EFM3D: A Benchmark for Measuring Progress Towards 3D Egocentric Foundation Models

JULIAN STRAUB

Daniel DeTone

Tianwei Shen

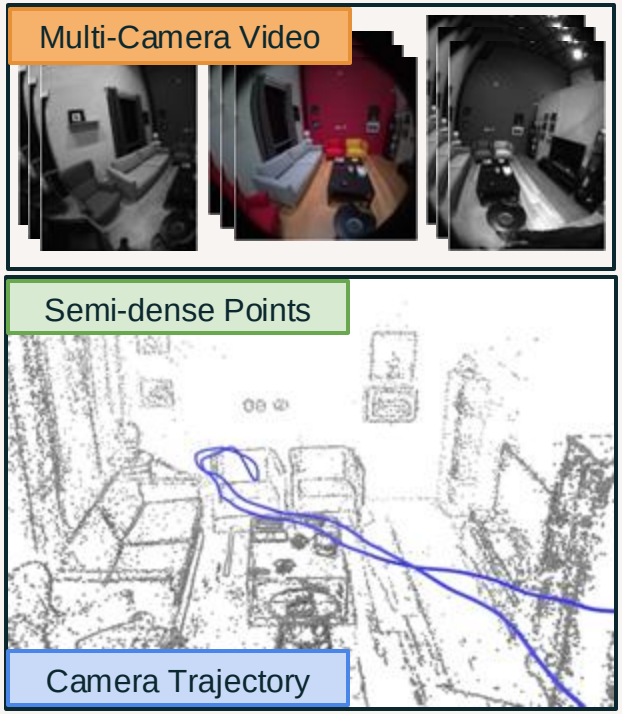
Nan Yang

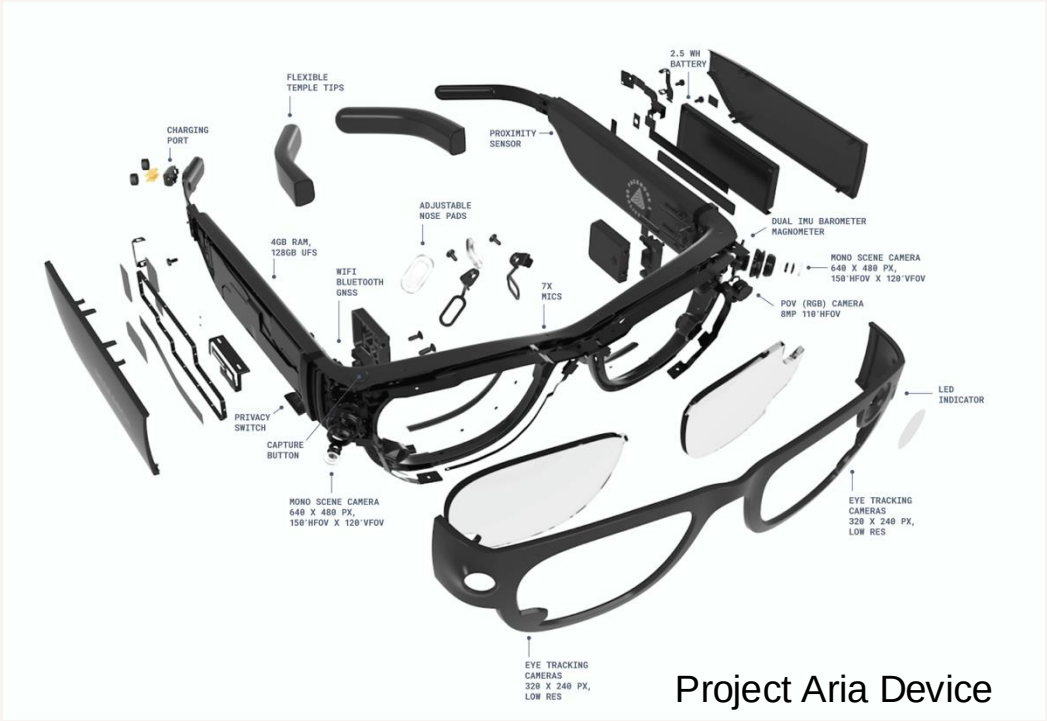
Chris Sweeney

Richard Newcombe

Reality Labs Research (RL-R)

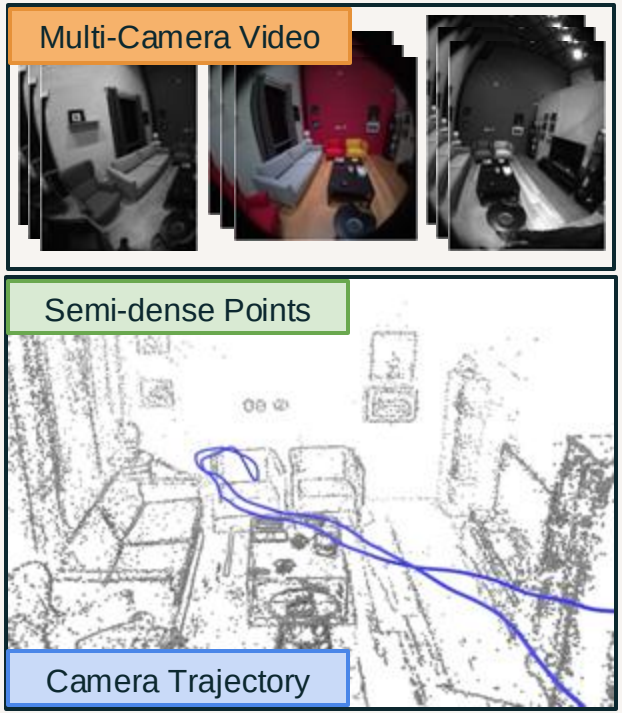
Egocentric Data is a New Category of Data





Egocentric Data

Egocentric Data is a New Category of Data

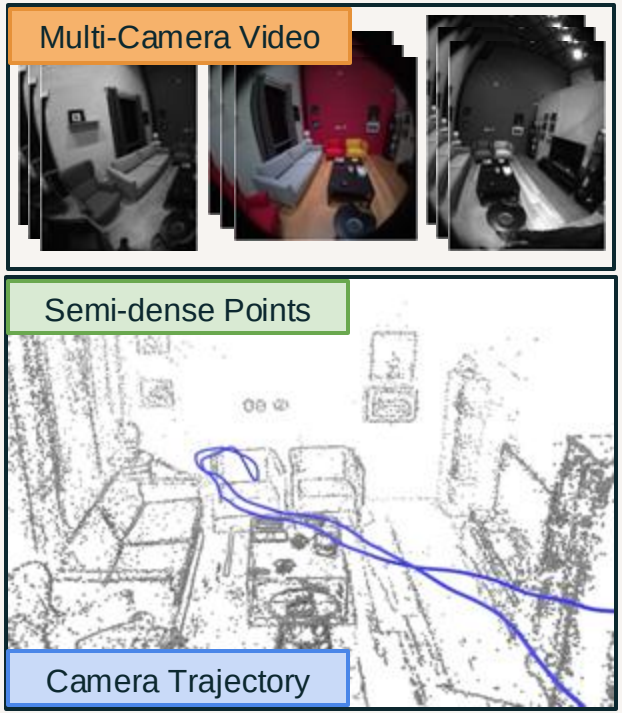


Egocentric Data

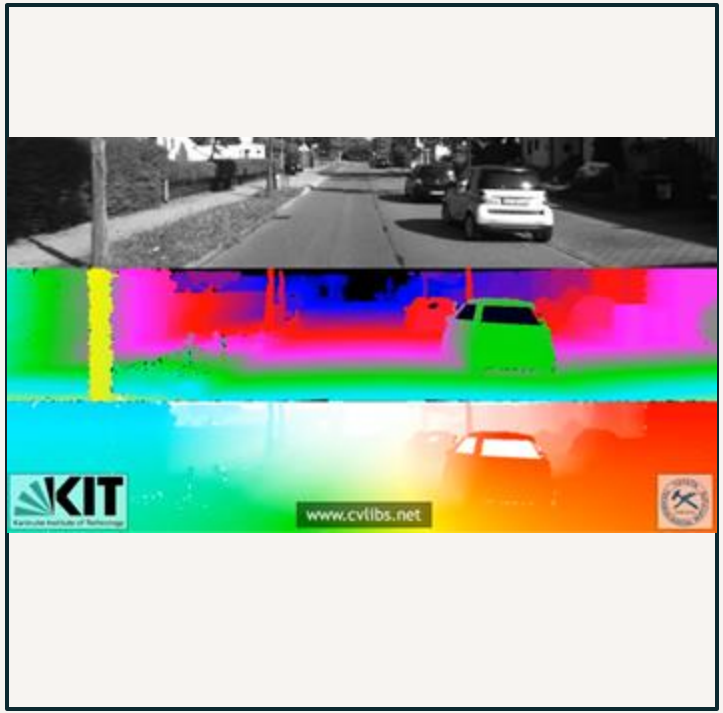
Key Properties of Egocentric data:

- Always-On
- Head-worn Natural Human Motion
- Partial Observations
- No Dense Depth
- Dynamic

Egocentric Data is a New Category of Data



Egocentric Data

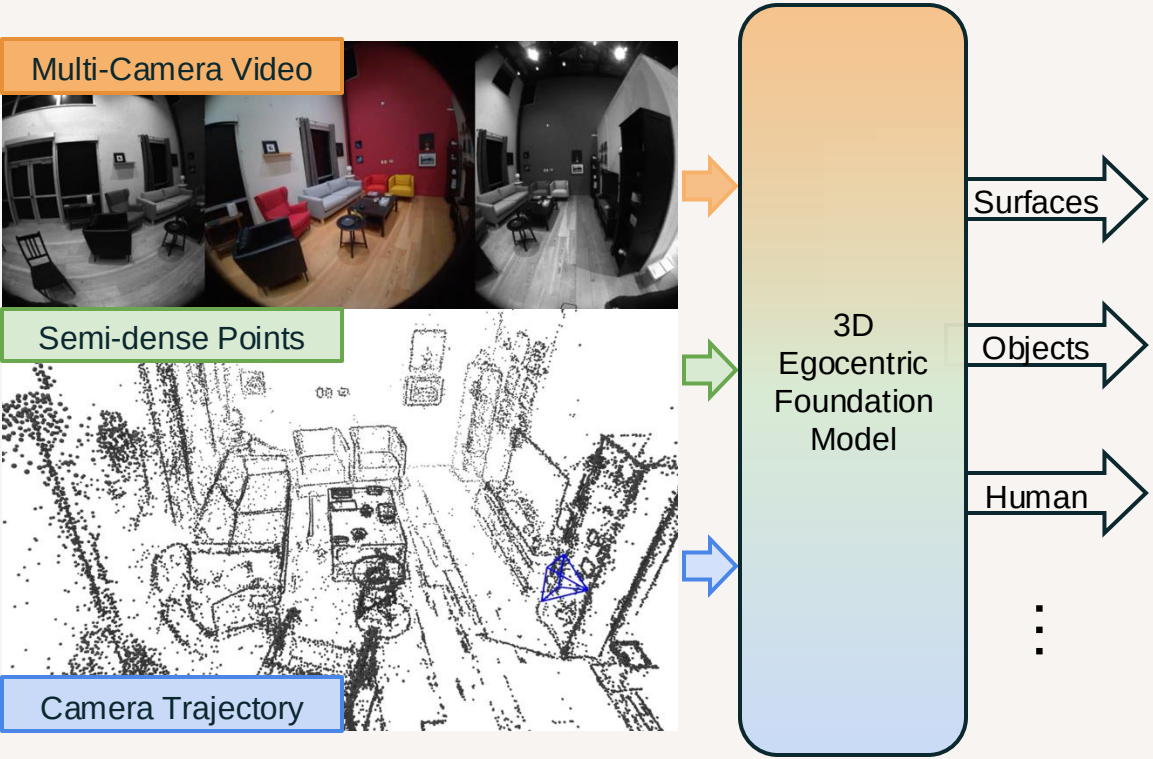


Autonomous Car Data

RGBD Indoor Scanning Data

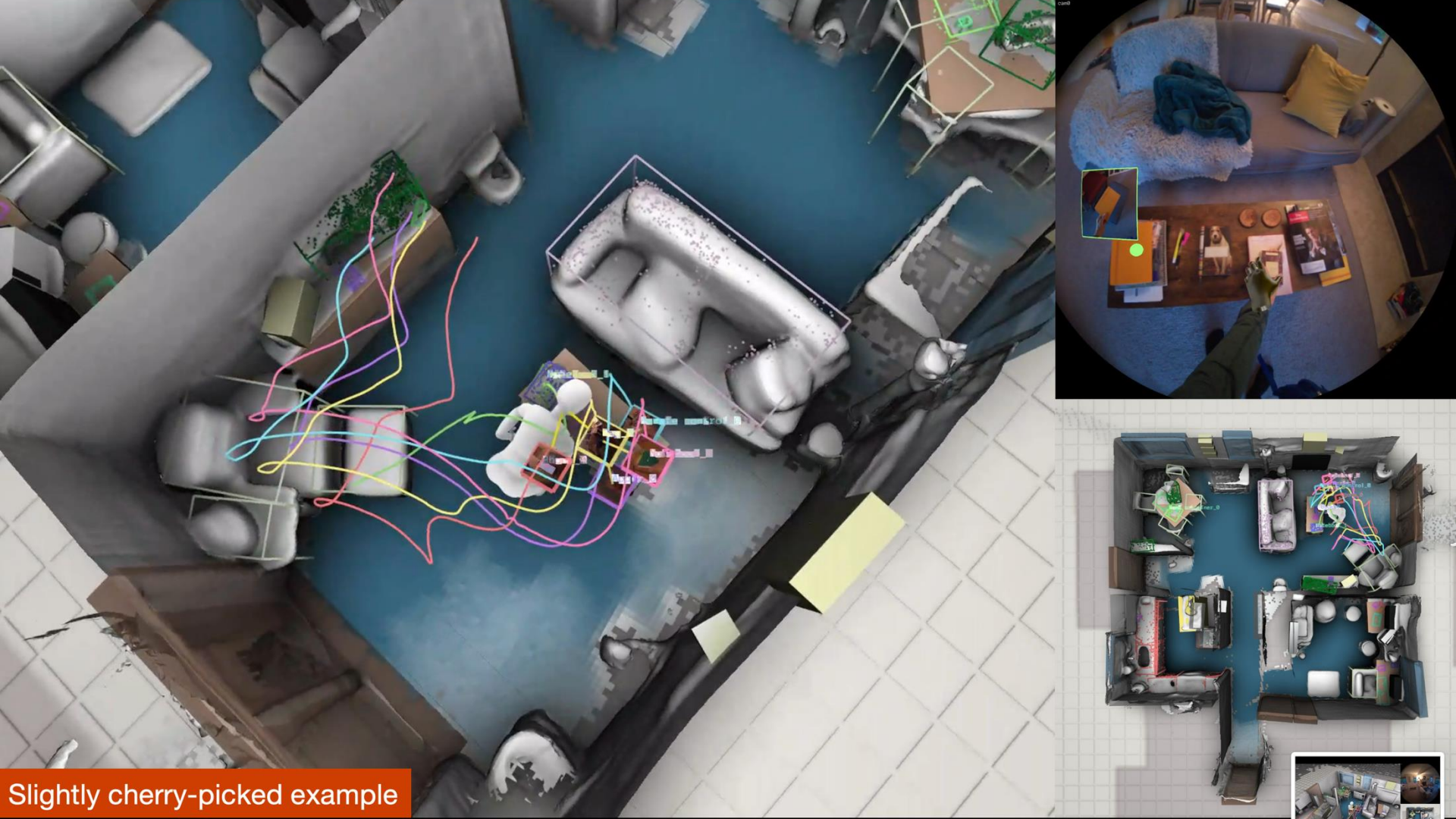
Egocentric Spatial AI is like shrinking autonomous cars 20x, their AI compute power by 1000x and flying them around all day long in 3D everywhere humans go, indoors and out, not just 2D roads.

3D Egocentric Foundation Models

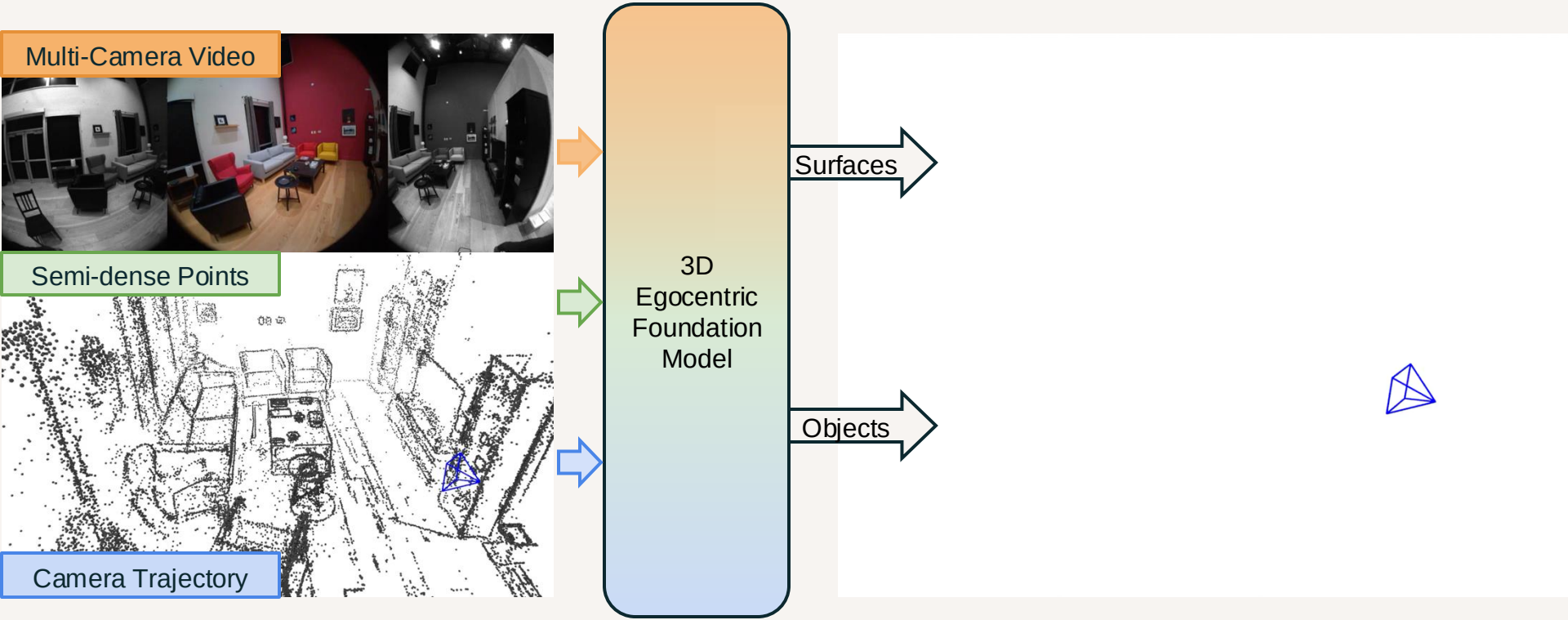


Key Properties

- Strong Learned Priors
- Scalable Incremental Inference
- Persistent Representation


Slightly cherry-picked example

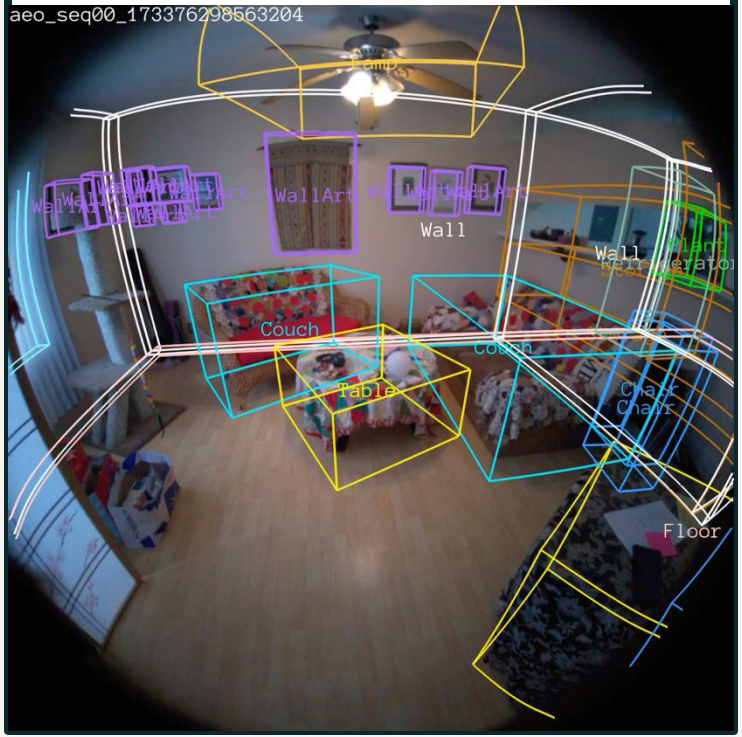
3D Egocentric Foundation Models



EFM3D Benchmark

Egocentric 3D Reconstruction

Egocentric 3D Object Detection



Aria Simulated Environments (ASE)

Training and Benchmarking Egocentric 3D Object Detection and Surface Estimation

10k scenes, 160h Project Aria recordings, 580k 3D OBBs, 29 classes

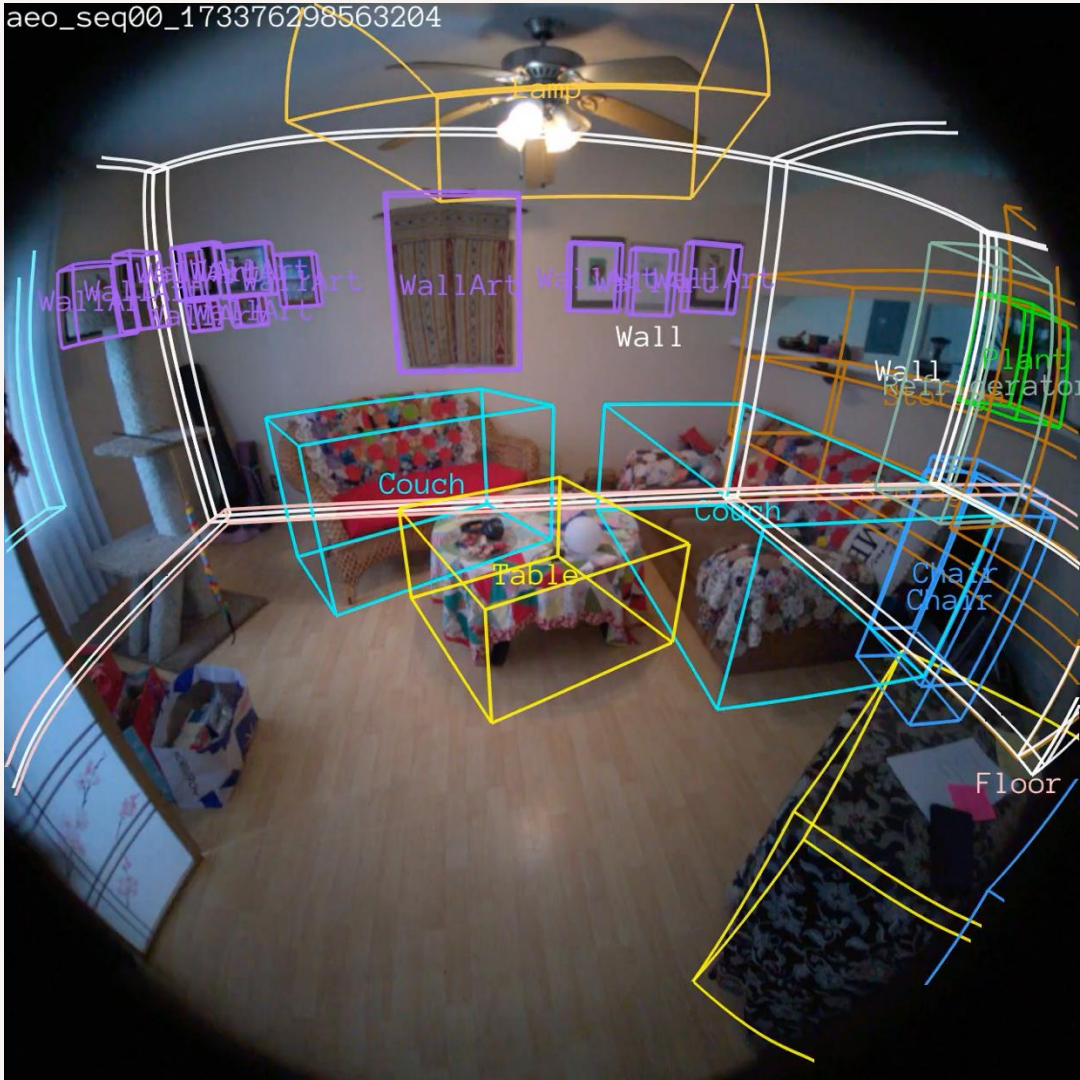
https://www.projectaria.com/datasets/ase/

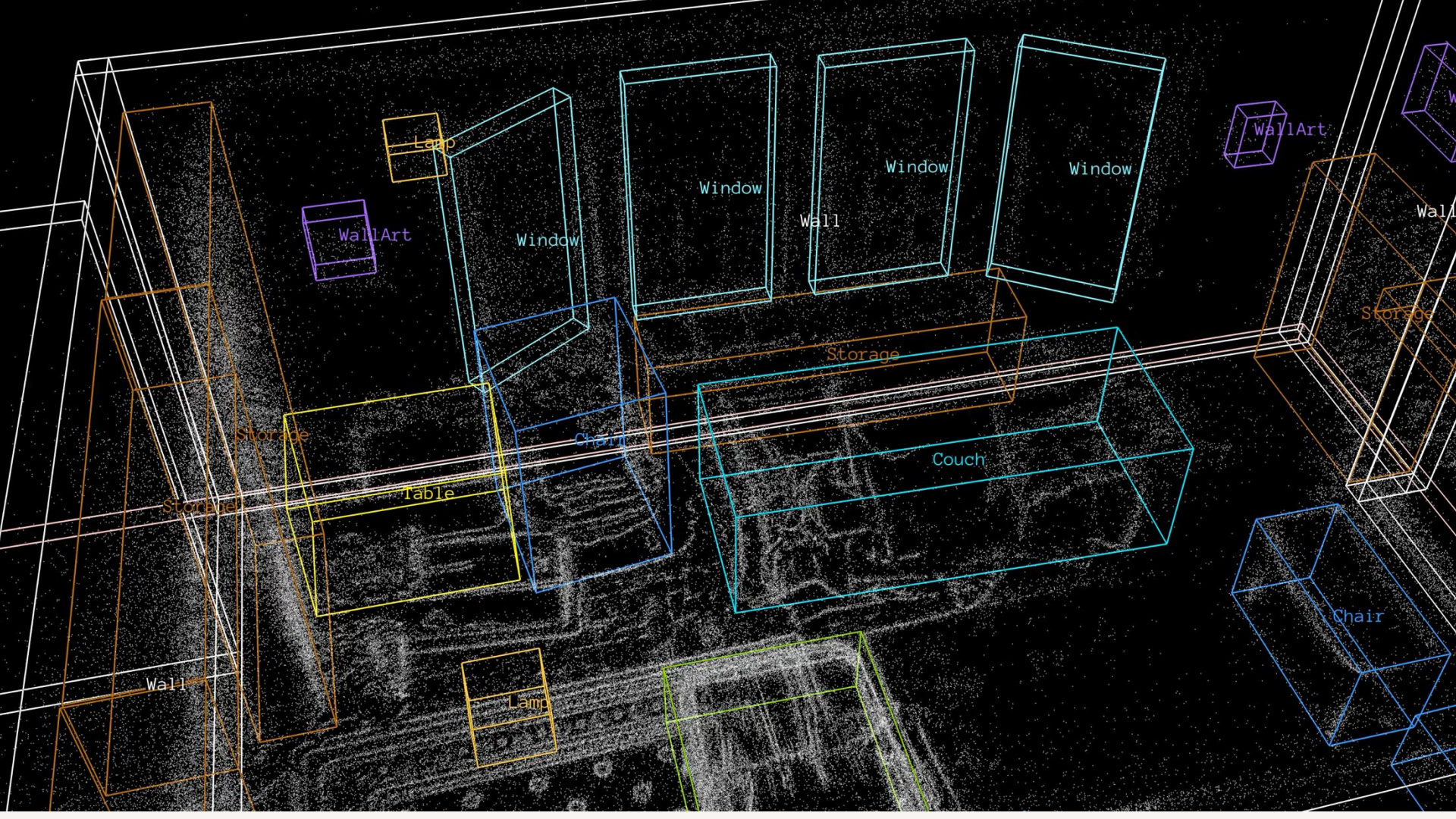
Aria Everyday Objects (AEO)

Benchmarking Egocentric 3D Object Detection

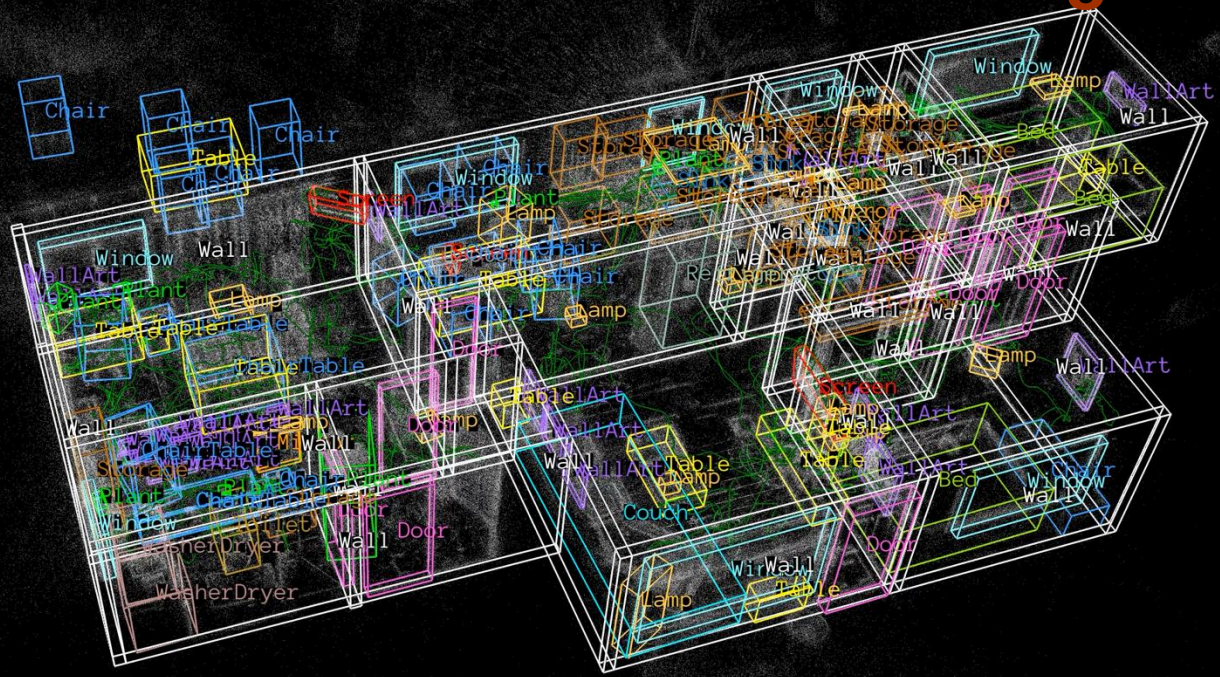
25 scenes, 45min Project Aria recordings,
1037 3D OBBs, 17 classes

https://www.projectaria.com/datasets/aeo/




Lamp
WallArt
Window
Window
Wall
Window
Window
WallArt
Wall
Storage
Storage
Chair
Storage
Table
Couch
Storage
Chair
Wall
Lamp

Coming Soon



Many more 3D BBs coming soon for the Nymeria Dataset!

https://www.projectaria.com/datasets/nymeria/

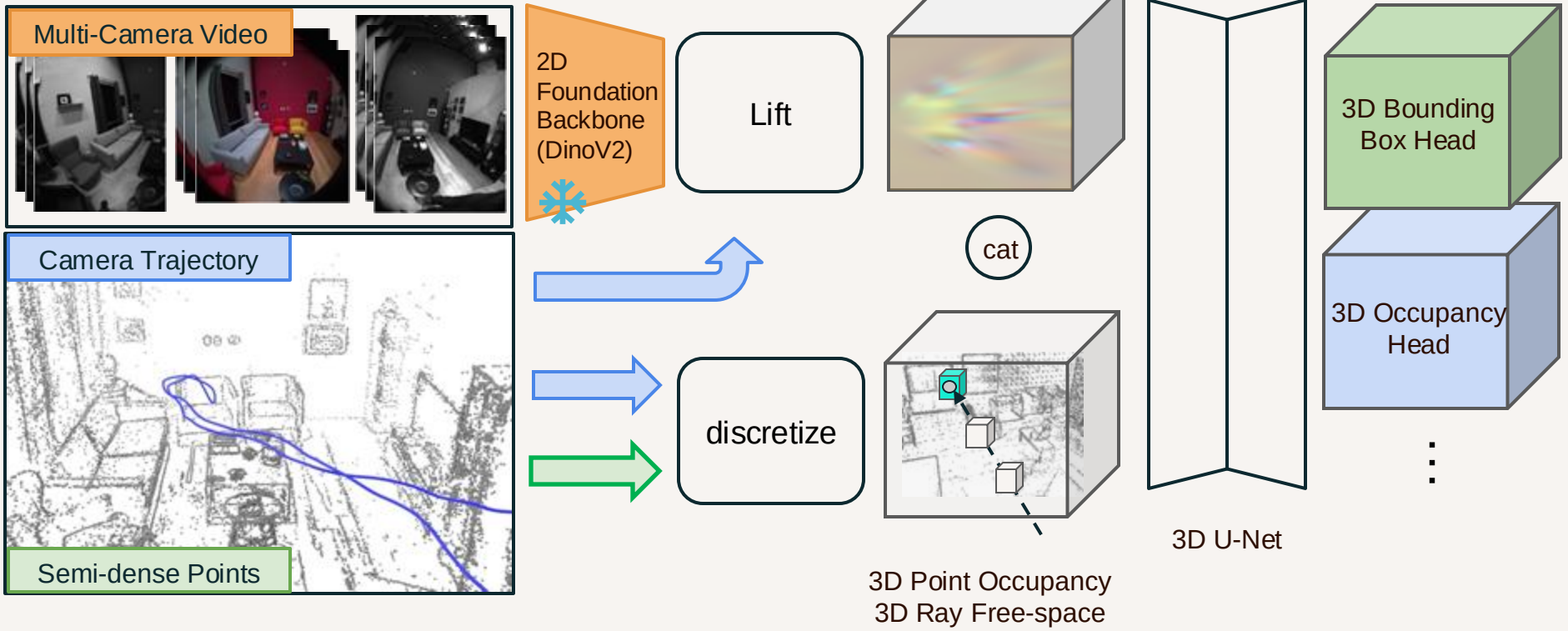
Aria Digital Twin (ADT)

Benchmarking Egocentric 3D Surface Estimation

1 scene, 12min of Project Aria data across 6 trajectories

https://www.projectaria.com/datasets/adt/

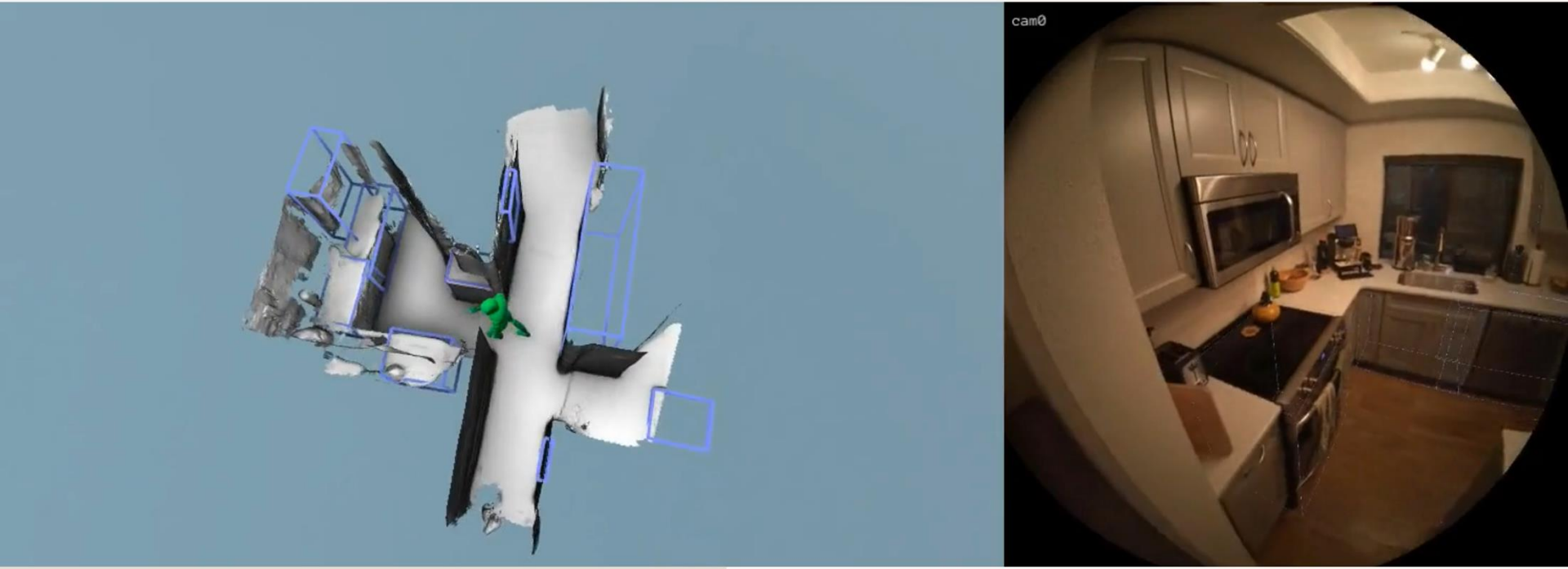
Egocentric Voxel Lifting (EVL) Model



EVL Inputs

Note that the Project Nymeria human mannequin in green is not part of the input and only shown for reference.

EVL Outputs



EVL + TSDF Fusion + 3DBB Tracking

Benchmarking EVL and other Baselines on EFM3D

EFM3D - Object Detection Benchmark

	Train Set	Modality	Decoder	ASE mAP Snippet	ASE mAP Sequence	AEO mAP Sequence
Cube R-CNN	OTS	frame	2D CNN	0.01	0.02	0.05
Cube R-CNN	ASE	frame	2D CNN	0.21	0.36	0.08
ImVoxelNet	ASE	snippet	3D CNN	0.30	0.64	0.15
3DETR	ASE	pts	Transformer	0.24	0.33	0.16
EVL (ours)	ASE	snip+pts	3D CNN	**0.40**	**0.75**	**0.22**

Lots of Opportunity for Improvement ☺

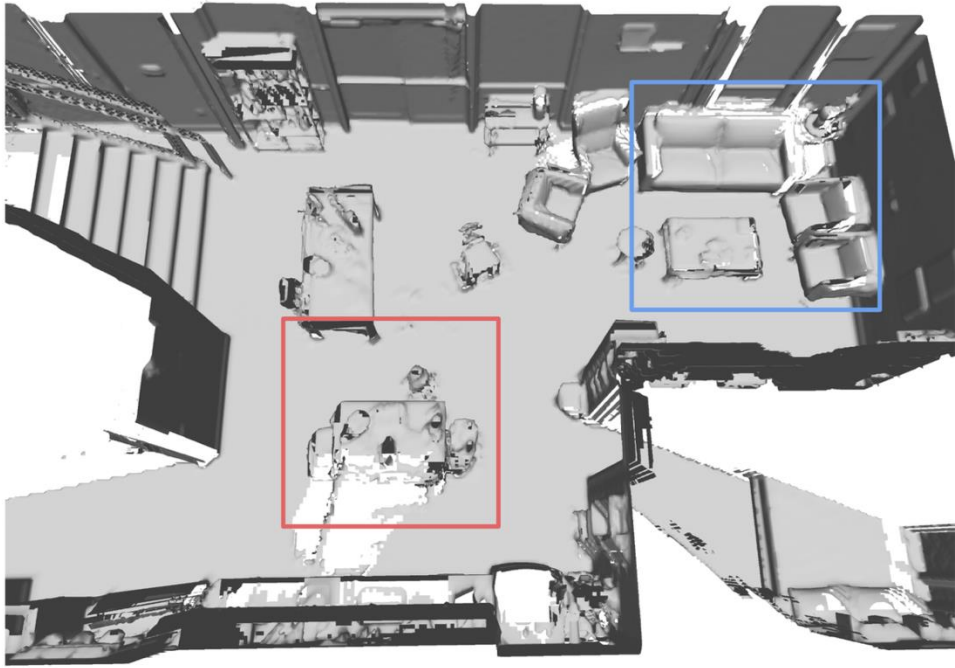
GT

ImVoxelNet

EVL

EFM3D - Surface Reconstruction Benchmark

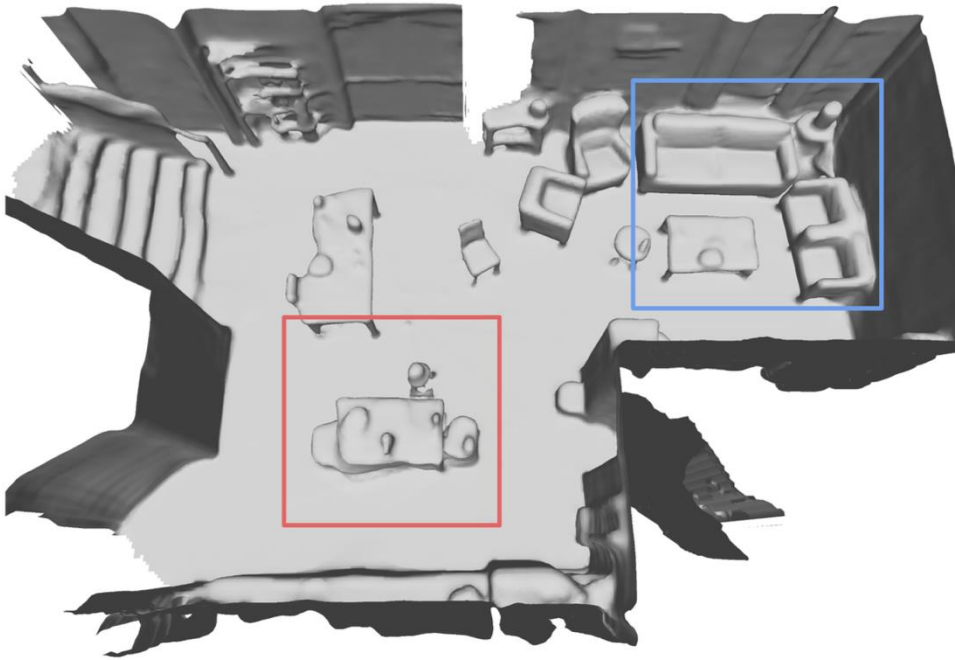
GT Mesh



Fused GT Depth

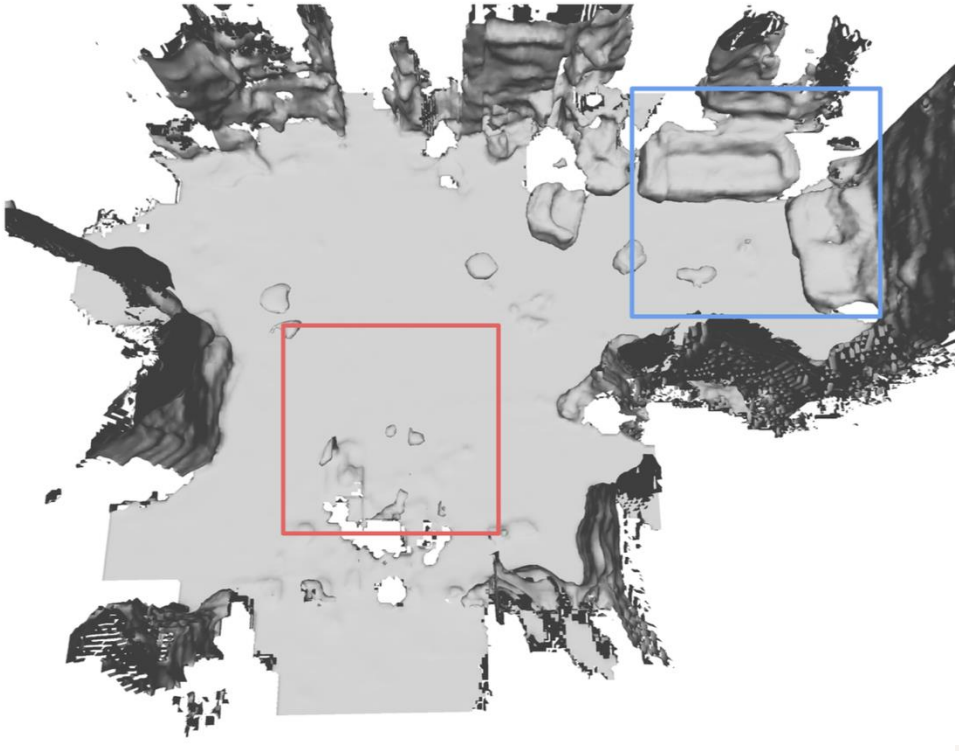
EFM3D - Surface Reconstruction Benchmark

GT Mesh

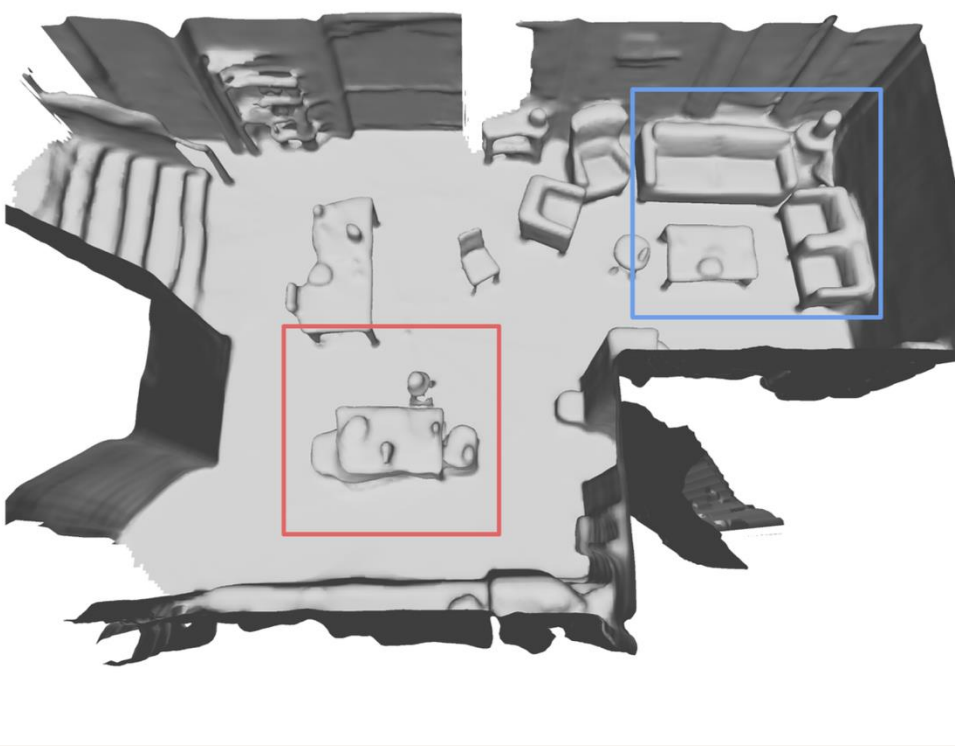


EVL

EFM3D - Surface Reconstruction Benchmark

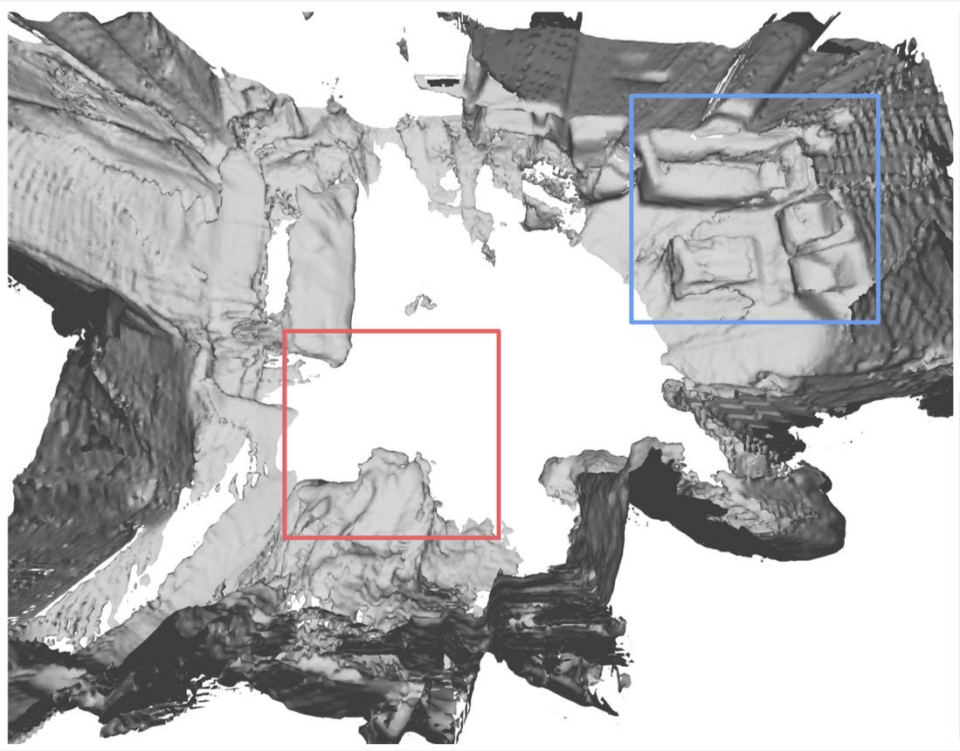


Neural Recon
(retrained on ASE)

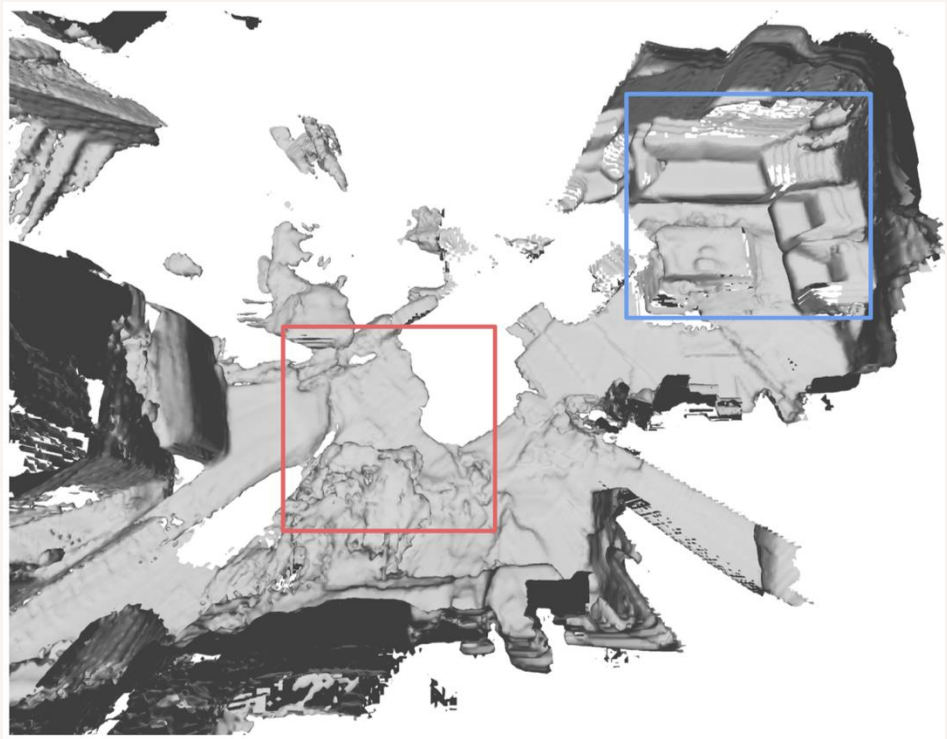


EVL

EFM3D - Surface Reconstruction Benchmark



Zoe Depth (OTS)



SimpleRecon (OTS)

What about Nerfs or Gaussian Splats (GS)?

- Sparse Partial Views are hard without learned priors.
- Dynamics in Egocentric Data lead to ”fog” Artifacts.

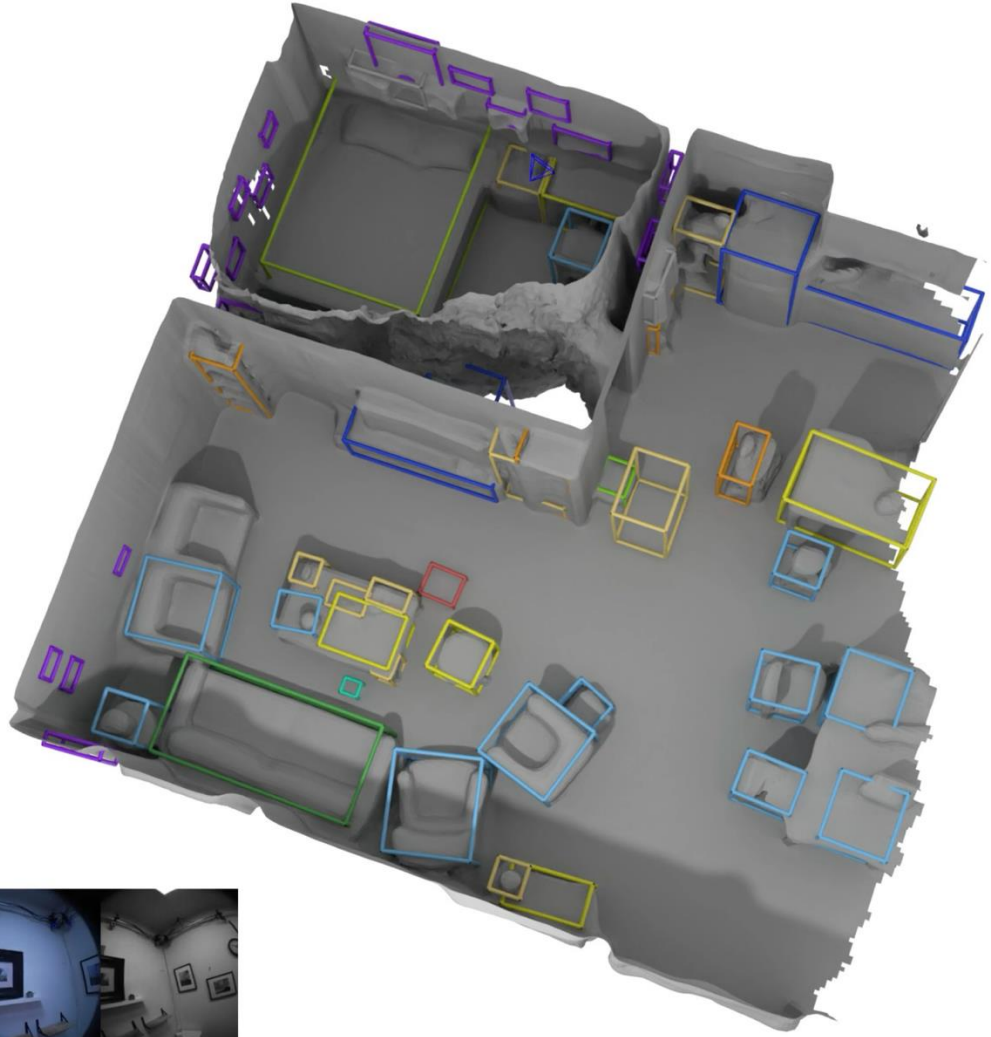
Learned Dynamics Masks | EgoLifter Using Mask | Baseline GS

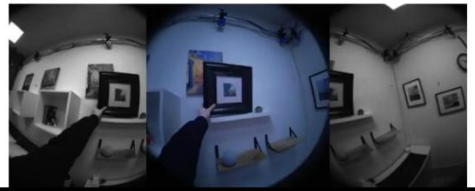
“EgoLifter: Open-world 3D Segmentation for Egocentric Perception”

Poster Session 1 on Tuesday

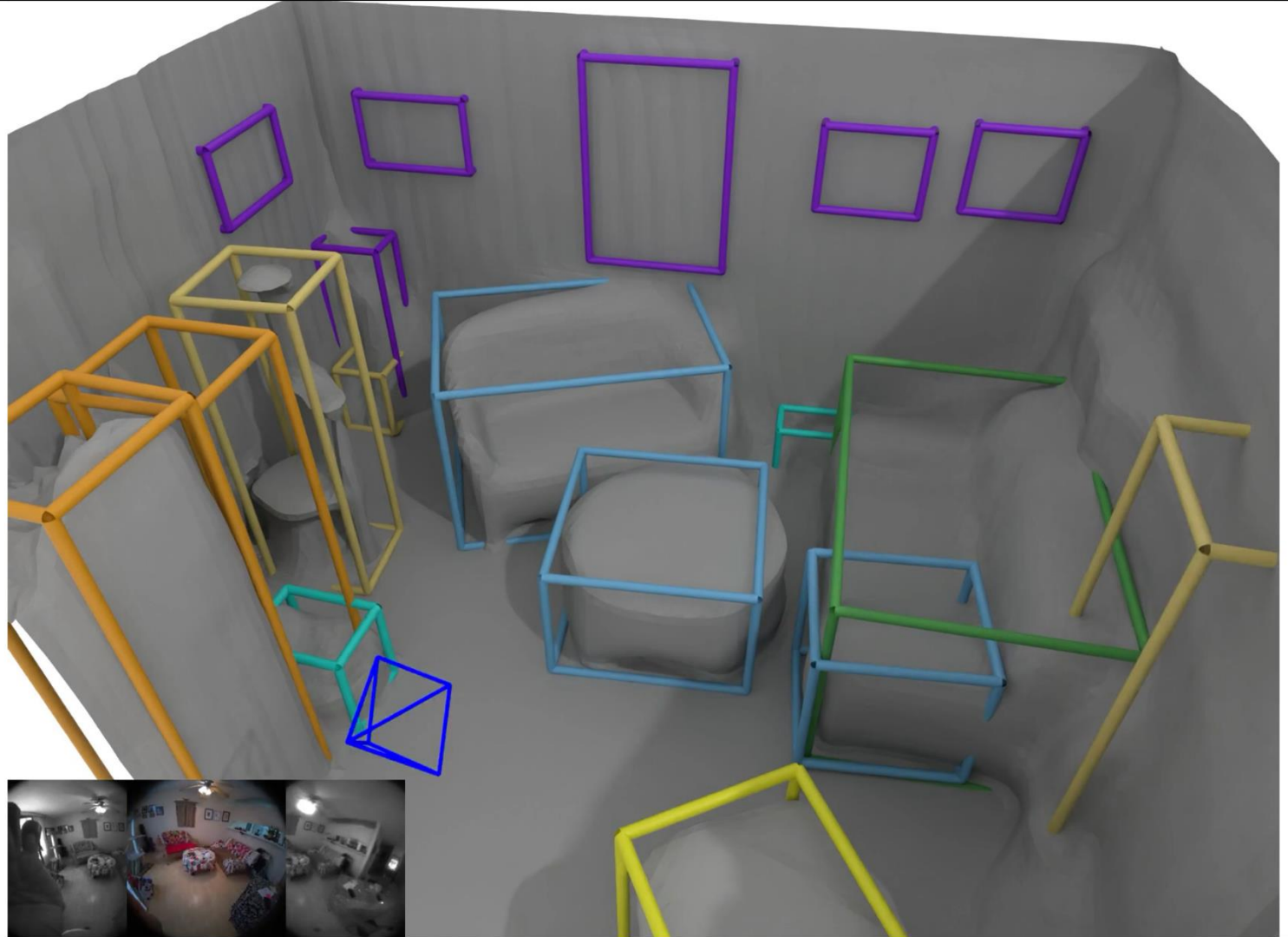
Probing EVL Sim-to-Real Performance on Project Aria Data

ADT
indoor
scene





AEO
indoor
scene

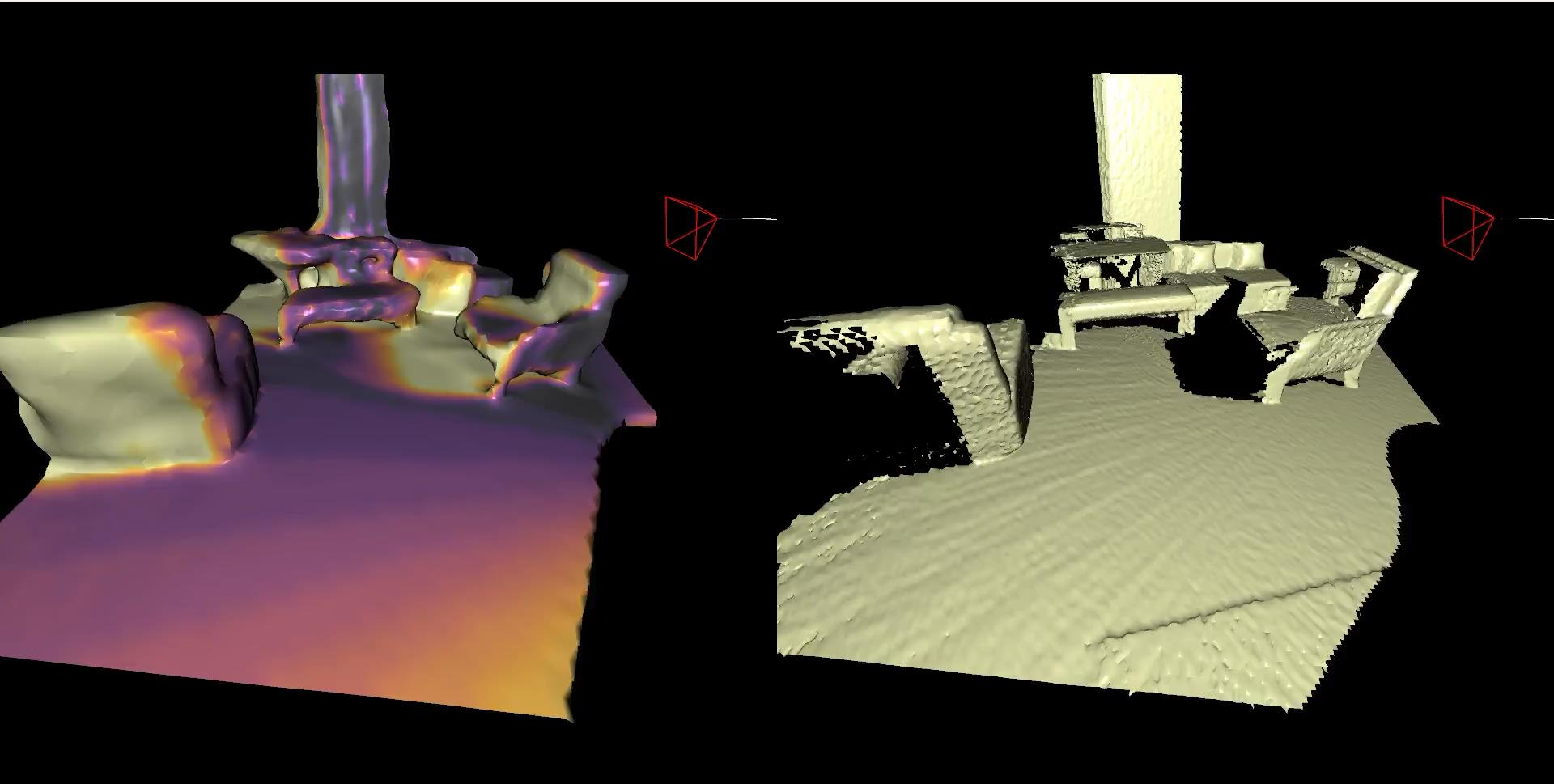


AEO
outdoor
scene

Scalability via learned priors and incremental fusion.

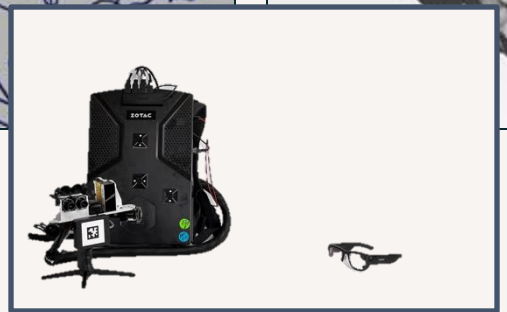
EVL Fused Mesh

GT Depth Fused Mesh

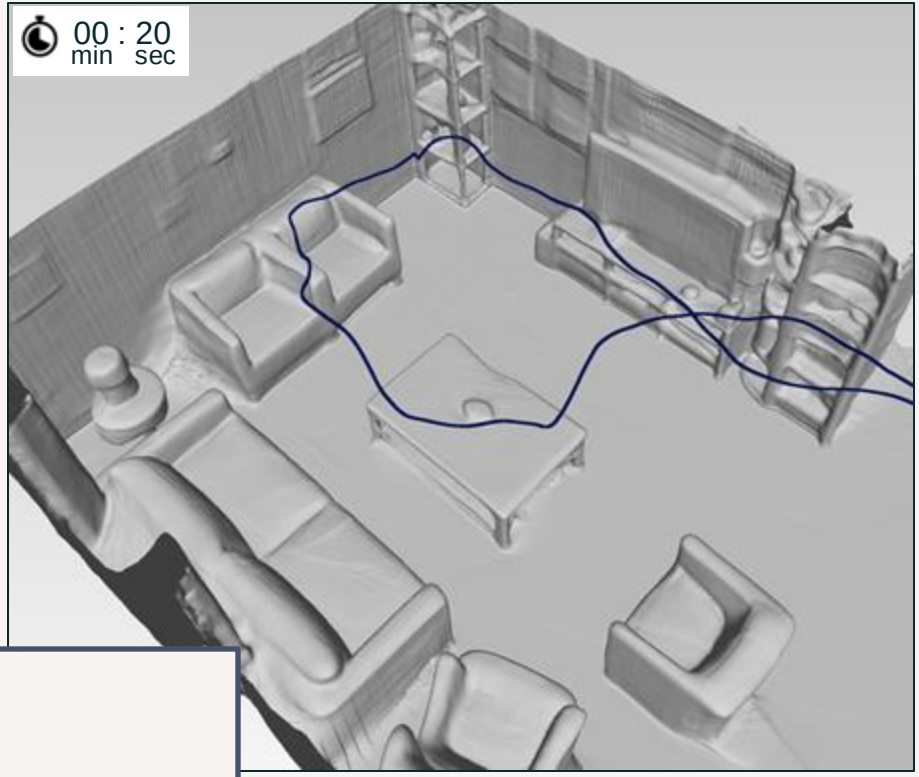


Replica

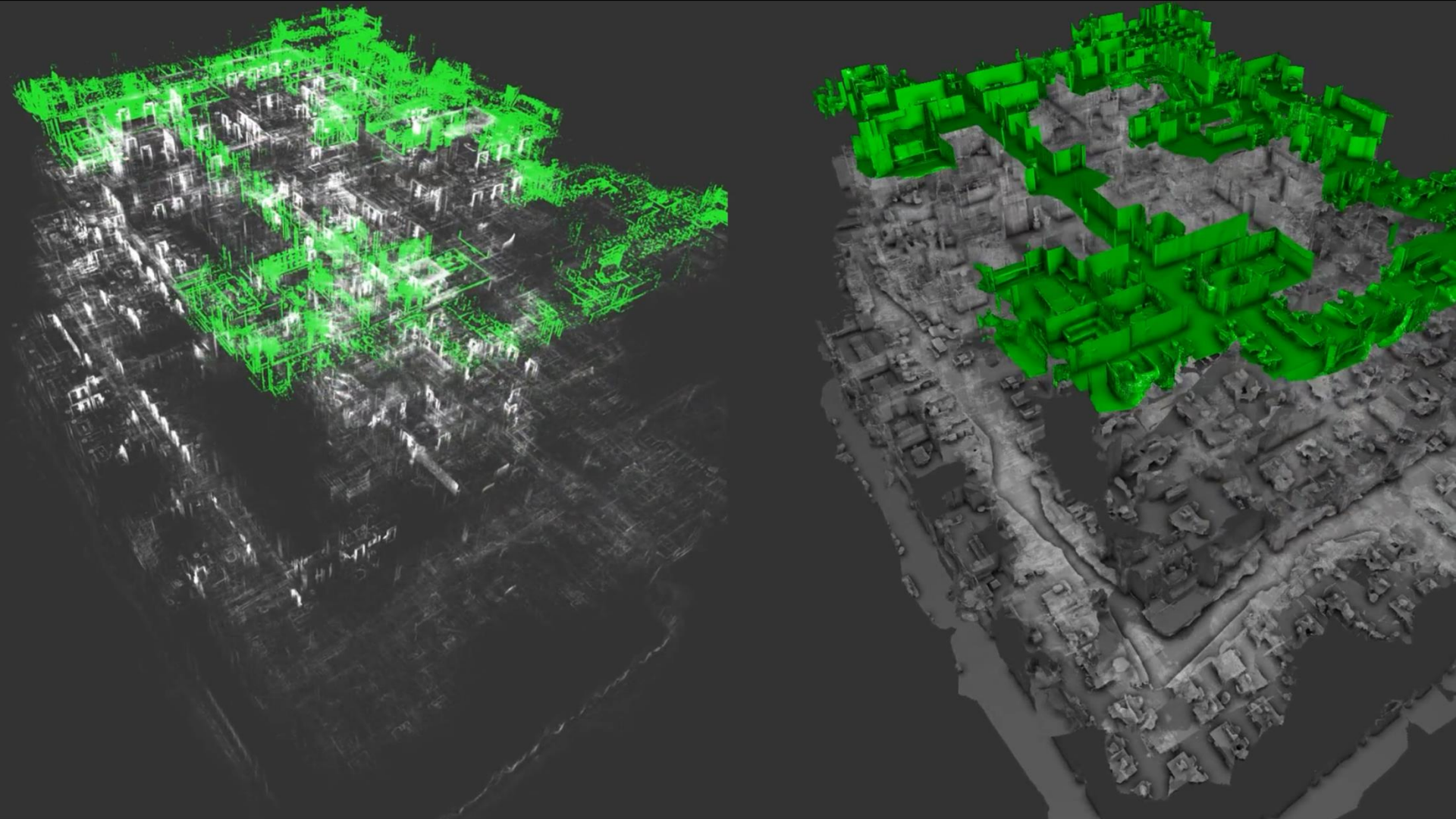
Camera + Depth Sensor +
Dedicated Scanning Motion



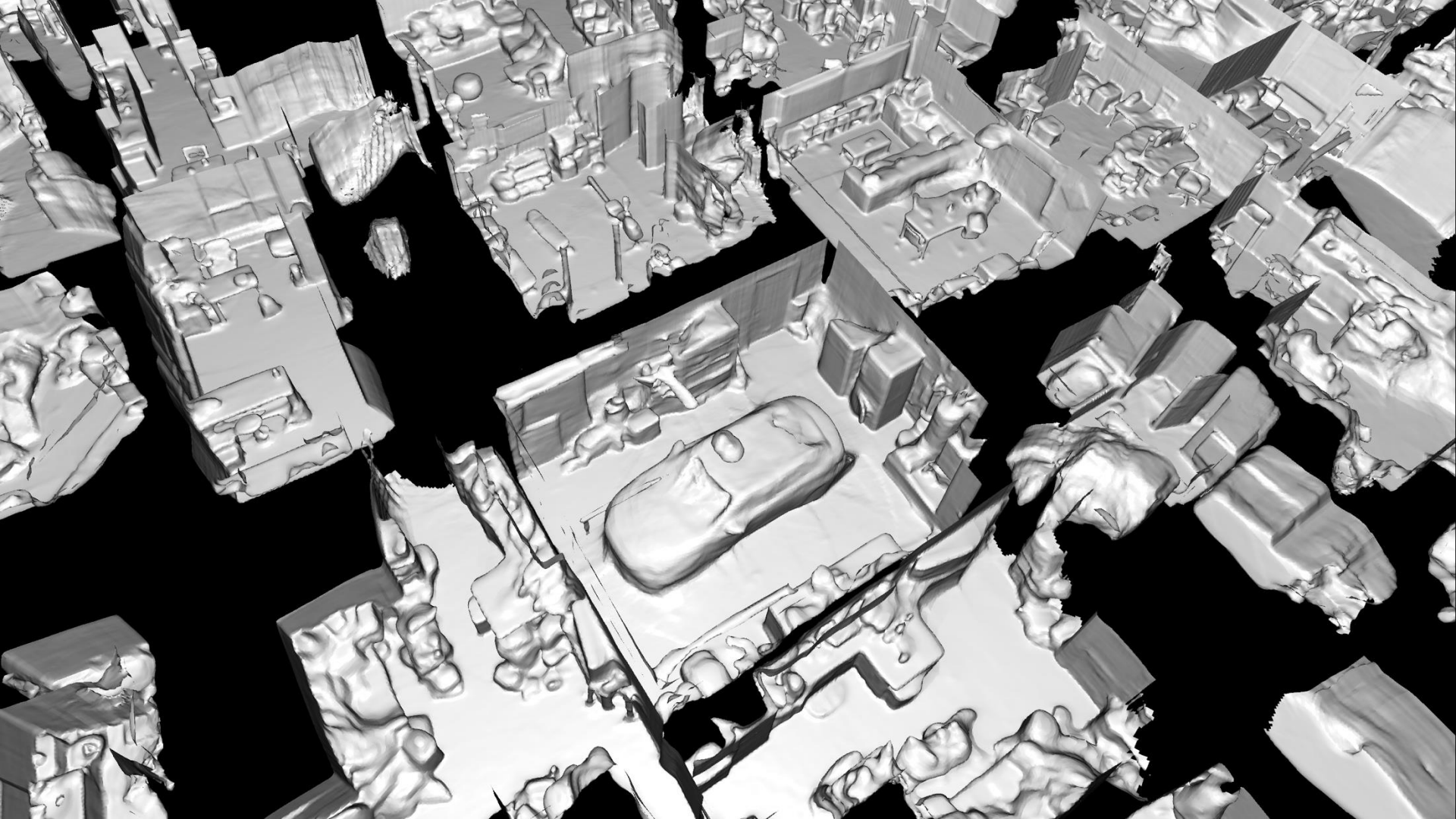
EVL



Aria + Casual Motion + EVL



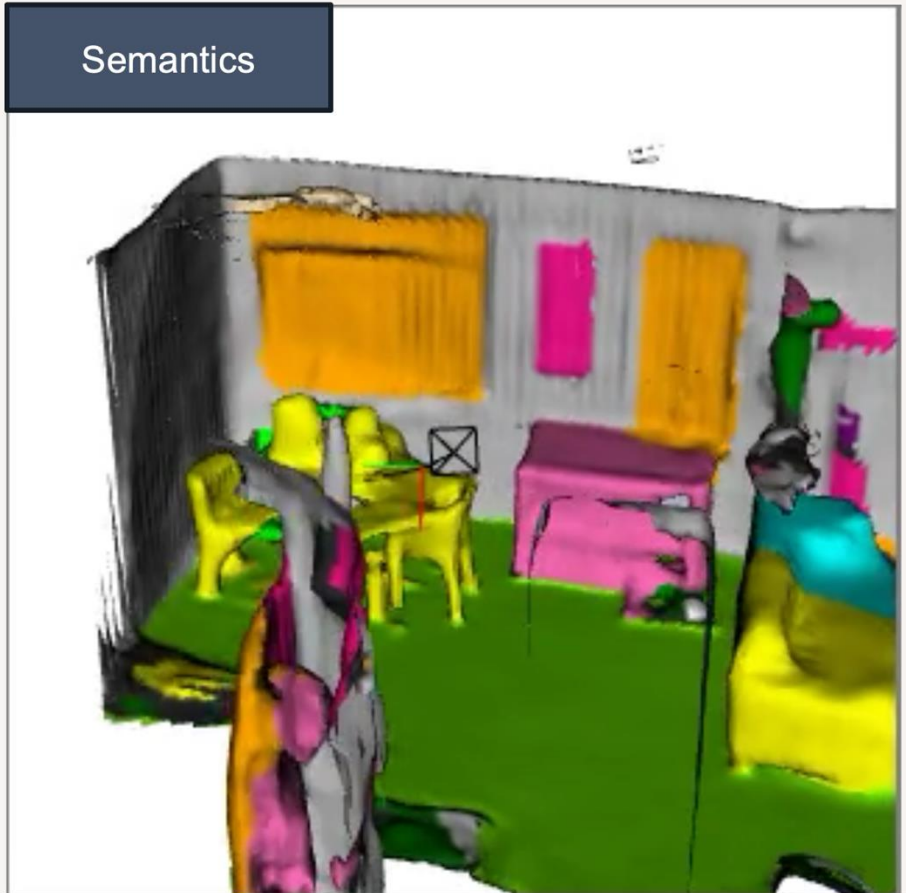
Reconstructing
1000 Project
Aria Recordings



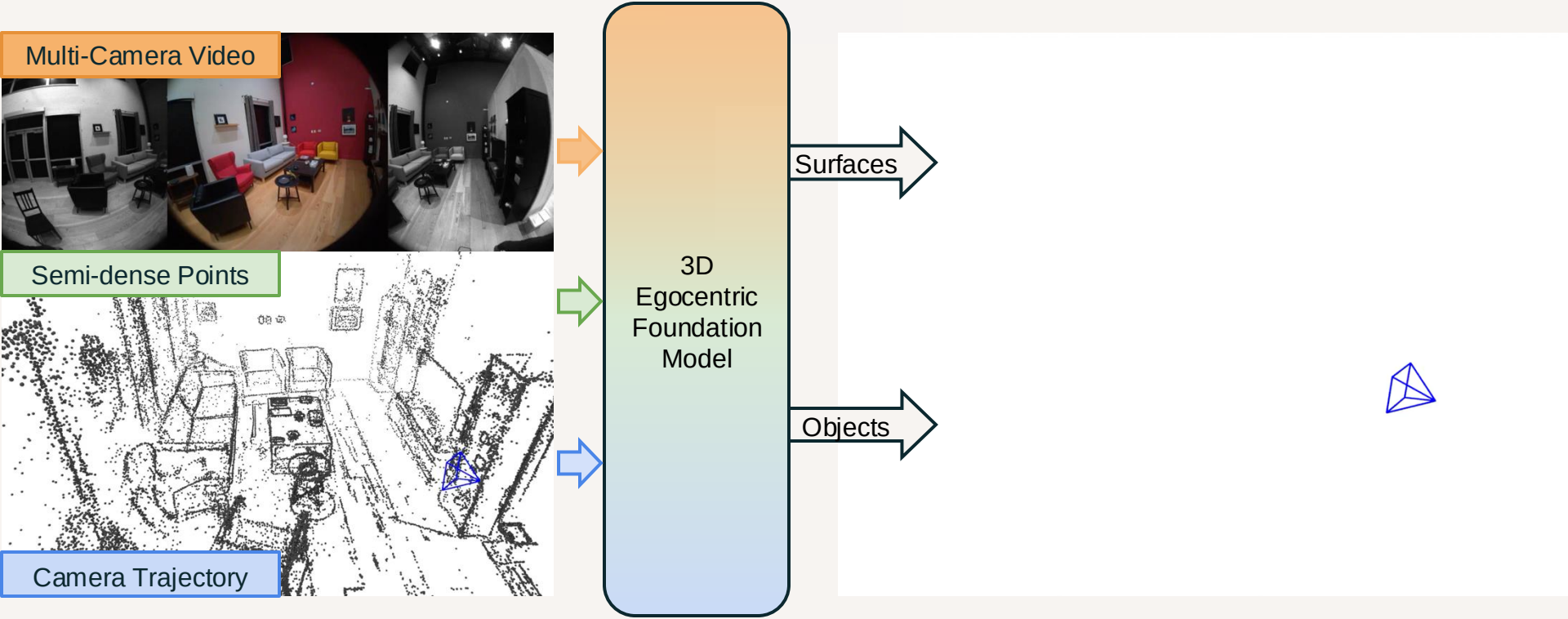
Sneak Peak: Appearance and Semantics

HDR RGB

Semantics



Conclusion



Egocentric data brings novel challenges:

- always-on sensor data
- human motion and partial observations
- dynamics

3D EFMs can address these:

- strong priors anchored in 3D
- incremental fusion into a persistent state

EFM3D Benchmark and EVL Model are Out Now!

- EFM3D Benchmark https://www.projectaria.com/research/efm3d/
- Aria Everyday Objects (AEO) https://www.projectaria.com/datasets/aeo/
- EVL Model https://github.com/facebookresearch/efm3d/