New Methods for Reconstruction and Rendering of 3D Real-world Scenes

Lingjie Liu

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

My Lab @ UPenn

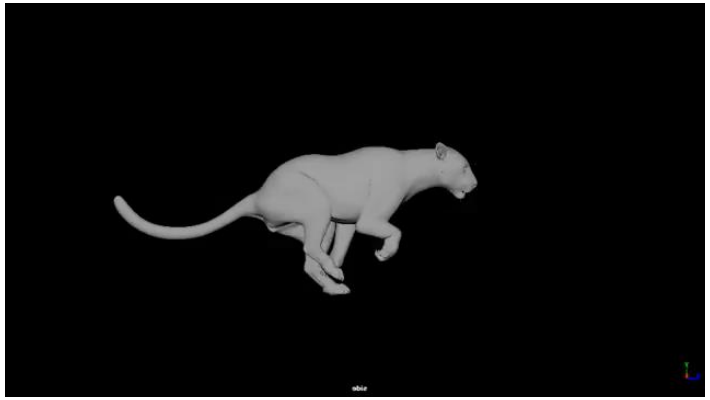
More Info:
https://lingjie0206.github.io/

Computer Graphics, Computer Vision & AI

I'm looking for
PhD students, postdocs &
visiting students

Reconstruction of 3D Real-world Scenes

Geometry
+ Appearance



Motion
+ Deformation

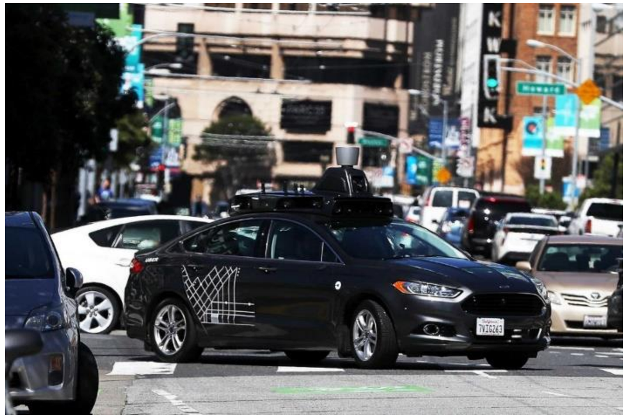
Photo-realistic Rendering

Photo-realistic Rendering

Why Are They Important?

AR / VR

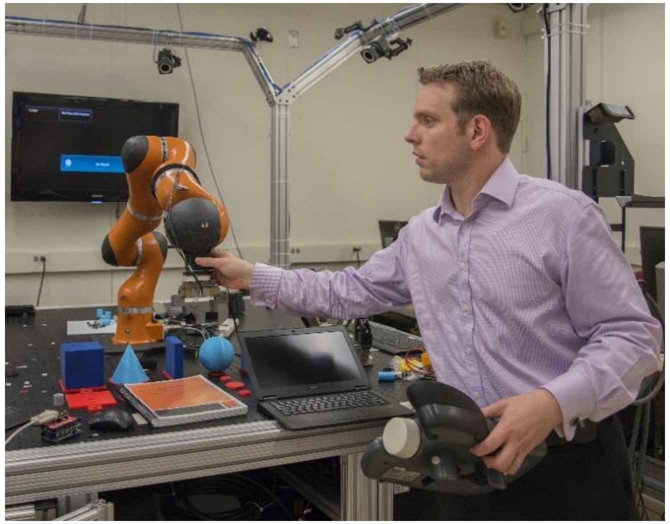
Gaming / Movie



Autonomous Driving

Robot Grasping

Healthcare



Human-robot Interaction

Why Challenging?

Classical Computer Graphics Pipeline

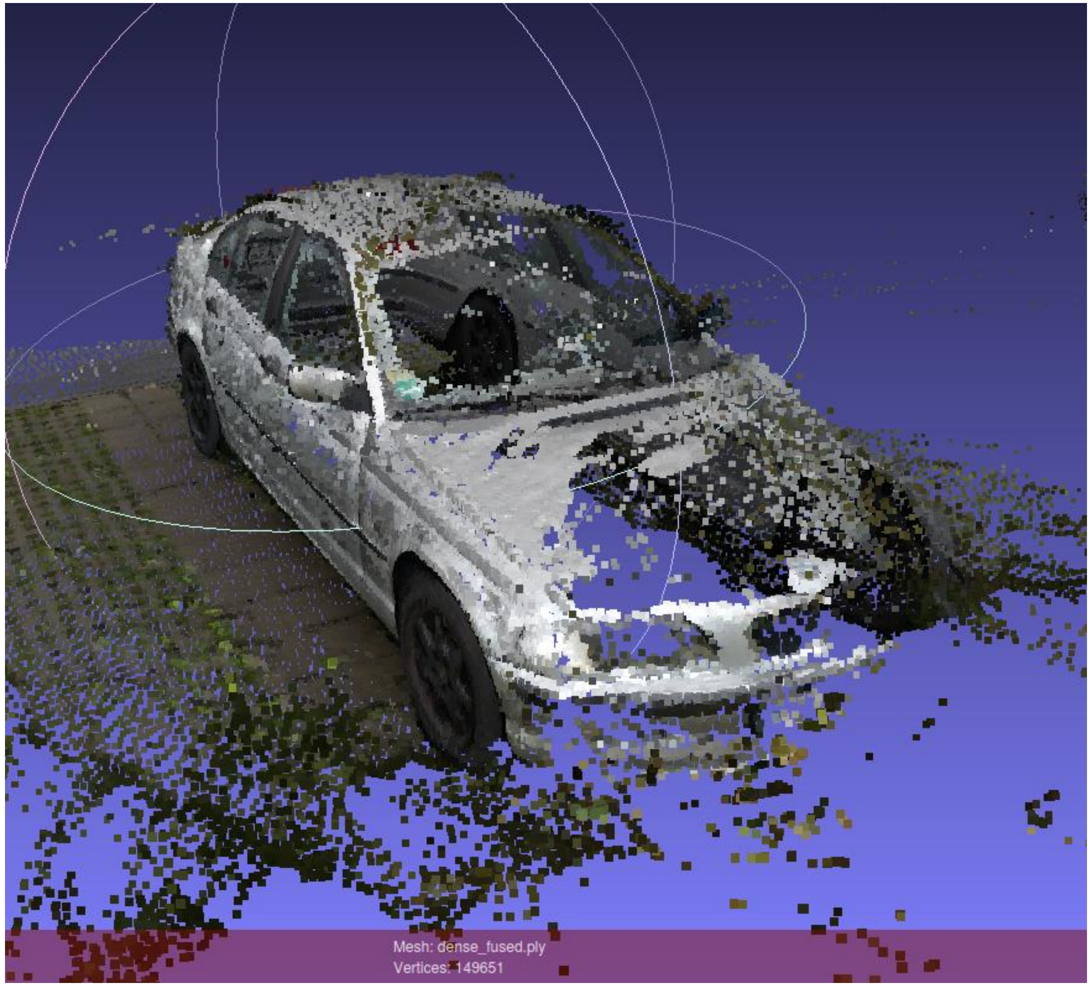
3D Reconstruction

Image-based 3D Reconstruction

Image Synthesis

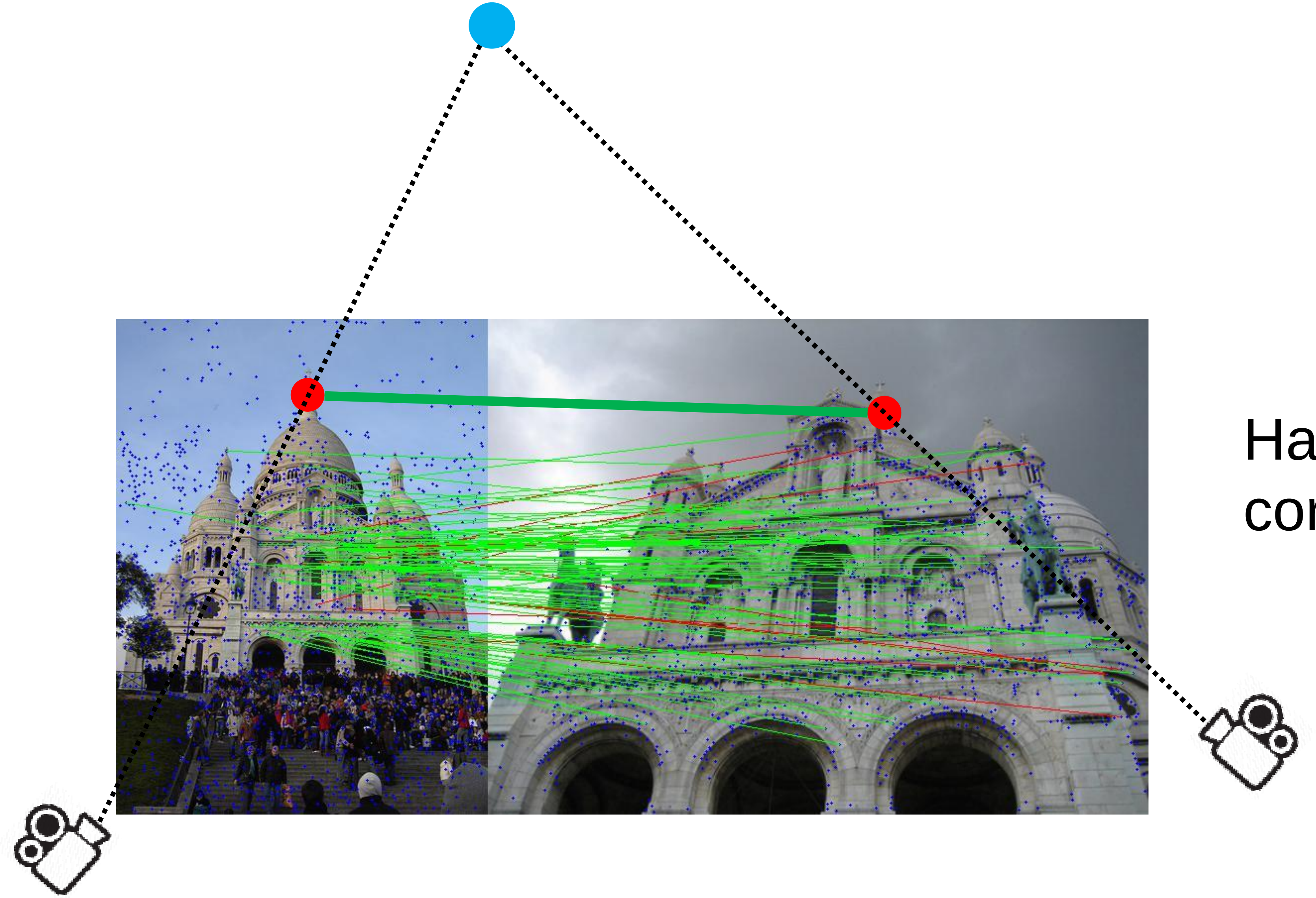
Computer Graphics Rendering

Image-based 3D Reconstruction



Colmap: (Input: 100 images)

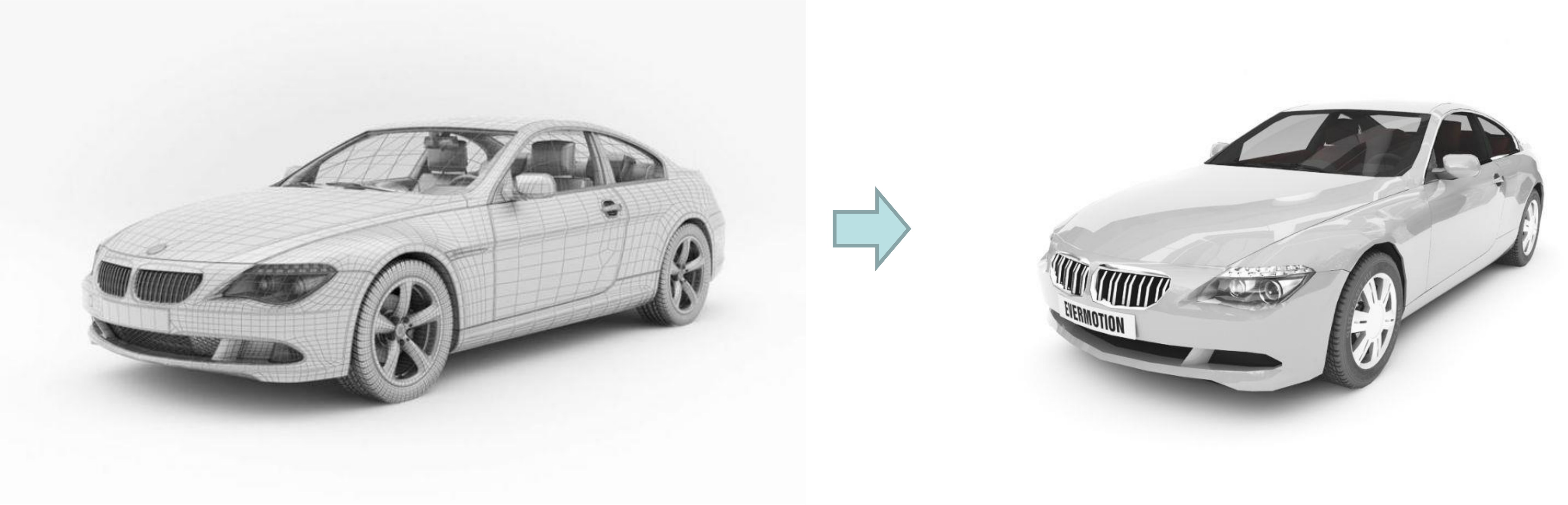
Challenges in Image-based Reconstruction



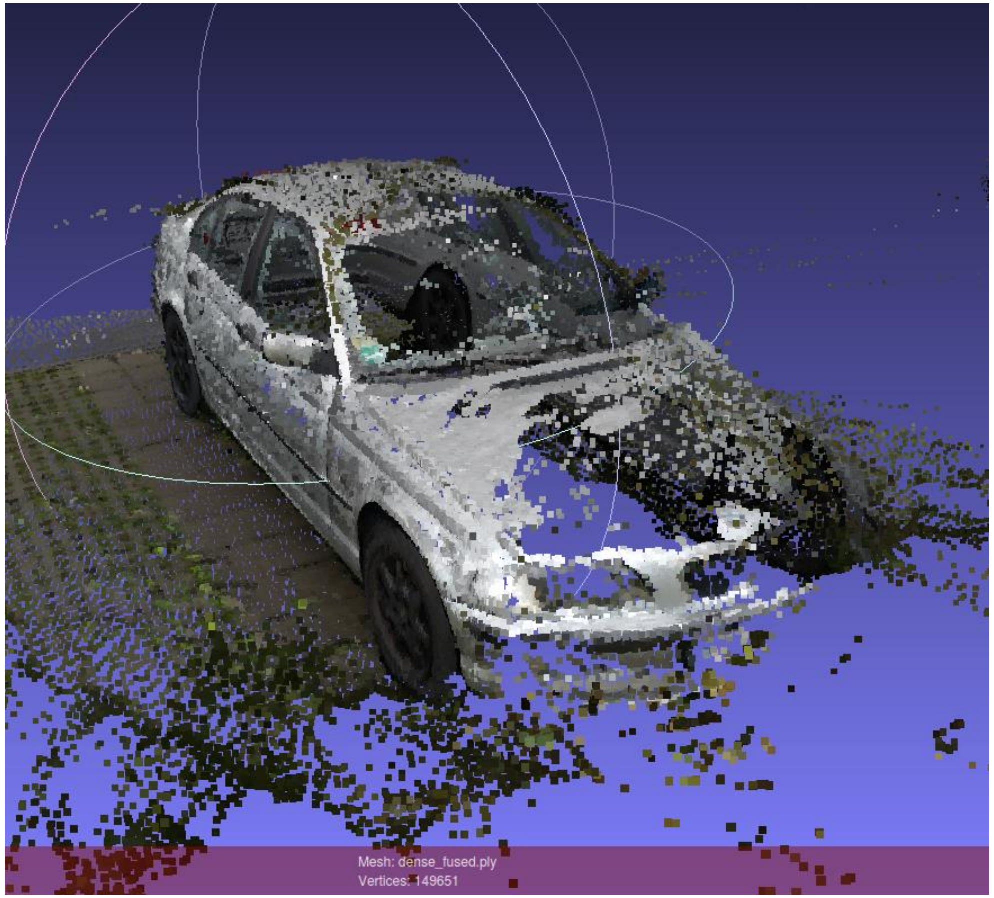
Hard to extract reliable correspondences!

Challenges in Image-based Reconstruction

Computer Graphics Rendering



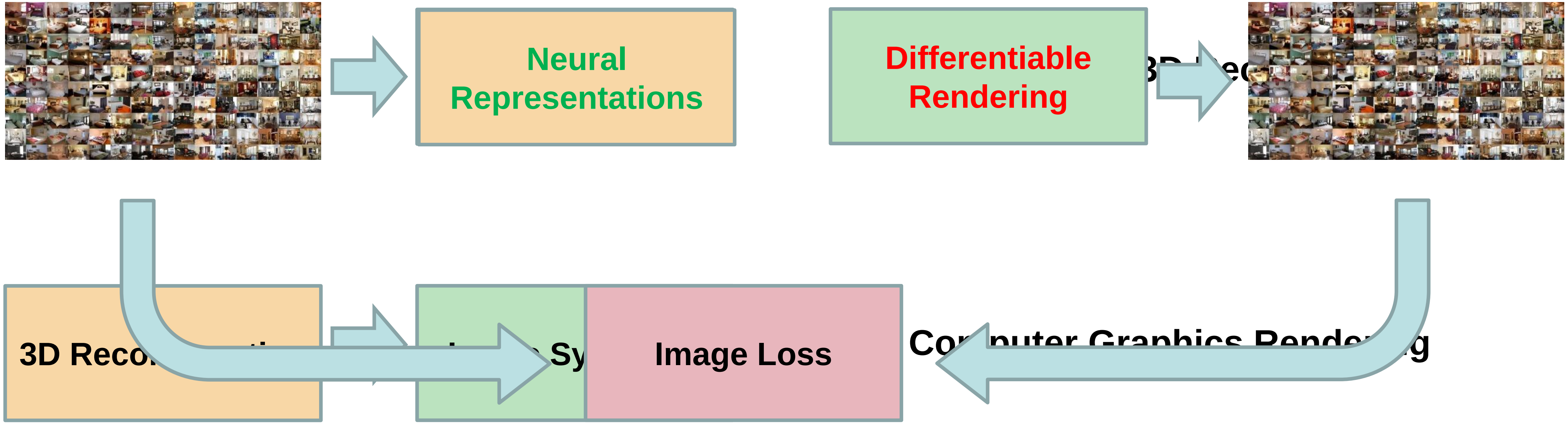
Challenges



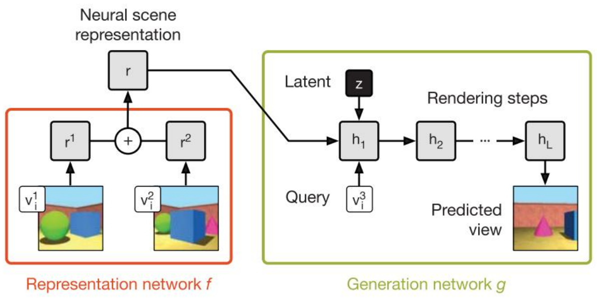
VS

Self-supervised Learning of 3D Scenes

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

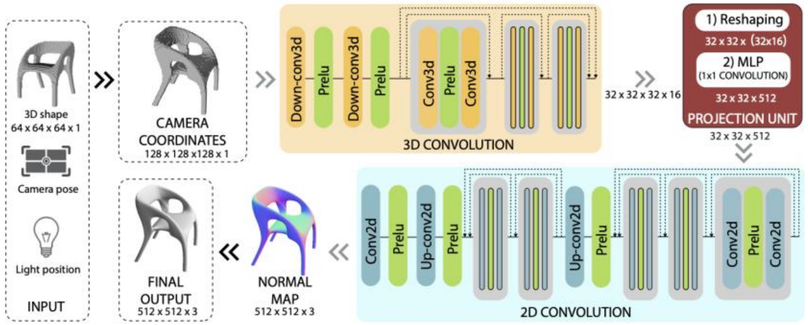


Neural 3D Scene Representations

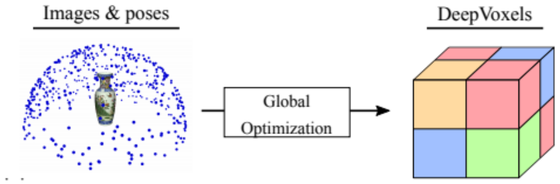


Generative Query Networks
[Eslami et al. 2018]

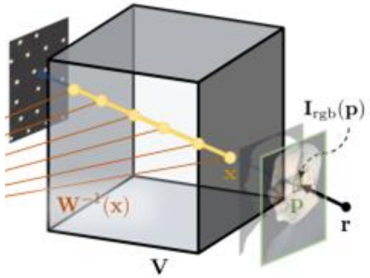
[Flynn et al., 2016; Zhou et al., 2018b; Mildenhall et al. 2019]
Multiplane Images (MPIs)



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



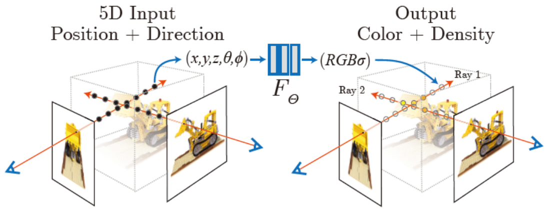
DeepVoxels
[Sitzmann et al. 2019]



Neural Volumes
[Lombardi et al. 2019]

Voxel Grids + Ray Marching

SRN [Sitzmann et al. 2019b]

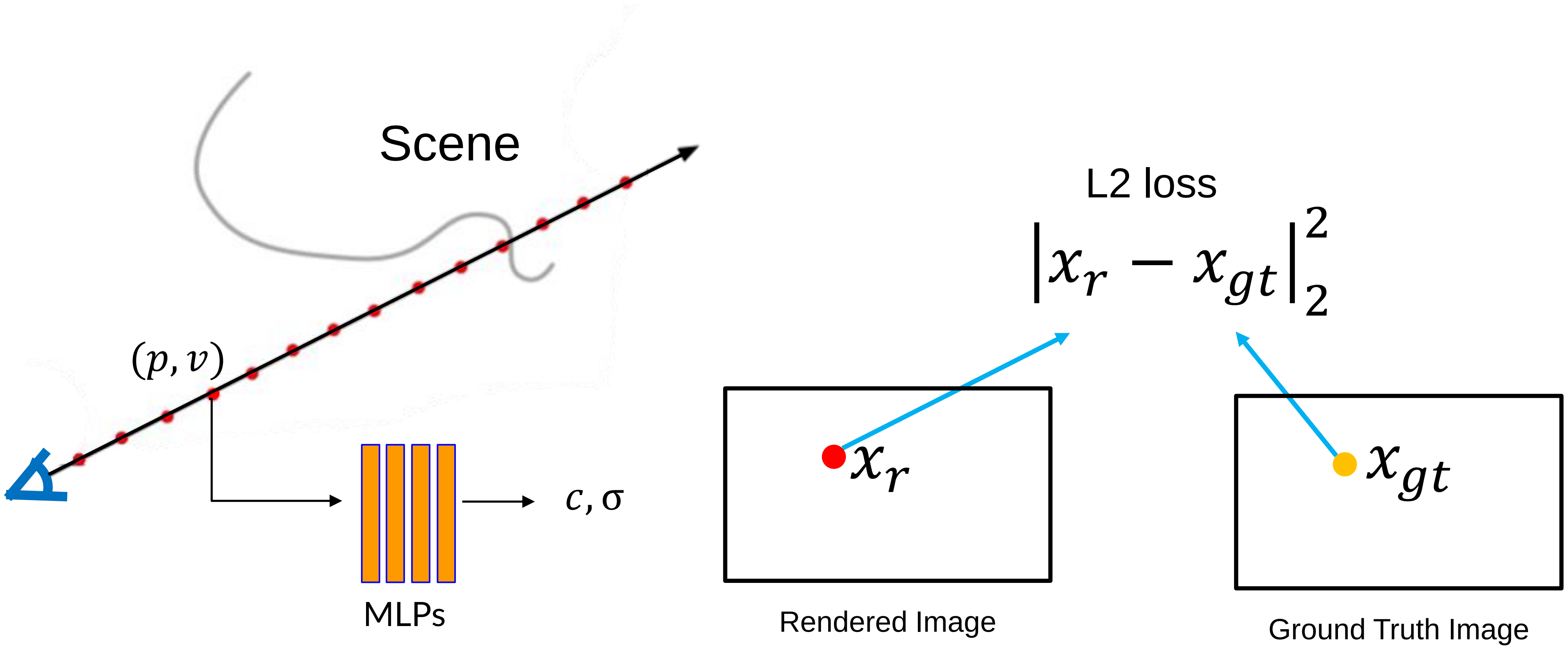


NeRF [Mildenhall et al. 2020]

IDR [Yariv et al. 2020]

Implicit Fields

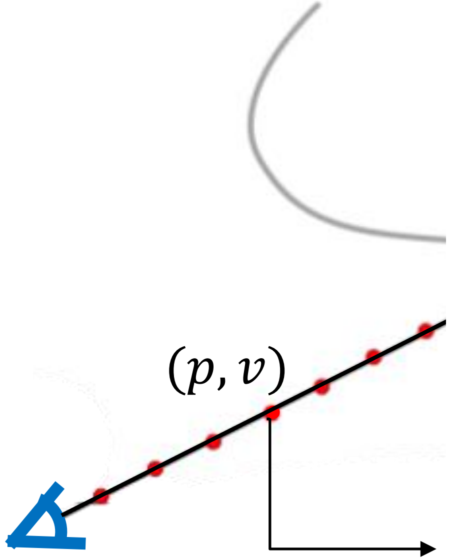
Neural Radiance Fields (NeRF)



[Mildenhall et al. 2020]

Neural Radiance Fields (NeRF)

- NeRF suffers from a slow rendering process.



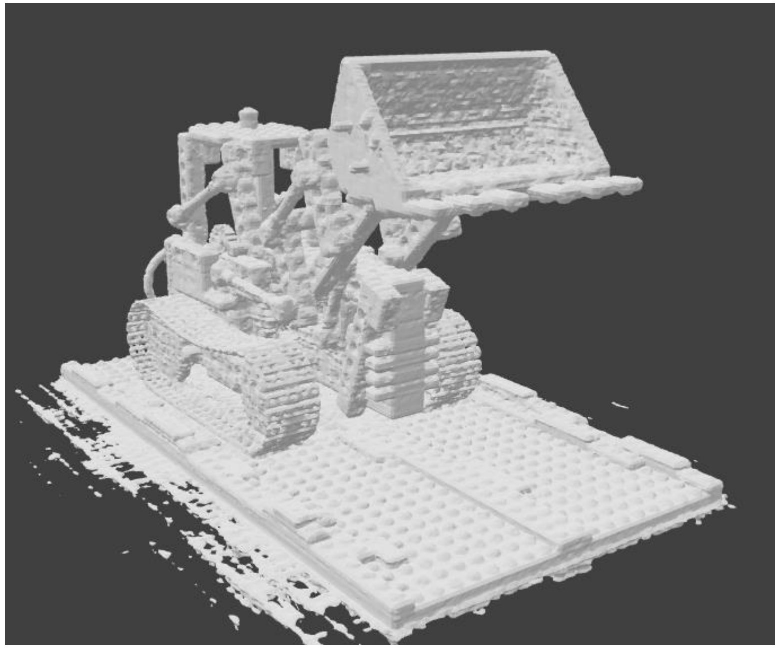
Rendering speed: 100 s/frame

[Mildenhall et al. 2020]

Image resolution: 1920x1080

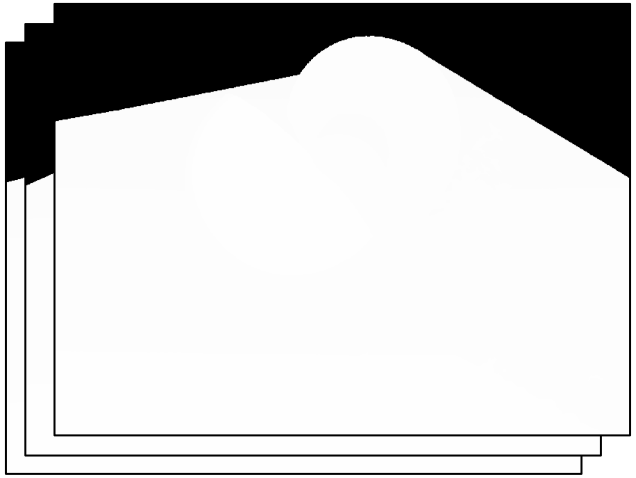
Surfaces Extracted from Learned Representation

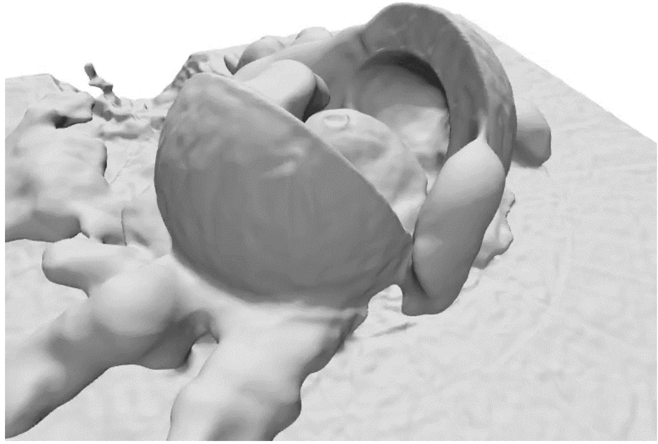
Scene Representations



Volume density used as scene representation lacks surface constraints

Background – Surface Reconstruction Methods





Images

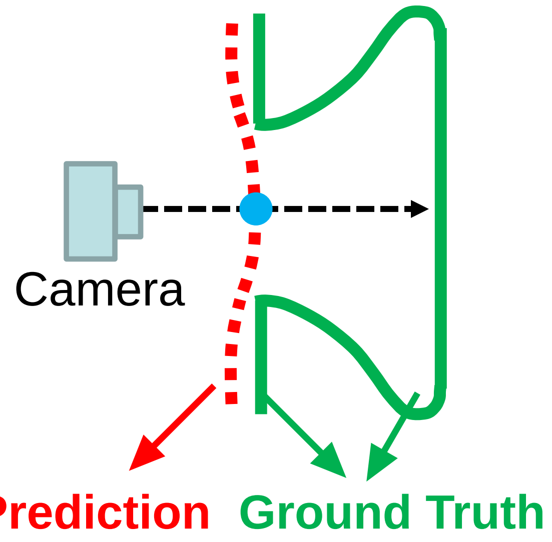
Masks

Supervise
Signed
Distance Fields
(SDF)

Geometry
(Represented as SDF)

IDR [Yariv et al. 2020]

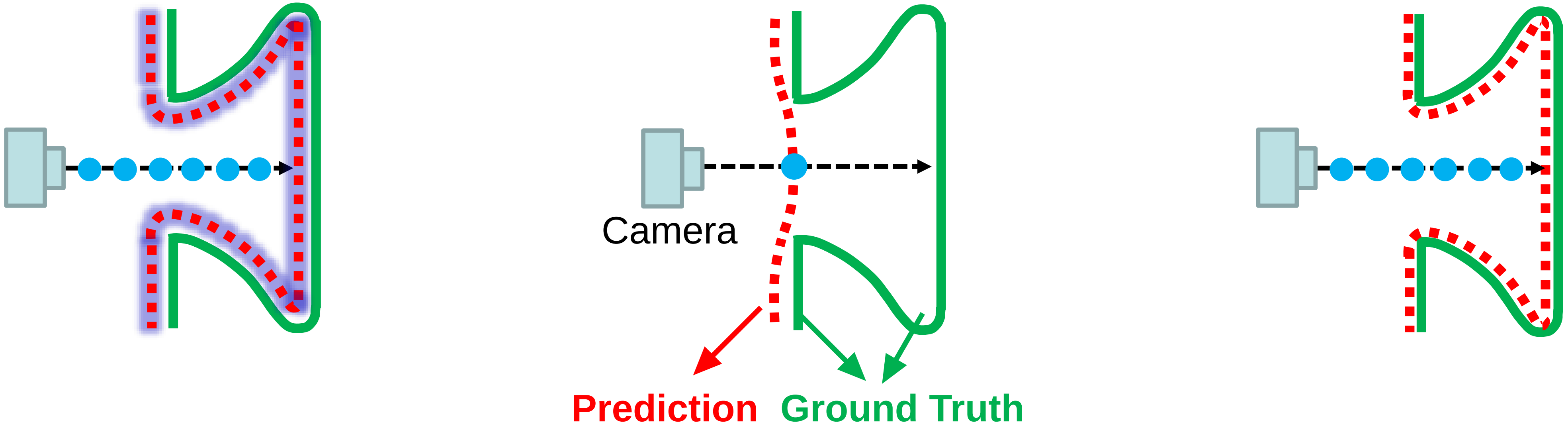
Background – Surface Reconstruction Methods



Surface Representation
+ Surface Rendering

Surface rendering is not suitable for learning scene representation

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

Challenge

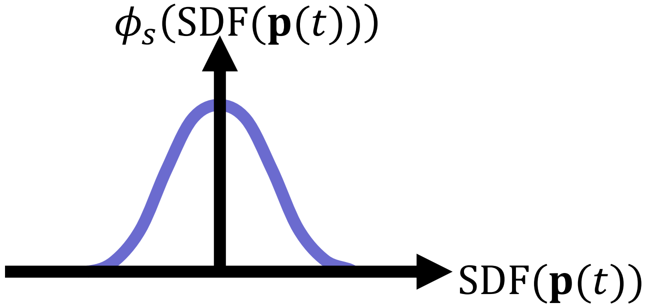
- Simply applying volume rendering to the density associated with SDF.

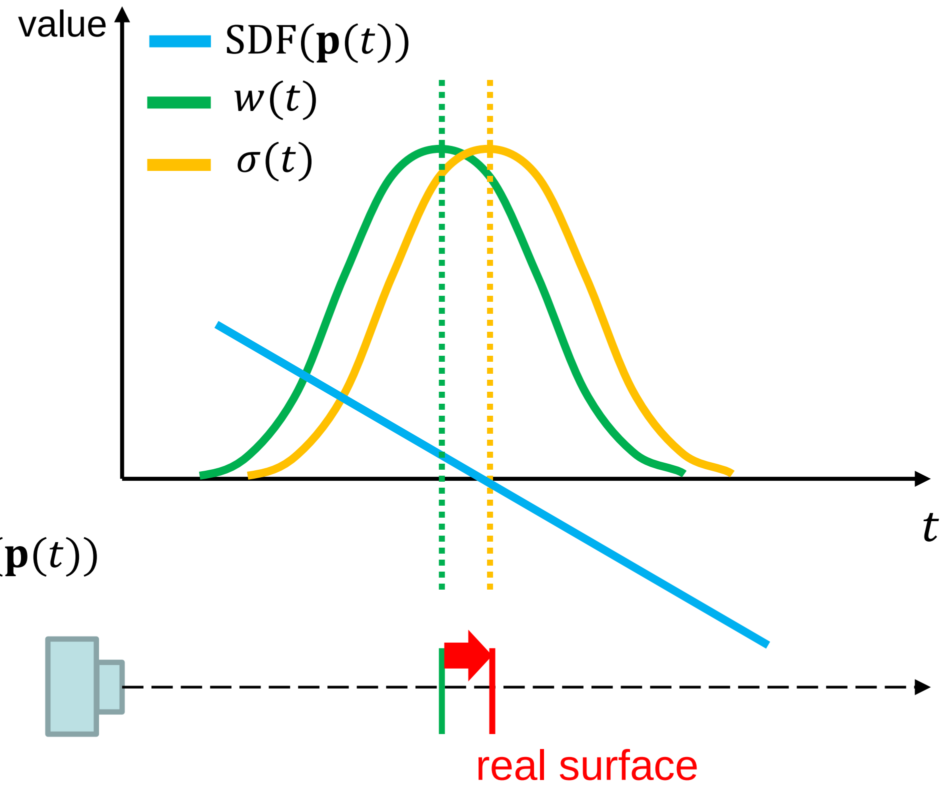
➢ $C = \int_0^{+\infty} w(t)c(t)\mathrm{d}t$

➢ $w(t) = T(t)\sigma(t)$

➢ $T(t) = e^{-\int_0^t \sigma(u)\mathrm{d}u}$

➢ $\sigma(t) = \phi_s(\mathrm{SDF}(\mathbf{p}(t)))$





- Problem: Introduce bias, i.e., the maximal weight term $w(t)$ is not attained on the surface intersection.

Our Solution

- Requirements on the weight function:

- Unbiased: w(t) attains a locally maximal value at a surface intersection point $\mathbf{p}(t^*)$, i.e. with $f(\mathbf{p}(t^*)) = 0$

- Occlusion-aware: Given any two depth values t_0 and t_1 satisfying $f(\mathbf{p}(t_0)) = f(\mathbf{p}(t_1))$, $w(t_0) > 0$, $w(t_1) > 0$, and $t_0 < t_1$, there is $w(t_0) > w(t_1)$.

Our Solution

- We first introduce a straightforward way to construct an unbiased weight function

$$w(t) = \frac{\phi_s(f(\mathbf{p}(t)))}{\int_0^{+\infty} \phi_s\left(f(\mathbf{p}(u))\right) \mathrm{d}u}$$

where $\phi_s(x) = se^{-sx}/(1+e^{-sx})^2$,

$f(\mathbf{p}(t))$ is the SDF value of point $\mathbf{p}(t)$

- However, this weight function is not occlusion-aware.

Our Solution

- We design a weight function that is both occlusion-aware and unbiased in the first order approximation of SDF by combining the following two equations.

Occlusion-aware but *NOT* unbiased

- $w(t) = T(t)\sigma(t)$
- $T(t) = e^{-\int_0^t \sigma(u)\mathrm{d}u}$
- $\sigma(t) = \phi(f(\mathbf{p}(t)))$

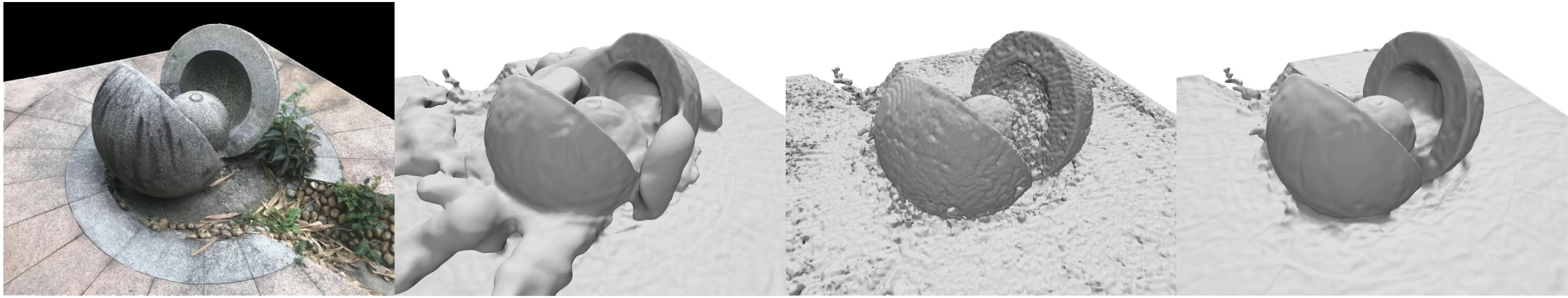
Unbiased but *NOT* occlusion-aware

$$w(t) = \frac{\phi_s(f(\mathbf{p}(t)))}{\int_0^{+\infty} \phi_s\left(f\left(\mathbf{p}(u)\right)\right) \mathrm{d}u}$$

where $\phi_s(x) = se^{-sx}/(1 + e^{-sx})^2$,

$f(\mathbf{p}(t))$ is the SDF value of point $\mathbf{p}(t)$

Comparison



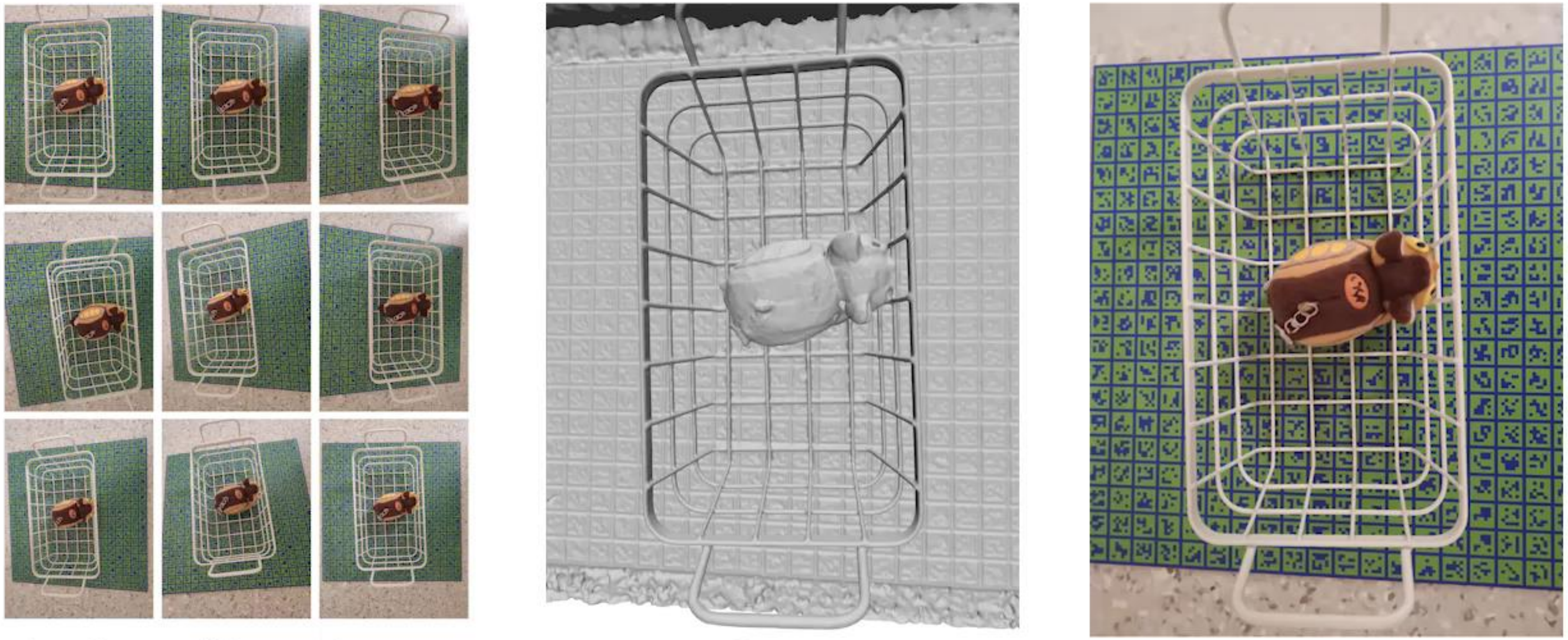
Reference Image

IDR
[Yariv et al. 2020]

NeRF
[Mildenhall et al. 2020]

NeuS
(Ours)

Results of NeuS

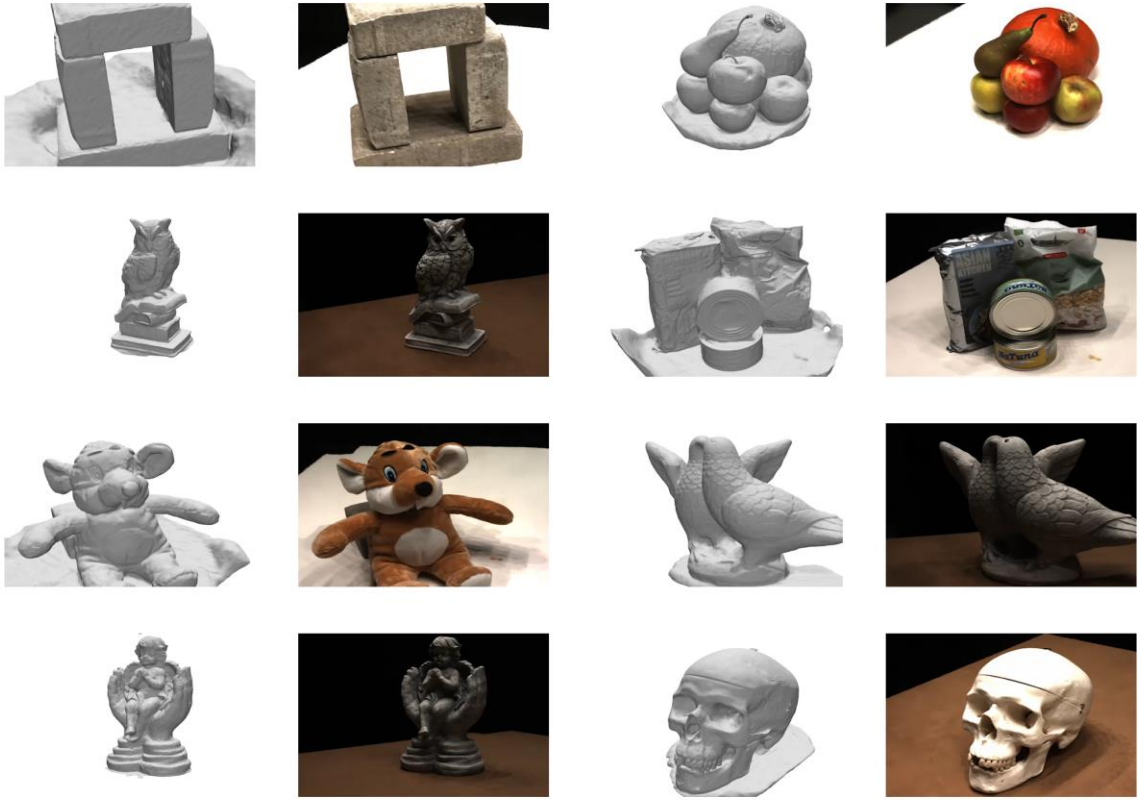


A subset of input images

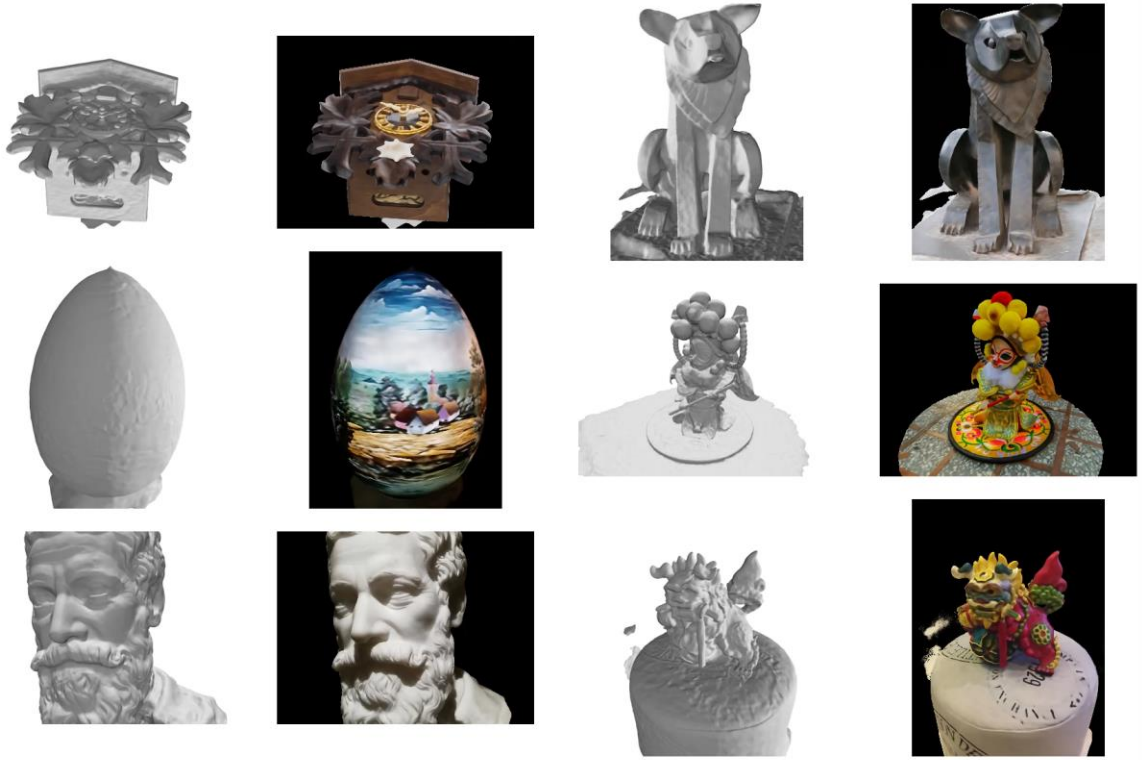
Our surface geometry (w/o mask supervision)

Our rendering (w/o mask supervision)

Results of NeuS

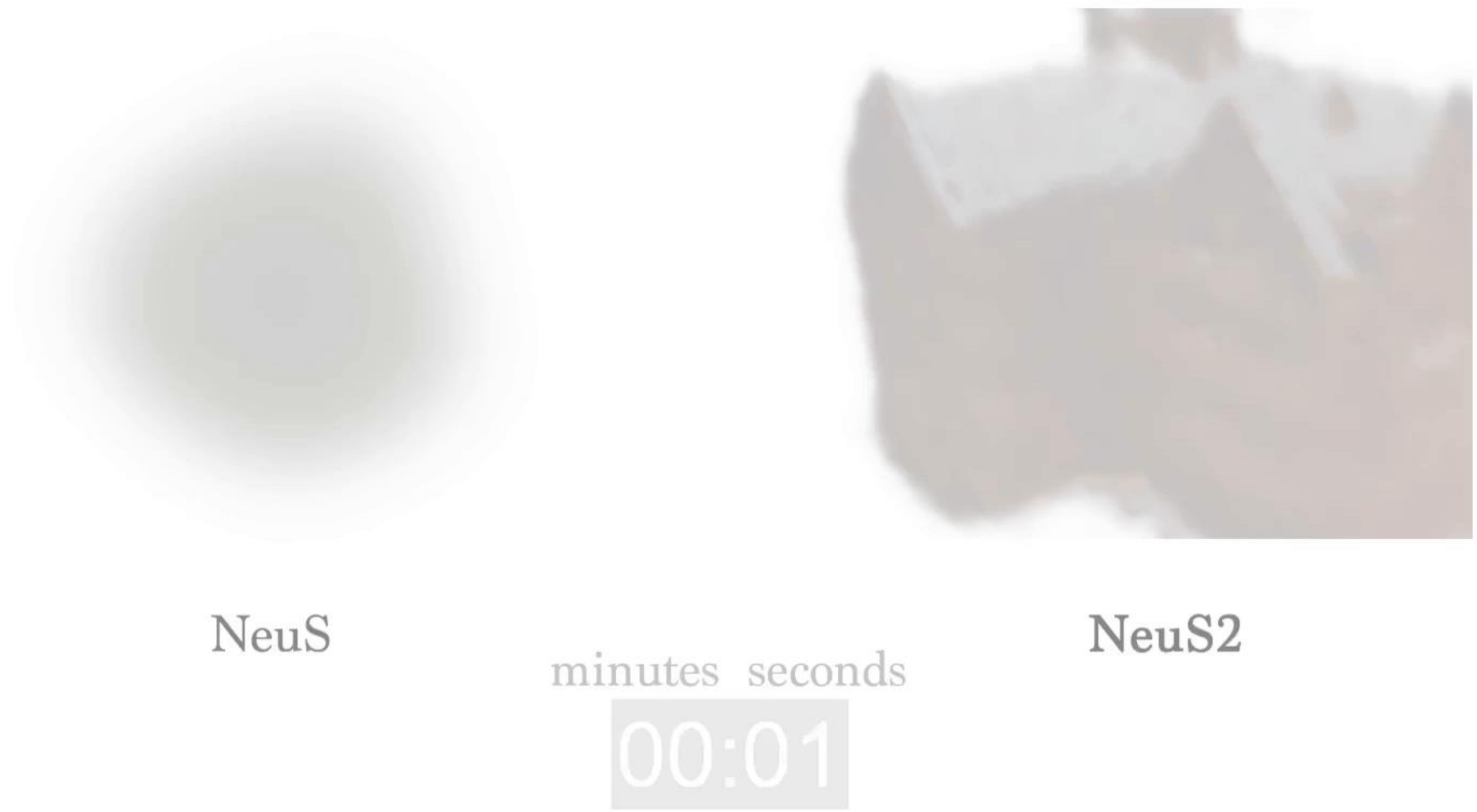


DTU dataset

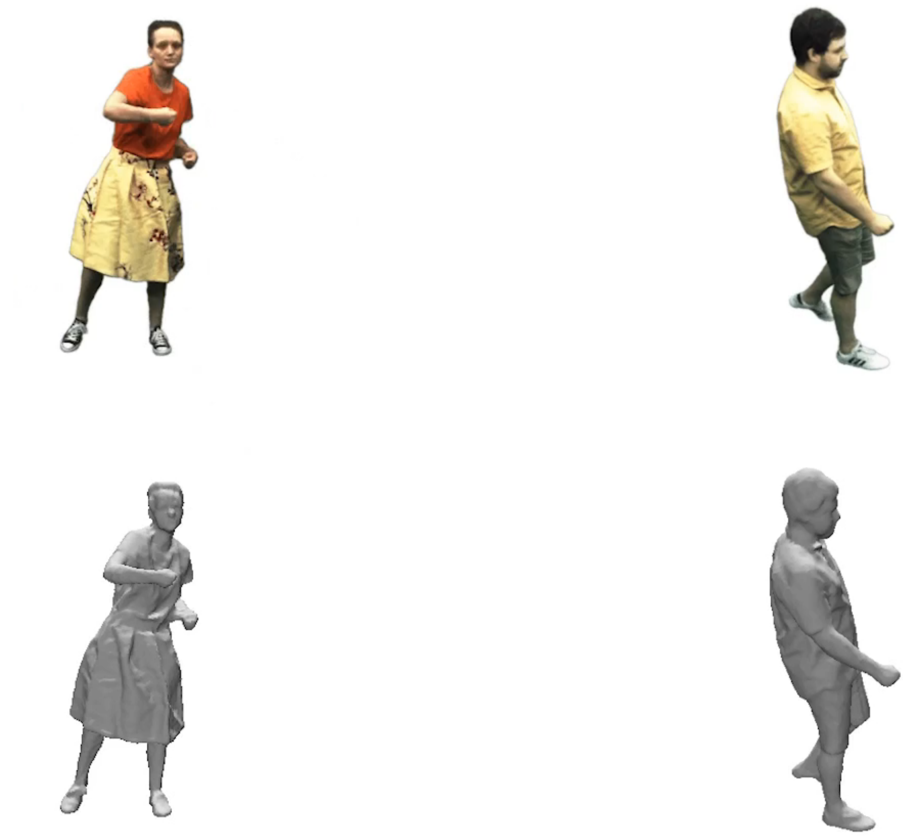


BlendedMVS dataset

Fast Training of NeuS

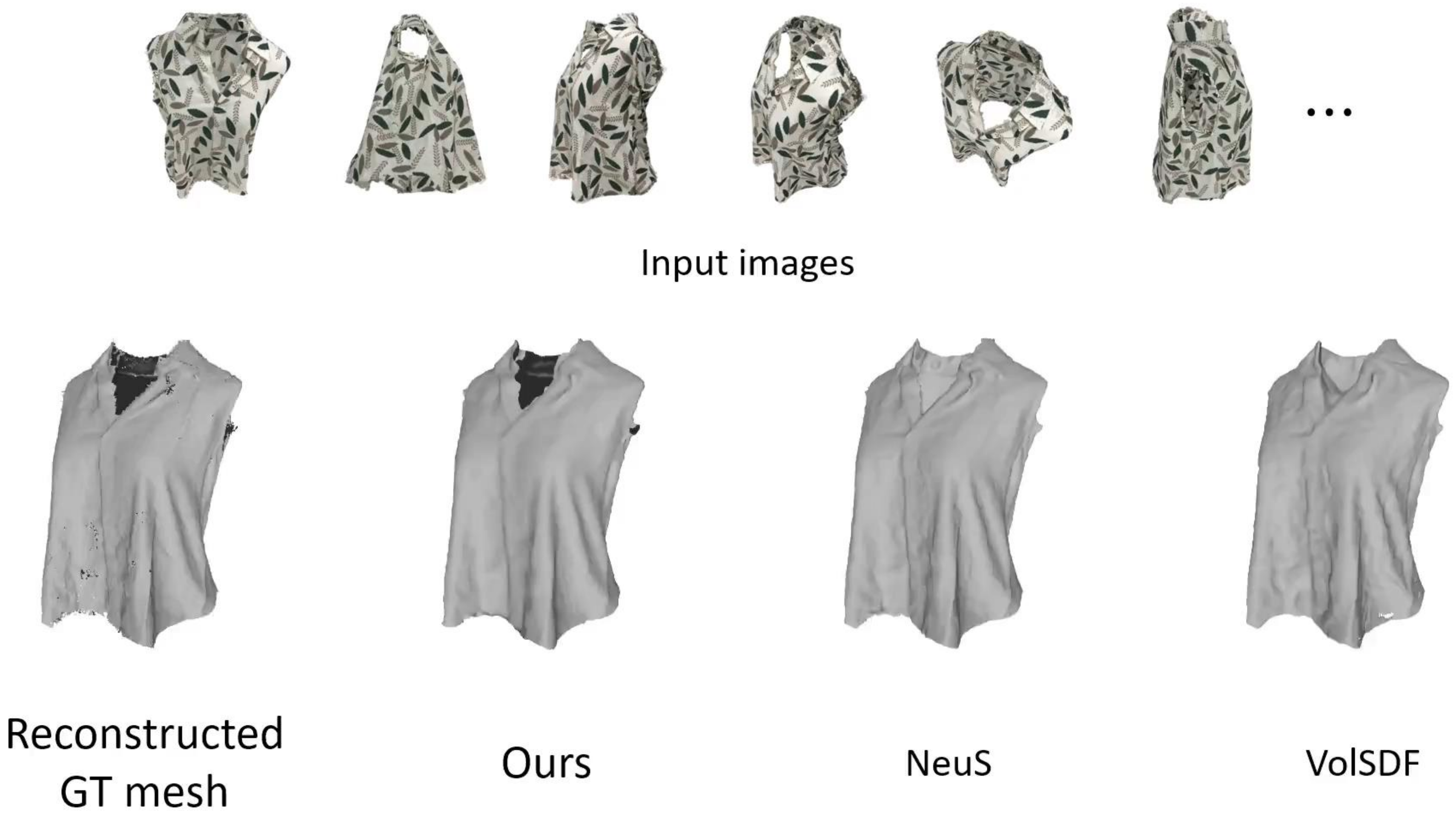


Fast Training of NeuS

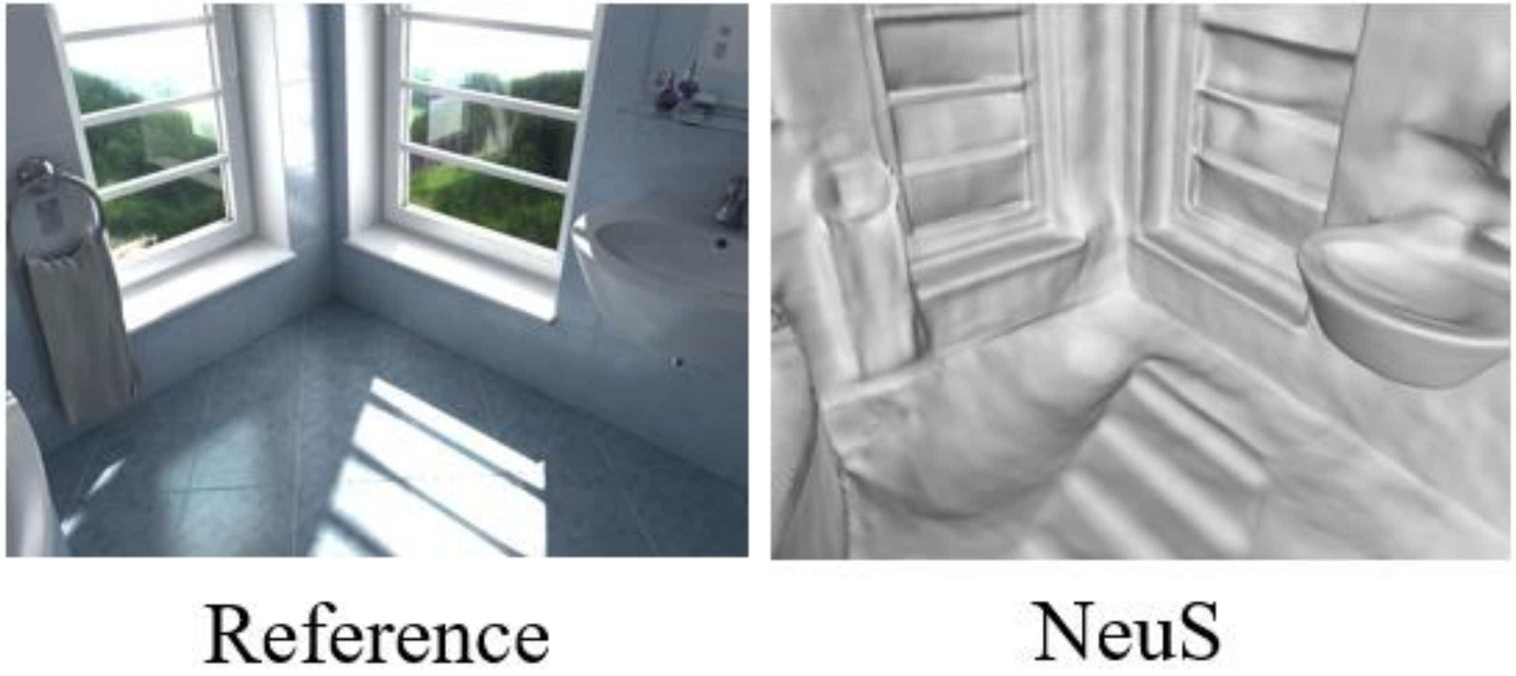


Dynamic Scenes (20 seconds per frame)

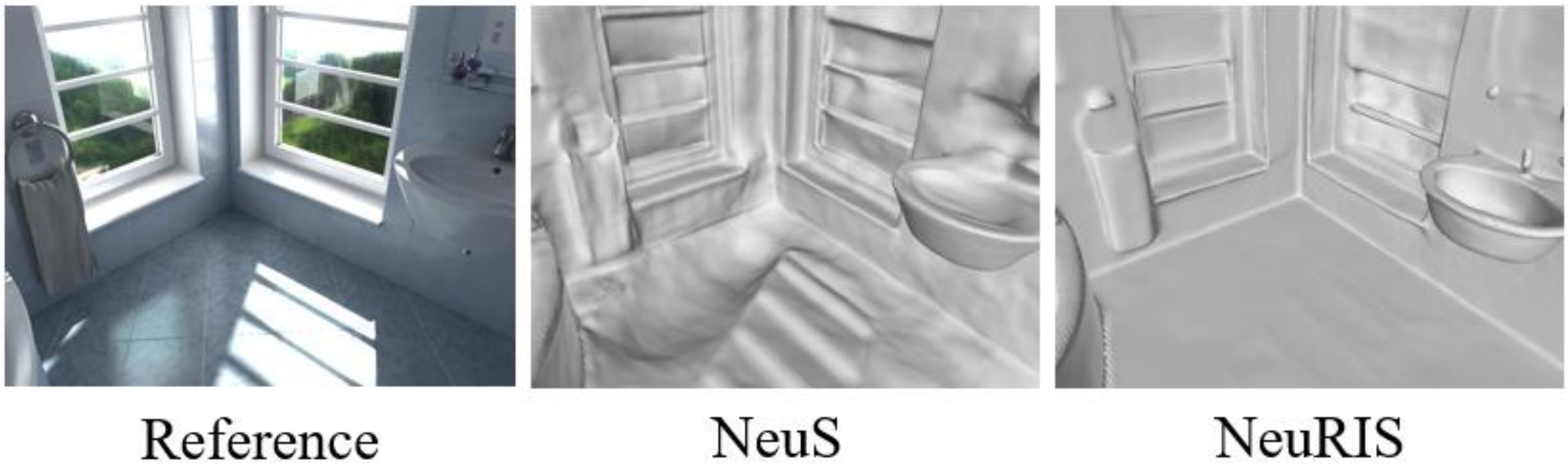
Neural SDF -> UDF



Indoor Scene Reconstruction



Indoor Scene Reconstruction

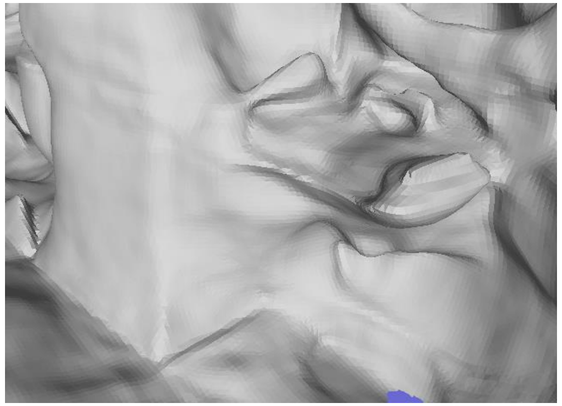


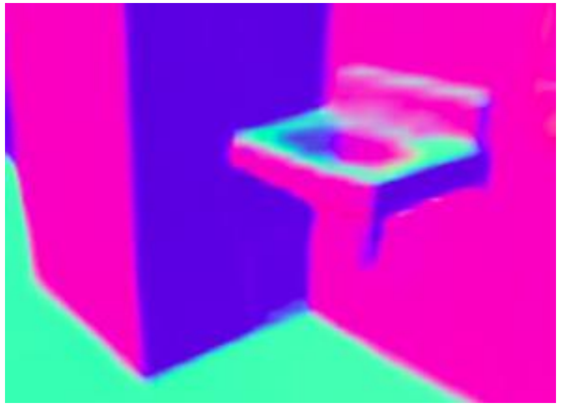
J. Wang, P. Wang, X. Long, C. Theobalt, T. Komura, ***L. Liu,*** W. Wang. NeuRIS: Neural Reconstruction of Indoor Scenes Using Normal Priors, ECCV 2022

Method

- **Normal priors**
- Invariant to translation and scaling, avoiding the scale ambiguity issue of depth estimation
- Encode geometry information of common indoor scenes


Reference

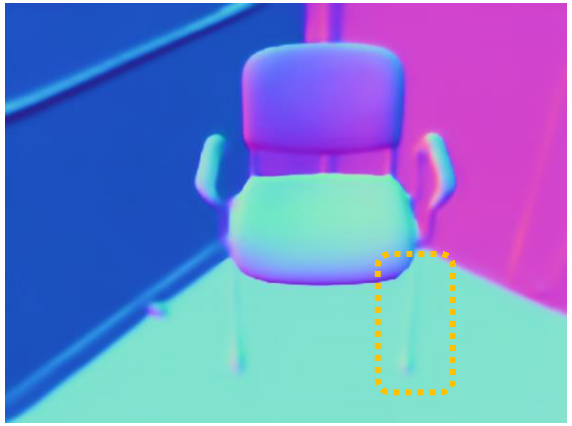
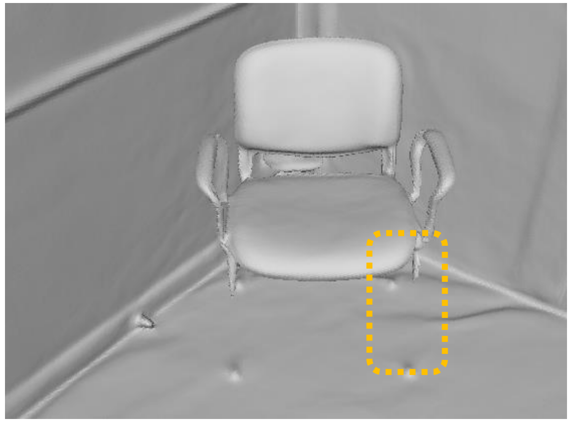
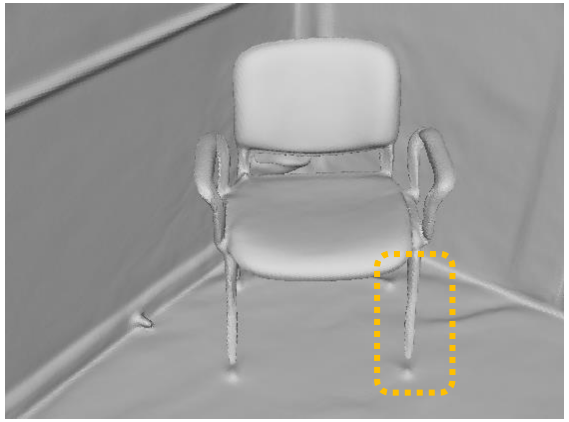

NeuS


Estimated normal


NeuS with normal priors

Method

- **Online geometry check**
- Normal estimation at object edges or thin structure areas is usually not accurate and may not produce correct geometry. Avoid using normal priors at those areas.

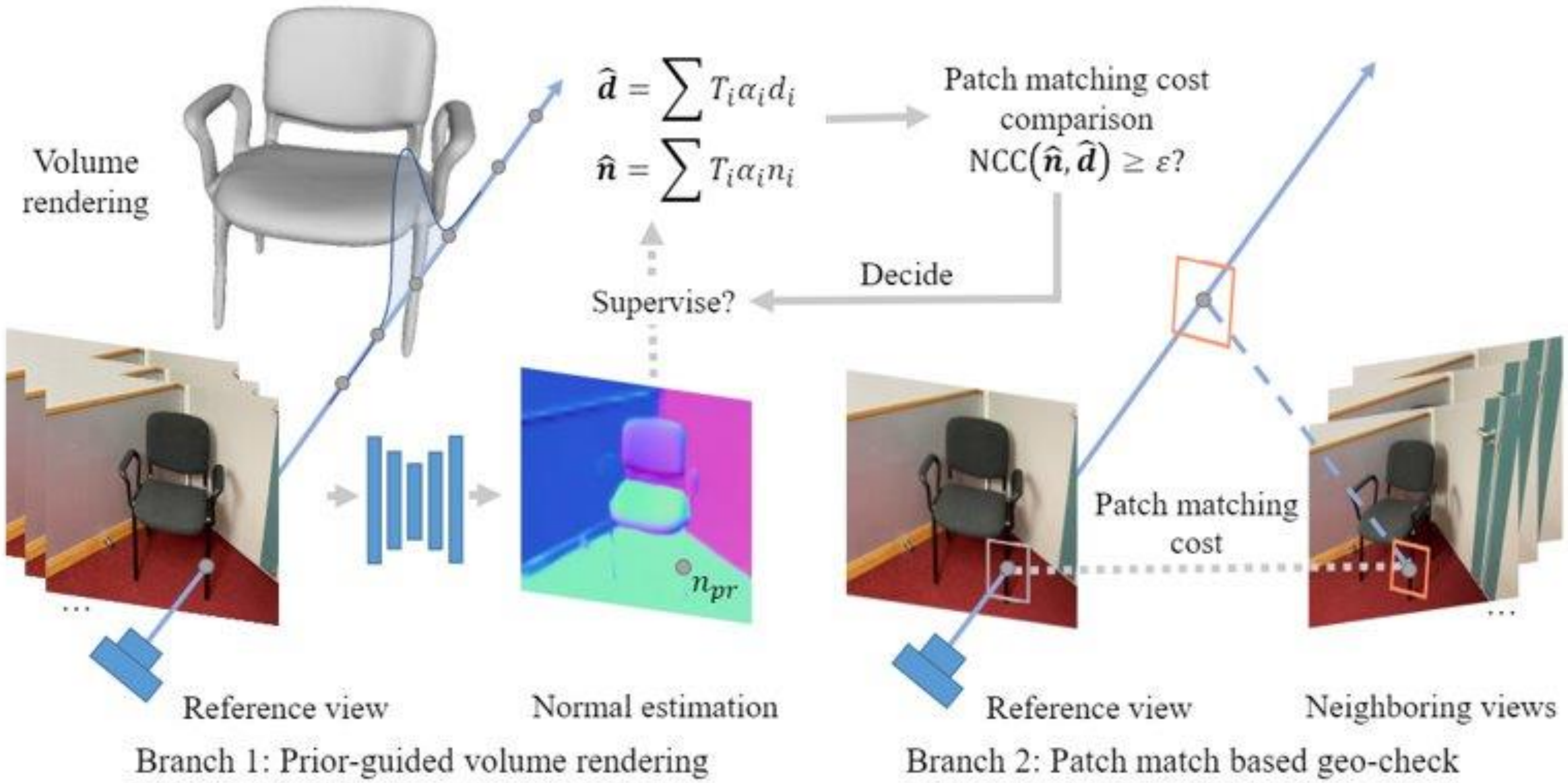





Reference | Estimated normal | NeuS with normal priors | NeuRIS (ours)

Method

Neural volume rendering using normal priors adaptively

(1) Use normal priors to guide the optimization process
(2) Check multi-view consistency to decide whether to use normal priors on the fly



NeuRIS: Neural Reconstruction of Indoor Scenes Using Normal Priors (ECCV’22)

Reference | Estimated normal | Reconstucted mesh

Learn an Animatable Human Model from Multi-view Videos

View 1

View 2

View 3

. . .

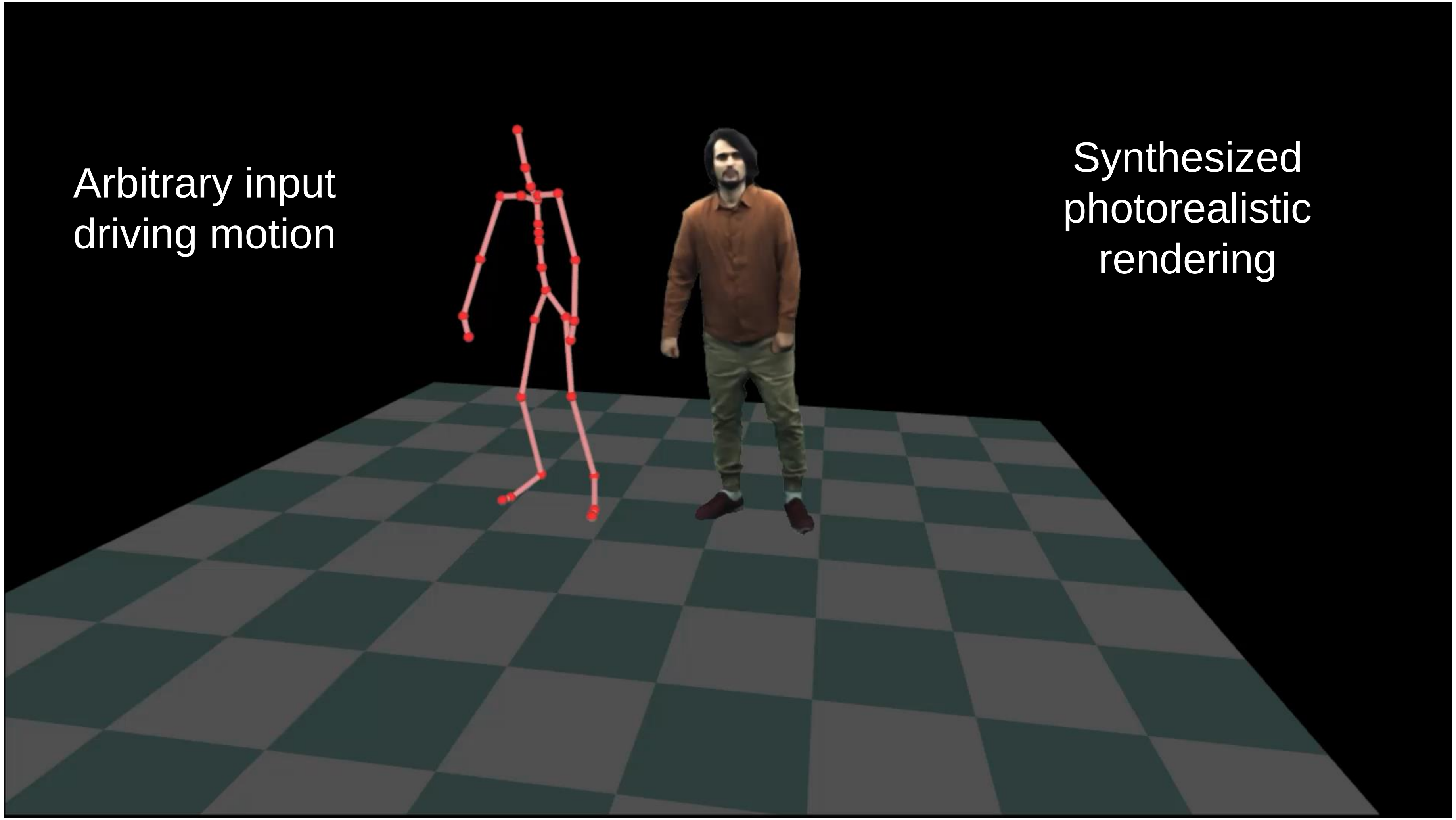
Multi-view Videos
(w/ camera parameters)

Animatable Human Representation

Learn Animatable Explicit Human Model from Multi-view Videos

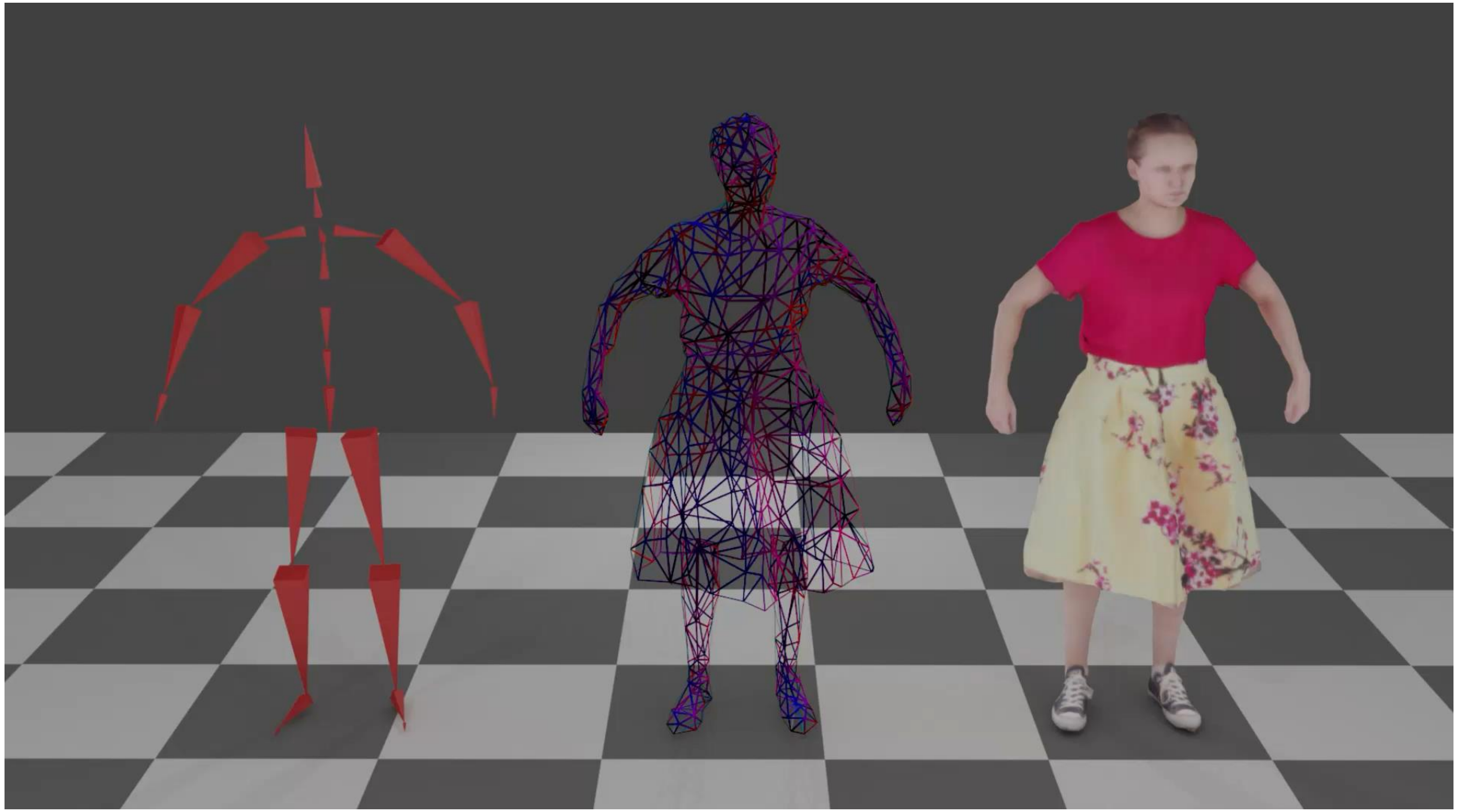
M. Habermann, ***L. Liu***, W. Xu, M. Zollhoefer, G. Pons-Moll, C. Theobalt.
Real-time Deep Dynamic Characters. SIGGRAPH 2021

Learn Animatable Explicit Human Model from Multi-view Videos



M. Habermann, ***L. Liu*,** W. Xu, M. Zollhoefer, G. Pons-Moll, C. Theobalt.
Real-time Deep Dynamic Characters. SIGGRAPH 2021

Require a Person-specific Scanned 3D Human Template

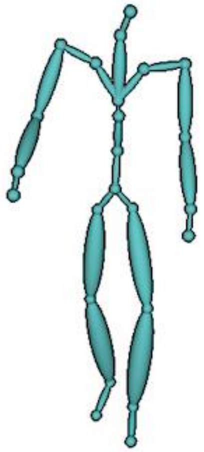


Meshes Have Limited Resolution

DDC

Ground Truth

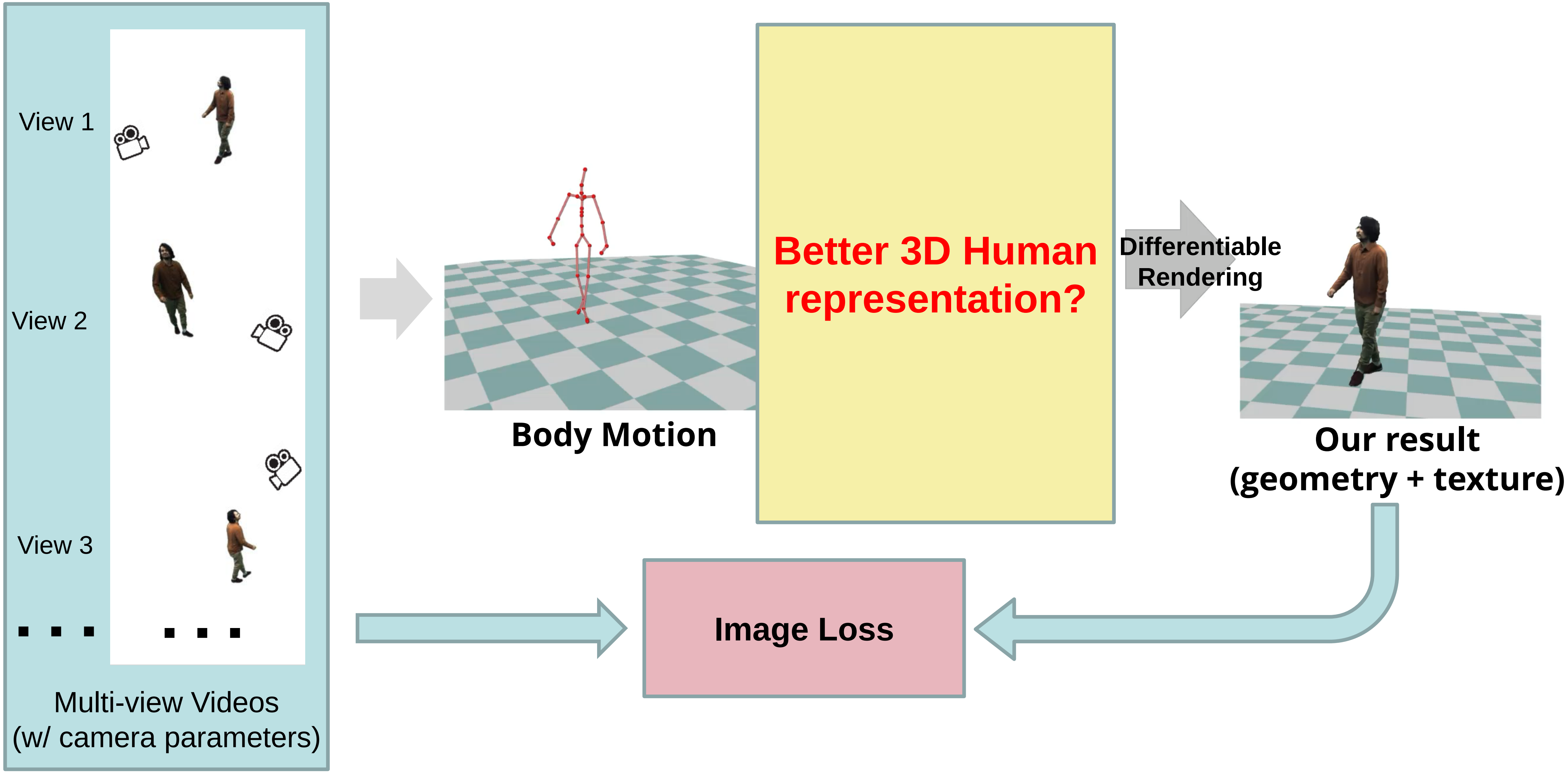
Neural Actor: Neural Free-view Synthesis of Human Actors with Pose Control



Arbitrary input driving poses

Synthesized results by Neural Actor

L. Liu, M. Habermann, V. Rudnev, K. Sarkar, J. Gu, C. Theobalt.
Neural Actor: Neural Free-view Synthesis of Human Actors with Pose Control, SIGGRAPH Asia 2021

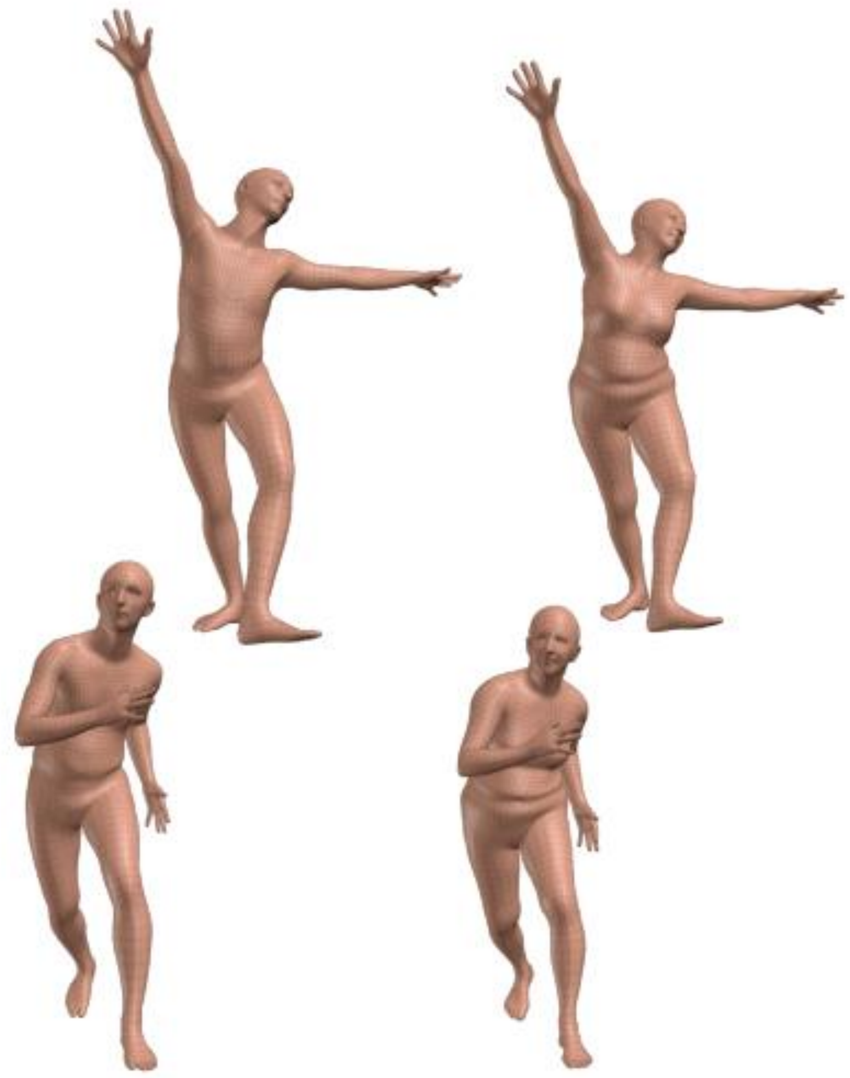

View 1
View 2
View 3
. . .
. . .
Multi-view Videos
(w/ camera parameters)
Body Motion
Better 3D Human representation?
Differentiable Rendering
Our result
(geometry + texture)
Image Loss

-specific

/ Animatable

l resolution

Better 3D Human representation?

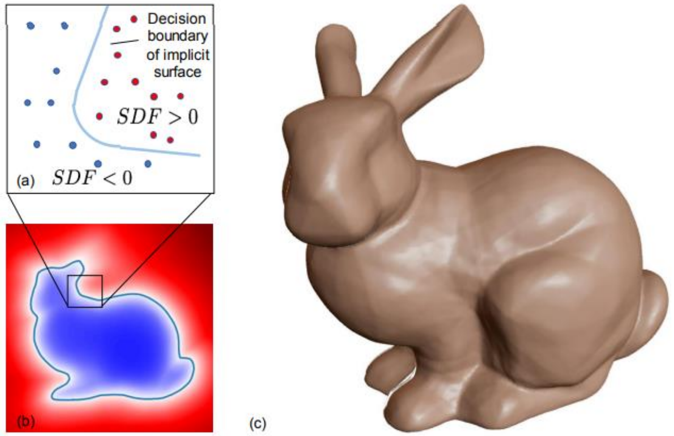


Not person-specific

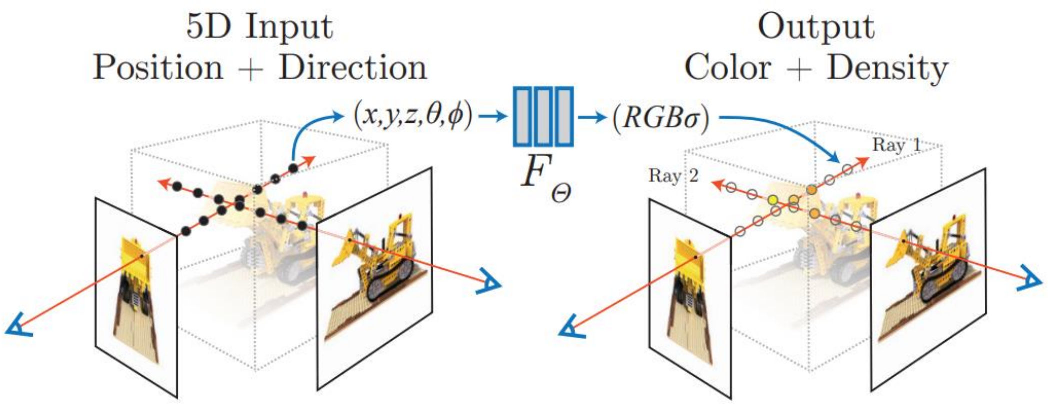
Articulated / Animatable

~~High spatial resolution~~ **No clothes**

Skinned Multi-person Linear Model (SMPL)



[Park et al. 2019]



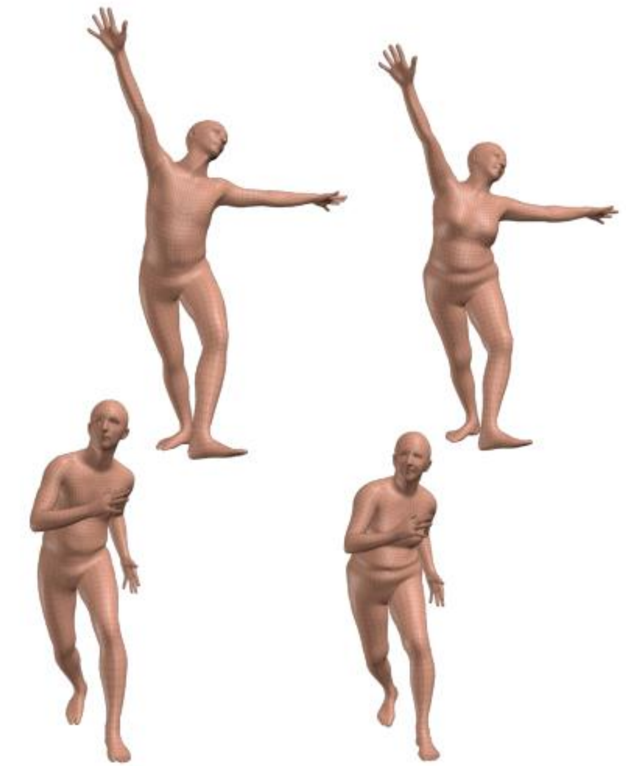
[Mildenhall et al. 2020]

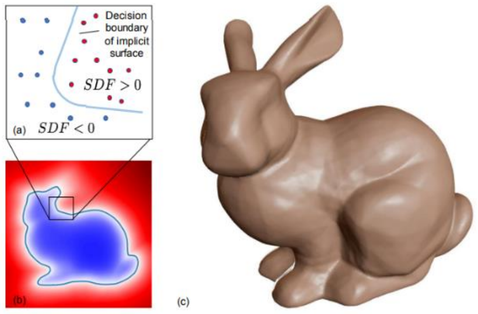
Neural Fields

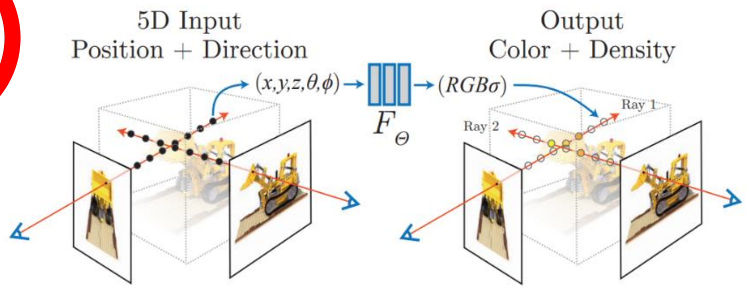
~~Not person-specific~~

~~Articulated / Animatable~~

High spatial resolution





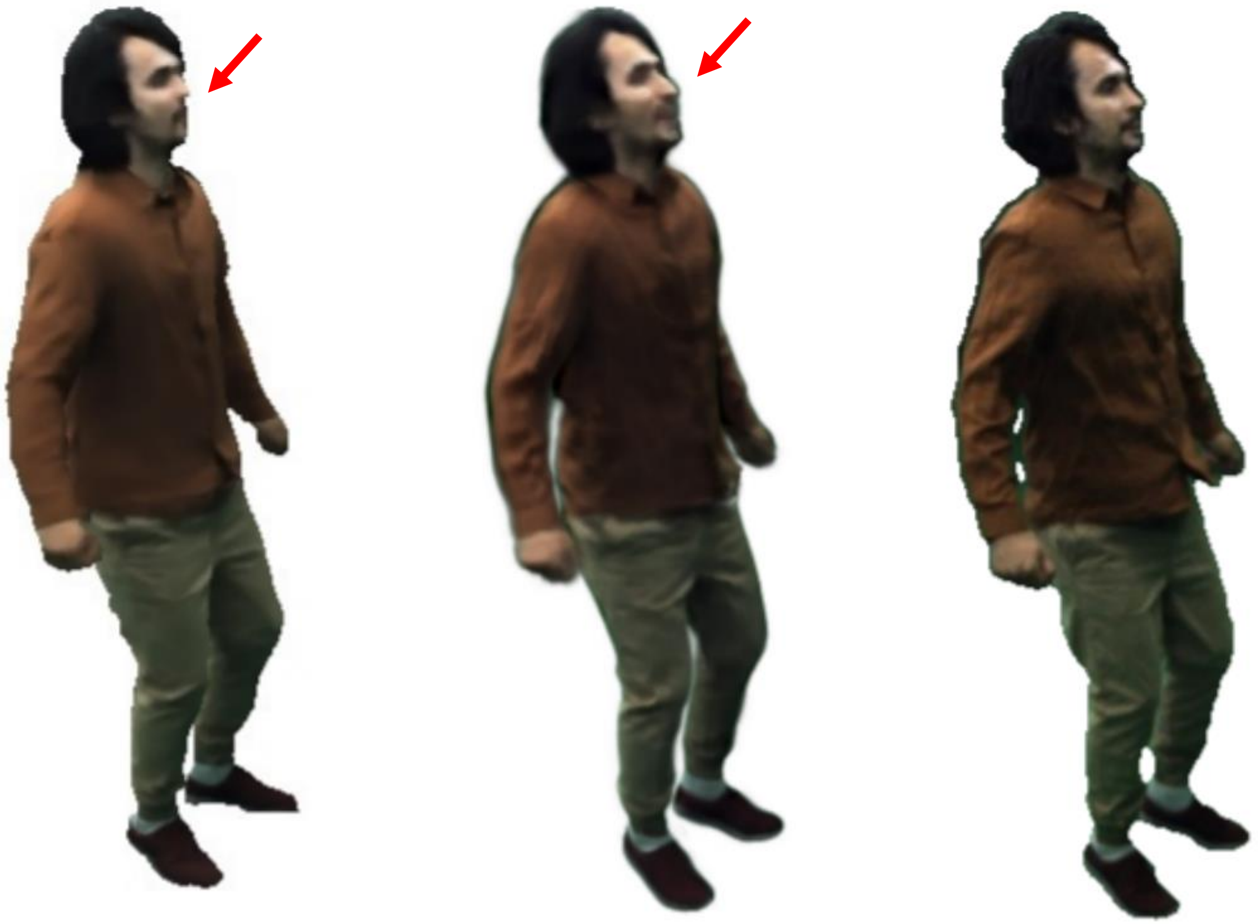


Skinned Multi-person Linear Model (SMPL)

Neural Fields

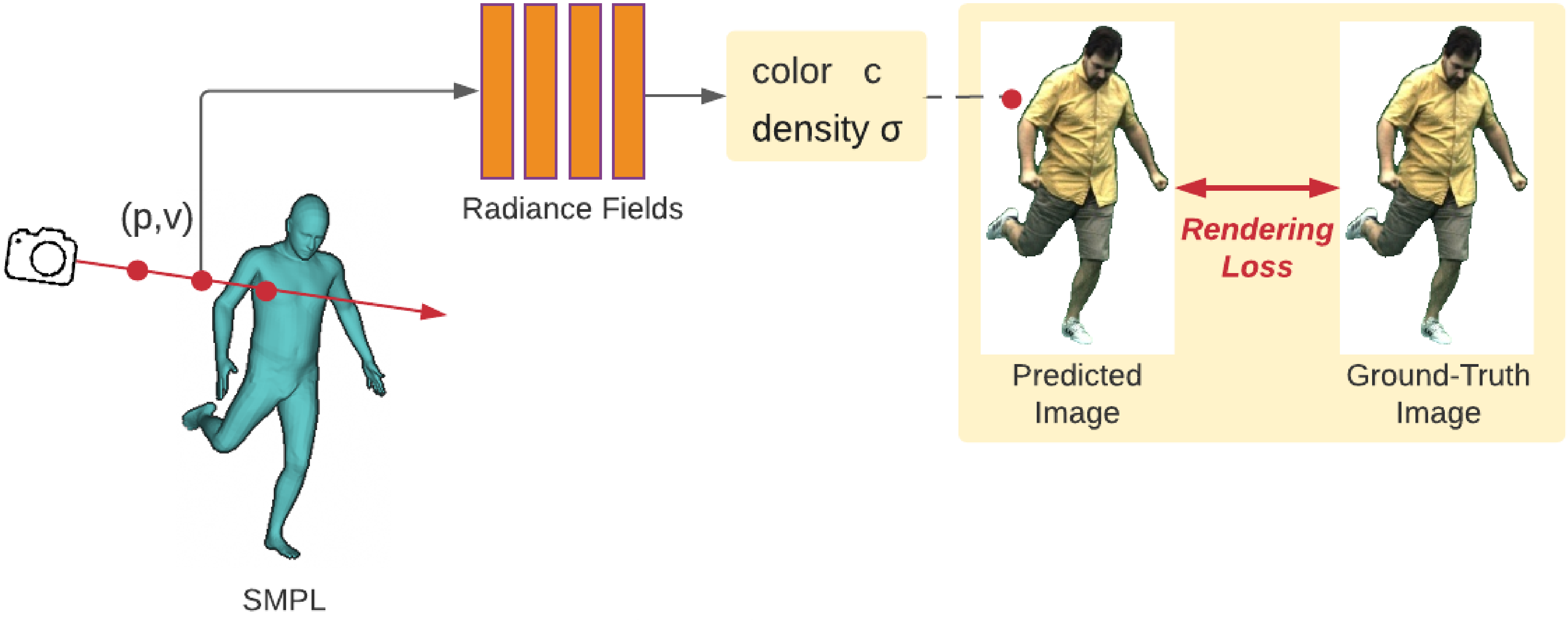
L. Liu, M. Habermann, V. Rudnev, K. Sarkar, J. Gu, C. Theobalt.
Neural Actor: Neural Free-view Synthesis of Human Actors with Pose Control, SIGGRAPH Asia 2021

Result of Neural Actor



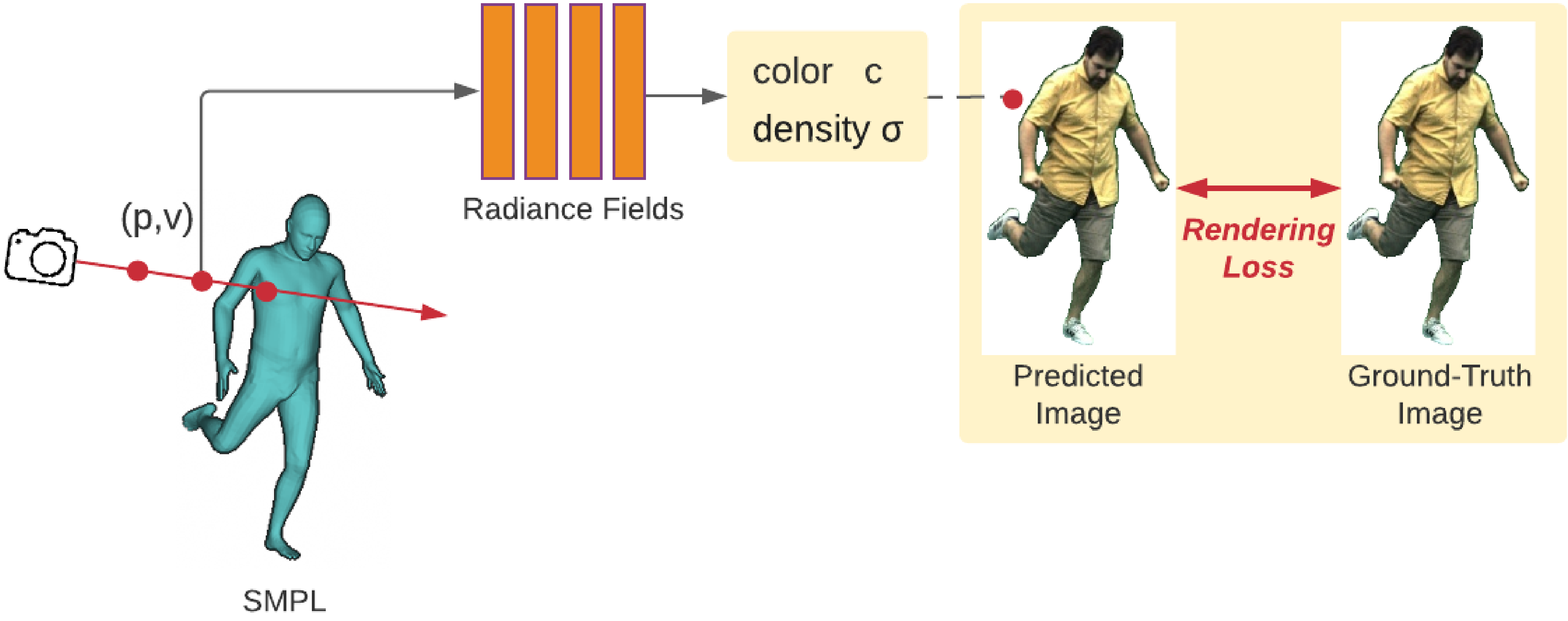
DDC   Neural Actor   Ground Truth

Model a Single Frame as NeRF

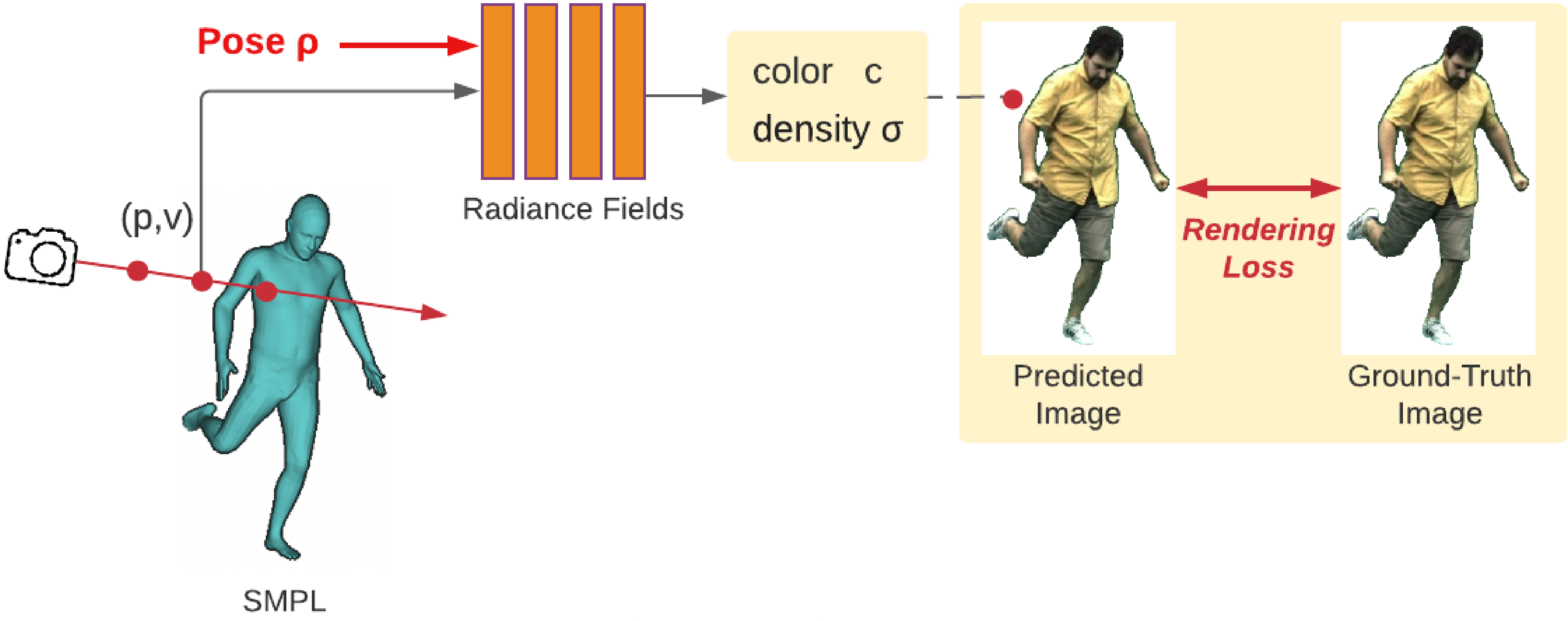


$$\Phi: (p, v) \rightarrow (c, \sigma)$$

How to Model a Moving Sequence of Human?



How to Model a Moving Sequence of Human?



$$\Phi: (p, v, \rho) \rightarrow (c, \sigma)$$

Pose as Conditioning

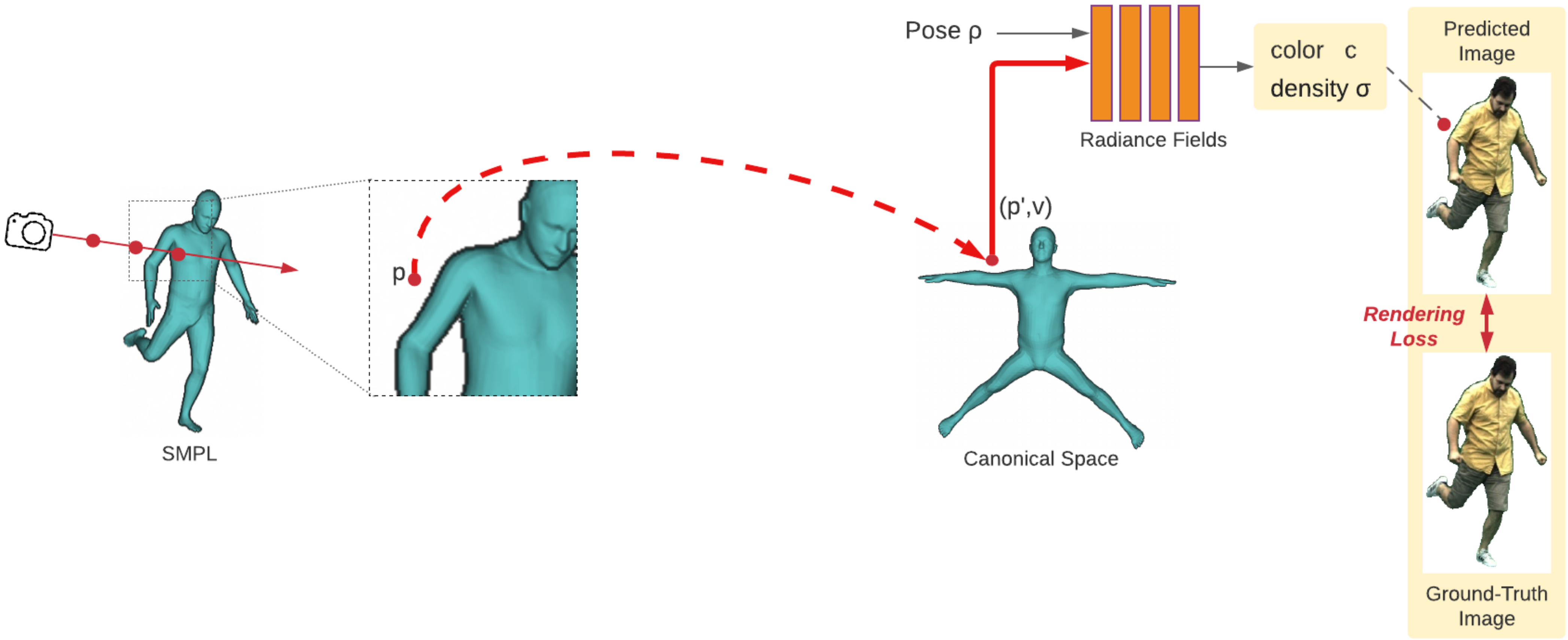
Pose as Conditioning

Ground Truth

What is the efficient way to model a moving sequence?

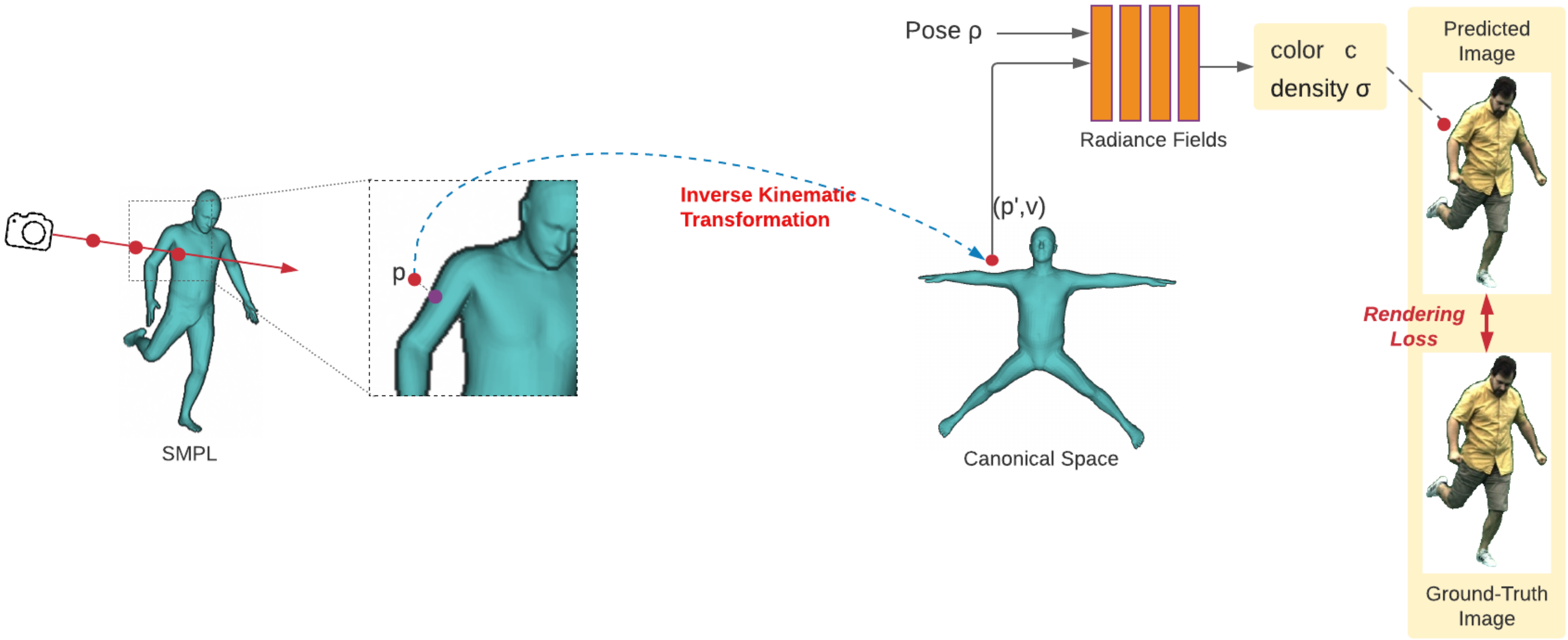
Shape and appearance of the person in each frame would not change much!

Geometry-guided Deformable Neural Fields

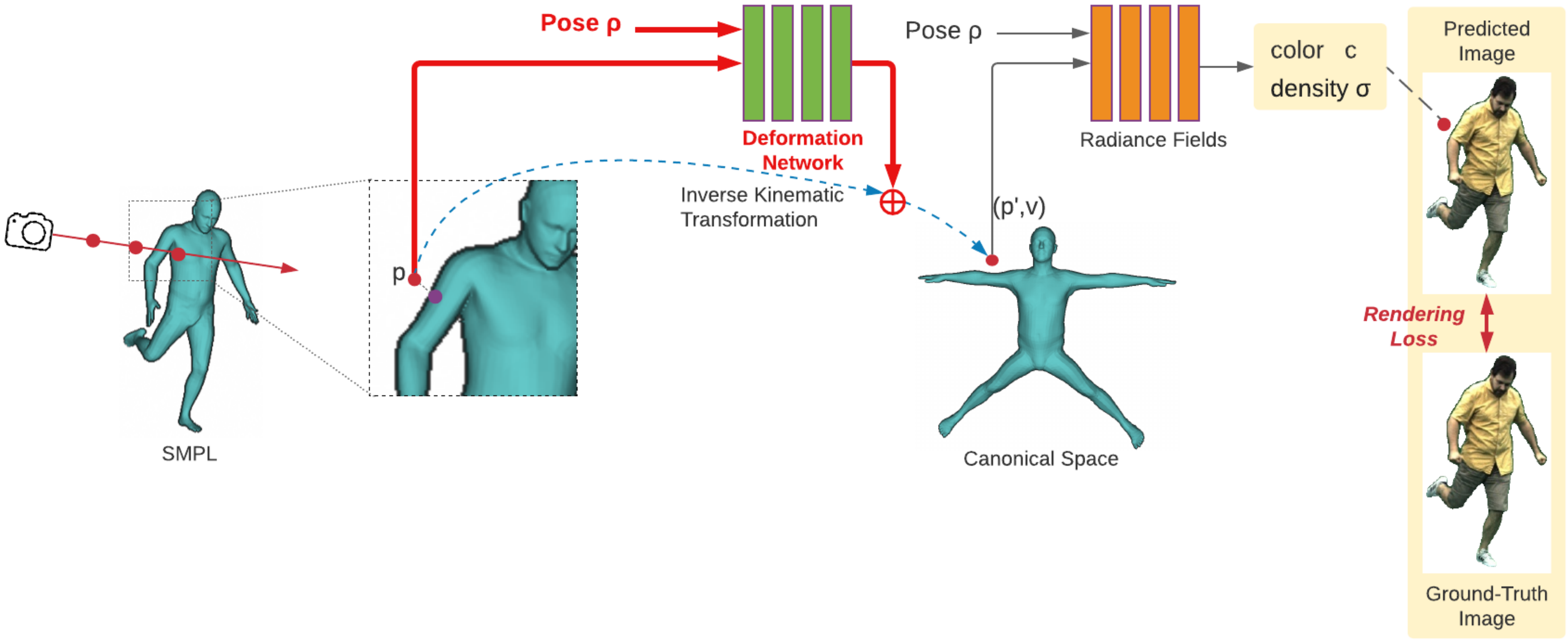


L. Liu, M. Habermann, V. Rudnev, K. Sarkar, J. Gu, C. Theobalt.
Neural Actor: Neural Free-view Synthesis of Human Actors with Pose Control, SIGGRAPH Asia 2021

Geometry-guided Deformable Neural Fields



Geometry-guided Deformable Neural Fields



Geometry-guided Deformable Neural Fields

Pose as Conditioning

Geometry-guided Deformable Neural Fields (One Proposed Component of Neural Actor)

Ground Truth

What Causes Blurriness?

- The mapping from the skeletal pose to dynamic geometry and appearance is not a bijection.
  - Complex dynamics of the surface
  - Pose tracking errors
  - Cloth-body interaction

Pose $\longrightarrow$ **Geometry + Appearance**

Many-to-many mapping

Model this mapping with a deterministic model with L2 loss?

With adversarial loss?

What Causes Blurriness?

- The mapping from the skeletal pose to dynamic geometry and appearance is not a bijection.
  - Complex dynamics of the surface
  - Pose tracking errors
  - Cloth-body interaction

Incorporate a latent variable:

Pose → **Geometry + Appearance**

Latent Variable

One-to-one mapping

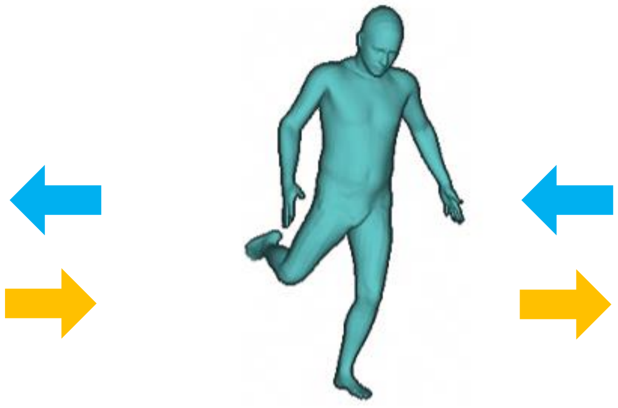
Many-to-many mapping

How to Choose Latent Variable?

Latent Variable ⟶ Geometry + Appearance

One-to-one mapping

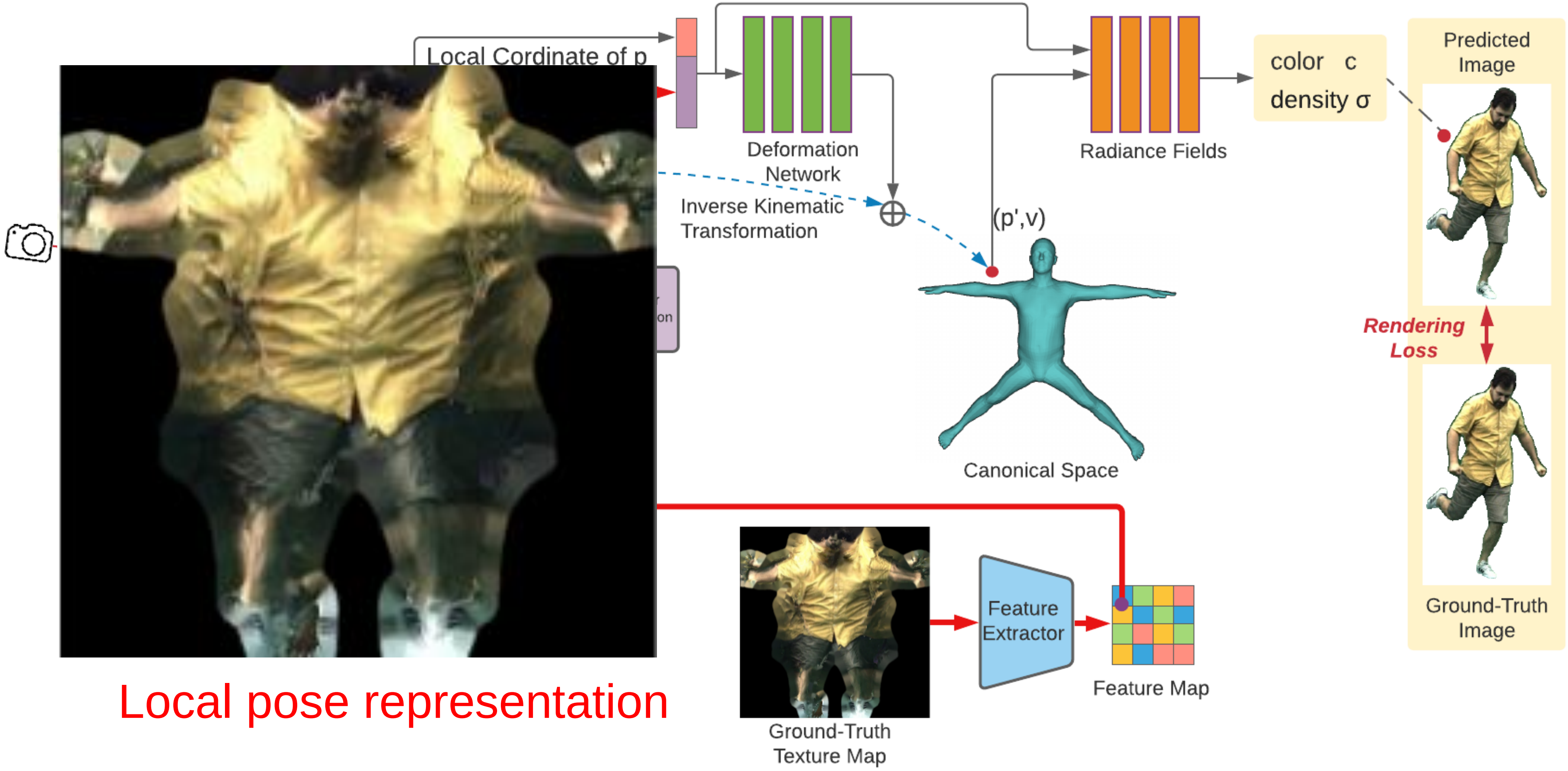
Texture Map



Tracked SMPL Mesh

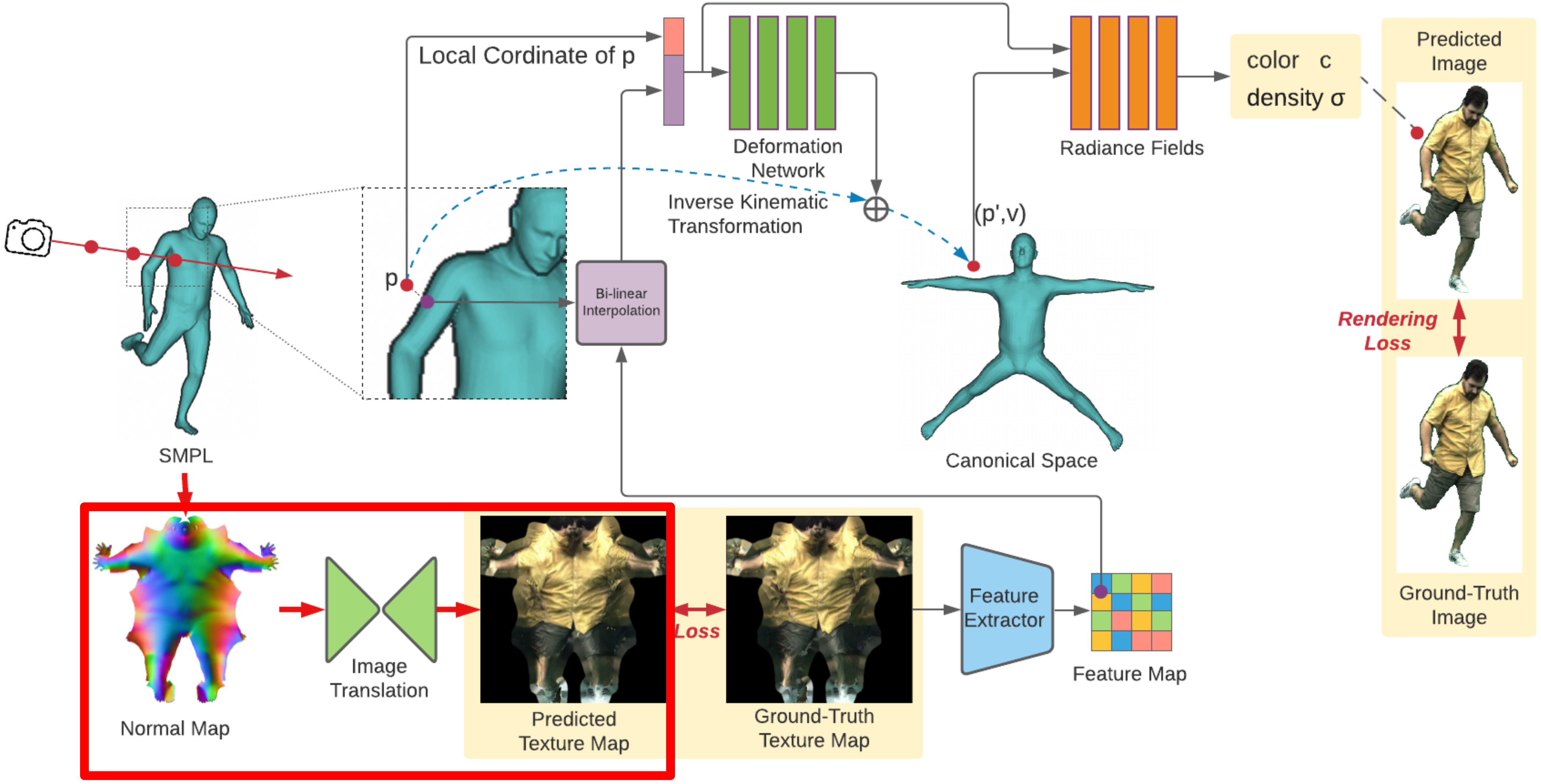
Multi-view Images

Texture Map as Latent Variable



Local pose representation

Texture Map as Latent Variable



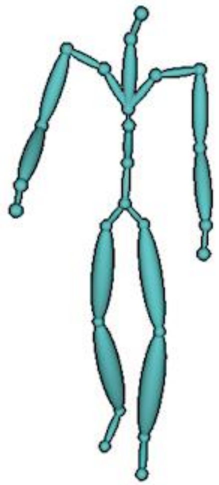
Texture Map as Latent Variable

Geometry-guided
Deformable Neural Fields
(One Proposed Component
of Neural Actor)

Neural Actor
(Our Full Model)

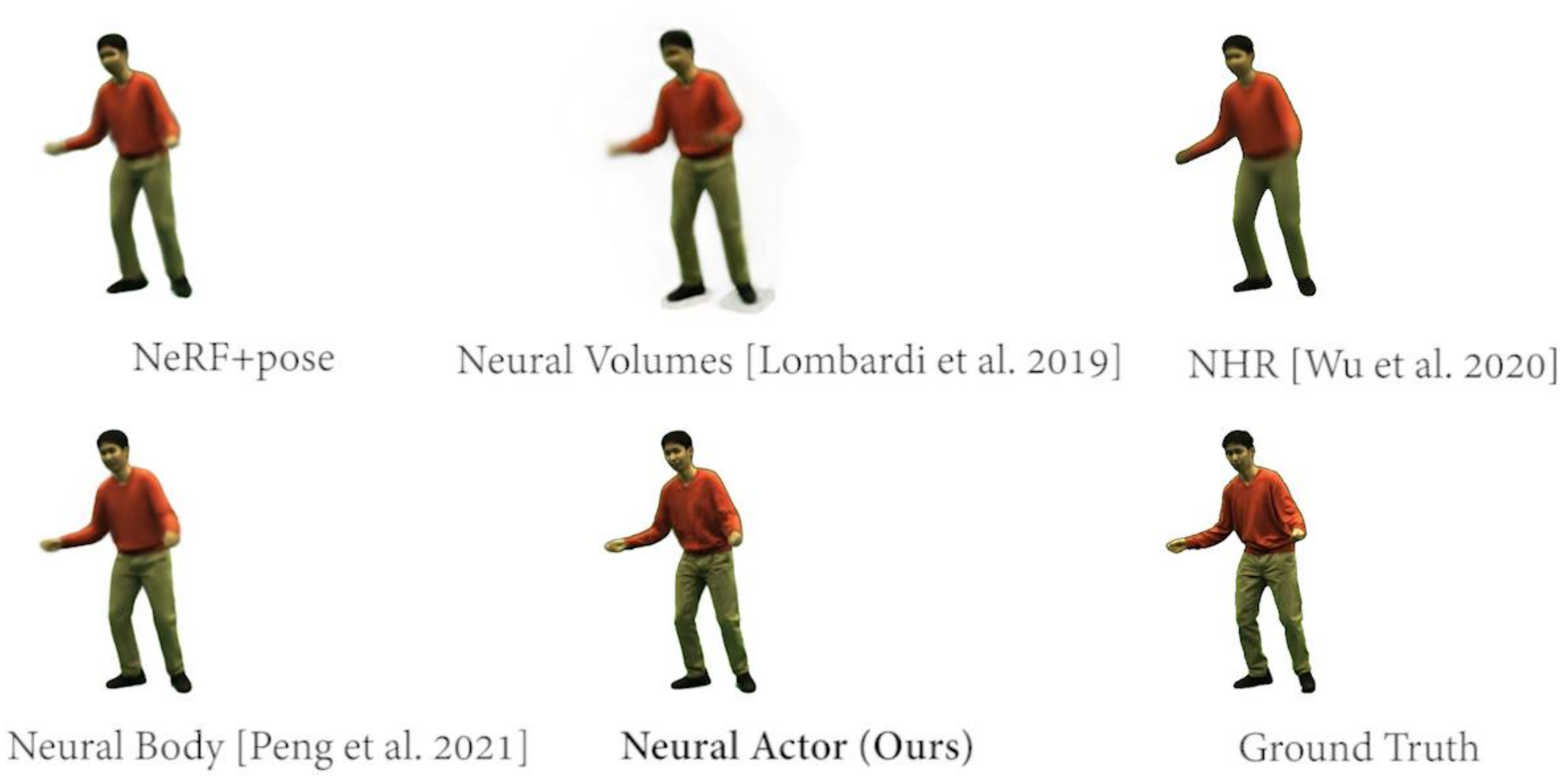
Ground Truth

Results of Neural Actor



Arbitrary input driving poses　　Synthesized results by Neural Actor

Comparisons

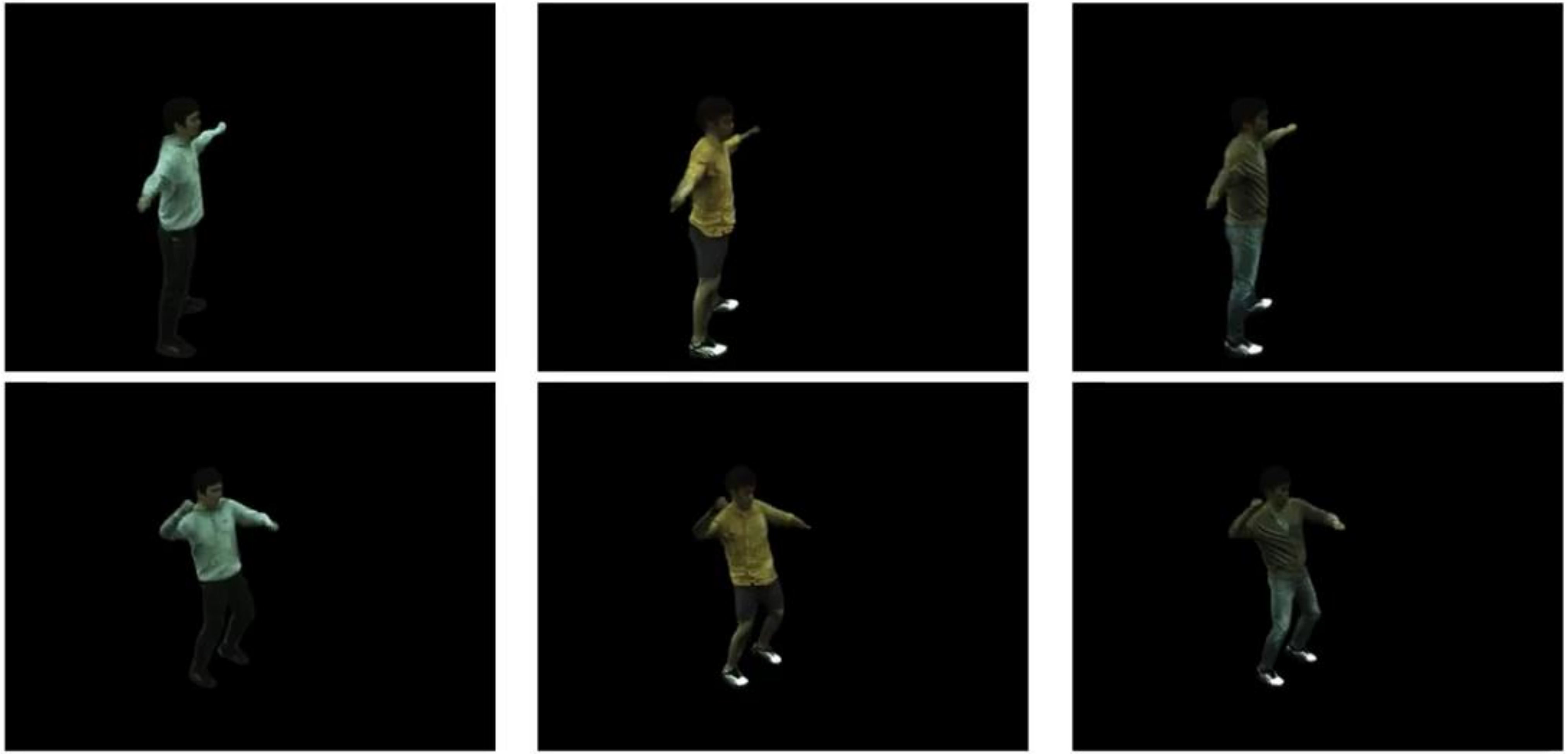


Results of Neural Actor

Results of Neural Actor

Results of Neural Actor

Person-specific Model -> Generalized Human Model



Y. Wang, Q. Gao, L. Liu, ***L. Liu***, C. Theobalt, B. Chen. Neural Novel Actor: Learning a Generalized Animatable Neural Representation for Human Actors, Arxiv 2022

Person-specific Model -> Generalized Human Model

GT

Neural Human Perfomer
NeurIPS 2021

Ours

Y. Wang, Q. Gao, L. Liu, ***L. Liu***, C. Theobalt, B. Chen. Neural Novel Actor: Learning a Generalized Animatable Neural Representation for Human Actors, Arxiv 2022

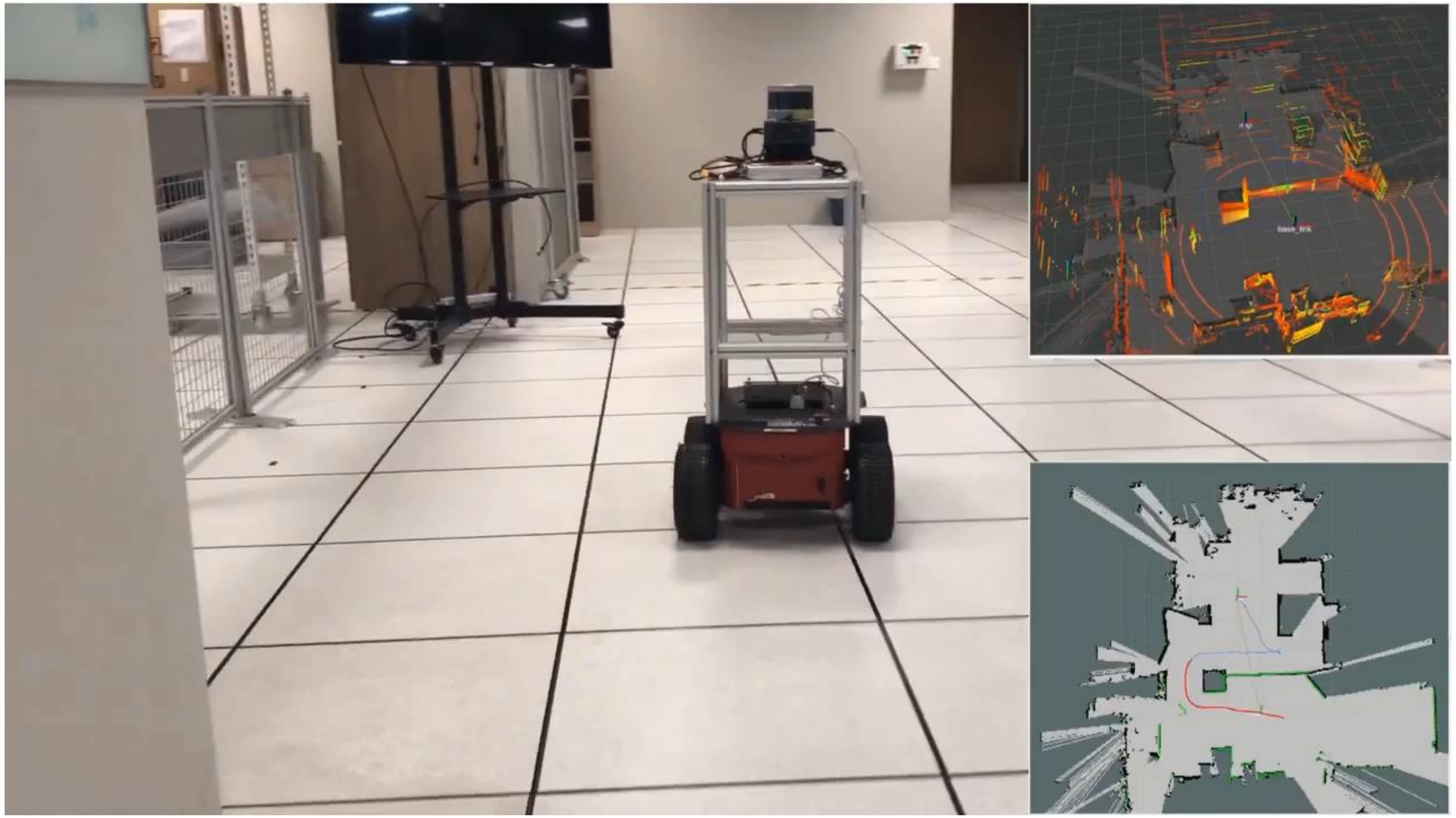
Future Directions

Future Directions

- Modeling and rendering more complex scenes.

Future Directions

- Efficiency, Accuracy



Future Directions

- Sparse-view / Single-view reconstruction

Future Directions

- Generalization -> To learn a prior

Future Directions

- Generalization -> To learn a prior -> Learn from in-the-wild data

Thank you!