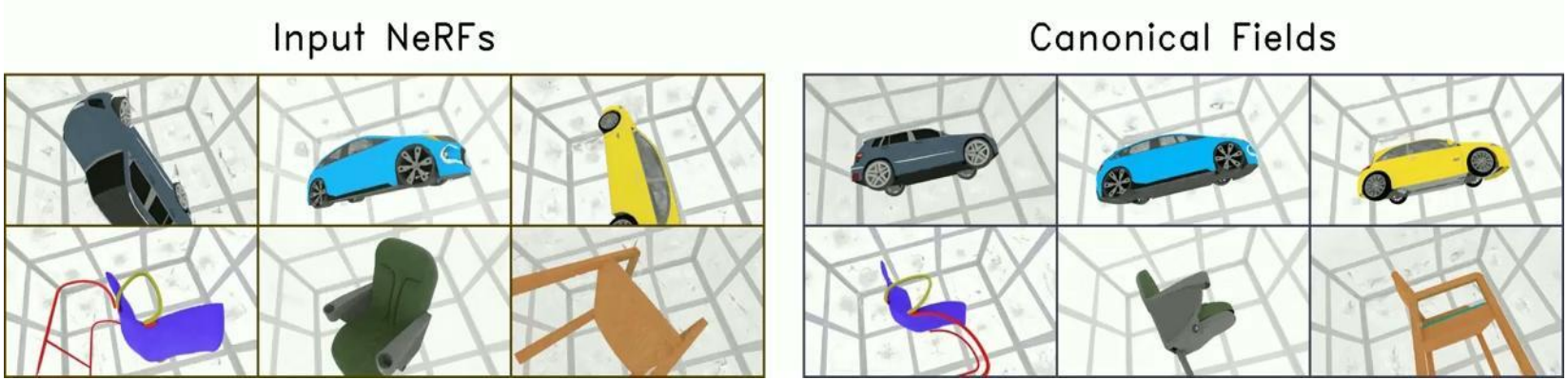
Canonical Fields: Self-Supervised Learning of Pose-Canonicalized Neural Fields

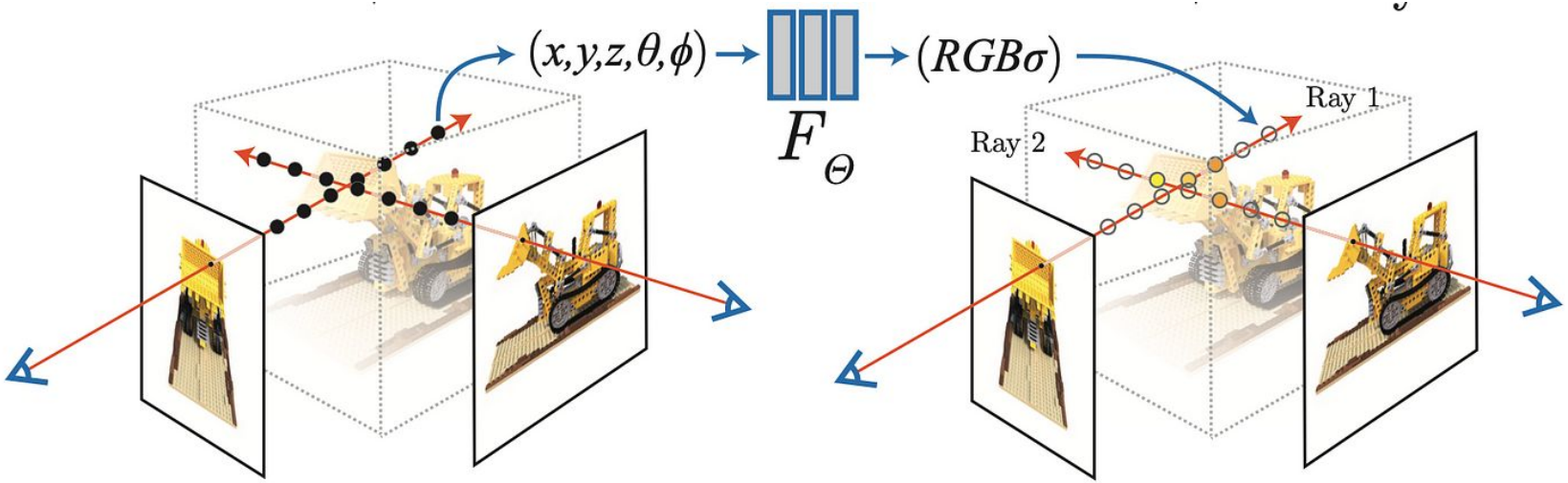
CVPR 2023 (Highlight)
TUE-PM-035



Rohith Agaram[1], Shaurya Dewan[1], Rahul Sajnani[2], Adrien Poulenard[3], Madhava Krishna[1], Srinath Sridhar[2]

1 INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD
2 BROWN
3 Stanford University

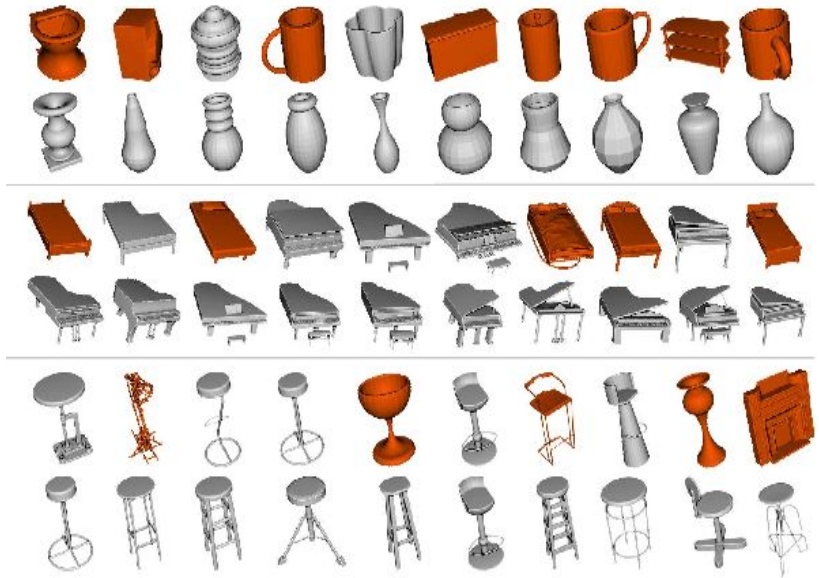
Neural Fields



Nerf (ECCV 2020)

Previous Methods

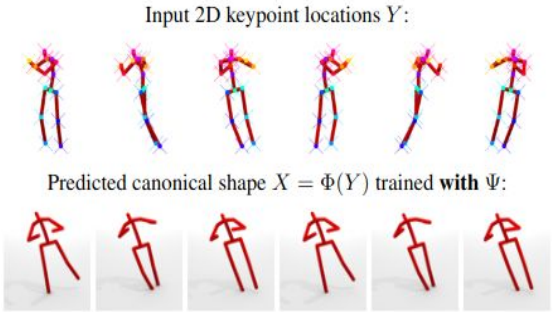
ShapeNet



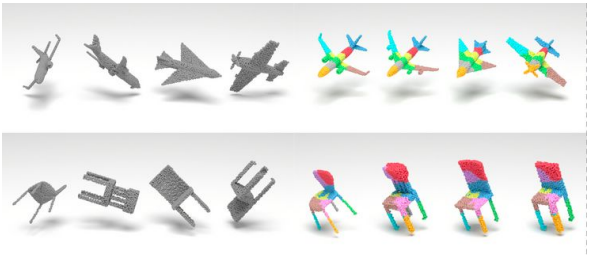
ModelNet40

Previous Methods

Images
C3dpo (ICCV 2019)

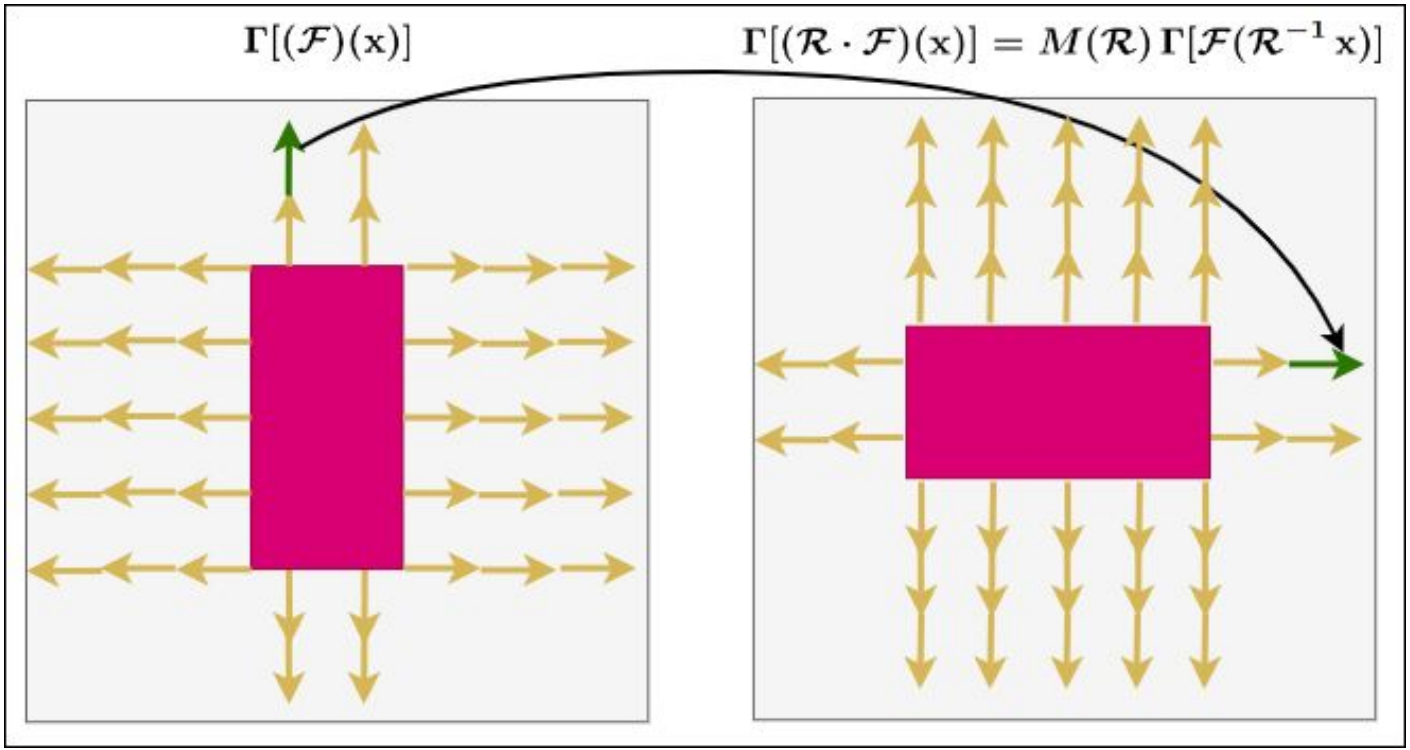


Keypoints
Canonical 3D Deformer Maps (NeurIPS 2020)

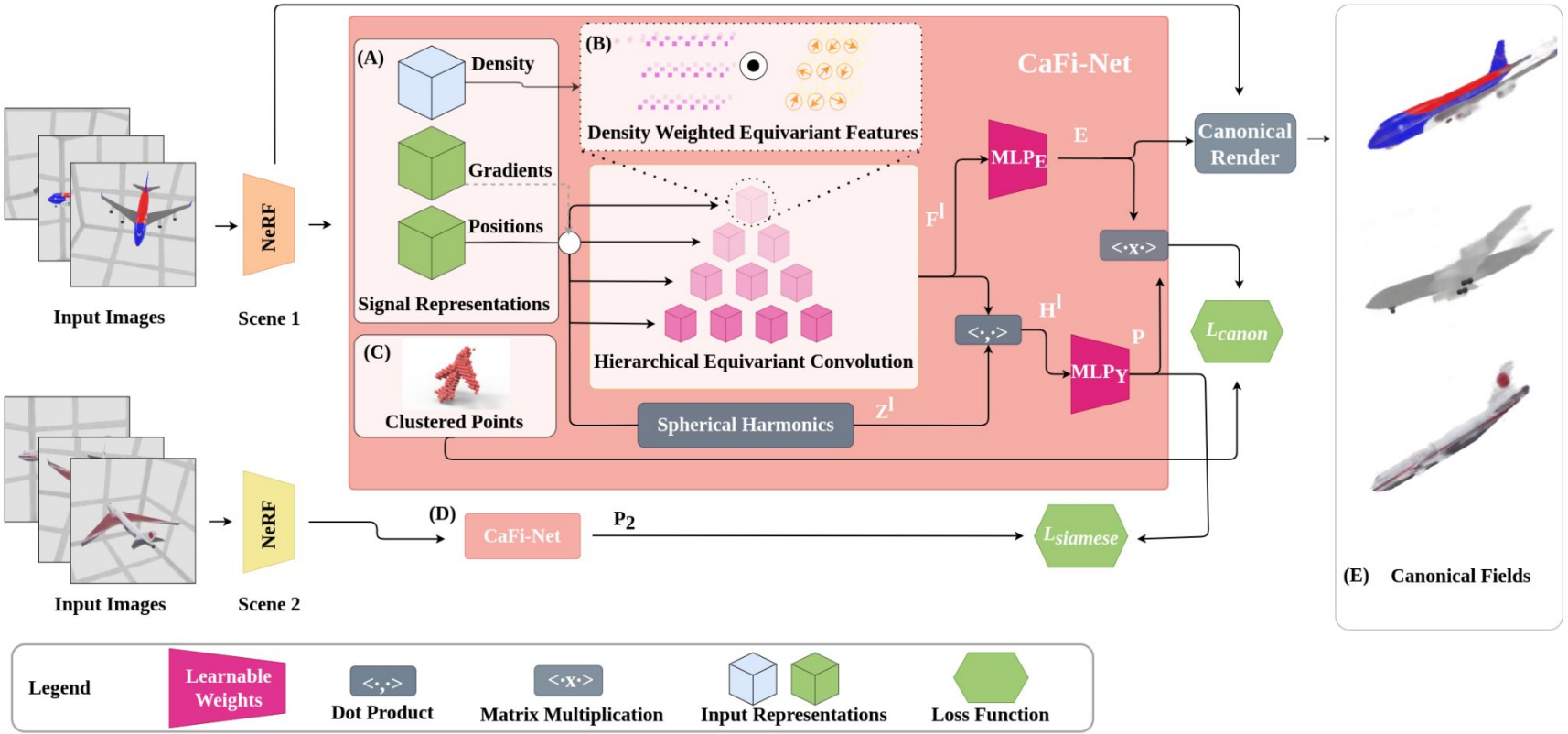


Pointclouds
Condor (CVPR 2022)

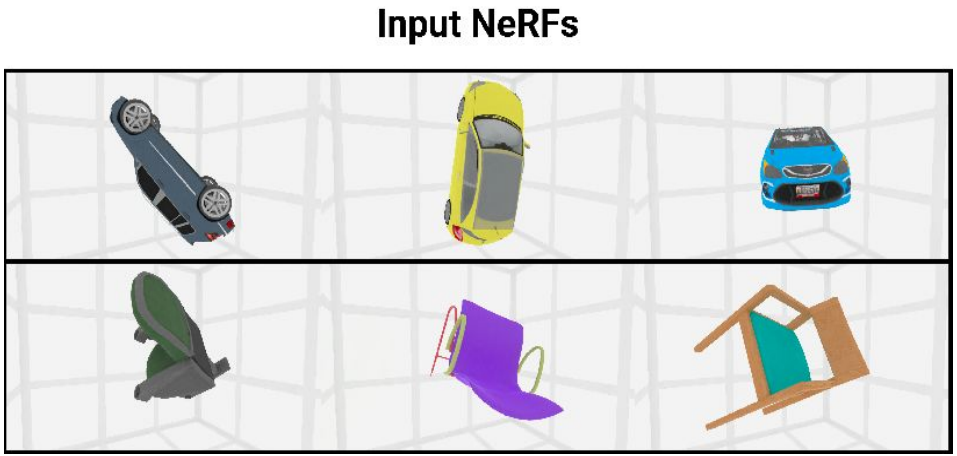
Equivariance in Vector Fields



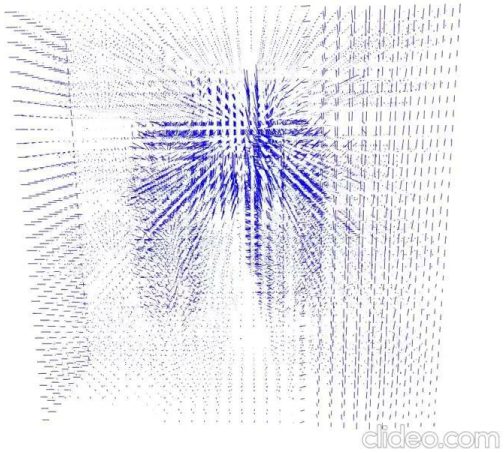
Our Method


Input Images
NeRF
Scene 1
(A)
Density
Gradients
Positions
Signal Representations
(B)
Density Weighted Equivariant Features
CaFi-Net
Hierarchical Equivariant Convolution
(C)
Clustered Points
Spherical Harmonics
F^l
MLP_E
E
H^l
MLP_Y
P
Z^l
Canonical Render
L_{canon}
Input Images
NeRF
Scene 2
(D)
CaFi-Net
P_2
$L_{siamese}$
(E) Canonical Fields
Legend
Learnable Weights
Dot Product
Matrix Multiplication
Input Representations
Loss Function

Dataset

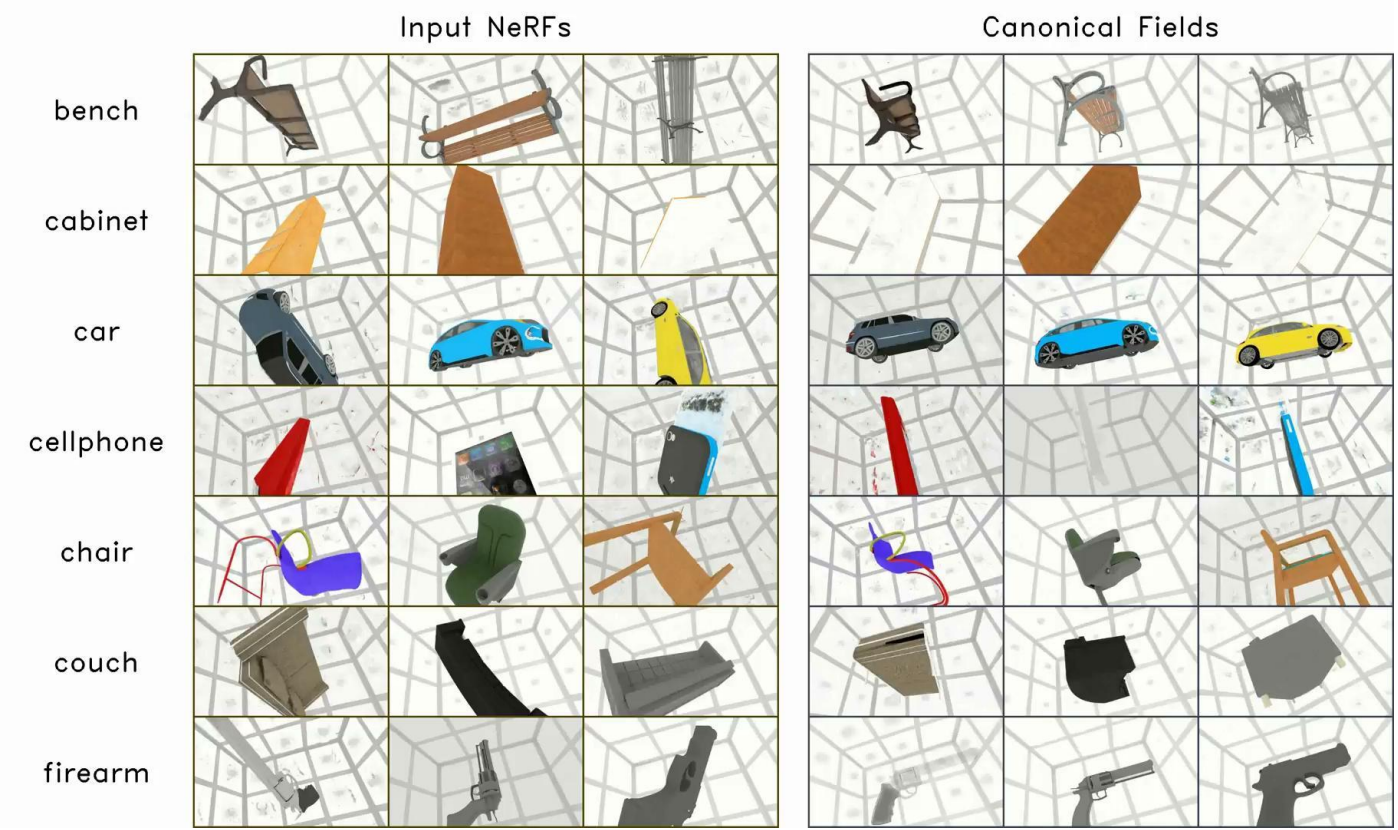


NeRF Renderings

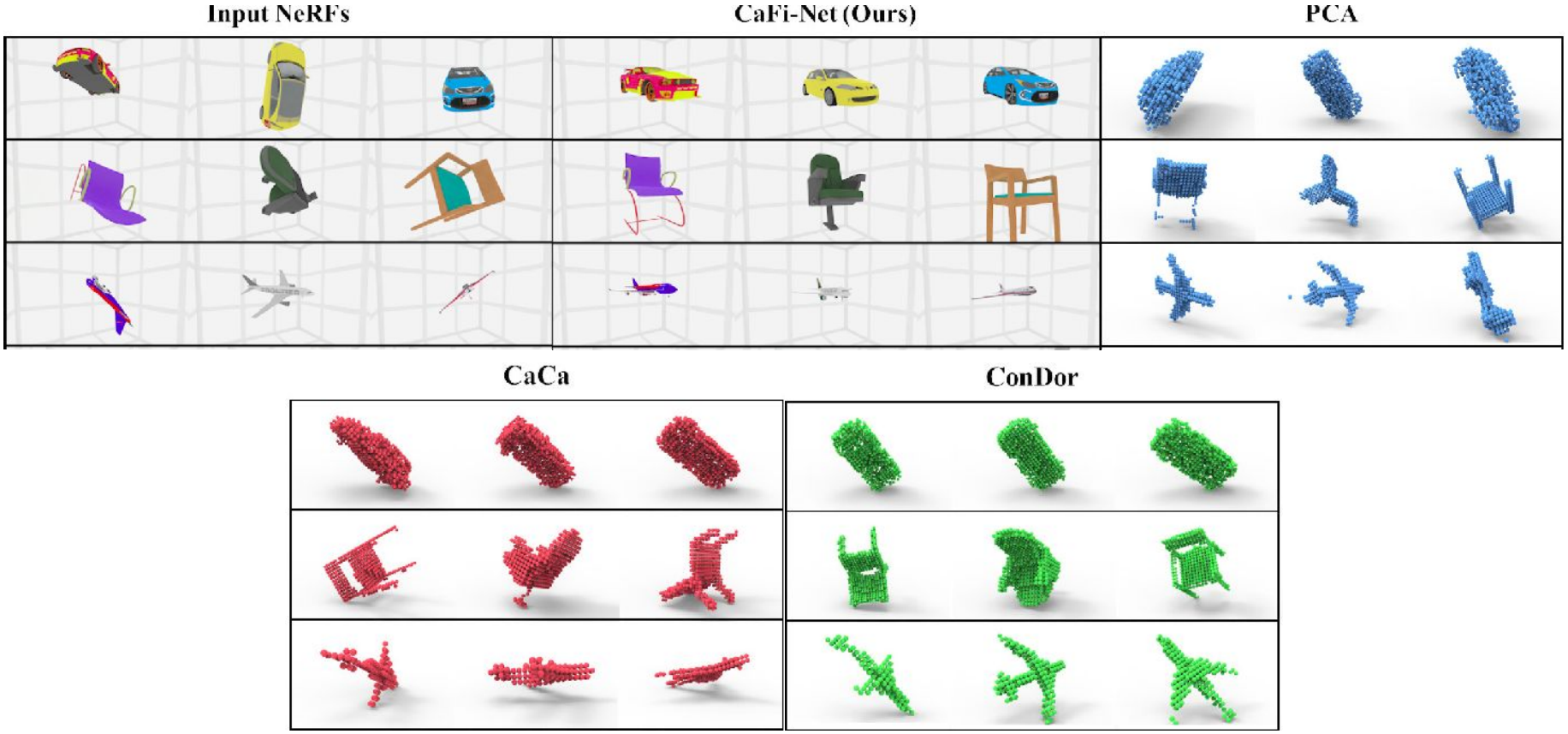


Gradient Field

Results


Input NeRFs
Canonical Fields
bench
cabinet
car
cellphone
chair
couch
firearm

Comparisons



Design Choices

Category	xyz vs. Gradient Signals	
	xyz	Gradient signal
Ground Truth Equivariance Consistency (GEC)↓		
bench	**2.17**	3.26
cellphone	1.84	**1.43**
chair	1.72	**0.94**
plane	2.07	**0.66**
Average	1.95	**1.57**
Instance-Level Consistency (IC)↓		
bench	**2.04**	2.62
cellphone	1.6	**1.28**
chair	1.4	**0.81**
plane	1.88	**0.36**
Average	1.73	**1.26**
Category-Level Consistency (CC)↓		
bench	**1.99**	2.77
cellphone	1.49	**1.26**
chair	1.52	**0.75**
plane	1.9	**0.63**
Average	1.72	**1.35**

(a) **Choice of Signal Representation** - Canonicalization metrics for using Gradients vs. xyz locations as input signal. Gradients capture the object surface that help in canonicalization.

Category	Local Average Density	
	w/o	w
Ground Truth Equivariance Consistency (GEC)↓		
bench	**3.26**	3.38
cellphone	**1.43**	1.63
chair	**0.94**	1.3
plane	0.66	**0.64**
Average	**1.57**	1.73
Instance-Level Consistency (IC)↓		
bench	**2.62**	2.72
cellphone	**1.28**	1.63
chair	**0.81**	2.64
plane	**0.36**	0.39
Average	**1.26**	1.85
Category-Level Consistency (CC)↓		
bench	**2.77**	2.82
cellphone	**1.26**	1.54
chair	**0.75**	1.07
plane	0.63	**0.51**
Average	**1.35**	1.49

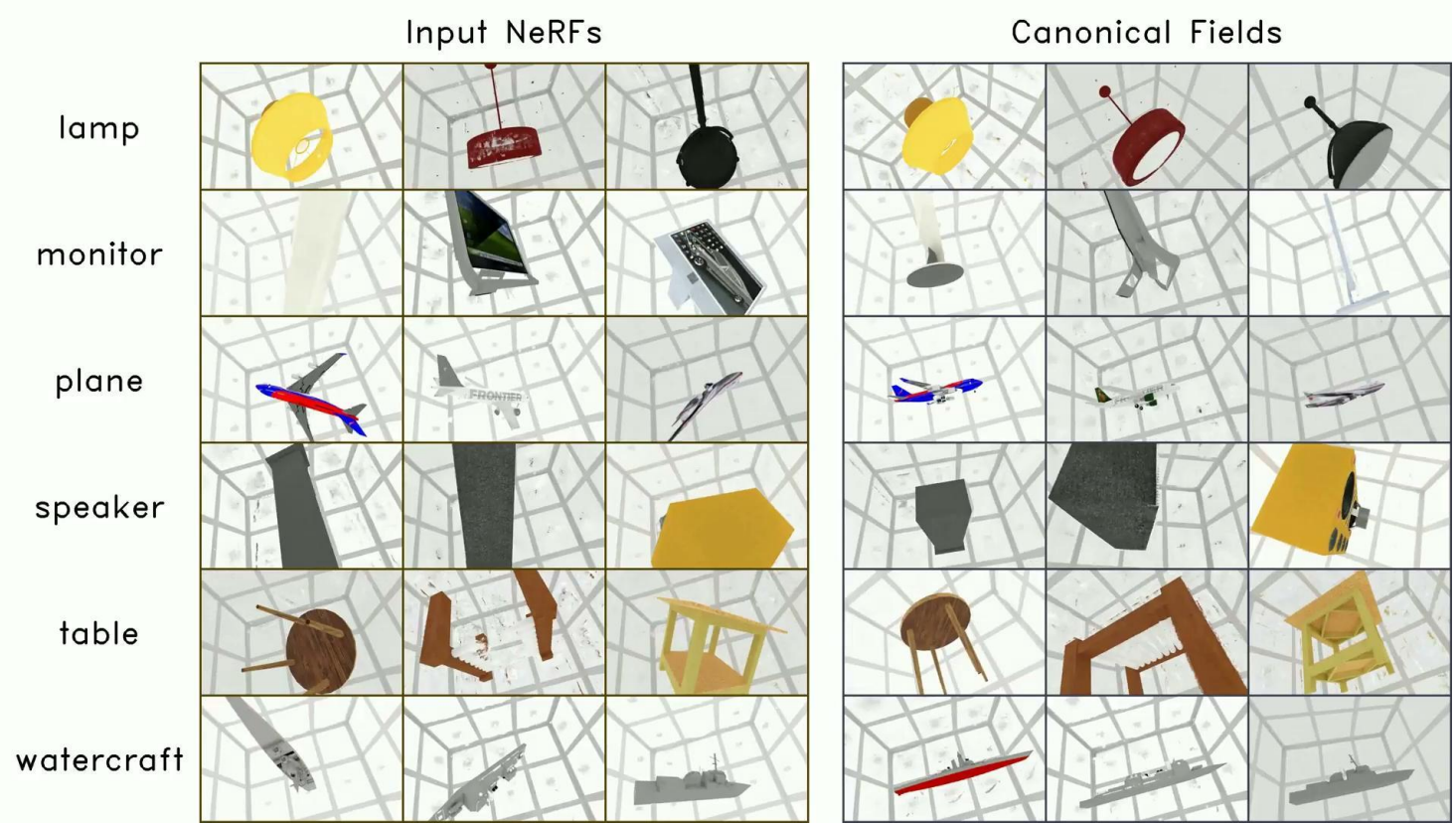
(b) **Weighing Equivariant Signals by Local Average Density** deteriorates the performance by smoothing out important details of the shape. We show canonicalization with (**w**) and without (**w/o**) weighing by the local averaged density.

Category	Siamese Training	
	w/o siamese	with siamese
Ground Truth Equivariance Consistency (GEC)↓		
bench	3.52	**3.26**
cellphone	1.61	**1.43**
chair	1.02	**0.94**
plane	1.31	**0.66**
Average	1.86	**1.57**
Instance-Level Consistency (IC)↓		
bench	2.73	**2.62**
cellphone	1.3	**1.28**
chair	0.82	**0.81**
plane	1.1	**0.36**
Average	1.48	**1.26**
Category-Level Consistency (CC)↓		
bench	2.95	**2.77**
cellphone	1.44	**1.26**
chair	0.83	**0.75**
plane	1.26	**0.63**
Average	1.62	**1.35**

(c) **Siamese Training** improves performance on all canonicalization metrics on average. We show canonicalization performance *with siamese* and without **(w/o) siamese** training. The average of Ground Truth Equivariance Consistency GEC metric reduces to 1.57 from 1.86

Limitations

Results


Input NeRFs
Canonical Fields
lamp
monitor
plane
speaker
table
watercraft

Acknowledgements

This work was supported by AFOSR grant FA9550-21-1-0214, NSF grant CNS-2038897, an AWS Cloud Credits award, NSF CloudBank, and a gift from Meta Reality Labs. We thank Chandradeep Pokhariya and Ishaan Shah.

Thank You