Refined Tensorial Radiance Field: Harnessing Coordinate-Based Networks for Novel View Synthesis from Sparse Inputs

Optimization and Statistical Inference LAB

Mingyu Kim[1], Jun-Seong Kim[2], Se-Young Yun[1] and Jin-Hwa Kim[3]

KAIST AI[1)], POSTECH EE[2)], NAVER AI Lab. & SNU AIIS[3)]

{callingu, yunseyoung}@kaist.ac.kr, gucka28@postech.ac.kr, j1nhwa.kim@navercorp.com

NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

The multi-plane encoding approach allows NeRFs to learn fine-grained details rapidly and achieves outstanding performance, however, it has limitations in representing the global context of the scene such as object shapes and dynamic motion over times when available training data is sparse.

In this work, we propose refined tensorial radiance fields that harness coordinate-based networks capturing low-frequency signals, while multi-plane network focuses on fine-grained details simultaneously. Empirically, the proposed outperform all baselines for the task with static and dynamic scenes under sparse inputs.

Motivation

The multi-plane encoding with TV loss still struggles to represent global context in the sparse-inputs (HexPlane, CVPR2023).

T=0.0 T=0.2 T=0.5 T=0.7 T=1.0

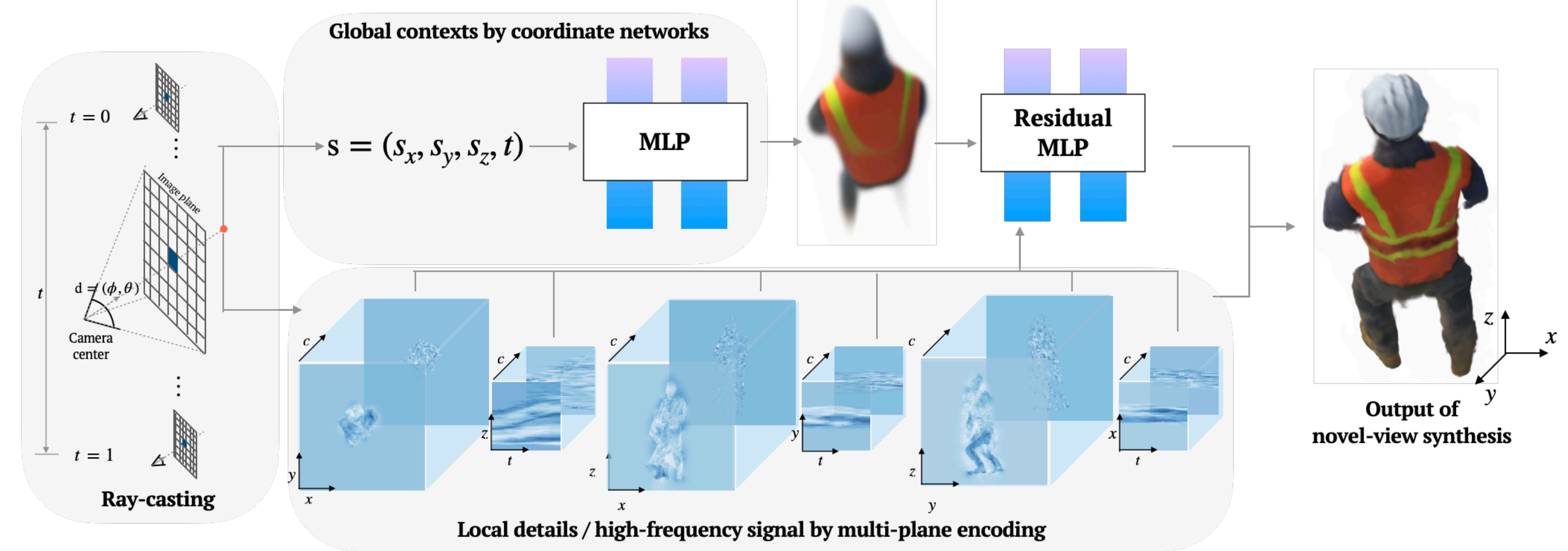
HexPlane (CVPR2023)

T=0.0 T=0.2 T=0.5 T=0.7 T=1.0

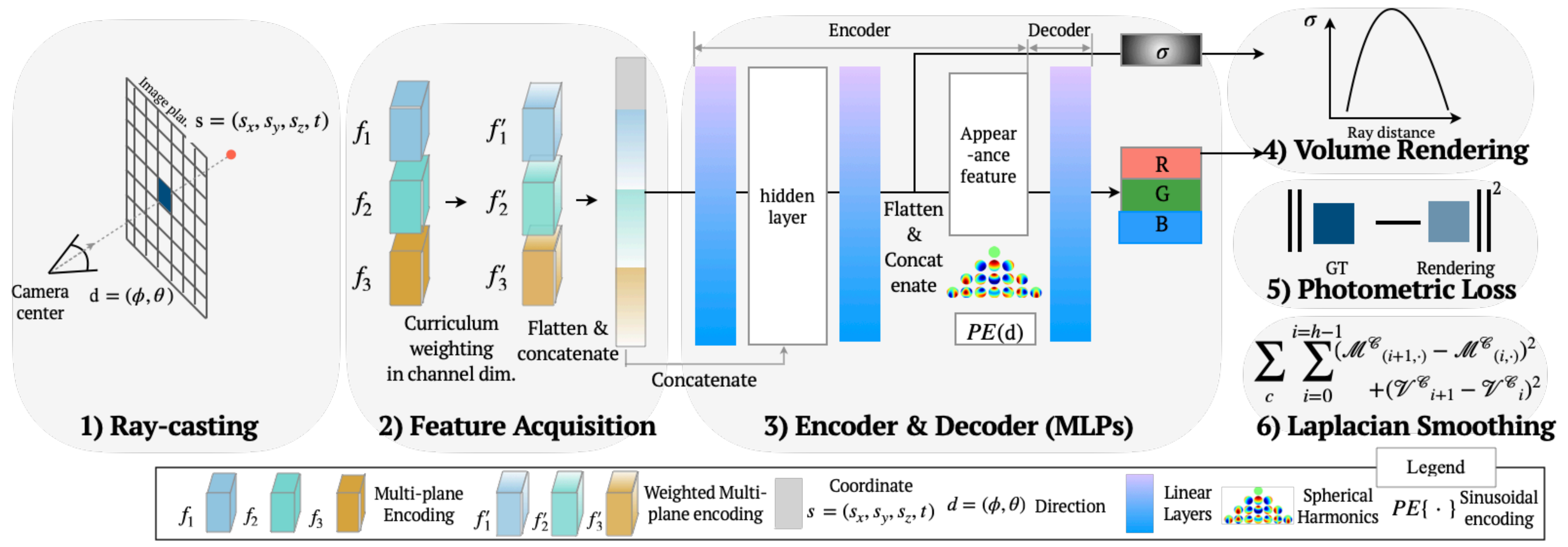
Proposed Method

Method

We devise incorporation ReLU-MLPs without positioning encoding with multi-plane grid features.



The detailed architecture is represented as follows.



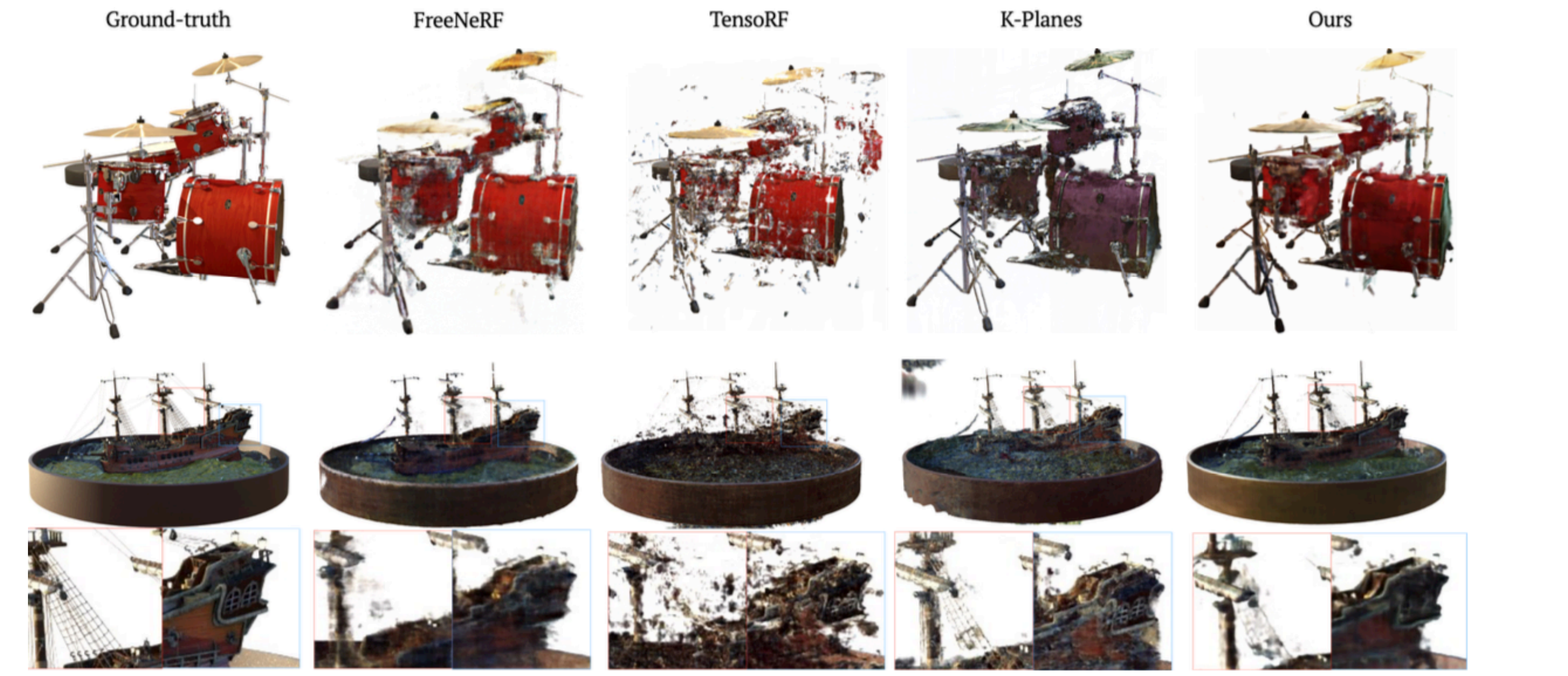
Quantitative Results

Evaluation on Static NeRF dataset (8 views)

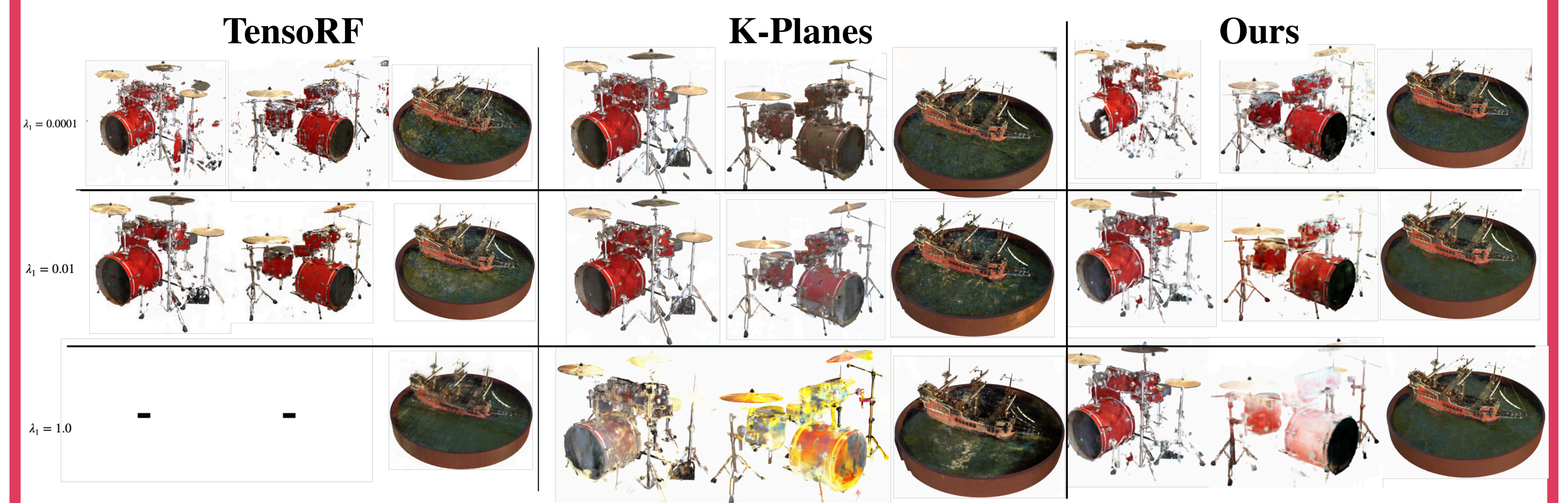
Models	PSNR ↑ chair	drums	ficus	hotdog	lego	materials	mic	ship	Avg. PSNR ↑
Simplified_NeRF	20.35	14.19	21.63	22.57	12.45	18.98	24.95	18.65	19.22
DietNeRF	21.32	14.16	13.08	11.64	16.12	12.20	24.70	19.34	16.57
HALO	24.77	18.67	21.42	10.22	22.41	21.00	24.94	21.67	20.64
FreeNeRF	26.08	19.99	18.43	28.91	24.12	21.74	24.89	23.01	23.40
DVGO	22.35	16.54	19.03	24.73	20.85	18.50	24.37	18.17	20.57
VGOS	22.10	18.57	19.08	24.74	20.90	18.42	24.18	18.16	20.77
iNGP	24.76	14.56	20.68	24.11	22.22	15.16	26.19	17.29	20.62
TensoRF	26.23	15.94	21.37	28.47	26.28	20.22	26.39	20.29	23.15
K-Planes	27.30	**20.43**	**23.82**	27.58	26.52	19.66	**27.30**	21.34	24.24
Ours	**28.02**	19.55	20.30	**29.25**	**26.73**	**21.93**	26.42	**24.27**	**24.56**

Qualitative Results

Qualitatively, the proposed method enables to represent both low-frequency and high-frequency details simultaneously, whereas baselines inevitably suffer from performance issues due to artifacts



While baselines struggle to learn consistently with varying TV regularization. the proposed method effectively handles novel-view synthesis, leading to reliable performance regardless of the regularization.



Conclusion

We developed a new method where the coordinate network captures global context, like object shapes and dynamic motions, and it also incorporates multi-plane encoding to precisely describe the finest details.