Joint Optimization of Neural Radiance Fields and Continuous Camera Motion from a Monocular Video

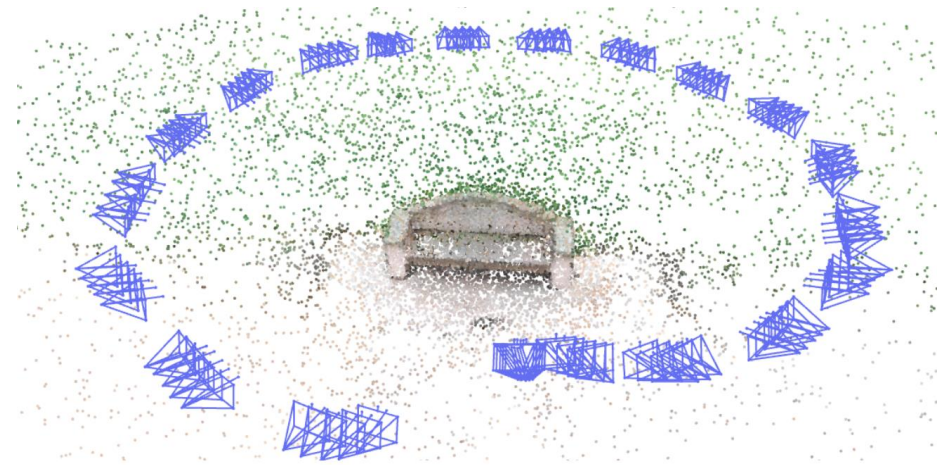Hoang Chuong Nguyen[1], Wei Mao, Jose M. Alvarez[2], Miaomiao Liu[1]

[1]Australian National University, [2]NVIDIA

CVPR 2025

Joint optimization of NeRF and camera poses from a monocular video
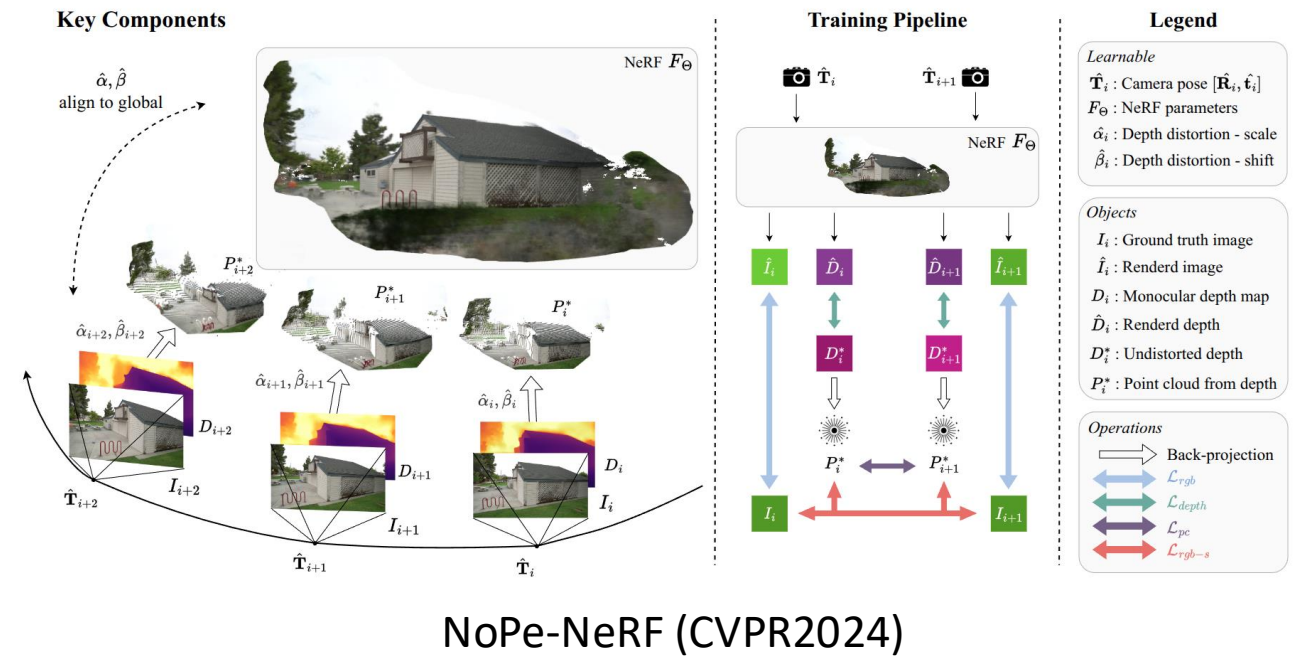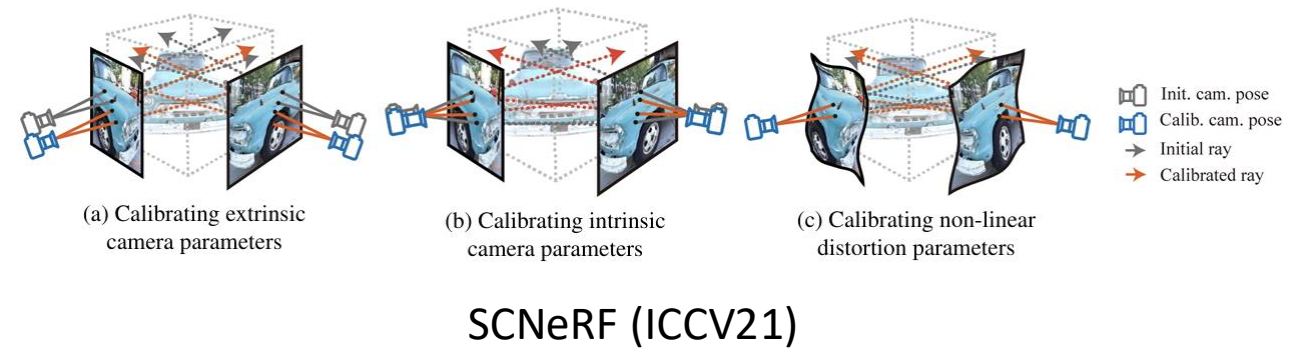

Input: Monocular video

Output: Scene geometry and camera poses. The scene geometry is represented via NeRF.

Related Works

- Prior methods jointly optimizes:
  - NeRF to represent the global scene geometry in the world space.
 - Camera-to-world mapppings.

→ Struggle in case large motions.

- Our contributions:
  - Optimize local scene geometry using a time-dependent NeRF.
  - Model camera continous motions.
  - Present a training pipeline:
    - handle large camera motions.
    - and accurately reconstruct the scene geometry.

(a) Calibrating extrinsic camera parameters

(b) Calibrating intrinsic camera parameters

(c) Calibrating non-linear distortion parameters

Init. cam. pose
Calib. cam. pose
Initial ray
Calibrated ray

SCNeRF (ICCV21)

NoPe-NeRF (CVPR2024)

Method

- Time-dependent NeRF:
  - Represents local scene at any timestep.
  - Enable the estimation of camera motions over a short time interval.

- Continous camera motion modelling:
  - Represent the continous motion as angular velocity and velocity at time t.
  - Avoid direct optimization of large motions.

- Camera motion over a longer time interval can be obtained by intergrating the predicted continous motions.
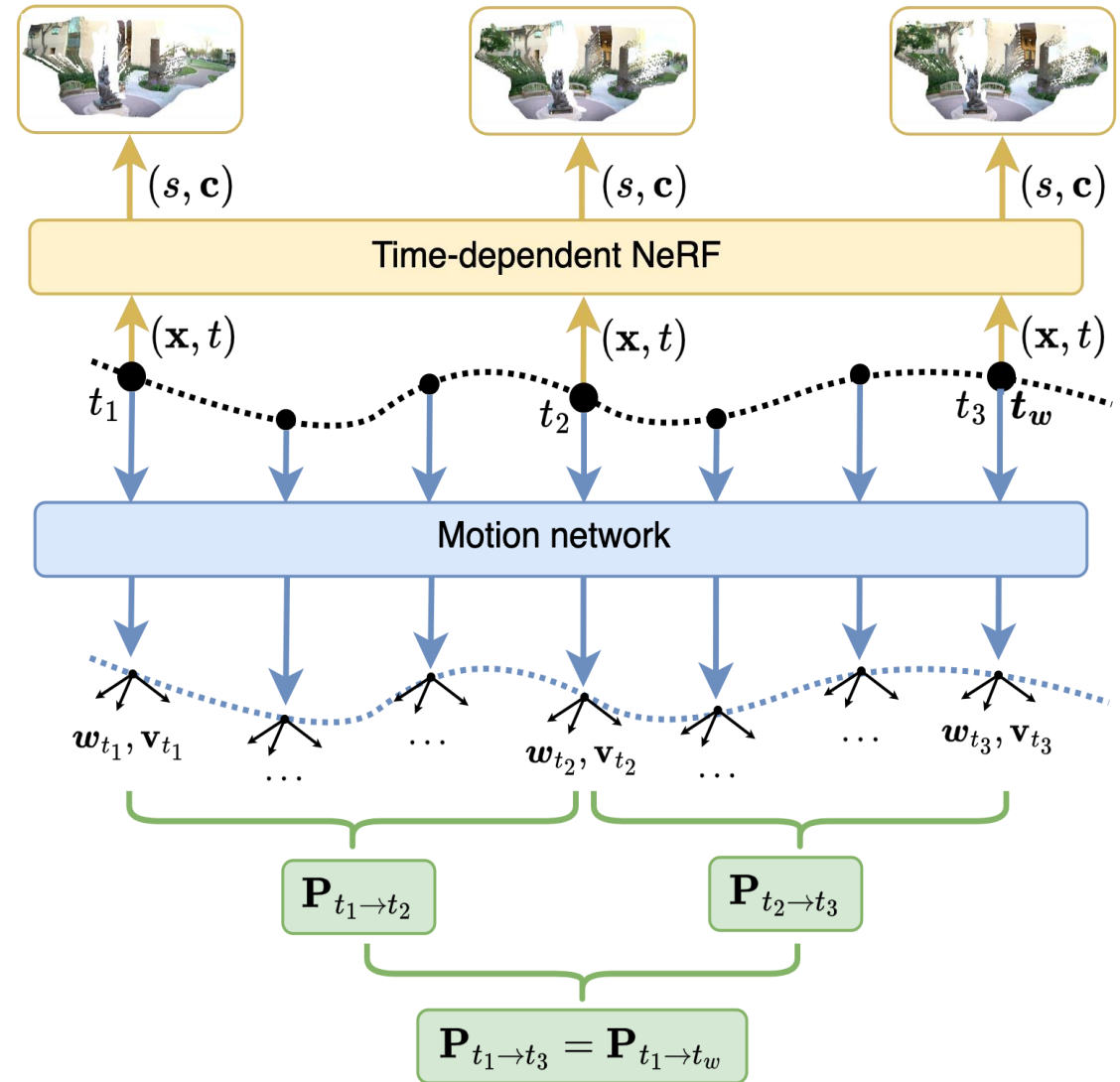
Training Pipeline

- Early training:
  - Goal: Learn accurate local scene geometry and camera continuous motion.
  - Jointly optimize the time-dependent NeRF and the motion network.

- Later training:
  - Goal: Learn accurate global scene geometry given the correct camera motions.
  - Select a world timestep t_w and compute the camera poses.
  - Fine-tune the time-dependent NeRF in the world space.
    - Follow NeRF standard training pipeline.
    - Input timestep is always t_w.

- End-to-end trainable.
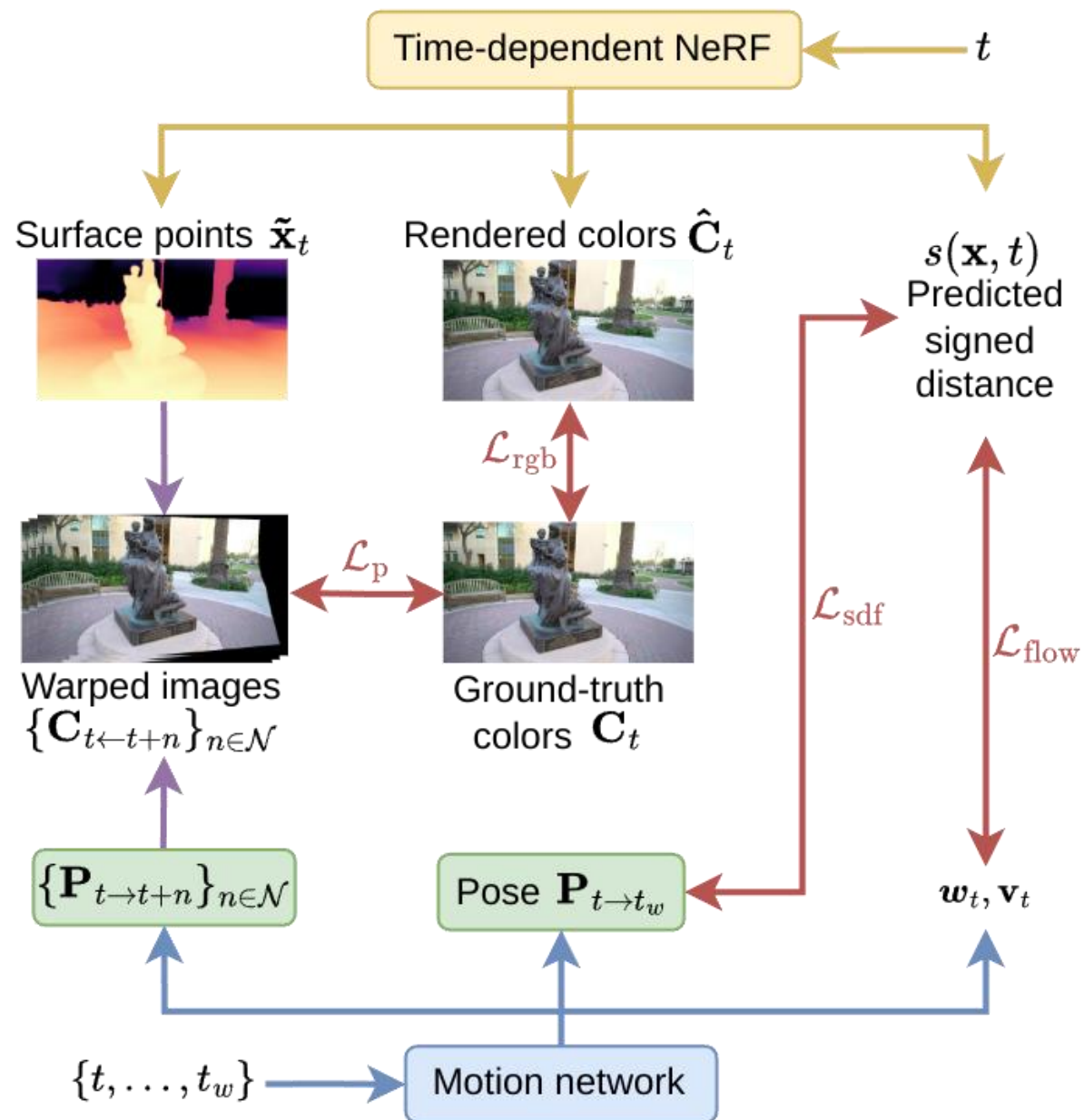
Results

|	Method	Novel-View Synthesis			Depth Estimation			Pose Estimation		
		PSNR↓	SSIM↑	LPIPS↓	AbRel↓	SqRel↓	δ_1 ↑	RPE_t↓	RPE_r↓	ATE↓
Scannet	NeRFmm	31.51	0.84	0.31	0.319	0.389	0.486	3.313	0.591	0.235
	NoPe-NeRF	31.87	0.83	0.38	_0.141_	0.137	_0.828_	0.808	0.180	0.030
	CF3DGS	_32.15_	_0.90_	**0.20**	0.157	_0.116_	0.803	_0.724_	_0.177_	_0.028_
	Ours	**34.09**	**0.91**	_0.24_	**0.063**	**0.025**	**0.952**	**0.632**	**0.160**	**0.020**
Co3D	NeRFmm	23.09	0.65	0.50	0.293	2.689	0.463	0.609	2.733	0.057
	NoPe-NeRF	_24.81_	_0.68_	0.46	_0.174_	_0.961_	0.728	0.337	1.581	0.049
	CF3DGS	21.57	_0.68_	0.36	0.212	3.749	_0.730_	_0.061_	_0.364_	_0.010_
	Ours	**27.78**	**0.76**	**0.34**	**0.034**	**0.195**	**0.974**	**0.025**	**0.070**	**0.002**

Ground-truth	Ours	NeRFmm [31]	NoPe-NeRF [2]	CF3DGS [6]
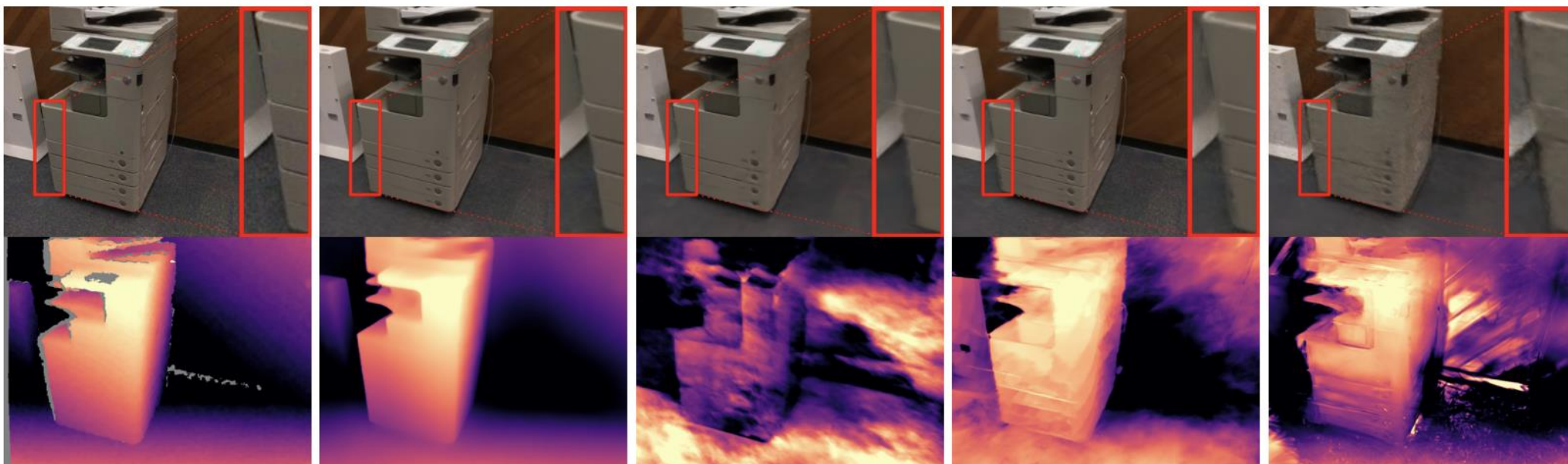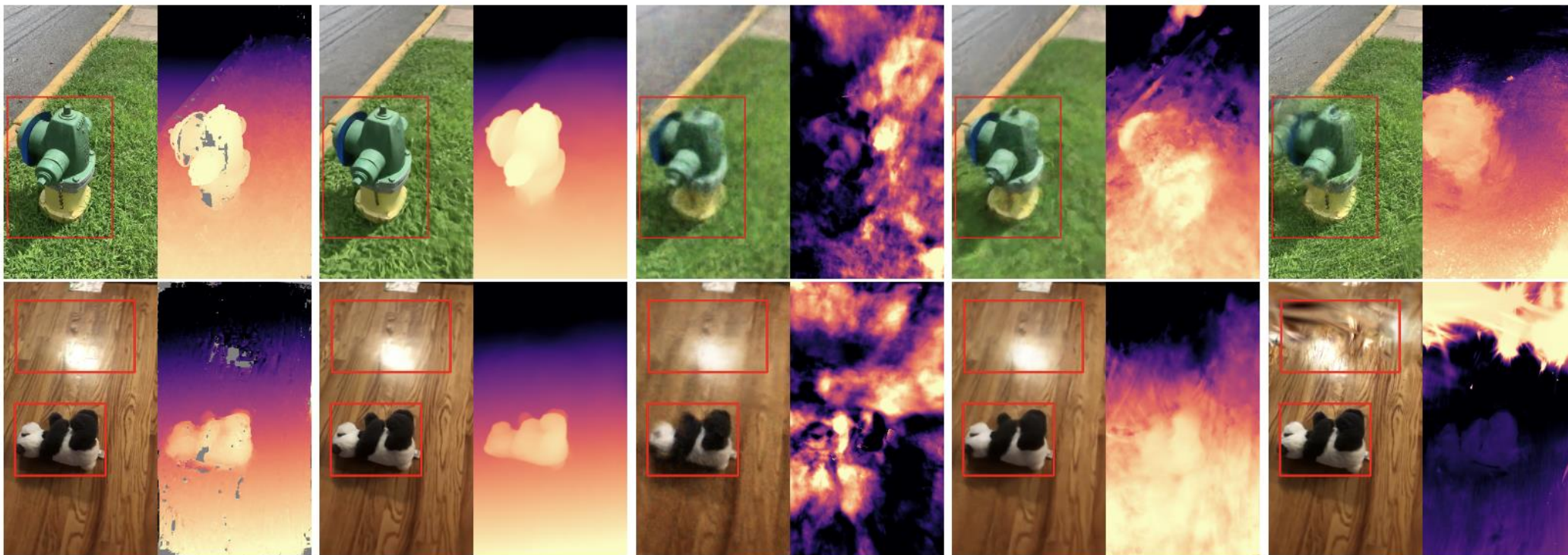
Learned poses, rendered images and rendered depth maps for the train set

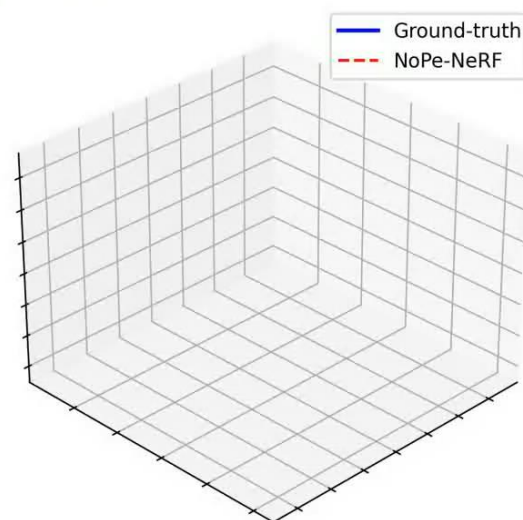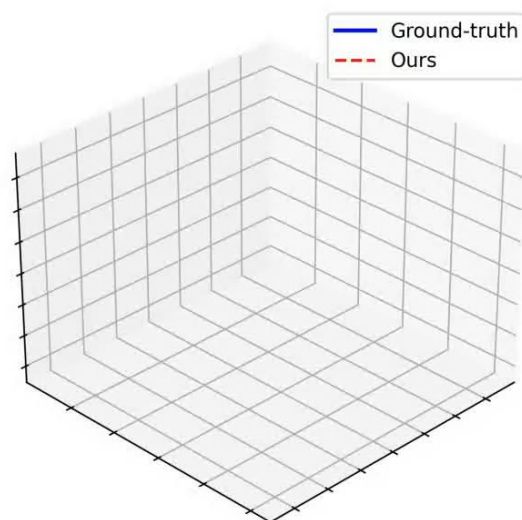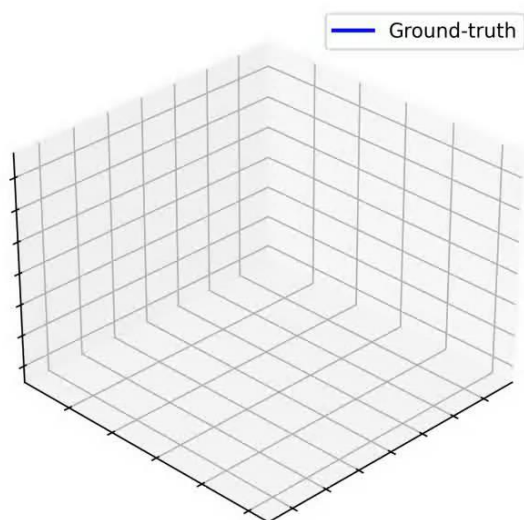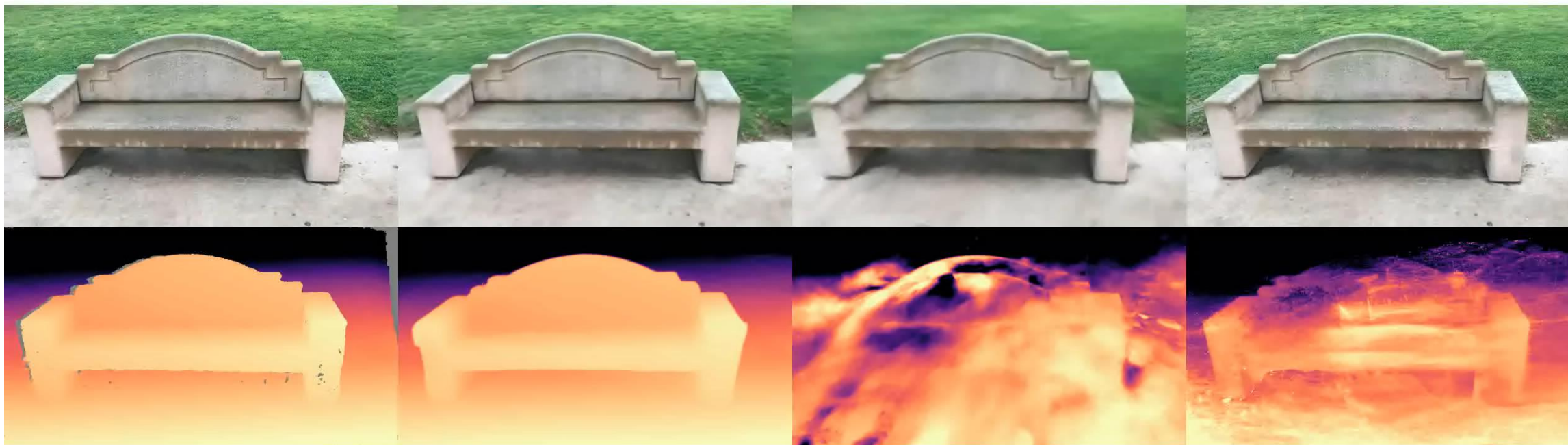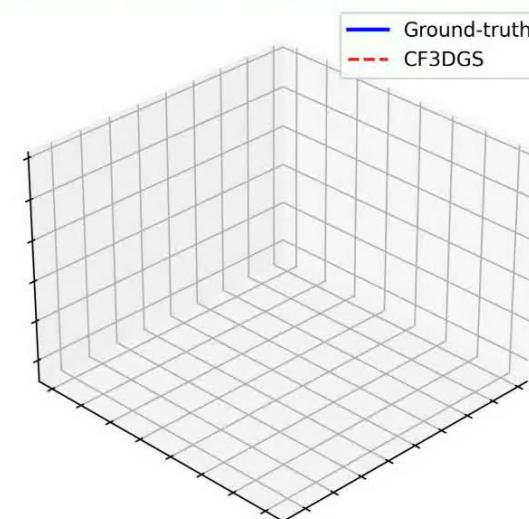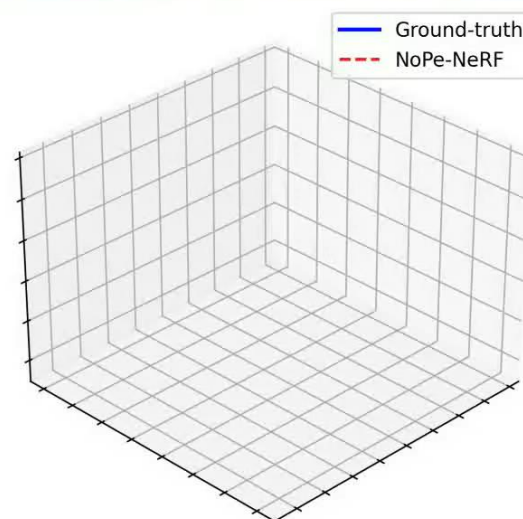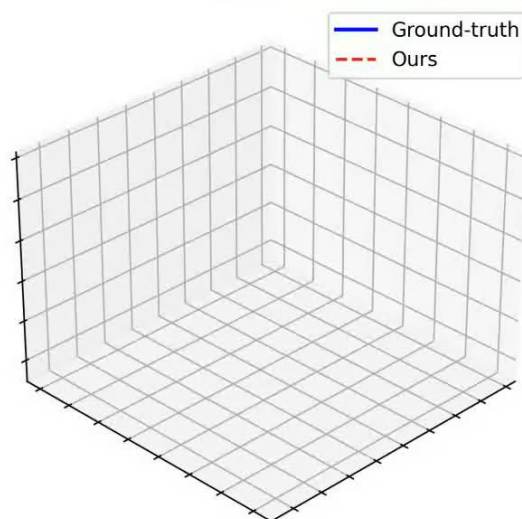| Ground-truth | Ours | NoPe−NeRF | CF3DGS |
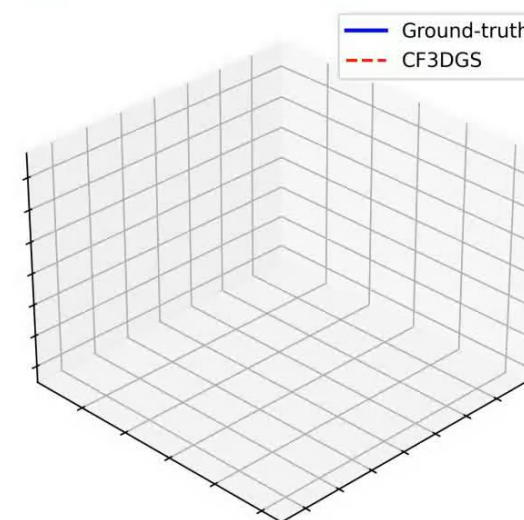
Learned poses, rendered images and rendered depth maps for the train set

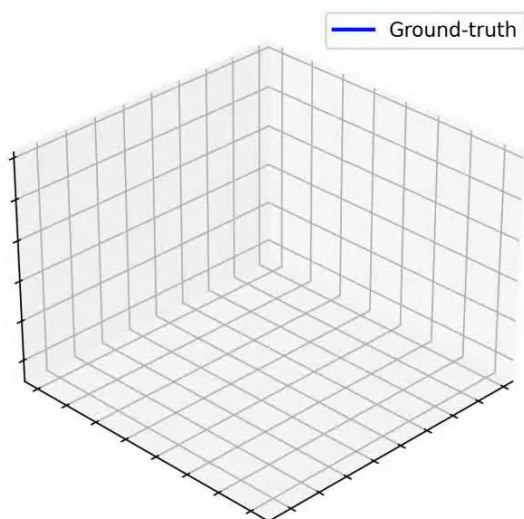Ground-truth   Ours   NoPe-NeRF   CF3DGS

Thanks for watching

Let's have a chat if you find this work interesting!

Morning session, Saturday June 14

Poster #79

Github page