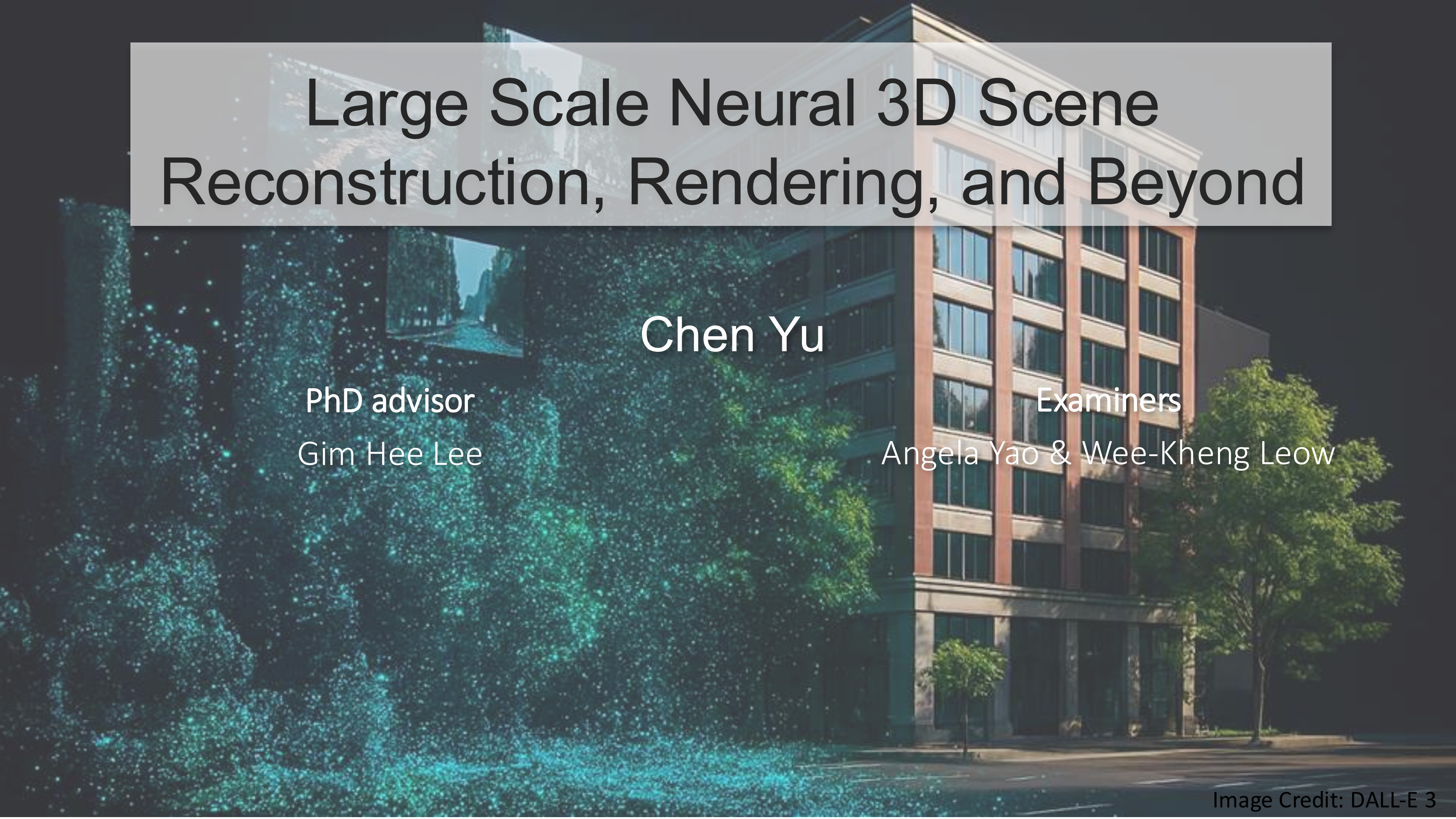


Large Scale Neural 3D Scene Reconstruction, Rendering, and Beyond

Chen Yu

PhD advisor
Gim Hee Lee

Examiners
Angela Yao & Wee-Kheng Leow

Image Credit: DALL-E 3

NUS National University of Singapore | School of Computing

CVRP

Overview of My PhD Research

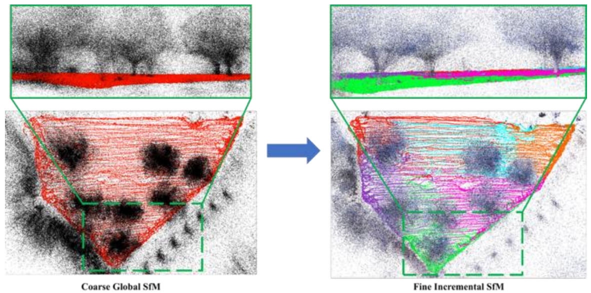
AdaSfM [ICRA 2023]

DBARF [CVPR 2023]

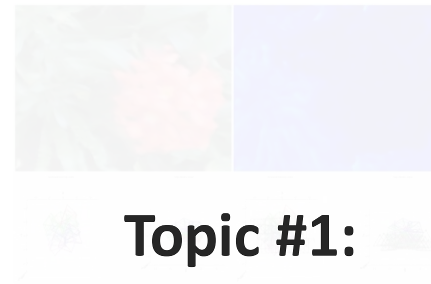
DOGS [NeurIPS 2024]

DReg-NeRF [ICCV 2023]

Overview of My PhD Research



AdaSfM [ICRA 2023]



Topic #1:

Fast and Distributed Neural 3D Reconstruction

DOGS [NeurIPS 2024]

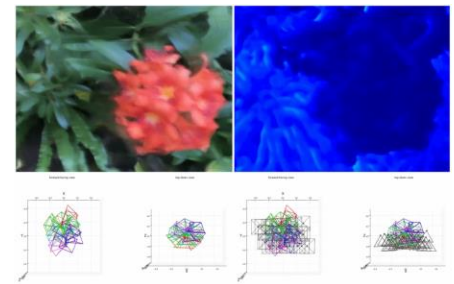
DReg-NeRF [ICCV 2023]

Overview of My PhD Research

AdaSfM [ICRA 2023]

Topic #2:

Robust Neural Rendering via
Pose-Aware Learning



DBARF [CVPR 2023]

DOGS [NeurIPS 2024]

DReg-NeRF [ICCV 2023]

Overview of My PhD Research

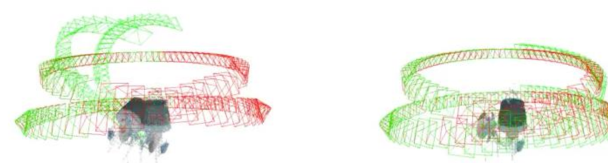
AdaSfM [ICRA 2023]

DBARF [CVPR 2023]

Topic #3:

Geometric Understanding with Neural Representation

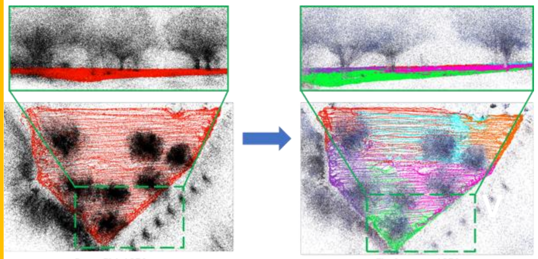
DOGS [NeurIPS 2024]



DReg-NeRF [ICCV 2023]

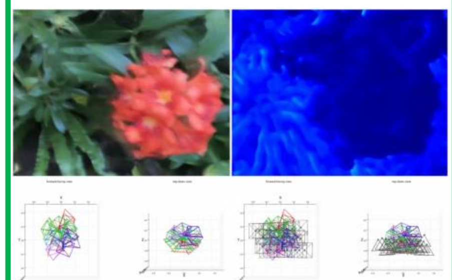
This Thesis

Develop Distributed System and Neural Scene Representations for 3D Reconstruction, Rendering, and Geometry Understanding

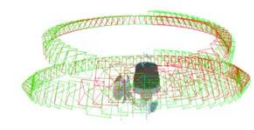


AdaSfM
ICRA 2023

DOGS
NeurIPS 2024

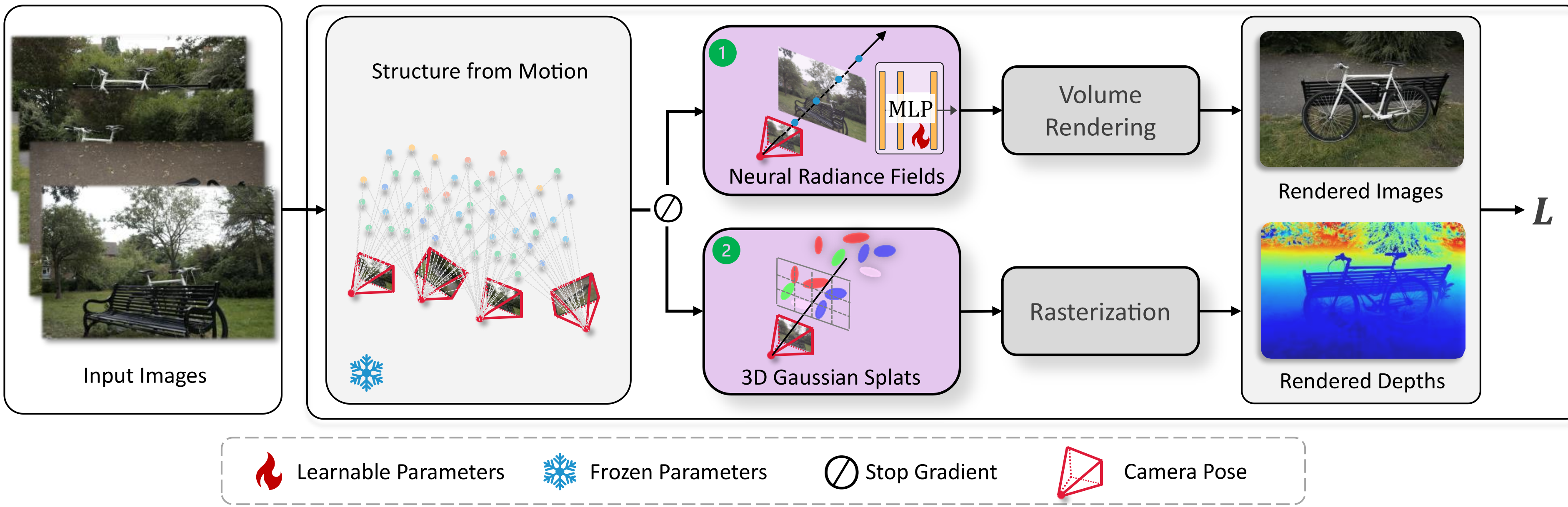


DBARF
CVPR 2023



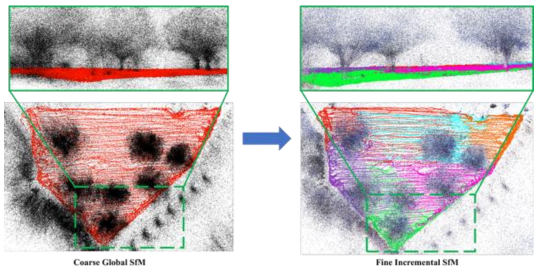
DReg-NeRF
ICCV 2023

The Modern Neural 3D Reconstruction System



Part #1

Develop Distributed System and Neural Scene Representations for 3D Reconstruction, Rendering, and Geometry Understanding



AdaSfM
ICRA 2023

DOGS
NeurIPS 2024

DBARF
CVPR 2023

DReg-NeRF
ICCV 2023

Yu Chen, *el, al*, AdaSfM: From Coarse Global to Fine Incremental Adaptive Structure from Motion. ICRA 2023

Key Challenges

Structure from Motion

- Reconstruct majorly from **2D images**
- Reconstruct 3D scenes **beyond single server**
- Reconstruct 3D scenes **at arbitrary scale**

Works well on small scale scenes, but **fragile** and slowly on **wild large-scale** areas

Seminal Papers of SfM

Photo Tourism

COLMAP

- Noah Snavely, Steven M. Seitz, Richard Szeliski, **Photo tourism: Exploring photo collections in 3D**. TOG 2006
- Johannes L. Schonberger, Jan-Michael Frahm. **Structure from Motion Revisited**. CVPR 2016

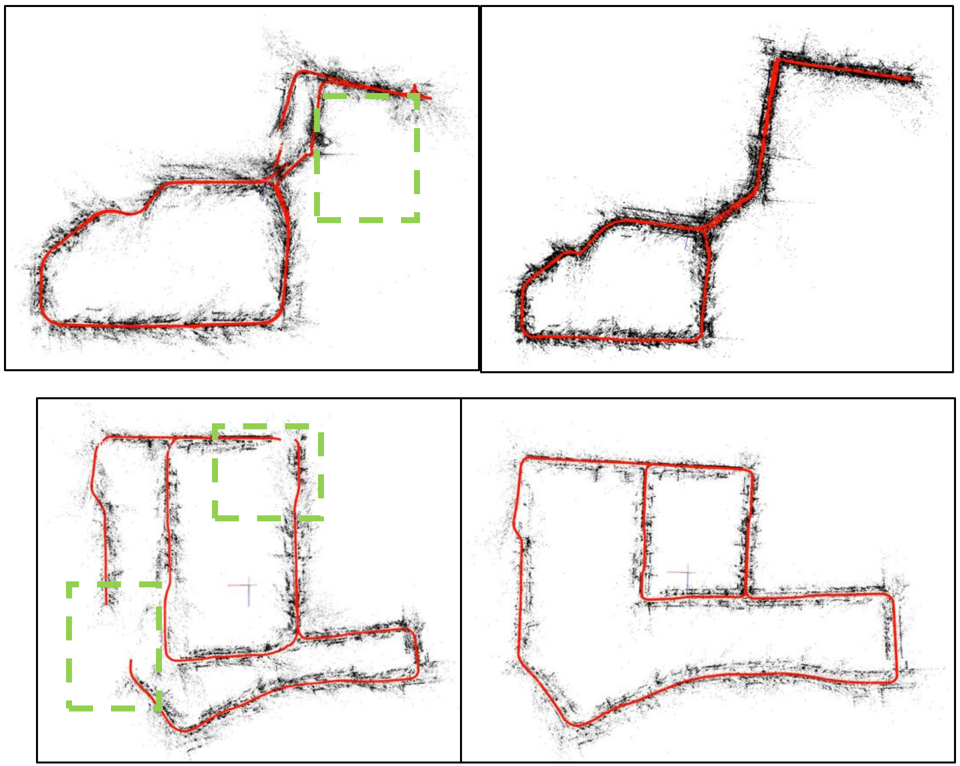
Main Idea

Robustness in Structure from Motion

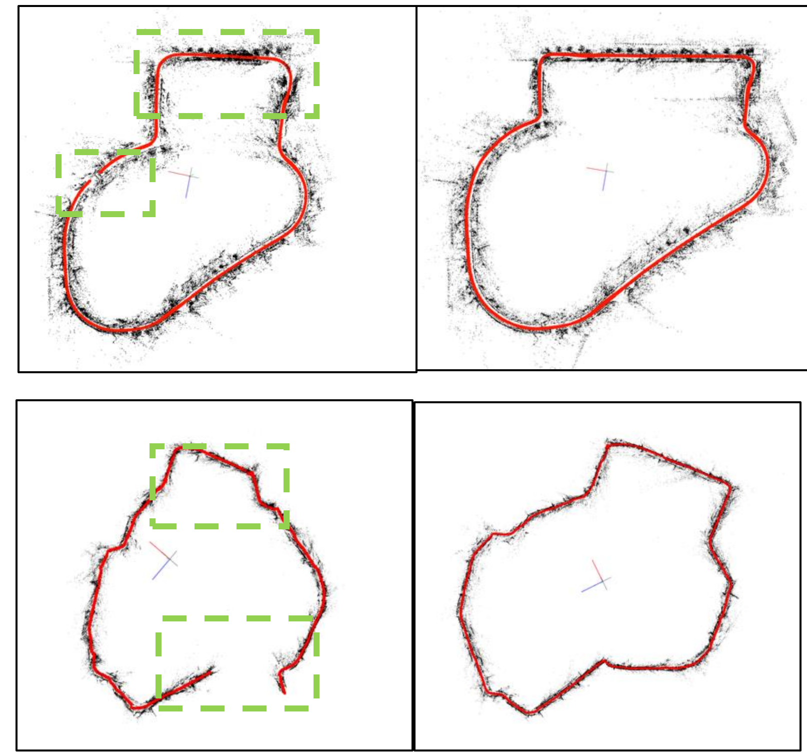
- **Fuse low-cost sensor data** into vision-based view graph
- Leveraging **global SfM guidance**

Main Idea

Effectiveness of Augmented View Graph



Global SfM from *raw view graph*　　Global SfM from *augmented view graph*

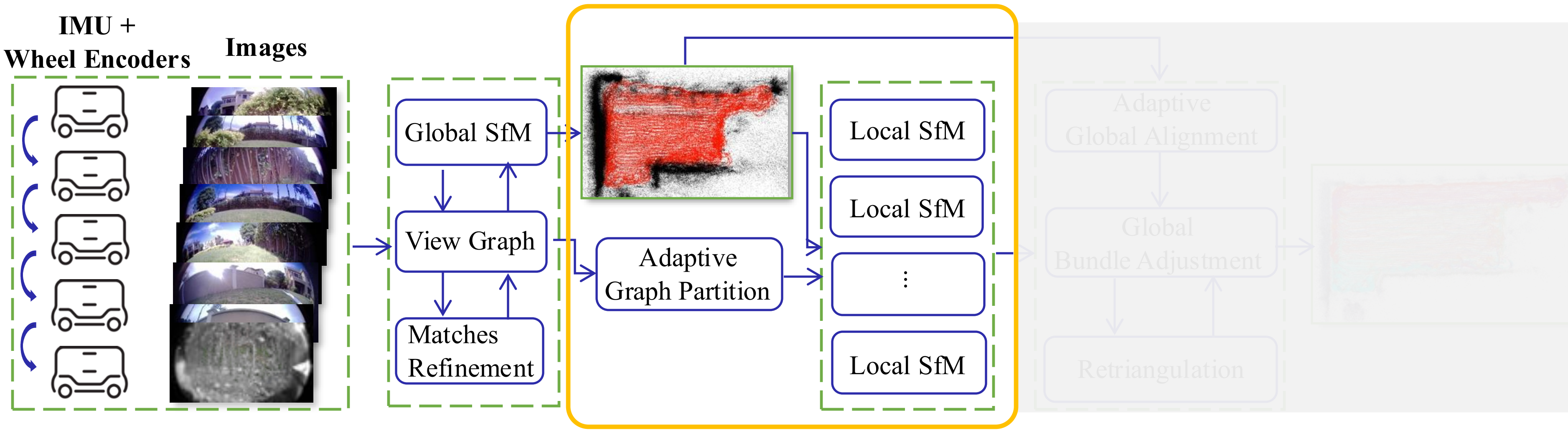


Global SfM from *raw view graph*　　Global SfM from *augmented view graph*

Data source: https://www.4seasons-dataset.com

Main Idea

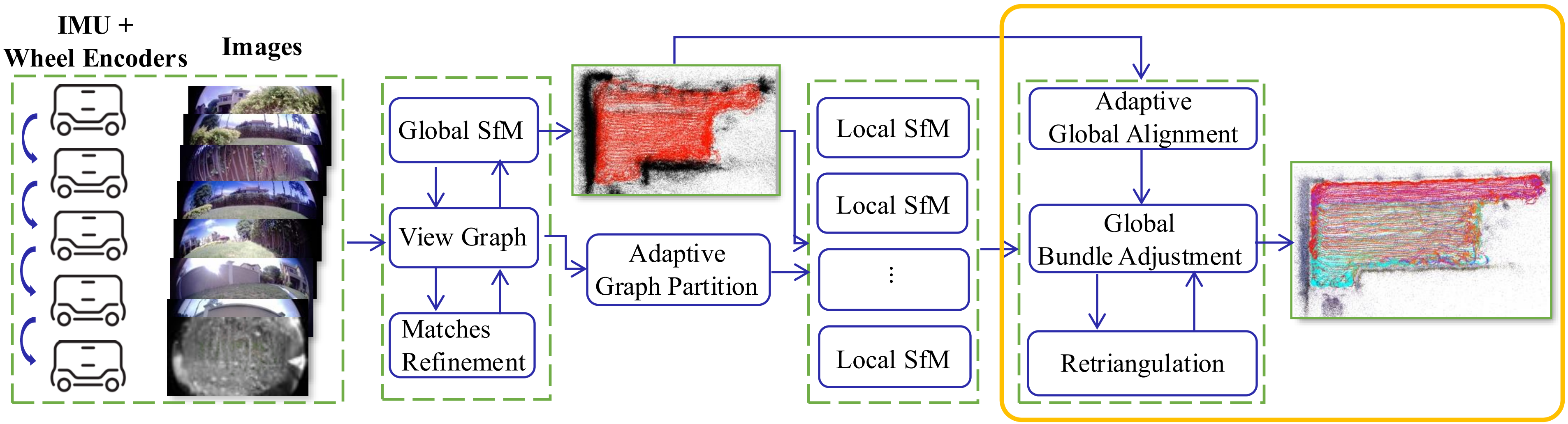
Efficiency in Structure from Motion



- **Divide-and-Conquer:** Split large scene into smaller blocks
- Leveraging **distribute computing resources**

Main Idea

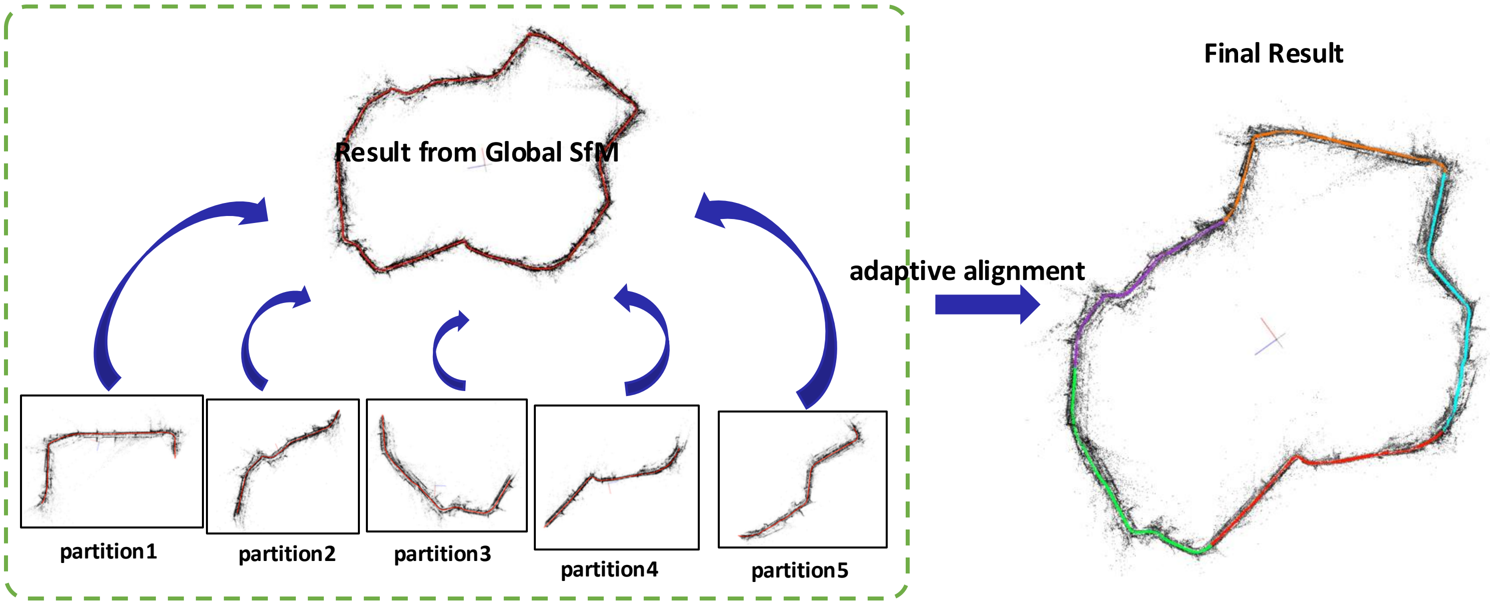
Robustness in Structure from Motion



- **Fuse low-cost sensor data** into vision-based view graph
- Leveraging **global SfM guidance**
- **Adaptive alignment** to handle scale ambiguity

Main Idea

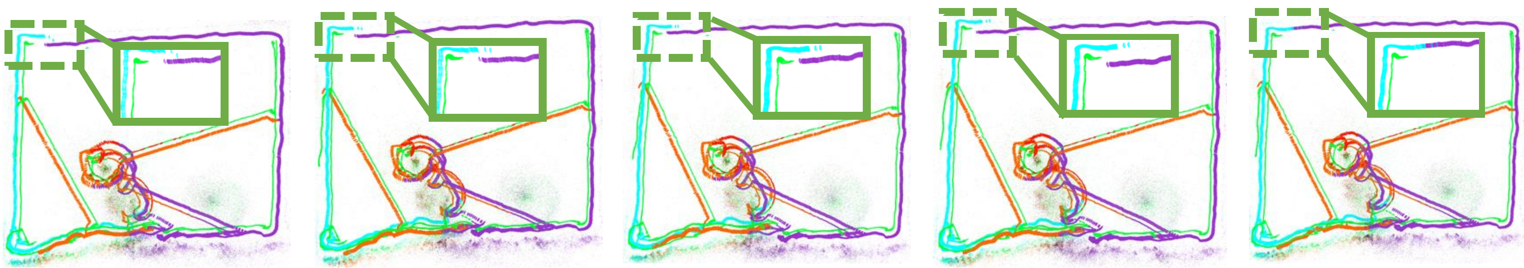
Robustness in Structure from Motion – Adaptive Alignment



Main Idea

Robustness in Structure from Motion – Adaptive Alignment

- Similarity transformation estimation is crucial!
  - Outliers in registered camera poses
  - Unknown absolute scale of inlier threshold

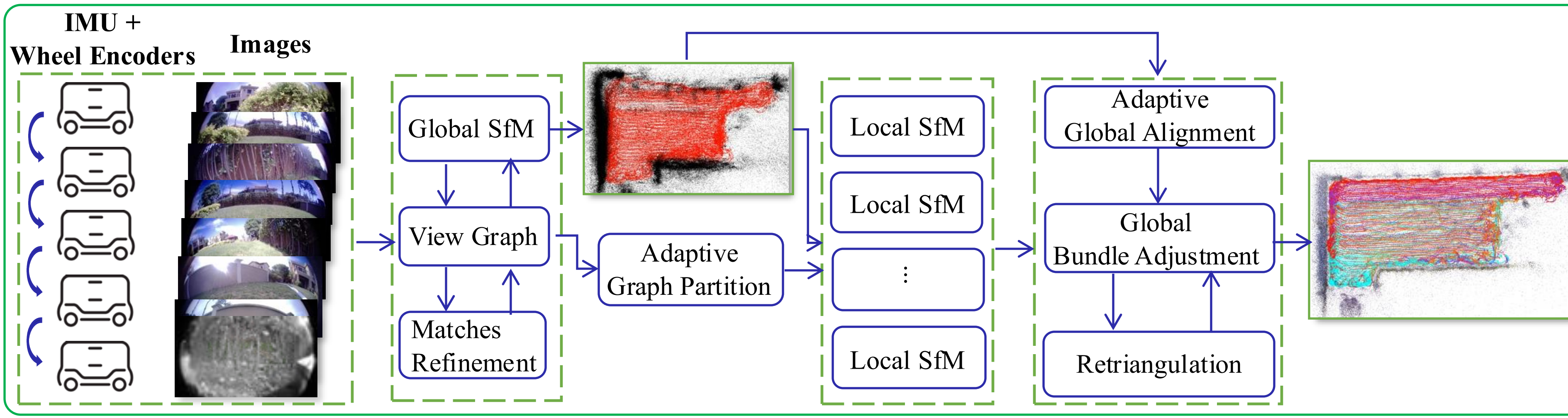


(a) $\tau_{\text{inlier}} = 0.5$ (b) $\tau_{\text{inlier}} = 1.0$ (c) $\tau_{\text{inlier}} = 1.5$ (d) $\tau_{\text{inlier}} = 2.0$ (e) $\tau_{\text{init}} = 1.0$

(a)-(d) are alignment results **by using different fixed inlier threshold within RANSAC**; **(e)** is the result **with our adaptive global alignment algorithm** with an initial inlier threshold 1.0.

Highlight

AdaSfM accelerates SfM by 3-5 times on single machine (~12 times on 3 servers) with higher camera pose accuracy



Results

Reconstruction on Autonomous Driving Datasets

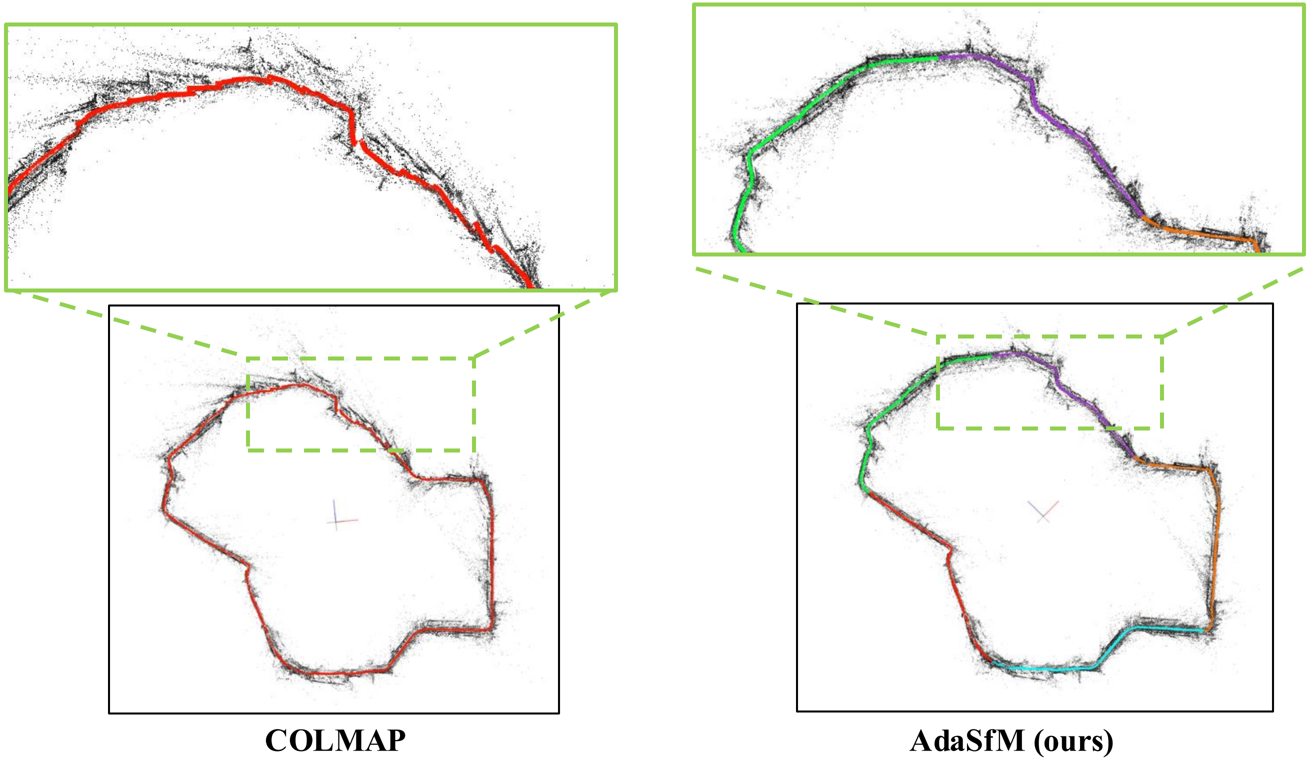
Scene	Sequence	COLMAP [8]					Ours (Global SfM)					Ours (final)				
		N_c	N_p	$\Delta\mathbf{R}$	$\Delta\mathbf{t}$	T	N_c	N_p	$\Delta\mathbf{R}$	$\Delta\mathbf{t}$	T	N_c	N_p	$\Delta\mathbf{R}$	$\Delta\mathbf{t}$	T
Neighborhood	recording_2020-10-07_14-53-52	6,326	137,135	**0.65**	1.78	334.90	6,036	66,777	2.52	1.17	14.68	6,033	109,483	0.74	**0.52**	123.96
	recording_2020-12-22_11-54-24	6,518	127,892	0.55	3.68	354.35	6,144	64,405	1.10	0.86	15.83	6,144	102,857	**0.51**	**0.62**	151.88
	recording_2020-03-26_13-32-55	7,414	148,848	**0.61**	1.24	603.13	5,982	70,066	0.92	**0.79**	17.10	5,982	111,807	1.11	0.98	157.76
	recording_2020-10-07_14-47-51	6,688	152,307	**0.56**	1.67	359.03	6,248	76,305	2.20	1.17	15.70	6,248	121,657	0.75	**0.74**	152.85
	recording_2021-02-25_13-25-15	6,174	138,807	0.75	1.05	325.65	5,238	62,879	1.00	1.14	15.12	5,238	106,609	**0.46**	**0.81**	202.85
	recording_2021-05-10_18-02-12	7,784	149,528	3.04	9.57	444.85	5,834	61,889	1.49	1.38	12.76	5,834	101,102	**0.47**	**0.59**	153.36
	recording_2021-05-10_18-32-32	7,174	141,864	2.77	19.15	416.34	6,046	89,010	**1.14**	**1.03**	23.81	6,046	142,430	1.49	1.34	264.75
Business Park	recording_2021-01-07_13-12-23	8,016	109,399	0.72	0.75	643.22	9,010	72,096	1.76	1.60	56.16	9,010	100,057	**0.66**	**0.51**	465.34
	recording_2020-10-08_09-30-57	11,520	127,013	**0.37**	1.57	1284.44	8,278	66,087	1.59	1.51	48.72	8,278	108,000	0.63	**0.45**	366.81
	recording_2021-02-25_14-16-43	7,414	148,848	**0.61**	1.24	603.13	5,982	70,066	0.92	**0.79**	17.10	5,982	111,807	1.11	0.98	157.76
Old Town	recording_2020-10-08_11-53-41	19,332	279,989	-	-	2454	12,910	181,569	2.23	2.81	45.72	12,048	279,127	**0.55**	**0.56**	254.71
	recording_2021-01-07_10-49-45	16.420	307,383	8.63	360.51	1496.6	12,728	194,340	2.56	3.14	53.18	12,728	327,348	**1.55**	**1.03**	238.82
	recording_2021-02-25_12-34-08	18,950	305,461	-	-	2392.98	12,387	182,940	2.02	3.14	40.97	12,387	302,833	**0.63**	**0.74**	683.97
Office Loop	recording_2020-03-24_17-36-22	10,188	209,942	1.17	3.40	822.38	9,522	126,680	2.28	2.38	31.87	9,377	214,285	**0.97**	**0.98**	166.54
	recording_2020-03-24_17-45-31	8,582	195,738	0.92	3.04	865.48	9,186	122,713	2.79	2.20	33.91	8,940	205,790	**0.84**	**0.85**	209.06
	recording_2020-04-07_10-20-31	10,350	223.649	4.22	42.44	795.68	10,184	138,446	2.53	1.78	39.83	10,184	224,499	**1.47**	**1.14**	253.24
	recording_2020-06-12_10-10-57	9,990	236,593	18.97	83.94	705.93	10,150	164,062	1.92	1.61	37.32	10,150	246,516	**0.76**	**0.87**	206.48
	recording_2021-01-07_12-04-03	9,164	475,950	**0.71**	2.58	1000.75	10,300	143,715	3.32	2.39	48.68	10,300	223,676	1.08	**0.67**	249.42
	recording_2021-02-25_13-51-57	9,574	214,695	**0.84**	2.84	773.32	9,426	122,746	3.80	2.68	28.96	9,426	204,289	1.01	**0.91**	173.29

Comparison of runtime and accuracy on the 4Seasons datasets. T denotes the runtime (in minutes). N_c, N_p denote the number of registered images and 3D points, respectively. **R, t** denotes the mean rotation error (in degrees) and translation error (in meters), respectively, and we highlight the best results in bold.

P. Wenzel, *et, al.* 4seasons: A cross-season dataset for multiweather SLAM in autonomous driving. GCPR 2020

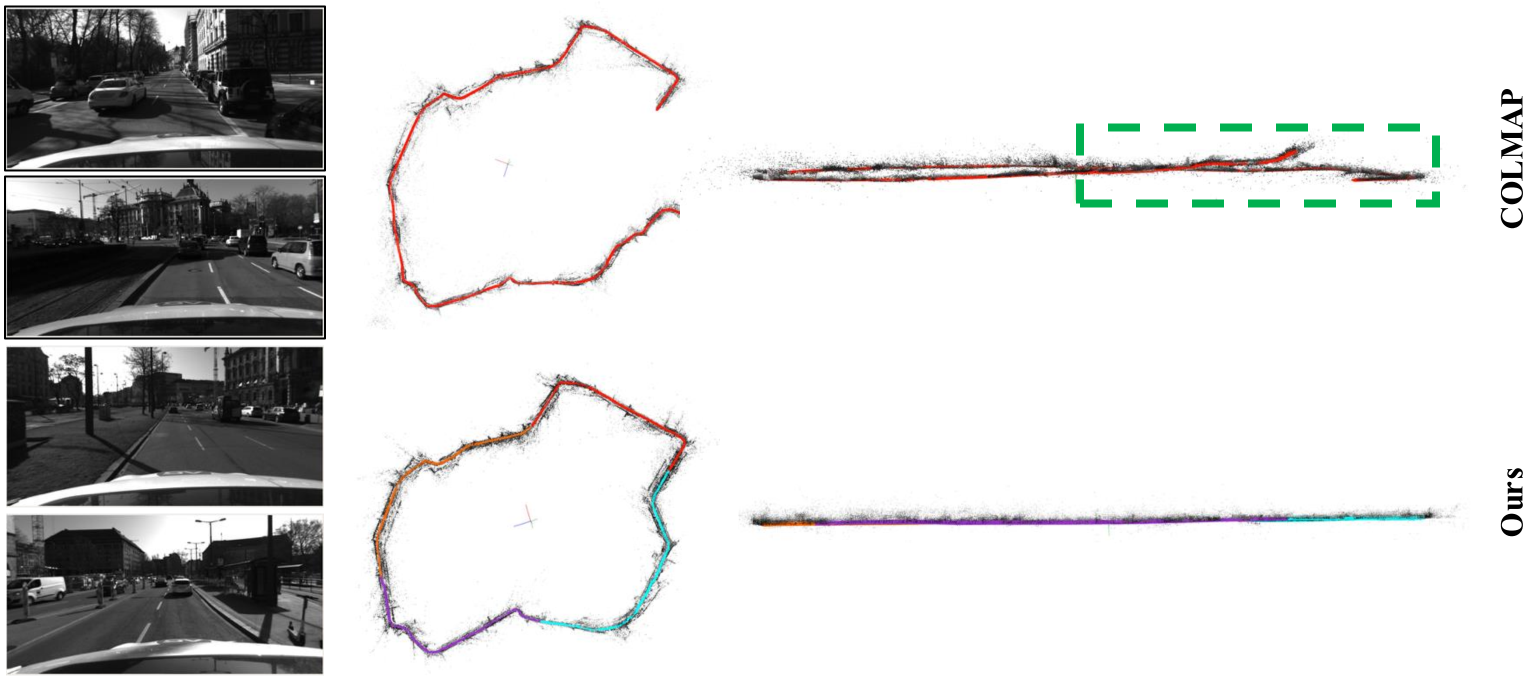
Reconstruction on Autonomous Driving Datasets

Comparison to COLMAP



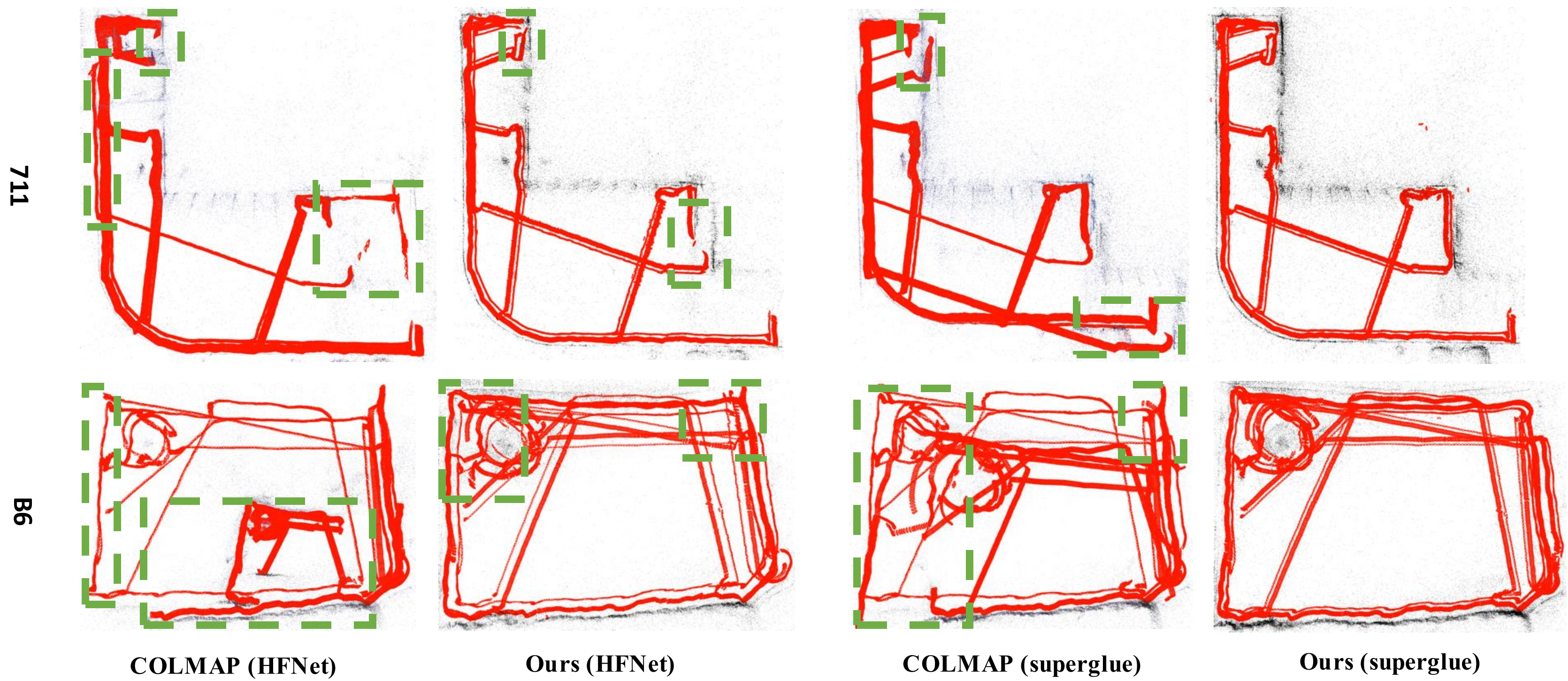
Reconstruction on Autonomous Driving Datasets

Comparison to COLMAP



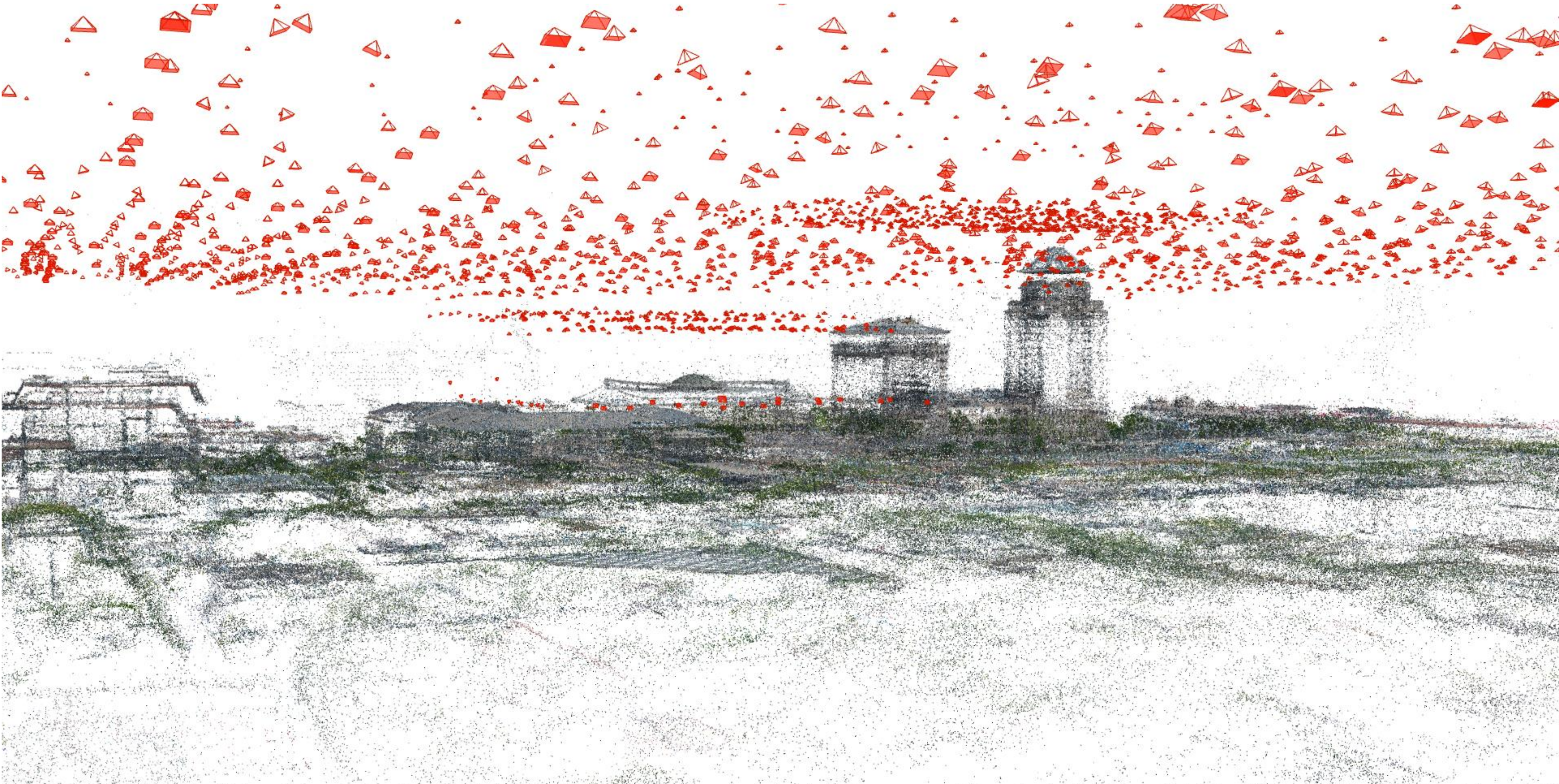
Reconstruction on Self-Collected Datasets

Comparison to COLMAP



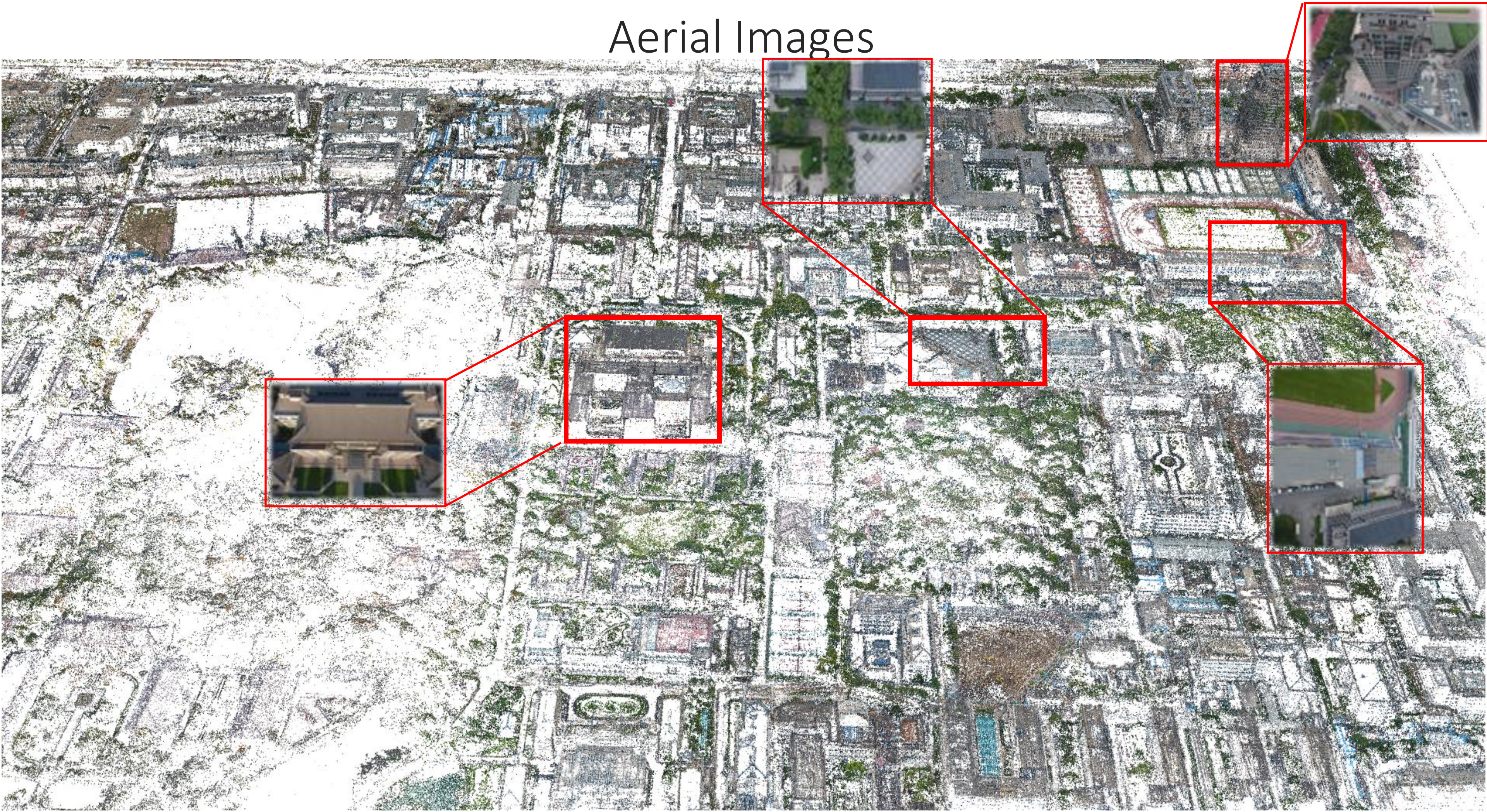
Reconstruction on Self-Collected Datasets

Aerial Images



Reconstruction on Self-Collected Datasets

Aerial Images



Part #1

Develop Distributed System and Neural Scene Representations for 3D Reconstruction, Rendering, and Geometry Understanding

AdaSfM
ICRA 2023

DOGS
NeurIPS 2024

DBARF
CVPR 2023

DReg-NeRF
ICCV 2023

NEURAL INFORMATION PROCESSING SYSTEMS

Yu Chen, Gim Hee Lee, **DOGS: Distributed-Oriented Gaussian Splatting for Large-Scale 3D Reconstruction Via Gaussian Consensus**. NeurIPS 2024

Key Challenges

Neural 3D Reconstruction using 3DGS

- Reconstruct majorly from **2D images**
- Reconstruct 3D scenes **beyond single server**
- Reconstruct 3D scenes **at arbitrary scale**

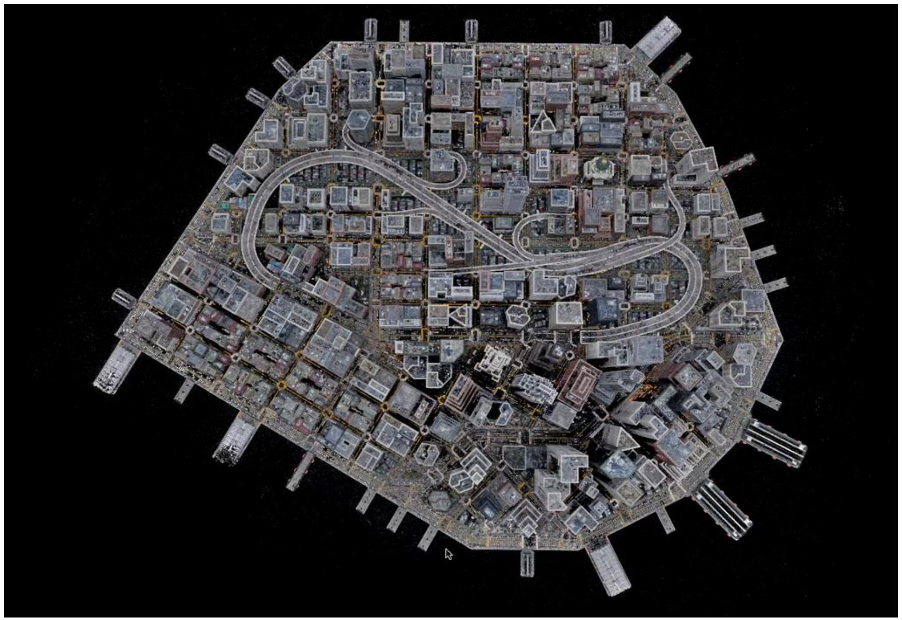
Works well on small scale scenes, but too slow when trained on **large-scale** areas

Key Challenges

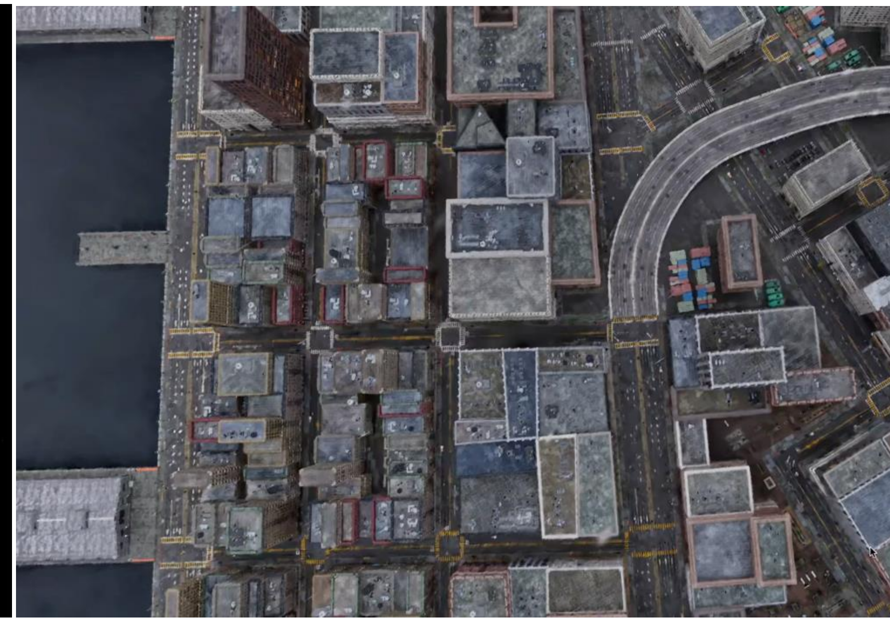
Room Scale / Object-centric Scenes with 3DGS

Highlight ☆☆☆

DOGS accelerates 3DGS training by 6+ times with better rendering quality on five compute nodes



Point Clouds After Training (12.5 M 3D Gaussians)

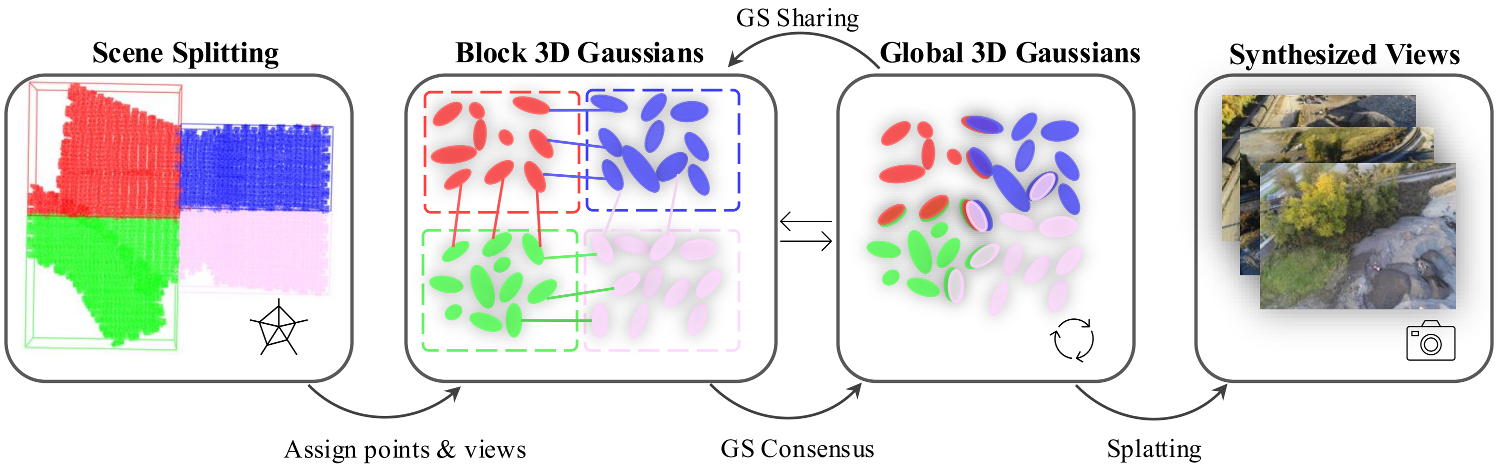


Novel View Rendering

Visualization Results Recorded on Web Viewer (MacBook m1 chip, 8GB memory)

Main Idea

Divide-and-Conquer



Algorithm Pipeline

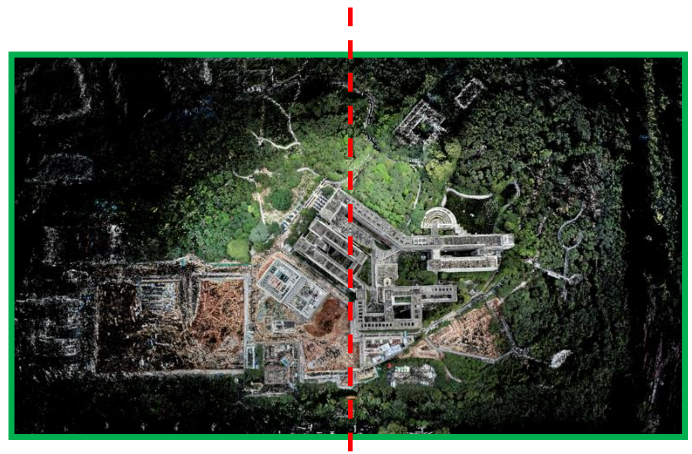
Main Idea

Divide-and-Conquer: Scene Splitting

Point Clouds in 3D Space

Main Idea

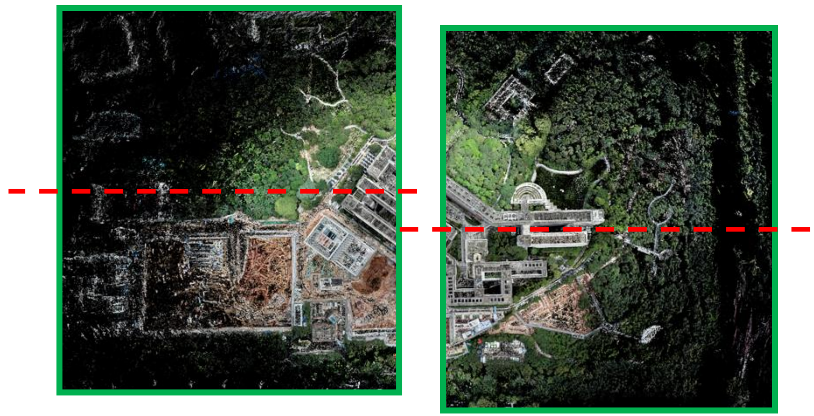
Divide-and-Conquer: Scene Splitting



Project 3D point clouds onto ground plane

Main Idea

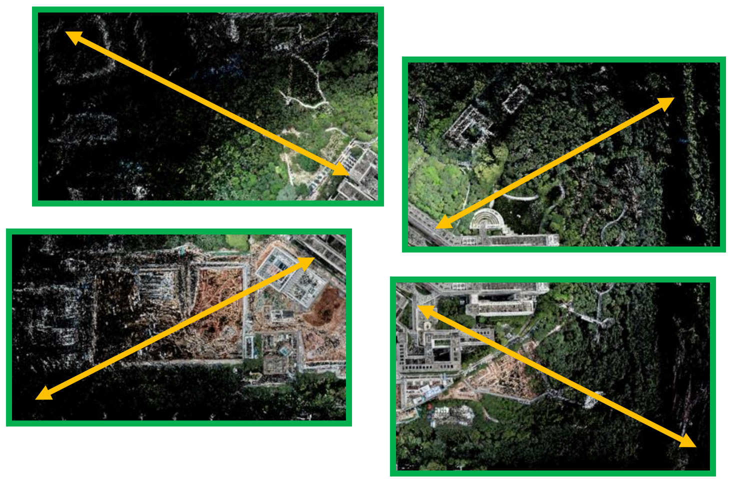
Divide-and-Conquer: Scene Splitting



Splitting along the longer axis

Main Idea

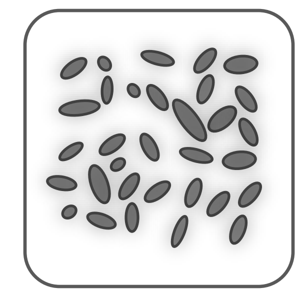
Divide-and-Conquer: Scene Splitting



Intersected Blocks

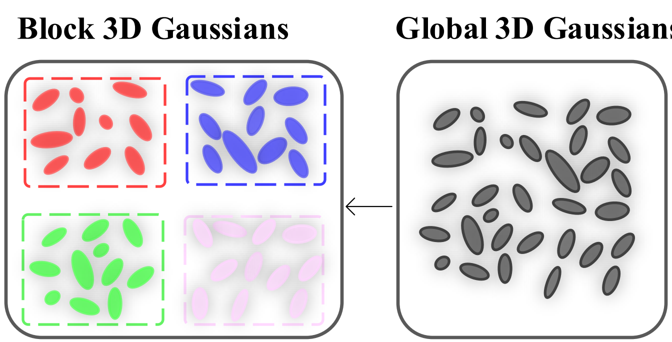
Main Idea

Divide-and-Conquer: Consensus and Sharing



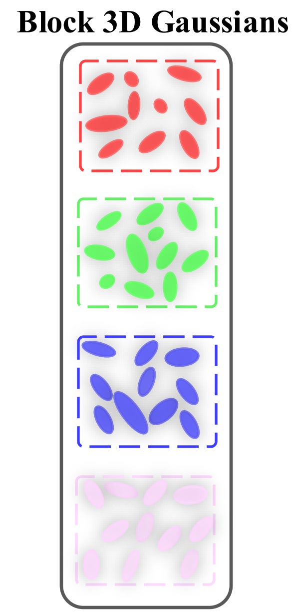
Main Idea

Divide-and-Conquer: Consensus and Sharing



Main Idea

Divide-and-Conquer: Consensus and Sharing

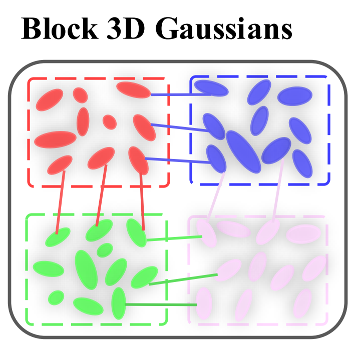


Distributedly trained
on K slave nodes

How to ensure the consistency of the **shared 3D Gaussians** in different blocks?

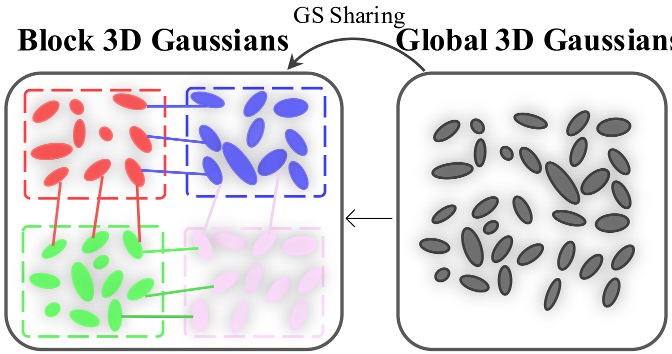
Main Idea

Divide-and-Conquer: Consensus and Sharing



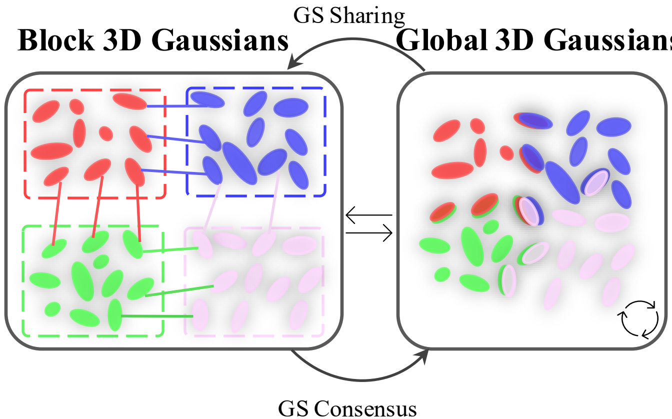
Main Idea

Divide-and-Conquer: Consensus and Sharing



Main Idea

Divide-and-Conquer: Consensus and Sharing



Results

Reconstruction on Large Scale Scenes

Higher Fidelity Novel View Synthesis

Scenes	Building			Rubble			Campus			Residence			Sci-Art		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Mega-NeRF [46]	20.92	0.547	0.454	24.06	0.553	0.508	23.42	0.537	0.636	22.08	0.628	0.401	25.60	0.770	0.312
Switch-NeRF [31]	21.54	0.579	0.397	24.31	0.562	0.478	23.62	0.541	0.616	22.57	0.654	0.352	26.51	0.795	0.271
3D-GS [18]	22.53	0.738	0.214	25.51	0.725	0.316	23.67	0.688	0.347	22.36	0.745	0.247	24.13	0.791	0.262
VastGaussian† [22]	21.80	0.728	0.225	25.20	0.742	0.264	23.82	0.695	0.329	21.01	0.699	0.261	22.64	0.761	0.261
Hierarchy-GS [19]	21.52	0.723	0.297	24.64	0.755	0.284	–	–	–	–	–	–	–	–	–
DoGaussian	22.73	0.759	0.204	25.78	0.765	0.257	24.01	0.681	0.377	21.94	0.740	0.244	24.42	0.804	0.219

Table 1: **Quantitative results of novel view synthesis on Mill19 [46] dataset and Urban-Scene3D [25] dataset**. ↑: higher is better, ↓: lower is better. The red, orange and yellow colors respectively denote the best, the second best, and the third best results. † denotes without applying the decoupled appearance encoding.

- Haithem Turkish, *et, al.* Mega-NeRF: Scalable Construction of Large-Scale NeRFs for Virtual Fly Throughs. CVPR 2022
- Zhenxing Mi, Dan Xu. Switch-NeRF: Learning Scene Decomposition with Mixture of Experts for Large-scale Neural Radiance Fields. ICLR 2023
- Lin Jiaqi, *et, al.* VastGaussian: Vast 3D Gaussians for Large Scene Reconstruction, CVPR 2024
- Kerbl, *et, al.* A Hierarchical 3D Gaussian Representation for Real-Time Rendering of Very Large Datasets, SIGGRAPH 2024

Reconstruction on Large Scale Scenes

Faster Training Speed than 3DGS

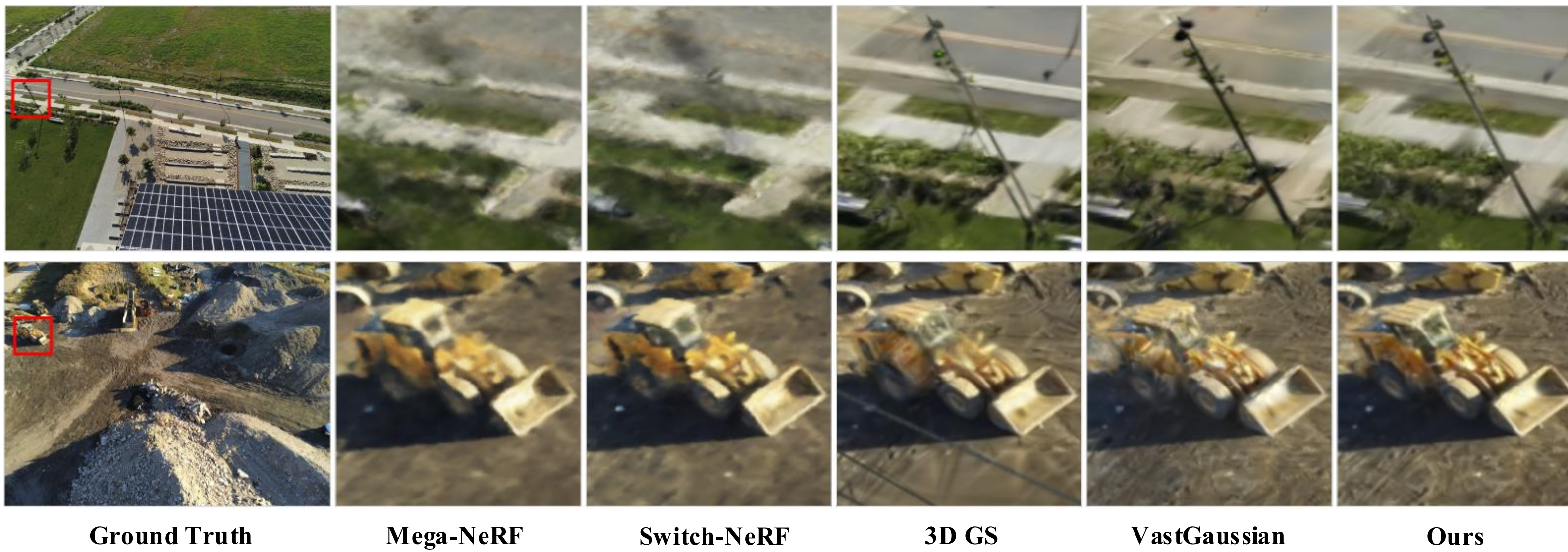
Scenes	Building				Rubble				Campus				Residence				Sci-Art			
	Train ↓	Points	Mem	FPS ↑	Train ↓	Points	Mem	FPS ↑	Train ↓	Points	Mem	FPS ↑	Train ↓	Points	Mem	FPS ↑	Train ↓	Points	Mem	FPS ↑
Mega-NeRF [46]	19:49	–	5.84	0.009	30:48	–	5.88	0.009	29:03	–	5.86	0.008	27:20	–	5.99	0.006	27:39	–	5.97	0.006
Switch-NeRF [31]	24:46	–	5.84	0.009	38:30	–	5.87	0.009	36:19	–	5.85	0.007	35:11	–	5.94	0.007	34:34	–	5.92	0.008
3D-GS [18]	21:37	7.99	4.62	90.09	18:40	3.85	2.18	166.67	23:03	13.6	7.69	59.52	23:13	5.35	3.23	142.86	21:33	2.31	1.61	240.96
VastGaussian† [22]	03:26	5.60	3.07	121.35	02:30	4.71	2.74	163.93	03:33	17.6	9.61	47.84	03:12	6.26	3.67	118.48	02:33	4.21	3.54	120.33
DoGaussian	03:51	6.89	3.39	122.33	02:25	4.74	2.54	147.06	04:15	8.27	4.29	99.85	04:33	7.64	6.11	82.34	04:23	5.67	3.53	107.87

Table 2: **Quantitative results of novel view synthesis on Mill19 dataset and UrbanScene3D dataset**. We present the training time (hh:mm), the number of final points (10^6), the allocated memory (GB), and the framerate (FPS) during evaluation. † denotes without applying the decoupled appearance encoding.

- Haithem Turkish, *et, al.* **Mega-NeRF**: Scalable Construction of Large-Scale NeRFs for Virtual Fly Throughs. CVPR 2022
- Zhenxing Mi, Dan Xu. **Switch-NeRF**: Learning Scene Decomposition with Mixture of Experts for Large-scale Neural Radiance Fields. ICLR 2023
- Lin Jiaqi, *et, al.* **VastGaussian**: Vast 3D Gaussians for Large Scene Reconstruction, CVPR 2024
- Kerbl, *et, al.* A Hierarchical 3D Gaussian Representation for Real-Time Rendering of Very Large Datasets, SIGGRAPH 2024

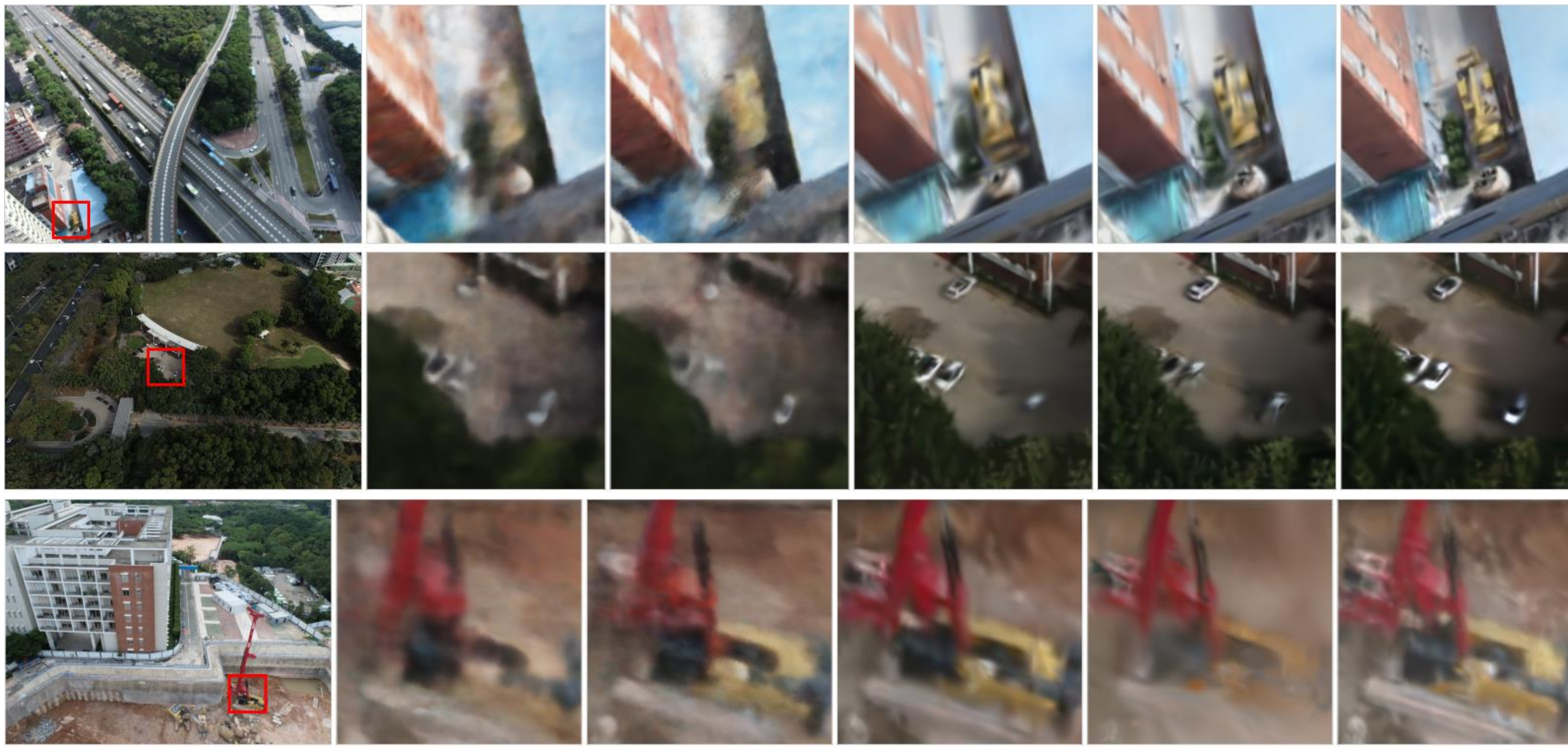
Reconstruction on Large Scale Scenes

Higher Fidelity Novel View Synthesis – Mill19



Reconstruction on Large Scale Scenes

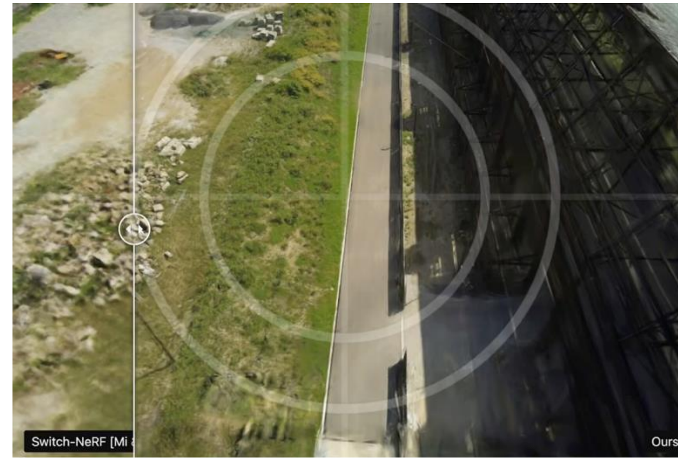
Higher Fidelity Novel View Synthesis – UrbanScene3D

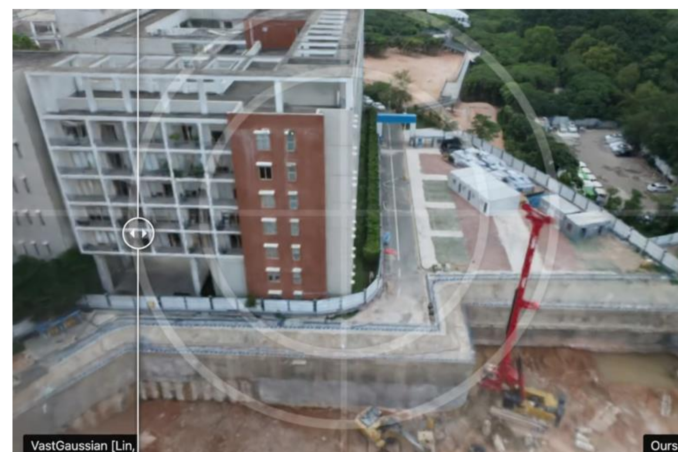


Ground Truth | **Mega-NeRF** | **Switch-NeRF** | **3D GS** | **VastGaussian** | **Ours**

Reconstruction on Large Scale Scenes

Higher Fidelity Novel View Synthesis





Reconstruction on Large Scale Scenes

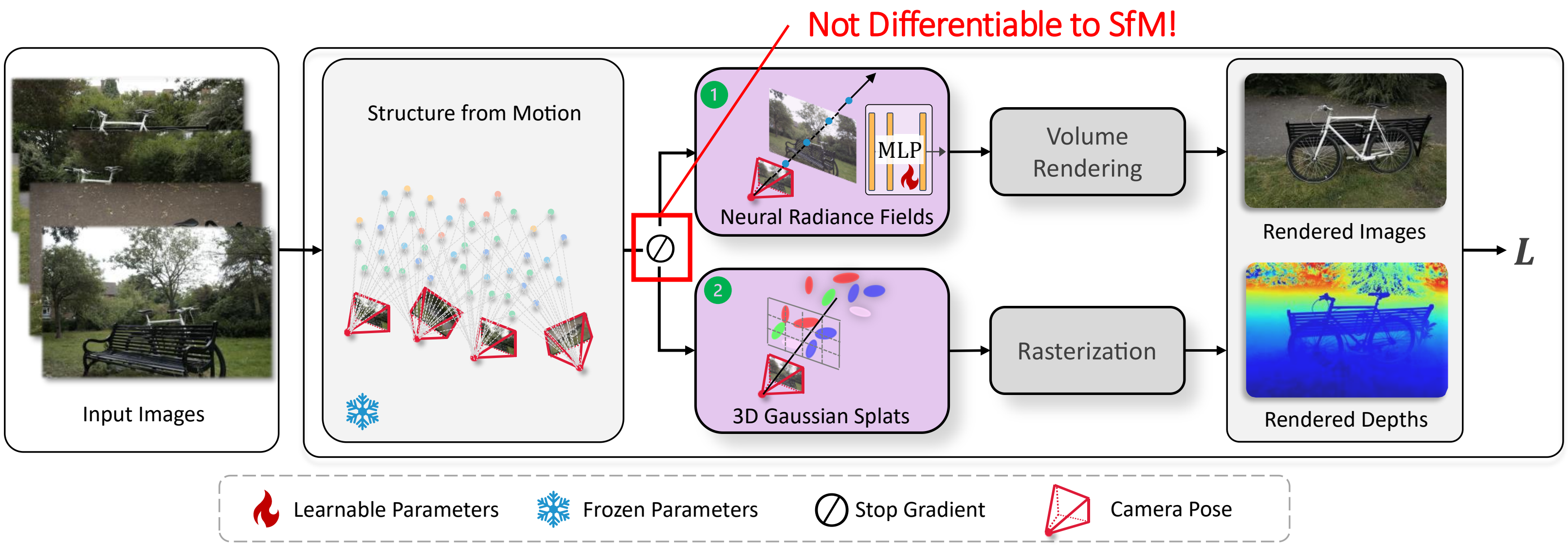
Importance of Consensus and Sharing

w.o. CS **full model** **w.o. CS** **full model**

- w.o. CS: our method without the 3D Gaussian consensus

Limitations

Decoupled Camera Pose Estimation and Scene Reconstruction

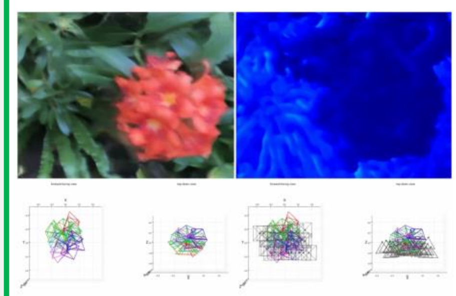


Part #2

Develop Distributed System and Neural Scene Representations for 3D Reconstruction, Rendering, and Geometry Understanding

AdaSfM
ICRA 2023

DOGS
NeurIPS 2024



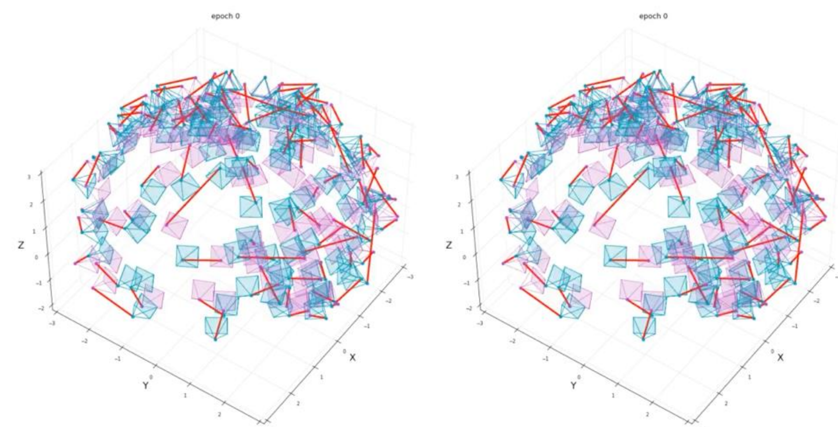
DBARF
CVPR 2023

DReg-NeRF
ICCV 2023

Yu Chen, Gim Hee Lee, **DBARF: Deep Bundle-Adjusting Generaliable Neural Radiance Fields**. CVPR 2023

Bundle-Adjusting Neural Radiance Fields

- **Pros**
  - Refine inaccurate camera poses
- **Cons**
  - Trained on per-scene NeRF
  - Camera poses initialization

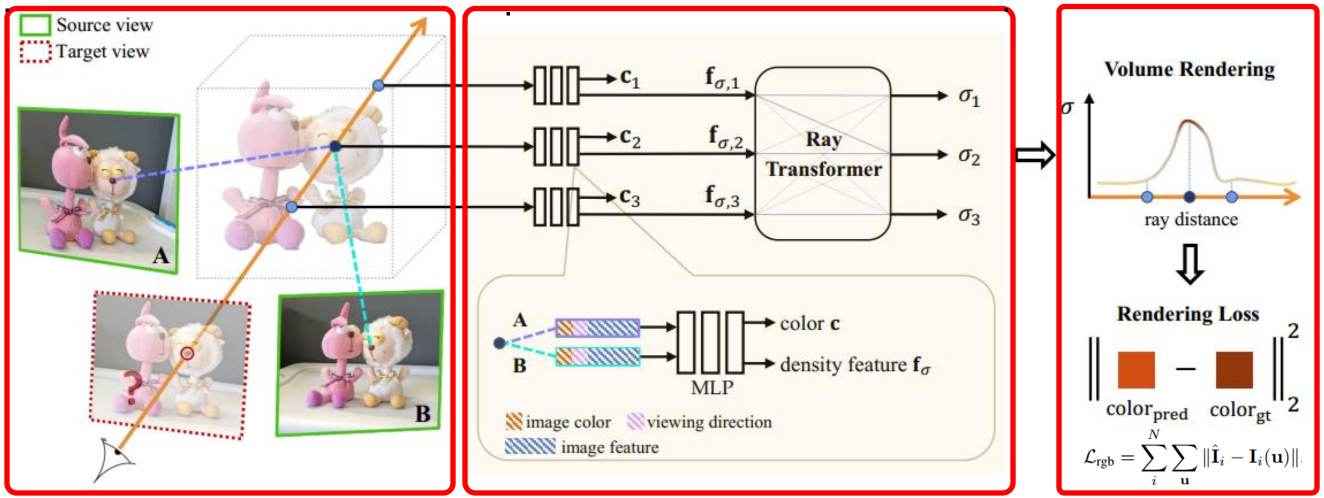


NeRF w/ full positional encoding **BARF**

Chen-Hsuan Lin, *et, al.* BARF: Bundle-Adjusting Neural Radiance Fields. ICCV 2021

Generalizable Neural Radiance Fields

1 $\mathbf{f} = \chi\big(\Pi(\mathbf{P}_j, \omega(\mathbf{X}_i^k, \mathbf{P}_i)), \mathbf{F}_j\big)$ 2 $g_k = f_a(\mathbf{f}_1^k, \mathbf{f}_2^k, \cdots, \mathbf{f}_M^k)$ 3 $\hat{\mathbf{I}}_{\text{target}} := \hat{\mathbf{I}}_i = h(g_1, \cdots, g_K; \mathbf{\Phi})$



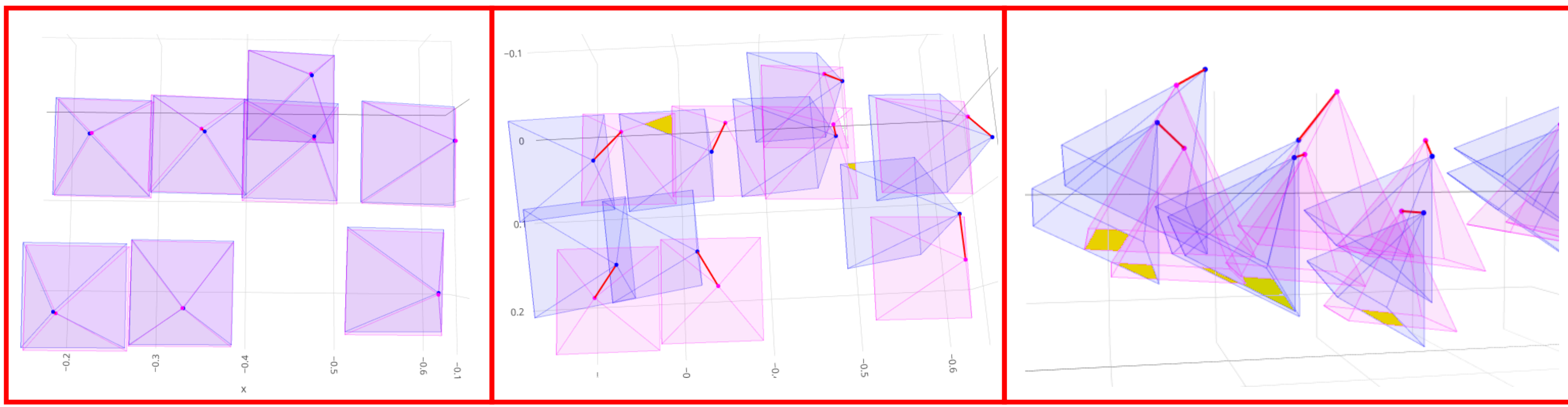
Can we bundle-adjust GeNeRF like BARF?

Qianqian Wang, *et, al.* IBRNet: Learning Multi-View Image-Based Rendering. CVPR 2021

Bundle Adjusting GeNeRF is Difficult

Initial Poses (BEV, iter=0) | **Optimized Poses (BEV, iter=10000)** | **Optimized Poses (SV, iter=20000)**

BARF→GeNeRF



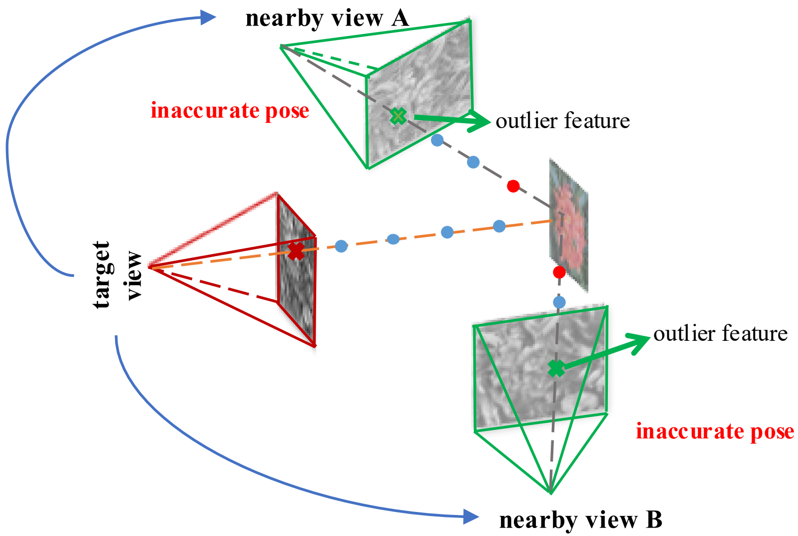
We adopted a pretrained GeNeRF model and constructed a $N \times 6$ learnable pose embedding like BARF. The pose embedding is jointly trained with the GeNeRF model and optimized by Adam with a learning rate $1e-5$.

Chen-Hsuan Lin, *et, al.* BARF: Bundle-Adjusting Neural Radiance Fields. ICCV 2021

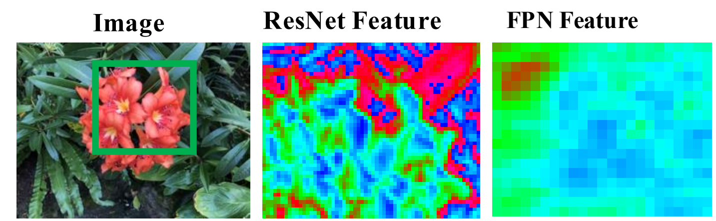
Key Idea

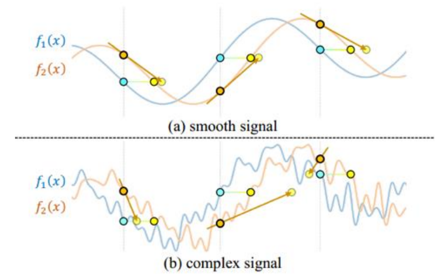
Smooth Cost Feature Map for Joint Relative Poses and GeNeRF Optimization

🤮 Occlusion contributes feature outliers



🤮 Image feature space is highly non-smooth





The overall update $\Delta\mathbf{p}$ can be more effectively estimated from pixel value differences with a smooth signal

1. Relative poses are all we need

$$\Pi(\mathbf{P}_j,\omega(\mathbf{X}_i^k,\mathbf{P}_i)) = \mathbf{K}_j\boxed{\mathbf{P}_j\mathbf{P}_i^{-1}}\mathbf{X}_i = \mathbf{K}_j\boxed{\mathbf{P}_{ij}}\mathbf{X}_i^k$$

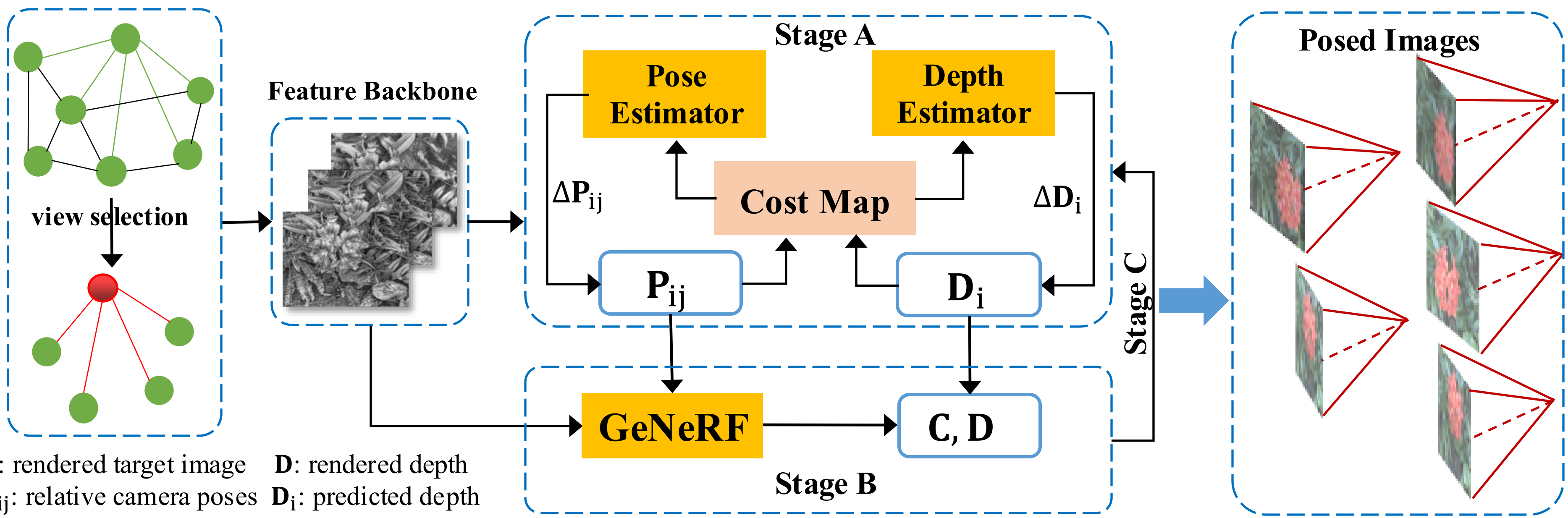
2. Smooth Cost Map as an Implicit Loss

$$\mathcal{C}=\sum_{\mathbf{u}_i}\sum_{j\in\mathcal{N}(i)}\rho\Big(\Big\|\underbrace{\boxed{\chi(\mathbf{K}_j\mathbf{P}_{ij}\mathbf{X}_i^k,\mathbf{F}_j)-\chi(\mathbf{u}_i,\mathbf{F}_i)}}_{\text{feature-metric error}}\Big\|\Big)$$

Chen-Hsuan Lin, *et, al.* BARF: Bundle-Adjusting Neural Radiance Fields. ICCV 2021

Key Idea

Smooth Cost Feature Map for Joint Relative Poses and GeNeRF Optimization



C: rendered target image **D**: rendered depth
P_{ij}: relative camera poses **D_i**: predicted depth

Results

Novel View Synthesis

Comparable Results Compared to NeRF with GT Poses

Scenes	PSNR ↑						SSIM ↑						LPIPS ↓					
	BARF [19]	GARF [4]	IBRNet [46]		Ours		BARF [19]	GARF [4]	IBRNet [46]		Ours		BARF [19]	GARF [4]	IBRNet [46]		Ours	
			✗	✔	✗	✔			✗	✔	✗	✔			✗	✔	✗	✔
fern	23.79	24.51	23.61	25.56	23.12	**25.97**	0.710	0.740	0.743	0.825	0.724	**0.840**	0.311	0.290	0.240	0.139	0.277	**0.120**
flower	23.37	**26.40**	22.92	23.94	21.89	23.95	0.698	0.790	0.849	0.895	0.793	**0.895**	0.211	0.110	0.123	0.074	0.176	**0.074**
fortress	29.08	29.09	29.05	31.18	28.13	**31.43**	0.823	0.820	0.850	0.918	0.820	**0.918**	0.132	0.150	0.087	0.046	0.126	**0.046**
horns	22.78	23.03	24.96	**28.46**	24.17	27.51	0.727	0.730	0.831	**0.913**	0.799	0.903	0.298	0.290	0.144	**0.070**	0.194	0.076
leaves	18.78	19.72	19.03	**21.28**	18.85	20.32	0.537	0.610	0.737	**0.807**	0.649	0.758	0.353	0.270	0.289	**0.137**	0.313	0.156
orchids	19.45	19.37	18.52	**20.83**	17.78	20.26	0.574	0.570	0.573	**0.722**	0.506	0.693	0.291	0.260	0.259	**0.142**	0.352	0.151
room	**31.95**	31.90	28.81	31.05	27.50	31.09	0.940	0.940	0.926	**0.950**	0.901	0.947	0.099	0.130	0.099	**0.060**	0.142	0.063
trex	22.55	22.86	23.51	**26.52**	22.70	22.82	0.767	0.800	0.818	**0.905**	0.783	0.848	0.206	0.190	0.160	**0.074**	0.207	0.120

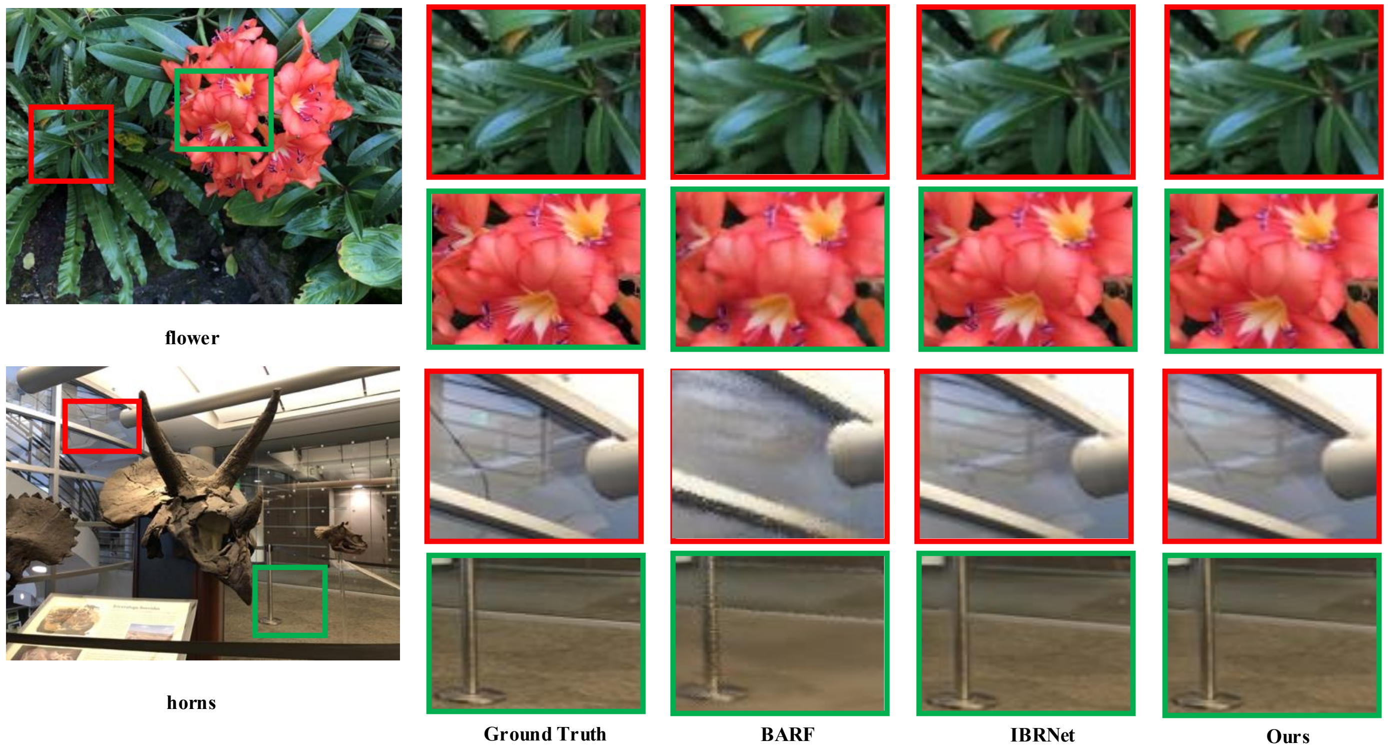
Table 1. Quantitative results of novel view synthesis on LLFF [26] forward-facing dataset. For IBRNet [46] and our method, the results with (✔) and without (✗) per-scene fine-tuning are given.

Scenes	**fern**	**flower**	**fortress**	**horns**	**leaves**	**orchids**	**room**	**trex**
Rotation (✗)	9.96	16.74	2.18	6.076	12.98	5.904	8.761	10.09
Rotation (✔)	0.89	1.39	0.586	0.819	4.63	1.164	0.530	1.057
translation (✗)	2.00	1.56	1.06	2.45	2.56	5.13	5.48	8.05
translation (✔)	0.34	0.32	0.23	0.29	0.85	0.57	0.36	0.46

Table 2. Quantitative results of camera pose accuracy on LLFF [26] forward-facing dataset. Rotation (degree) and translation (scaled by 10^2, without known absolute scale) errors with (✔) and without (✗) per-scene fine-tuning are given.

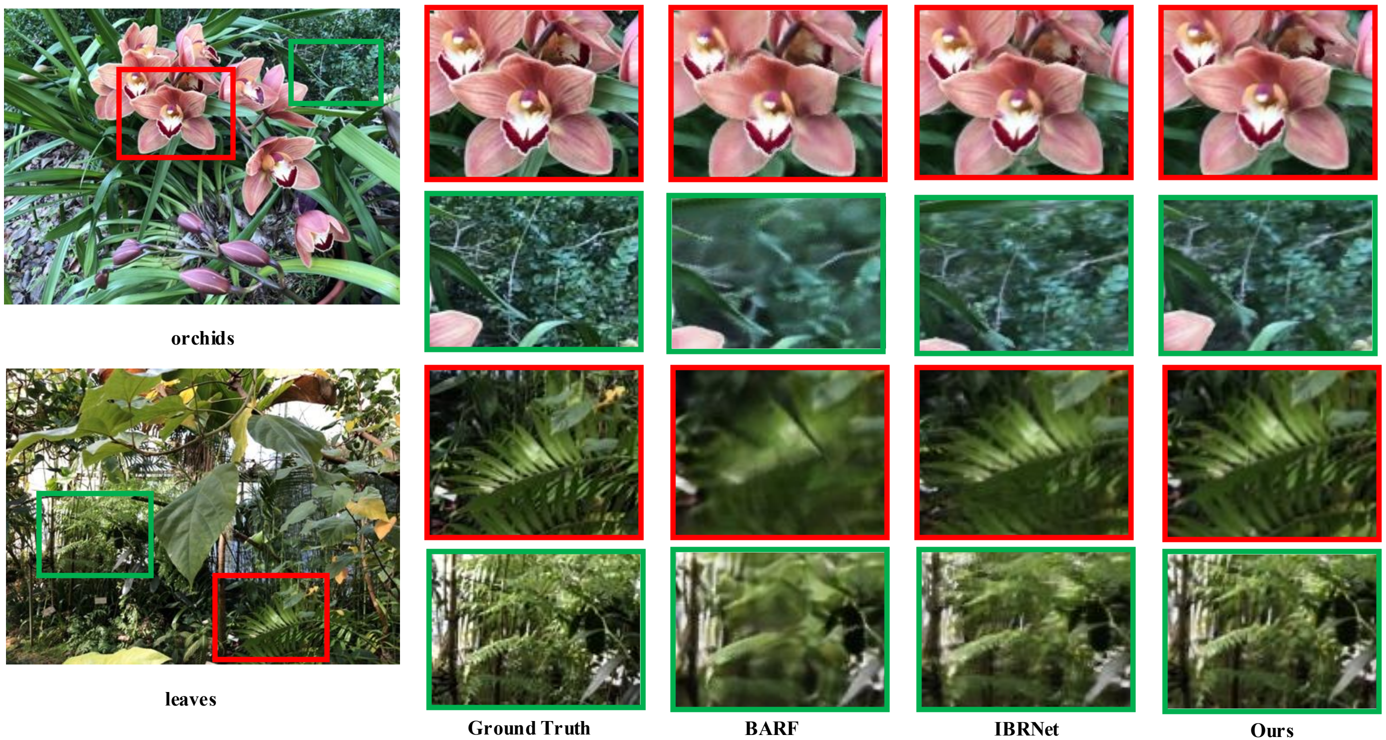
Novel View Synthesis

Comparable Results Compared to NeRF with GT Poses



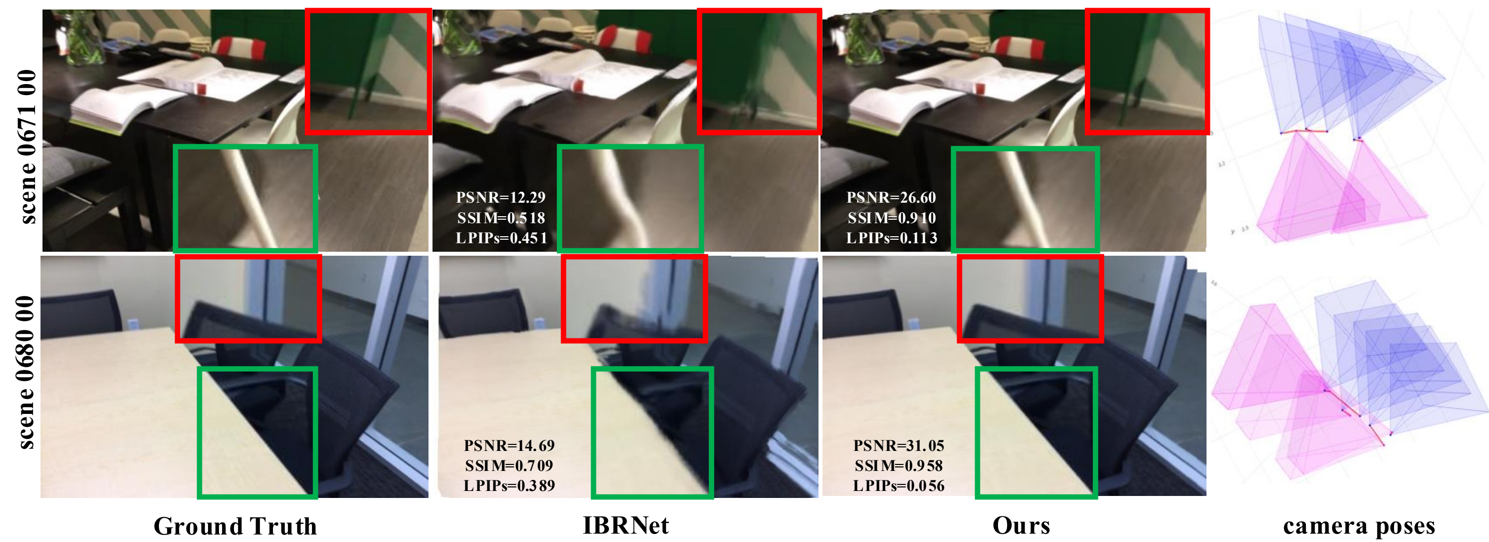
Novel View Synthesis

Comparable Results Compared to NeRF with GT Poses



Novel View Synthesis

When poor camera poses are provided to NeRF - ScanNet



Novel View Synthesis

When poor camera poses are provided to NeRF - ScanNet

Scenes	PSNR ↑		SSIM ↑		LPIPS ↓	
	IBRNet [6]	Ours	IBRNet [6]	Ours	IBRNet [6]	Ours
scene0671-00	12.29	26.60	0.518	0.910	0.451	0.113
scene0673-03	11.31	23.56	0.457	0.859	0.615	0.156
scene0675-00	10.55	19.95	0.590	0.875	0.589	0.207
scene0680-00	14.69	31.05	0.709	0.958	0.389	0.056
scene0684-00	18.46	33.61	0.737	0.975	0.296	0.052
scene0675-01	10.33	23.56	0.595	0.899	0.548	0.166
scene0684-01	14.69	33.01	0.678	0.967	0.426	0.056

Quantitative results of novel view synthesis on ScanNet dataset after finetuning.

Part #3

Develop Distributed System and Neural Scene Representations for 3D Reconstruction, Rendering, and Geometry Understanding

AdaSfM
ICRA 2023

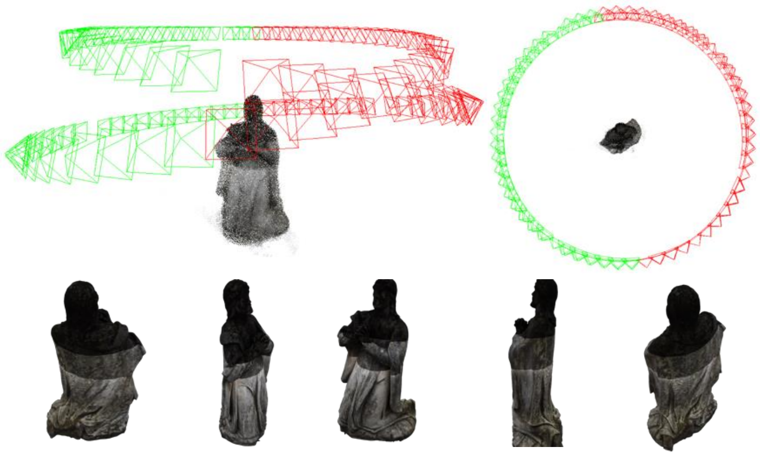
DOGS
NeurIPS 2024

DBARF
CVPR 2023

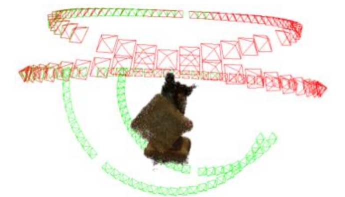
DReg-NeRF
ICCV 2023

Yu Chen, Gim Hee Lee, **DReg-NeRF: Deep Registration for Neural Radiance Fields**. ICCV 2023

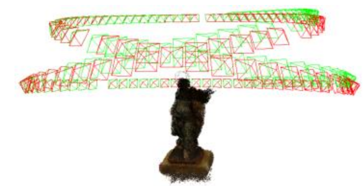
NeRF Registration



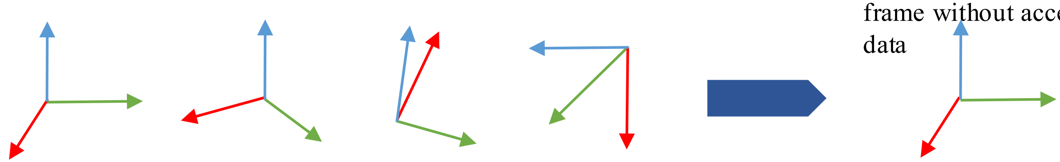
(a) Objaverse



(b) NeRF models are trained in different coordinate frames



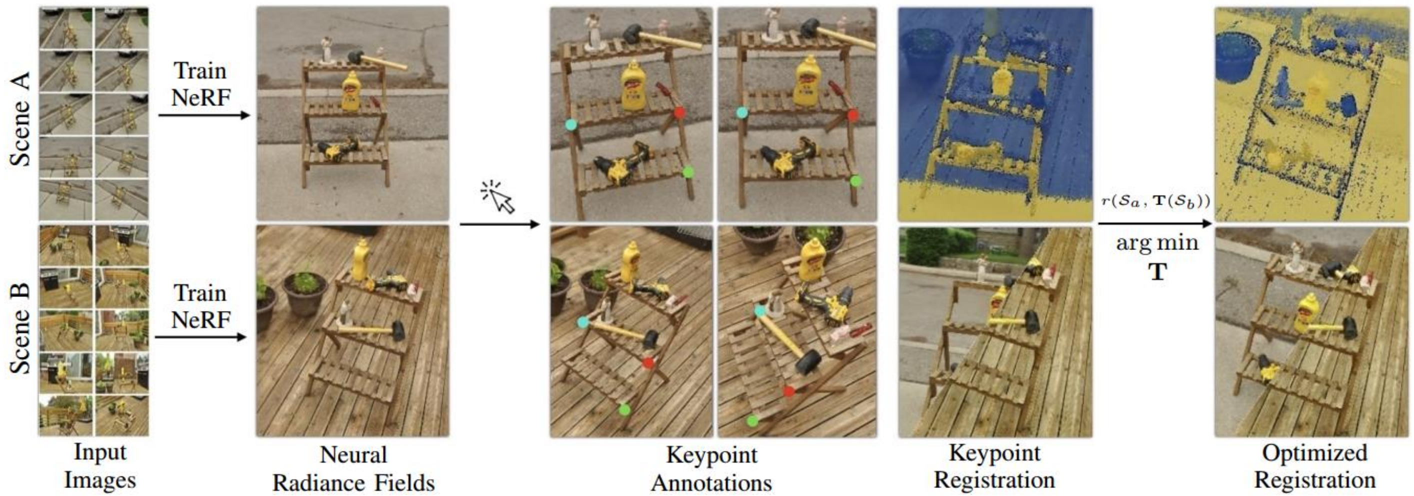
(c) DReg-NeRF aligns NeRF blocks into the same coordinate frame without accessing raw image data



Register a pair of NeRF models that are trained in different coordinate frames into a same coordinate frame

Seminal Work

nerf2nerf: Pairwise Registration of Nueral Radiance Fields



- Limitations
  - Initialized from human annotated keypoints
  - Based on traditional optimization method

Lily Goli, et, al. nerf2nerf: Pairwise Registration of Nueral Radiance Fields. ICRA 2023

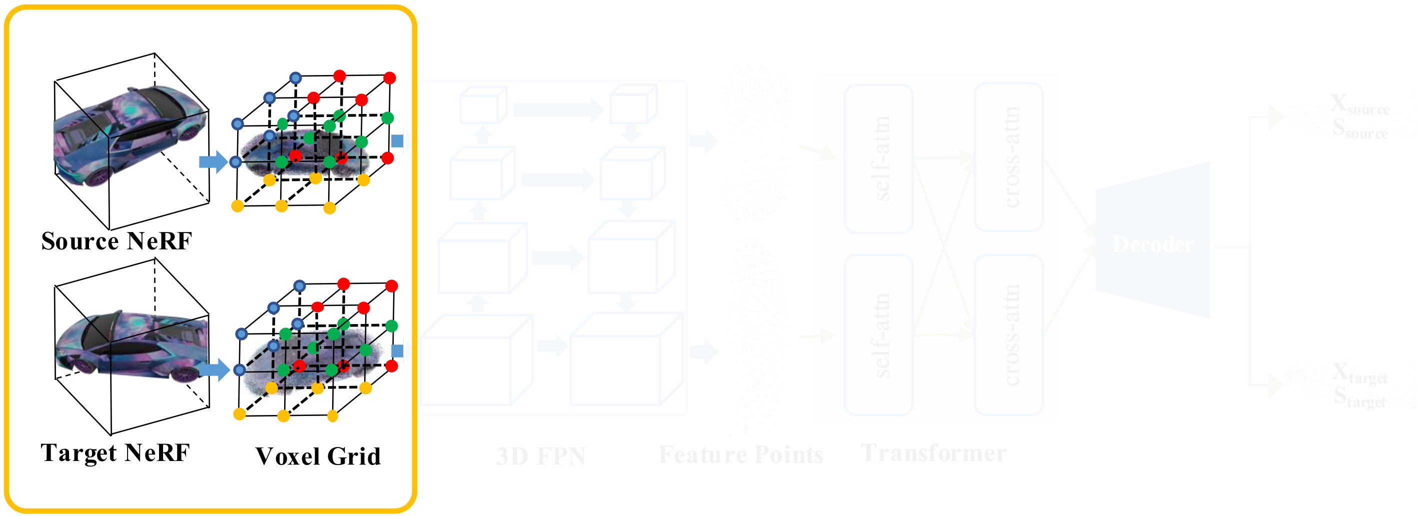
Key Idea

Data Driven; End-to-End

- **NeRF2NeRF**
  - Initialized from human annotated keypoints
  - Based on traditional optimization method
- **Ours (DReg-NeRF)**
  - Generalizable
  - End-to-end

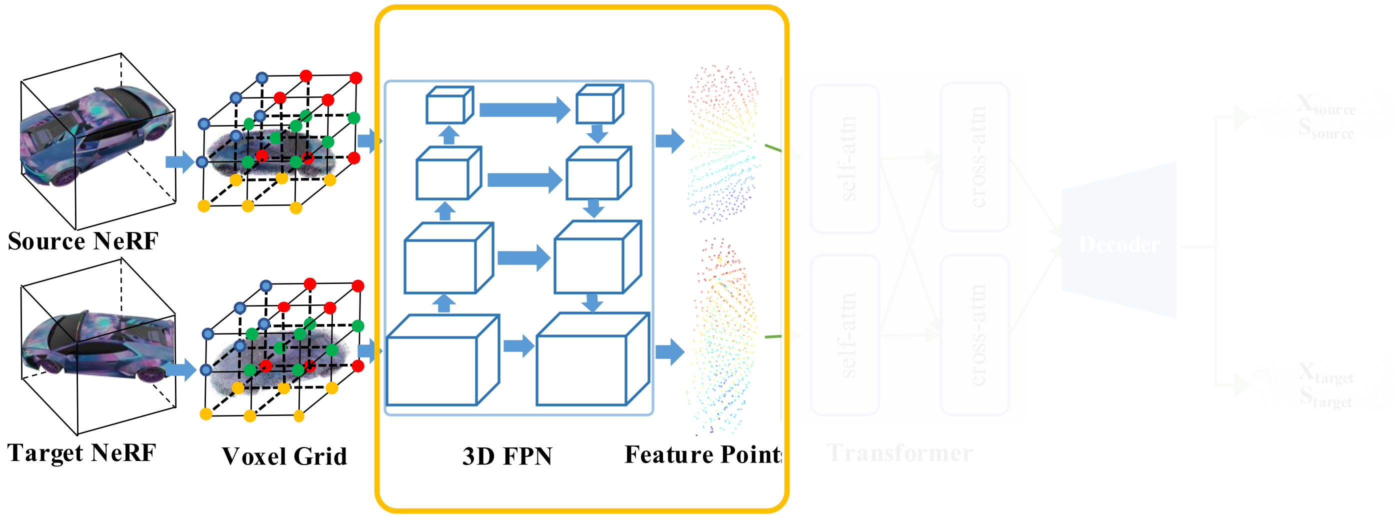
Key Idea

Data Driven; End-to-End



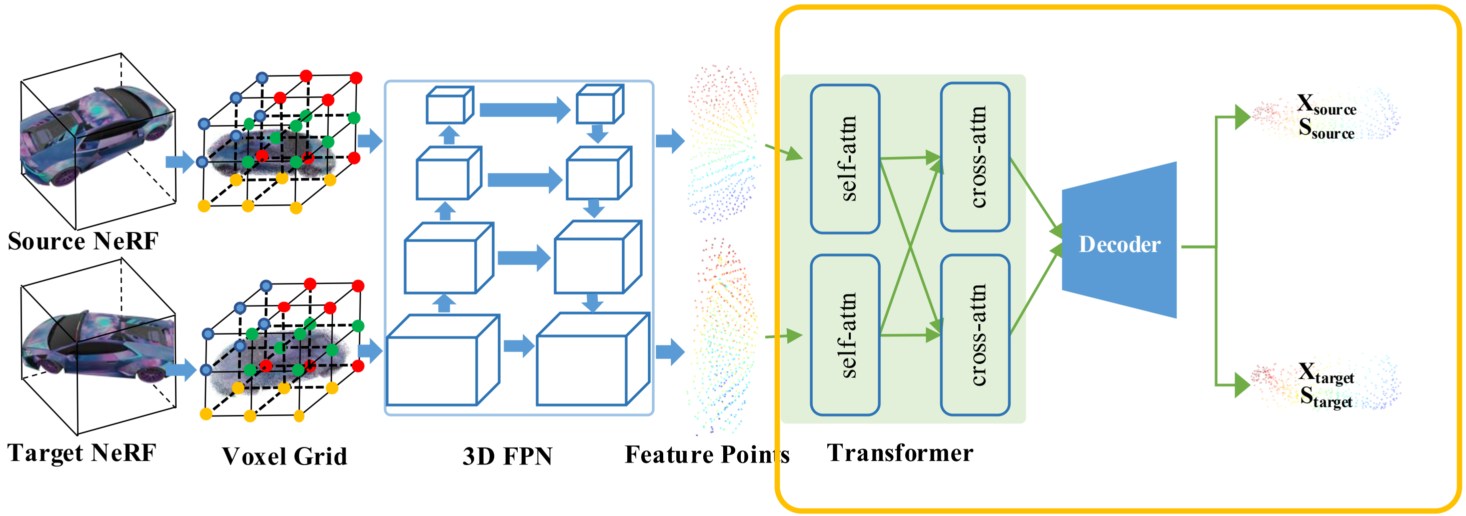
Key Idea

NeRF Feature Extraction



Key Idea

Enhanced Feature with Transformer



Dataset Construction

Absence of Training Data for NeRF Registration

- NeRF2NeRF
  - Not data driven
  - Contain only 6 synthesized scenes

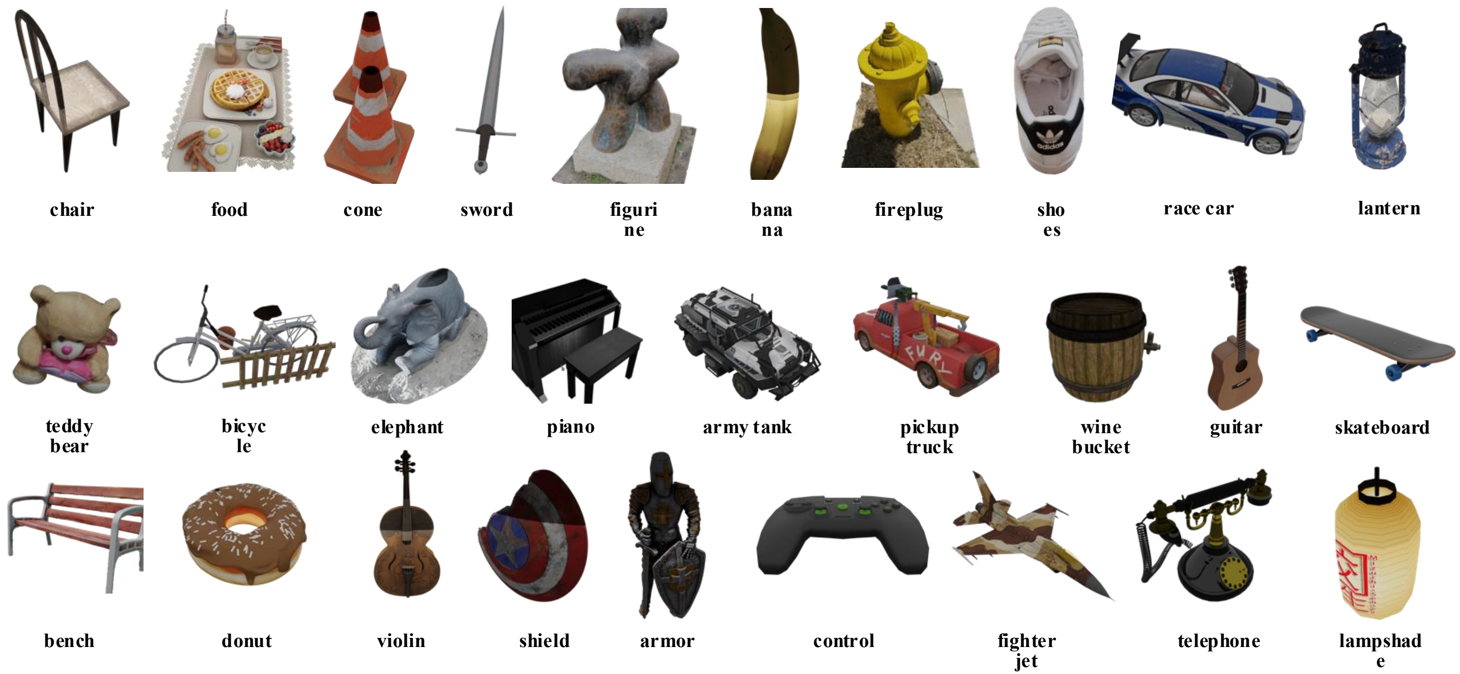
Matt Deitke, *et, al.* Objaverse: A Universe of Annotated 3D Objects. CVPR 2023

Dataset Construction

Matt Deitke, *et, al.* Objaverse: A Universe of Annotated 3D Objects. CVPR 2023

Dataset Construction

Selected 30+ categories, each category contains 40-80 objects

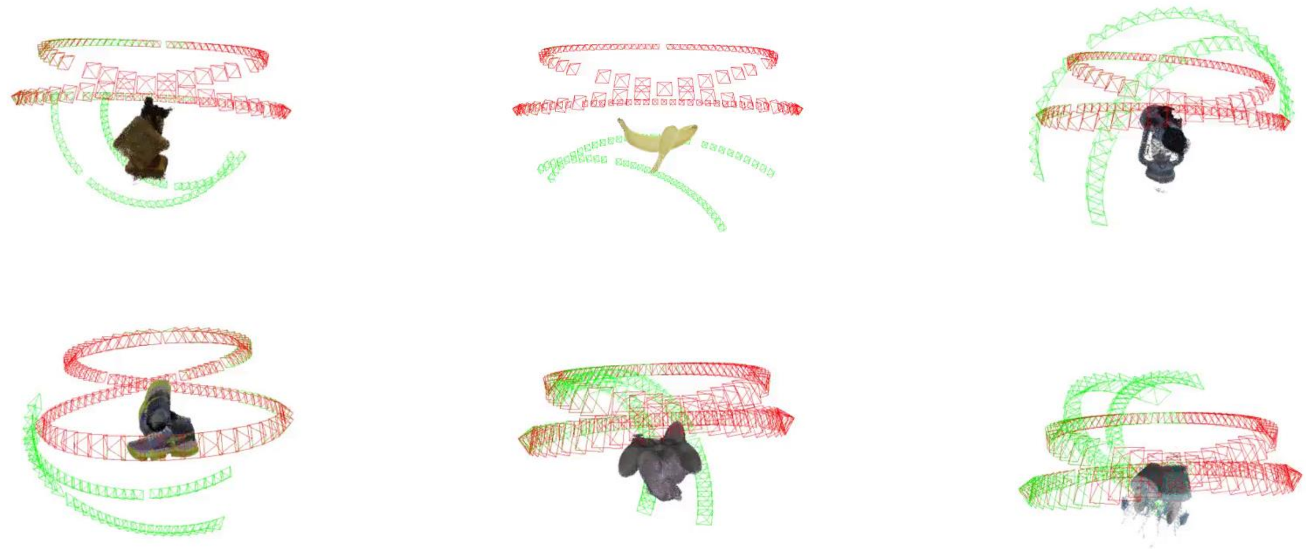


- Collected 1700+ 3D objects from Objaverse dataset
  - Render **120 images** from distinct view points per object with **blender**
  - Take one week with 8 concurrent processes

Matt Deitke, *et, al.* Objaverse: A Universe of Annotated 3D Objects. CVPR 2023

Dataset Construction

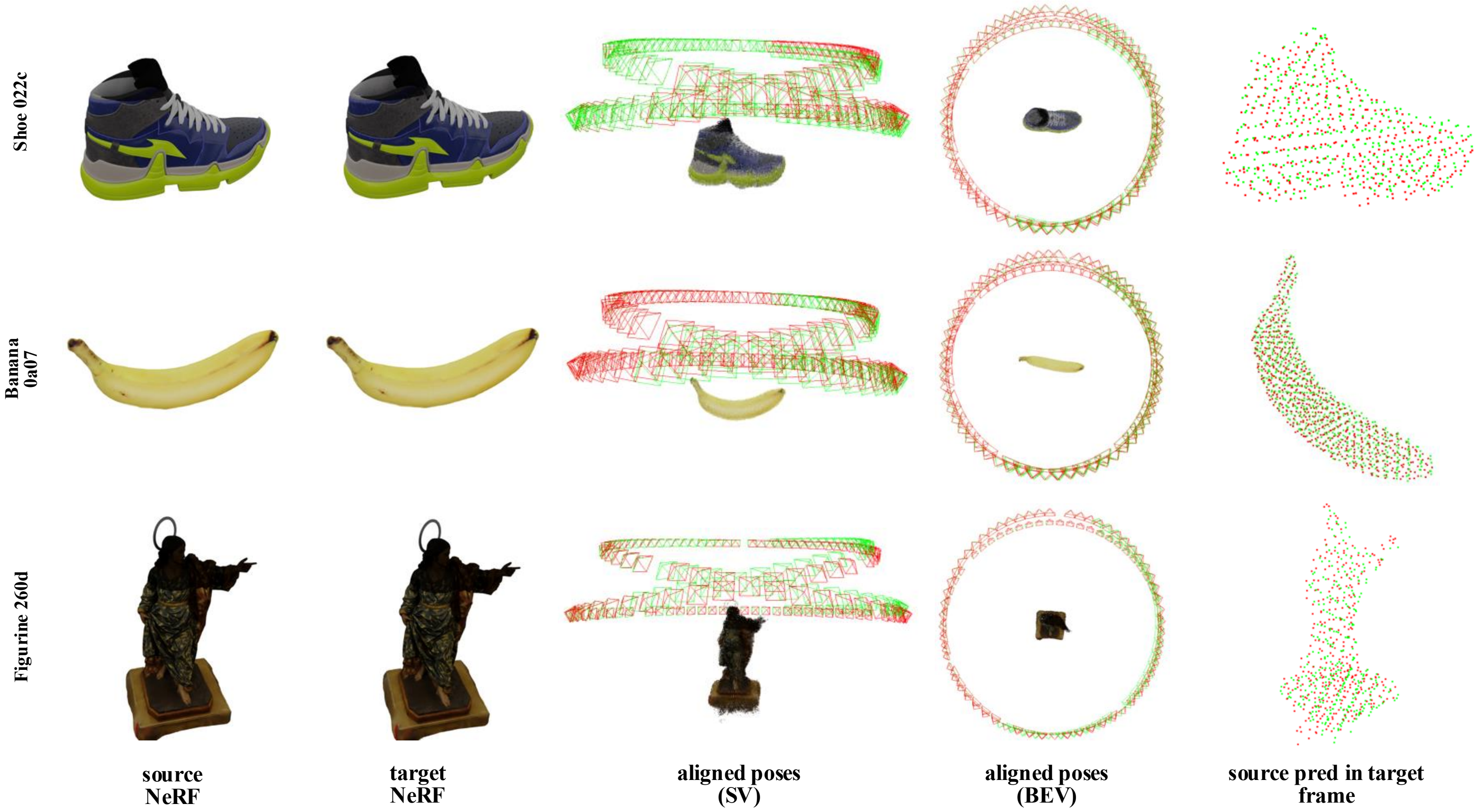
Trained 1,700+ pairwise NeRF models



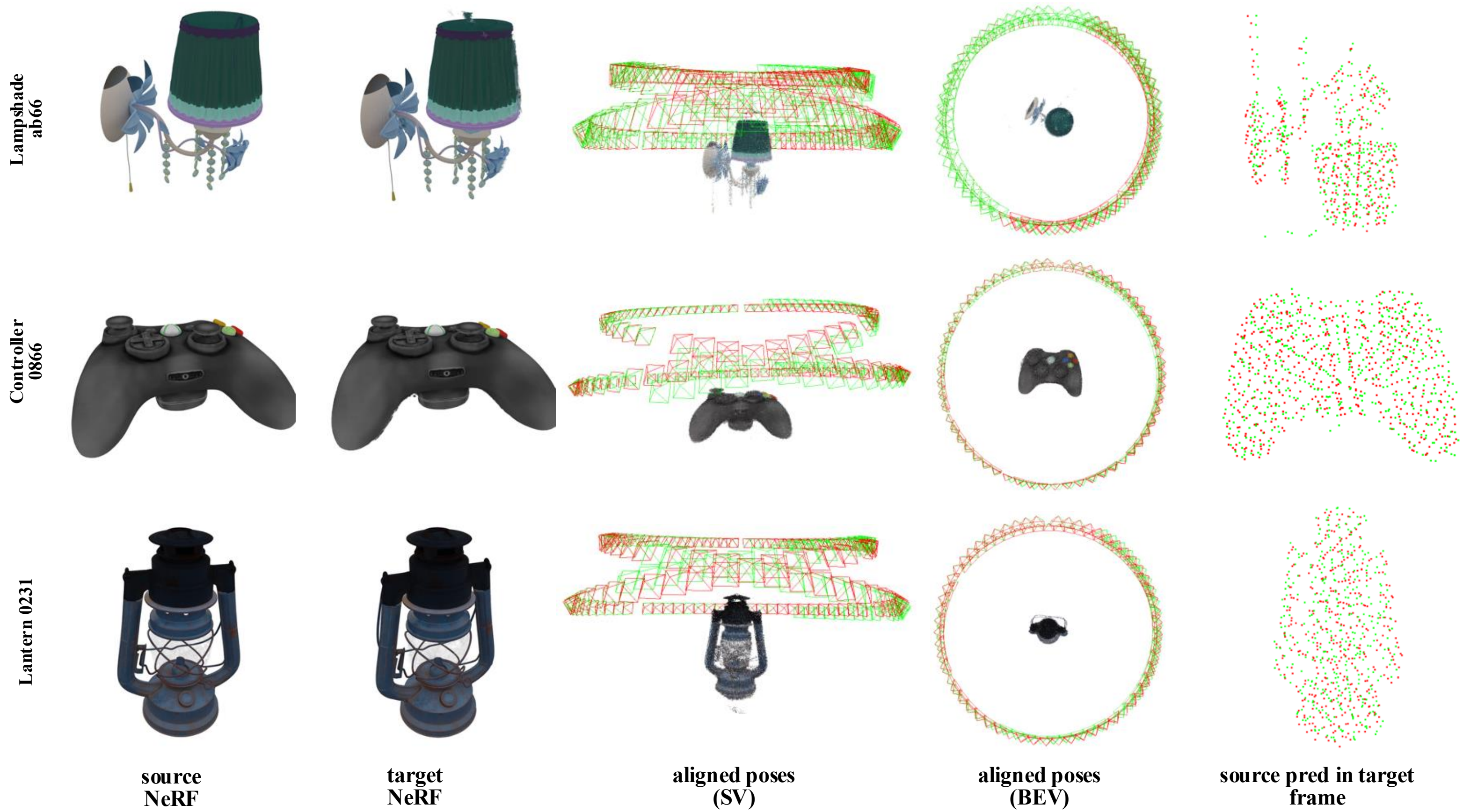
- Trained 1700+ pairwise NeRF models with the collected objects
  - Images are split into two blocks by KMeans
  - Perturb the coordinate frame with a randomly generated 3D transformation
  - Take one week with a 4090 GPU

Matt Deitke, *et, al.* Objaverse: A Universe of Annotated 3D Objects. CVPR 2023

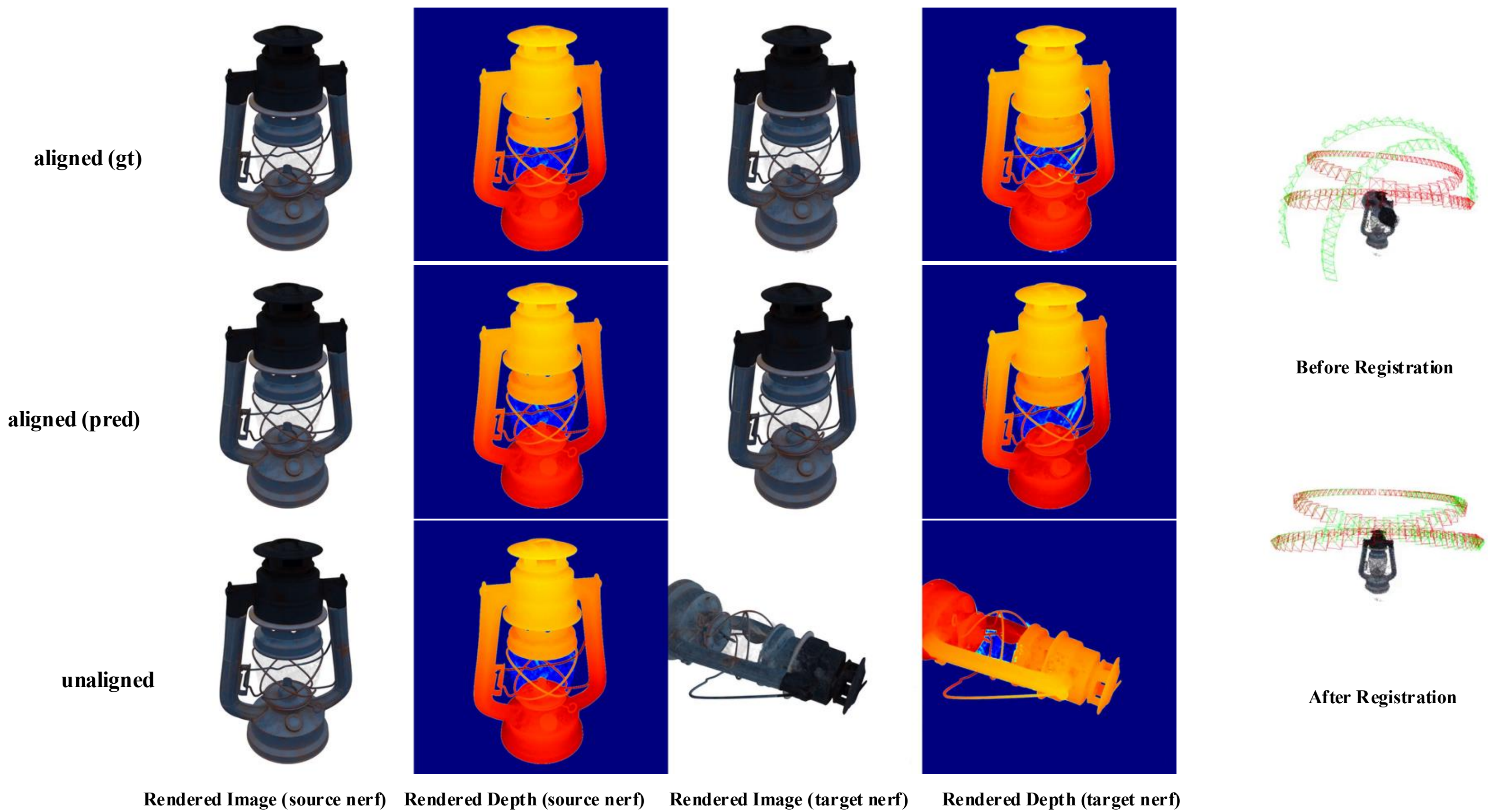
Generalize to Unseen Objects



Generalize to Unseen Objects

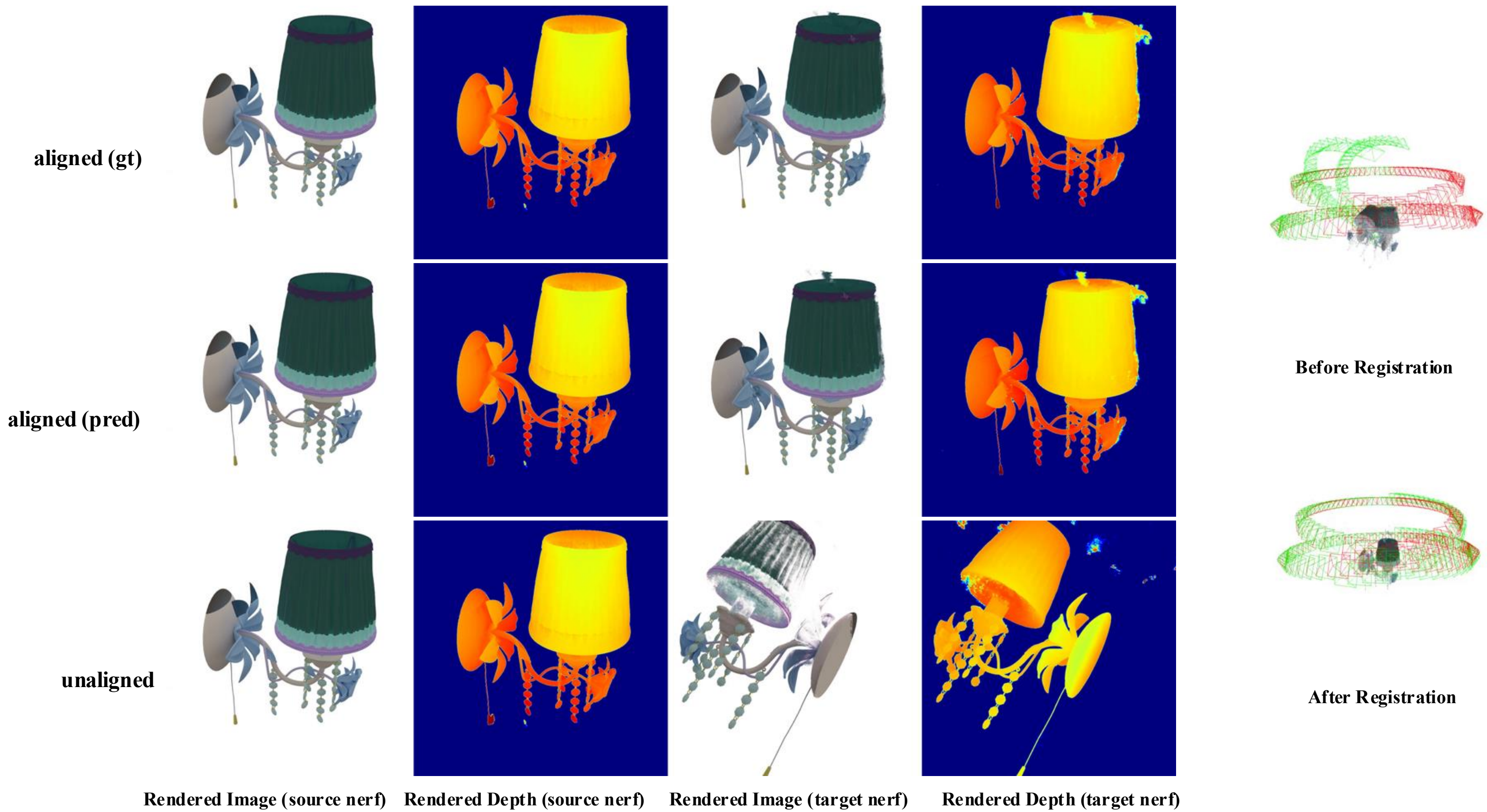


Generalize to Unseen Objects



Lantern 0231

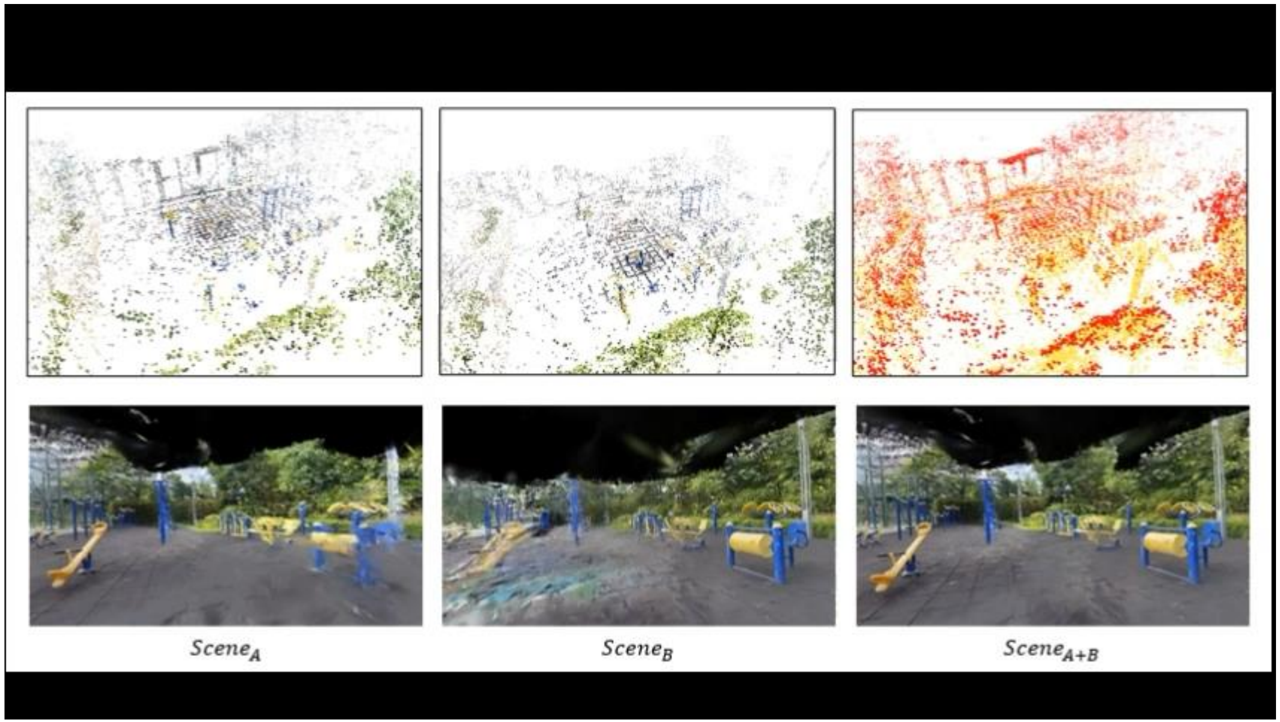
Generalize to Unseen Objects



Lampshade ab66

Inspired Follow-up Work

GaussReg – Scene Level



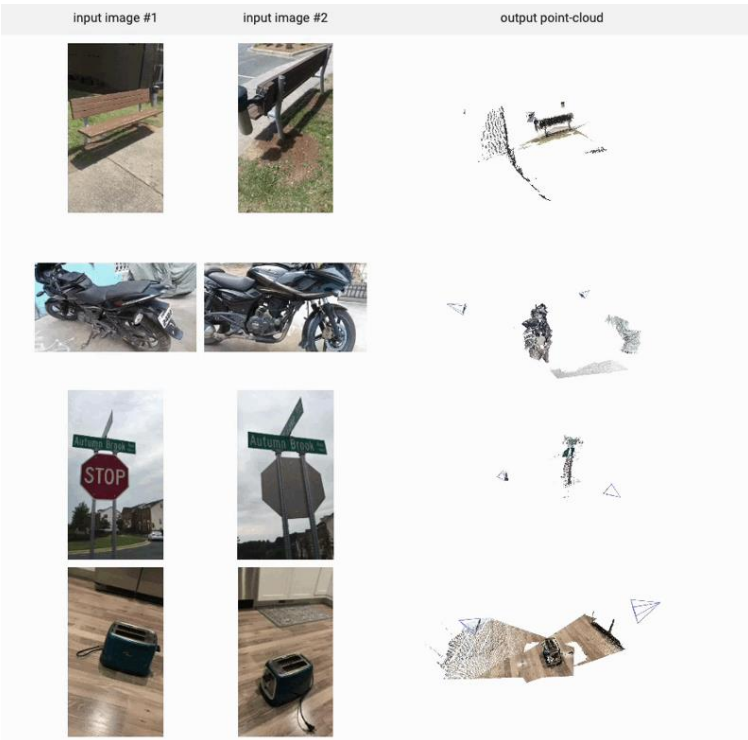
Jiahao Chang, *et, al.* GaussReg: Fast 3D Registration with Gaussian Splatting. ECCV 2024

What is Next?

From 3D reconstruction to 3D foundation model

From DUSt3R to VGGT

Two Views to Multiple Views



DUSt3R

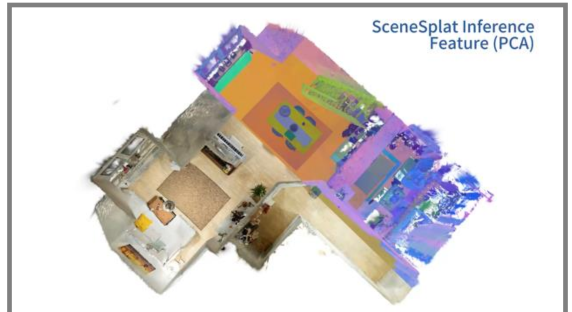
VGGT

- Shuzhe Wang, *et.al*. **DUSt3R: Geometric 3D Vision Made Easy.** CVPR 2024
- Jianyuan Wang, *et.al*. **VGGT: Visual Geometry Grounded Transformer**. CVPR 2025 Best Paper Award

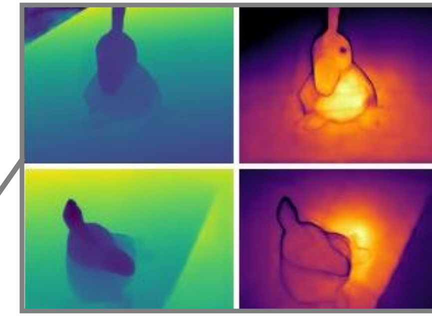
3D Foundation Model

Camera Pose Estimation

Scene Understanding



Depth Estimation



Unconstrained
image collections

Unified Model

Dense Reconstruction

Scene Generation

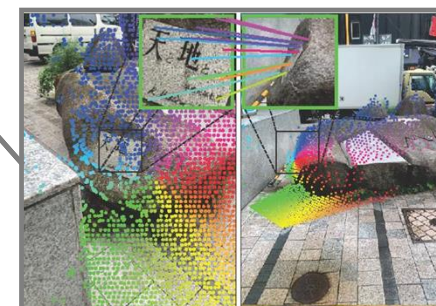
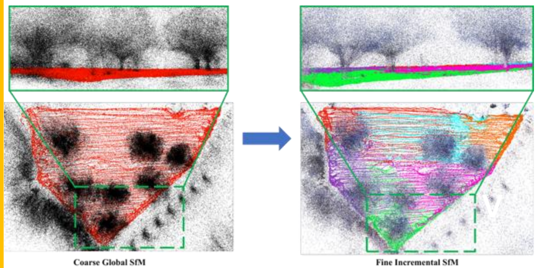


Image Matching

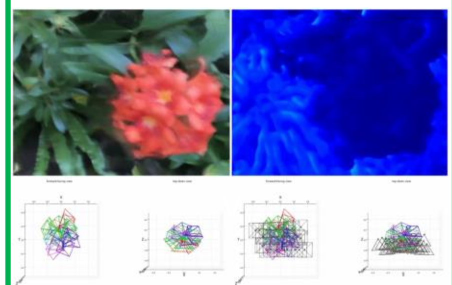
Large Scale Neural 3D Scene Reconstruction, Rendering, and Beyond

Chen Yu

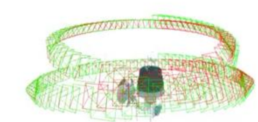


AdaSfM
ICRA 2023

DOGS
NeurIPS 2024



DBARF
CVPR 2023



DReg-NeRF
ICCV 2023

NUS National University of Singapore | School of Computing

CVRP

Acknowledgements

Supervisor

Gim Hee Lee

Examiners

Angela Yao

Leow Wee Kheng

Collaborators

Li Jianming

Song Shu

Yu Zihao

Yu Tianning

Low Wengfei

Yan Zhiwen

Rolandos Potamias

Evangelos Ververas

Song Jifei

Deng Jiankang

Thanks for you listening!