THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

CVPR SEATTLE, WA JUNE 17-21, 2024

Benchmarking Implicit Neural Representation and Geometric Rendering in Real-Time RGB-D SLAM

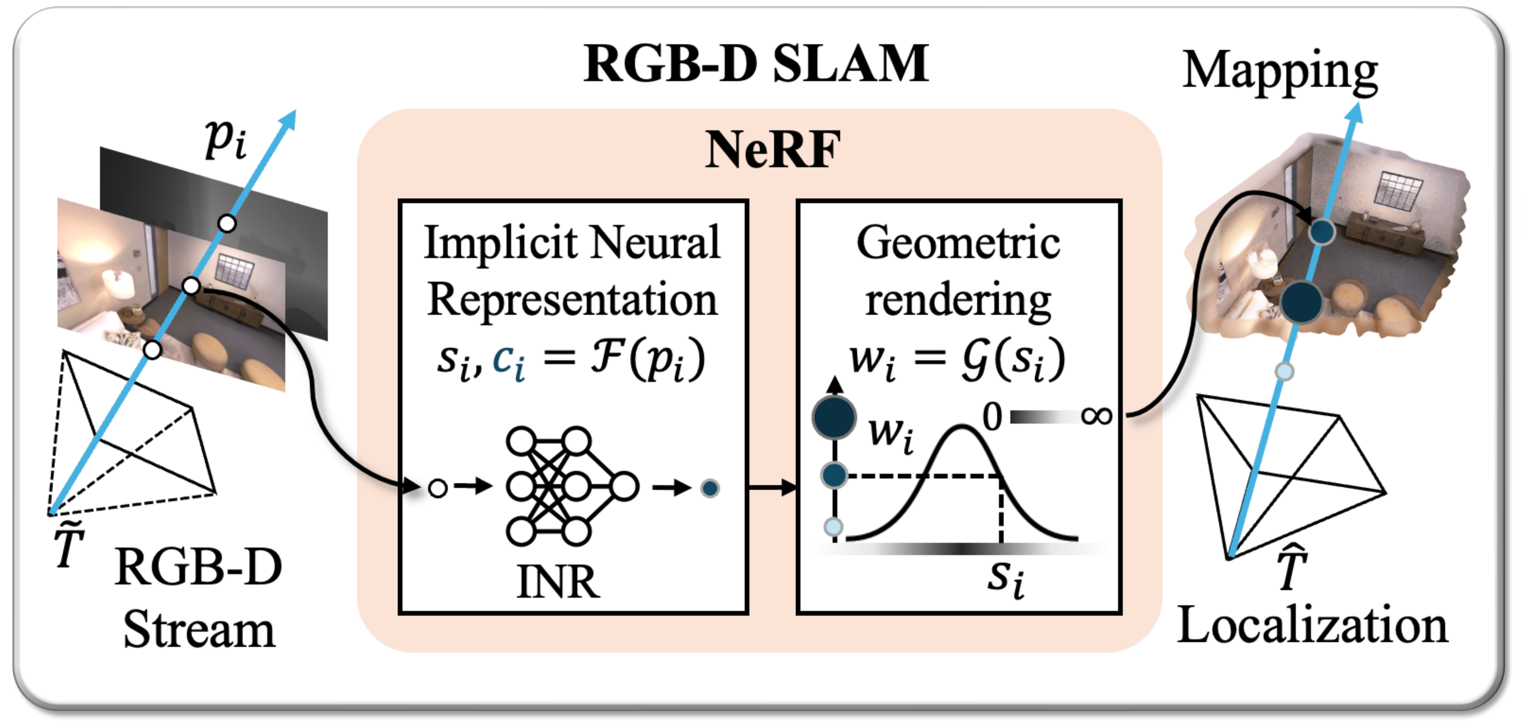
Tongyan Hua[1], Lin Wang[1,2,*]

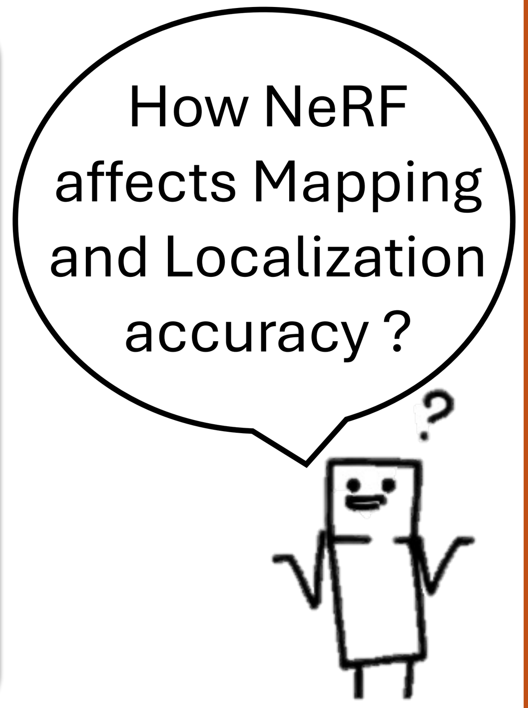
[1]HKUST(GZ), [2]HKUST

Project Page: https://vlis2022.github.io/nerf-slam-benchmark/

VLIS LAB   Project Page   Tongyan Hua

How the Hell does NeRF Matter for SLAM ?





Challenges in NeRF-based SLAM systems

➤ **No Unified SLAM framework**

NeRF Variant 1 | NeRF Variant 2 | NeRF Variant 3

Difficult to track progress introduced by NeRF variants due to varied SLAM system strategies .

➤ **No dissection of NeRF variants**

$\mathcal{F}_1$ $\mathcal{G}_1$ | $\mathcal{F}_2$ $\mathcal{G}_2$ | $\mathcal{F}_3$ $\mathcal{G}_3$

Different components of variants affect SLAM performance and varied on scene basis.

Why Benchmarking?

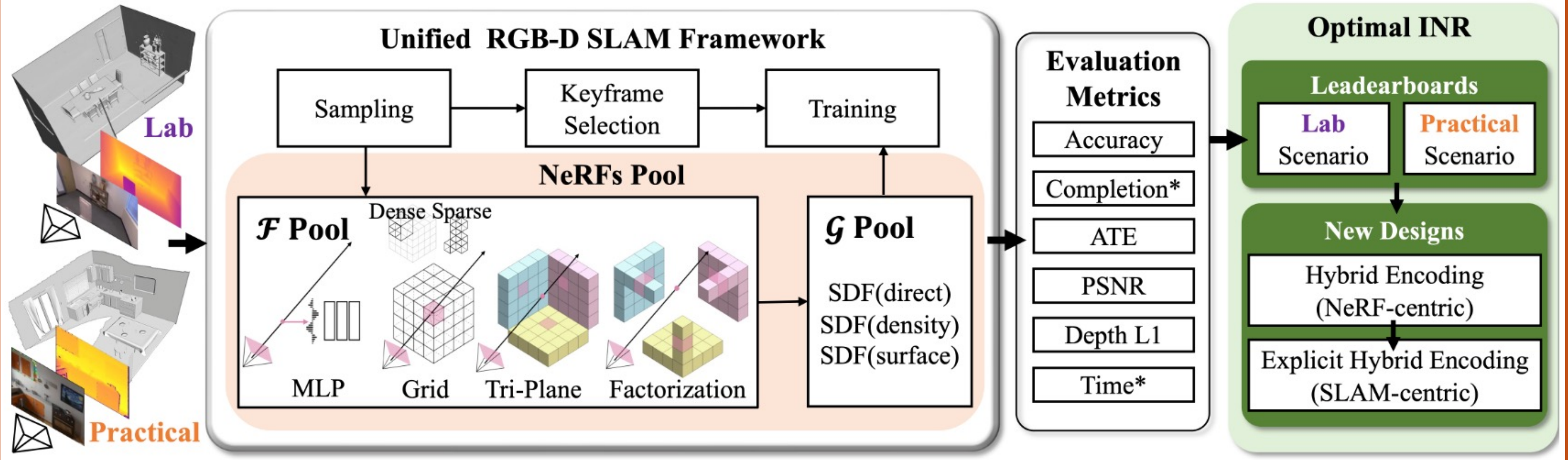
➤ **Unified framework for NeRFs**

NeRF Variant 1 | NeRF Variant 2 | NeRF Variant 3

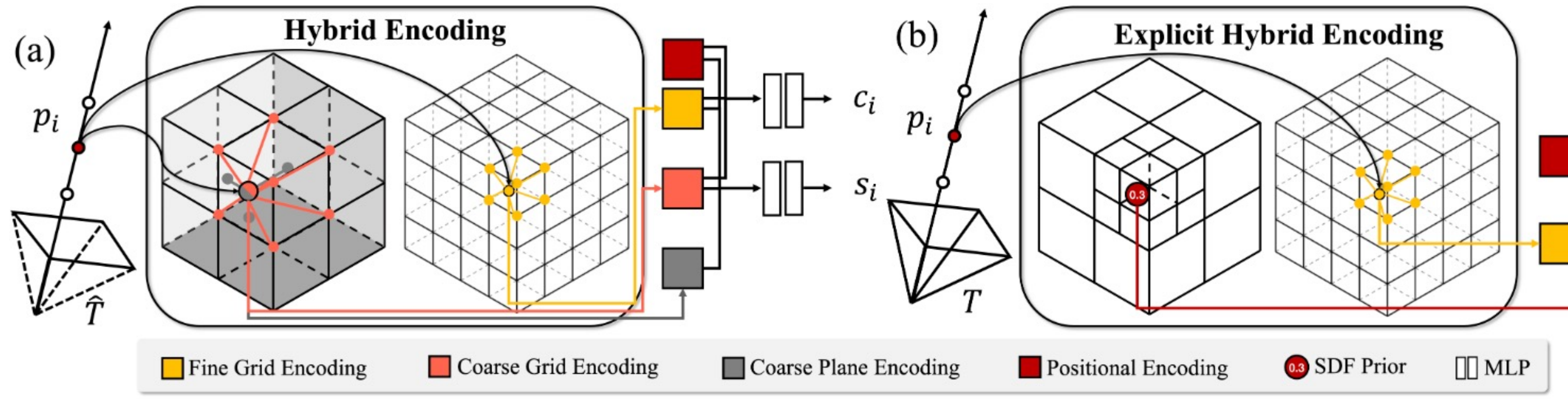
➤ **Systematic $\mathcal{F}$ + $\mathcal{G}$ evaluation**

$\mathcal{F}_1$ $\mathcal{G}_1$ | $\mathcal{F}_1$ $\mathcal{G}_2$ | $\mathcal{F}_1$ $\mathcal{G}_3$ ... ...

Methodology



The proposed pipeline for NeRF-SLAM benchmark.



Inspired New Designs.

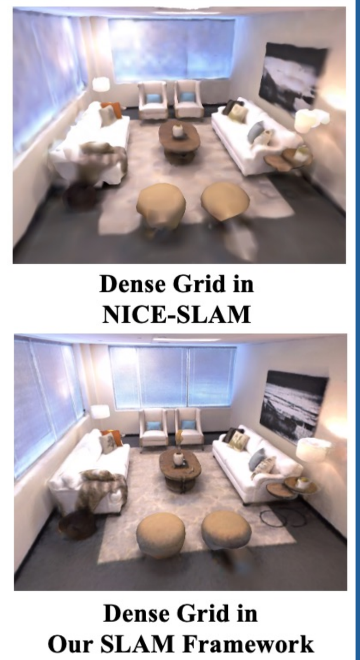
Hybrid Encoding & Explicit Hybrid Encoding

1. **Hybrid Encoding**, a strategy **blending** the **comprehensiveness** of decomposition with the **precision** of grid-based methods;

2. **Explicit hybrid encoding**, a strategy substituting the coarse-level feature grid with an octree structure and **simplifying the encoding process** by using a single-level dense grid, boosting mapping efficiency.

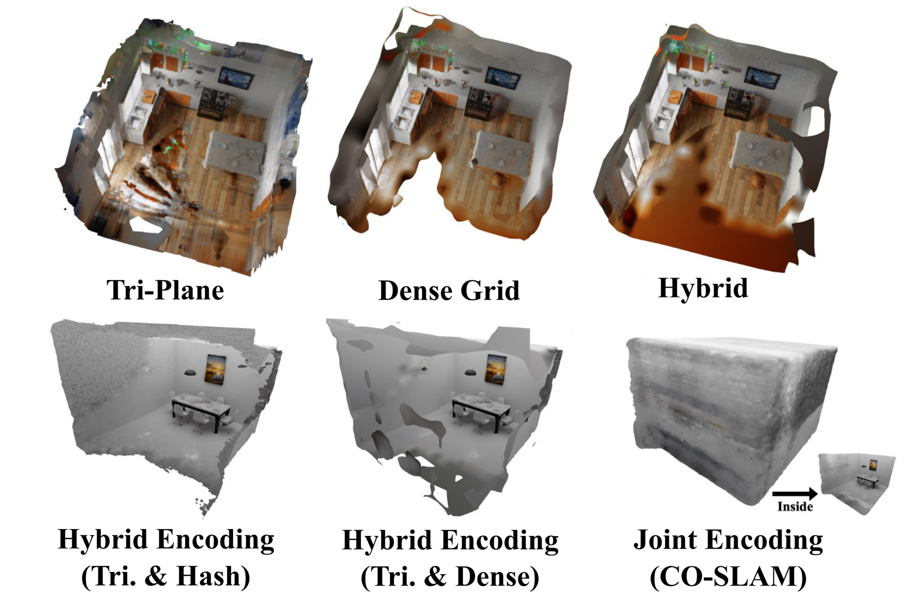
Results

$\mathcal{G}$	$\mathcal{F}$	From results				From processes			
		Acc.[cm]↓	Comp.[cm]↓	Comp.[%]↑	ATE[cm]↓	PSNR[db]↑	Depth L1[cm]↓	Tracking[ms]↓	Mapping[ms]↓
SDF(direct)	MLP	11.95	8.36	76.82	14.41	24.20	3.12	293	421
	Dense[1]	1.65	5.62	83.93	1.37	27.88	**1.50**	288	**286**
	Sparse[3]	1.76	5.66	83.61	1.40	**28.23**	1.65	**197**	300
	Tri.	1.69	5.64	83.60	1.42	27.52	1.80	352	820
	Fact.	1.69	5.60	83.69	1.50	27.52	1.74	419	911
SDF(density)	MLP	9.64	9.92	72.22	24.58	23.51	7.01	250	419
	Dense[2]	**1.60**	**5.58**	**84.01**	**1.31**	27.77	4.42	253	585
	Sparse[4]	1.69	5.65	83.72	1.40	28.10	4.45	207	310
	Tri.[5]	1.80	5.59	83.82	1.49	27.55	4.51	376	829
	Fact.	1.75	5.60	83.73	1.55	27.54	4.48	414	897
SDF(surface)	MLP	30.44	24.28	20.21	44.68	17.13	98.24	342	585
	Dense	32.83	20.71	42.28	135.39	16.21	176.15	418	669
	Sparse	48.22	25.73	30.73	176.68	16.49	174.94	258	359
	Tri.	30.28	18.06	41.77	87.71	16.05	180.68	490	897
	Fact.	30.75	20.45	31.31	85.63	16.43	168.81	584	1009



Leaderboard of lab scenario: Dense Grid Representation excels.

$\mathcal{G}$	$\mathcal{F}$	Indicators			Targets		
		Acc.↓ [cm]	Comp.↓ [cm]	Comp.↑ [%]	ATE↓ [cm]	PSNR↑ [db]	Depth L1↓ [cm]
SDF(direct)	MLP	4.37	5.16	79.22	3.69	22.56	3.56
	Dense	2.69	4.69	83.45	1.96	25.15	1.70
	Sparse	2.84	4.81	82.64	2.12	25.23	1.84
	Tri.	2.29	**4.42**	**84.01**	1.89	24.67	1.85
	Fact.	2.62	4.47	83.54	1.94	24.69	1.87
	Hybrid	2.40	4.64	83.48	**1.86**	**25.25**	**1.68**
SDF(density)	MLP	3.98	5.12	78.82	3.87	22.56	4.80
	Dense	2.70	4.72	83.27	1.87	25.04	4.40
	Sparse	2.84	4.71	82.79	1.96	25.08	4.46
	Tri.	2.12	4.62	83.90	1.90	24.62	4.42
	Fact.	**2.11**	4.45	**84.01**	2.01	24.63	4.41
	Hybrid	2.34	4.71	83.25	1.91	25.05	4.40



Leaderboard of Practical scenario: novel Hybrid Encoding excels.

	Lab		Practical	
	NICE-SLAM [54]	Ours	CO-SLAM [40]	Ours
Depth L1↓	3.53	**1.50**	3.02	**1.68**
Acc.↓	2.85	**1.65**	2.95	**2.40**
Comp.↓	**3.00**	5.62	**2.96**	4.64
Comp.%↑	**89.33**	83.93	**86.88**	83.48
ATE↓	1.95	**1.37**	-	**1.86**
Res×Dim↓	3×32	**2×2**	16×2	**2×2**

Comparison with existing SOTA & Inspired Explicit Hybrid Encoding.

We demonstrated the superior efficacy of dense grid representations and introduced (explicit) hybrid encoding strategies for improvement.