Common Pets in 3D: Dynamic New-View Synthesis of Real-Life Deformable Categories

Samarth Sinha, Roman Shapovalov, Jeremy Reizenstein,
Ignacio Rocco, Natalia Neverova, Andrea Vedaldi, David Novotny

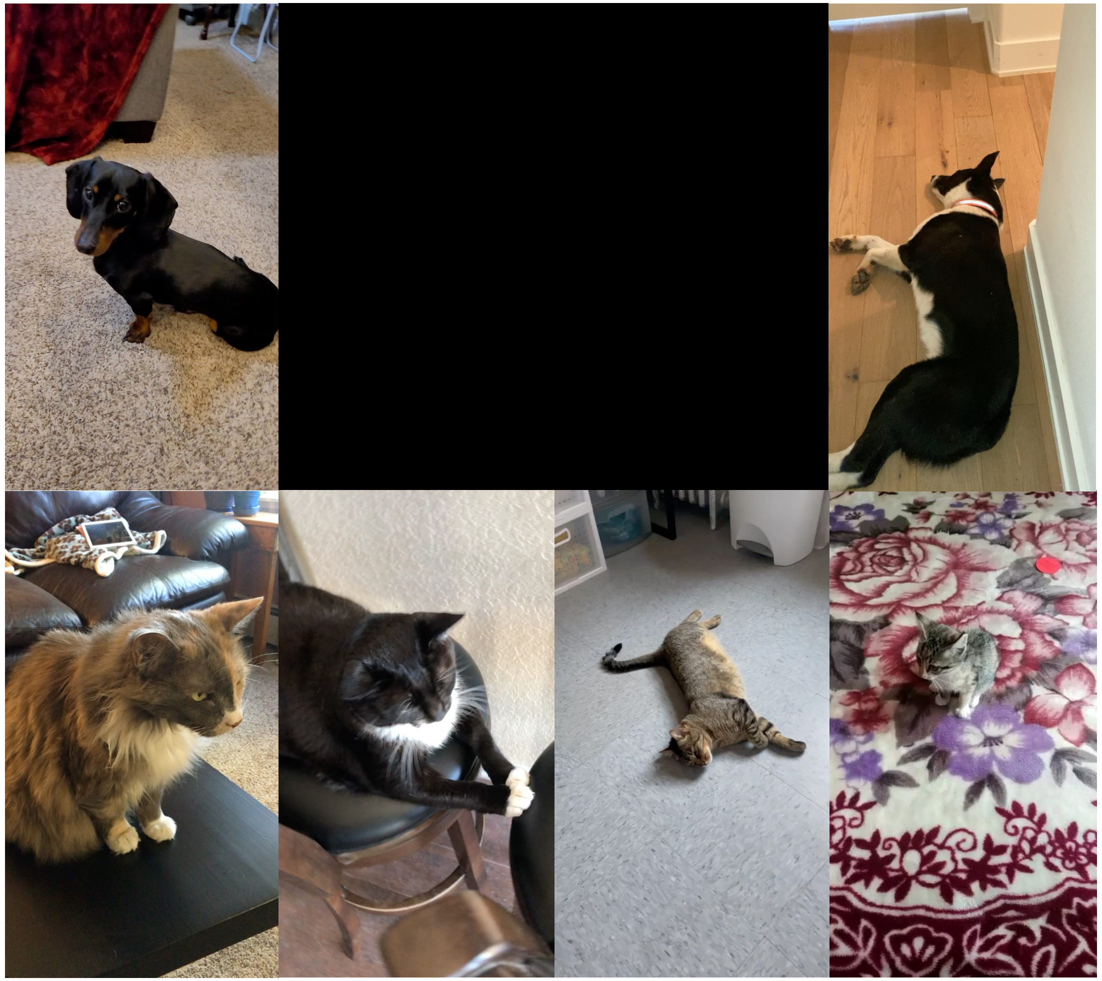
Poster Session: Tuesday-PM-3772

https://cop3d.github.io/

Common Pets in 3D Dataset: Overview

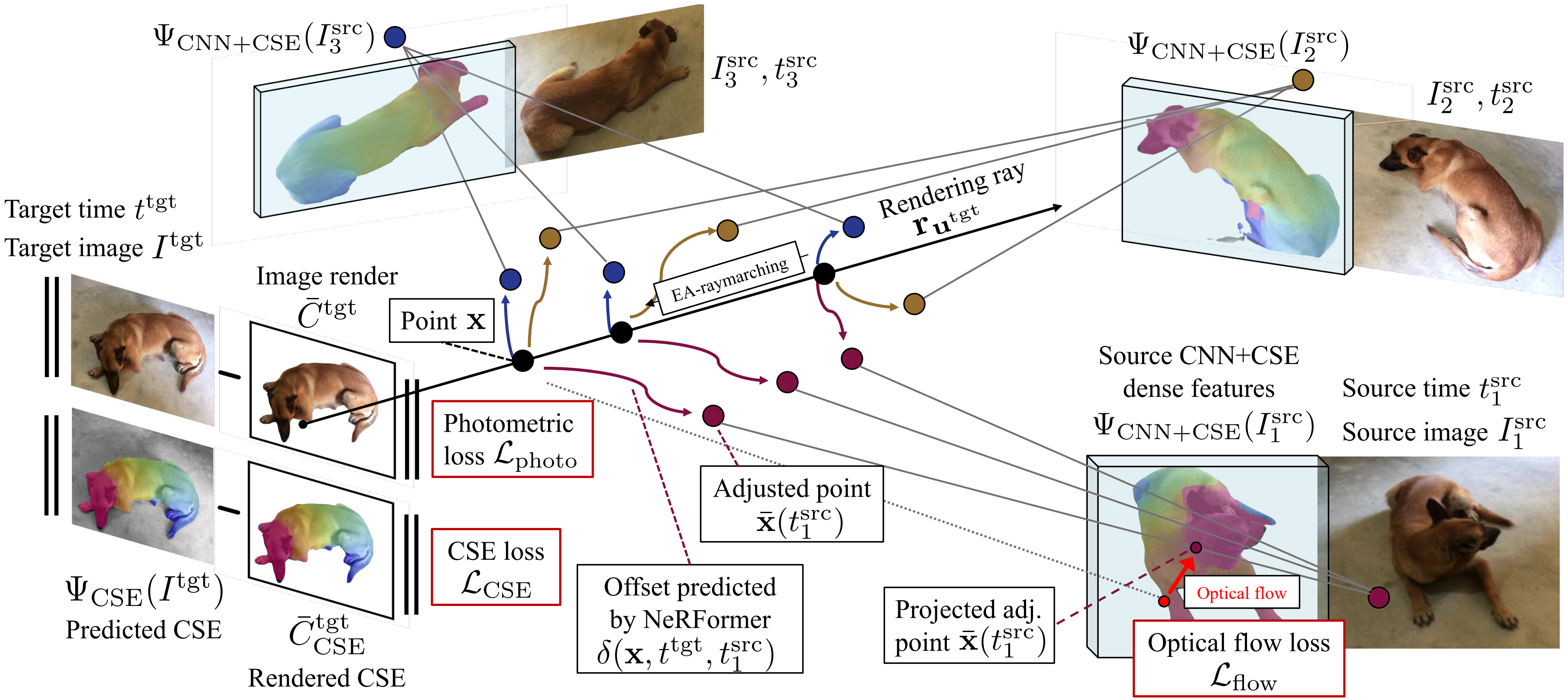
- First large-scale 3D dataset of dynamic objects
- Over 4,200 videos of cats and dogs
- Videos collected in-the-wild by crowdsourcing
- Collected with the camera panning around the pet
- Can enable category-centric 3D reconstruction

Common Pets in 3D Dataset: Collection



- First large-scale 3D dataset of dynamic objects
- Over 4,200 videos of cats and dogs
- Videos collected in-the-wild by crowdsourcing
- Collected with the camera panning around the pet
- Can enable category-centric 3D reconstruction

Tracker NeRF



- Learns category-specific priors from a large-scale dataset
- Pretrained prior provides good initialization for single-scene overfitting

Experiments: Reconstructions

CoP3D

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Common Pets in 3D: Dynamic New-View Synthesis of Real-Life Deformable Categories

Please visit the poster or read the paper for further details!

Poster Session: Tuesday-PM-3772

https://cop3d.github.io/

Meta

Common Pets in 3D: Dynamic New-View Synthesis of Real-Life Deformable Categories

Samarth Sinha, Roman Shapovalov, Jeremy Reizenstein,
Ignacio Rocco, Natalia Neverova, Andrea Vedaldi, David Novotny

Poster Session: Tuesday-PM-3772

https://cop3d.github.io/

3D Reconstruction Datasets

Dynamic NeRF datasets

- Easy to collect and annotate
- Static and Dynamic objects
- Only single scenes
- Used only for overfitting

3D Reconstruction Datasets

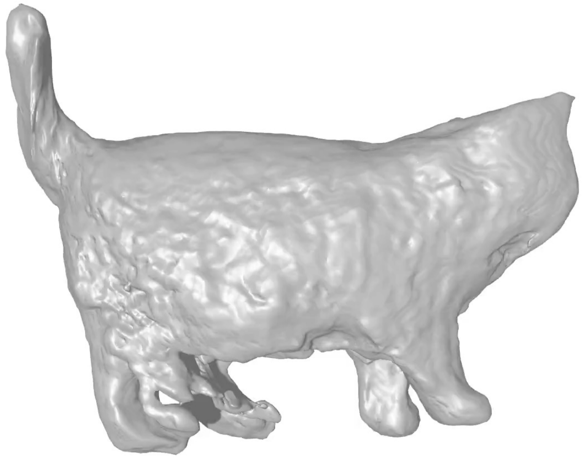
Dynamic NeRF datasets

- Easy to collect and annotate
- Static and Dynamic objects
- Only single scenes
- Used only for overfitting

Common Objects in 3D (Reizenstein et al.)

- Large scale 3D dataset
- Enables learning category priors
- Only static objects
- Challenging to collect and annotate

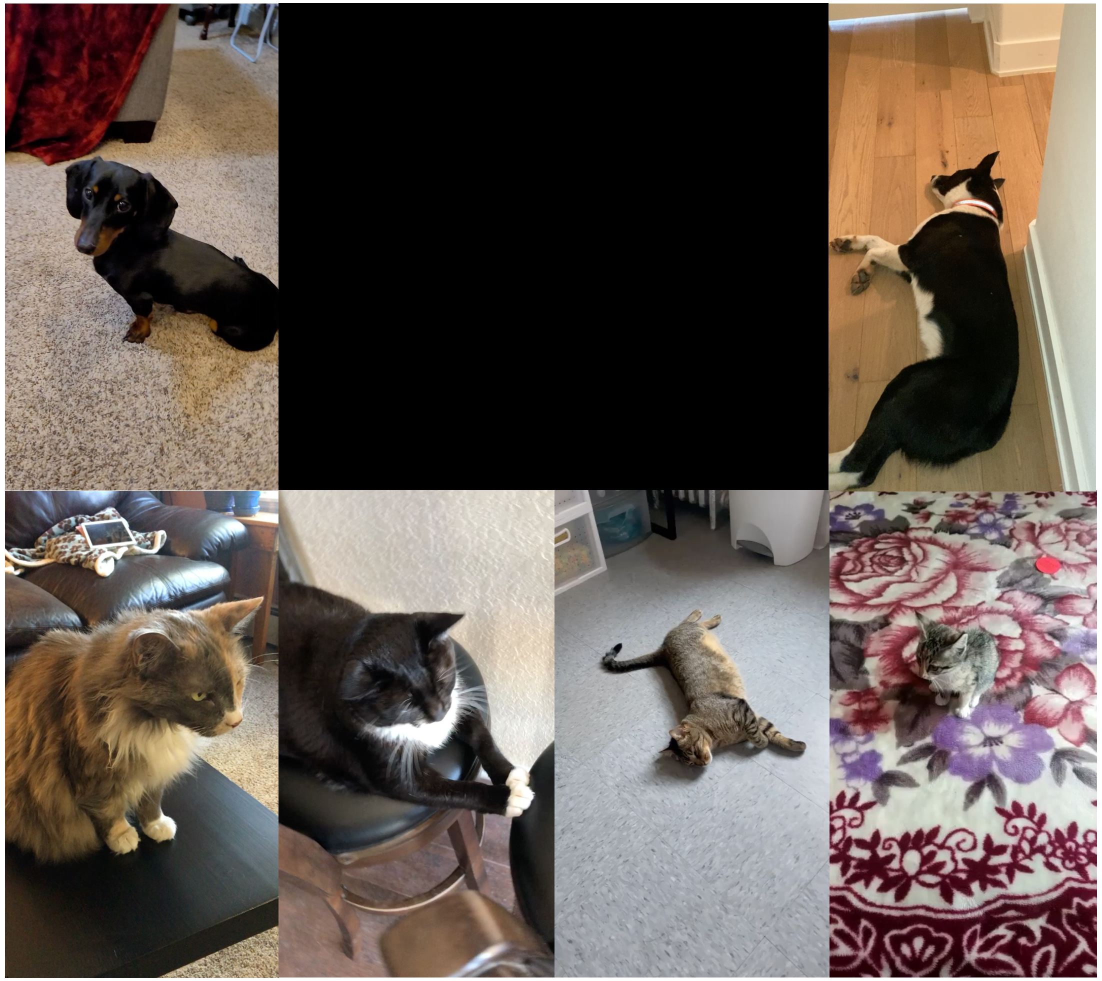
Non-Rigid 3D Reconstruction



Common Pets in 3D Dataset: Overview

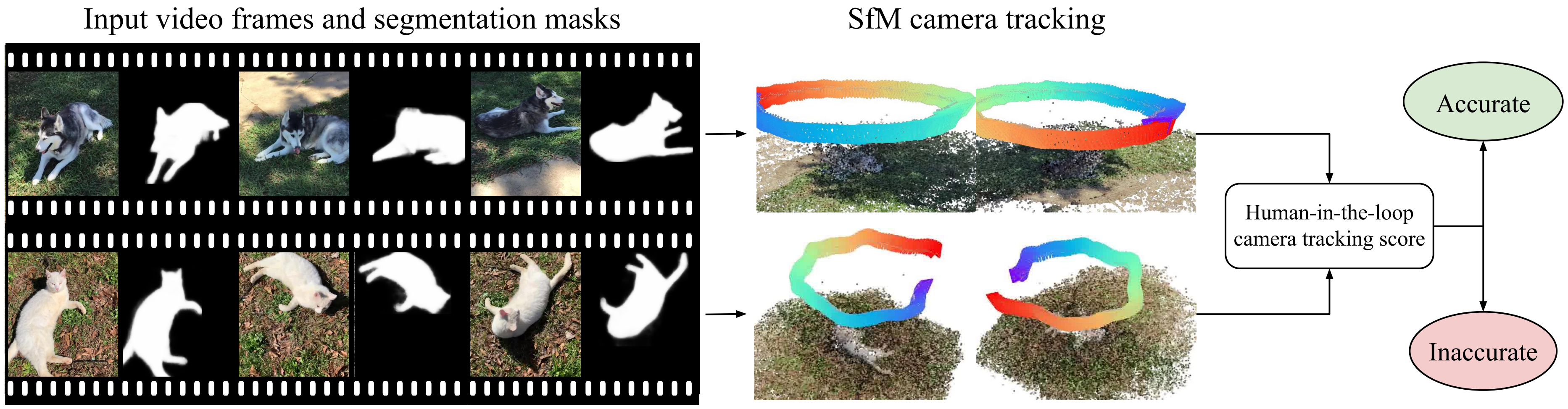
- First large-scale 3D dataset of dynamic objects
- Over 4,200 videos of cats and dogs
- Videos collected in-the-wild by crowdsourcing
- Collected with the camera panning around the pet
- Can enable category-centric 3D reconstruction

Common Pets in 3D Dataset: Collection



- First large-scale 3D dataset of dynamic objects
- Over 4,200 videos of cats and dogs
- Videos collected in-the-wild by crowdsourcing
- Collected with the camera panning around the pet
- Can enable category-centric 3D reconstruction

Common Pets in 3D Dataset: Annotation



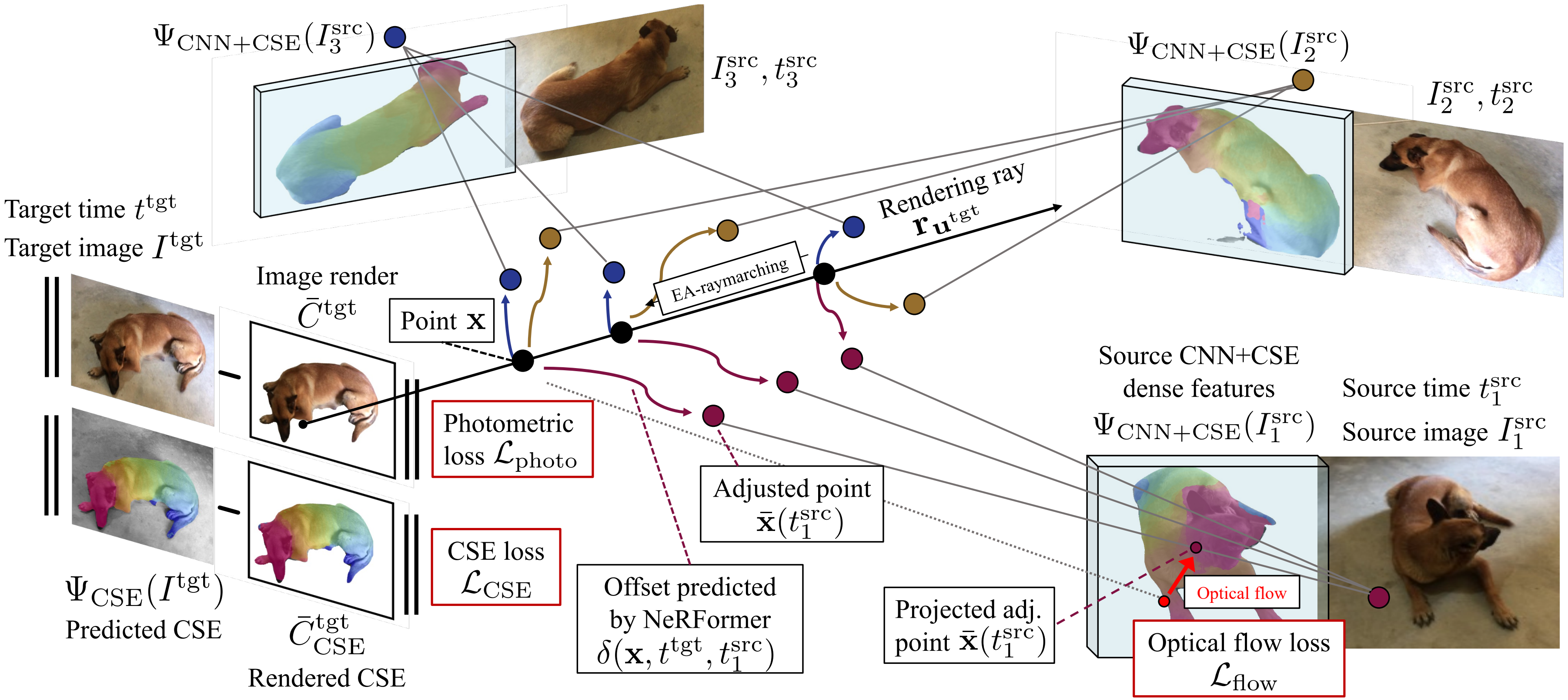
Full 3D annotation available including:

- 200 frames per scene
- 3D camera trajectory
- Camera quality scores
- Segmentation masks

Tracker NeRF: Overview

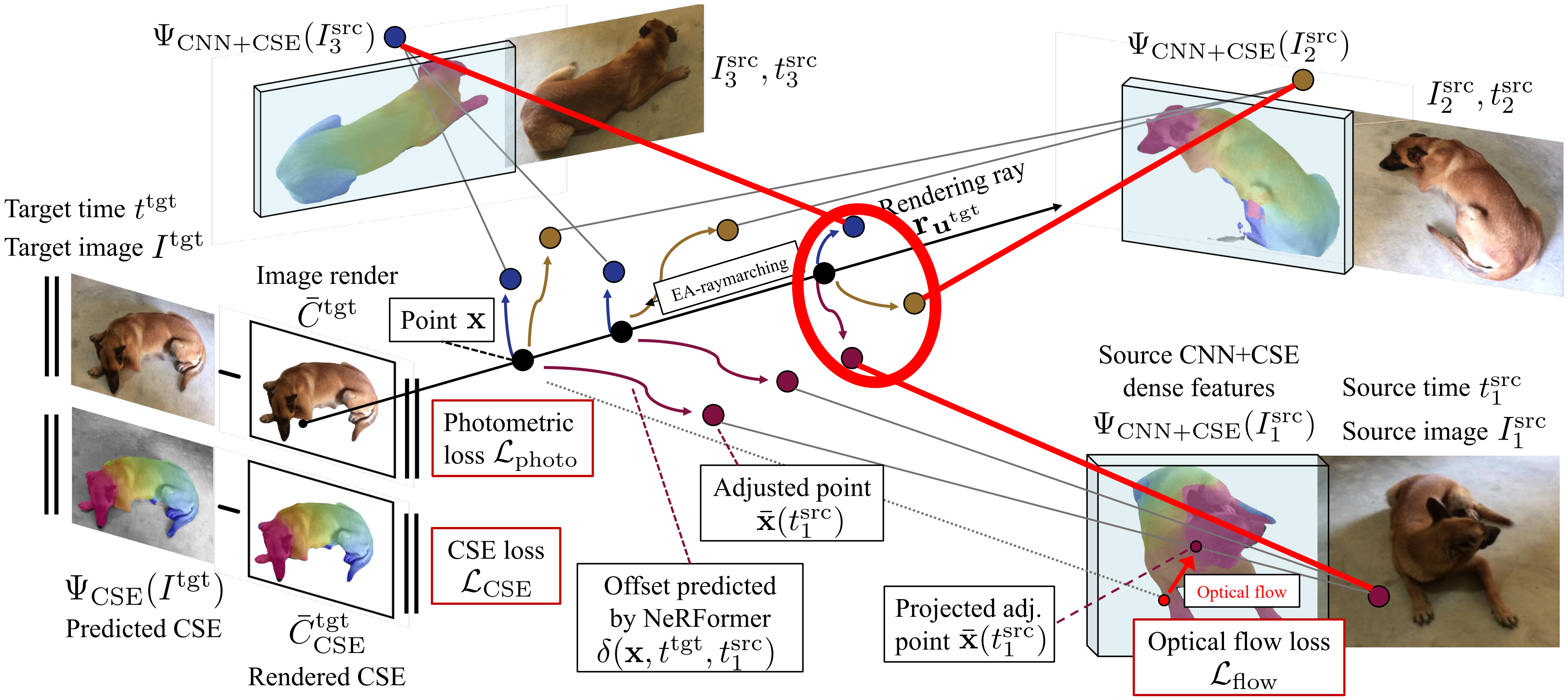
- Learns category-specific priors from a large-scale dataset
- Pretrained prior provides good initialization for single-scene overfitting

Tracker NeRF: Method



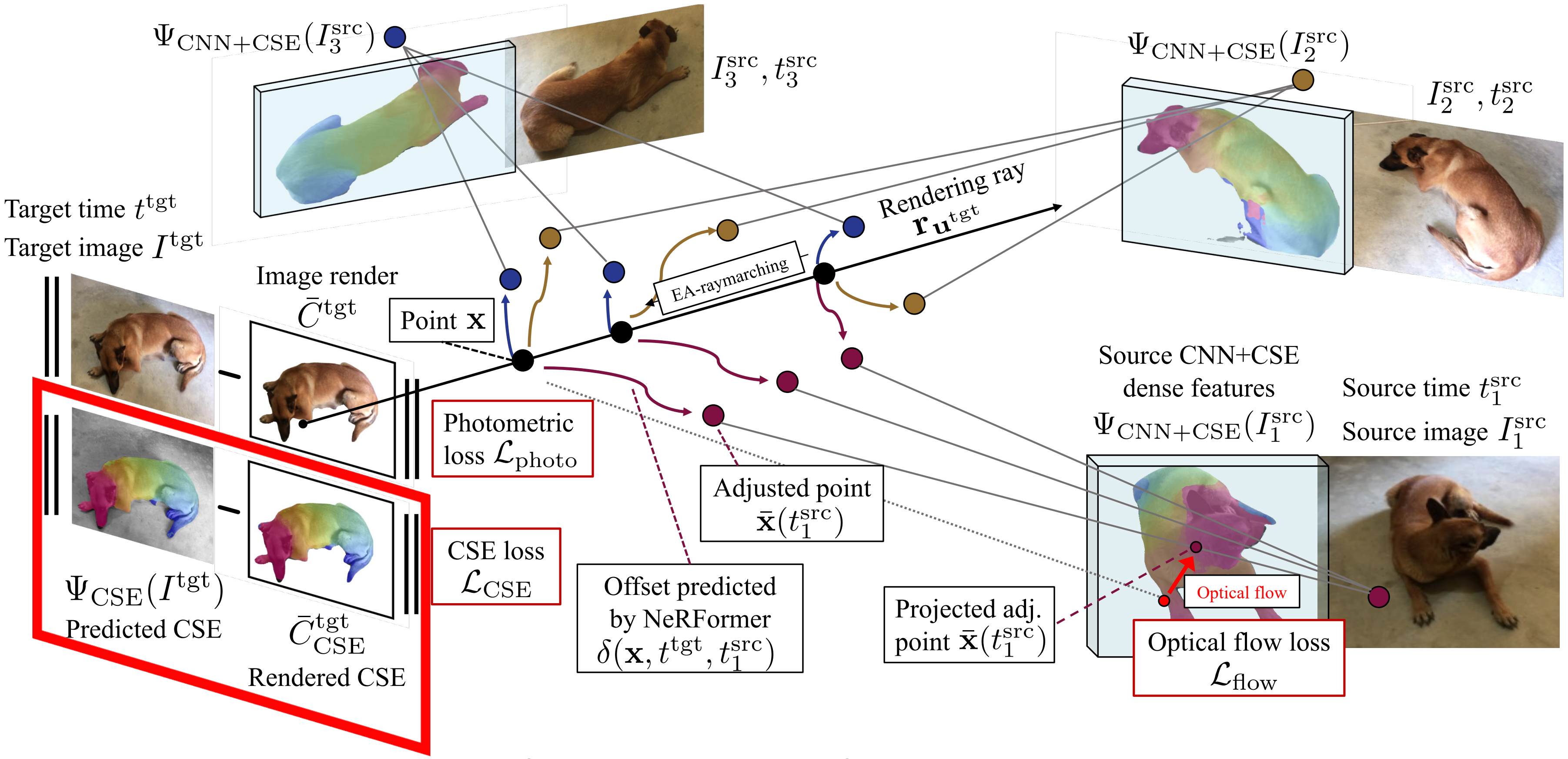
- Tracker NeRF utilizes a NeRFormer as its base architecture (Reizenstein et al., ICCV 2021)
- Tracker NeRF utilizes supervision from 2D optical flow to learn dynamic priors and Continuous Surface Embeddings (CSEs) to learn priors over the category

Tracker NeRF: Method



- For a point sampled along the target ray, Tracker NeRF predicts the 3D offsets for each source view
- That point is then projected into the source views and the features are sampled at the new location
- The prediction is supervised using 2D optical flow

Tracker NeRF: Method

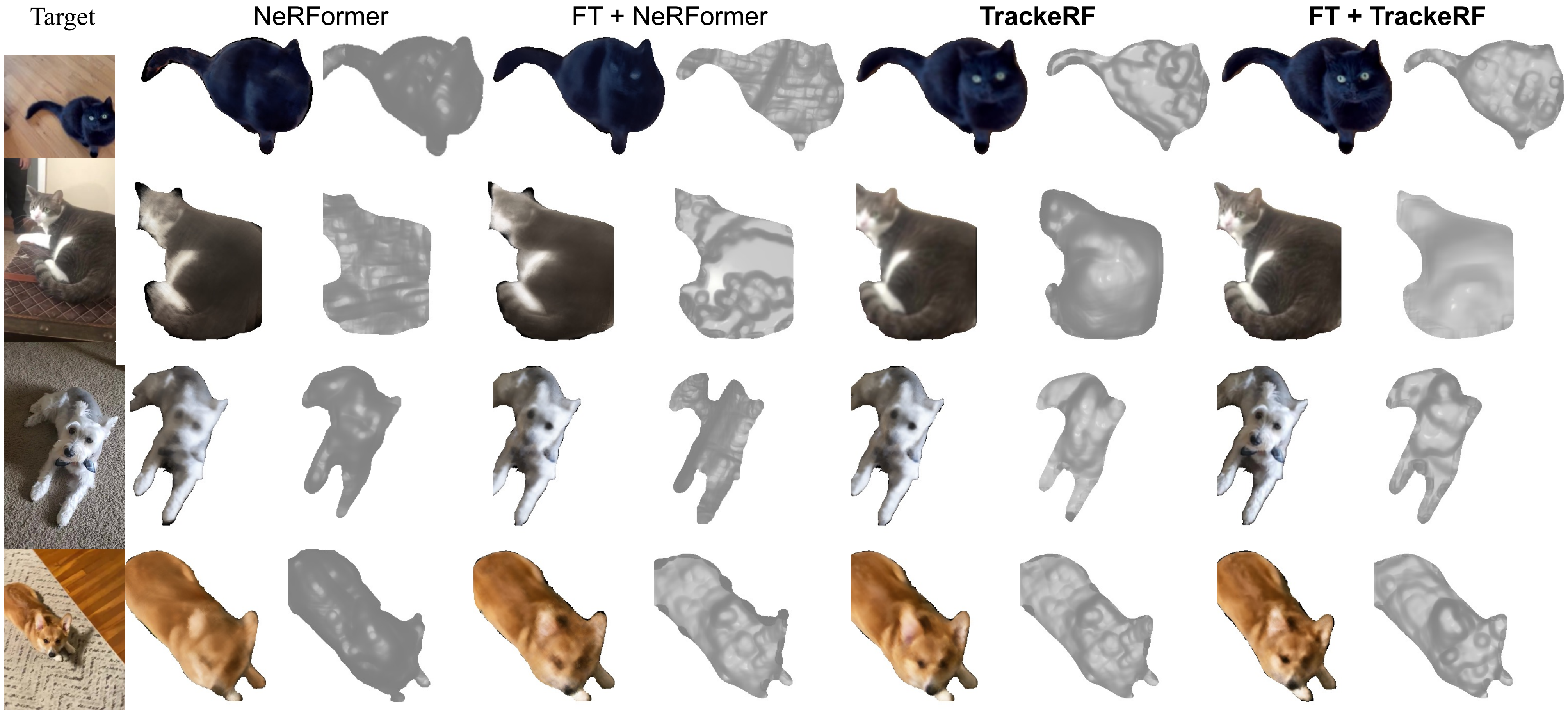


- Tracker NeRF also predicts the surface embedding for each 3D point
- The surface embedding can then be rendered and supervised by the ground-truth embedding for the target views

Experiments: Single-scene Overfitting

↓ Method	PSNR	LPIPS	IoU	ℓ_1^{RGB}
SRN+AD	17.2	0.24	0.78	0.27
SRN+TWCE	16.8	0.19	0.75	0.29
TimeNeRF	17.3	0.18	0.72	0.20
NeRF+TWCE	17.3	0.19	0.73	0.46
NeRFormer+TWCE	18.6	0.17	0.82	0.21
NSFF	20.2	0.17	—	0.19
TrackeRF (ours)	21.4	0.15	**0.91**	**0.17**
FT+NeRF+TWCE	17.7	0.19	0.82	0.30
FT+NeRFormer+TWCE	20.5	0.15	0.88	0.20
FT+TrackeRF (ours)	**23.1**	**0.13**	**0.91**	**0.17**

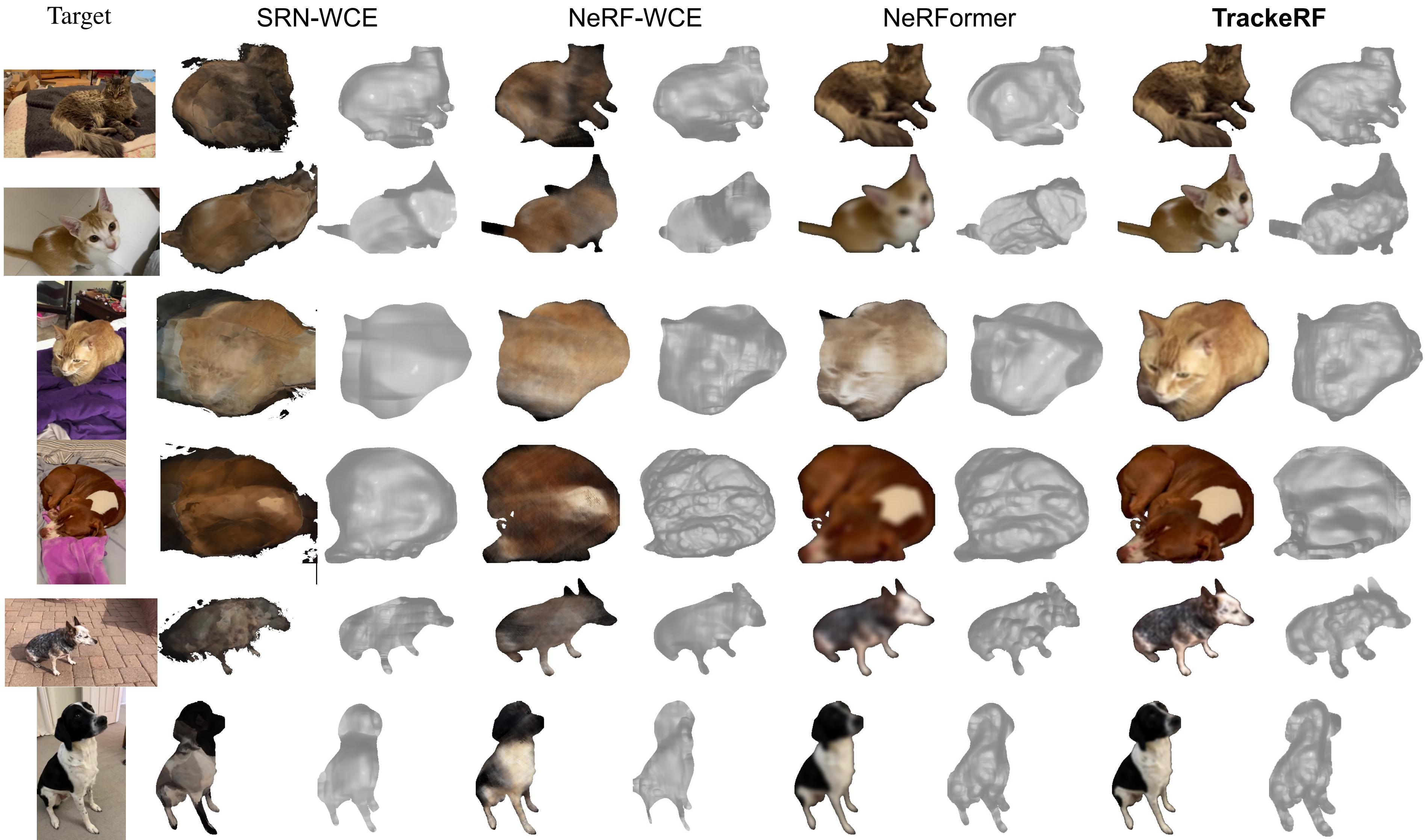
Experiments: Single-scene Overfitting



Experiments: Category Reconstruction

	(a) Average statistics								(b) PSNR @ # source views									
Frame set →	`train-unseen`				`test-unseen`				`train-unseen`					`test-unseen`				
↓ Method	PSNR	LPIPS	ℓ_1^{RGB}	IoU	PSNR	LPIPS	ℓ_1^{RGB}	IoU	25	20	15	10	5	25	20	15	10	5
TrackeRF (ours)	**19.1**	**0.16**	**0.33**	**0.80**	**19.6**	**0.17**	**0.31**	**0.82**	**19.7**	**19.6**	**19.6**	**18.8**	**17.6**	**21.0**	**20.0**	**20.0**	**18.2**	**17.9**
NeRFormer+TWCE [38]	16.3	0.19	0.39	0.73	16.6	0.18	0.36	0.76	16.5	17.2	16.6	16.3	14.9	17.7	16.9	17.0	15.6	15.7
NeRF+TWCE [10]	14.6	0.20	0.48	0.60	14.2	0.20	0.44	0.60	14.8	15.1	14.6	14.6	14.0	15.9	15.3	15.0	14.4	15.0
SRN+TWCE	13.9	0.18	0.52	0.53	14.2	0.18	0.49	0.53	13.7	14.6	14.3	13.7	13.1	15.0	14.4	14.3	13.3	14.2
SRN+AD [43]	15.5	0.19	0.40	0.66		-			15.0	16.2	15.5	15.1			-			

Experiments: Category Reconstruction



Experiments: Reconstructions

Experiments: Reconstructions

Common Pets in 3D: Dynamic New-View Synthesis of Real-Life Deformable Categories

Please visit the poster or read the paper for further details!

Poster Session: Tuesday-PM-3772

https://cop3d.github.io/