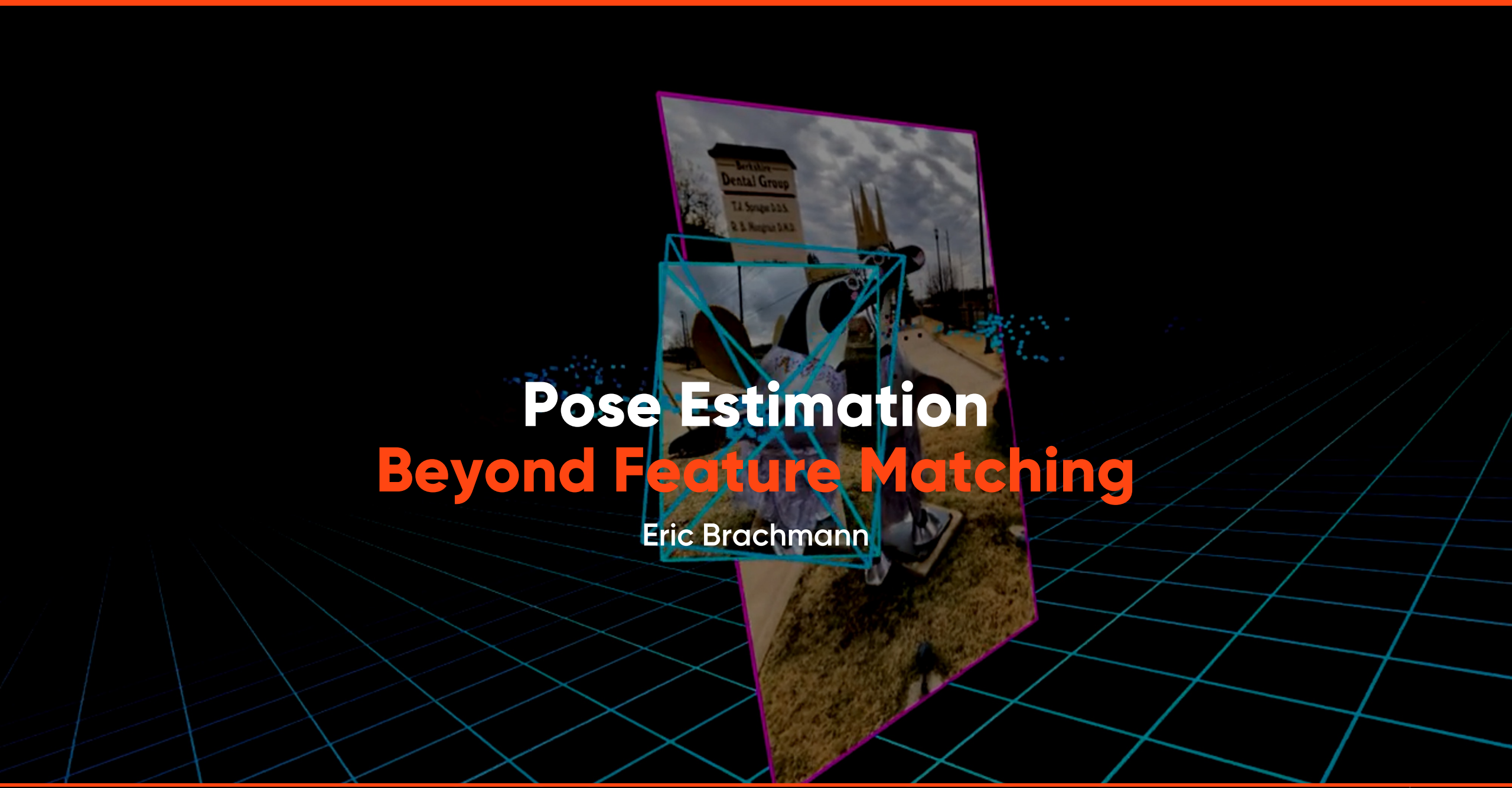


Pose Estimation
Beyond Feature Matching

Eric Brachmann

Eric Brachmann

Staff Scientist

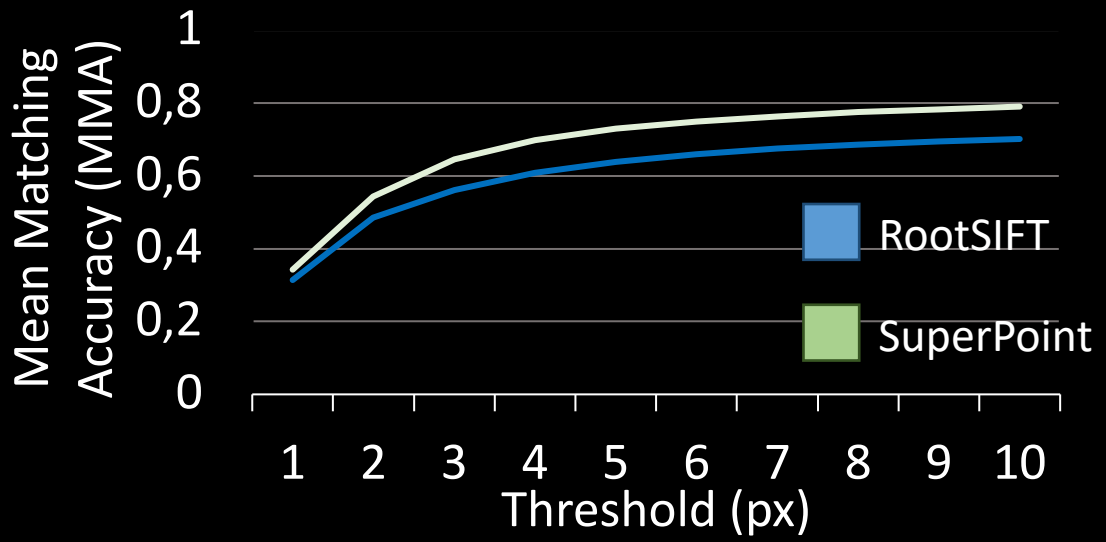
ebrach.github.io

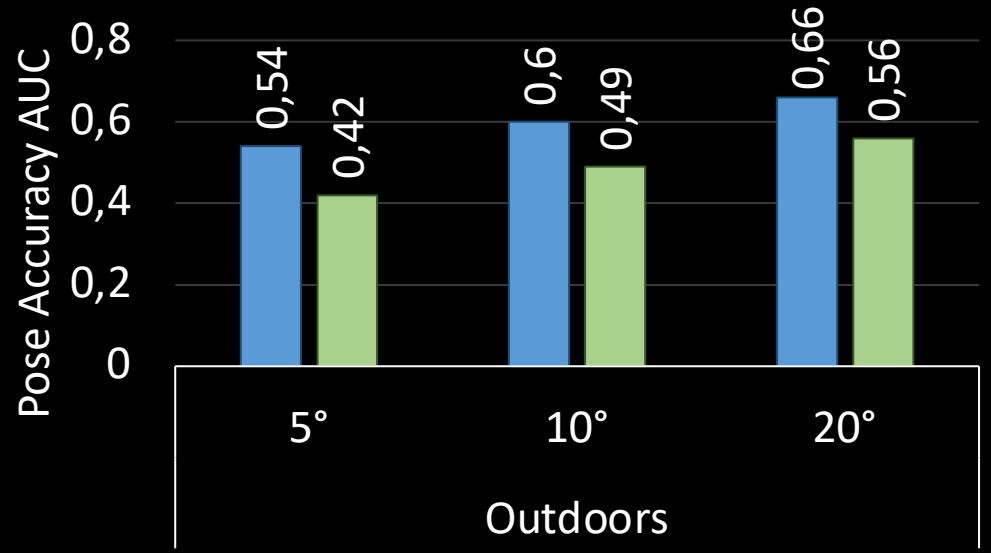
@eric_brachmann

ebrachmann@nianticlabs.com

linkedin.com/in/eric-brachmann

NIANTIC

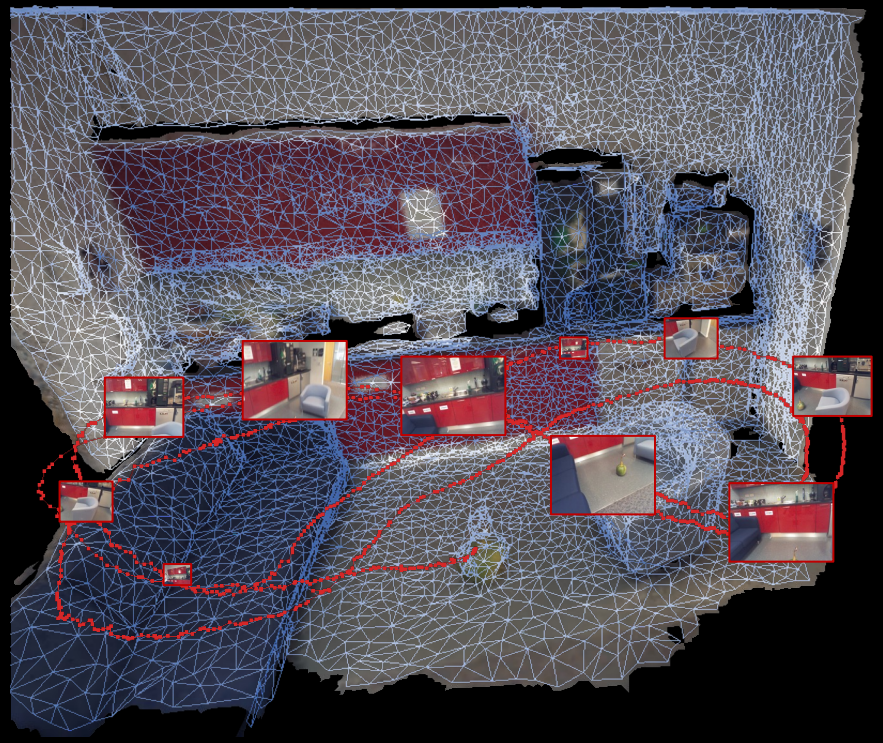




"Reinforced Feature Points: Optimizing Feature Detection and Description for a High-Level Task" Bhowmik et al., CVPR'20 (oral)

Mapping

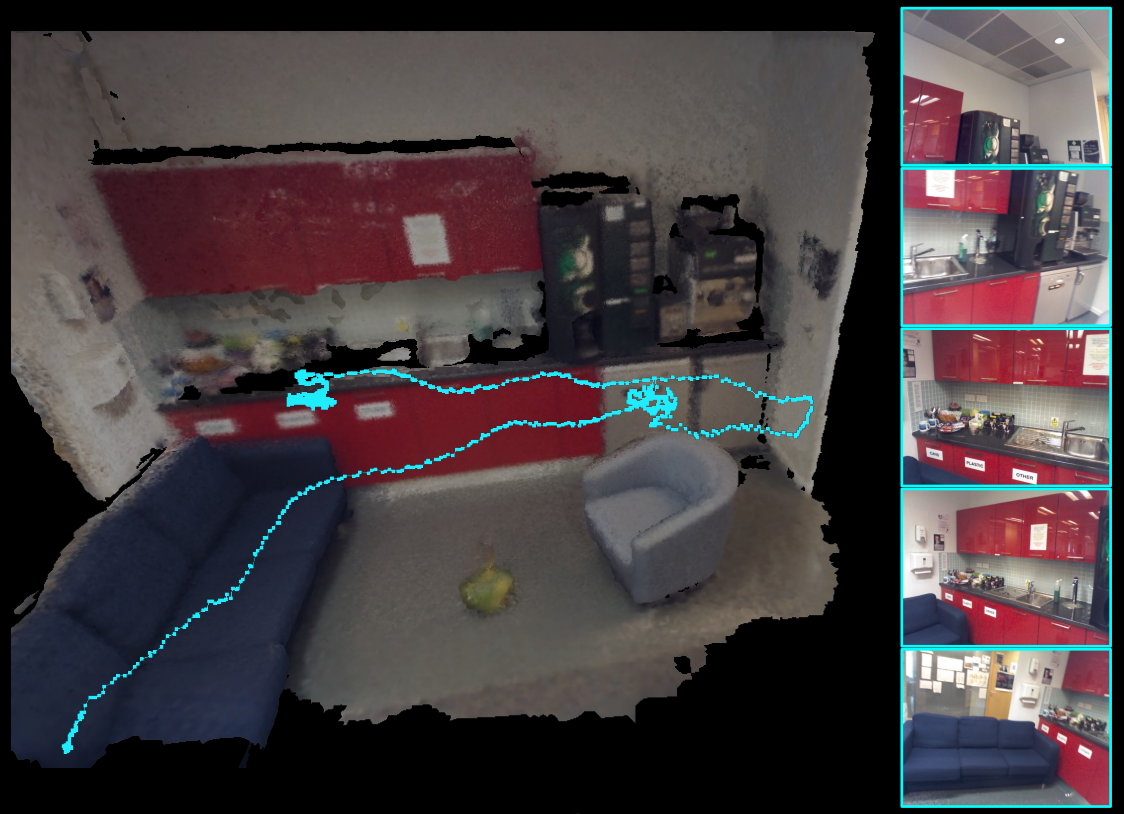
Build the Scene Representation



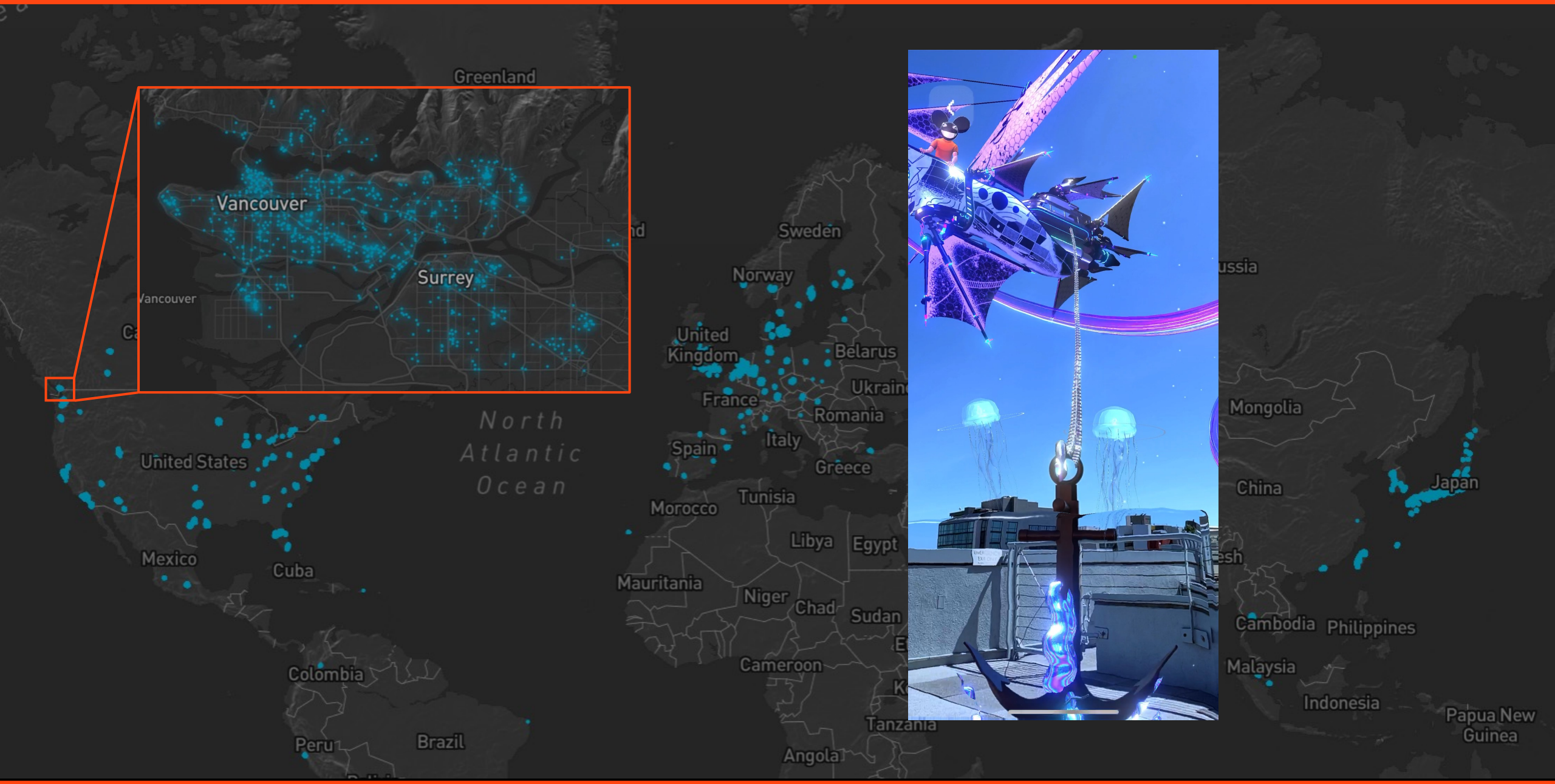
Obtain Posed Images
Triangulate Scene
Bundle Adjustment

Re-Localisation

Register Query Frames

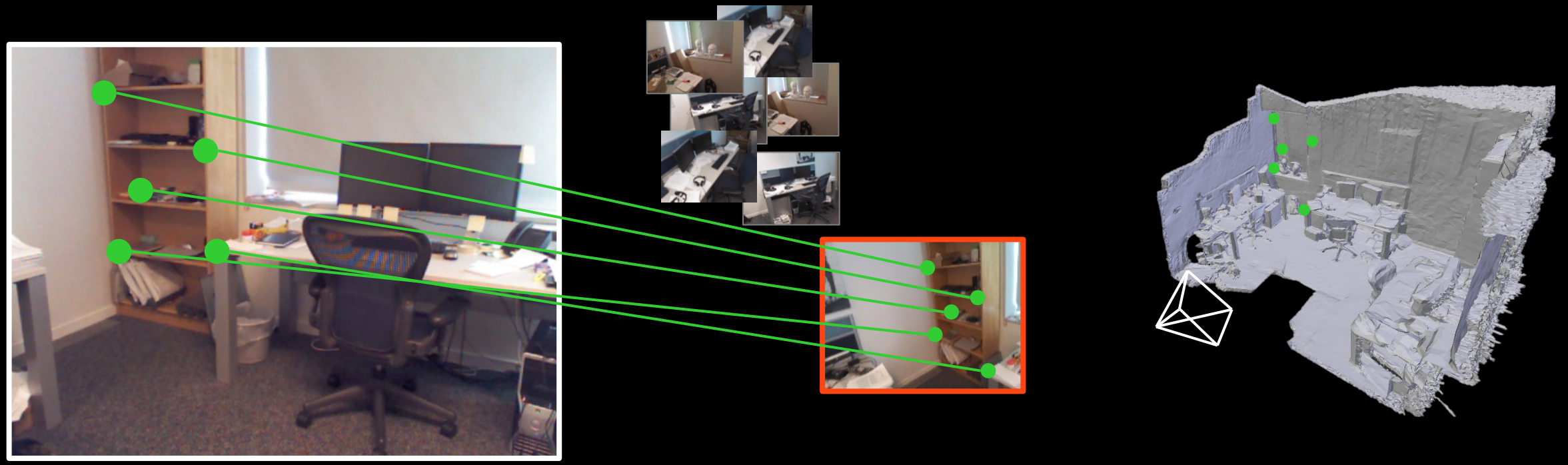


NN Retrieval
Discrete Feature Matching
Pose Optimization


Greenland
Vancouver
Surrey
Vancouver
Sweden
Norway
United Kingdom
Belarus
France
Romania
Italy
Spain
Greece
Tunisia
Morocco
Libya
Egypt
Mauritania
Niger
Chad
Sudan
Cameroon
Tanzania
Angola
North Atlantic Ocean
United States
Mexico
Cuba
Colombia
Peru
Brazil
Mongolia
China
Japan
Cambodia
Philippines
Malaysia
Indonesia
Papua New Guinea

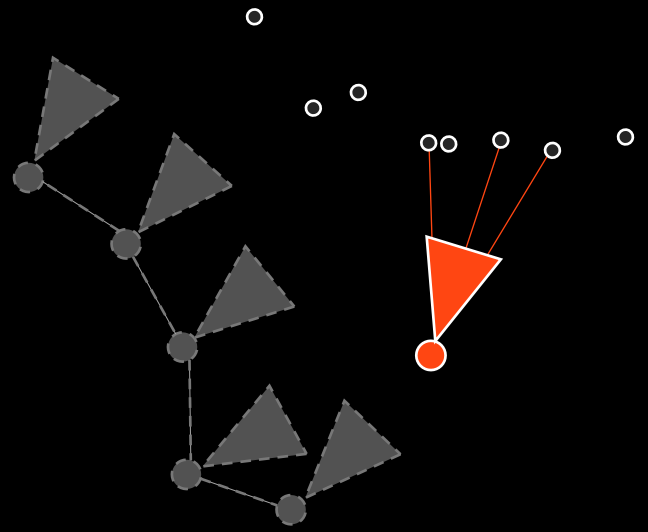
Query Image I	Image Retrieval	Feature Extraction	Feature Matching	RANSAC	6D Pose $\mathbf{h}$

RANSAC: Estimate Poses (PnP); Score Poses (Inliers)



e.g. “From Coarse to Fine: Robust Hierarchical Localization at Large Scale”, Sarlin et al., CVPR’19

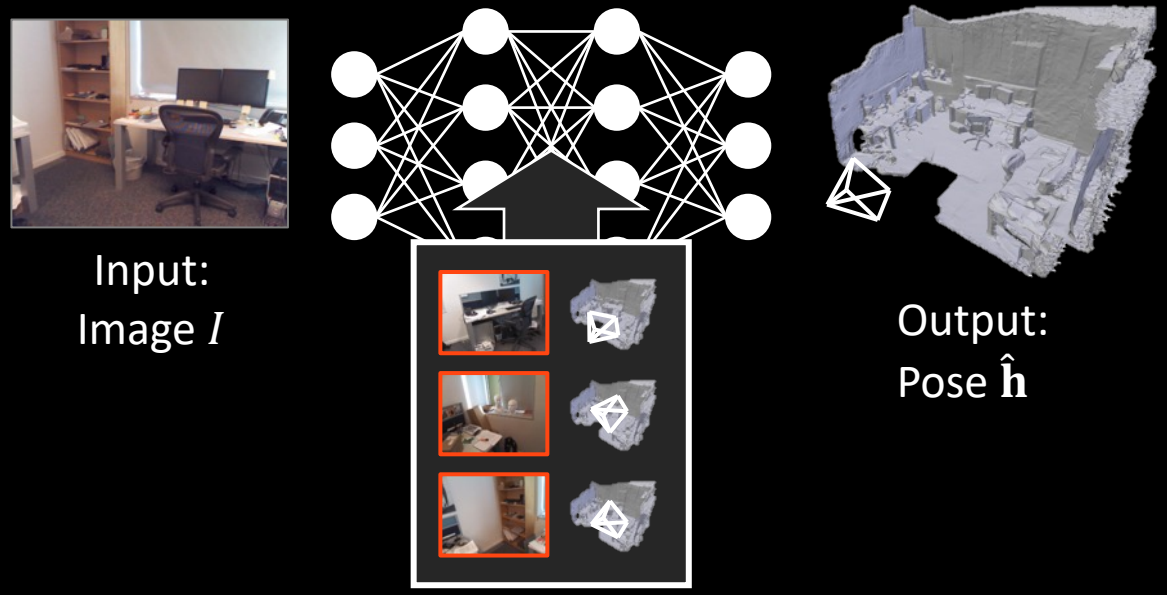
Feature Matching

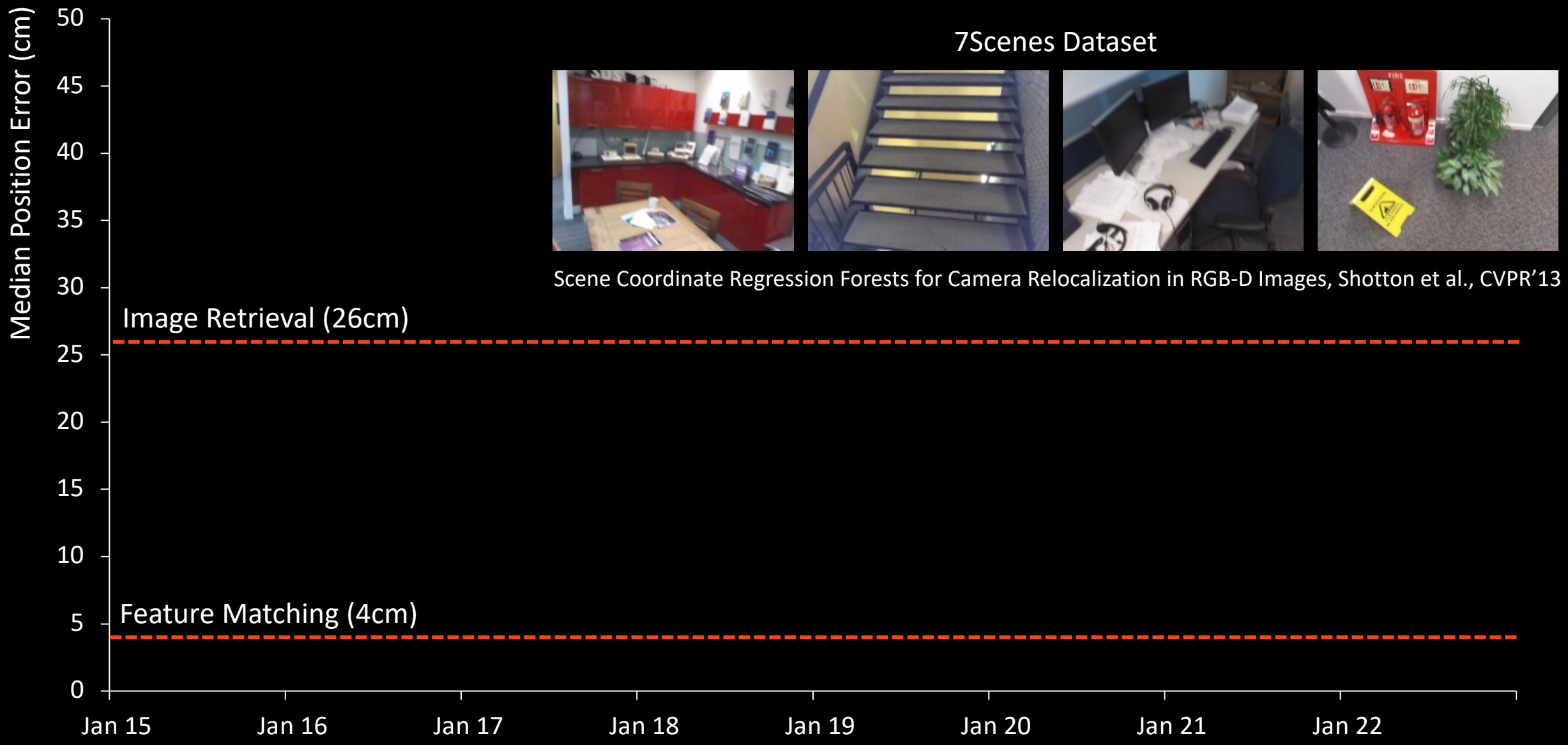


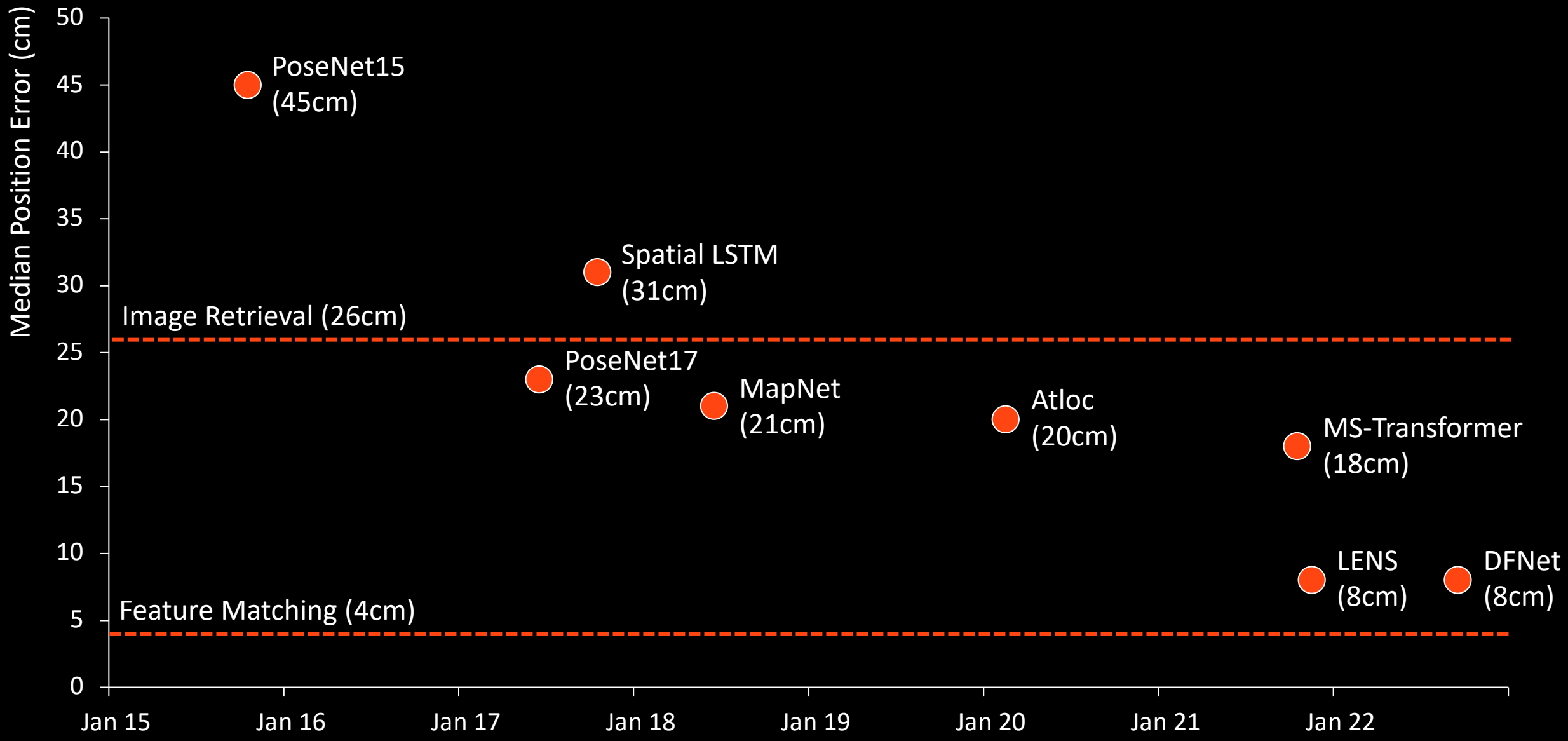
- Moderate query time
- Slow at mapping time
- High Memory Demand
- High Accuracy

Query Image I	Image Retrieval	Feature Extraction	Feature Matching	RANSAC: Estimate Poses (PnP); Score Poses (Inliers)	6D Pose $\mathbf{h}$

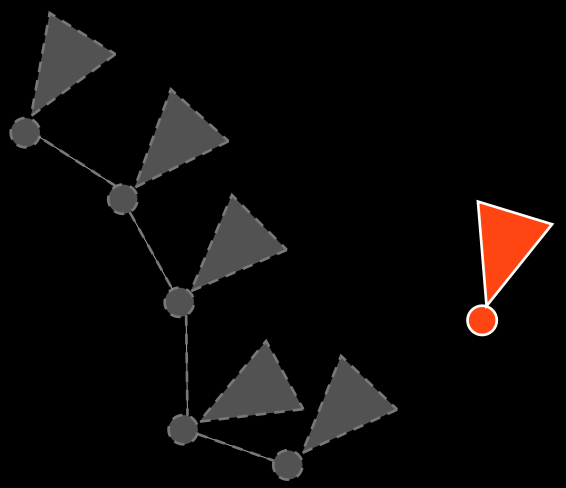
Query Image I	Absolute Pose Regression	6D Pose $\mathbf{h}$




Median Position Error (cm)
50
45
40
35
30
25
20
15
10
5
0
Jan 15
Jan 16
Jan 17
Jan 18
Jan 19
Jan 20
Jan 21
Jan 22
7Scenes Dataset
Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images, Shotton et al., CVPR'13
Image Retrieval (26cm)
Feature Matching (4cm)

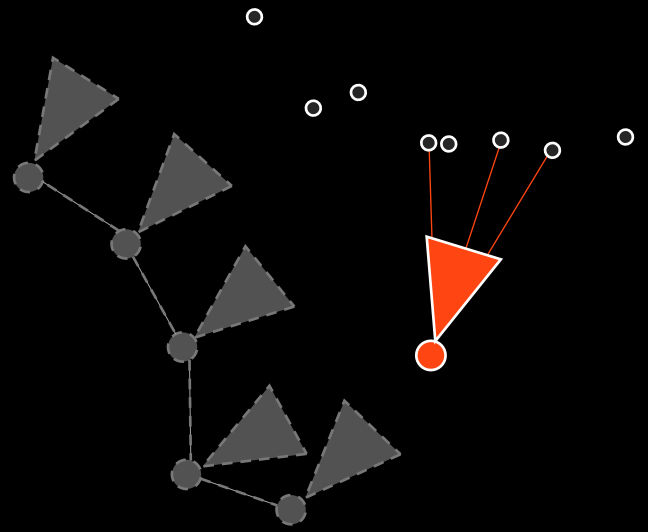

Median Position Error (cm)
50
45
40
35
30
25
20
15
10
5
0
PoseNet15 (45cm)
Spatial LSTM (31cm)
Image Retrieval (26cm)
PoseNet17 (23cm)
MapNet (21cm)
Atloc (20cm)
MS-Transformer (18cm)
LENS (8cm)
DFNet (8cm)
Feature Matching (4cm)
Jan 15
Jan 16
Jan 17
Jan 18
Jan 19
Jan 20
Jan 21
Jan 22

Absolute Pose Regression



- Fast at query time
- Slow at mapping time
- Moderate memory demand
- Moderate accuracy

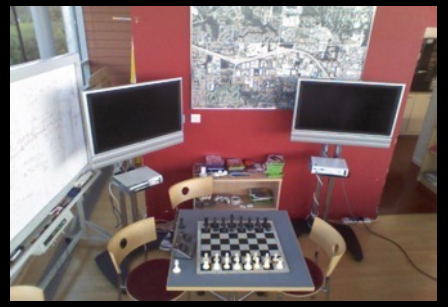
Feature Matching



- Moderate query time
- Slow at mapping time
- High memory demand
- High accuracy

Query Image I	Image Retrieval	Feature Extraction	Feature Matching	RANSAC	Estimate Poses (PnP) / Score Poses (Inliers)	6D Pose $\mathbf{h}$

Query Image I	Image Retrieval	Correspondence Regression	RANSAC	Estimate Poses (PnP) / Score Poses (Inliers)	6D Pose $\mathbf{h}$

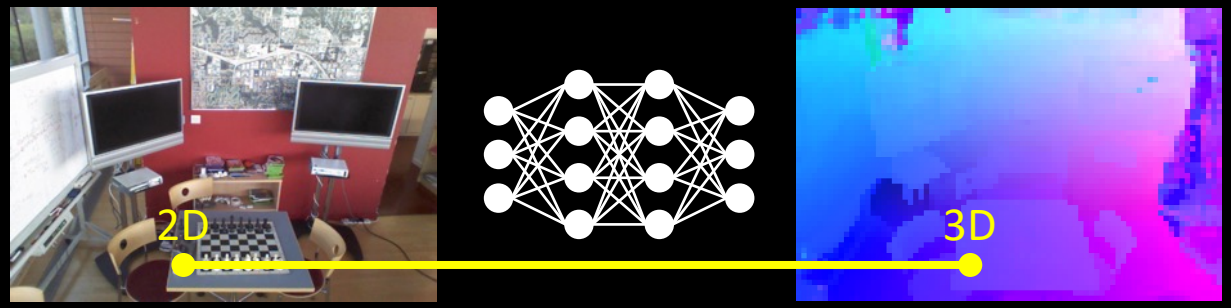


Query Image

Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images, Shotton et al., CVPR'13

Query Image I	Image Retrieval	Feature Extraction	Feature Matching	RANSAC	Estimate Poses (PnP) / Score Poses (Inliers)	6D Pose $\mathbf{h}$

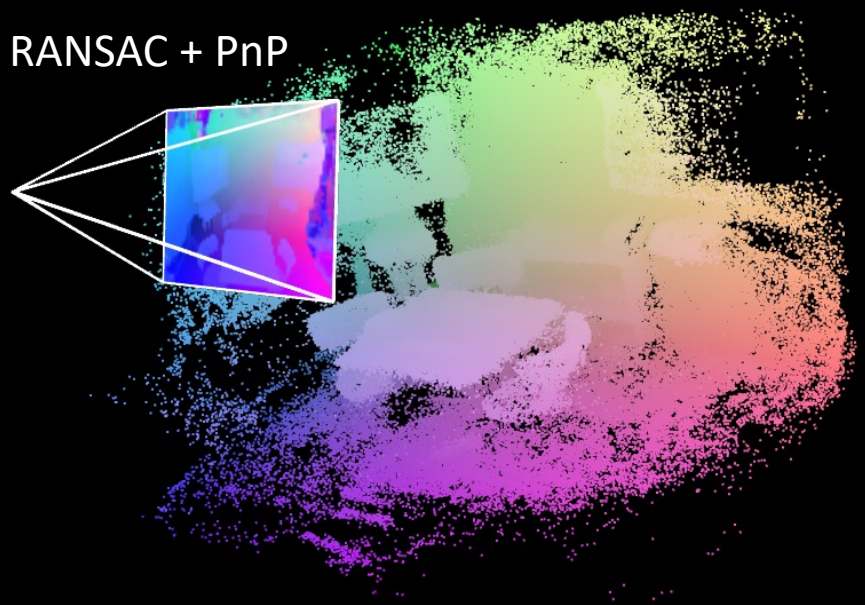
Query Image I	Image Retrieval	Correspondence Regression	RANSAC	Estimate Poses (PnP) / Score Poses (Inliers)	6D Pose $\mathbf{h}$



Query Image

Scene Coordinate Prediction

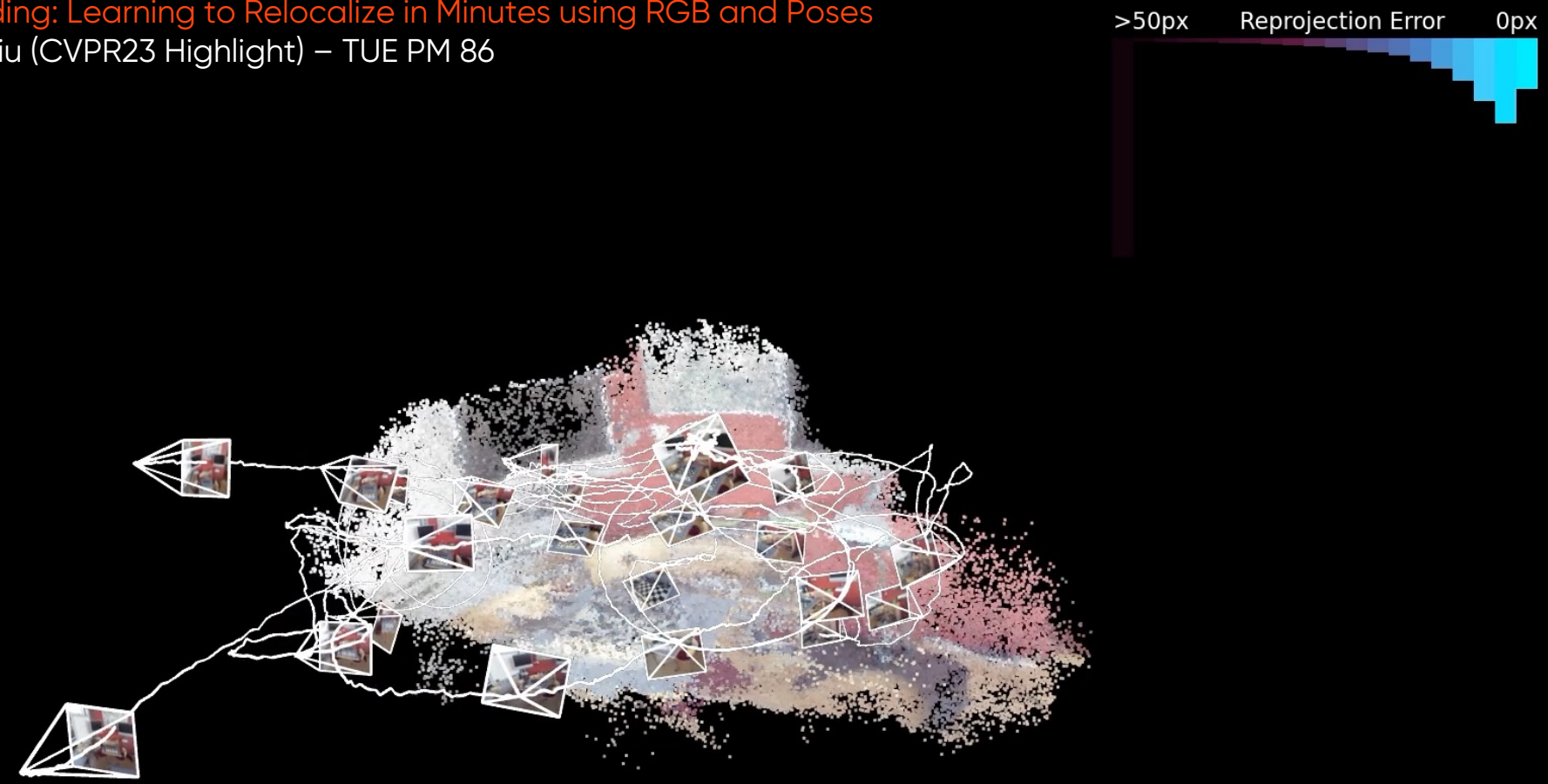
RANSAC + PnP



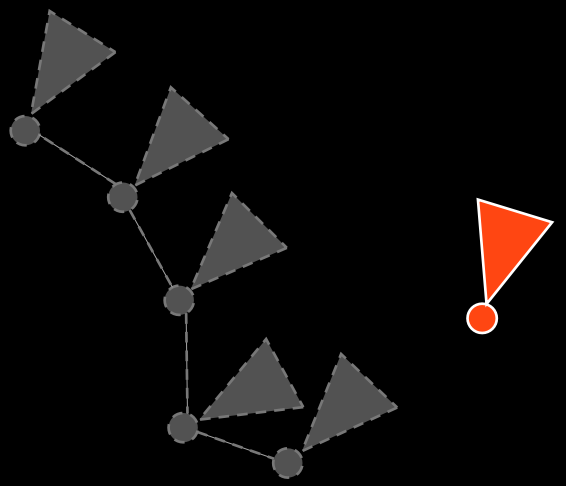
Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images, Shotton et al., CVPR'13

Accelerated Coordinate Encoding: Learning to Relocalize in Minutes using RGB and Poses

Brachmann, Cavallari, Prisacariu (CVPR23 Highlight) – TUE PM 86

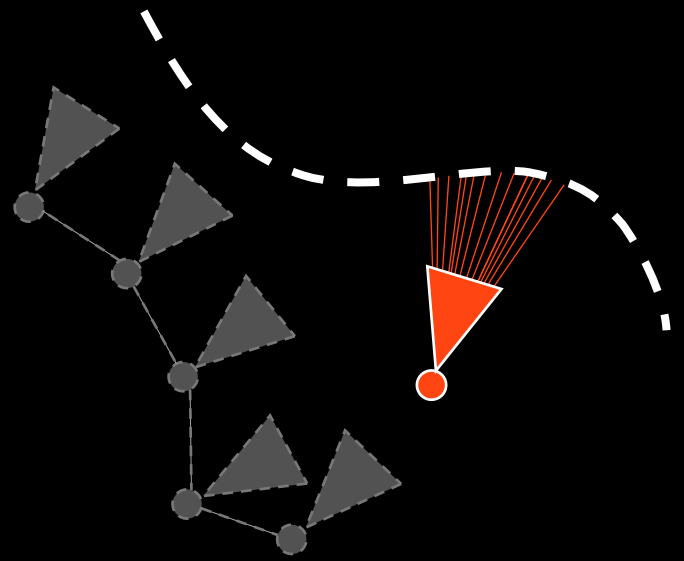


Absolute Pose Regression



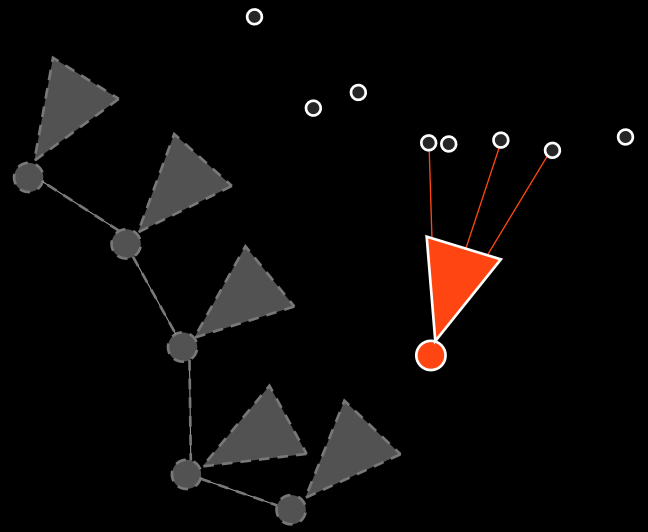
- Fast at query time
- Slow at mapping time
- Moderate memory demand
- Moderate accuracy

Correspondence Regression

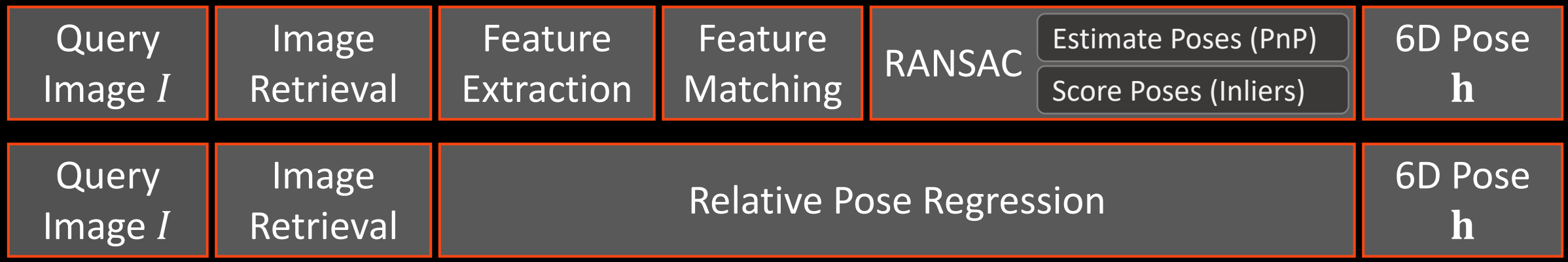


- Fast at query time
- Moderate mapping time
- Low memory demand
- High accuracy

Feature Matching



- Moderate query time
- Slow at mapping time
- High memory demand
- High accuracy


Query Image I
Image Retrieval
Feature Extraction
Feature Matching
RANSAC
Estimate Poses (PnP)
Score Poses (Inliers)
6D Pose $\mathbf{h}$
Query Image I
Image Retrieval
Relative Pose Regression
6D Pose $\mathbf{h}$

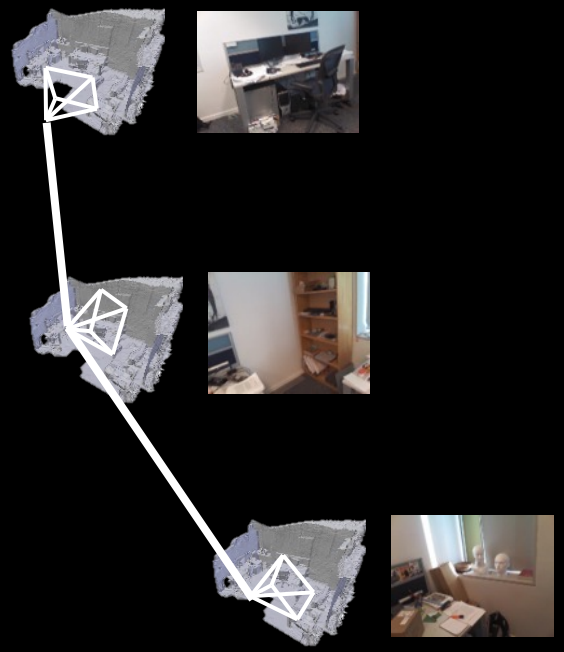
Preparation
Scene-Agnostic Training

Pre-Train Image Retrieval

Pre-Train Relative Pose Regressor

Mapping
Scene-Specific Training

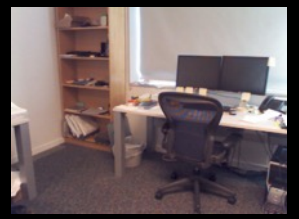
Obtain Posed Images
Build Retrieval Index



Mapping Time: Very Low

Re-Localisation
Evaluation

Retrieve NN
Refine pose



„RelocNet: Continuous Metric Learning Relocalisation using Neural Nets", Balntas et al., ECCV'18

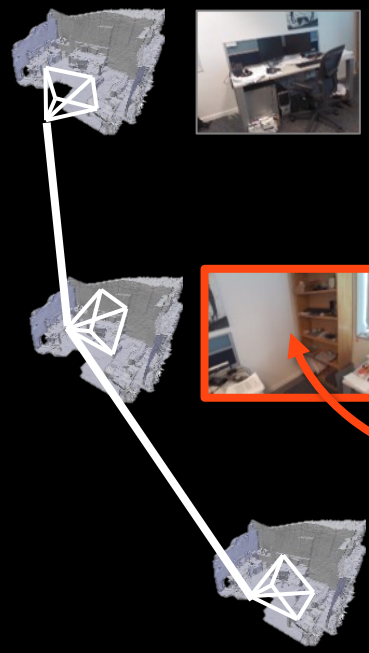
Preparation
Scene-Agnostic Training

Pre-Train Image Retrieval

Pre-Train Relative Pose Regressor

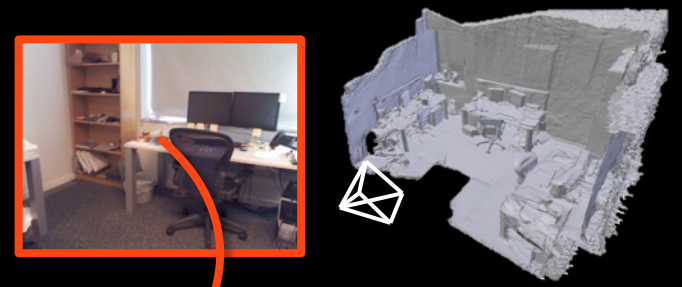
Mapping
Scene-Specific Training

Obtain Posed Images
Build Retrieval Index



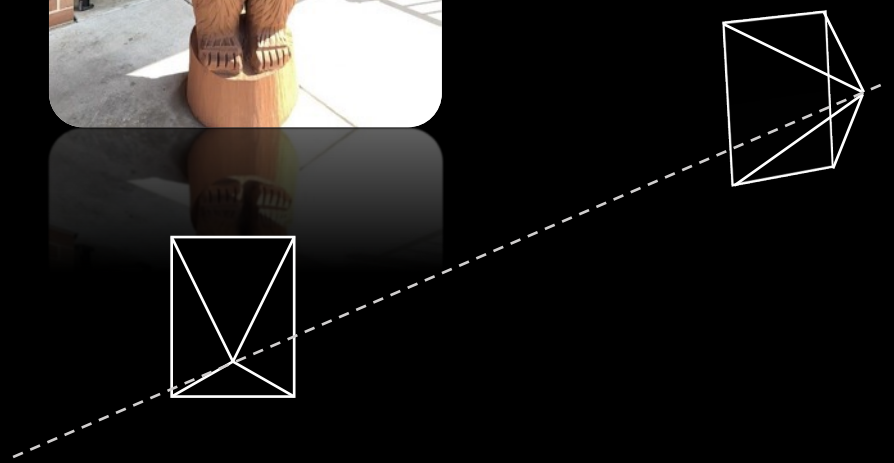
Re-Localisation
Evaluation

Retrieve NN
Refine pose



„RelocNet: Continuous Metric Learning Relocalisation using Neural Nets", Balntas et al., ECCV'18

Reference Image



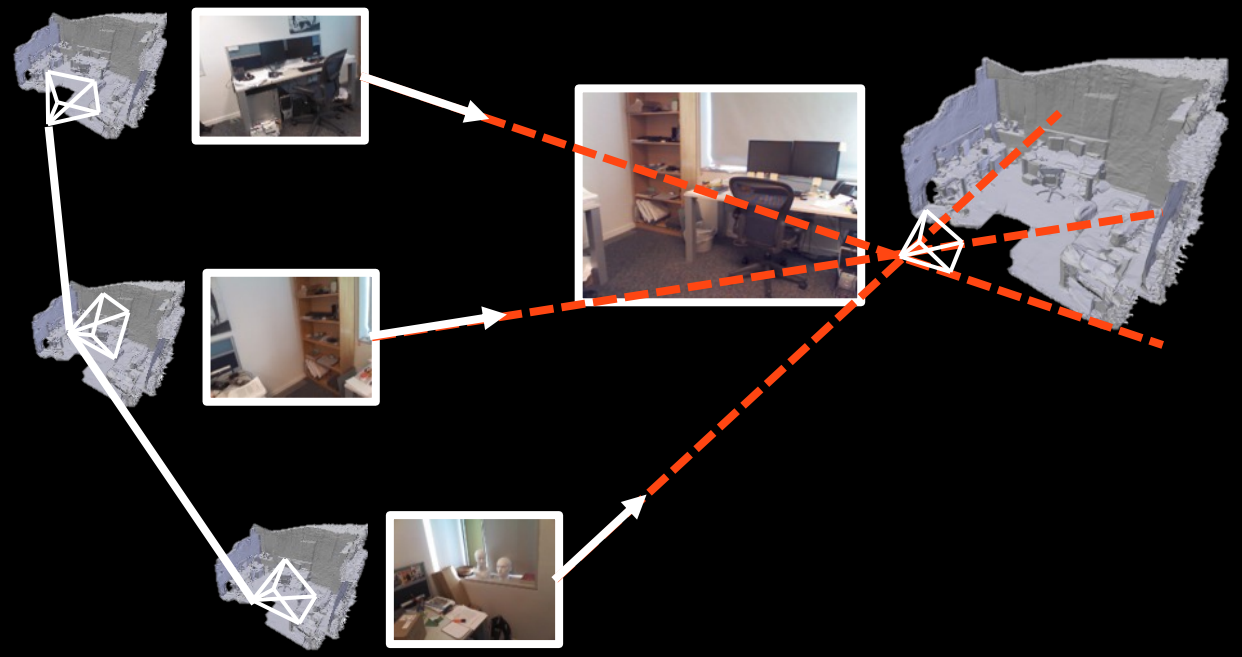
Query Image

RelocNet
Median error on 7Scenes

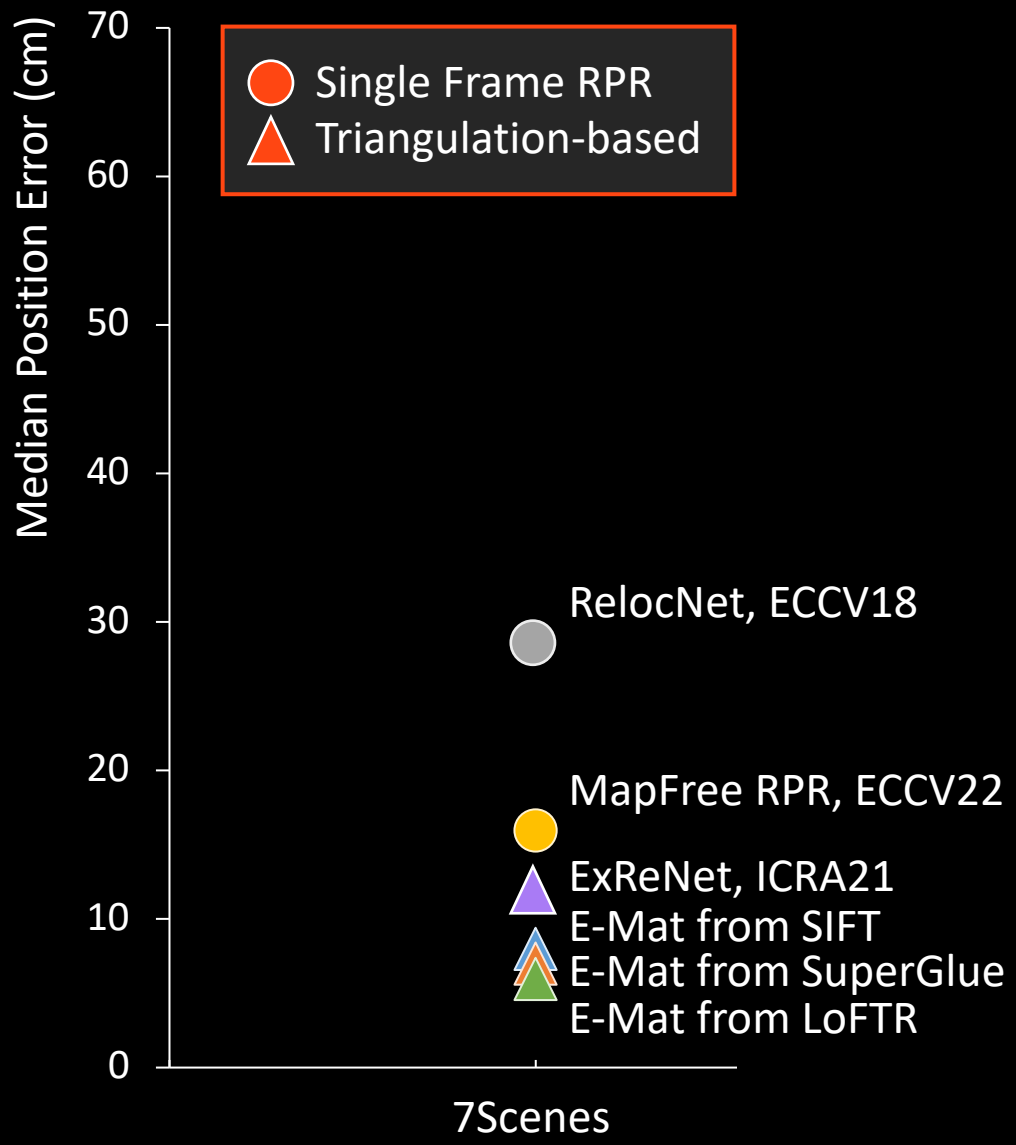
Train on 7Scenes: 21cm
Train on ScanNet: 29cm

RelocNet: Continuous Metric Learning Relocalisation using Neural Nets
Balntas et al., ECCV 2018

Preparation **Scene-Agnostic Training**	**Mapping** **Scene-Specific Training**	**Re-Localisation** Evaluation
Pre-Train Image Retrieval Pre-Train Relative Pose Regressor	Obtain Posed Images Build Retrieval Index	Retrieve NN Refine pose

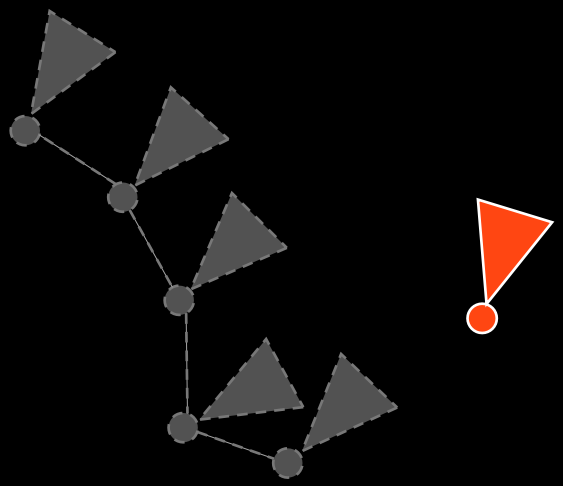


"To Learn or Not to Learn: Visual Localization from Essential Matrices", Zhou at al., ICRA'20



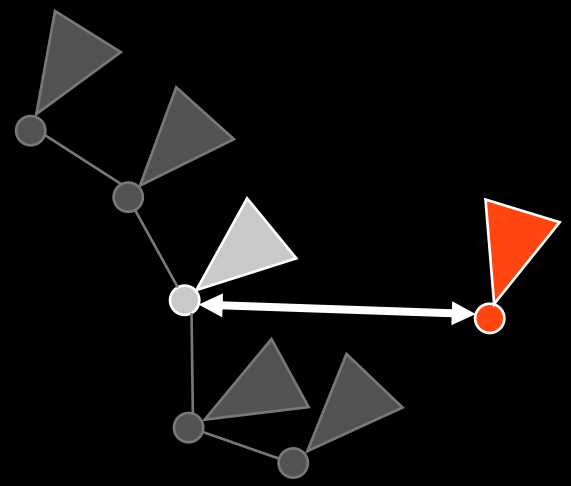
Data from Arnold et al., “Map-free Visual Relocalization: Metric Pose Relative to a Single Image”, ECCV22

Absolute Pose Regression



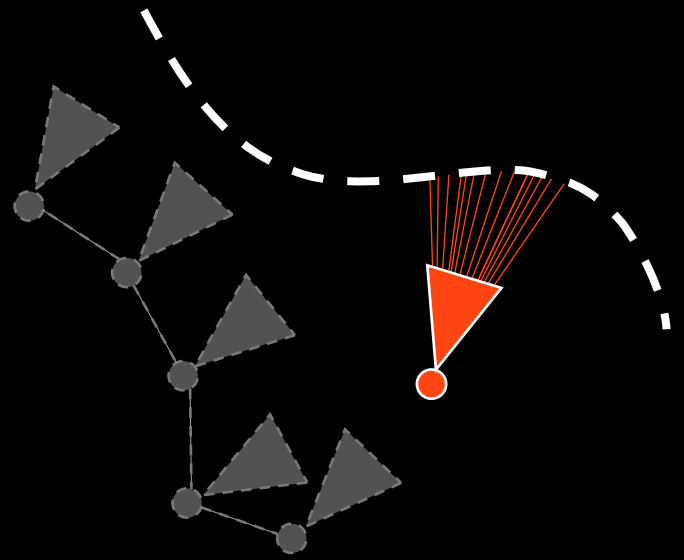
- Fast at query time
- Slow at mapping time
- Moderate memory demand
- Moderate accuracy

Relative Pose Regression



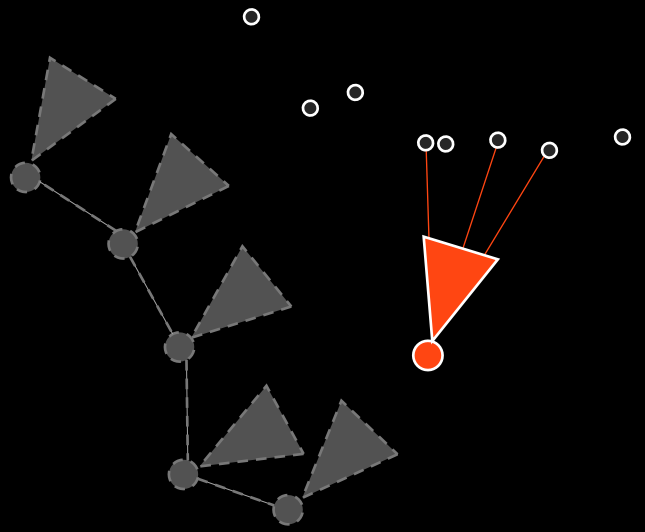
- Fast at query time
- Fast mapping time
- Low memory demand
- Moderate accuracy

Correspondence Regression

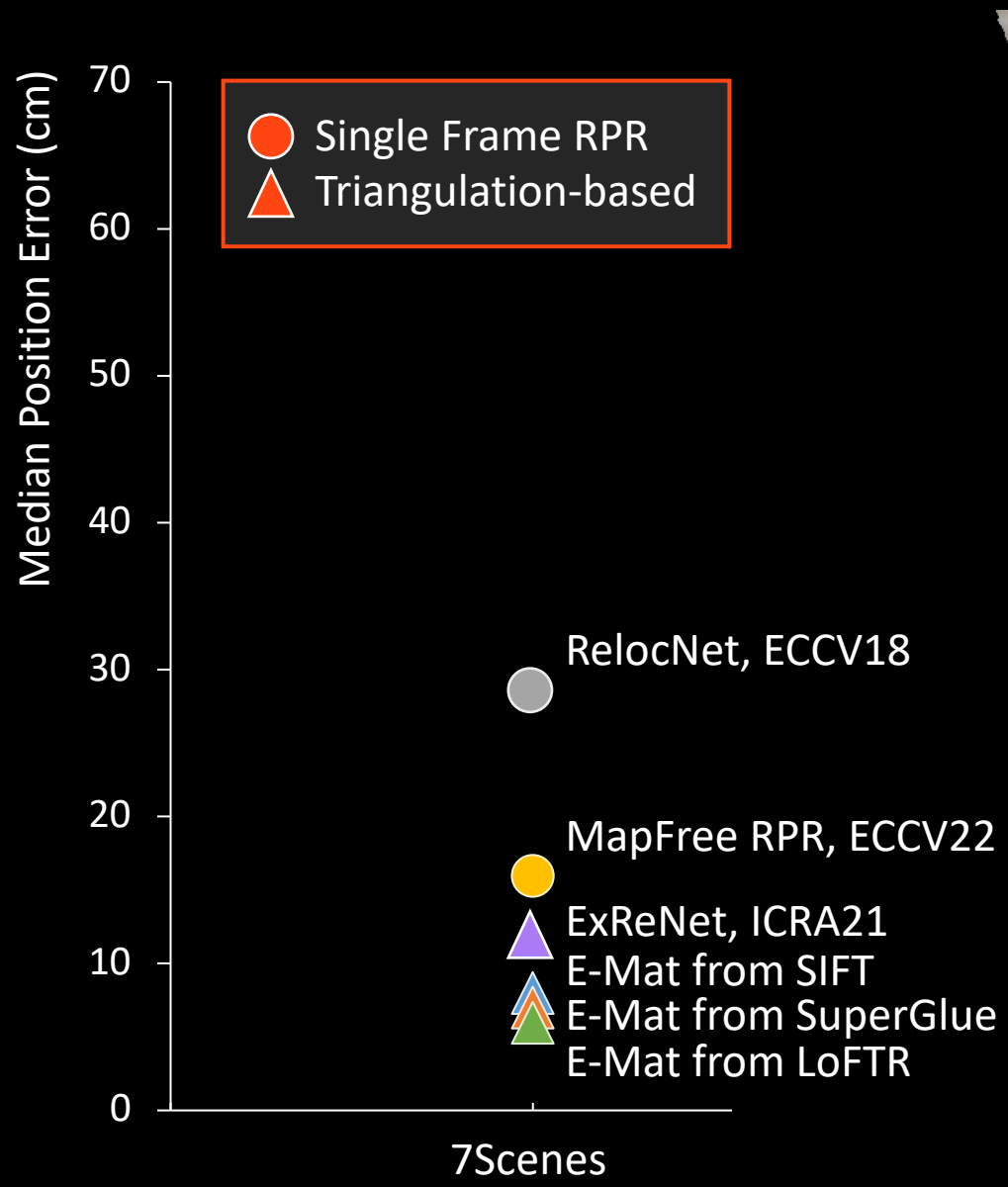


- Fast at query time
- Moderate mapping time
- Low memory demand
- High accuracy

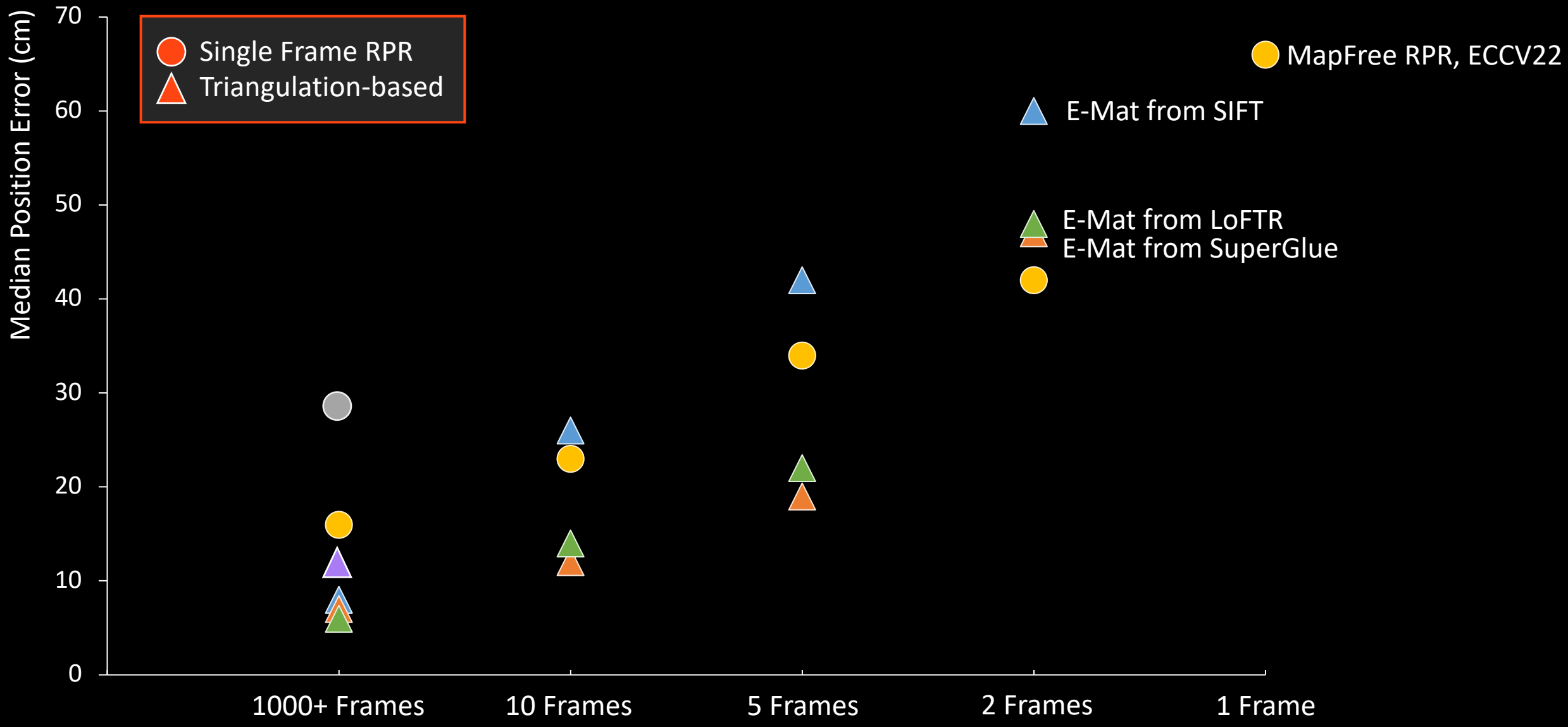
Feature Matching



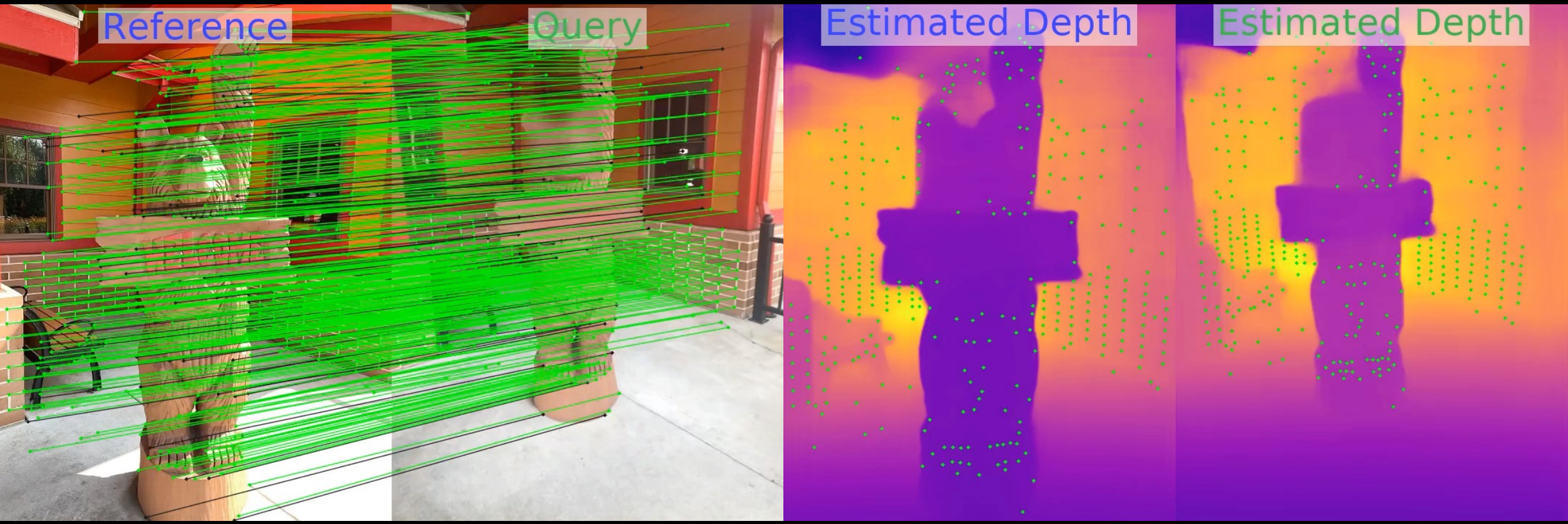
- Moderate query time
- Slow at mapping time
- High memory demand
- High accuracy



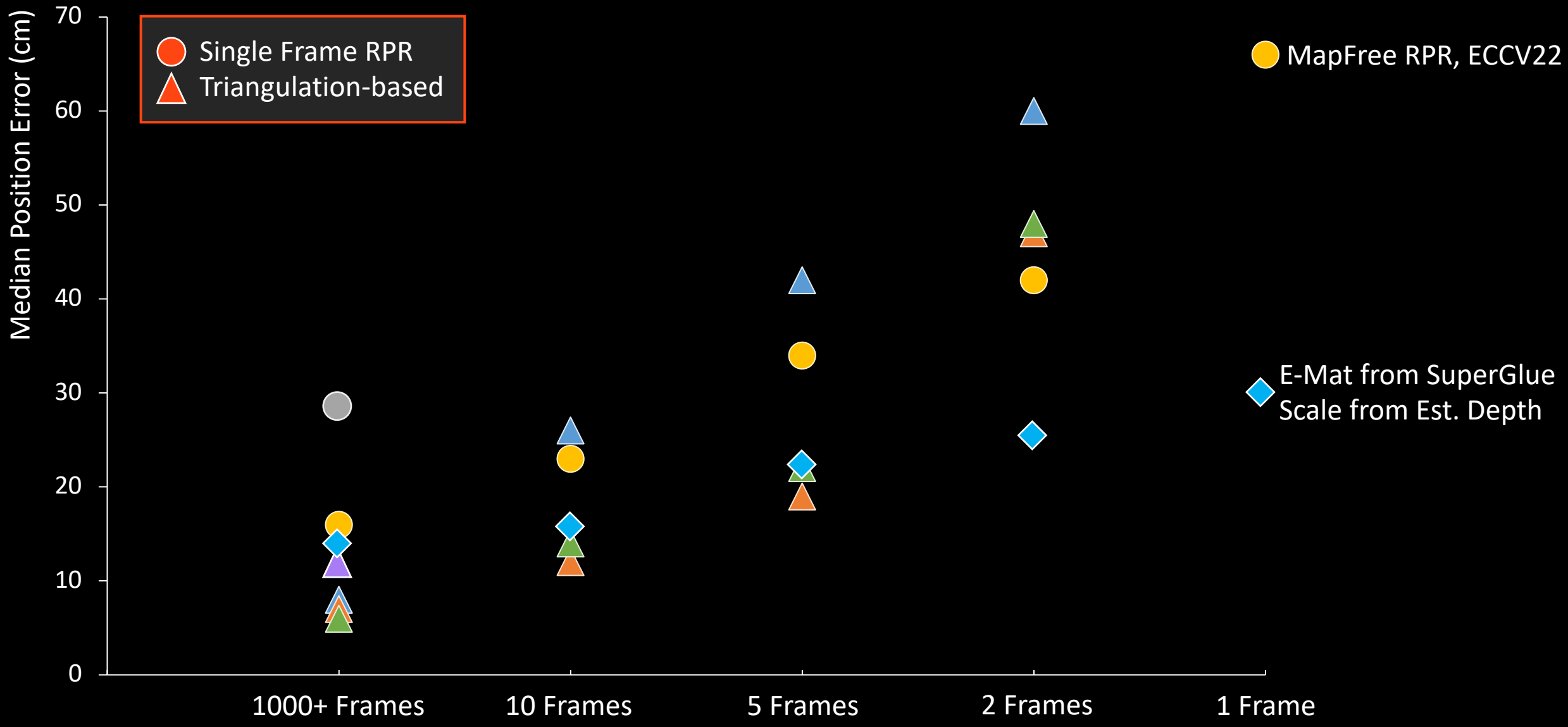
Data from Arnold et al., “Map-free Visual Relocalization: Metric Pose Relative to a Single Image”, ECCV22



Data from Arnold et al., “Map-free Visual Relocalization: Metric Pose Relative to a Single Image”, ECCV22

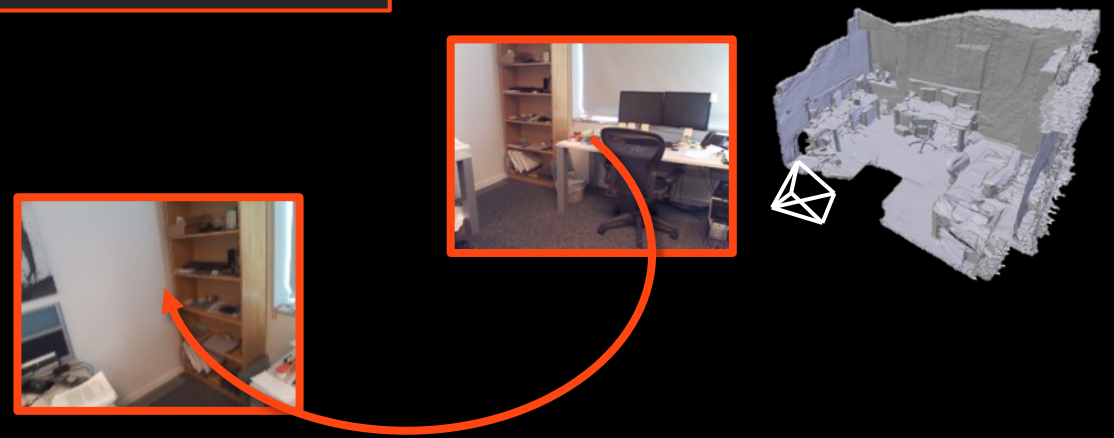


Reference Frame | Pseudo Ground Truth | DPT Depth + SuperGlue



Data from Arnold et al., “Map-free Visual Relocalization: Metric Pose Relative to a Single Image”, ECCV22

Preparation **Scene-Agnostic Training**	**Mapping** **Scene-Specific Training**	**Re-Localisation** Evaluation
Pre-Train Image Retrieval Pre-Train Relative Pose Regressor	~~Obtain Posed Images~~ ~~Build Retrieval Index~~ Shoot Reference Image **Mapping Time: Zero**	Estimate Relative Pose



"Map-free Visual Relocalization: Metric Pose Relative to a Single Image", ECCV22

Map-Free Relocalization

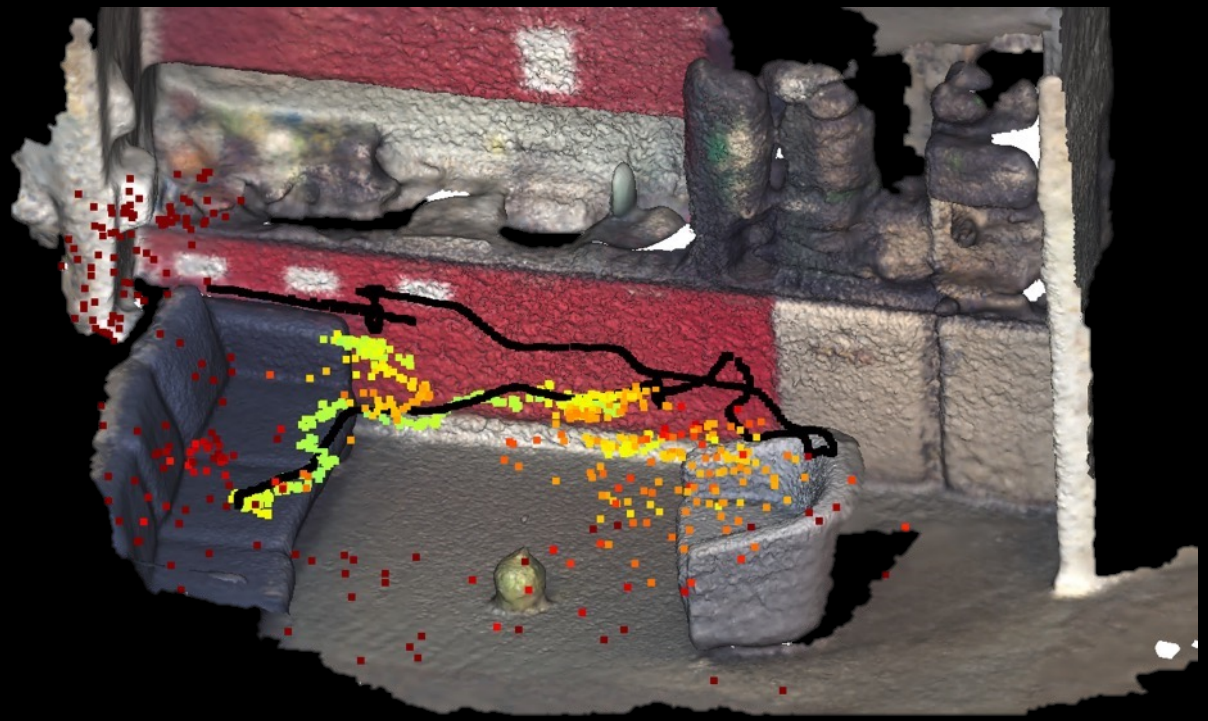
Reference Image

Query Images

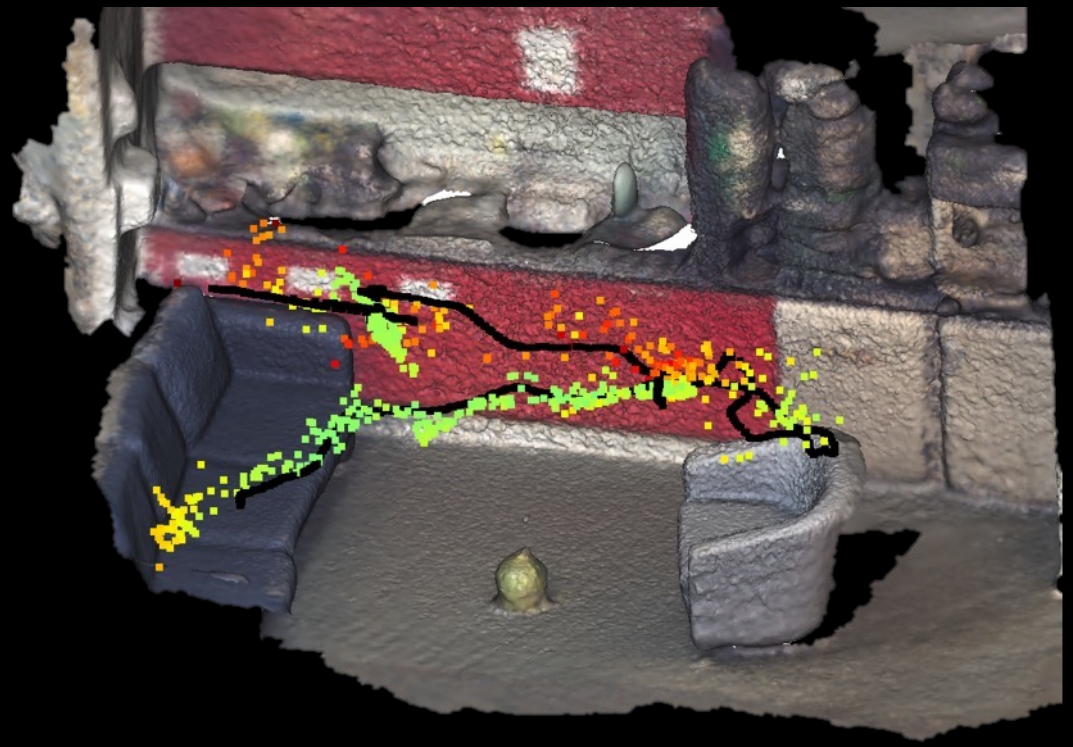


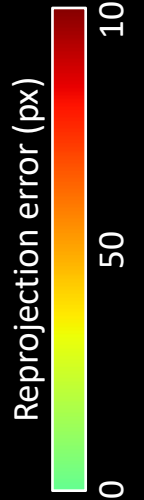


MapFree RPR, ECCV22



E-Mat from SuperGlue, Scale from Est. Depth







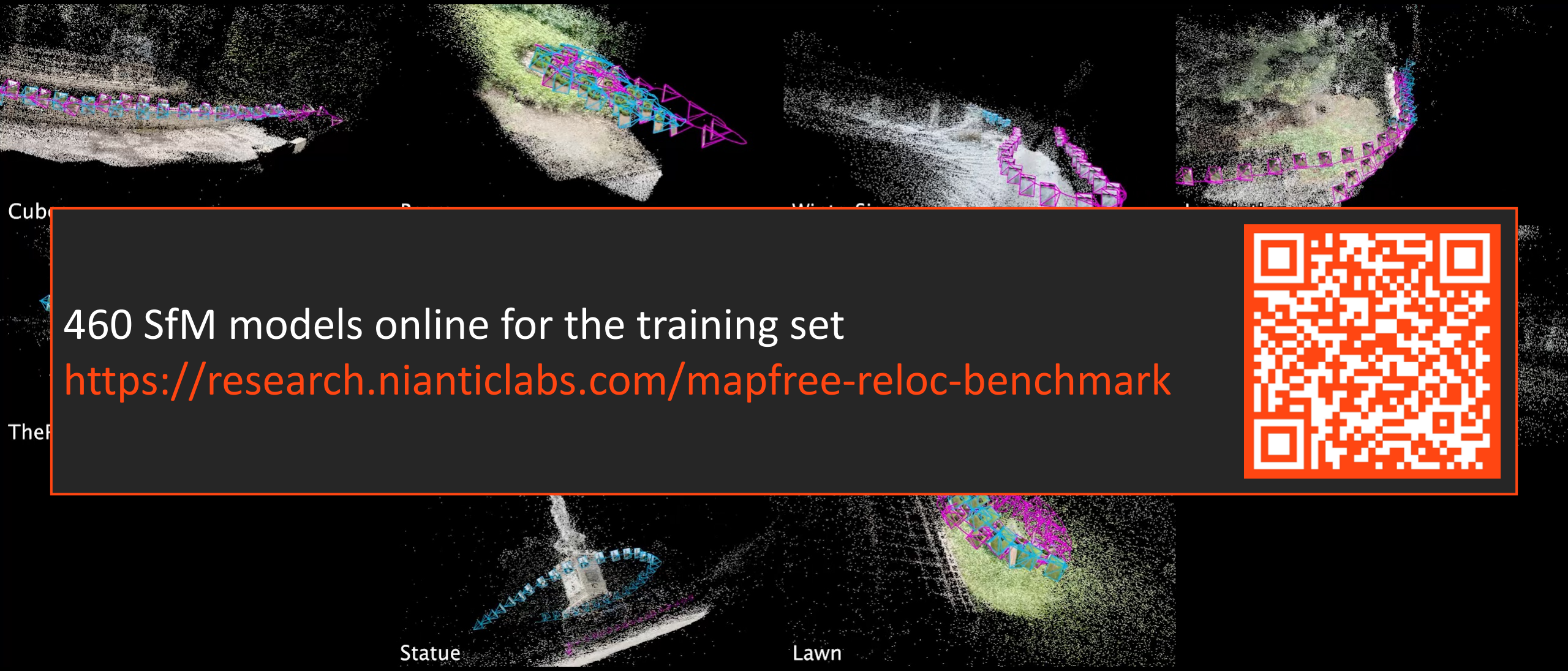




Area Guide
住居表示街区案内図

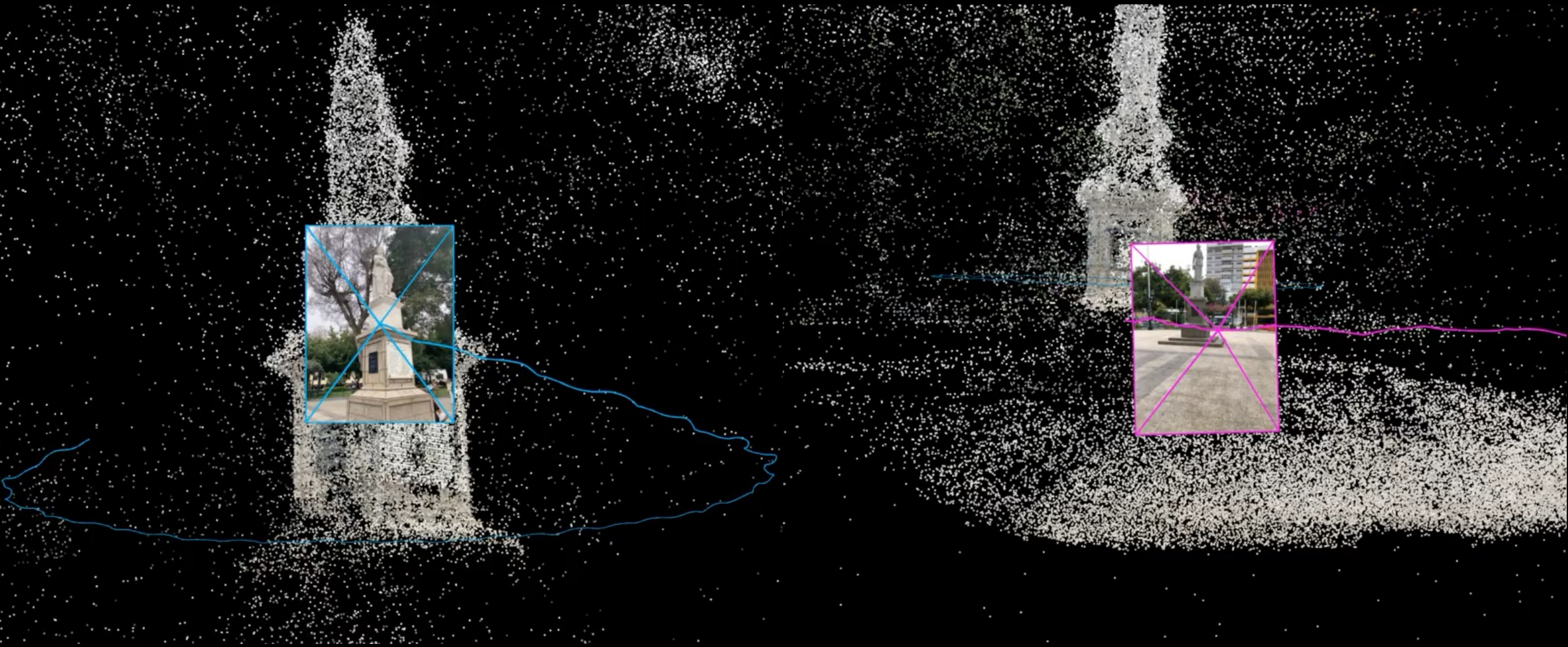


Niantic's Map-Free Dataset

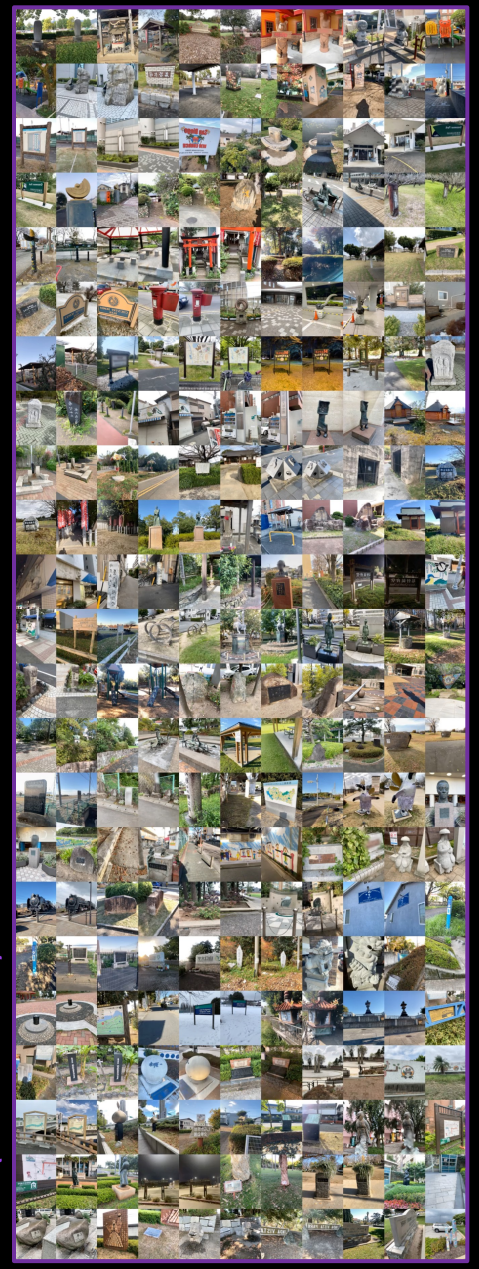


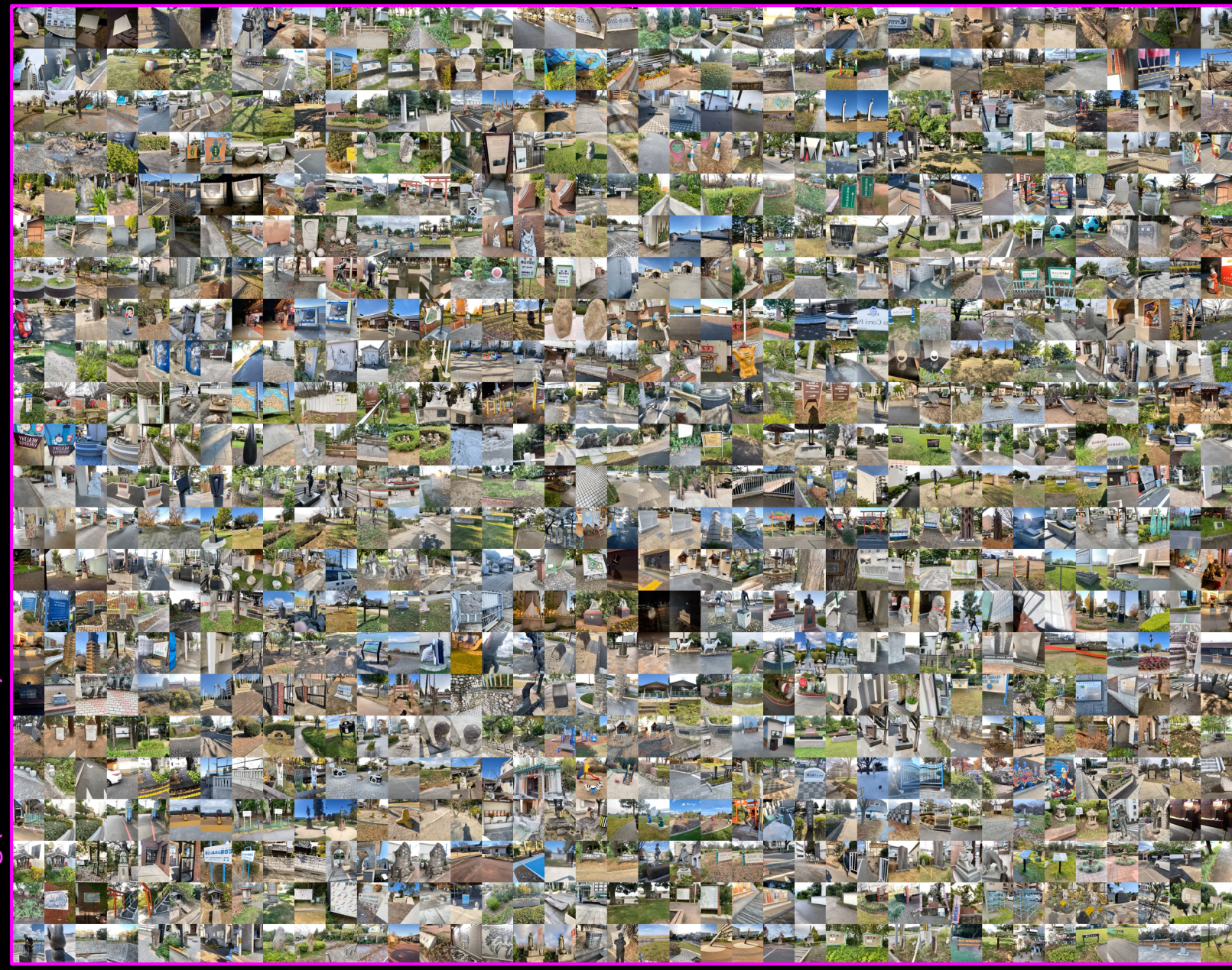
460 SfM models online for the training set

https://research.nianticlabs.com/mapfree-reloc-benchmark




Validation (65 Scenes)


Test (130 Scenes)


Training (460 Scenes)


Reference

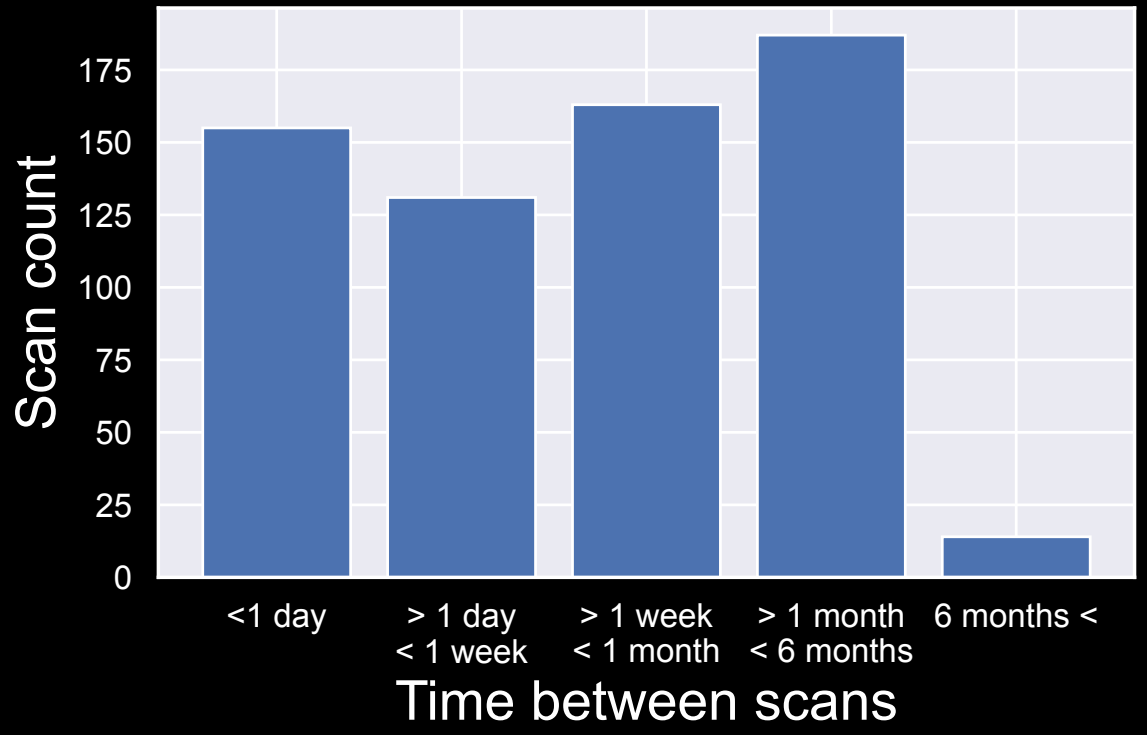

Query


Reference

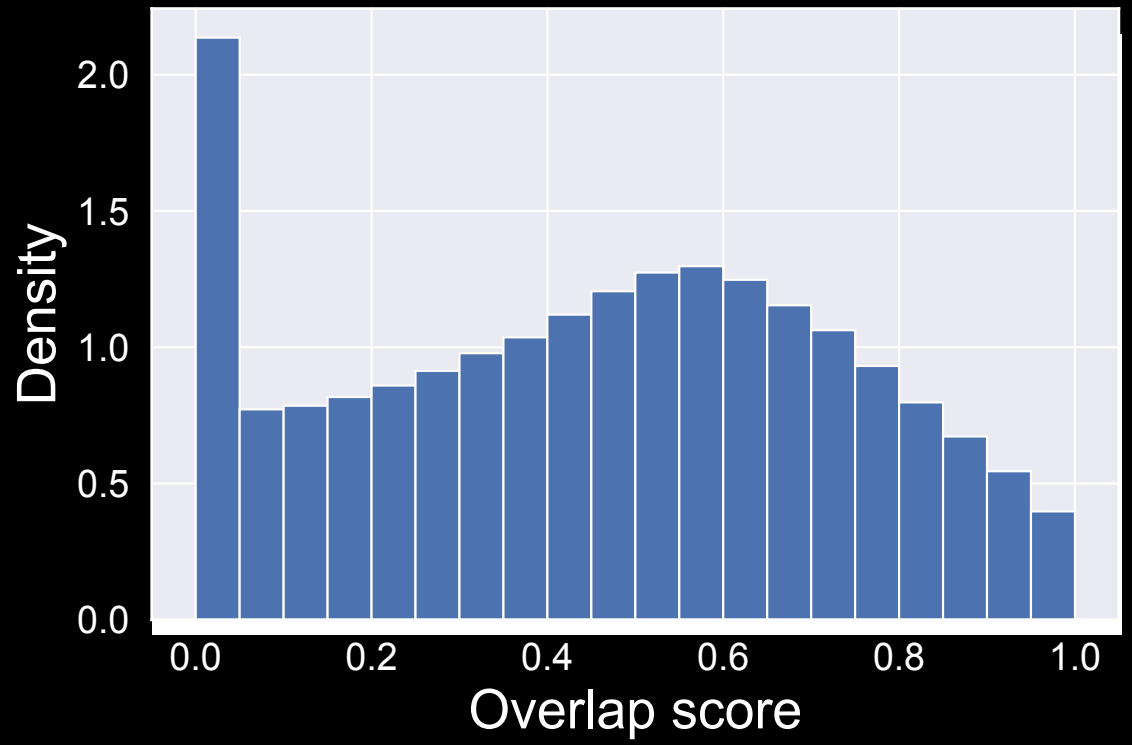

Query


Reference

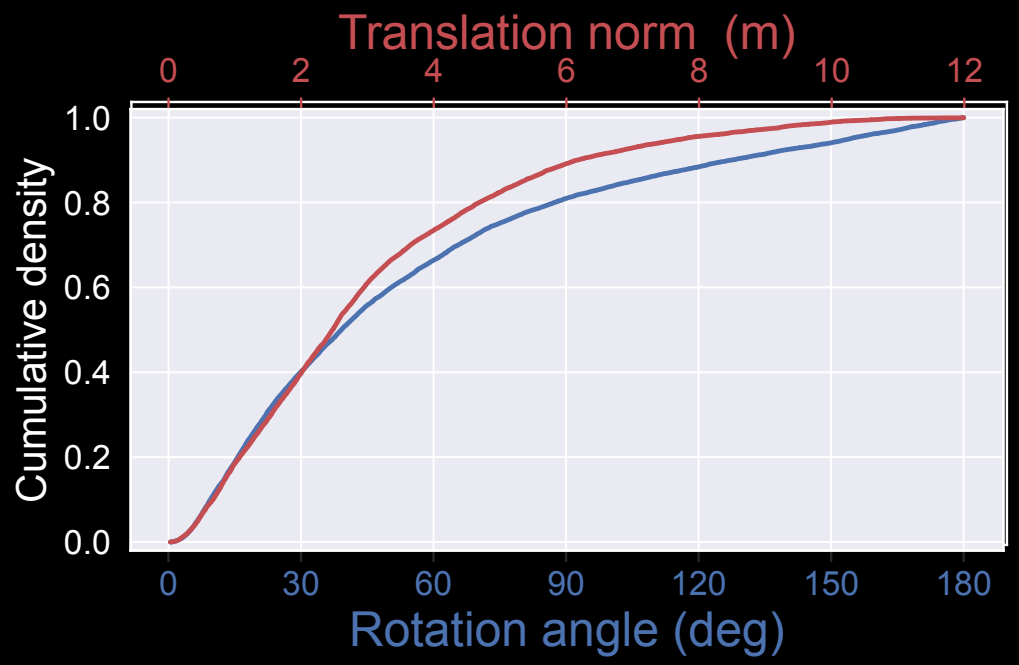

Query


Scan count
175
150
125
100
75
50
25
0
<1 day
> 1 day
< 1 week
> 1 week
< 1 month
> 1 month
< 6 months
6 months <
Time between scans

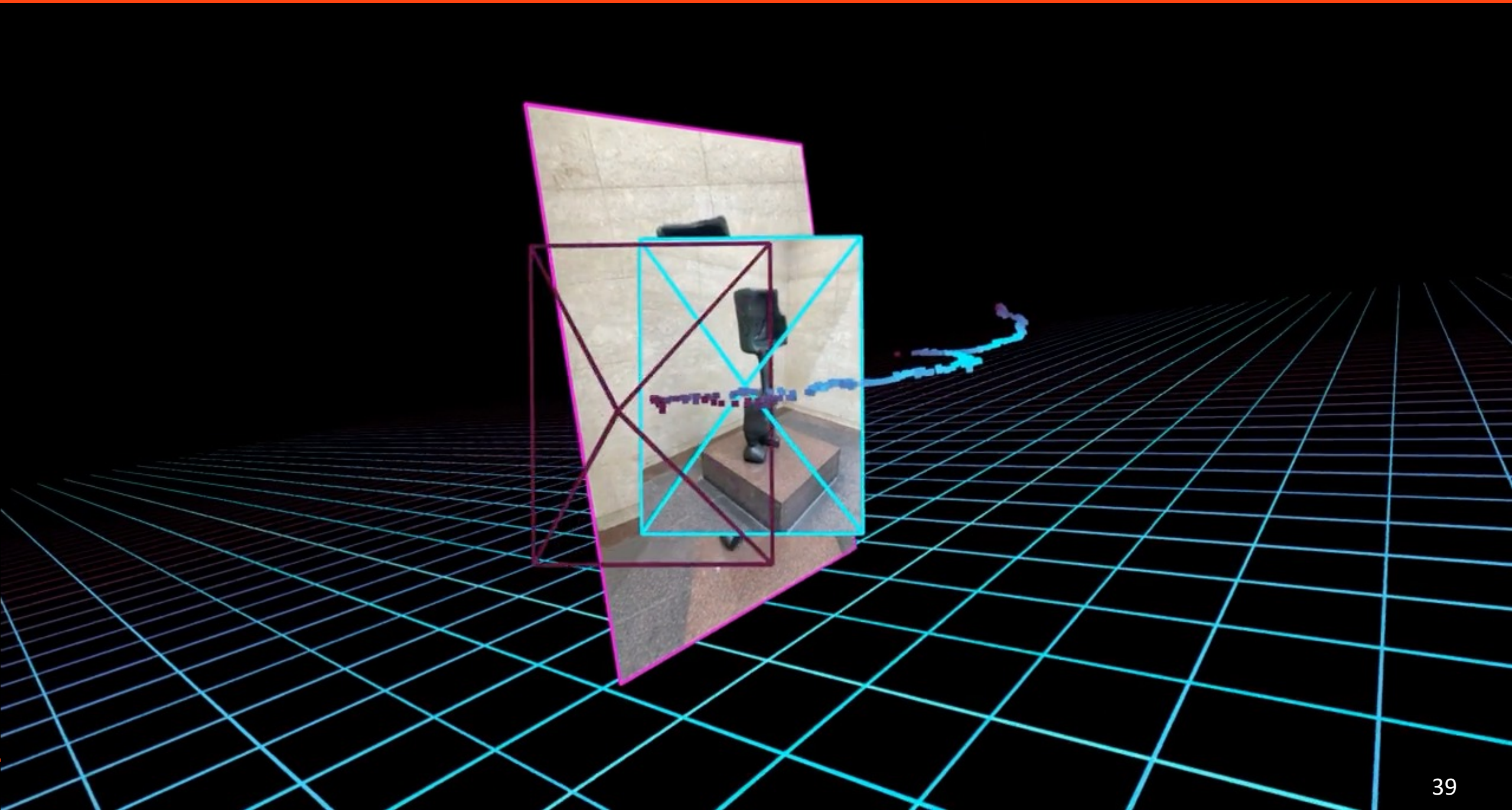

Reference
Query

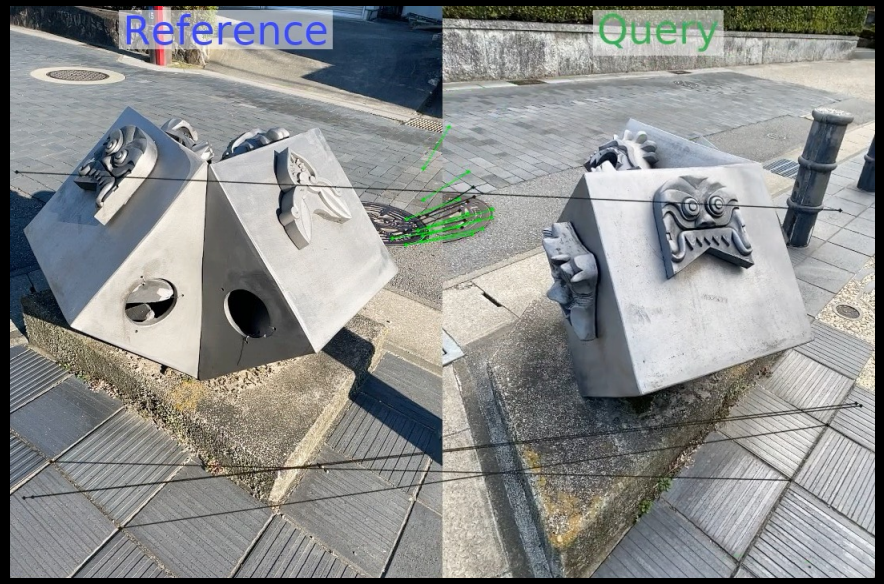

Density
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Overlap score

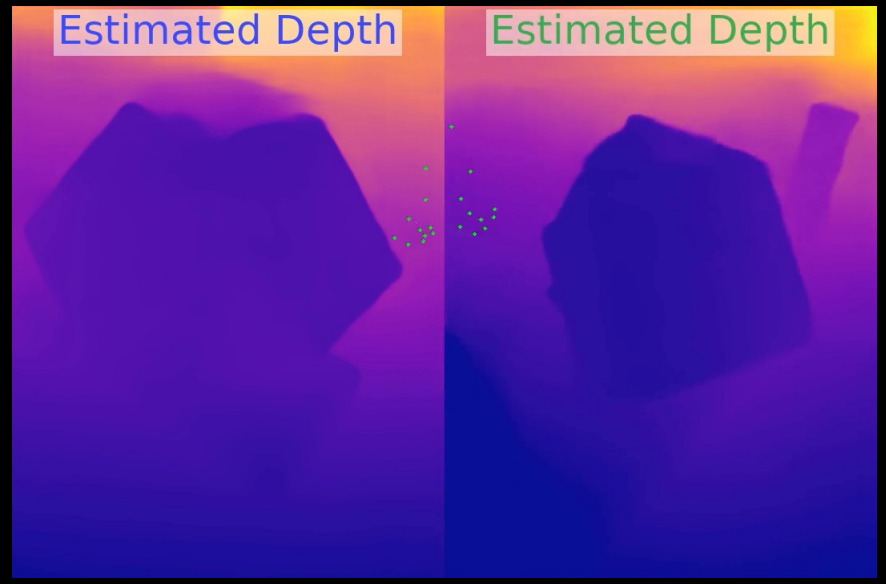

Reference
Query


Translation norm (m)
0
2
4
6
8
10
12
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
0
30
60
90
120
150
180
Rotation angle (deg)


Reference
Query



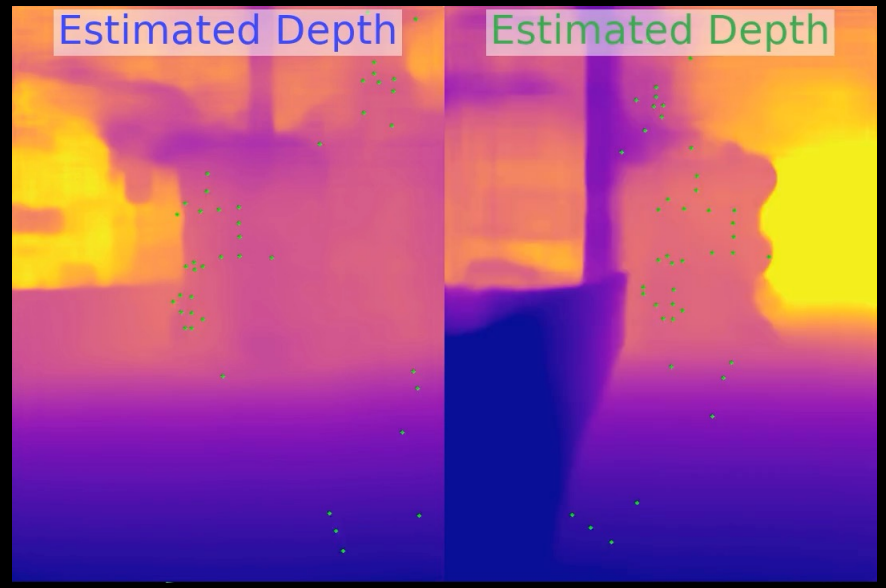




Reference Frame

Pseudo Ground Truth

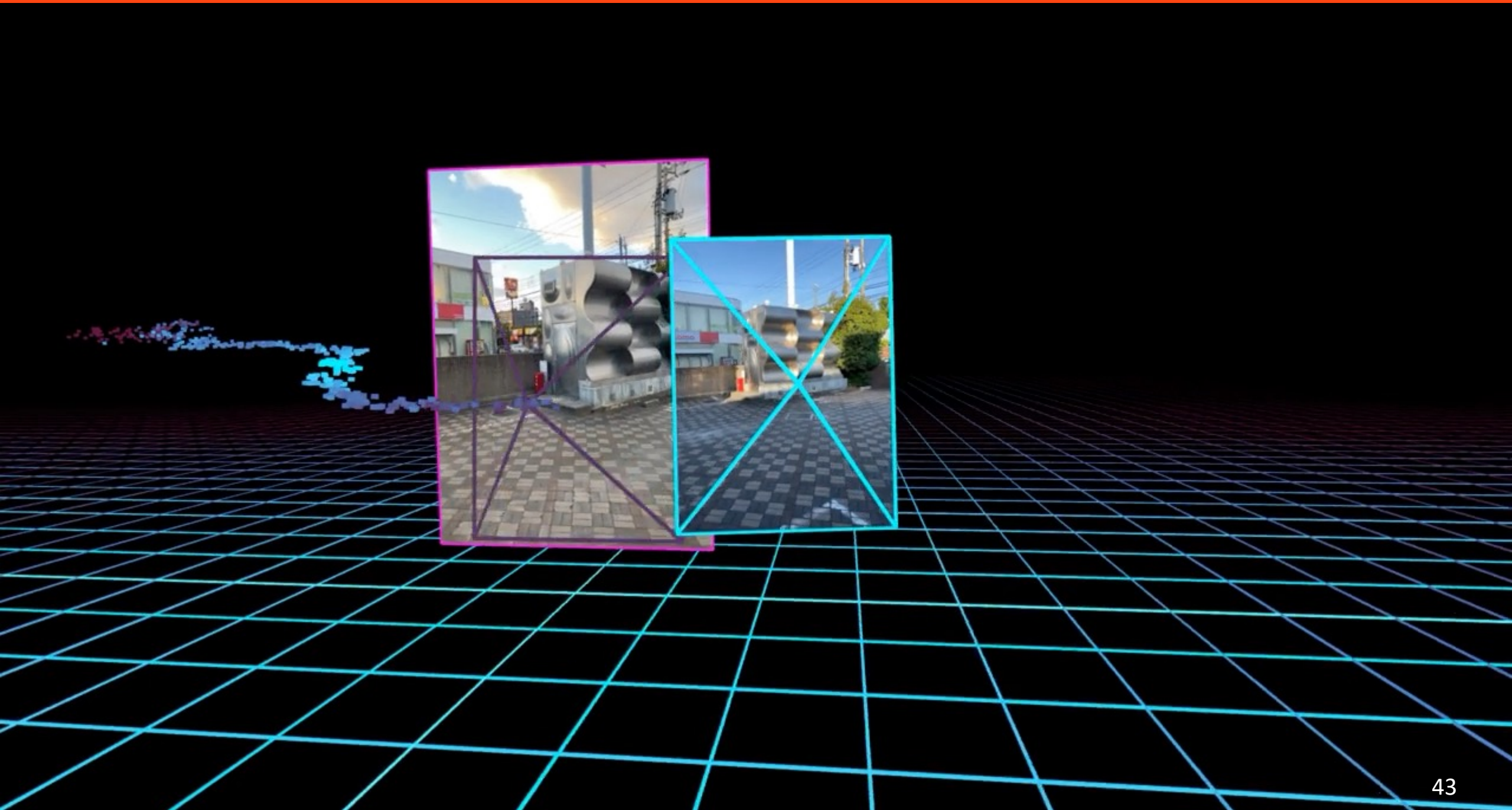
DPT Depth + SuperGlue

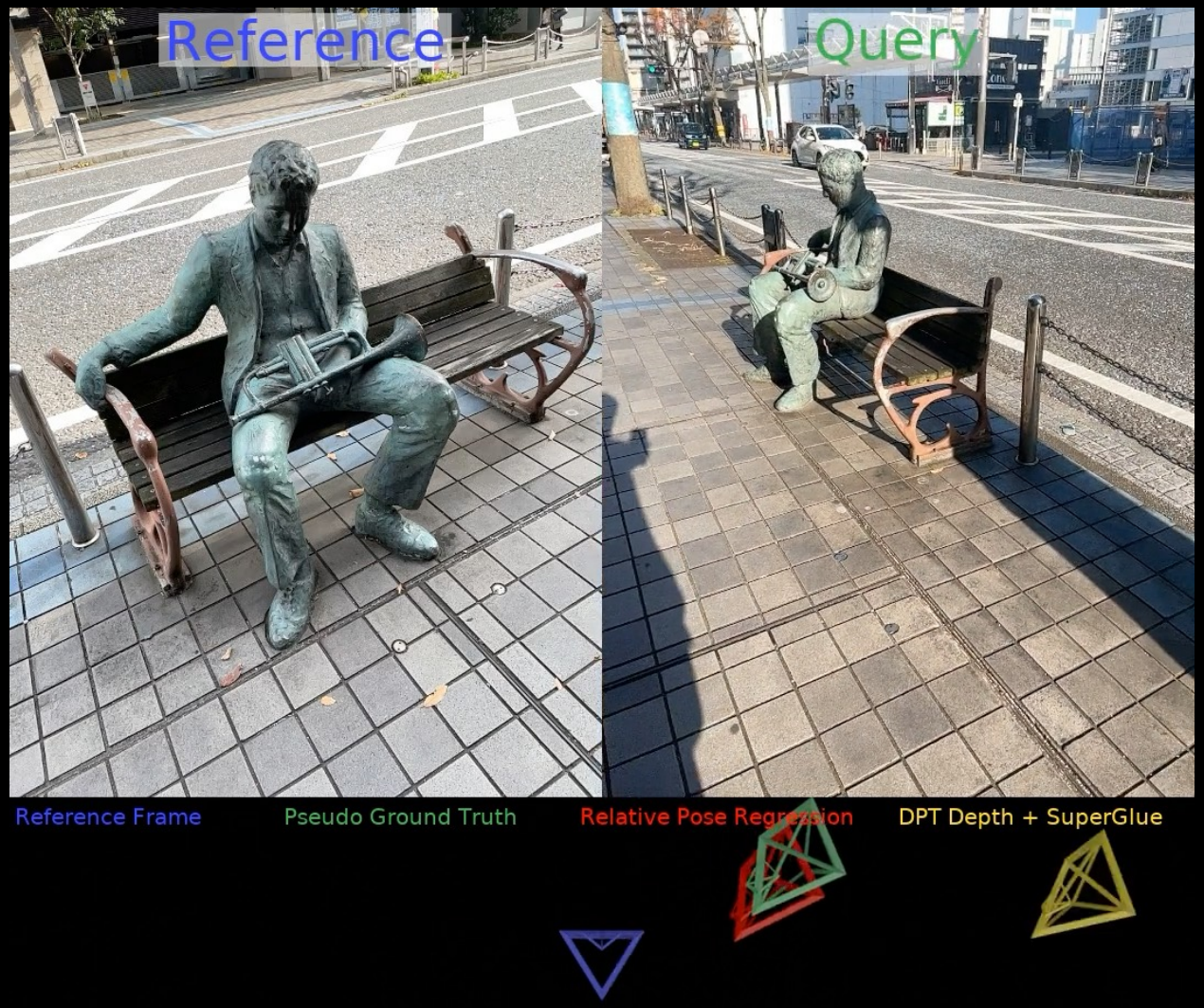


Reference Frame

Pseudo Ground Truth

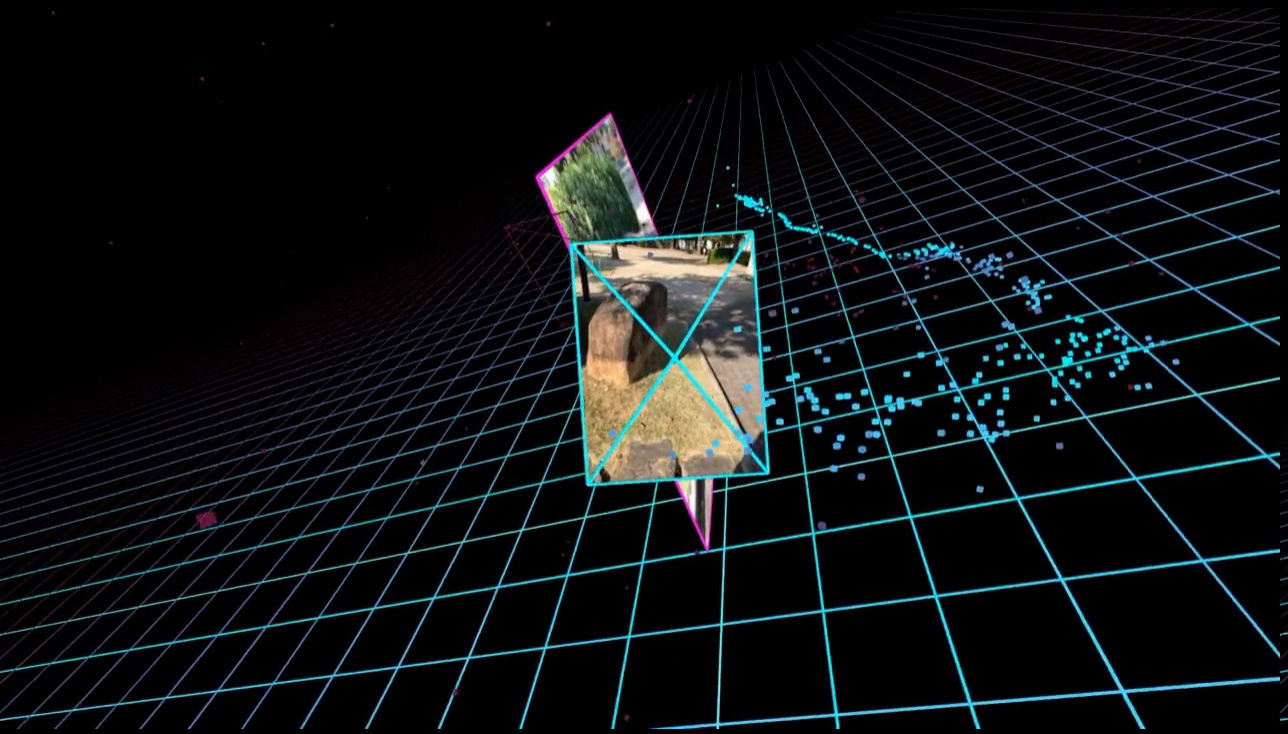
DPT Depth + SuperGlue



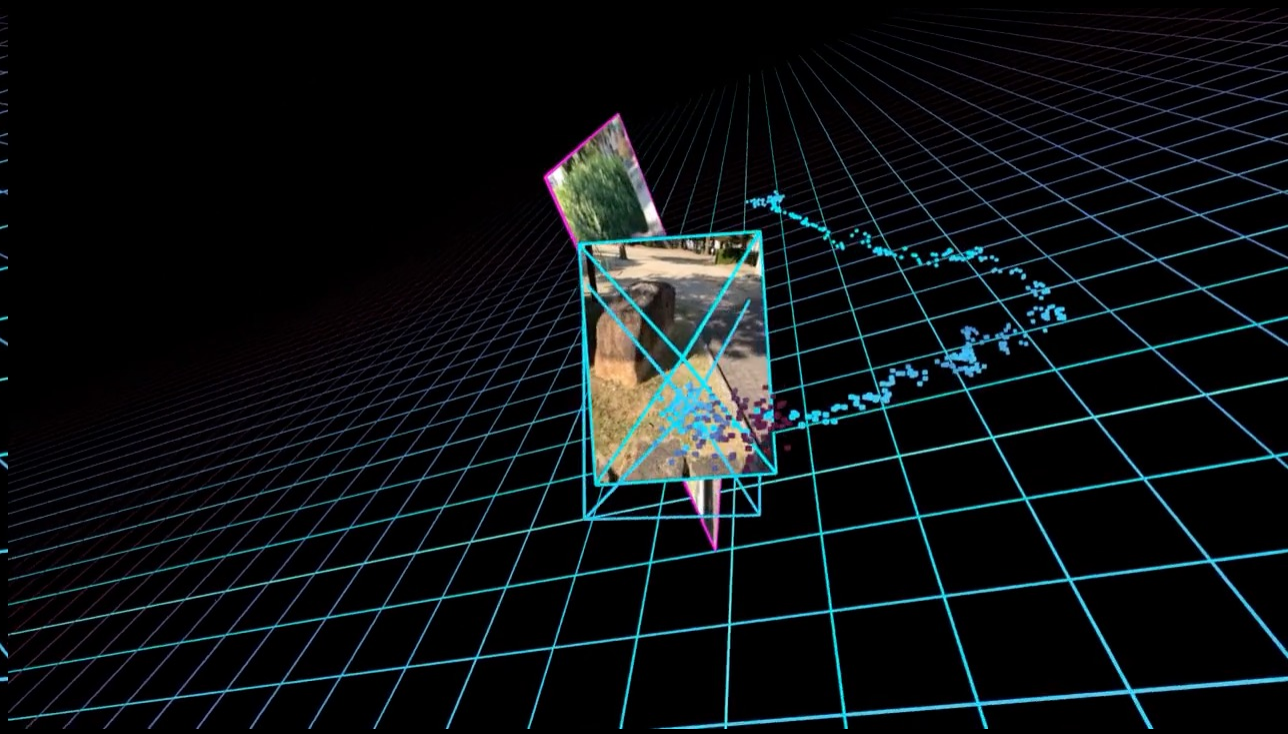

Reference
Query
Reference Frame
Pseudo Ground Truth
Relative Pose Regression
DPT Depth + SuperGlue


Reference
Query
Reference Frame
Pseudo Ground Truth
Relative Pose Regression
DPT Depth + SuperGlue

SuperGlue + Depth



Map-Free RPR



Reference Frame Pseudo Ground Truth Relative Pose Regression DPT Depth + SuperGlue

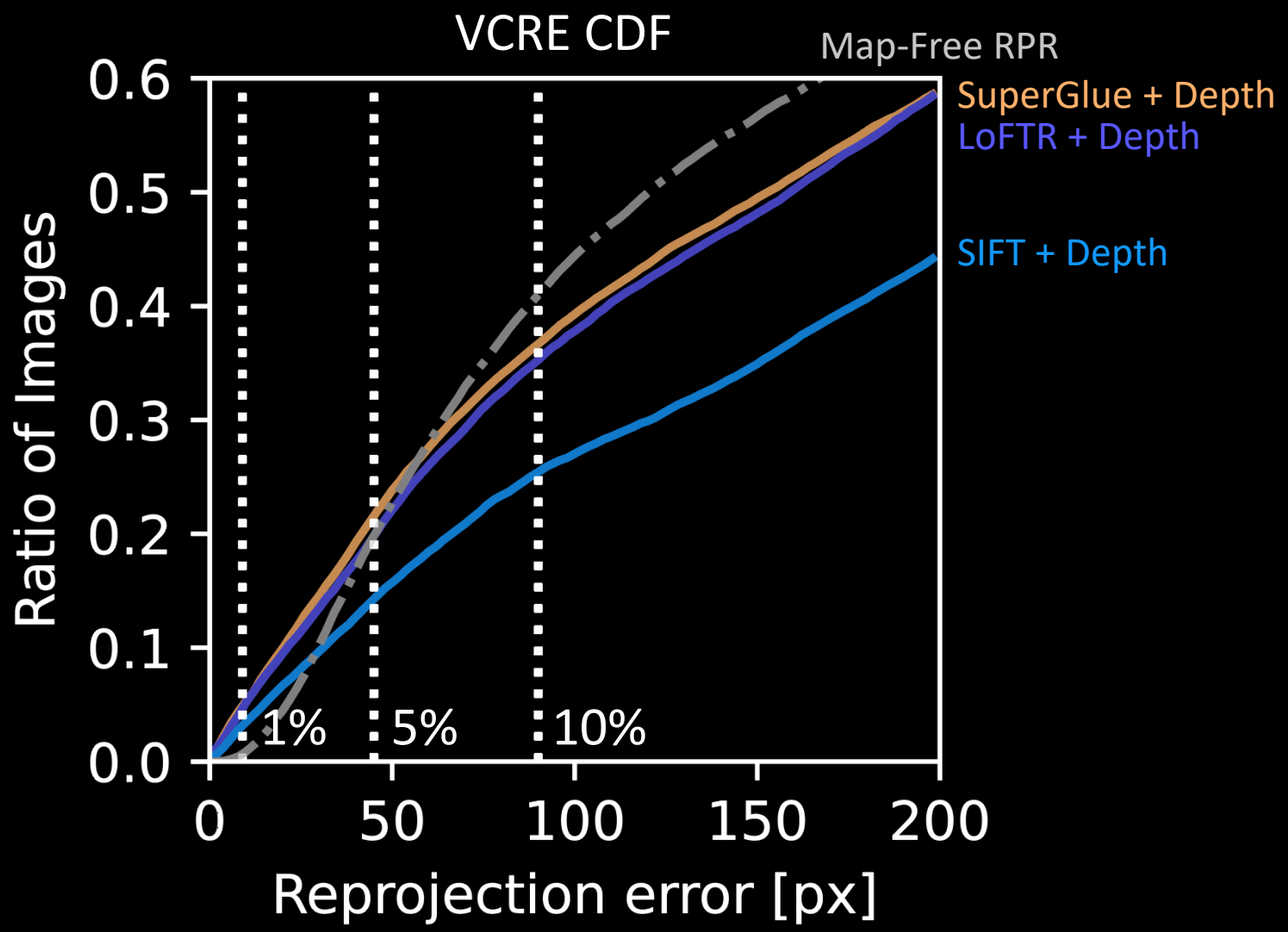
Reference view

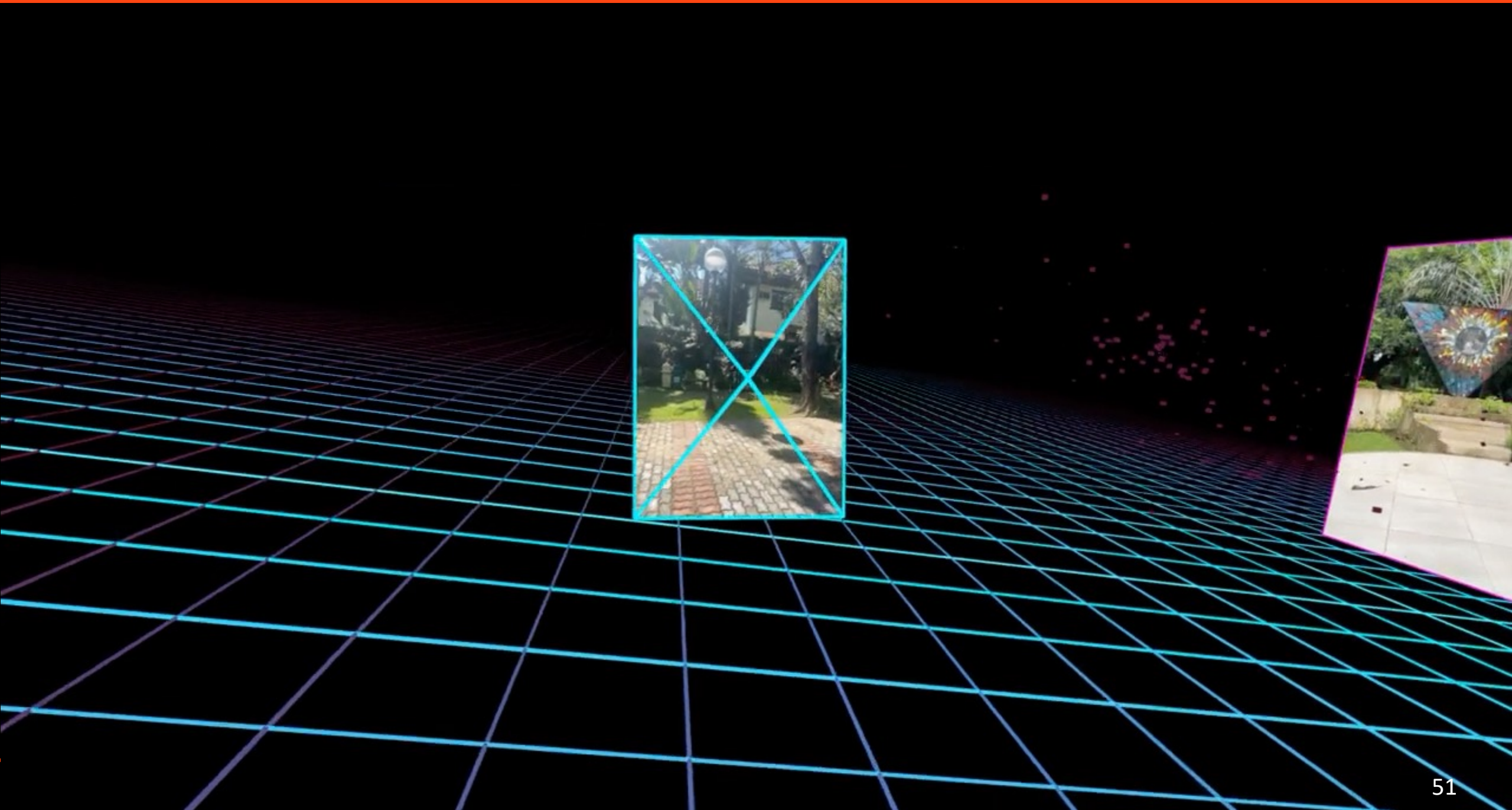
Ground Truth

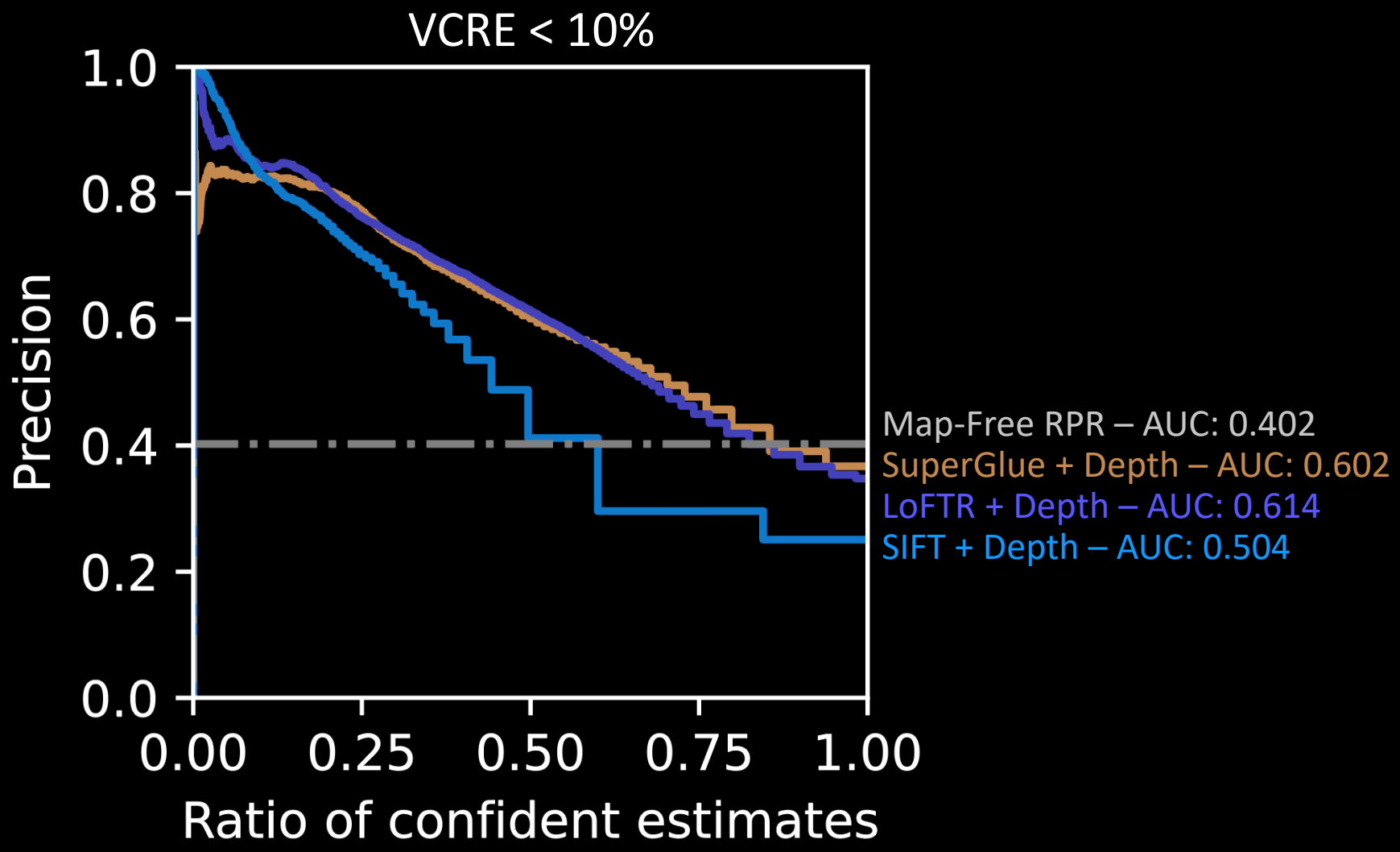
Relocalized with

Feature Matching + D.Scale

All Top


VCRE CDF
Map-Free RPR
SuperGlue + Depth
LoFTR + Depth
SIFT + Depth
Ratio of Images
0.0
0.1
0.2
0.3
0.4
0.5
0.6
1%
5%
10%
0
50
100
150
200
Reprojection error [px]




VCRE < 10%
Precision
1.0
0.8
0.6
0.4
0.2
0.0
0.00
0.25
0.50
0.75
1.00
Ratio of confident estimates
Map-Free RPR – AUC: 0.402
SuperGlue + Depth – AUC: 0.602
LoFTR + Depth – AUC: 0.614
SIFT + Depth – AUC: 0.504

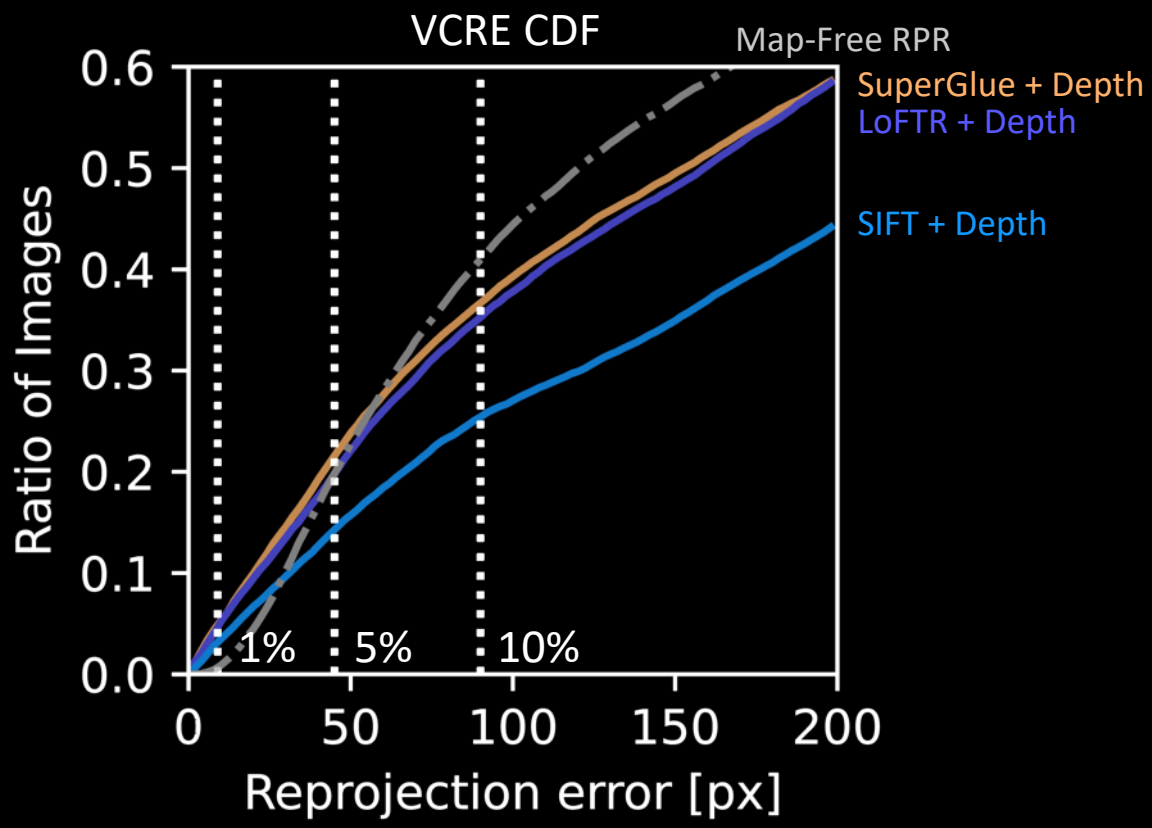
Evaluation Leaderboard

Method	AUC (VCRE < 90px)	Precision (VCRE < 90px)	Median Reproj. Error (px)	AUC (Err < 25cm, 5°)	Precision (Err < 25cm, 5°)	Median Trans. Error (m)	Median Rot. Error (°)	Estimates for % of frames
DPT-KITTI & LoFTR (PnP)	0.618	33.4%	166.7	0.269	9.8%	2.31	39.4	100.0
DPT-KITTI & LoFTR (Ess.Mat. + D.Scale)	0.614	34.7%	167.6	0.346	15.4%	1.98	30.5	100.0
DPT-KITTI & SuperGlue (Ess.Mat. + D.Scale)	0.602	36.1%	160.3	0.346	16.8%	1.88	25.4	100.0
DPT-KITTI & SuperGlue (PnP)	0.598	36.0%	156.9	0.252	10.7%	2.10	32.5	100.0
DPT-KITTI & SIFT (Ess.Mat. + D.Scale)	0.504	25.0%	222.8	0.252	10.3%	2.93	61.4	100.0
DPT-KITTI & SIFT (PnP)	0.468	25.1%	192.0	0.190	7.8%	3.35	63.7	100.0
RPR [R(6D) +t]	0.402	40.2%	147.1	0.060	6.0%	1.68	22.5	100.0
RPR [3D-3D]	0.387	38.7%	148.7	0.060	6.0%	1.69	22.9	100.0
RPR [R(α, β, γ) + s · t(θ, ϕ)]	0.354	35.4%	166.3	0.105	10.5%	1.83	23.2	100.0

Dataset, Baselines, Evaluation Code, Leaderboard

https://research.nianticlabs.com/mapfree-reloc-benchmark

Pose Estimation
Beyond Feature Matching



https://research.nianticlabs.com/mapfree-reloc-benchmark