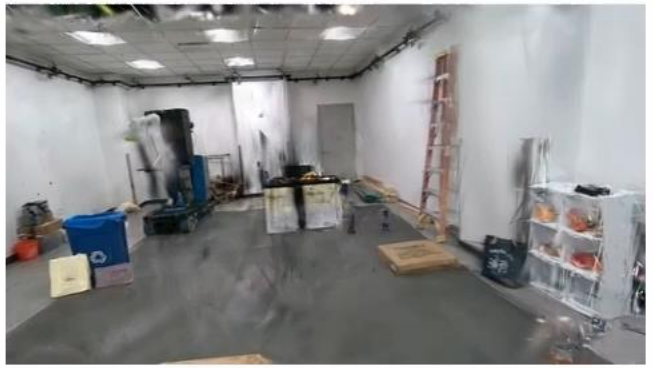
Umbrella

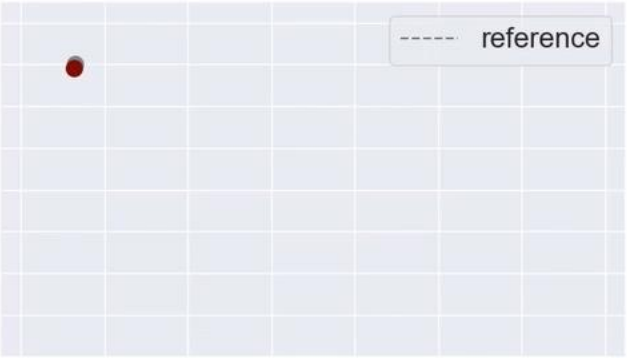
Input Frames

Traj. Error Colormap

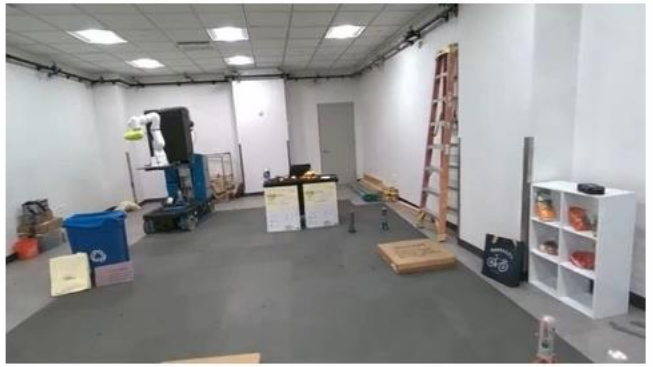


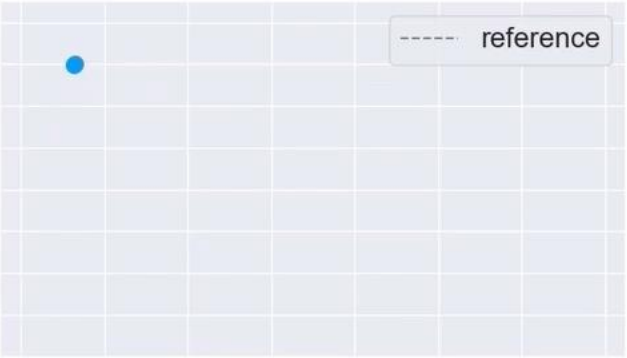
MonoGS [CVPR' 24]





Splat-SLAM [arXiv' 24]





WildGS-SLAM (Ours)

Results

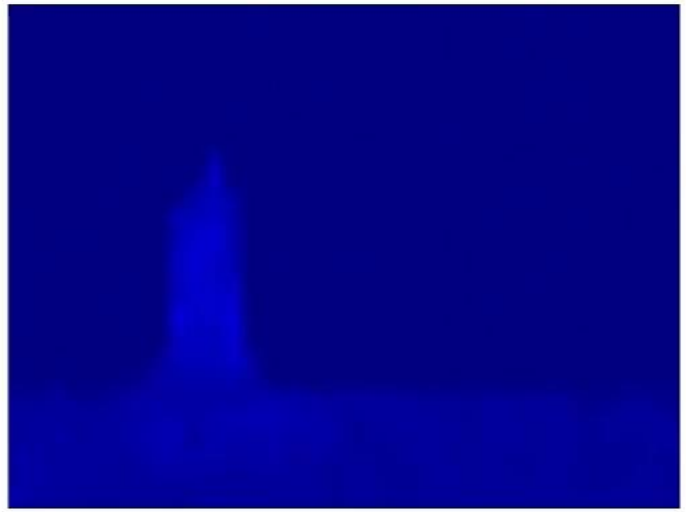
Tower

Input Frames

Rendered (MonoGS)

Rendered (Splat-SLAM)

Rendered (Ours)



Uncertainty (Ours)

An Ideal 3D Reconstruction Pipeline

Instant, Pose-Free, Real-World 3D Everywhere

Feedforward

Pose-Agnostic

Dynamic

Fast Rendering

Arbitrary Lengths

Lighting-Robust

Opinion

3D static reconstruction is almost at the last mile

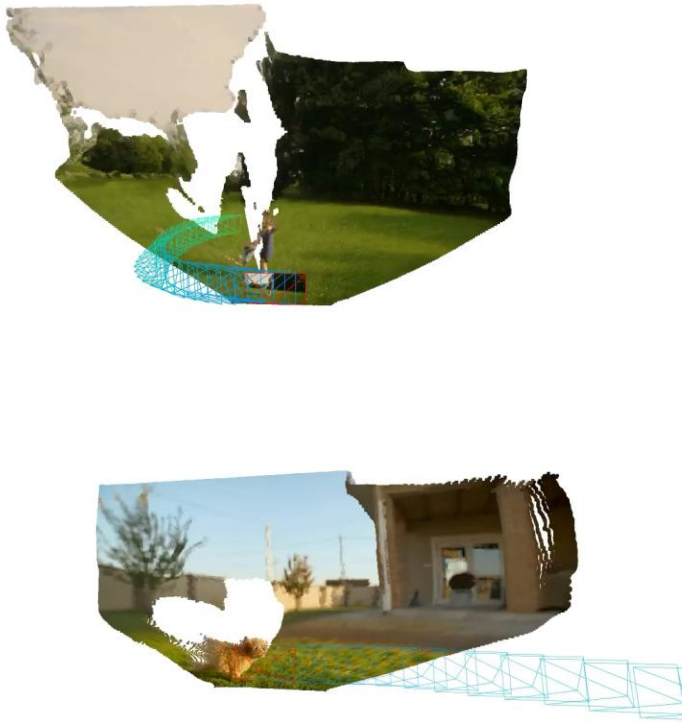
VGGT

[CVPR'25 Oral]

Opinion

3D static reconstruction is almost at the last mile

VGGT
[CVPR'25 Oral]



MegaSaM
[CVPR'25 Oral]

Challenges & Opportunities

How to robustly handle long sequences?

CUT3R

Static Videos

Dynamic Videos

Unstructured Photo Collections

CUT3R

[CVPR'25 Oral]

Challenges & Opportunities

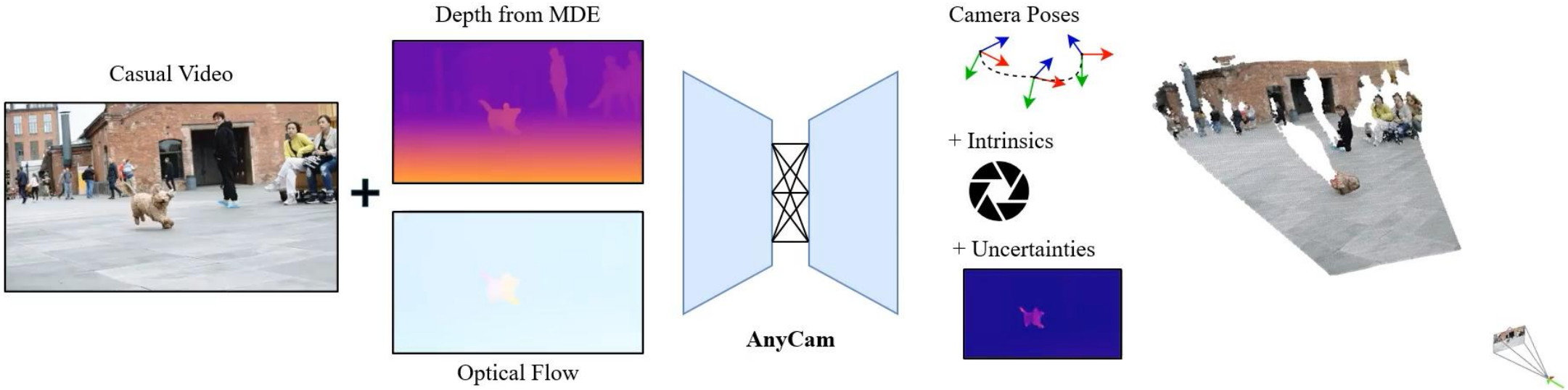
4D (3D dynamic) reconstruction is still very hard, especially for NVS

MegaSaM
[CVPR'25 Oral]

Gaussian-Flow
[CVPR'24 Highlight]

Challenges & Opportunities

Feedforward pose estimation is not solved



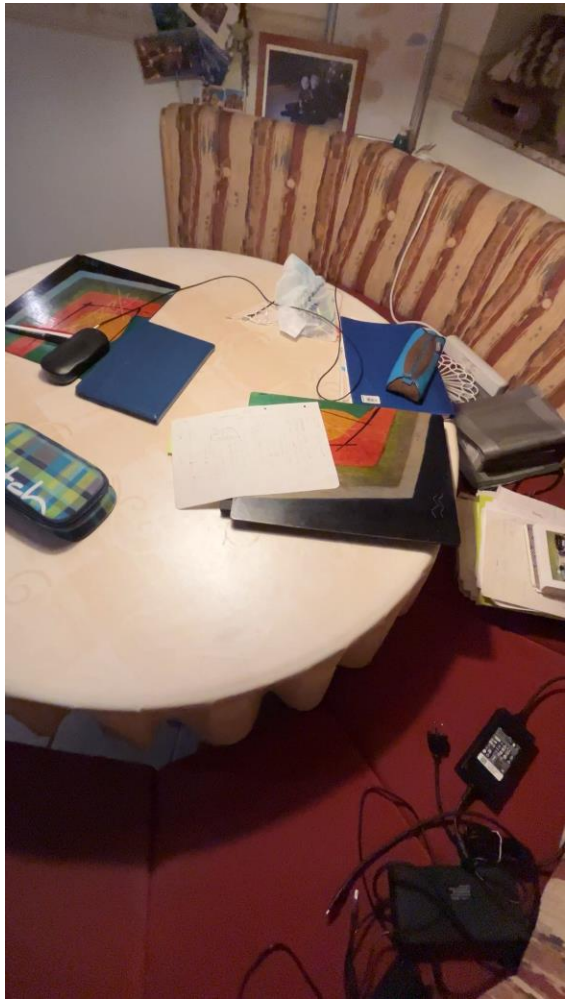
AnyCam

[CVPR'25]

Challenges & Opportunities

How to continuously update your 3D reconstruction?

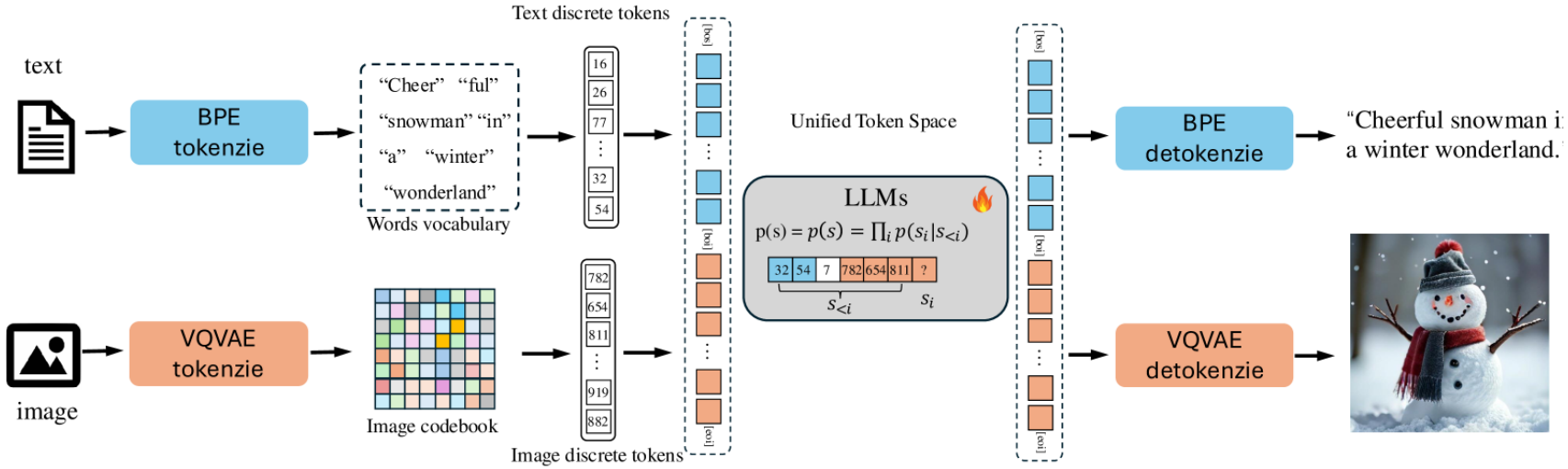
1 PM



7 PM

Challenges & Opportunities

Can we unify 3D reconstruction/SLAM into LLM?



Liquid

[arXiv'24]

Challenges & Opportunities

Interaction with 3D scenes at speed!

World Labs Demo

“People overestimate what they can do in **one year**, and underestimate what they can do in **ten years**.”

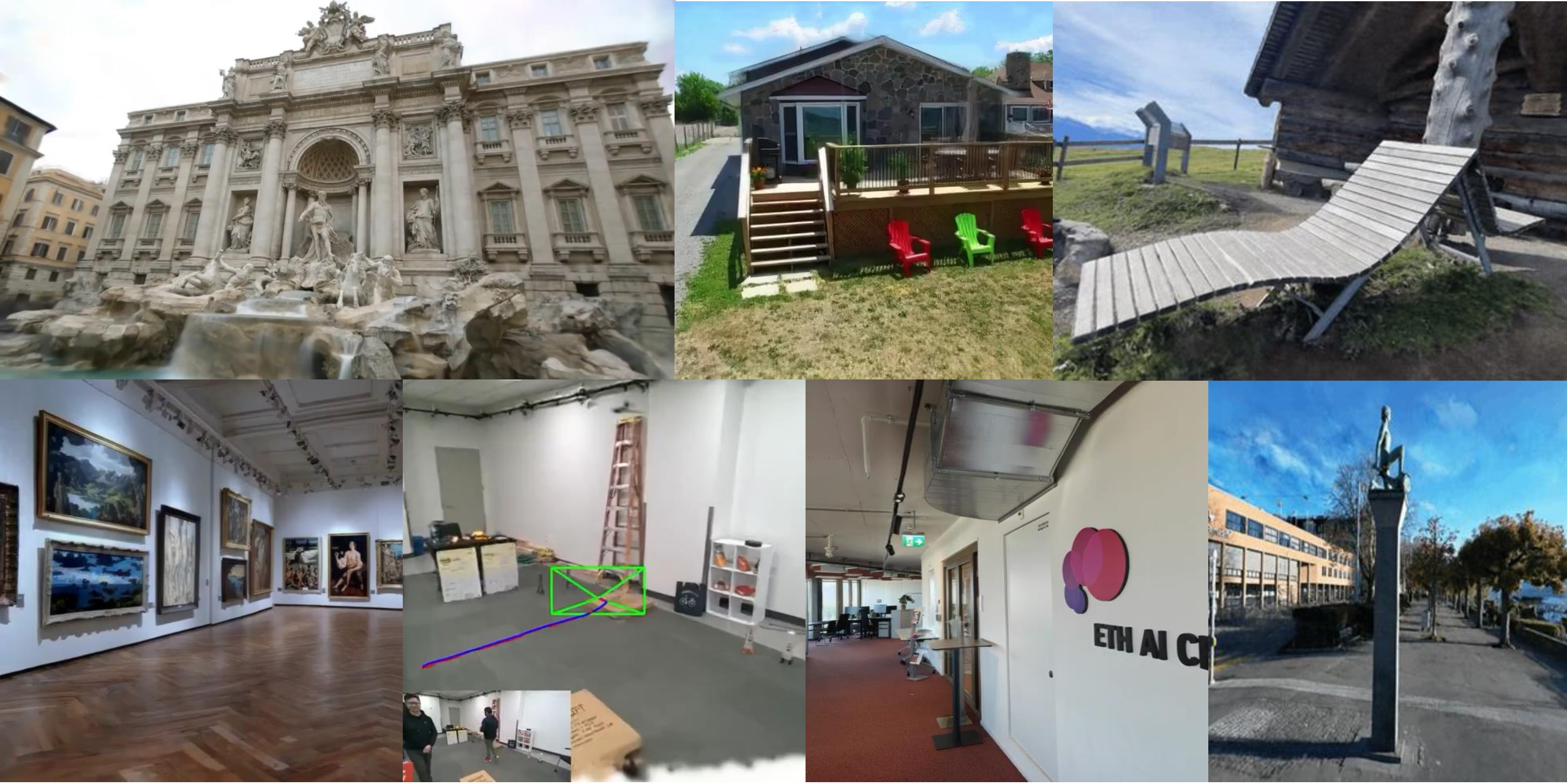
--- *Bill Gates*

“People overestimate what they can do in **one year**, and underestimate what they can do in **ten years**.”

--- *Bill Gates*

“People overestimate what they can do in **one week**, and underestimate what they can do in **one year**.”

--- *Lars Mescheder*



pengsongyou.github.io songyou@google.com @songyoupeng