Neural Representation and Rendering of 3D Real-world Scenes

Lingjie Liu

Postdoc at Max Planck Institute for Informatics
Incoming Assistant Professor at the University of Pennsylvania

max planck institut informatik

Penn UNIVERSITY of PENNSYLVANIA

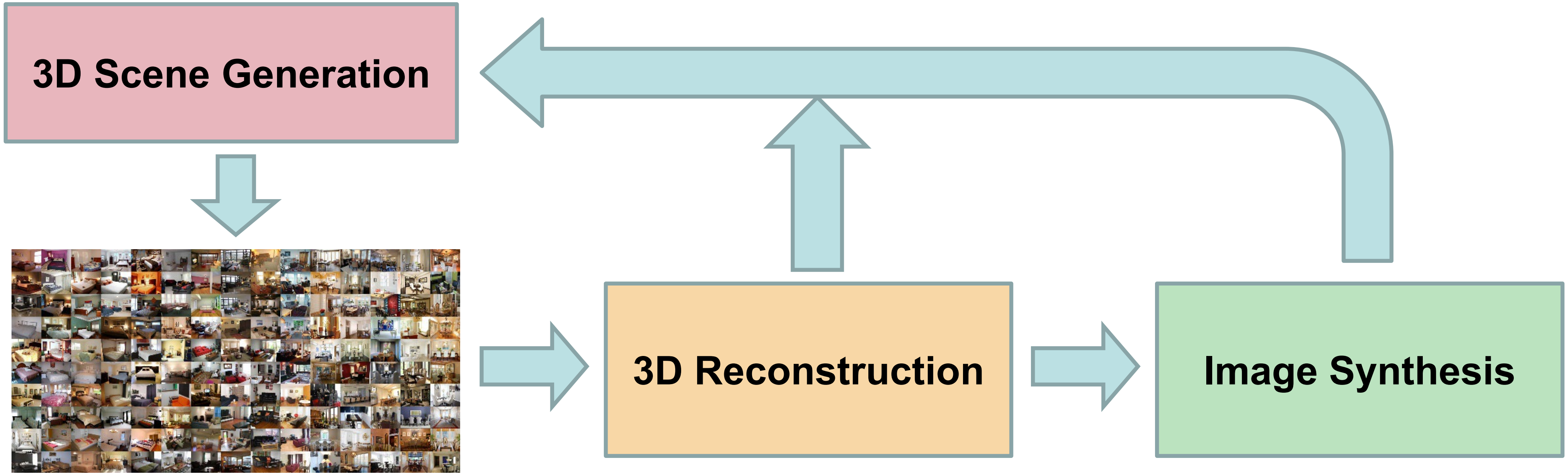
We Live in a World that is 3D and Contains Dynamics

We Digitize Our World in 3D

Future AI: Towards 3D Aware

Lingjie Liu

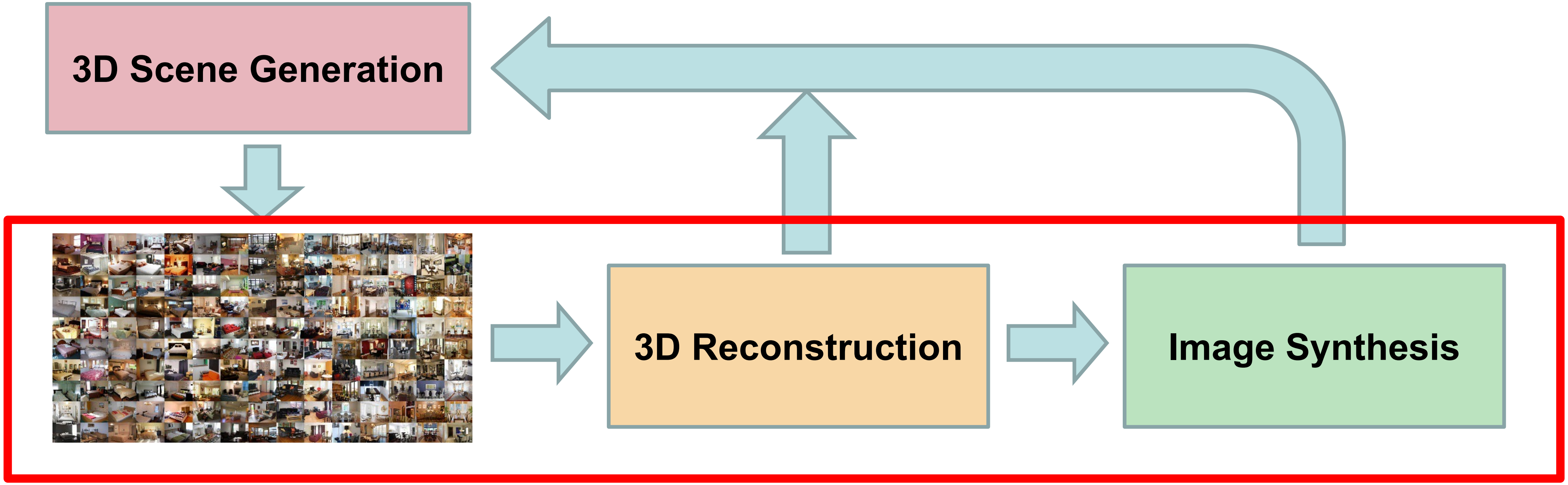
Long-term Vision



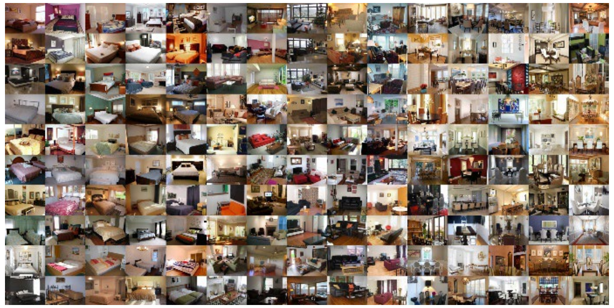
Why are they challenging?

max planck institut informatik

3D Reconstruction and Image Synthesis are Challenging



Classical Computer Graphics Pipeline



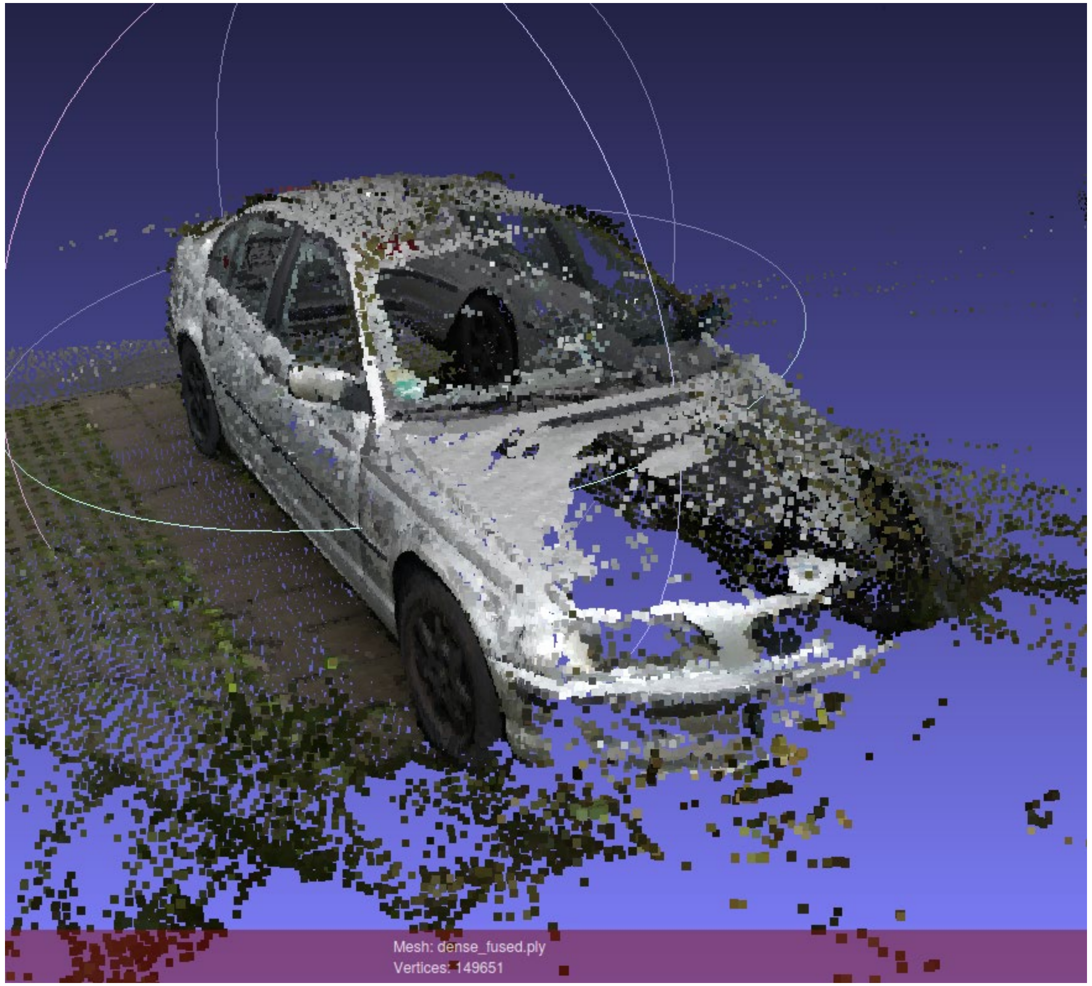
3D Reconstruction

Image Synthesis

Image-based 3D Reconstruction

Computer Graphics Rendering

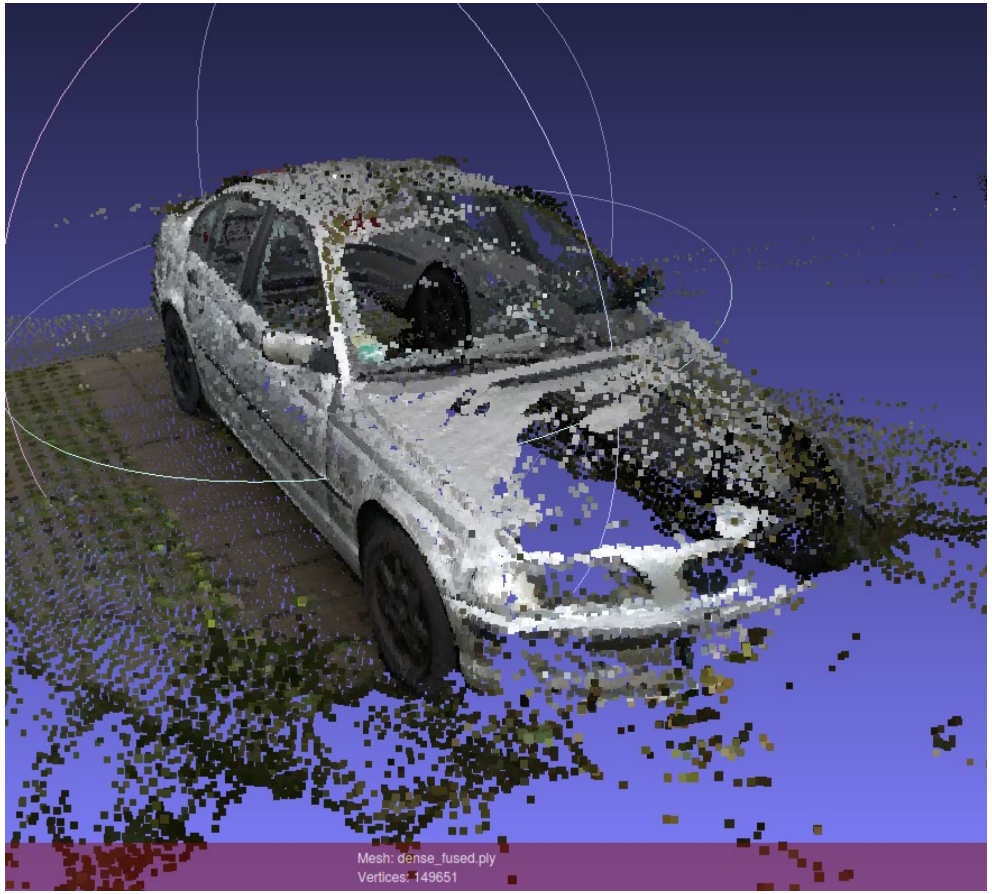
Image-based 3D Reconstruction



COLMAP [Johannes et al. 2016, Schoenberger et al. 2016]
(Input: 100 images)

Computer Graphics Rendering

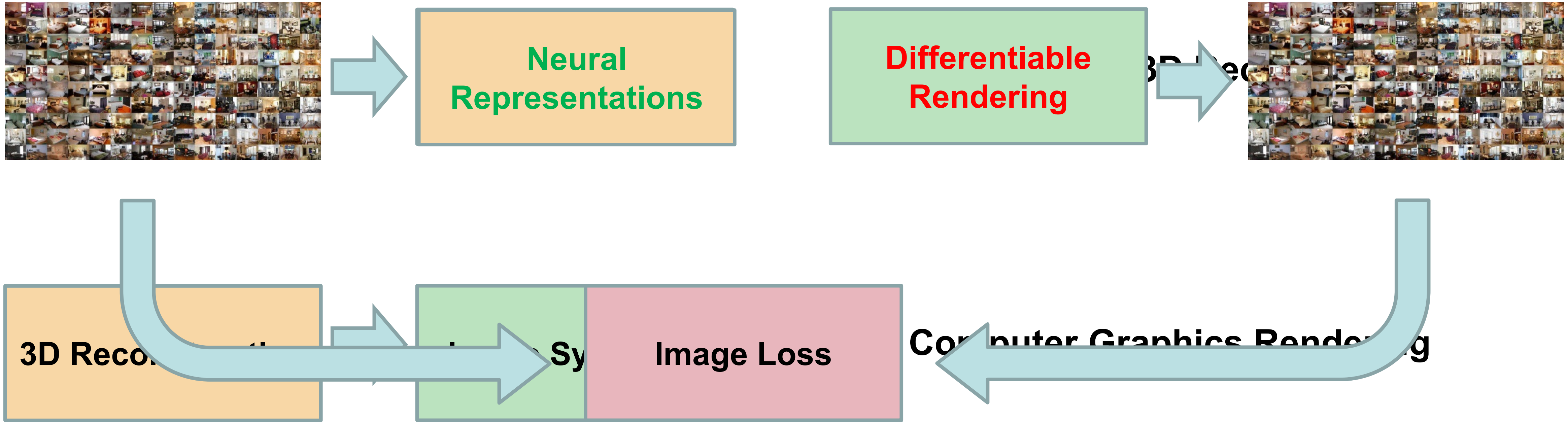
Rendering requires very high-quality 3D models



VS

Self-supervised Learning of 3D Scenes

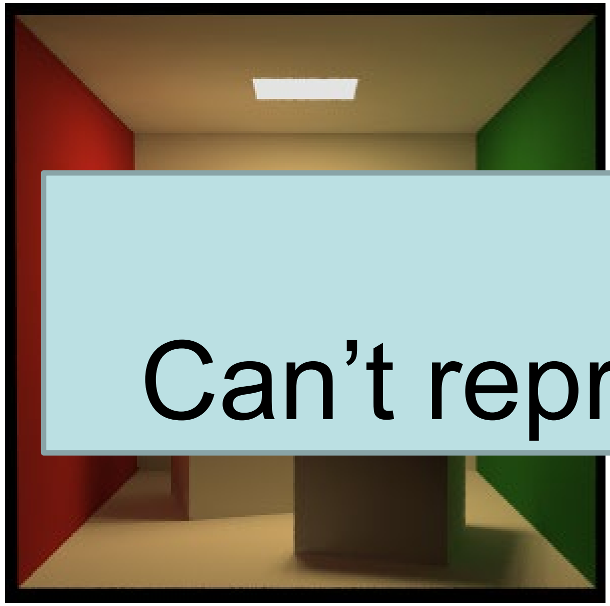
Allow the gradients of 3D objects to be calculated and propagated through images



Classical (Handcrafted) Representations

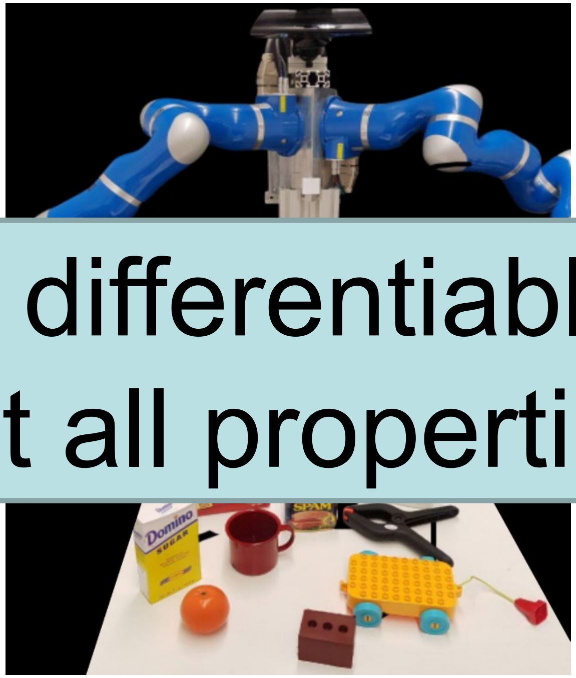
- Features are hand-crafted for each application.

Not differentiable.
Can’t represent all properties of scenes.



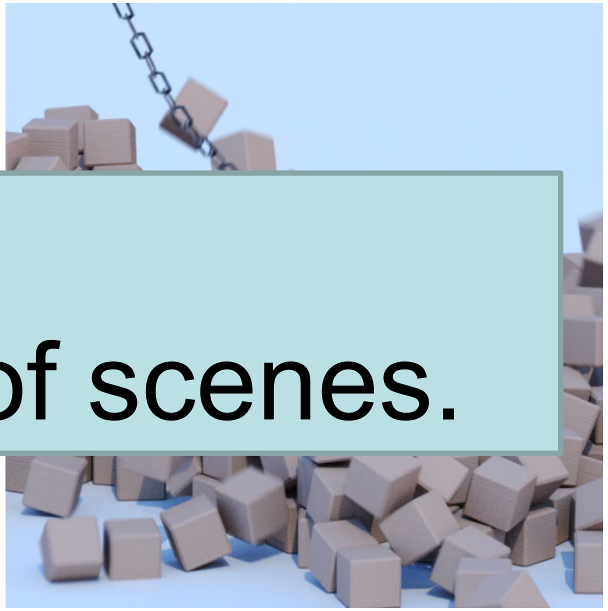
Graphics Rendering

Albedo, specular color, …



Robotics

Geometry, affordance, …

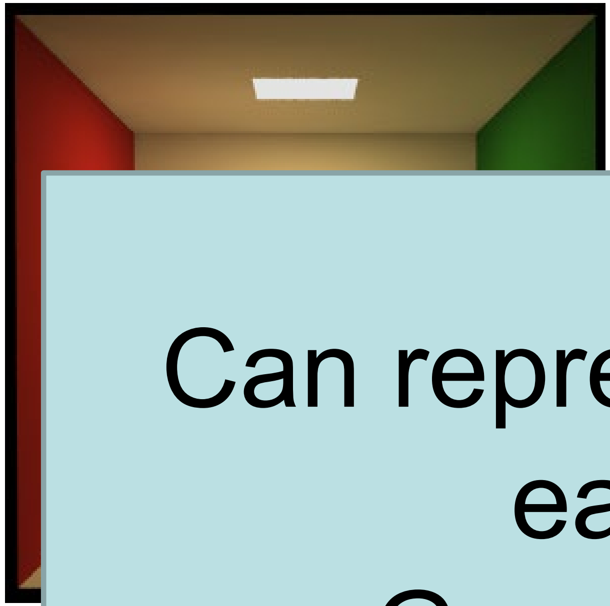


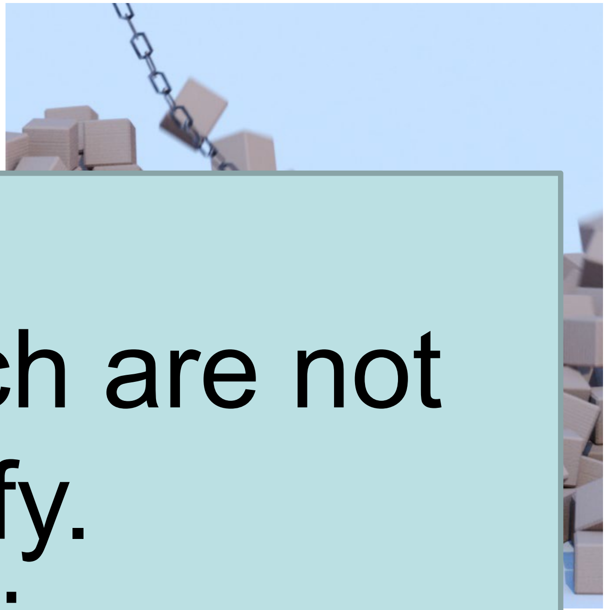
Physics Simulation

Fiction coefficients, mass, …

Neural Scene Representations

- Goal: Learning scene representations from data.

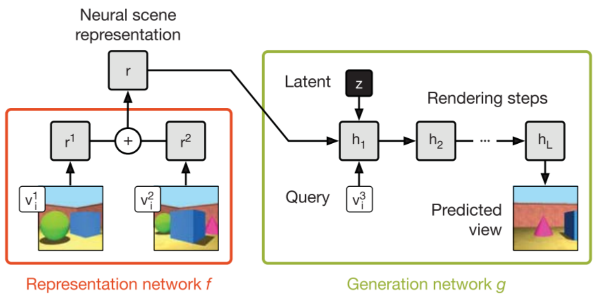




Differentiable.
Can represent properties which are not easy to explicitly specify.
Generic to all the applications.

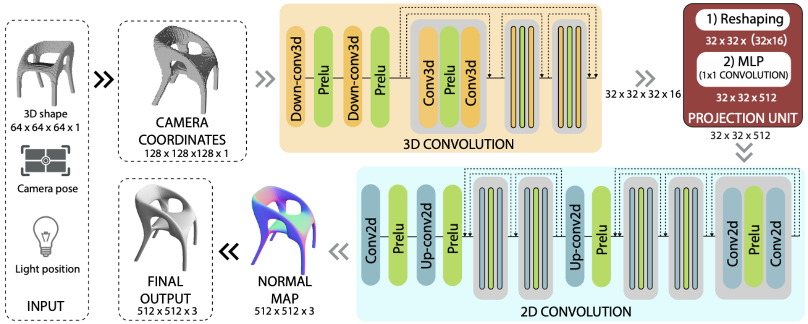
Graphics Rendering	Robotics	Physics Simulation
Albedo, specular color, …	Geometry, affordance, …	Fiction coefficients, mass, …

Neural Scene Representations

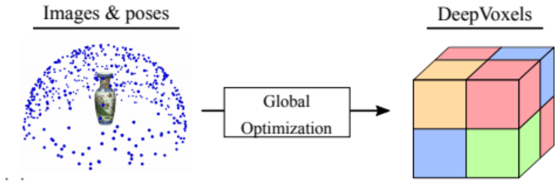


Generative Query Networks
[Eslami et al. 2018]

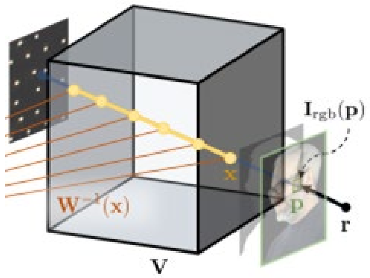
[Flynn et al., 2016; Zhou et al., 2018b; Mildenhall et al. 2019]
Multiplane Images (MPIs)



RenderNet [Nguyen-Phuoc et al. 2018]
Voxel Grids + CNN decoder



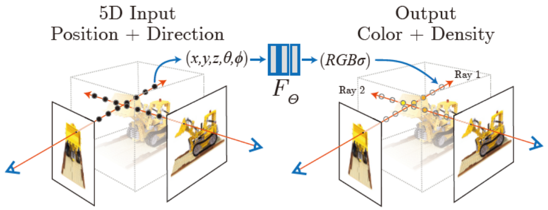
DeepVoxels
[Sitzmann et al. 2019]



Neural Volumes
[Lombardi et al. 2019]

Voxel Grids + Ray Marching

SRN [Sitzmann et al. 2019b]



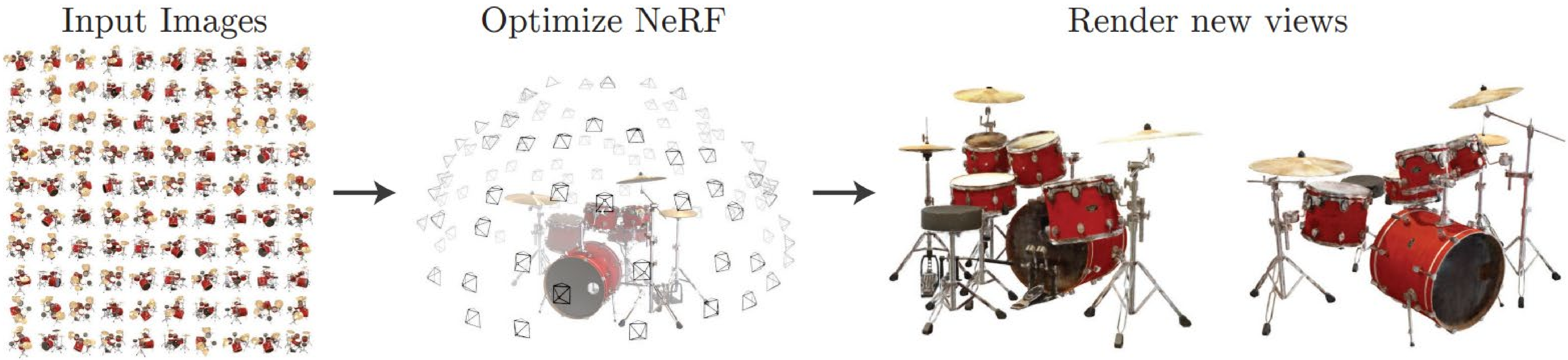
NeRF [Mildenhall et al. 2020]

IDR [Yariv et al. 2020]

Implicit Fields

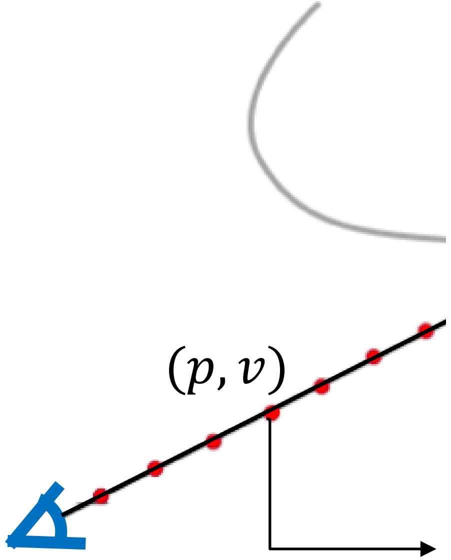
Lingjie Liu

NeRF [Midenhall et al. 2020]



Lingjie Liu

NeRF suffers from a slow rendering process



Rendering speed: 100 s/frame

[Mildenhall et al. 2020]

Image resolution: 1920x1080

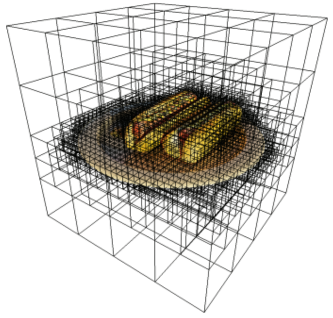
Neural Sparse Voxel Fields (NSVF)

NeRF (Mildenhall et al. 2020)
(Rendering speed: 100 s/frame)

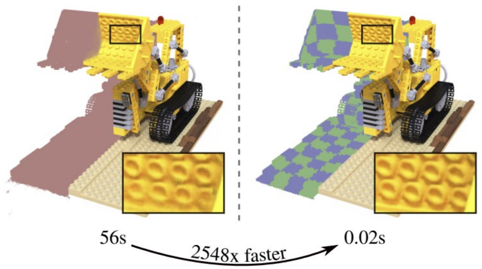
Ours (NSVF)
(Rendering speed: 2.62 s/frame)

L. Liu, J. Gu, K.Z. Lin, T.S. Chua, C. Theobalt. Neural Sparse Voxel Fields, NeurIPS 2020 Spotlight

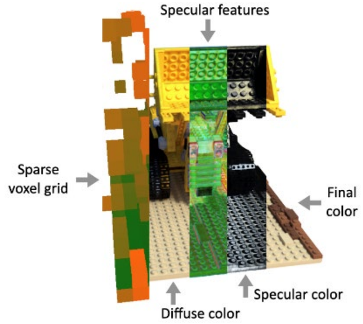
Other Works for Fast Rendering and Fast Training



Yu et. al, 2021

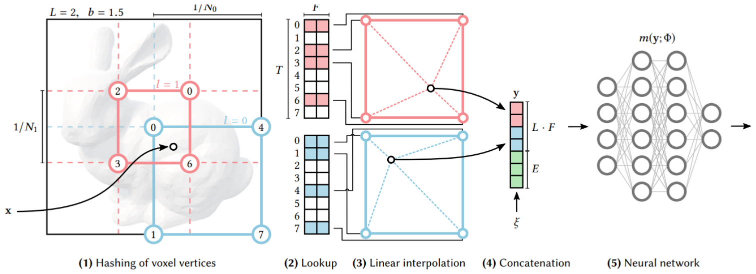


Reiser et. al, 2021



Hedman et. al, 2021

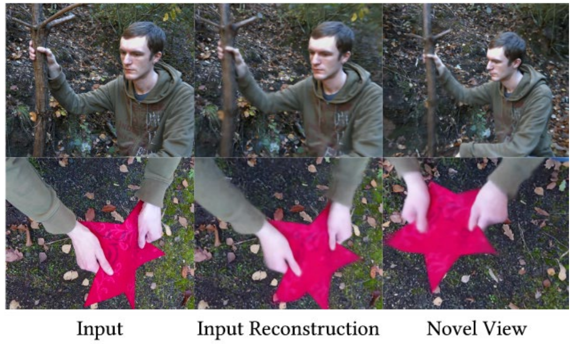
Garbin et. al, 2021



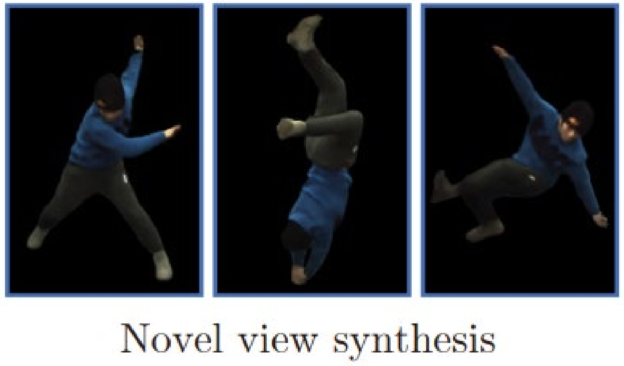
Müller, et. al, 2021

. . .

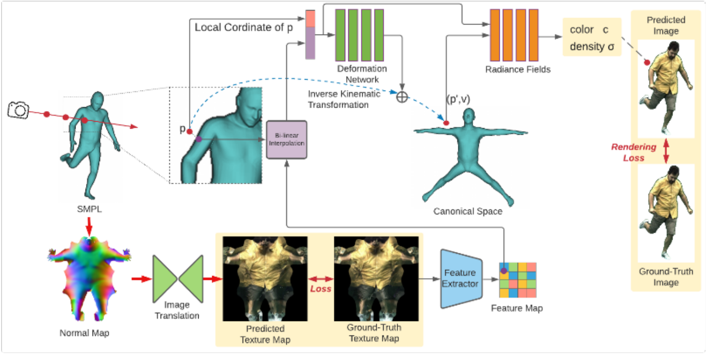
Dynamic Scene Representations



Tretschk et al. 2019, Park et al. 2020, Pumarola et al. 2020,
Li et al. 2020, Xian et al. 2020



Peng et al. 2020, 2021



Neural Actor [Liu et al. 2021]

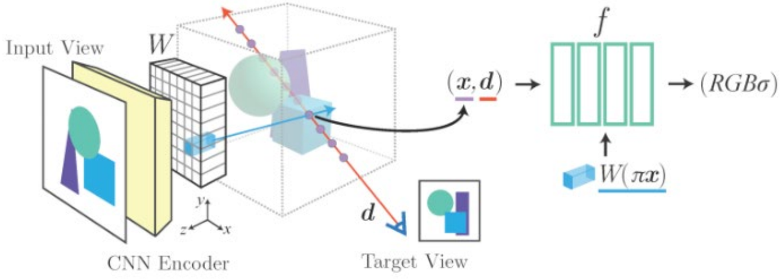
Zheng et al. 2022

HumanNeRF [Zhao et al. 2022]

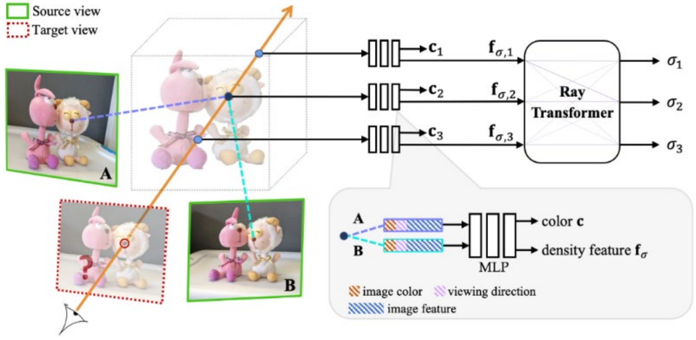
Artemis [Luo et al. 2022]

. . .

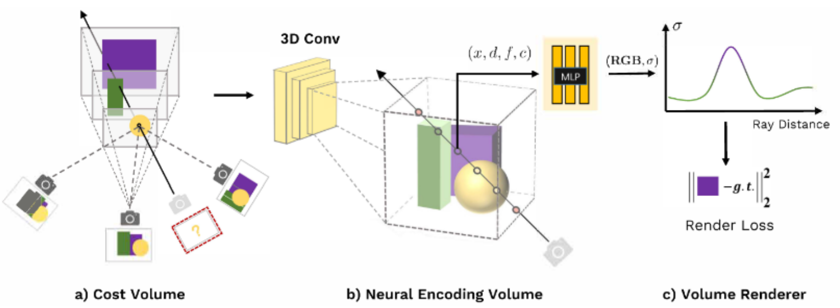
Generalization



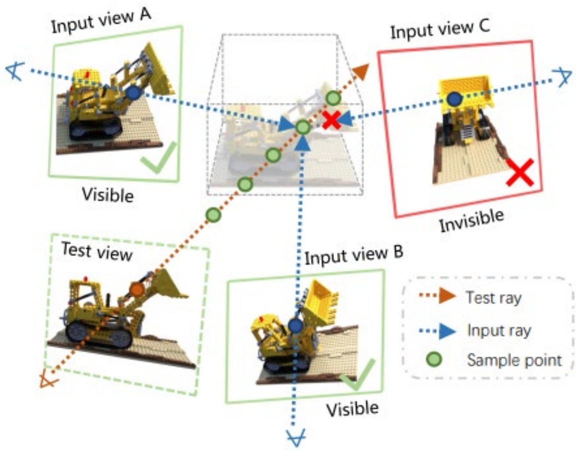
PixelNeRF [Yu et al. 2021]



IBRNet [Wang et al. 2021]



MVSNeRF [Chen et al. 2021]



NeuRay [Liu et al. 2021]

. . .

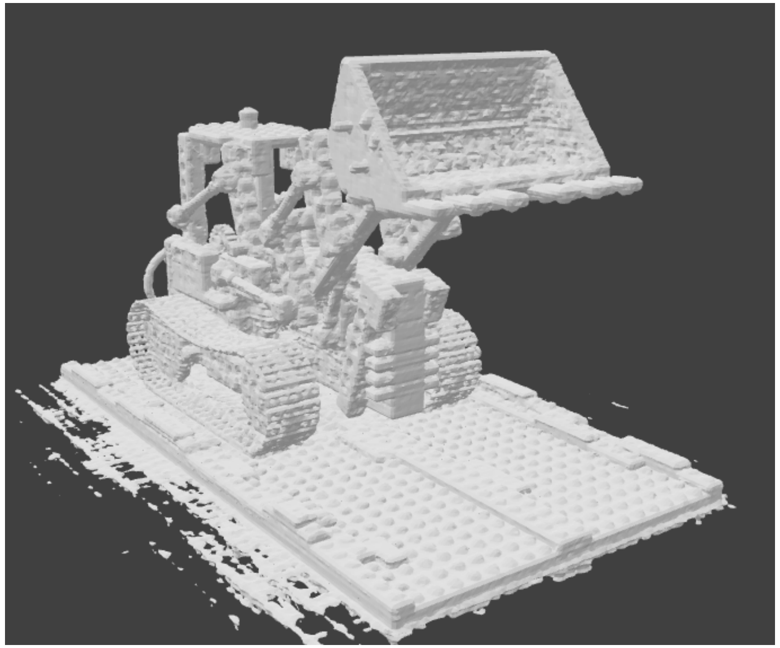
Other Topics

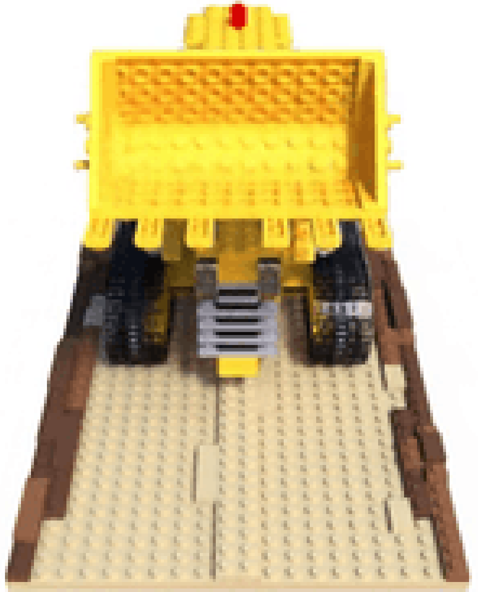
- Pose estimation
- Relighting
- Editing and Composition
- …

Neural Surface Reconstruction

...

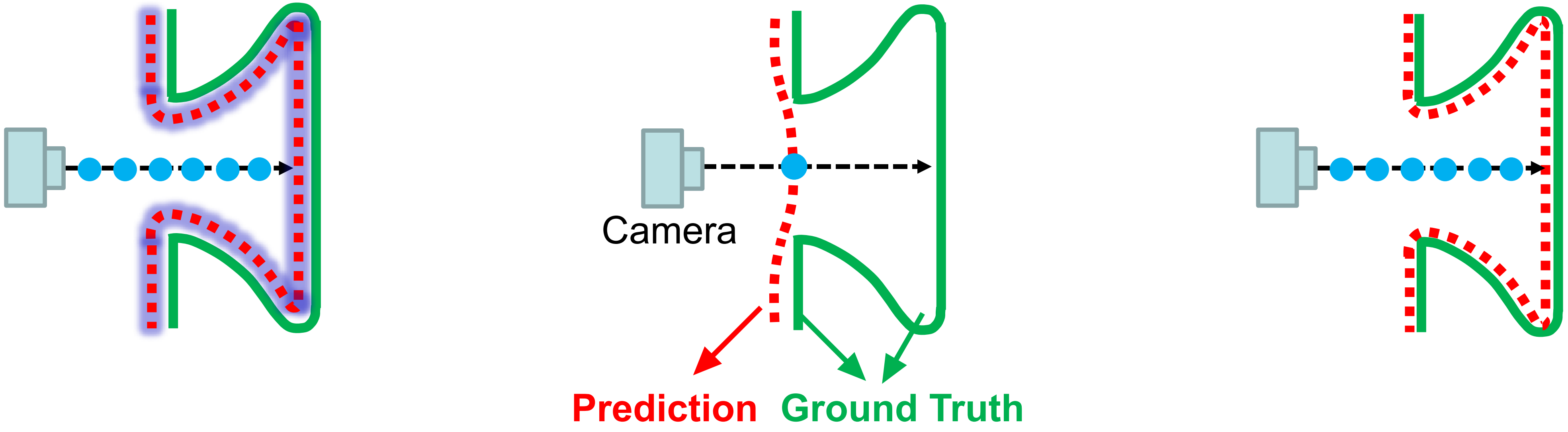
→ Scene Representations →





Volume density used as scene representation lacks surface constraints

Surface Representation + Volume Rendering



Volume Representation + Volume Rendering

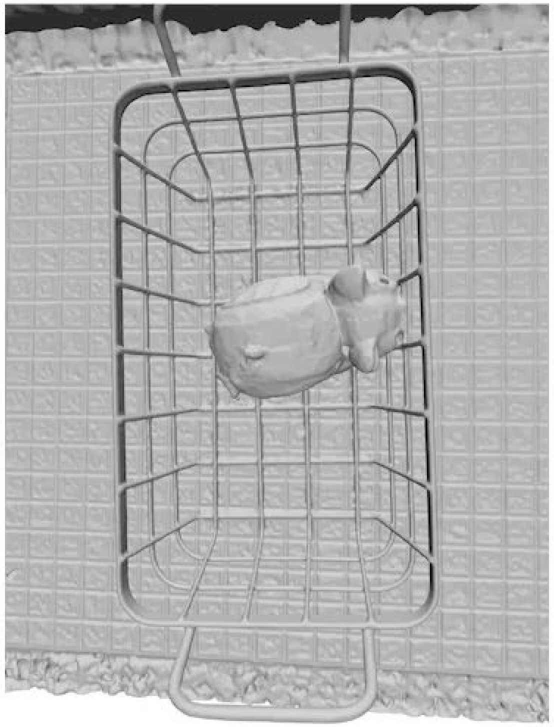
Surface Representation + Surface Rendering

Surface Representation + Volume Rendering

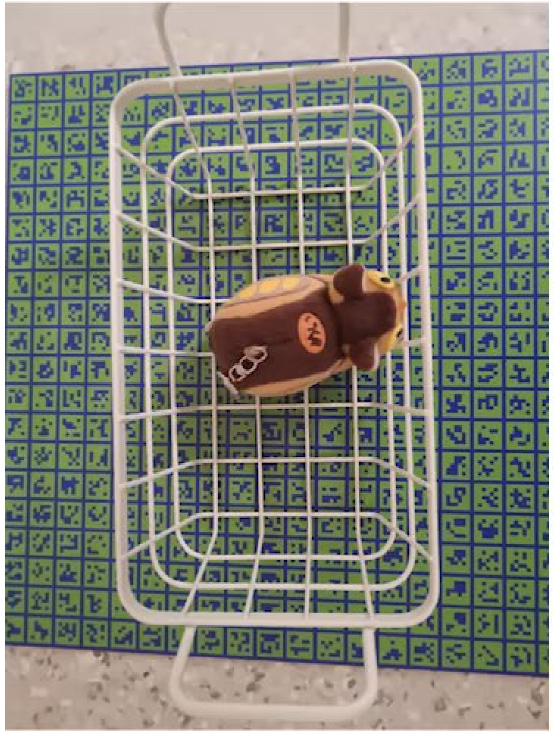
P. Wang, ***L. Liu,*** Y. Liu, C. Theobalt, T. Komura, W. Wang. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction, NeurIPS 2021 Spotlight

Results of NeuS

A subset of input images



Our surface geometry
(w/o mask supervision)

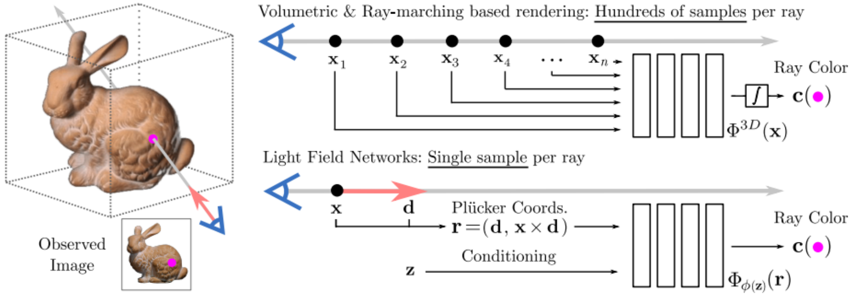


Our rendering
(w/o mask supervision)

max planck institut
informatik

Other New Neural Scene Representations

Light Fields



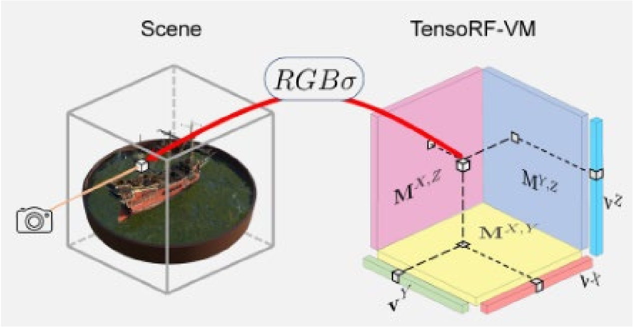
LFN [Sitzmann et al. 2022]

ProLiF [Wang et al. 2022]

. . .

TriPlane

EG3D [Or-El et al. 2022]



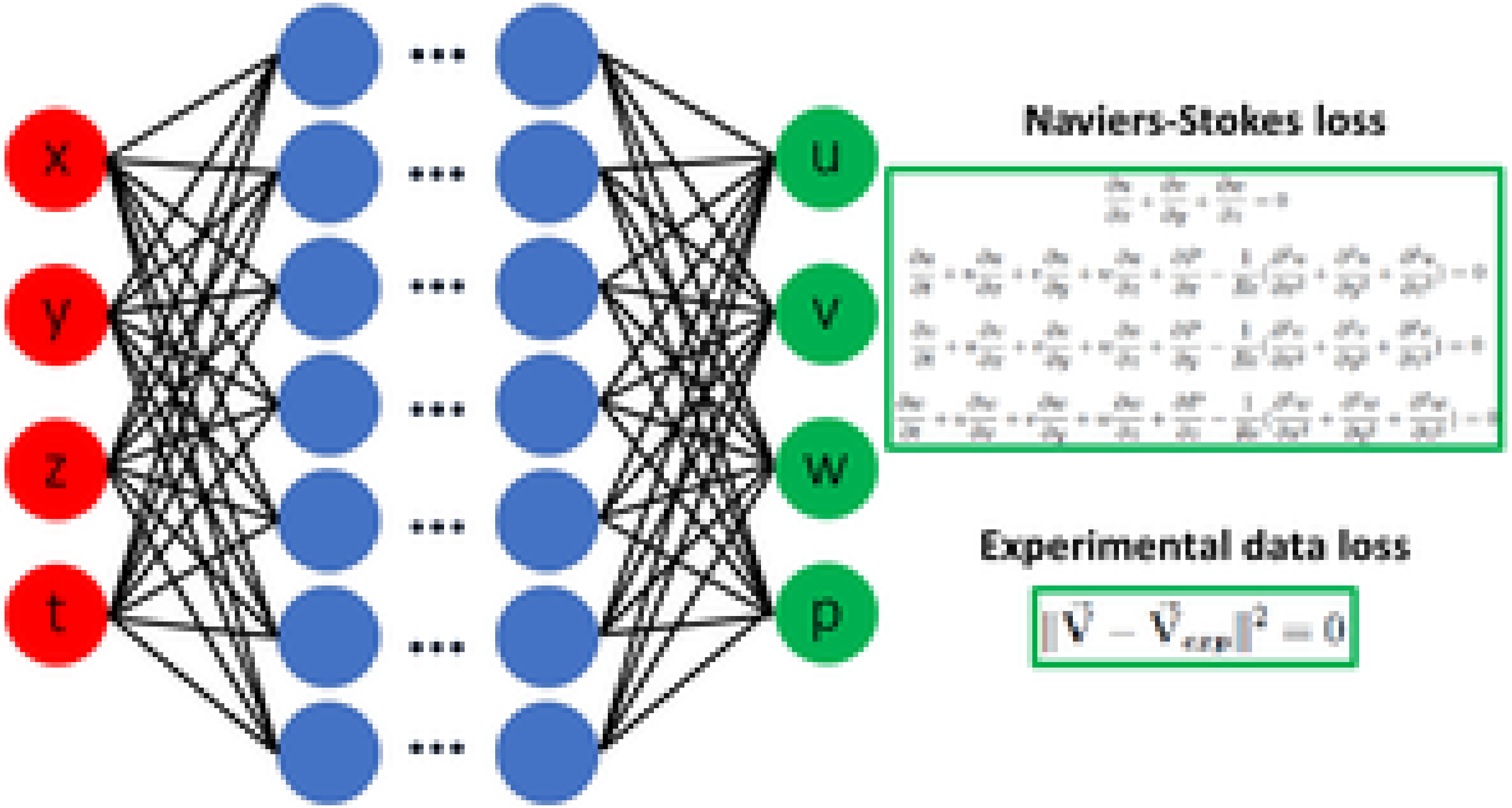
TensoRF [Chen et al. 2022]

. . .

Lingjie Liu

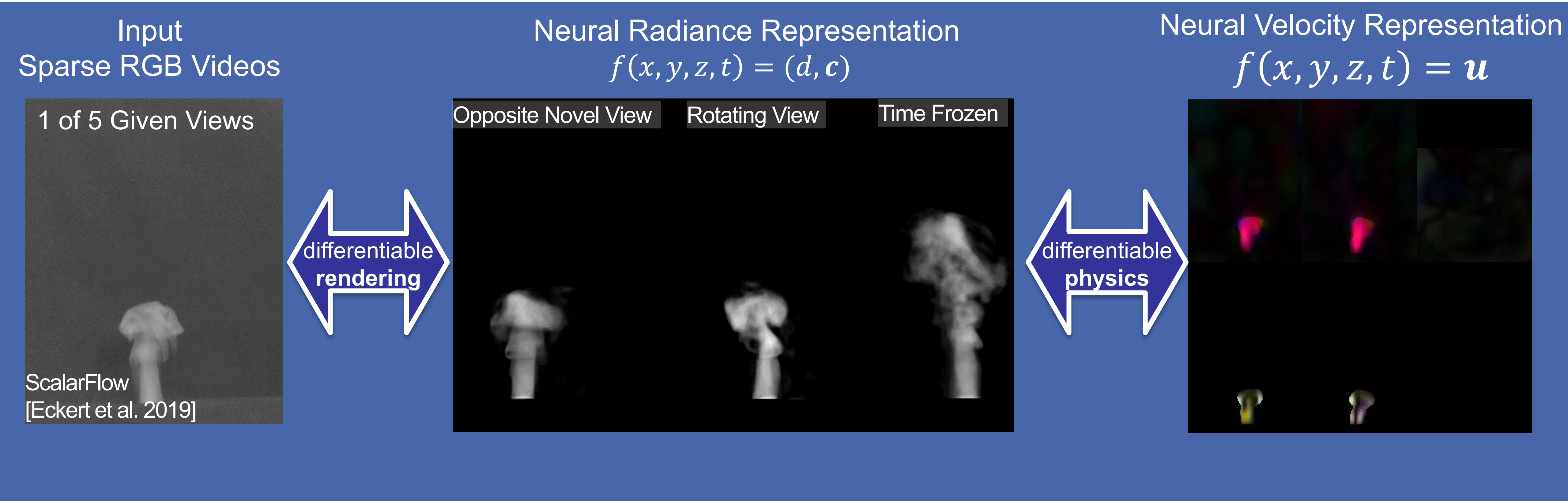
Other New Neural Scene Representations

- PINN



Other New Neural Scene Representations

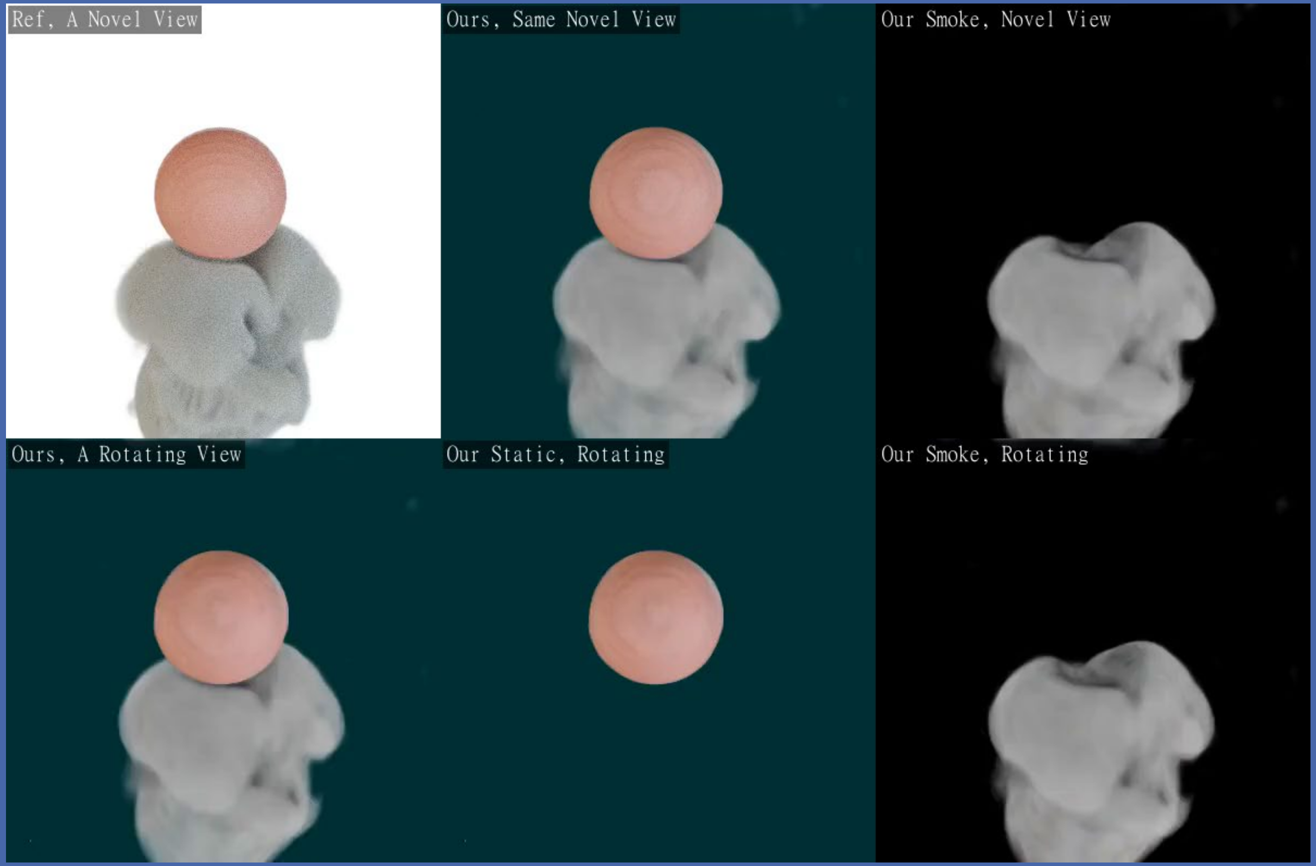
- Physics Informed Neural Fields



M. Chu, ***L. Liu***, Q. Zheng, E. Franz, H.P. Seidel, C. Theobalt, R. Zayer.
Physics Informed Neural Fields for Smoke Reconstruction with Sparse Data, SIGGRAPH 2022 (Journal track)

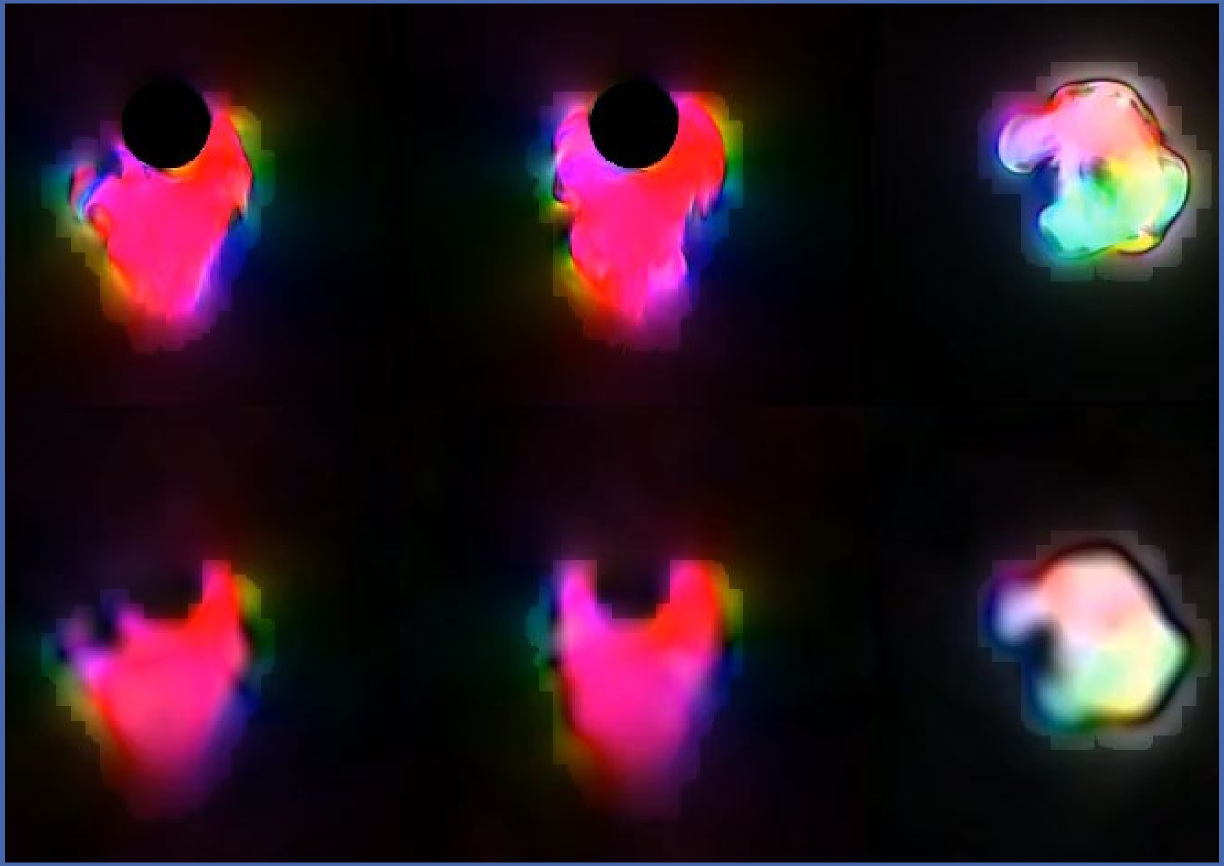
Result Comparisons: Hybrid Scene

Reference, Novel View — Ours, Novel View — Our smoke, Novel View

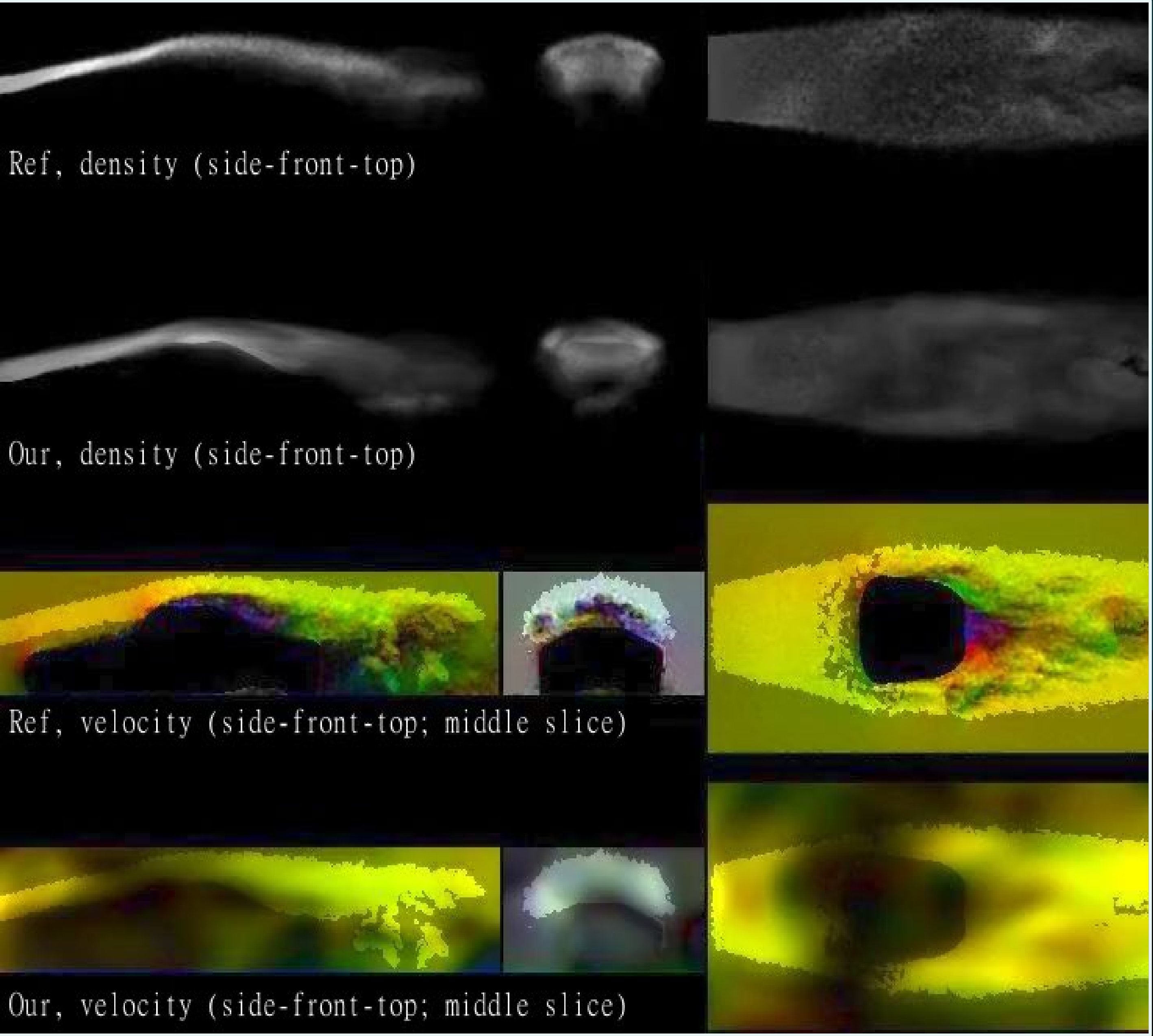


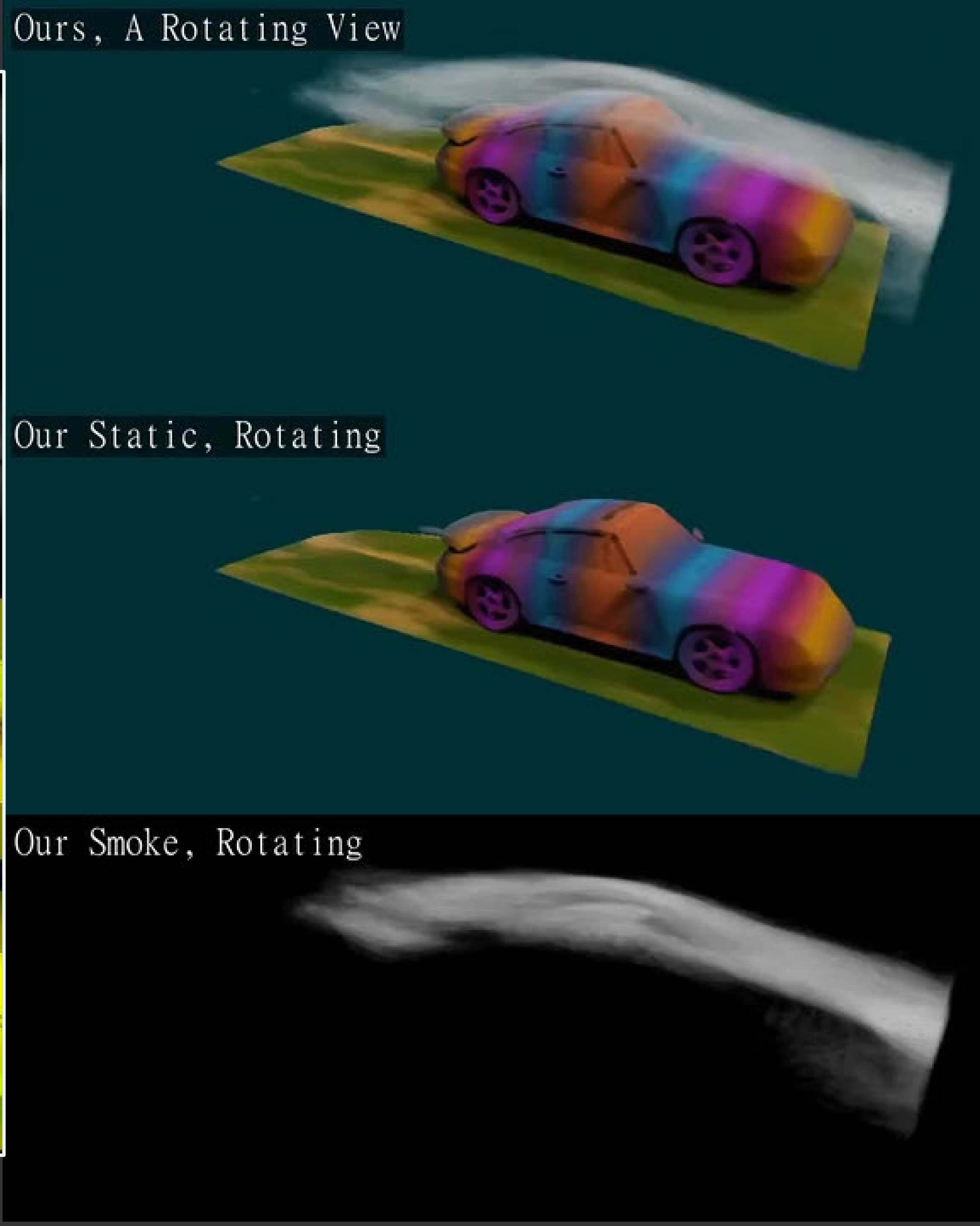
Ours, Rotating View — Our static, Rotating View — Our smoke, Rotating View

Reference velocity

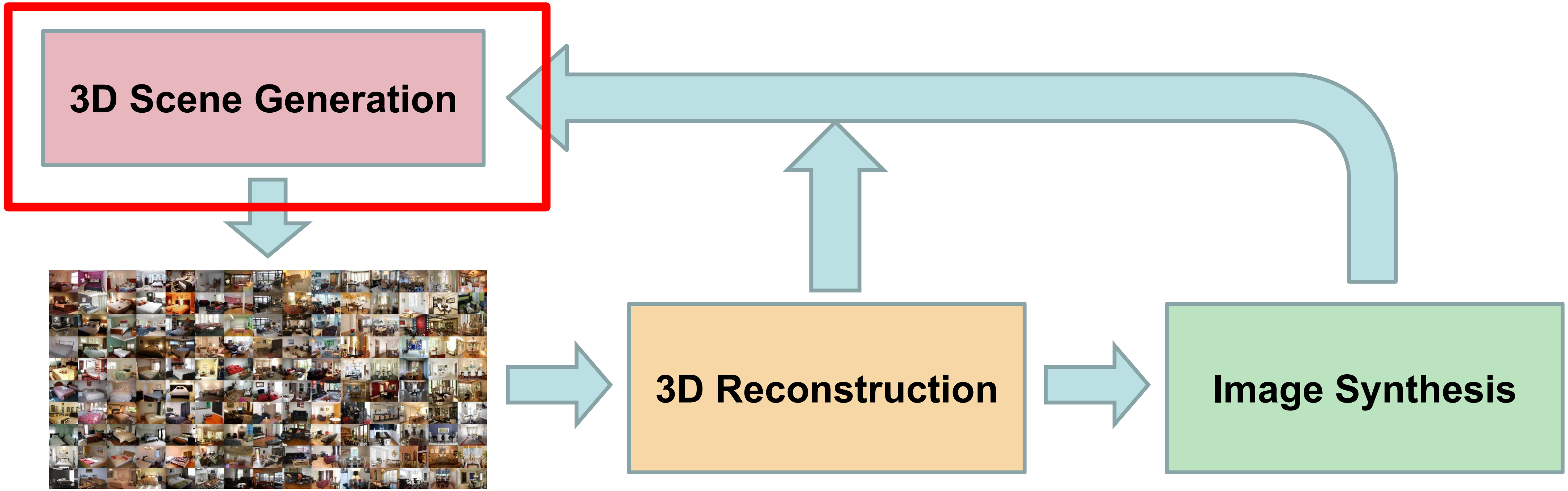


Our velocity

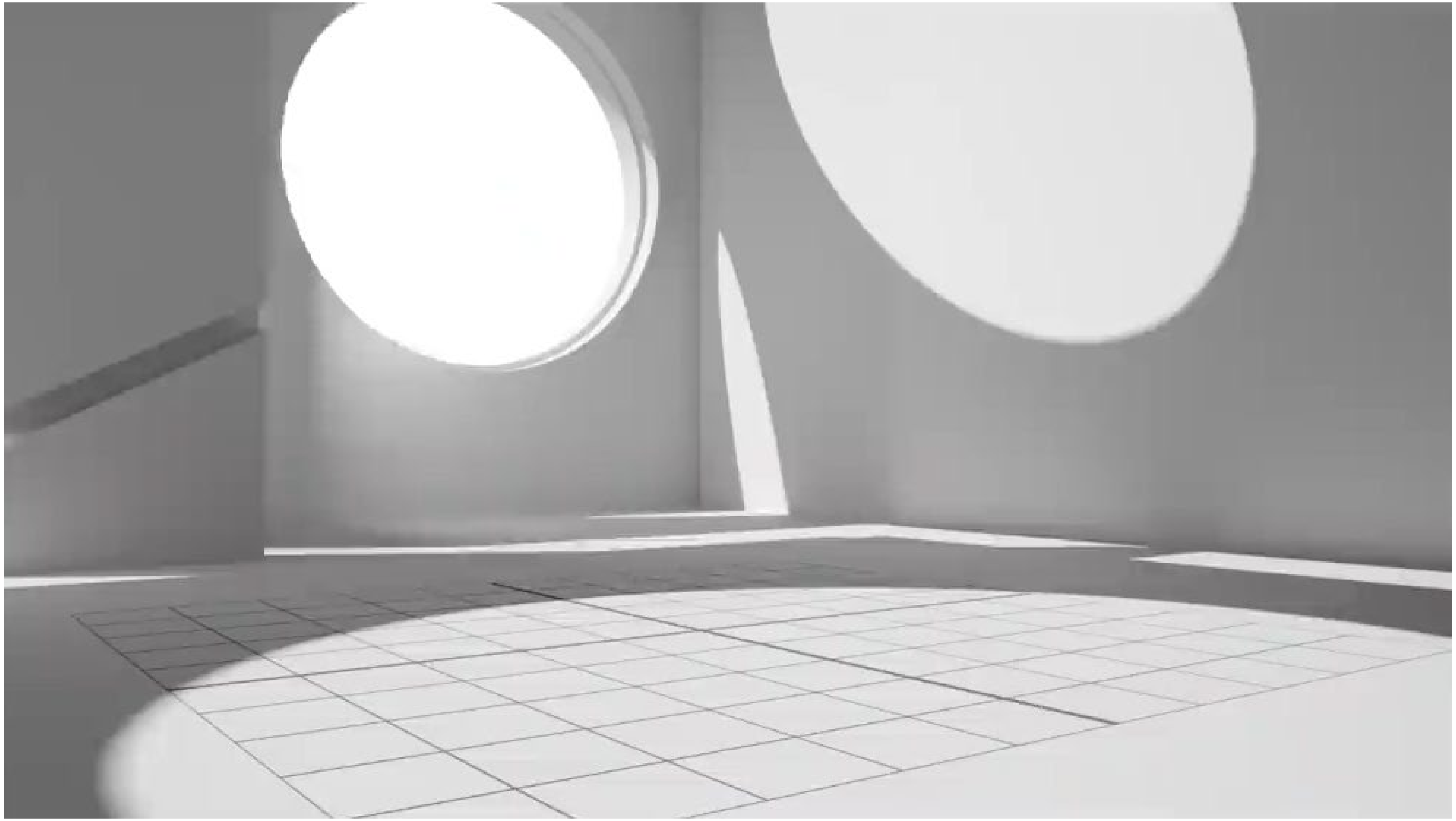

Ref, density (side-front-top)
Our, density (side-front-top)
Ref, velocity (side-front-top; middle slice)
Our, velocity (side-front-top; middle slice)


Ours, A Rotating View
Our Static, Rotating
Our Smoke, Rotating

Scene Generation is challenging



Scene Generation is challenging



How to Generate New 3D Scenes?

Training Data?   Multi-view Images?

How to Generate New 3D Scenes?

Training Data?

Single-view Images? 😃

How to Generate New 3D Scenes?

Model?

2D GANs?

Results of the state-of-the-art GAN model (StyleGAN2)

Generate merely 2D images, without 3D information

Lingjie Liu

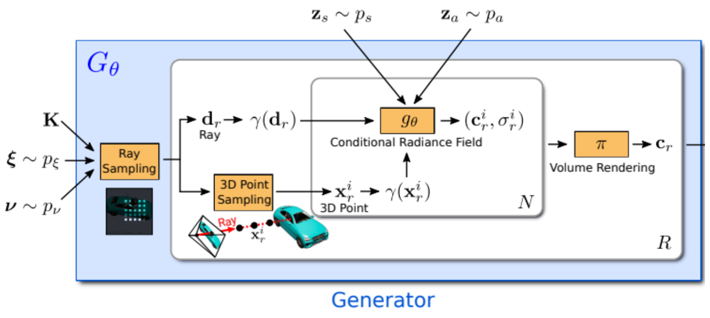
How to Generate New 3D Scenes?

Model?

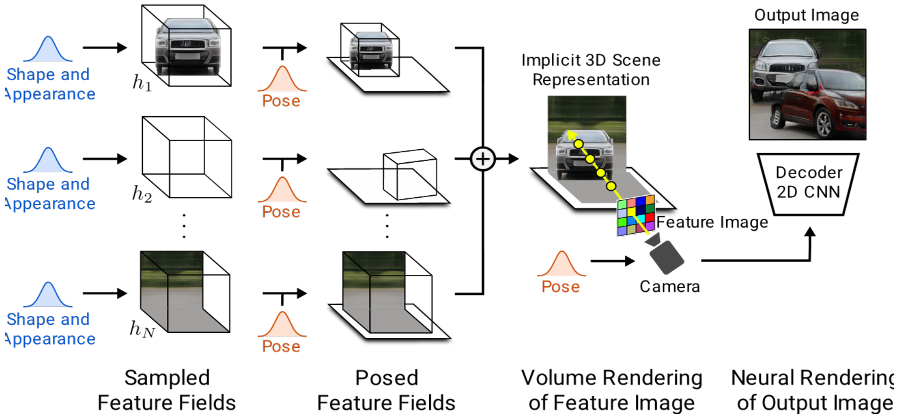
3D GANs?

J. Gu, ***L. Liu,*** P. Wang, C. Theobalt.
StyleNeRF: A Style-based 3D-Aware Generator for High-resolution Image Synthesis, ICLR 2022

Other 3D GAN Models

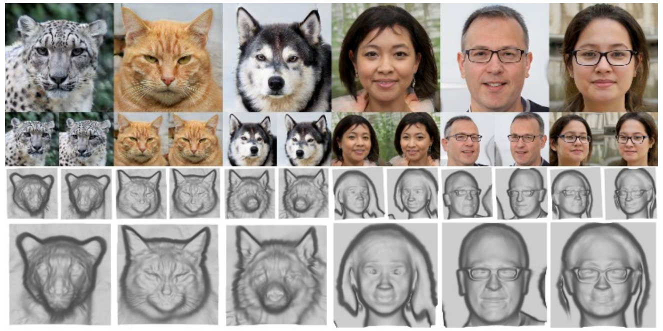


GRAF [Schwarz et al. 2020]

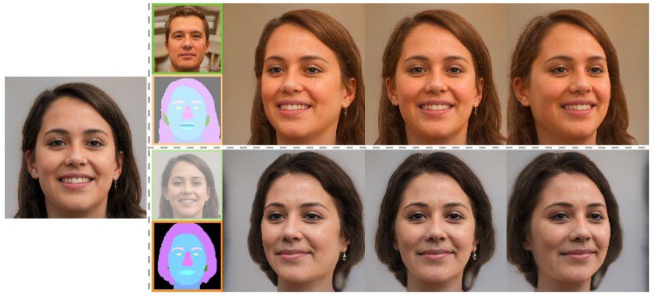


GIRAFFE [Niemeyer et al. 2020]

EG3D [Or-El et al. 2022]



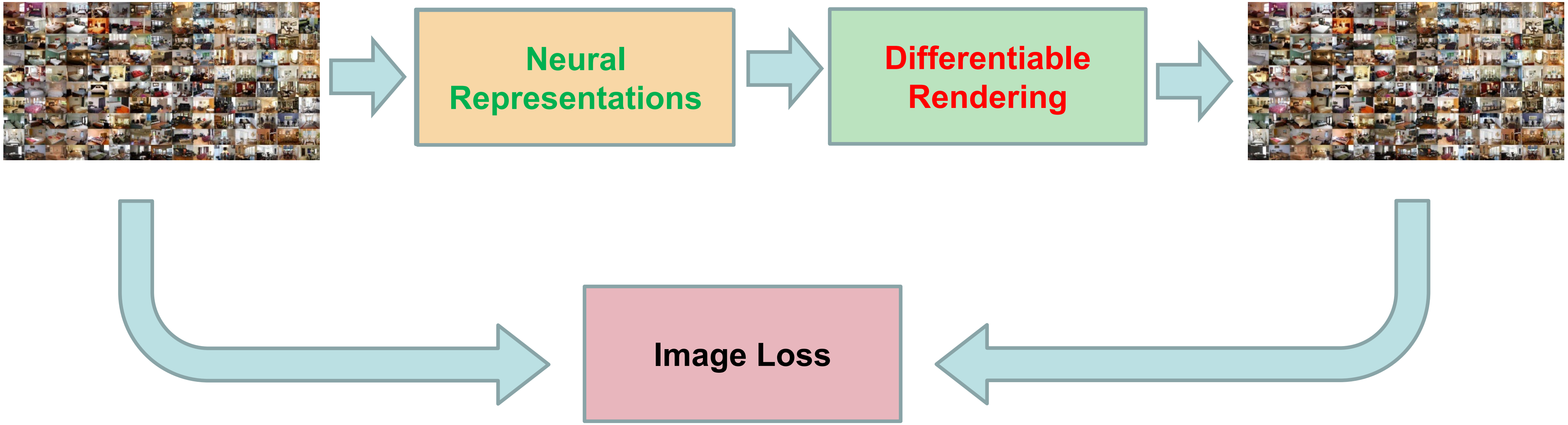
StyleSDF [Or-El et al. 2022]



Jiang et al. 2022, Sun et al. 2022

. . .

What’s Next?


Neural
Representations
Differentiable
Rendering
Image Loss

Modeling More Complex Scenes

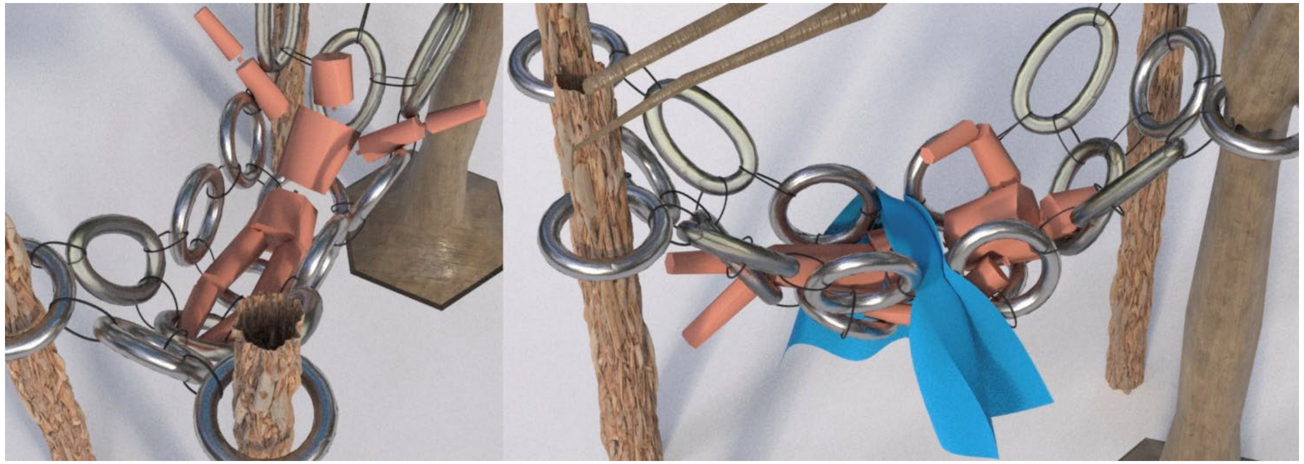
Modeling More Complex Scenes

Human-scene Interaction

Physics and environment constraints

Generalizability

Neural Representations for Simulation



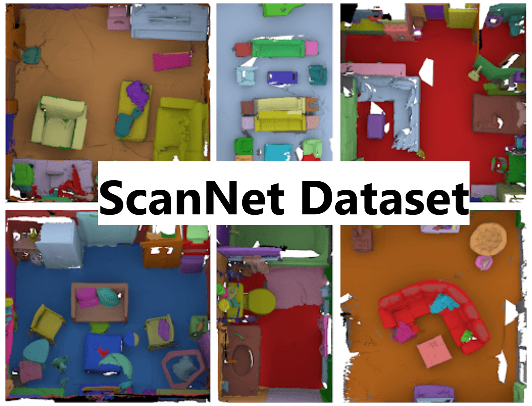
Large-scale 3D Data Generation

The size of 2D datasets can be as large as millions

The ImageNet dataset contains **millions** of images

Existing 3D data is far from sufficient

12k synthetic models



1k indoor scenes

1k indoor scenes

7.2k demonstrations
of a robot performing kitchen tasks

max planck institut informatik

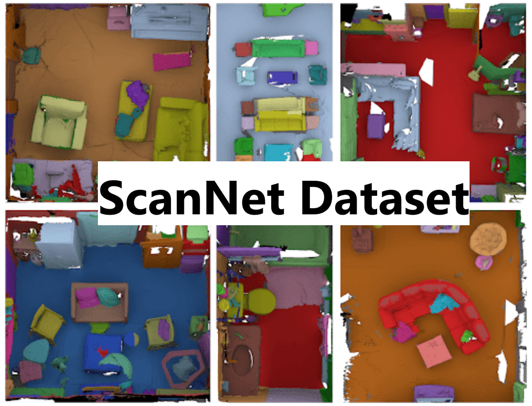
Large-scale 3D Data Generation

The size of 2D datasets can be as large as millions

The ImageNet dataset contains **millions** of images

Learn 3D generative models from 2D image datasets

12k synthetic models



1k indoor scenes

1k indoor scenes

7.2k demonstrations of a robot performing kitchen tasks

Lingjie Liu

Large-scale 3D Data Generation

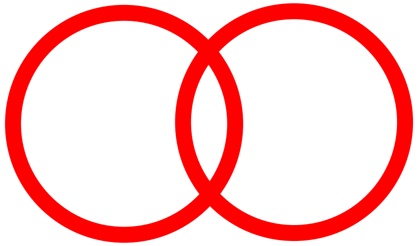
The ImageNet dataset contains **millions** of images

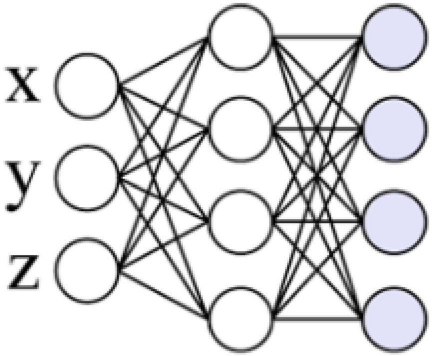
Learn 3D generative models from 2D image datasets

This is the first time that a generative model can synthesize high-resolution images from novel views while preserving high 3D consistency

Multi-modal Learning

Large-scale multimodal learning models



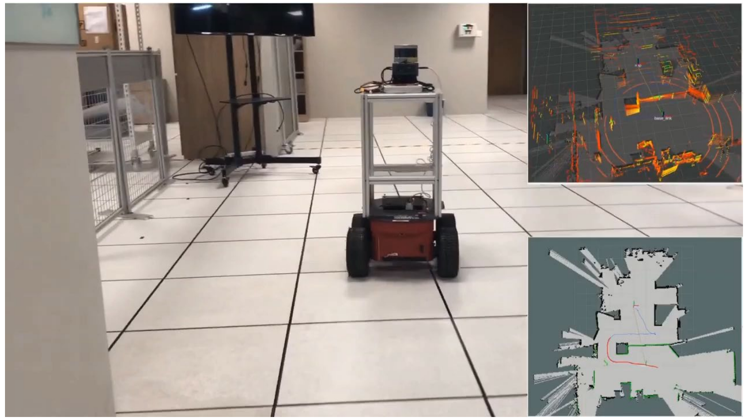


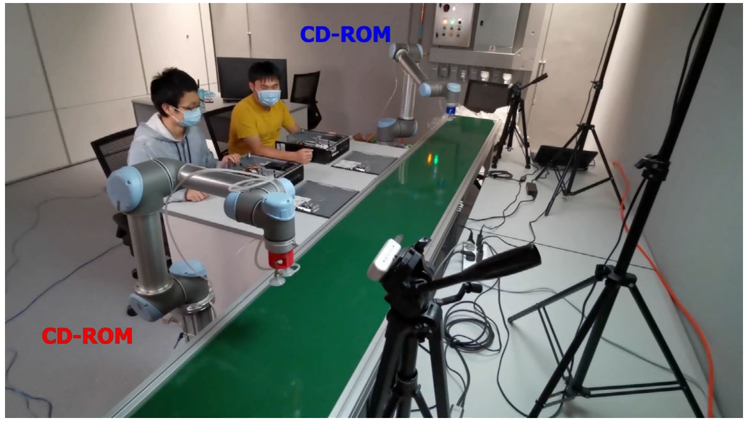
Neural scene representations

Lingjie Liu

Applications

- Extending neural scene representation and rendering techniques to other domains outside of computer graphics, such as robotics.





6DOF estimation, 3D/4D scene reconstruction and segmentation, motion prediction
for robot grasping, robot navigation and planning, human-robot interaction.

Thank you!