Differentiable Point-Based Radiance Fields for Efficient View Synthesis

Qiang Zhang, Szymon Rusinkiewicz, Seung-Hwan Baek, Felix Heide

Siggraph Asia 2022

Presenter: Jason Yuan (jcyuan)

Motivation

Novel view synthesis but…

Fast inference

Fast training

Low memory

This would enable…

Novel view synthesis for video

Related Work

NeRF: MLPs + volumetric rendering

- Slow training and inference

PlenOxels: explicit volume instead of MLP.

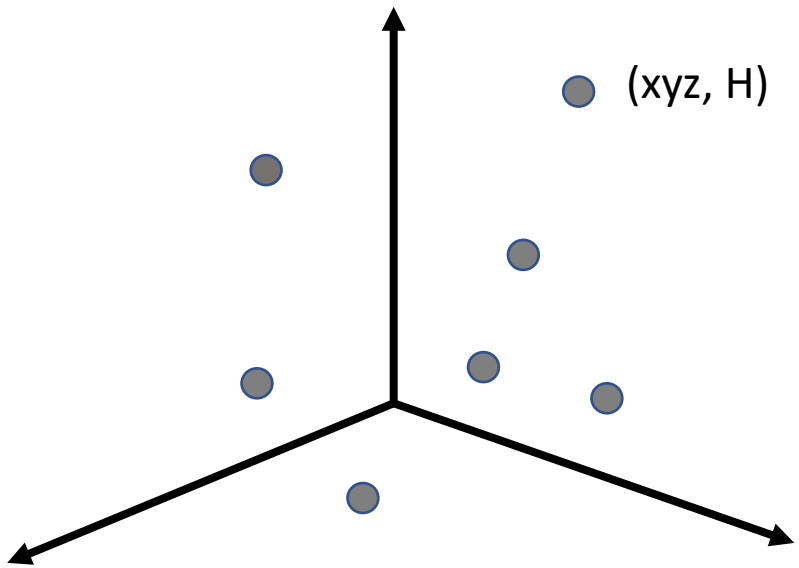
- Big memory requirement

STNeRF: for videos, conditions MLP on time t.

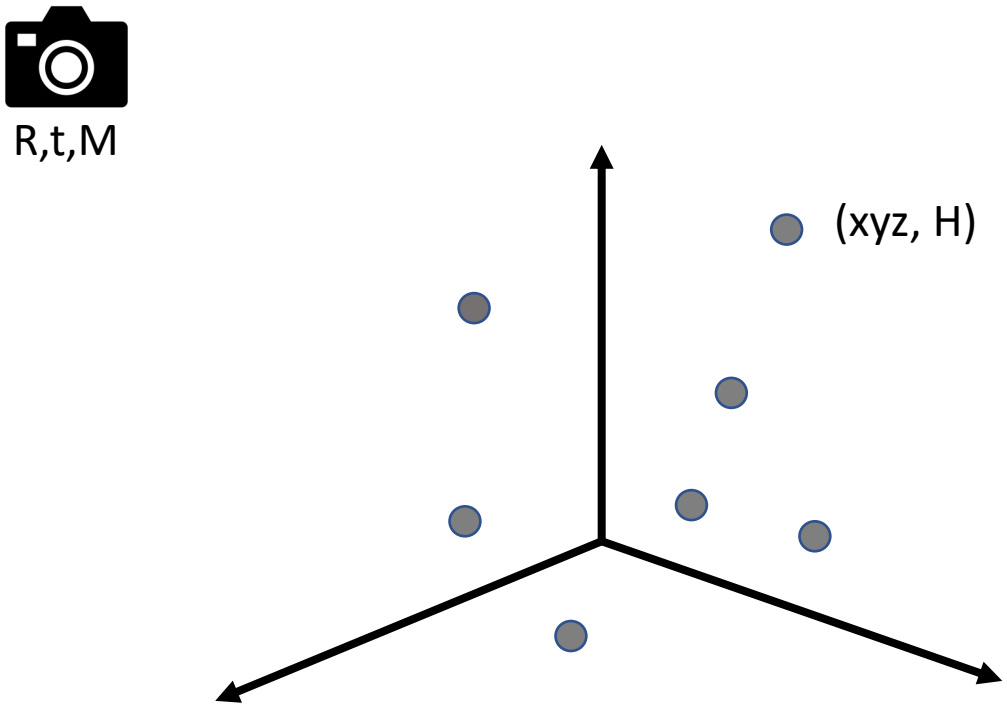
- Quality worse than frame-by-frame models

Method

Model: 3D point cloud (xyz) w Radiance parameter (H). *Learnable* parameters.



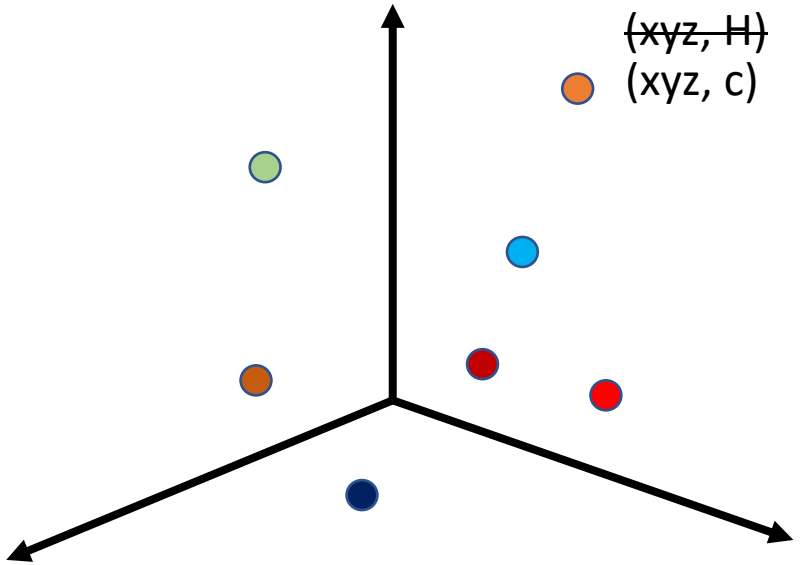
Input: viewing direction (i.e. camera intrinsics/extrinsics info, parameters R,t,M)



view direction (R,t) + radiance parameter (H) -> view-dependent color with Spherical Harmonics.

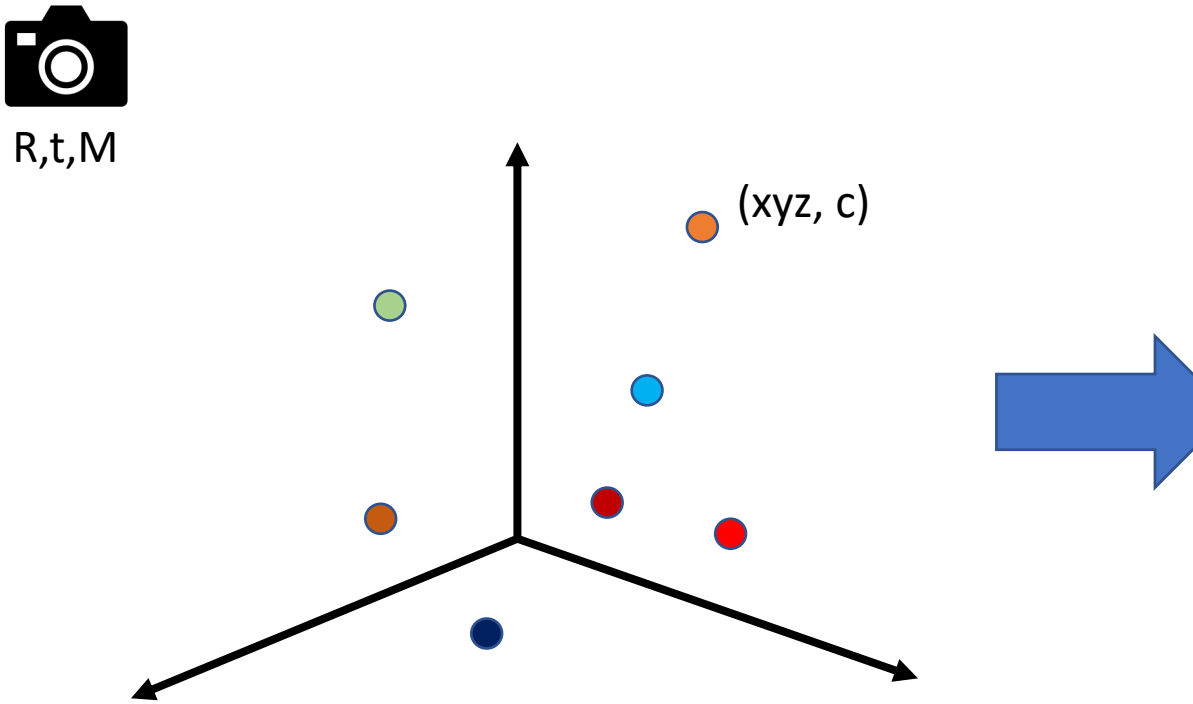
$$v_i^j = \frac{R_j P_i + t_j}{\|R_j P_i + t_j\|}, \qquad c_i^j = \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} h_{i,lm}\, Y_l^m(v_i^j)$$

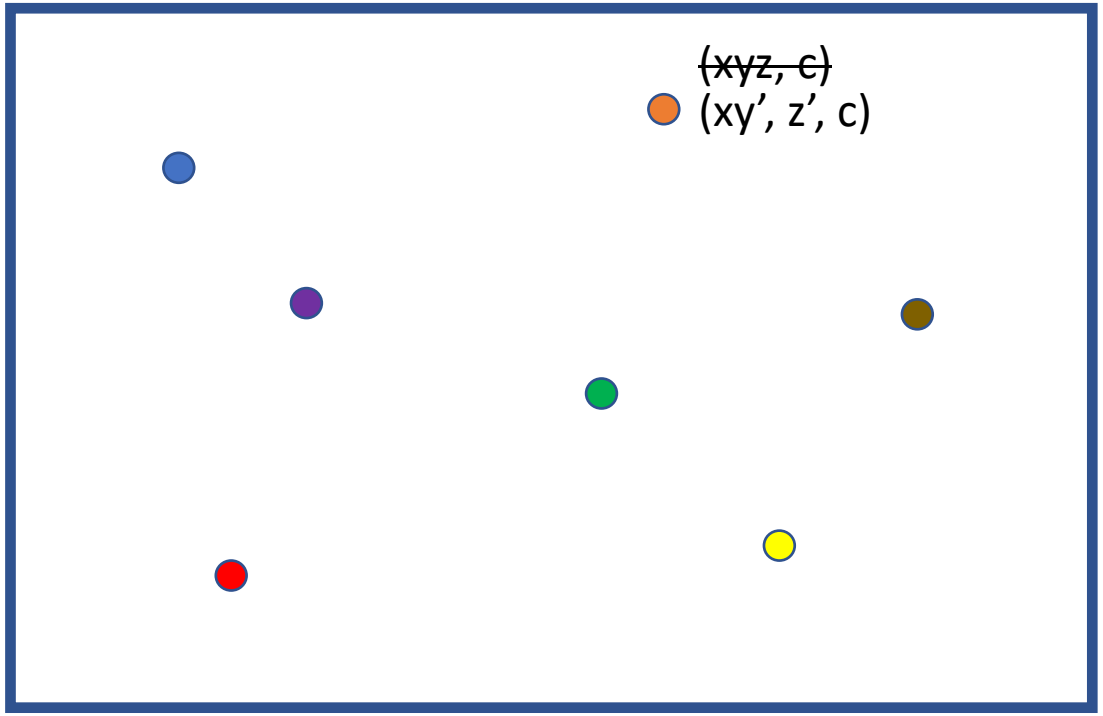
R,t,M

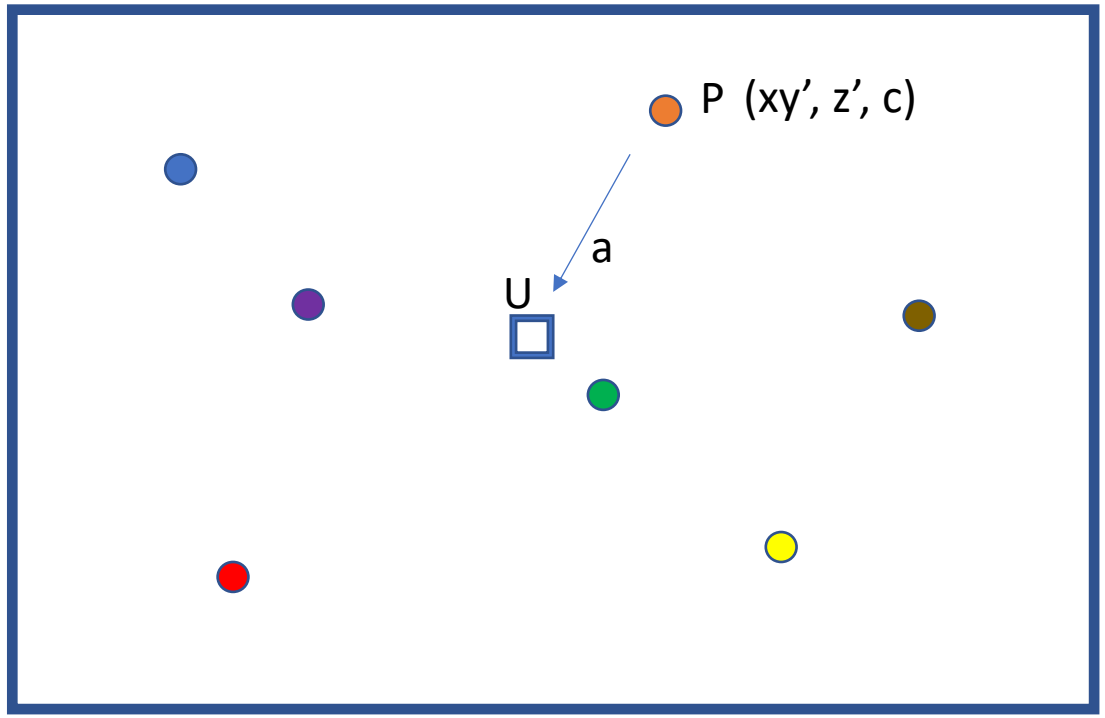


view information (R,t,M) + 3D positions (xyz) -> project into 2D and retain depth

$$p_i^j = \left(M_j(R_j P_i + t_j)\right)^{\downarrow}$$



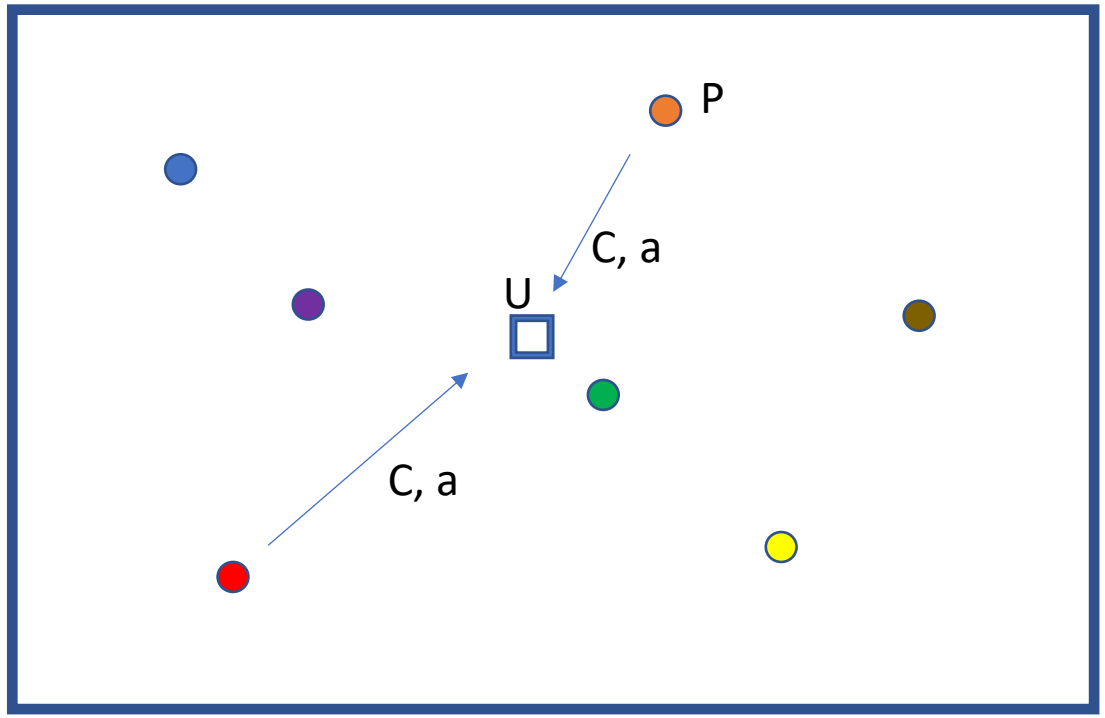




Splat Render the points:

For a given pixel (U), compute effect of each Point (P) -> Gaussian Kernel

$$\alpha_i^j(u) = \frac{1}{\sqrt{2\pi r^2}} e^{-\frac{\|p_i^j - u\|^2}{2r^2}}$$



Splat Render the points:

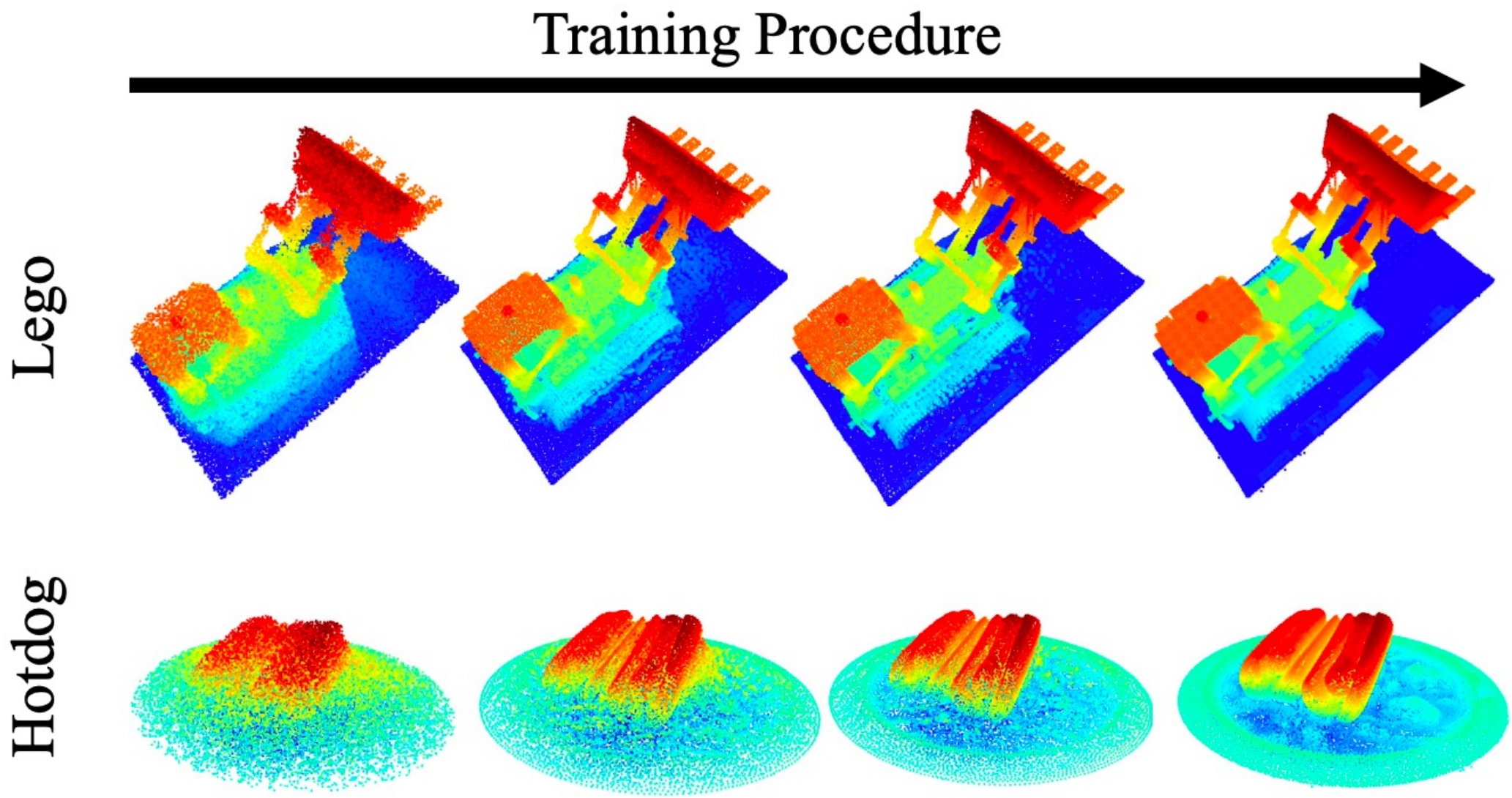
Do regular alpha blending. After sorting by depth Z.

where $A_i^j(u)$ represents the net contribution of the i-th point:

$$A_i^j(u) = \alpha_i^j(u) \prod_{k=1}^{i-1} \left(1 - \alpha_k^j(u)\right), \tag{7}$$

Training: SGD

$$\mathcal{L} = \sum_{j=1}^{N} ||I_j - \hat{I}_j||_2^2 + \lambda TV(\hat{I}_j). \quad (8)$$


Training Procedure
Lego
Hotdog

In addition to SGD…

Aggregate	Filter	Add
Aggregate clusters of points into a single point by averaging the parameters.	Filter outliers, points too far from other points.	Add new points

- Summary 
  - Model scene as a 3D point cloud (xyz) with view dependent radiance (H, spherical harmonics).
  - Learn these parameters. Optimize with SGD.
  - Use “splat rendering” to render in a single pass, unlike NERF which takes many passes. More efficient.

Results

Blender Dataset

Ground Truth

Point-Based
3 min
32 FPS
9 MB

NeRF
20 h
0.08 FPS
14 MB

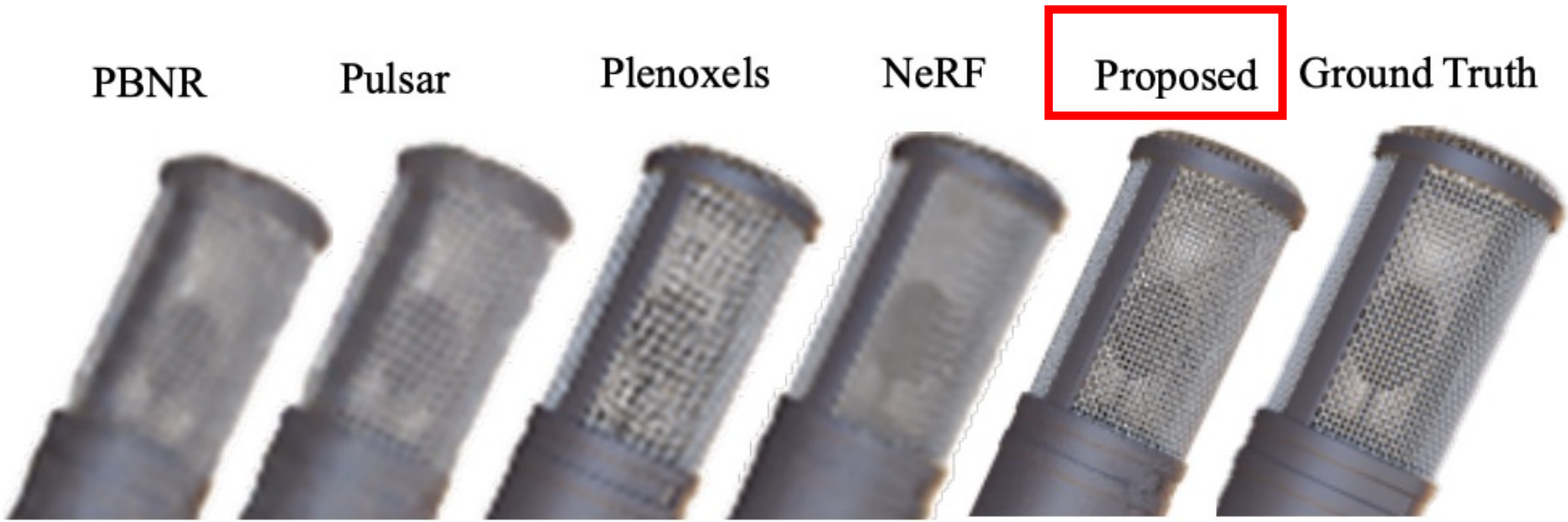
Plenoxels
11 min
15 FPS
1.1 GB

Ground Truth

Point-Based
3 min
32 FPS
9 MB

NeRF
20 h
0.08 FPS
14 MB

Plenoxels
11 min
15 FPS
1.1 GB


PBNR
Pulsar
Plenoxels
NeRF
Proposed
Ground Truth

Synthetic Dataset	Pretraining	Training	Inference	Model Size	Rendering Quality		
					PSNR↑	SSIM↑	LPIPS↓
NeRF [Mildenhall et al. 2020]	None	20 h	1/12 fps	14 MB	31.0 dB	0.947	0.081
IBRNet [Wang et al. 2021]	1 day	30 min	1/25 fps	15 MB	28.1 dB	0.942	0.072
MVSNeRF [Chen et al. 2021b]	20 h	15 min	1/14 fps	14 MB	27.0 dB	0.931	0.168
Plenoxels [Yu et al. 2021a]	None	11 min	15 fps	1.1 GB	**31.7 dB**	**0.958**	**0.050**
Plenoxels_s [Yu et al. 2021a]	None	8.5 min	18 fps	234 MB	28.5 dB	0.926	0.100
Pulsar [Lassner and Zollhofer 2021]	None	95 min	4 fps	228 MB	26.9 dB	0.923	0.184
PBNR [Kopanas et al. 2021]	None	3 h	4 fps	2.96 GB	27.4 dB	0.932	0.164
Ours	**None**	**3 min**	**32 fps**	**9 MB**	30.3 dB	0.945	0.078

Video Dataset

Ground Truth

Proposed

St-NeRF

NeRF-t


Proposed
STNeRF
NeRF-t

STNeRF	Train	Render	Model	Rendering Quality		
				PSNR↑	SSIM↑	LPIPS↓
NeRF	40 h	1/25 fps	14 MB	23.7 dB	0.853	0.304
NeRF-t	100 h	1/26 fps	16 MB	28.9 dB	0.913	0.259
STNeRF	50 h	1/30 fps	**12 MB**	32.1 dB	0.918	0.224
Ours	**30 min**	**25 fps**	110 MB	**34.6 dB**	**0.927**	**0.207**

- Technical Approach, Strengths:
  - A new approach combining point-based radiance and differentiable splatting.
  - Outperforms SOTA neural rendering on memory, train time, inference time.
  - Enables rendering of videos with frame-by-frame modeling.
  - SOTA render quality on novel view synthesis for video.

- Limitations / Weaknesses
  - A mask is needed for initializing the point clouds
  - The splatting technique doesn’t work well for semi-transparent objects, like fur