NOVA3R

From Pixel-Aligned Reconstruction to Non-Pixel-Aligned Worlds

Tracks
Persistent visual evidence

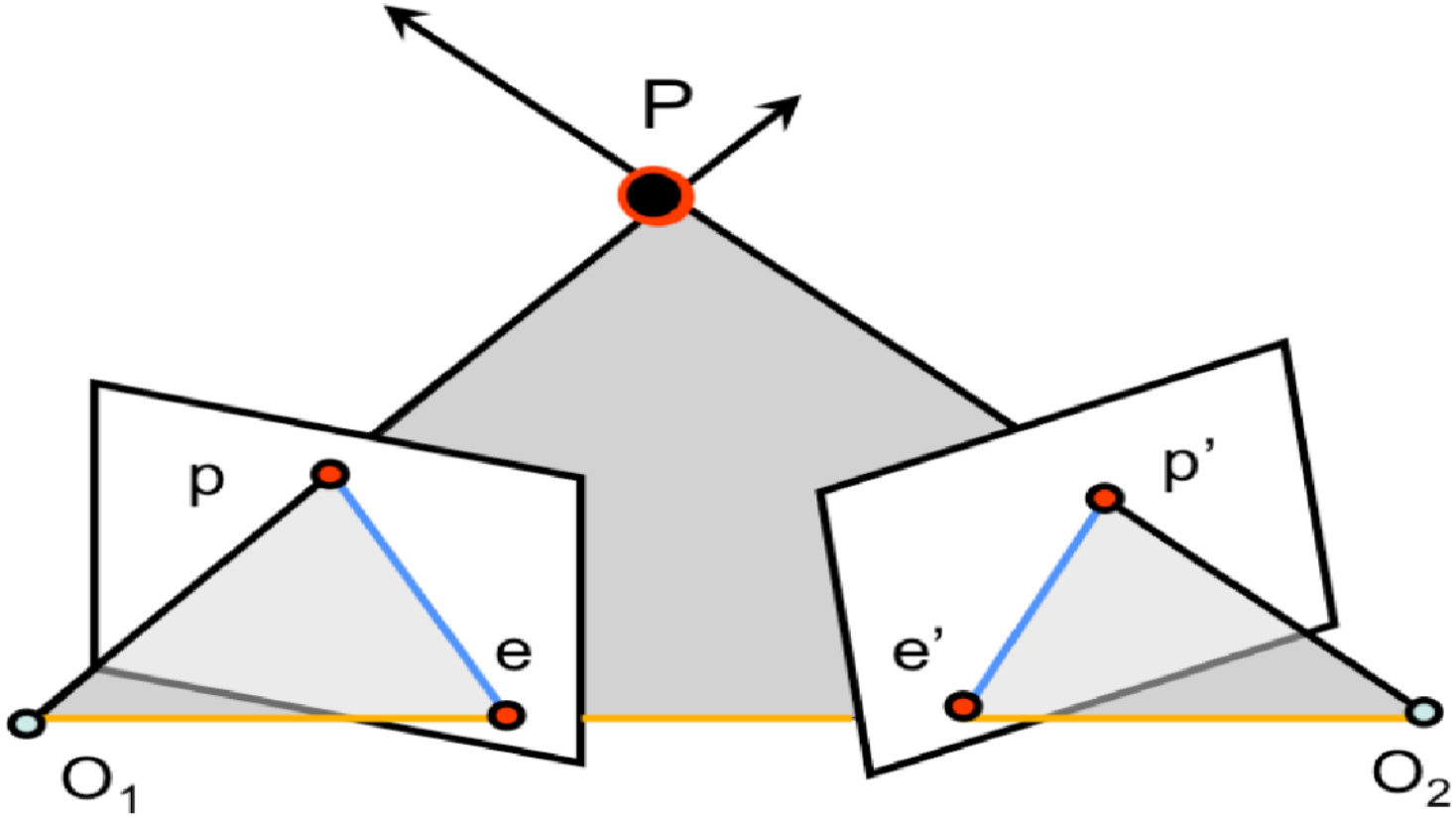
Dynamics
Structured 4D motion

Worlds
Complete 3D representations

Weirong Chen

Computer Vision Group, TUM

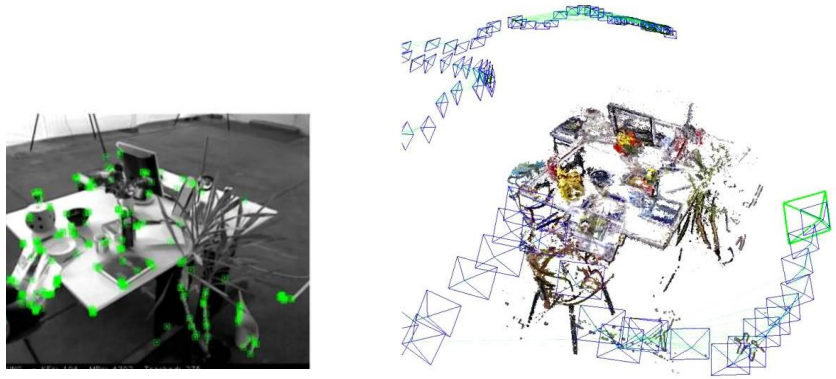
From the lectures: Two-view Geometry (Pixel-aligned)



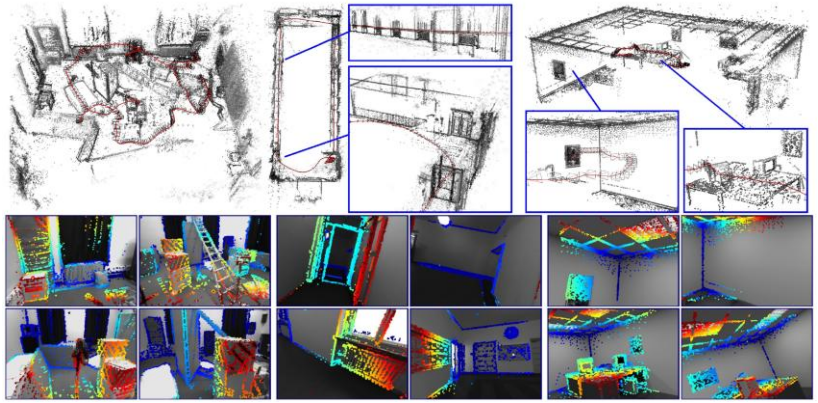
Epipolar Geometry

Structure-from-Motion / SLAM

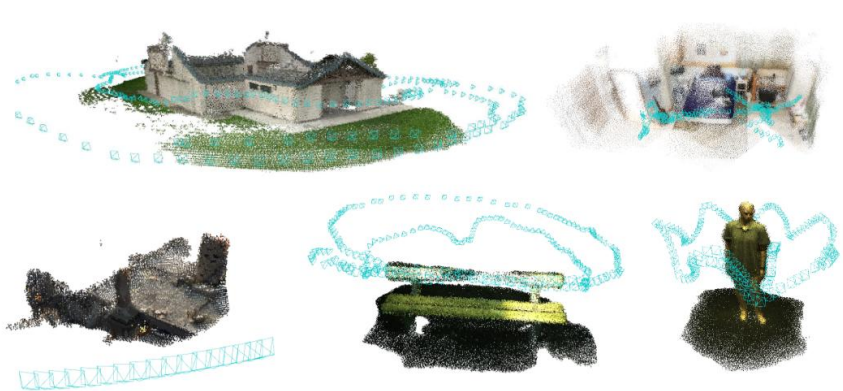
COLMAP [Schönberger et al. 2016]



ORB-SLAM2 [Mur-Artal et al. 2017]



DSO [Engel et al. 2017]



DROID-SLAM [Teed et al. 2021]

At the core: **Epipolar Geometry**

Beyond Two-view Geometry

Pixel-aligned

Non-pixel-aligned

Temporal

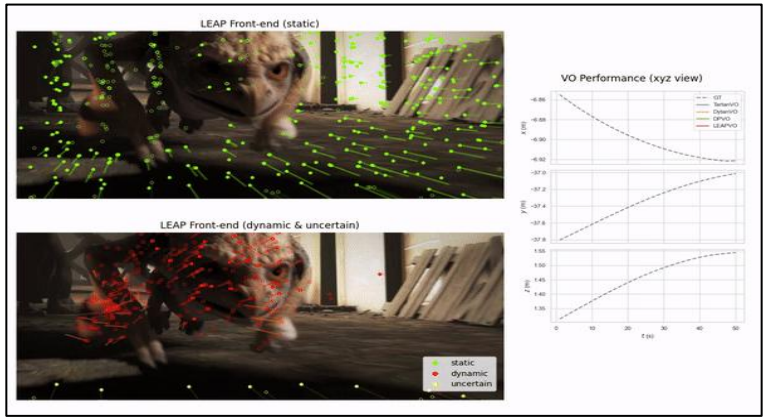
Visual anchors for geometry and motion estimation

Dynamic

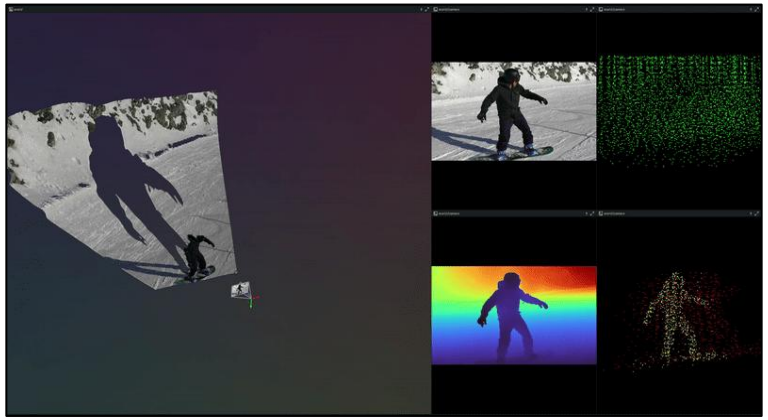
Consistent camera / object motion over time

Complete

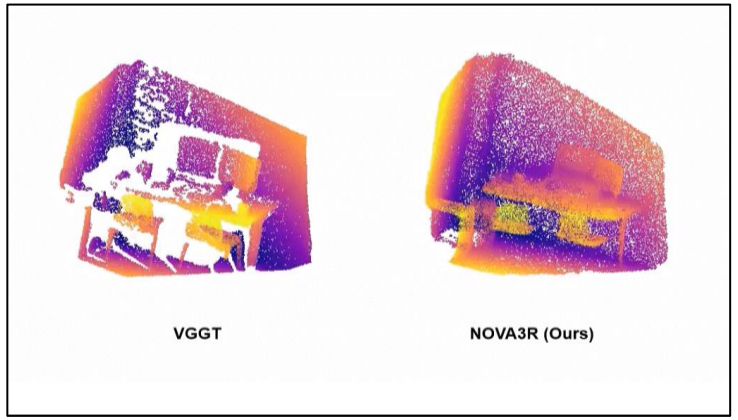
Infer hidden 3D structure beyond direct observations



LEAP-VO
[CVPR 2024]



BA-Track
[ICCV 2025]



NOVA3R
[ICLR 2026]

Pixel-aligned: Learning dynamic motion from point tracks

Observed pixel motion

Camera-induced motion

Static geometry + camera poses

Object-induced motion

Dynamic 4D trajectories

Reconstruct dynamics

Key role

Decouple camera and object motion to recover camera poses, static geometry, and dynamic 4D point trajectories

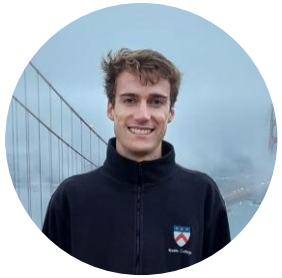
ICCV HONOLULU HAWAII
OCT 19-23, 2025

Back on Track: Bundle Adjustment for Dynamic Scene Reconstruction

Weirong Chen Ganlin Zhang Felix Wimbauer Rui Wang

Nikita Araslanov Andrea Vedaldi Daniel Cremers







Munich Center for Machine Learning

UNIVERSITY OF OXFORD

Microsoft

Dynamic Scene Reconstruction

Input
Casual Video

Output
Camera Poses + Dynamic Reconstruction

Can we reconstruct both ***static and dynamic*** geometry in an ***efficient*** way?

Why Motion Decoupling ?

(1) Camera Movement

(2) Dynamic Object Motion

Why Motion Decoupling ?

Static Pixel Motion

Observed Motion = **Camera-induced Motion**

✅ Static Background:

Epipolar geometry holds → Bundle adjustment works 👍

Why Motion Decoupling ?

Dynamic Pixel Motion

Observed Motion = **Camera-induced Motion** + **Object-induced Motion**

❌ Dynamic Objects:

Epipolar geometry breaks → Bundle adjustment fails 😢

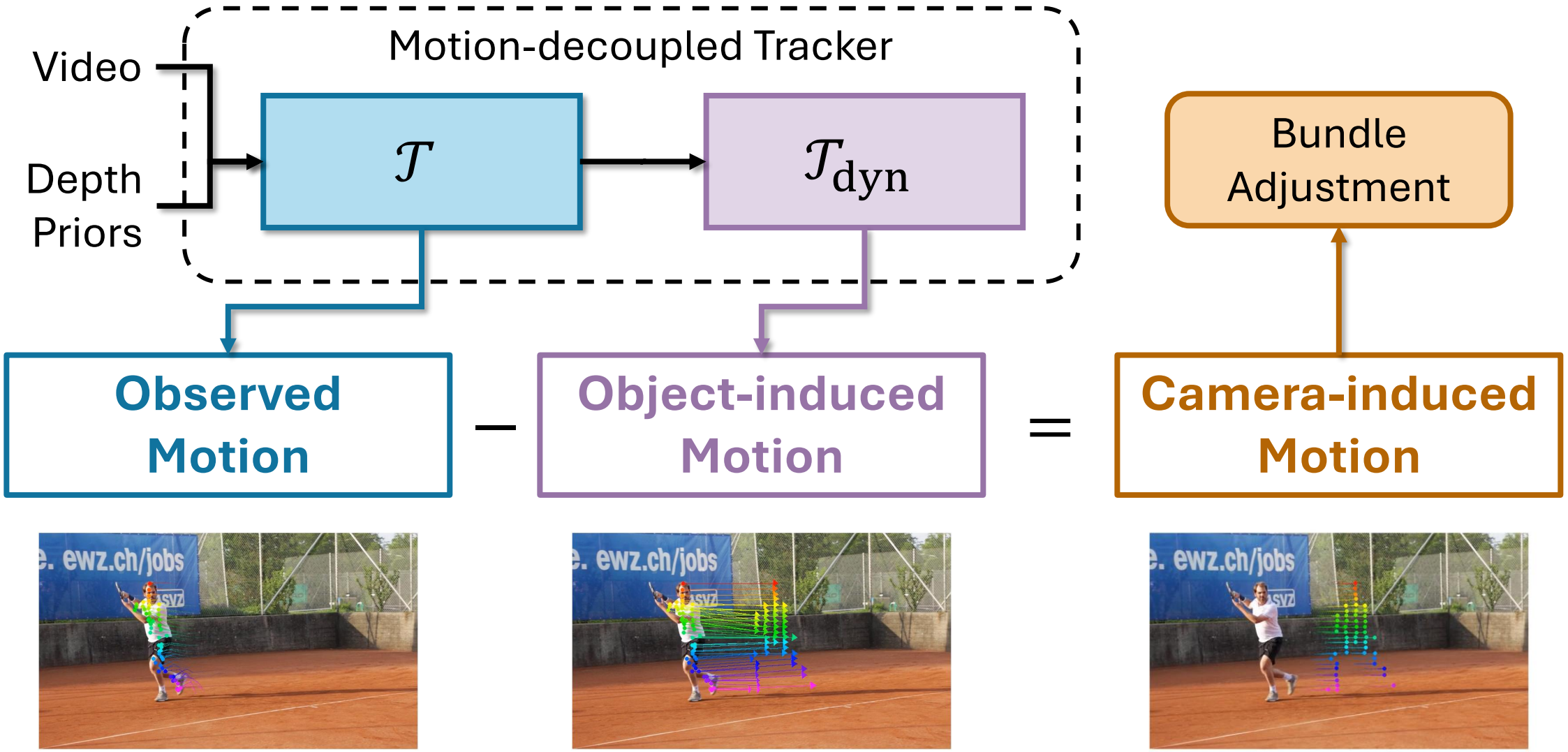
Why Motion Decoupling ?

Observed Motion − **Object-induced Motion** = **Camera-induced Motion**

💡 Our Motivation

Use *camera-induced component* for *bundle adjustment*

Motion Decoupling with Point Tracking



Motion Decoupling with Point Tracking

Observed Motion

Camera-induced Motion

(decoupling in local window)

BA-Track Pipeline

Input Video

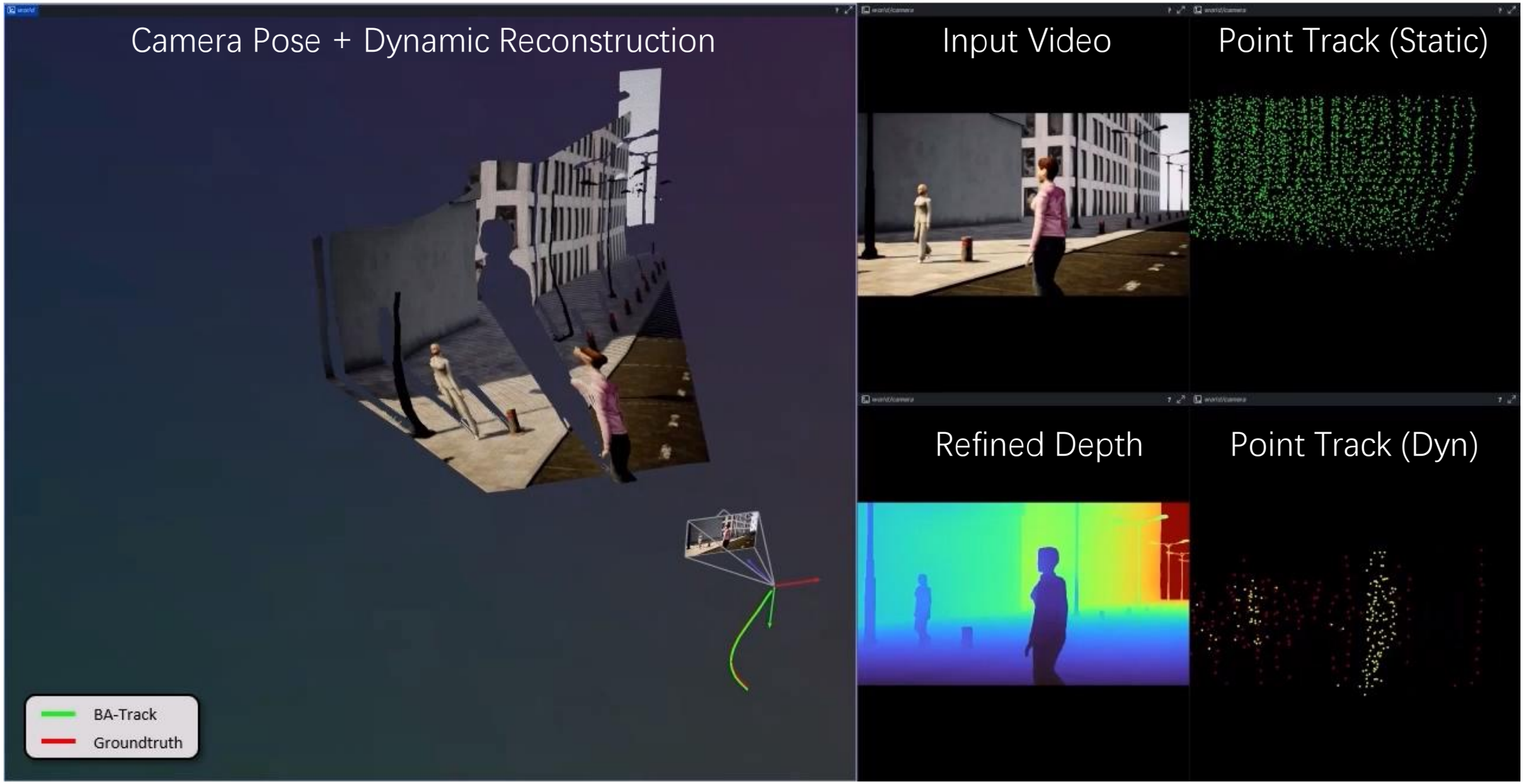
Depth Priors

I. Motion-decoupled Tracker

II. Bundle Adjustment

III. Global Refinement

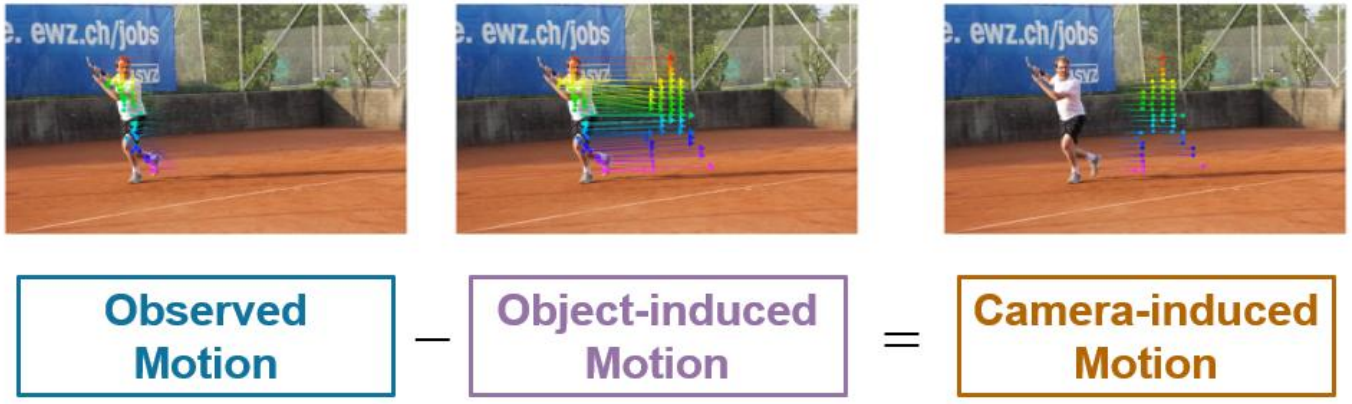
Qualitative Results



BA-Track: From Discarding to Exploiting Motion

- **Dual-network** captures 3D motion priors for effective motion decoupling
- **Motion decoupling** restores **epipolar constraint** for dynamic points

BA-Track brings bundle adjustment
back to dynamic scenes



Non-pixel-aligned: Global decoding with 3D generation

Reconstruction

Explains observed geometry and motion

3D-native representation

Global, view-agnostic scene state

Generation

Infers missing / occluded 3D structure

Key role

Reconstruction grounds the model; generation completes the world

NOVA3R: Non-pixel-aligned Visual Transformer for Amodal 3D Reconstruction

Weirong Chen Chuanxia Zheng Ganlin Zhang

Andrea Vedaldi Daniel Cremers





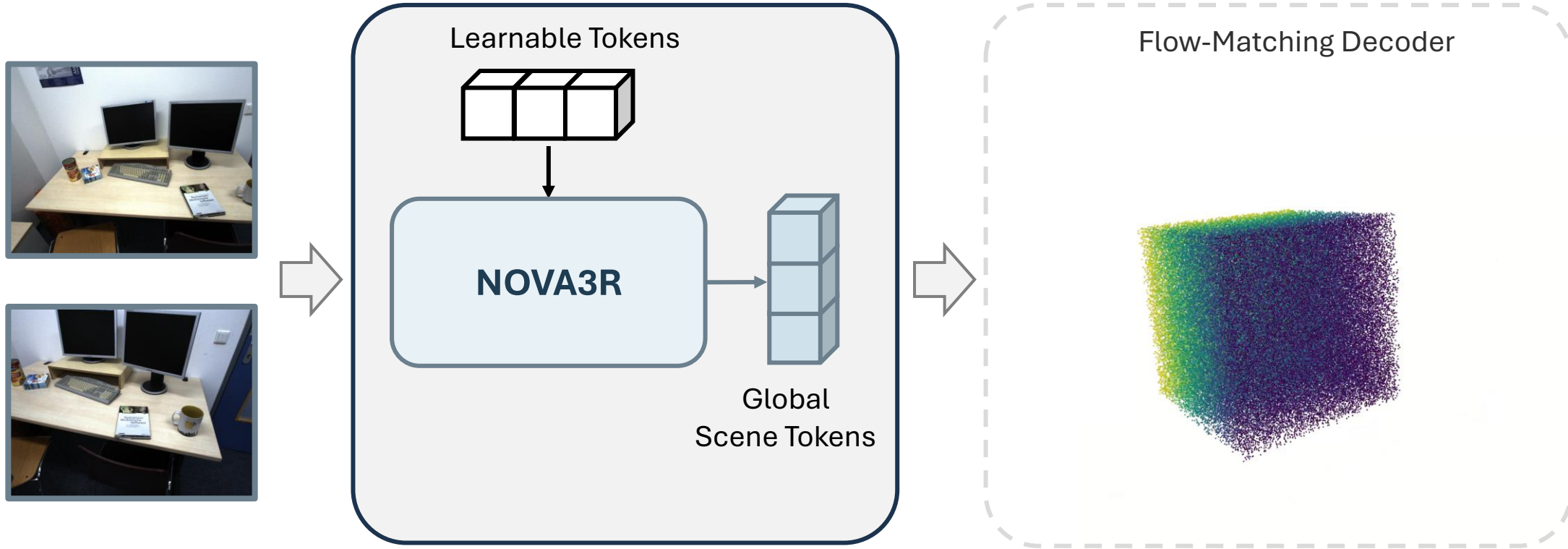
Munich Center for Machine Learning

UNIVERSITY OF OXFORD

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

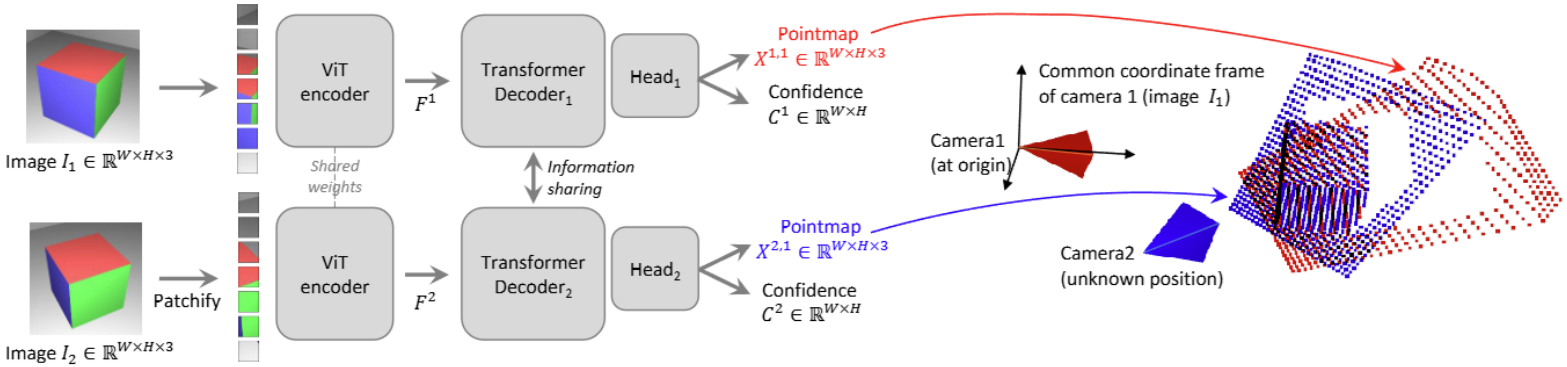
Task: Non-pixel-aligned Reconstruction

Goal: Complete, Non-overlapping 3D Reconstruction from Unposed Images

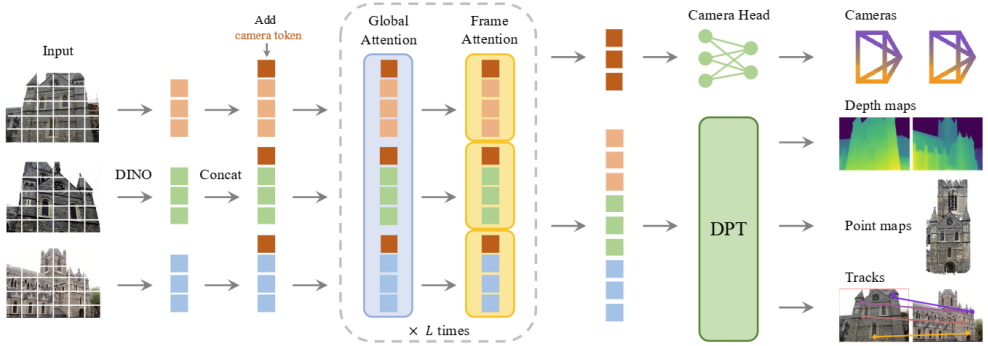


Multi-view Images **Non-pixel-aligned Representation** **Global 3D Reconstruction**

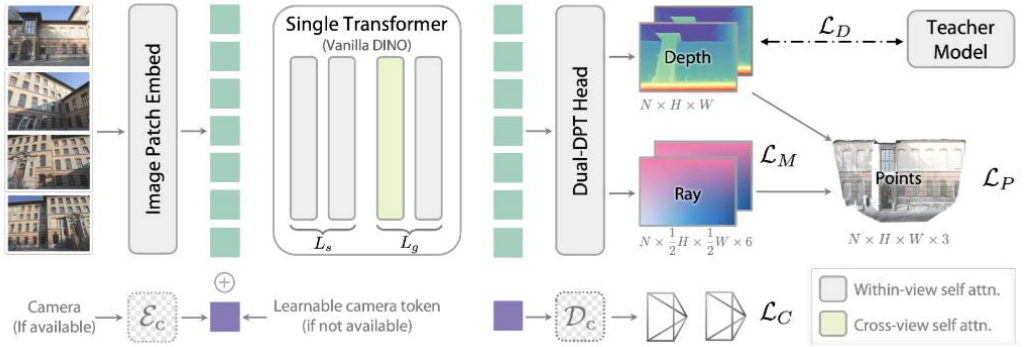
Motivation



Dust3r
[Wang et al. CVPR 2024]



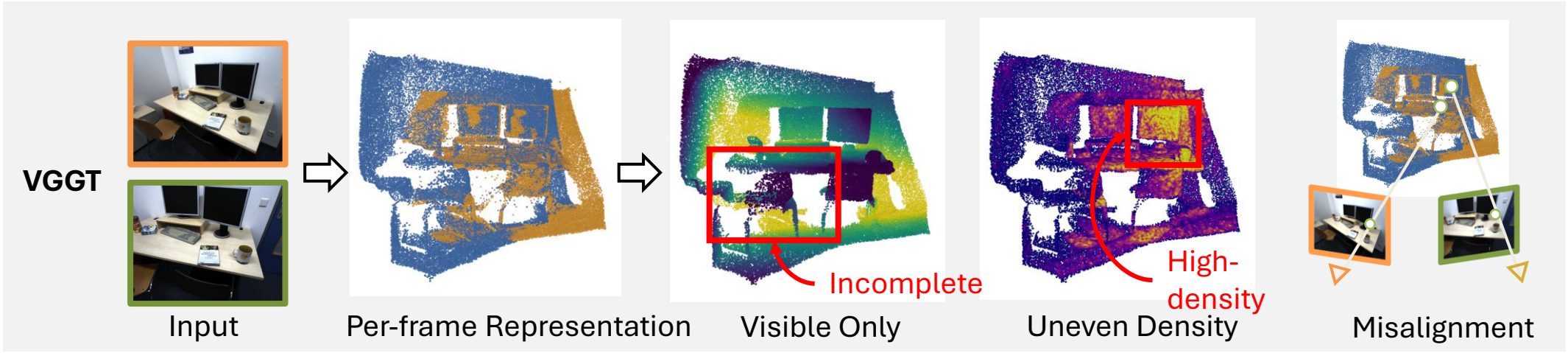
VGGT
[Wang et al. CVPR 2025]



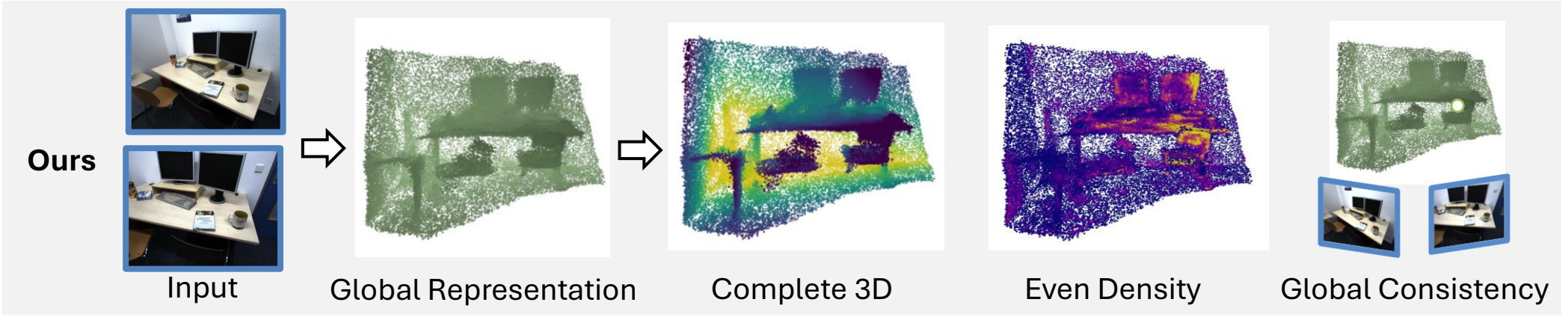
Depth Anything 3
[Lin et al. ICLR 2026]

Motivation

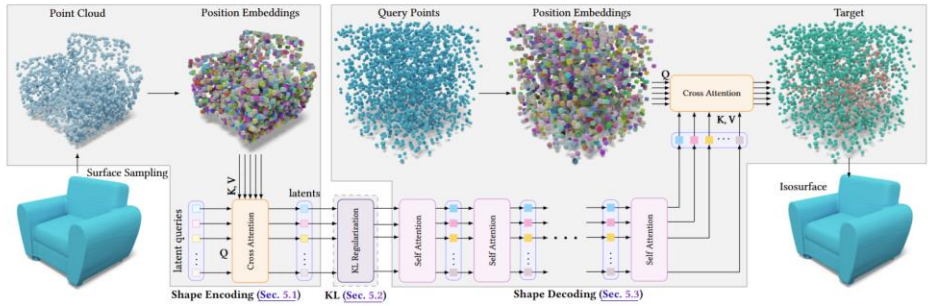
Problem: Pixel-aligned methods often recover **per-frame visible geometry**, leading to **incompleteness**, **redundancy**, and **misalignment**.



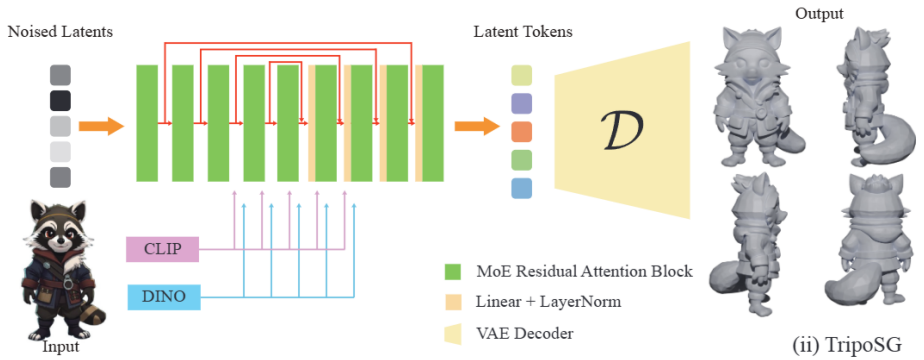
💡 **Insight**: Recover the **underlying scene**, not just visible pixels



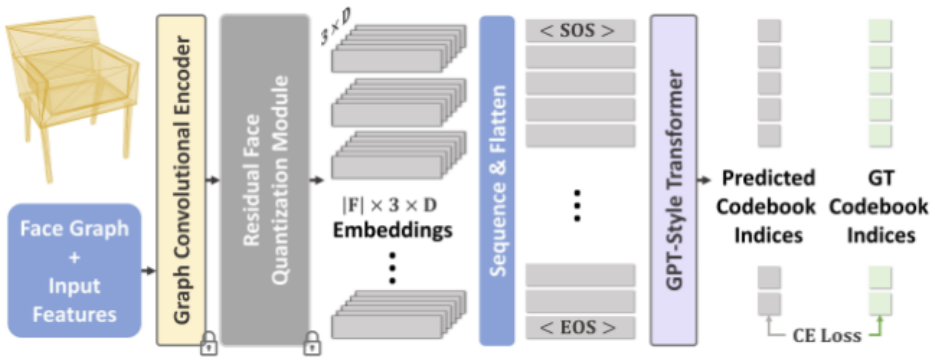
Object-level 3D Generation



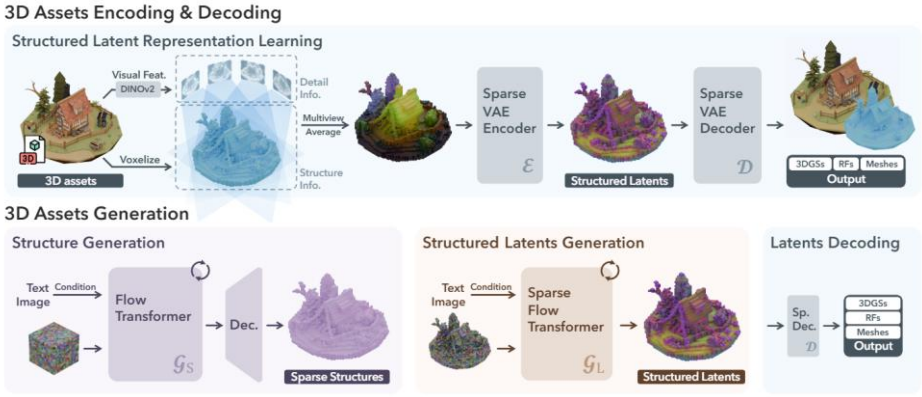
3DShape2VecSet
[Zhang et al. SIGGRAPH 2023]



TripoSG
[Li et al. T-PAMI 2025]



MeshGPT
[Siddiqui et al. CVPR 2024]

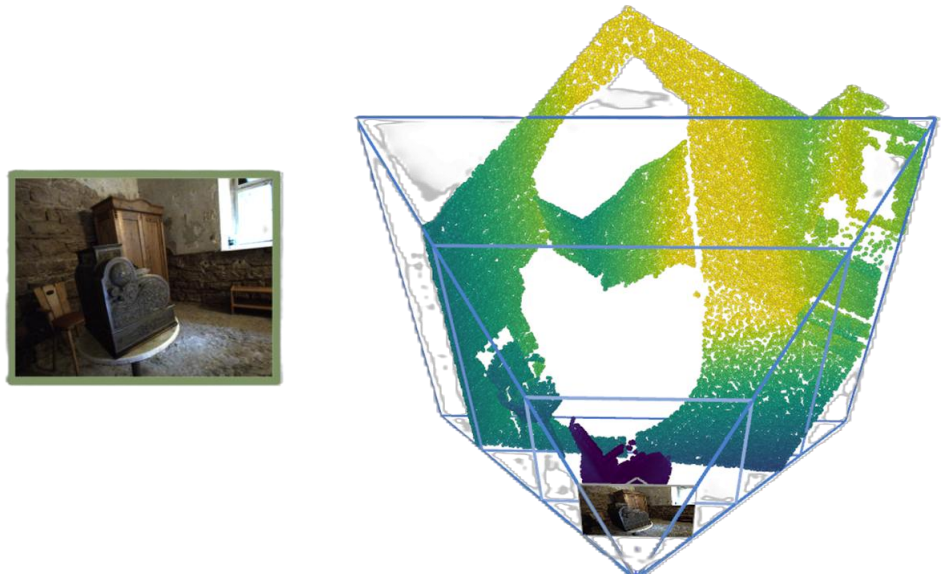


TRELLIS
[Xiang et al. CVPR 2025]

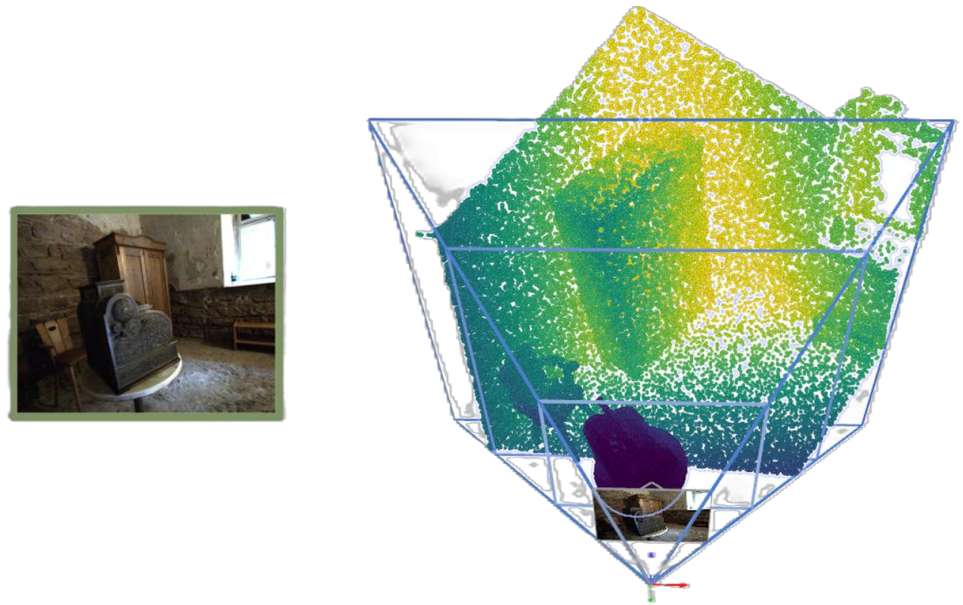
Can we bridge feed-forward reconstruction with non-pixel-aligned 3D generation?

Scene-level Completeness

Recover complete geometry within the input view frustum

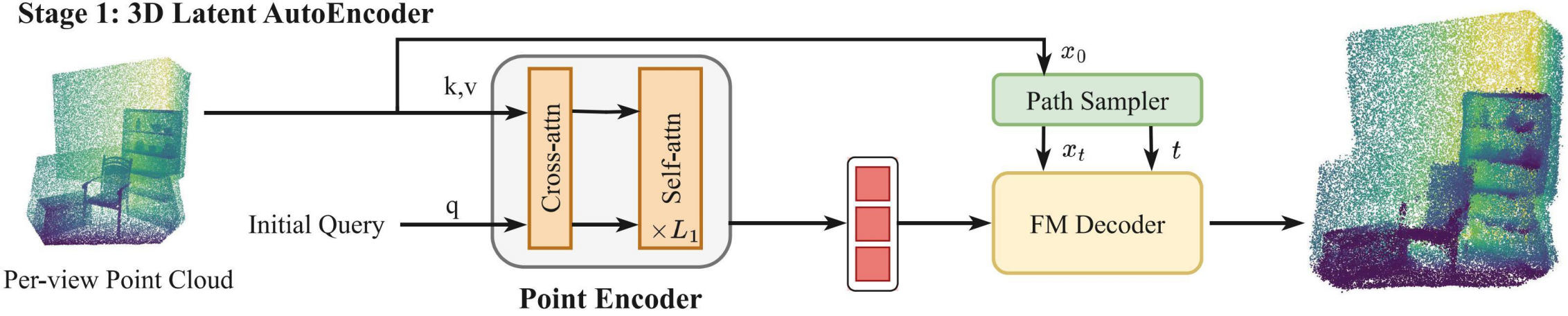


Visible-only Geometry
(Previous Methods)

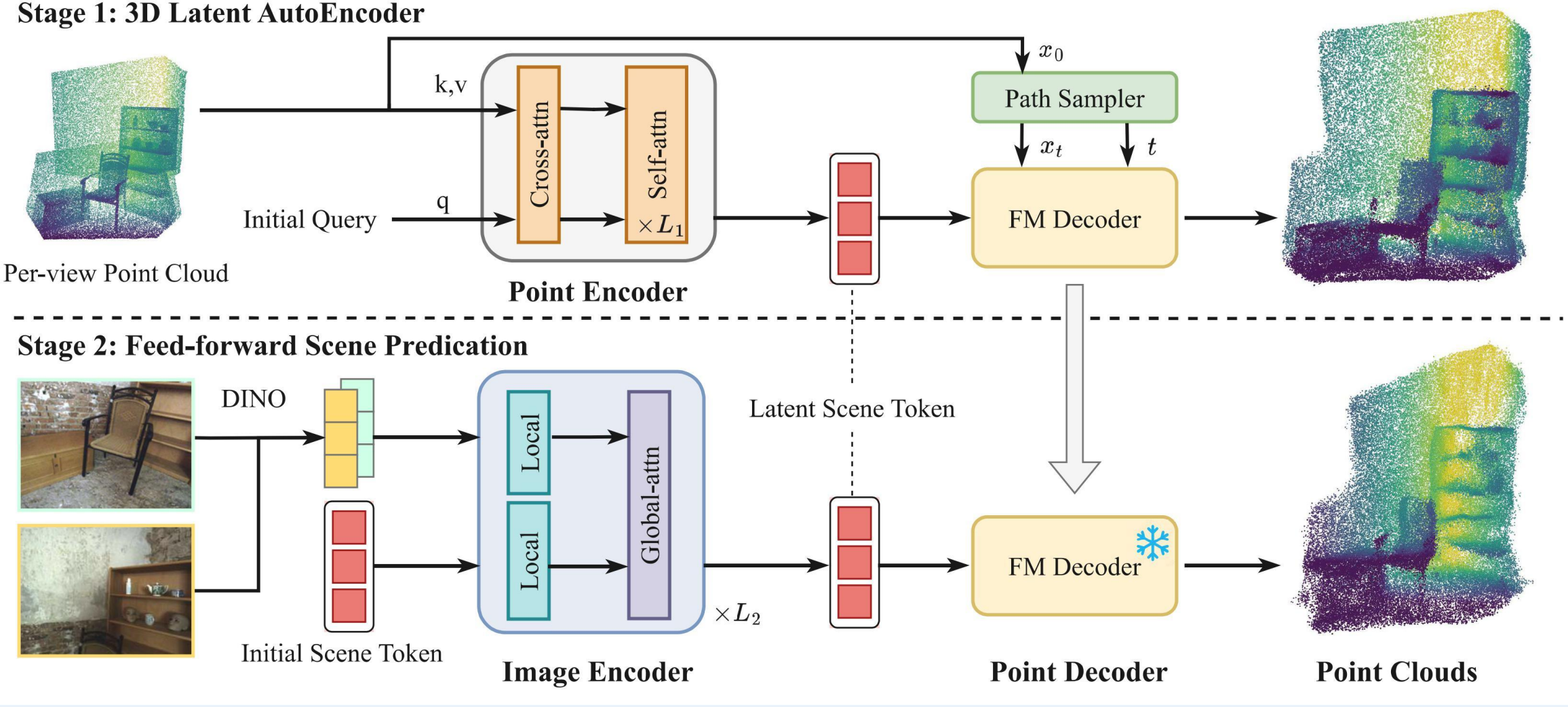


Complete Geometry
(Ours)

Scene-level Completeness

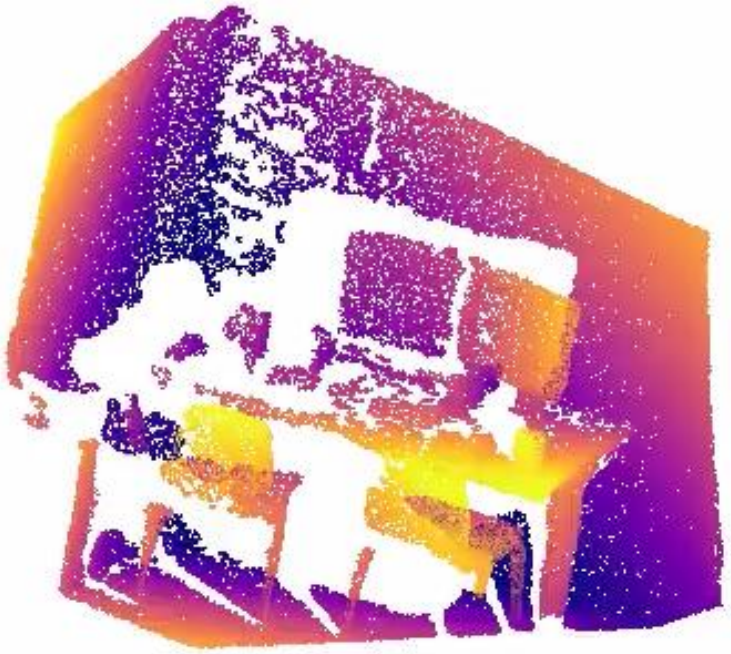


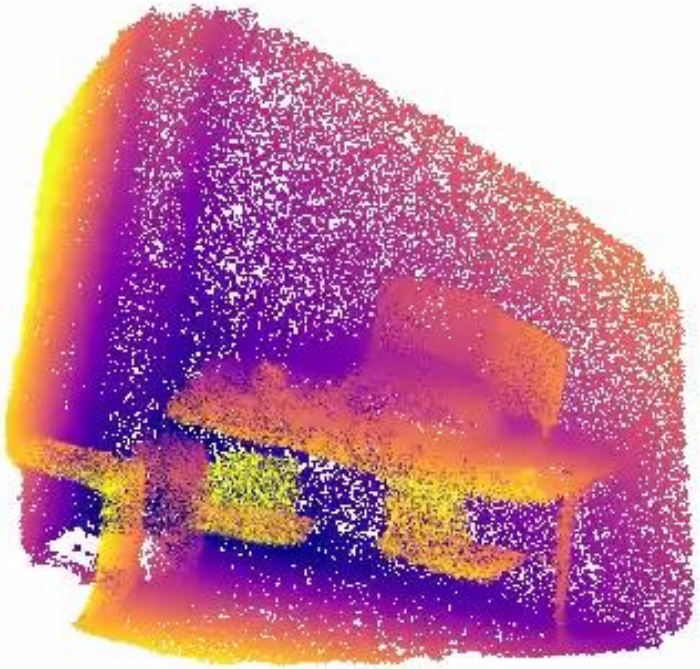
Scene-level Completeness



Results

NOVA3R reconstructs both **visible** and **occluded** geometry within the view frustum

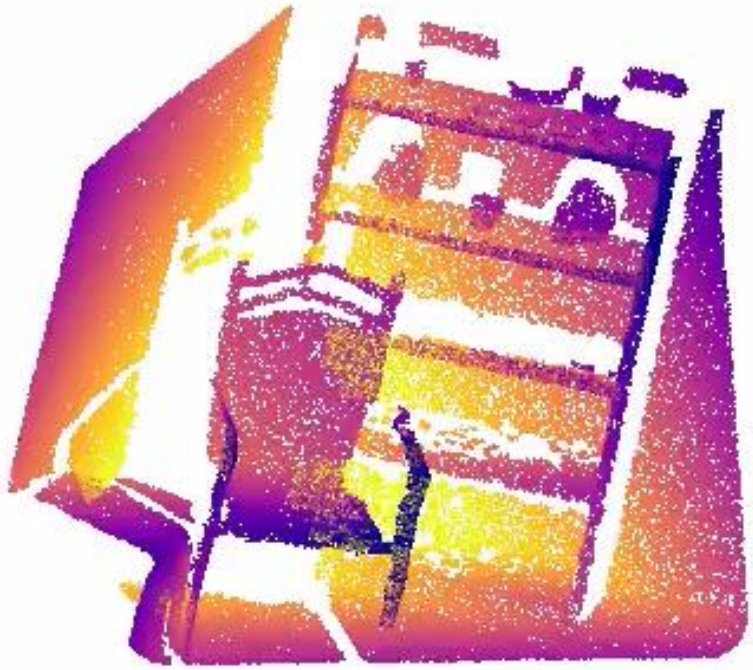


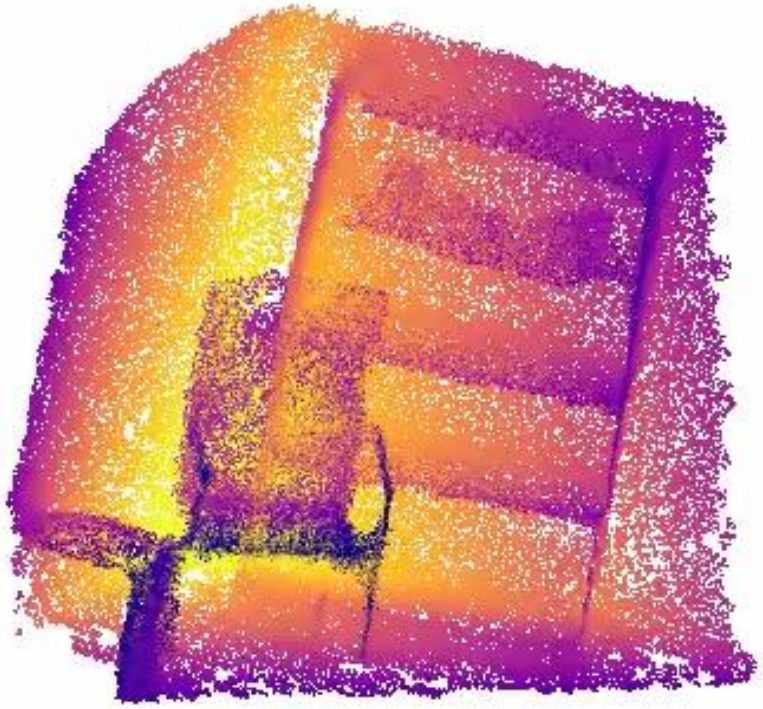


Input **VGGT** **NOVA3R (Ours)**

Results

NOVA3R reconstructs both **visible** and **occluded** geometry within the view frustum

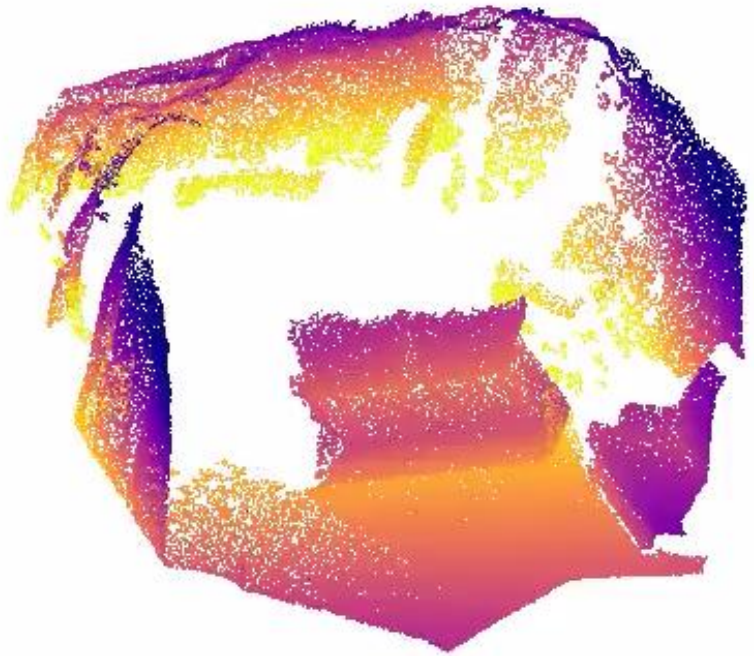


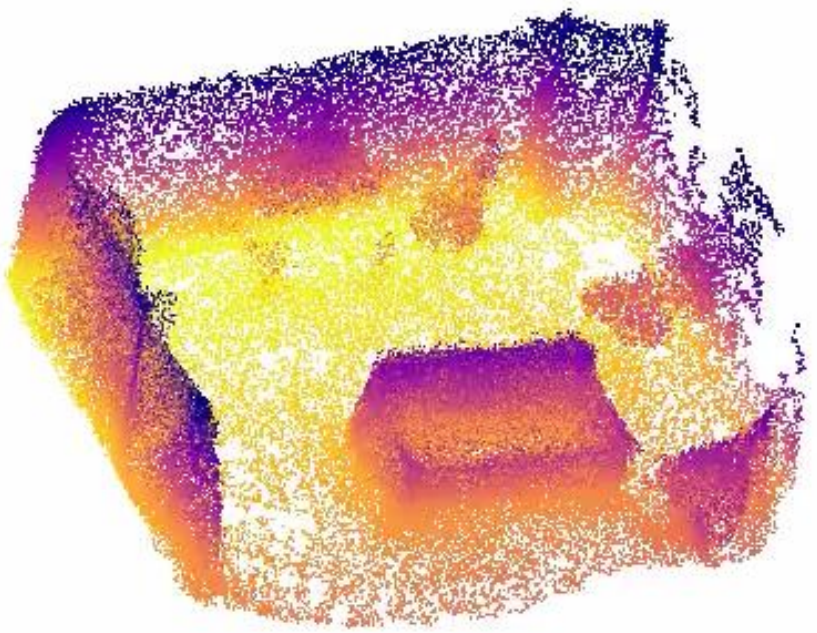


Input **VGGT** **NOVA3R (Ours)**

Results

NOVA3R reconstructs both **visible** and **occluded** geometry within the view frustum

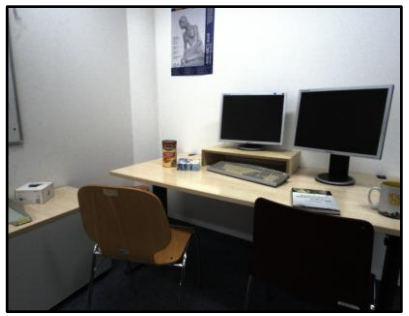




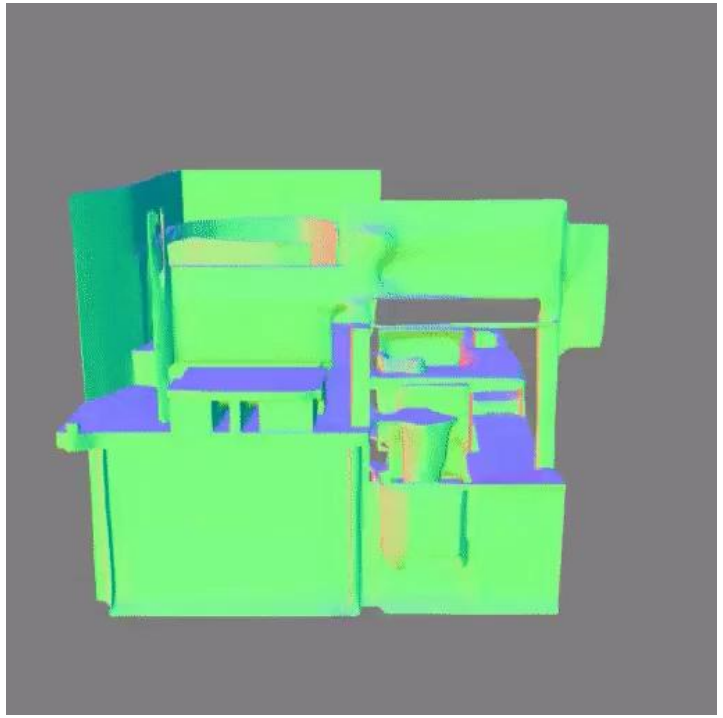
Input **VGGT** **NOVA3R (Ours)**

Results

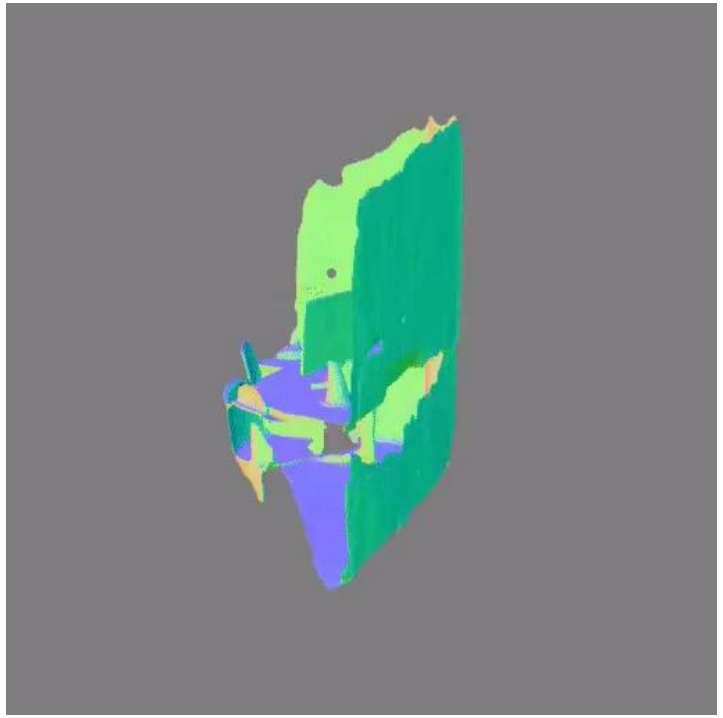
NOVA3R bridges 3D object generation to **real-world scene** with **faithful synthesis**



Input



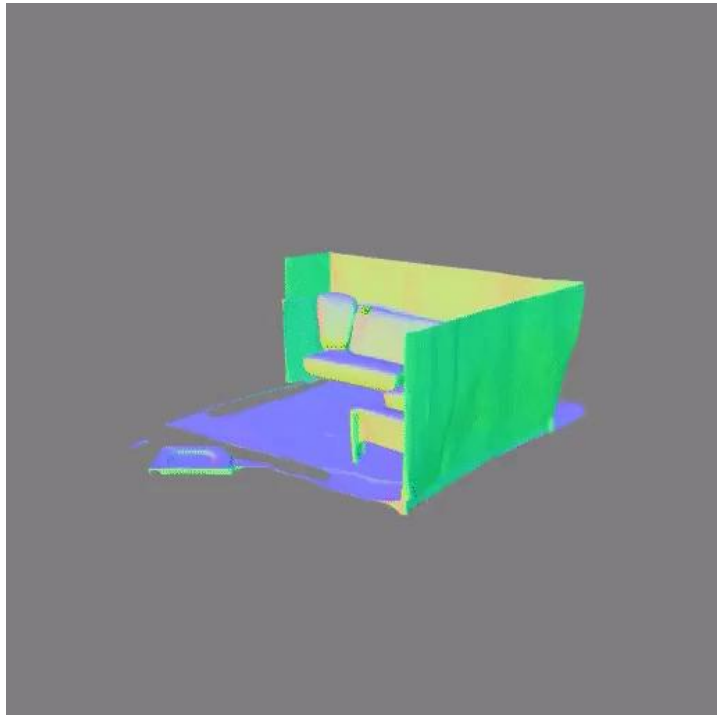
TRELLIS

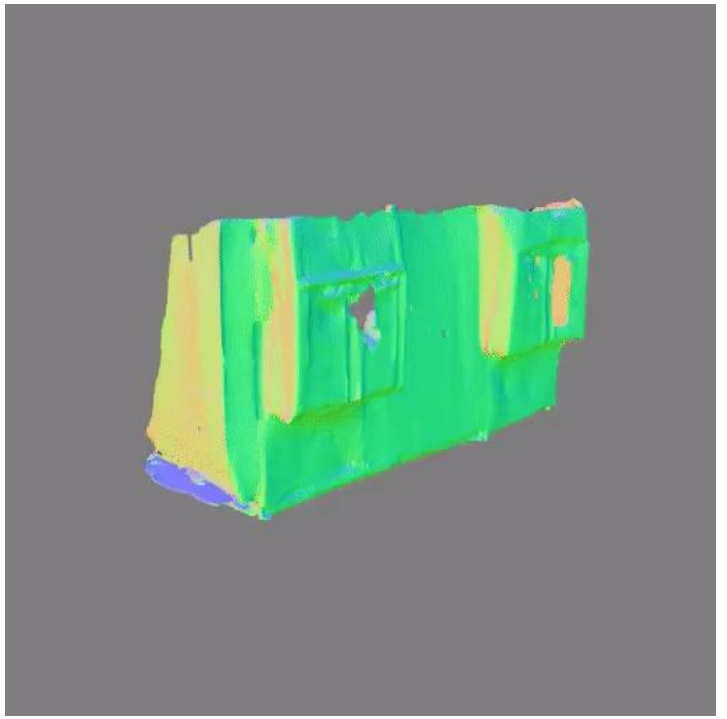


NOVA3R + TRELLIS

Results

NOVA3R bridges 3D object generation to **real-world scene** with **faithful synthesis**





Input

TRELLIS

NOVA3R + TRELLIS

Results

NOVA3R bridges 3D object generation to **real-world scene** with **faithful synthesis**

Input

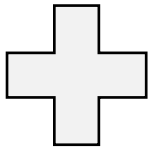
TRELLIS

NOVA3R + TRELLIS

NOVA3R: Bridging reconstruction and generation

Feed-forward Encoder

Efficient multi-view encoding



Generative Decoder

Complete 3D generation

- **Unified** framework for object- and scene-level complete 3D reconstruction
- More **complete**, **uniform**, and **physically plausible** geometry
- Bridging **feed-forward** reconstruction and **latent 3D generation**