Recent Trends in 3D Reconstruction for General Non-Rigid Scenes

State-Of-The-Art Report

Raza Yunus, Jan Eric Lenssen, Michael Niemeyer, Christian Rupprecht, Yiyi Liao, Christian Theobalt, Gerard Pons-Moll, Jia-Bin Huang, Vladislav Golyanik and Eddy Ilg

Presented by Raza Yunus

The 45th Annual Conference of the European Association for Computer Graphics is organized by CYENS Centre of Excellence in collaboration with the University of Cyprus and the Cyprus University of Technology.

Talk Schedule

Introduction

State-Of-The-Art Methods - Part 1

5 minute break

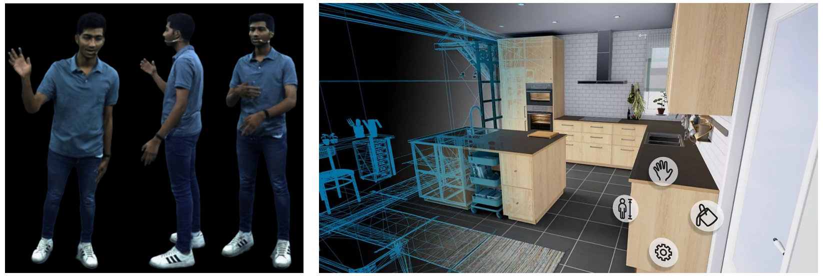
State-Of-The-Art Methods - Part 2 & 3

Future Directions

Motivation & Applications

The world is dynamic! Needs to be modelled in various applications.

Motivation & Applications



Telepresence / VR

Movie & Gaming Industry

Robotics / AR

Motivation & Applications

Recent advances are making non-rigid 3D reconstruction methods more and more powerful!

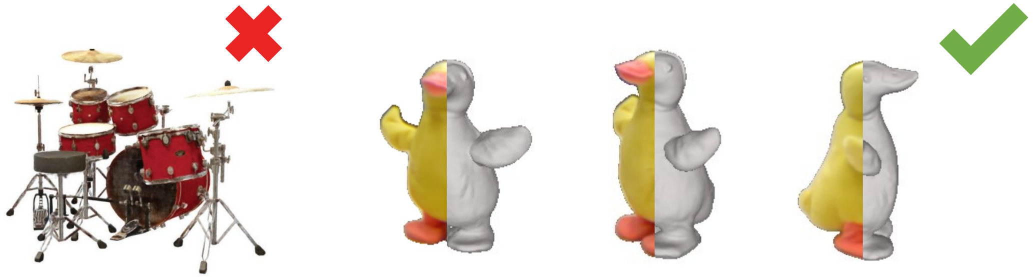
Robotics / AR

Scope

Recent Trends in 3D Reconstruction of General Non-Rigid Scenes

Scope

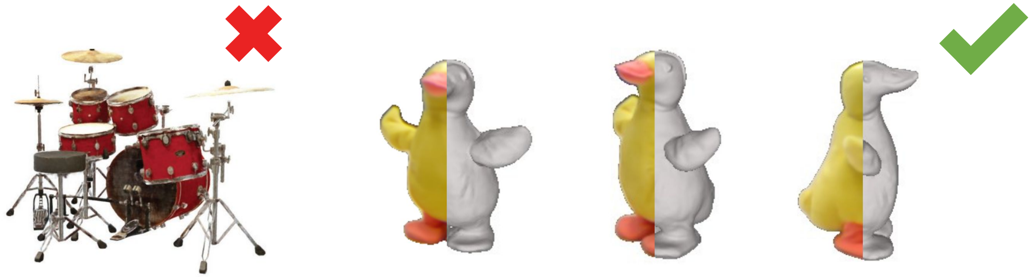
Recent Trends in 3D Reconstruction of General Non-Rigid Scenes



- Focus on methods that consider non-rigid deformations during reconstruction

Scope

Recent Trends in 3D Reconstruction of General Non-Rigid Scenes

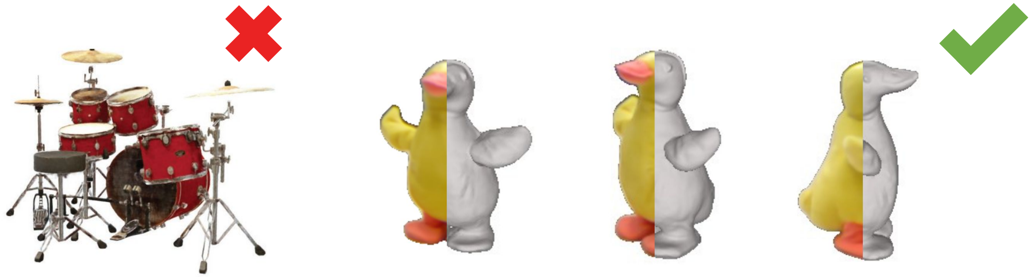


- Focus on methods that consider non-rigid deformations during reconstruction

- No domain-specific methods

Scope

Recent Trends in 3D Reconstruction of General Non-Rigid Scenes



- Focus on methods that consider non-rigid deformations during reconstruction

- No domain-specific methods

➔ Covers methods mostly from the last three years

➔ We refer to older Eurographics STARs for a survey of earlier techniques:

Tewari *et al.* (2022)

Tretschk *et al.* (2023)

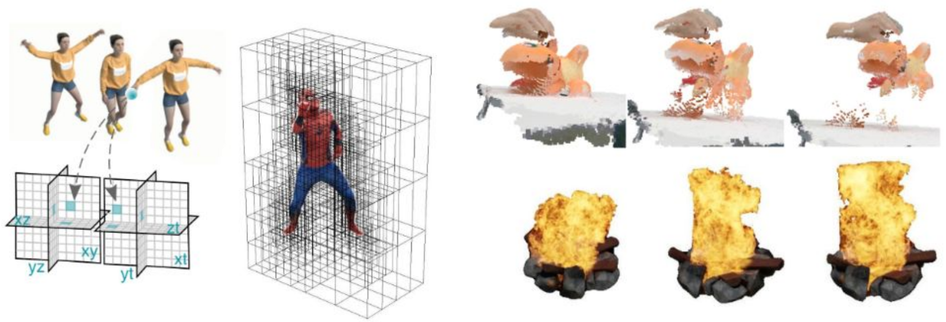
State-of-the-Art Report

State-of-the-Art Report

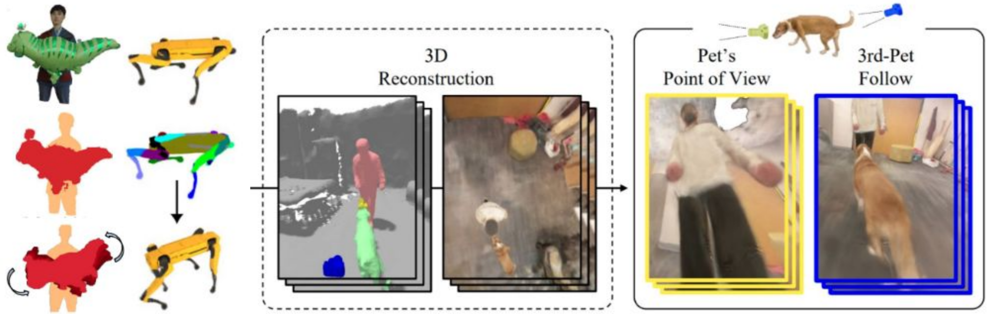
Over 150 methods divided into four categories

State-of-the-Art Report

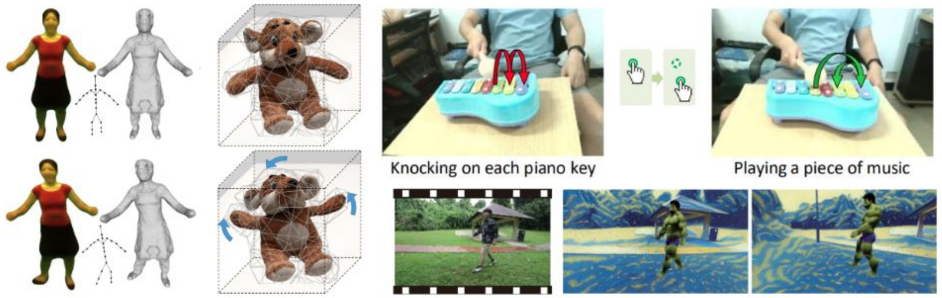
Over 150 methods divided into four categories



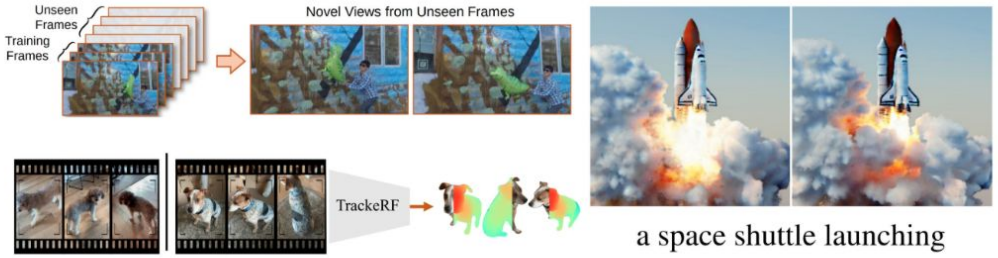
3D Non-Rigid Reconstruction and View Synthesis



Decompositional Scene Analysis



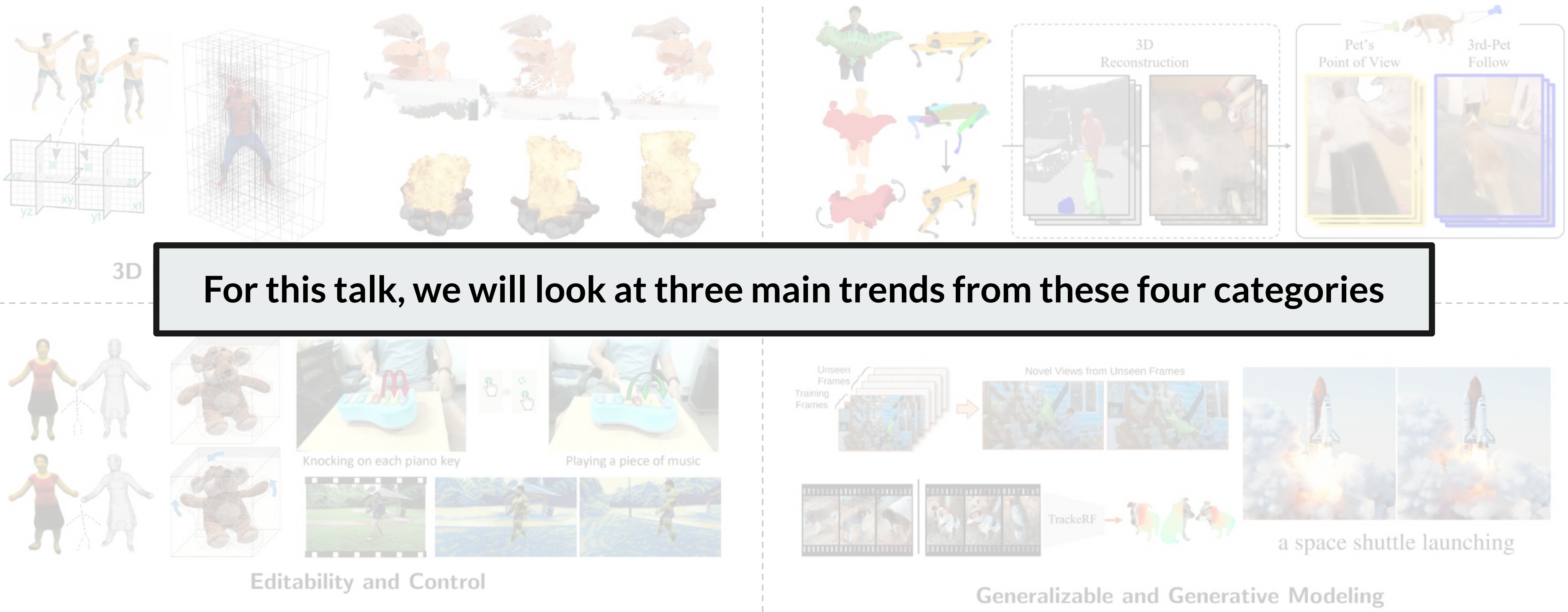
Editability and Control



Generalizable and Generative Modeling

State-of-the-Art Report

Over 150 methods divided into four categories



For this talk, we will look at three main trends from these four categories

Trends

Trends

1. Speed and Quality Advancements

Trends

1. Speed and Quality Advancements

2. Handling of Large Deformations / Long-Term 3D Correspondences

Trends

1. Speed and Quality Advancements

2. Handling of Large Deformations / Long-Term 3D Correspondences

3. Modelling Articulated Motion for General Objects

Trends

- Speed and Quality Advancements

- Ha

First, let’s have a brief look at the different aspects of non-rigid 3D reconstruction

dences

- Modelling Articulated Motion for General Objects

Task
Non-Rigid 3D Reconstruction and View Synthesis

Task

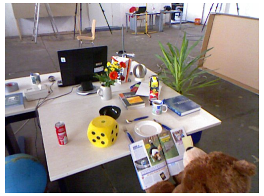
Non-Rigid 3D Reconstruction and View Synthesis

Time

Observations

Sensors and Capture Settings

Sensors and Capture Settings



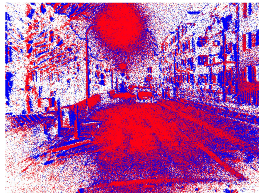
RGB

Depth

Passive Depth

Structured Depth

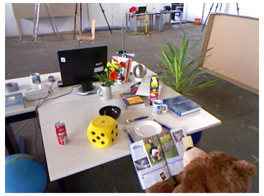
Time-of-Flight



Event

Sensor Types

Sensors and Capture Settings



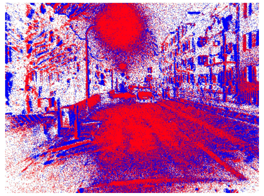
RGB

Depth

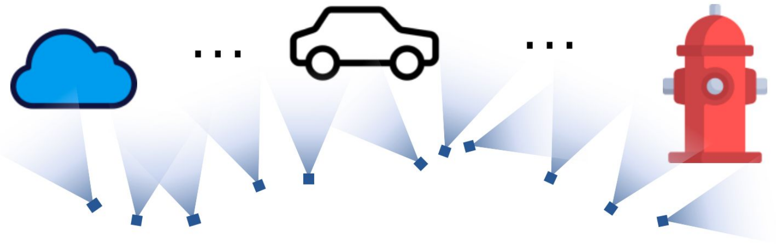
Passive Depth

Structured Depth

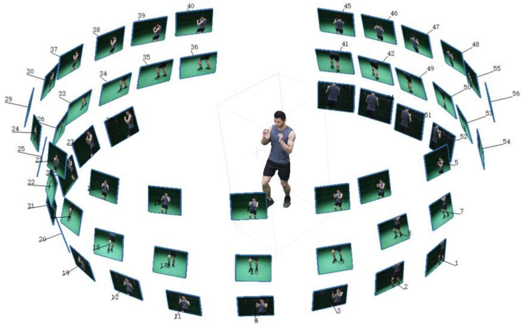
Time-of-Flight



Event



Monocular

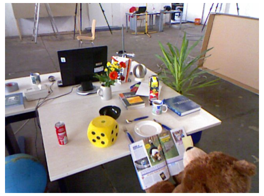


Multi-view

Sensor Types

Capture Settings

Sensors and Capture Settings



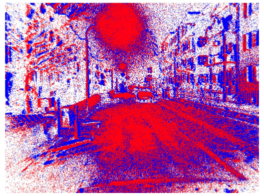
RGB

Depth

Passive Depth

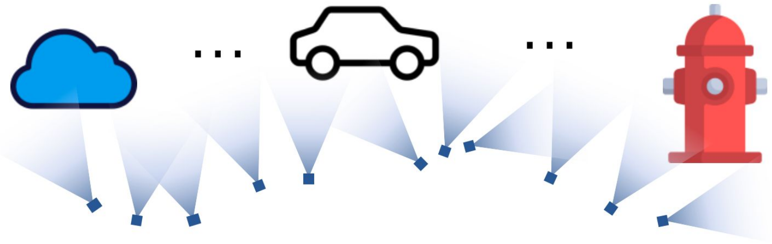
Structured Depth

Time-of-Flight

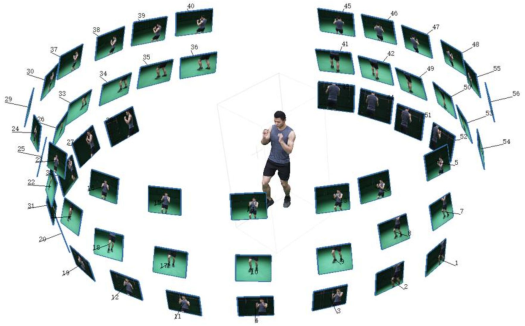


Event

Sensor Types

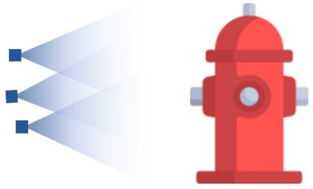


Monocular

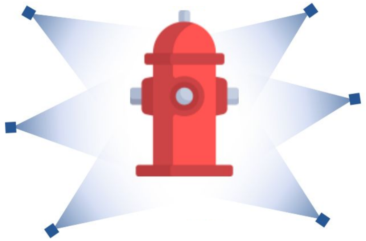


Multi-view

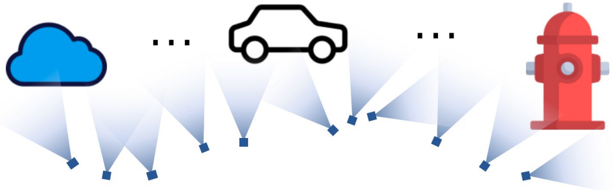
Capture Settings



Forward Facing



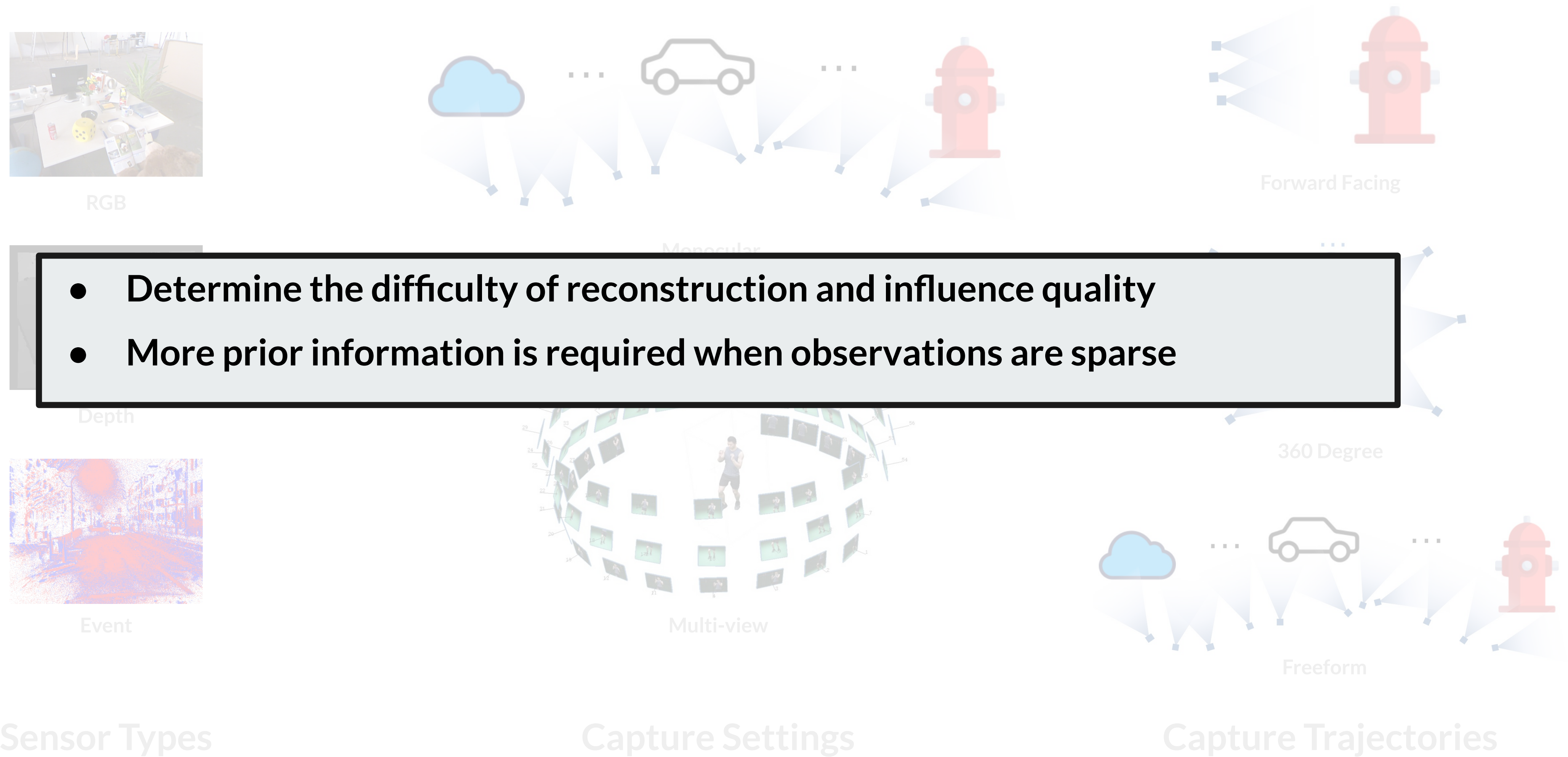
360 Degree



Freeform

Capture Trajectories

Sensors and Capture Settings



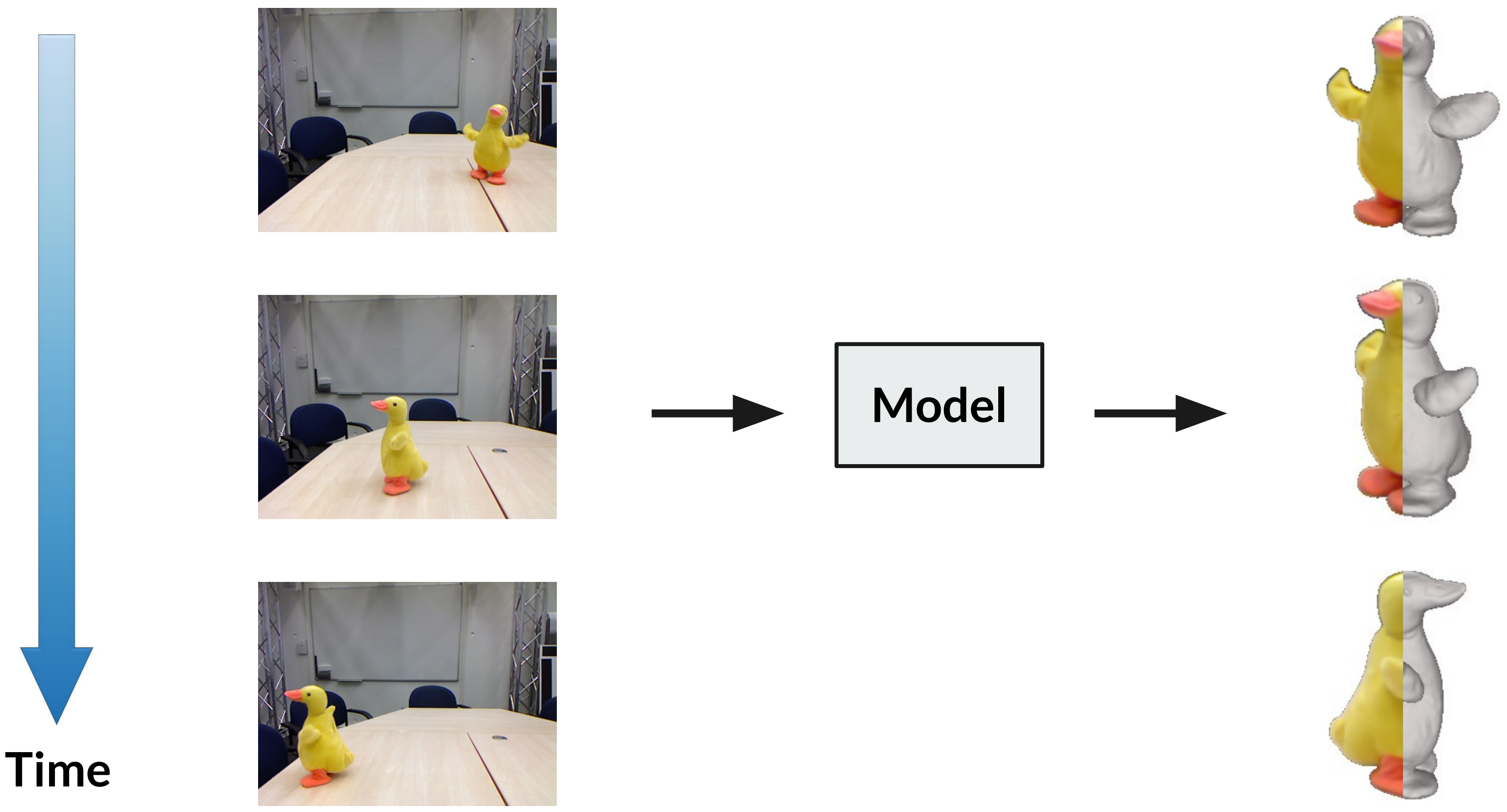
- Determine the difficulty of reconstruction and influence quality
- More prior information is required when observations are sparse

Task
Non-Rigid 3D Reconstruction and View Synthesis

Time

Observations

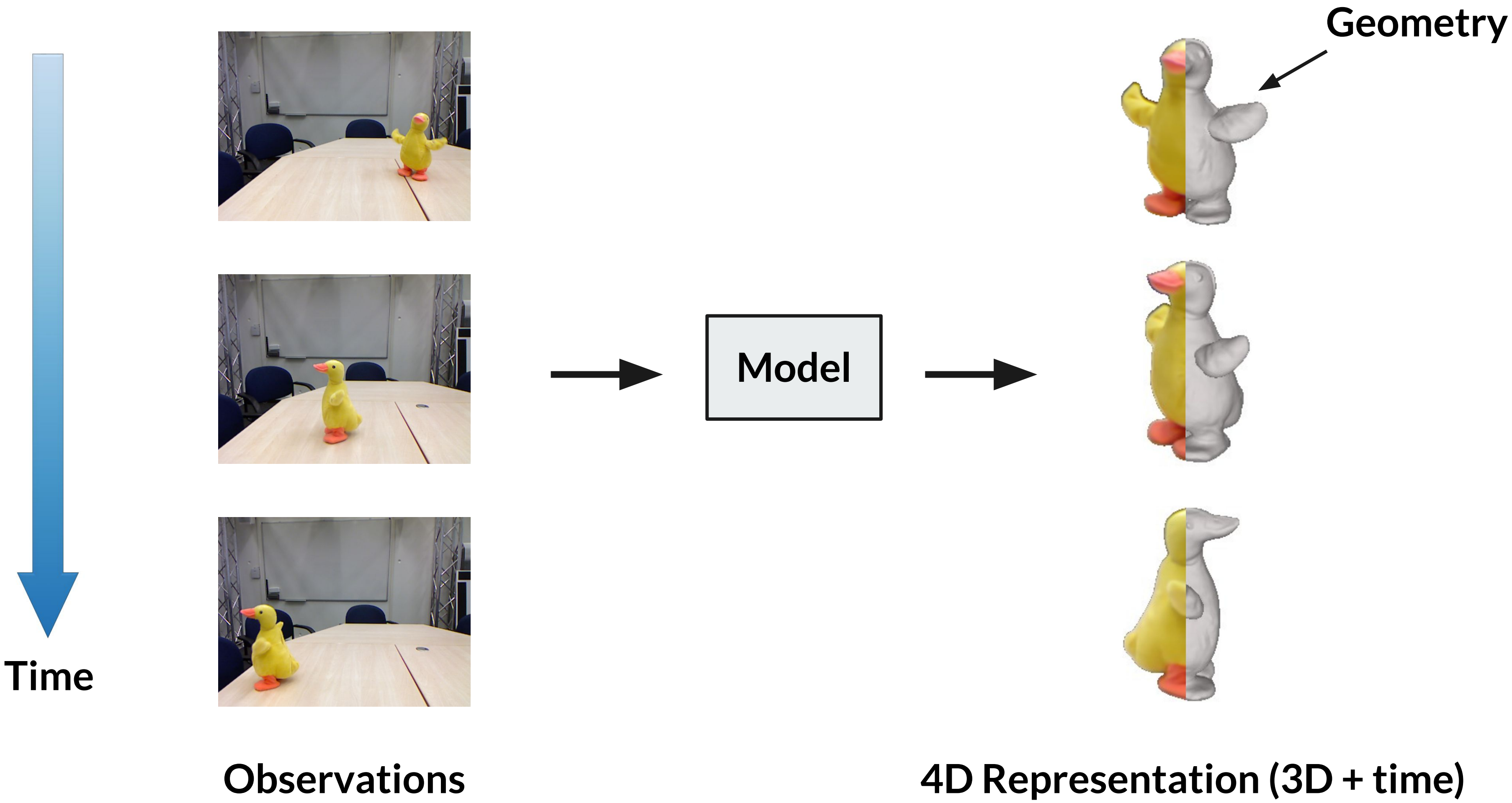
Task
Non-Rigid 3D Reconstruction and View Synthesis



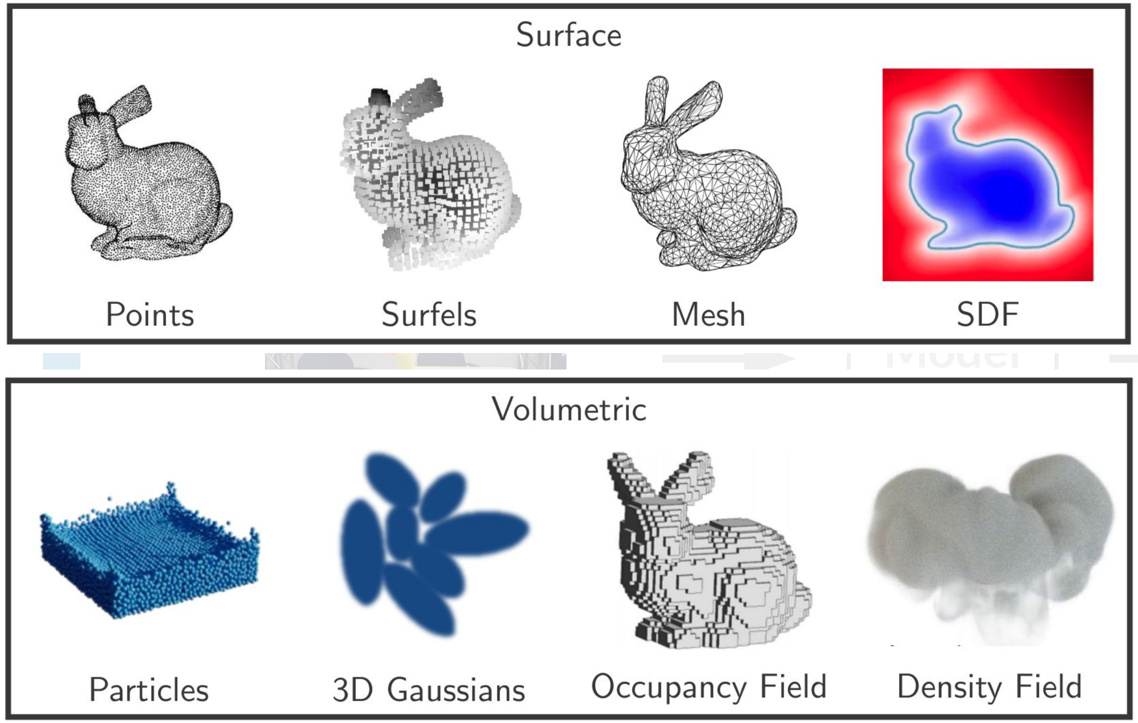
Observations

4D Representation (3D + time)

Task
Non-Rigid 3D Reconstruction and View Synthesis



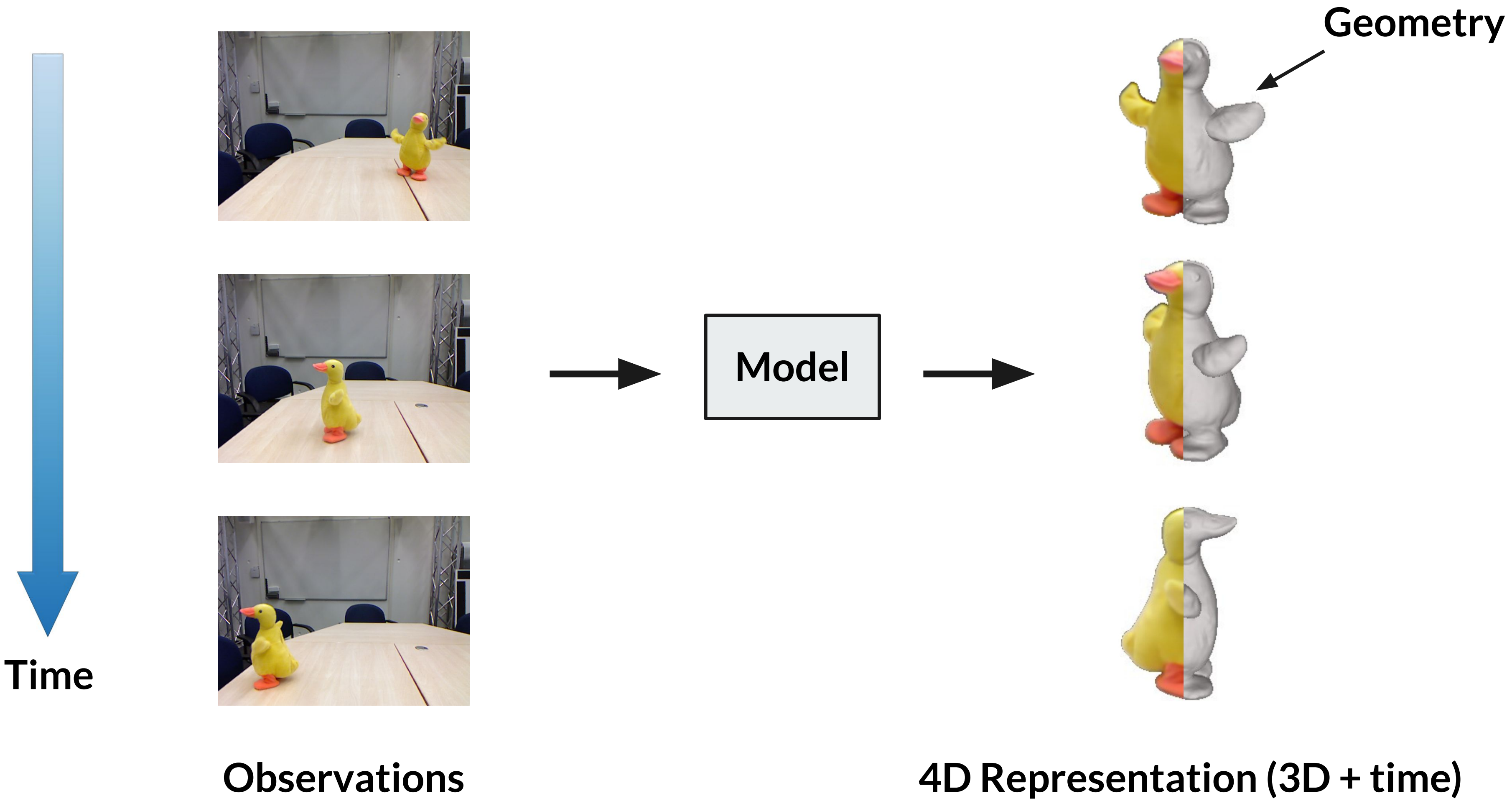
Task
Non-Rigid 3D Reconstruction and View Synthesis



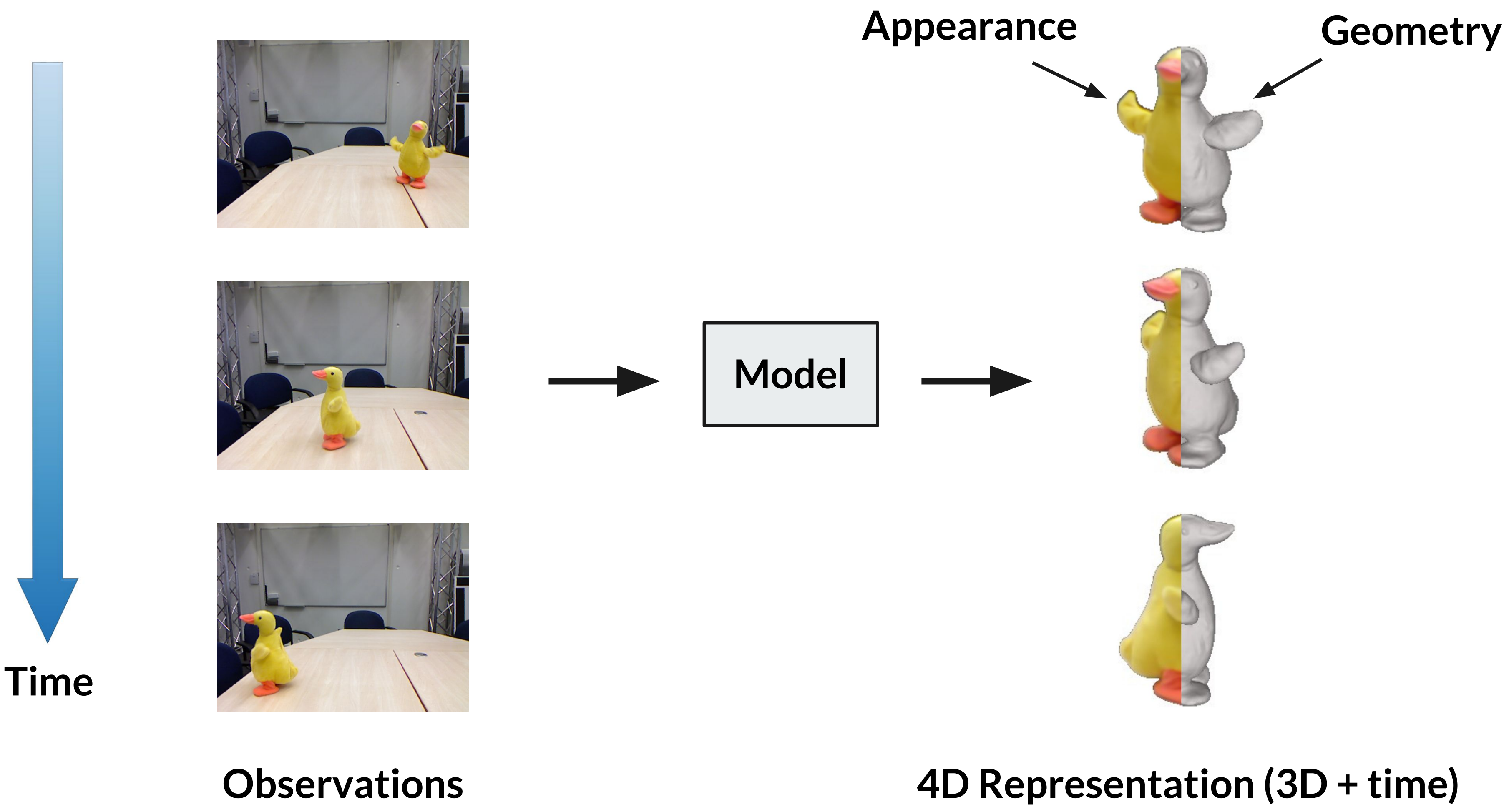
Geometry

4D Representation (3D + time)

Task
Non-Rigid 3D Reconstruction and View Synthesis

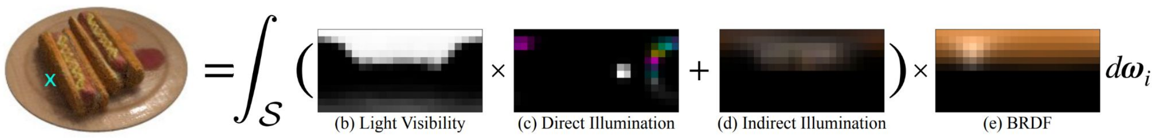


Task
Non-Rigid 3D Reconstruction and View Synthesis



Task

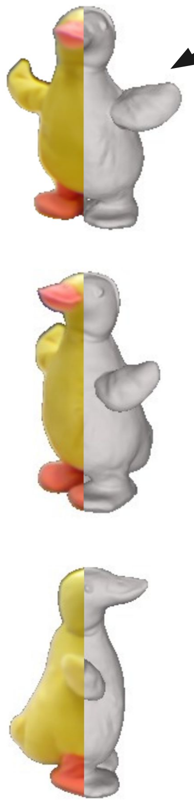
Non-Rigid 3D Reconstruction and View Synthesis



Interaction between environment lighting and surface materials!

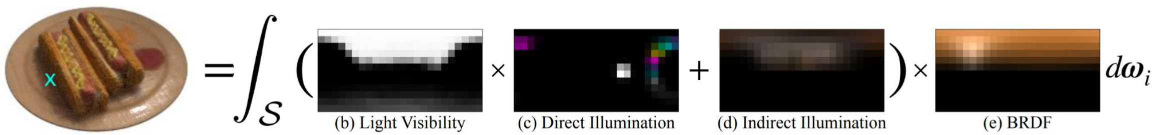
Appearance

Geometry



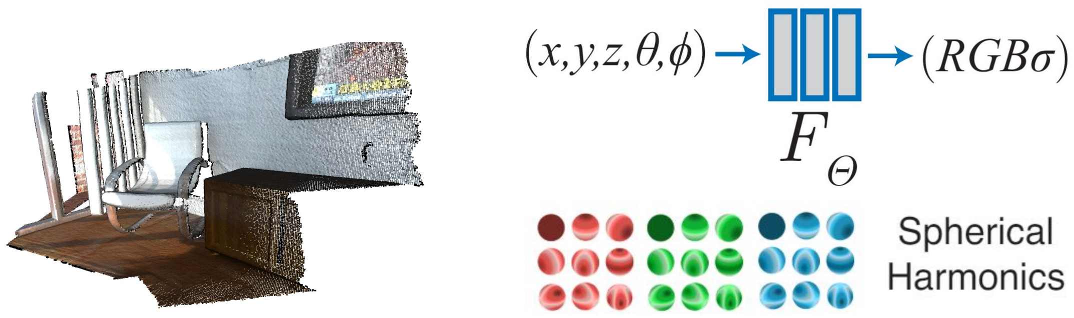
4D Representation (3D + time)

Task
Non-Rigid 3D Reconstruction and View Synthesis



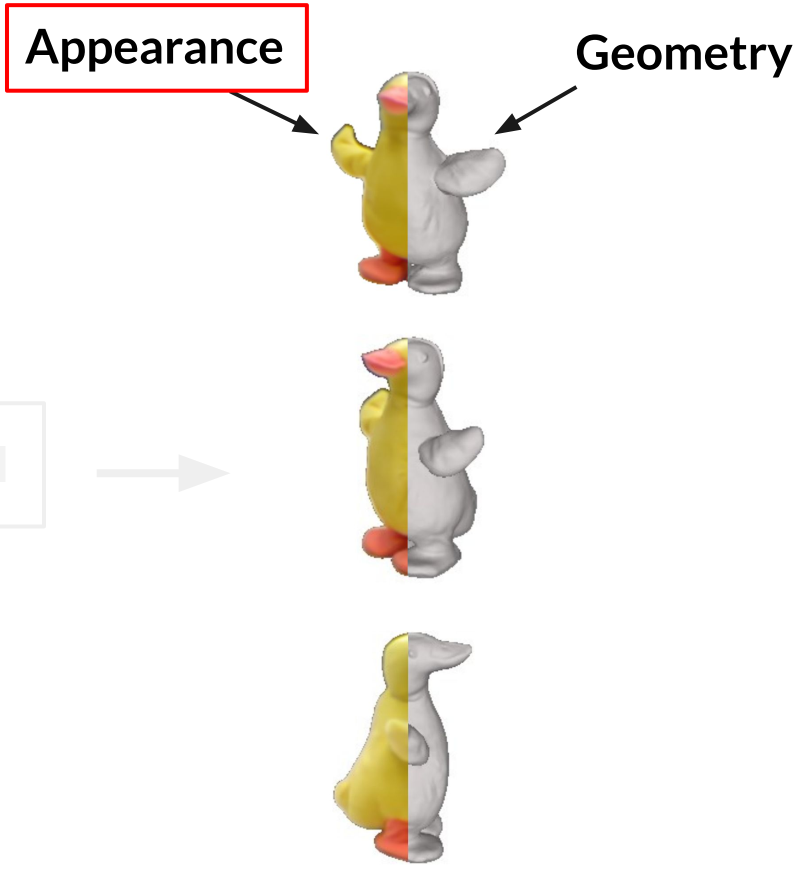
Interaction between environment lighting and surface materials!

Simplifications:



Lambertian Surface Model

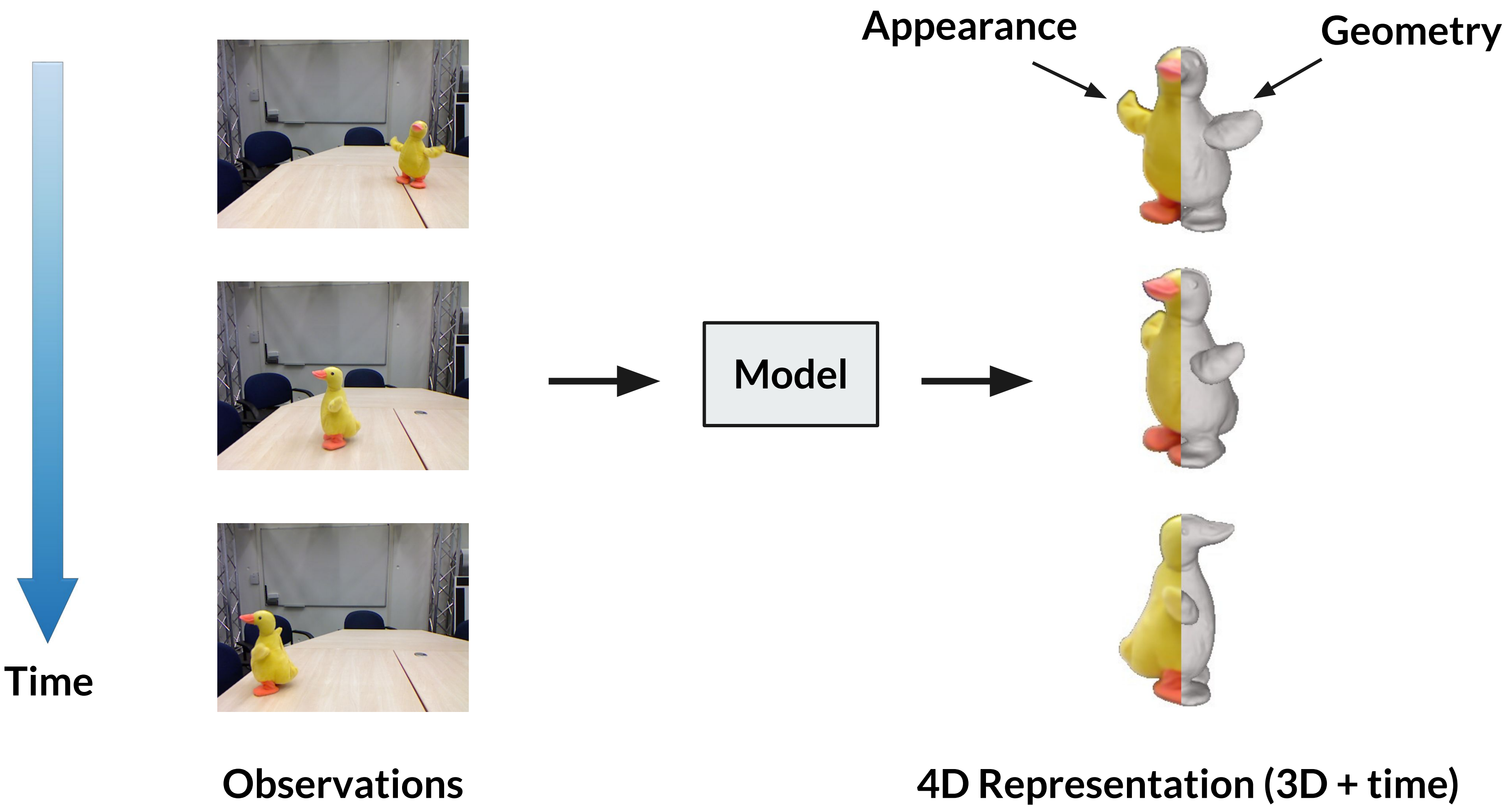
View-Dependent Outgoing Irradiance



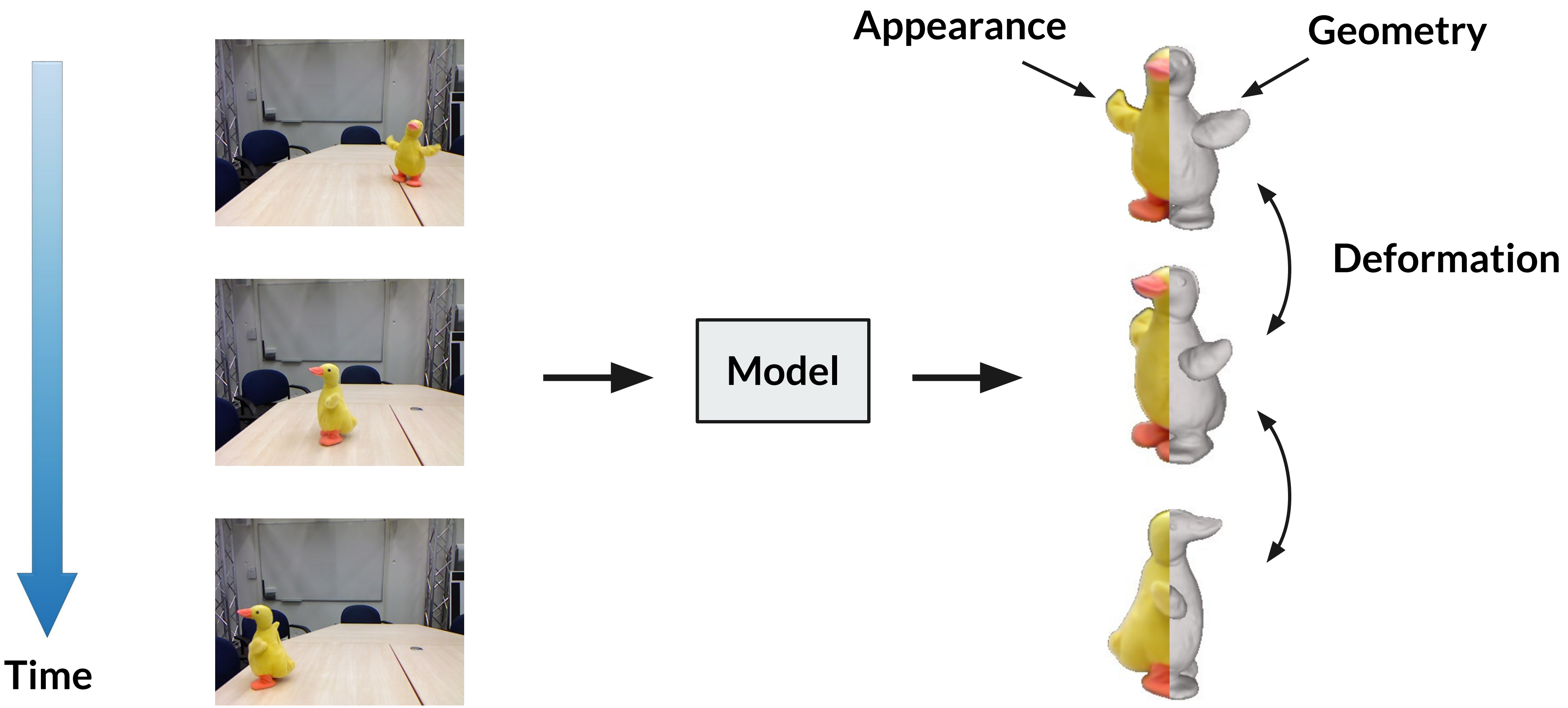
4D Representation (3D + time)

Task

Non-Rigid 3D Reconstruction and View Synthesis



Task
Non-Rigid 3D Reconstruction and View Synthesis

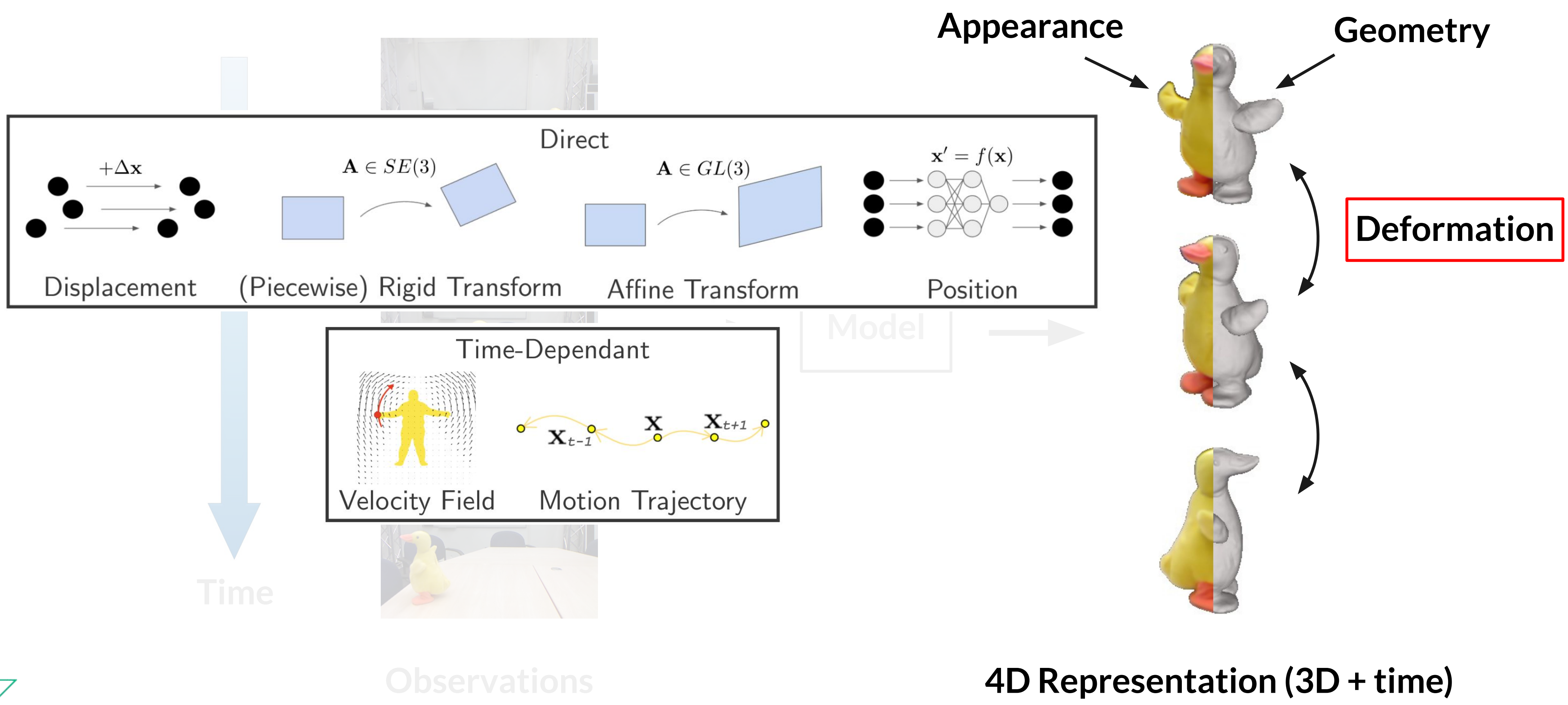


Observations

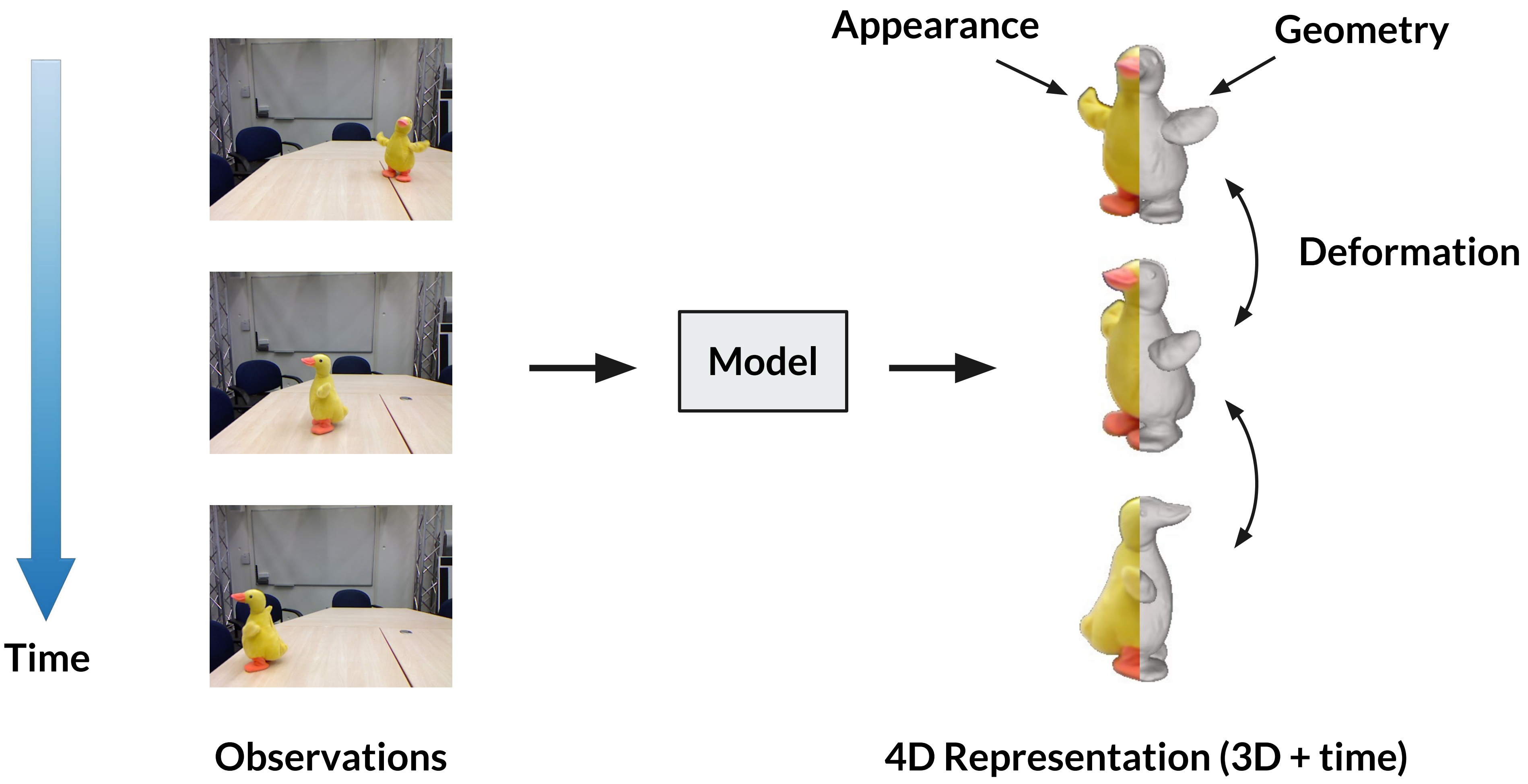
4D Representation (3D + time)

Task

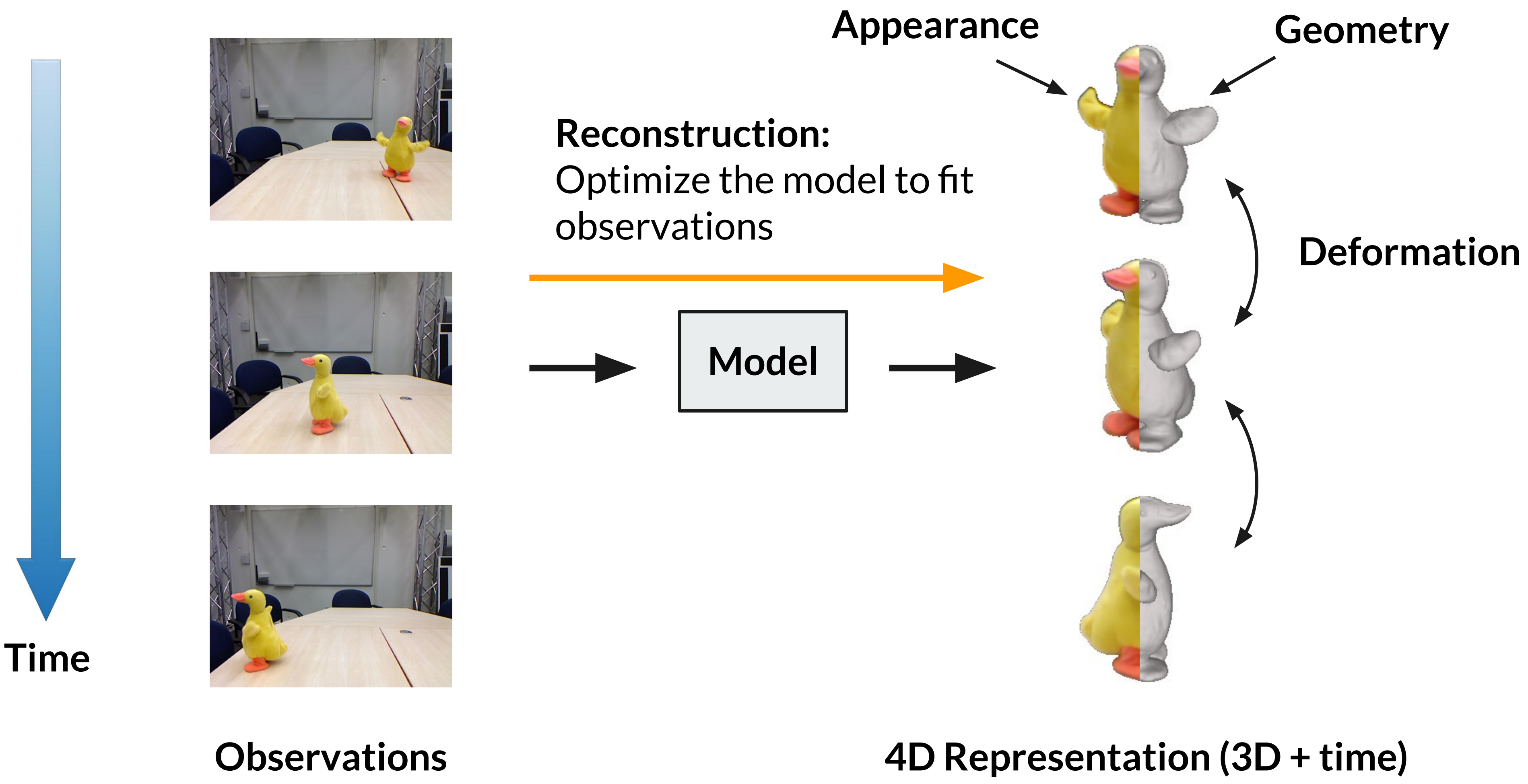
Non-Rigid 3D Reconstruction and View Synthesis



Task
Non-Rigid 3D Reconstruction and View Synthesis



Task
Non-Rigid 3D Reconstruction and View Synthesis



Task

Non-Rigid 3D Reconstruction and View Synthesis

Time

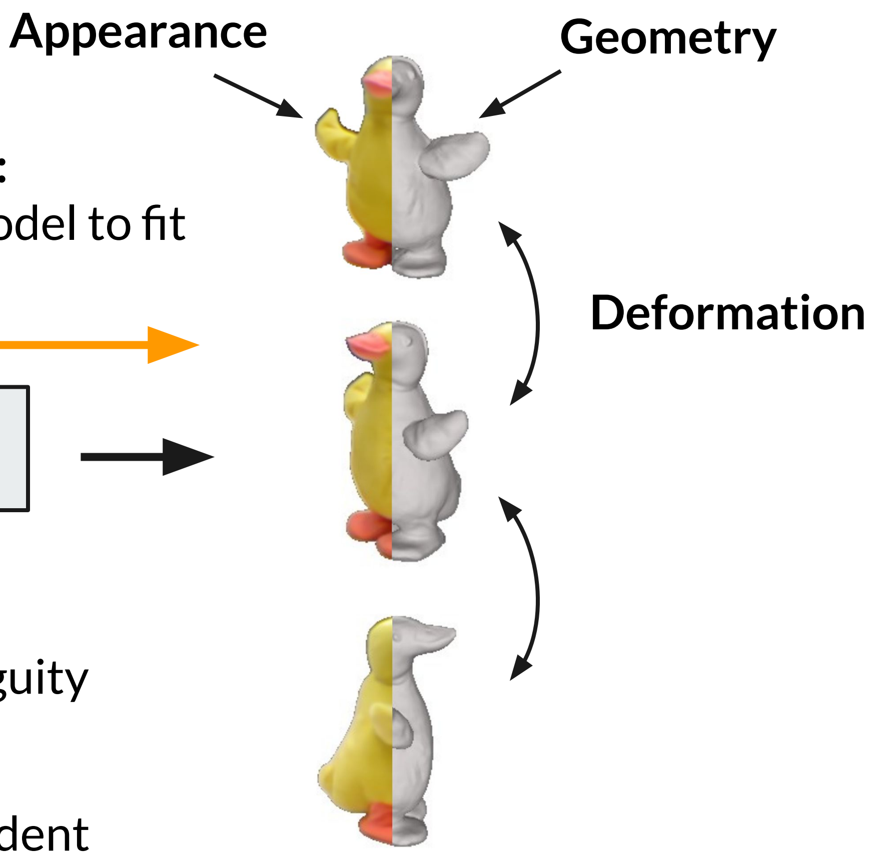
Observations

Reconstruction:
Optimize the model to fit observations

Model

Challenges:

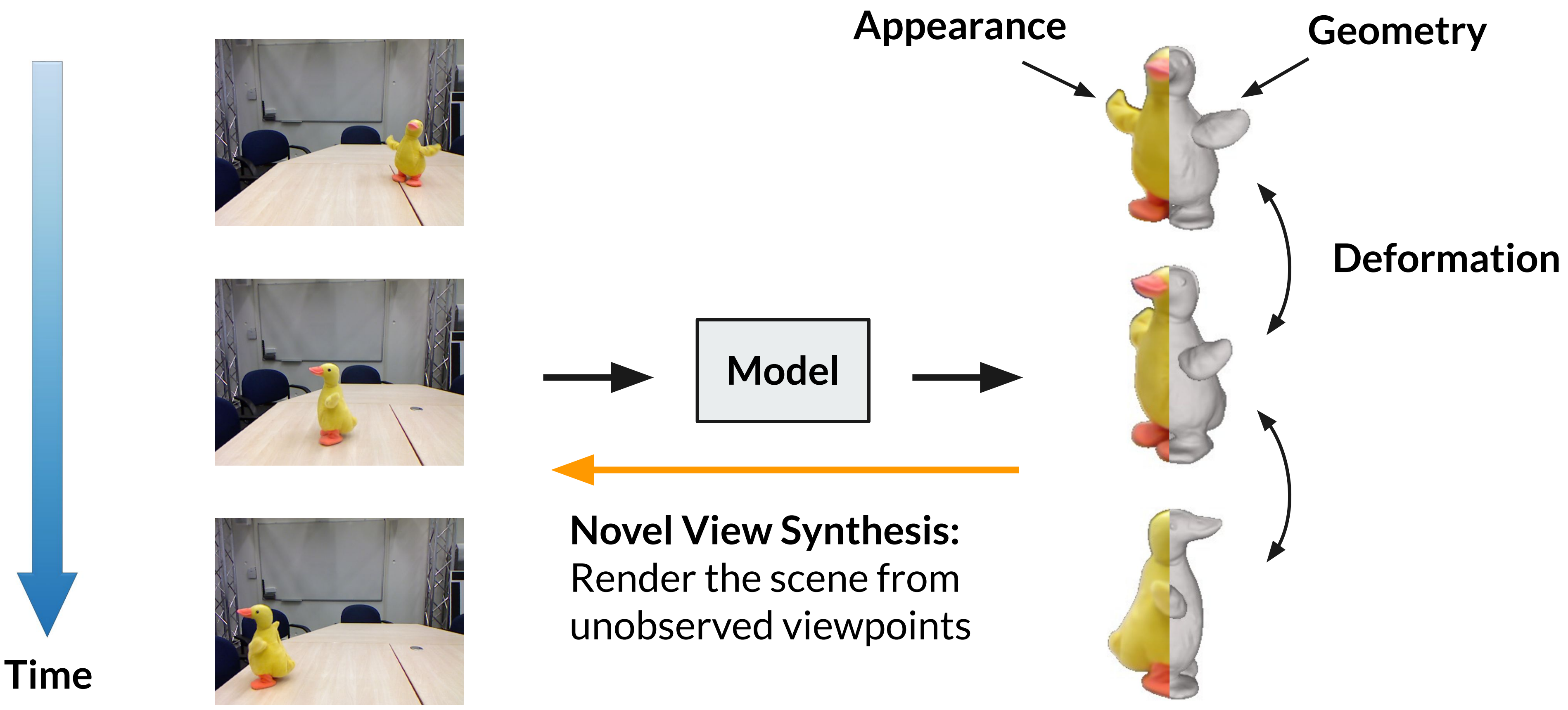
- Depth Ambiguity
- Occlusions
- View-Dependent Appearance
- …



4D Representation (3D + time)

Task

Non-Rigid 3D Reconstruction and View Synthesis

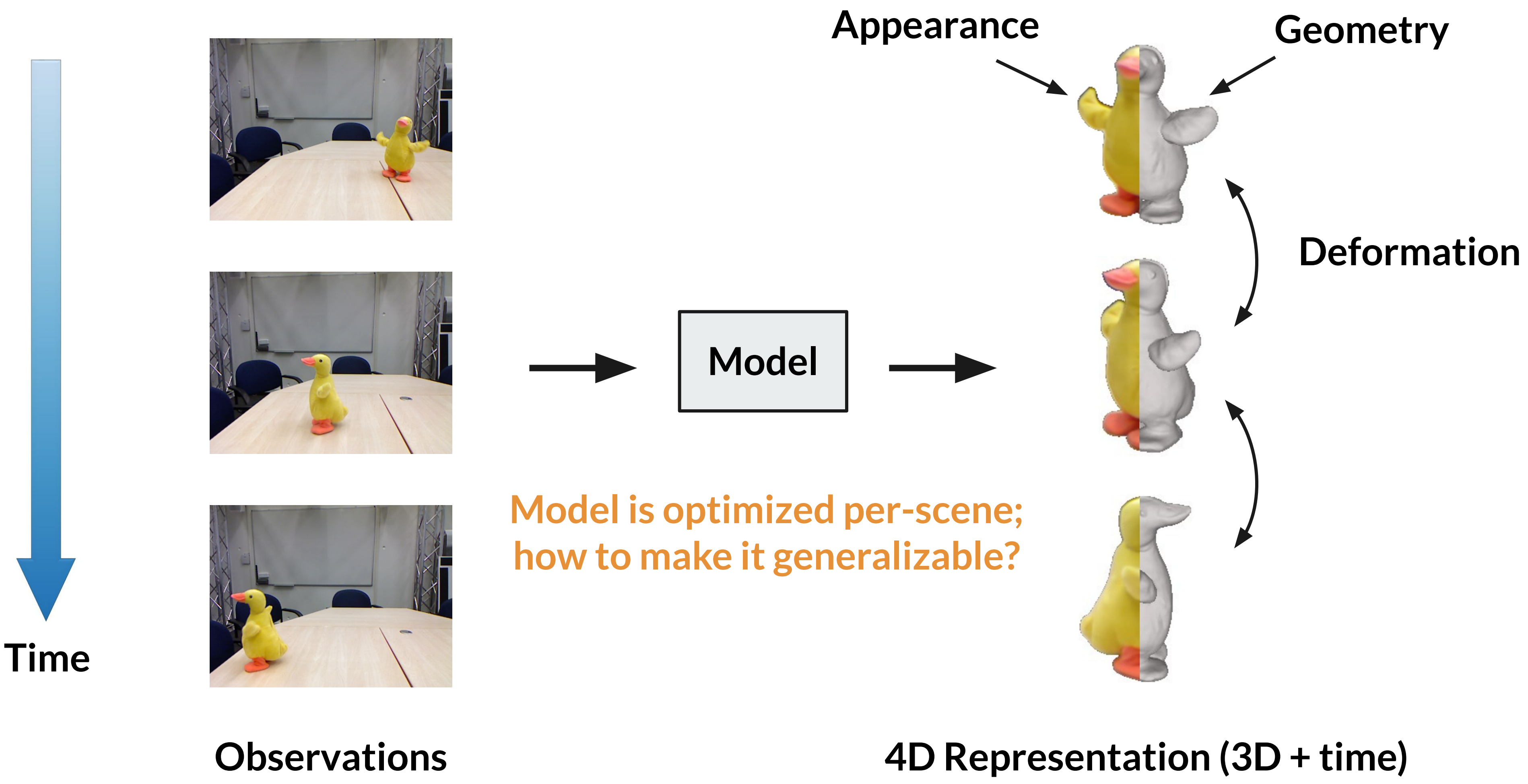


Observations

4D Representation (3D + time)

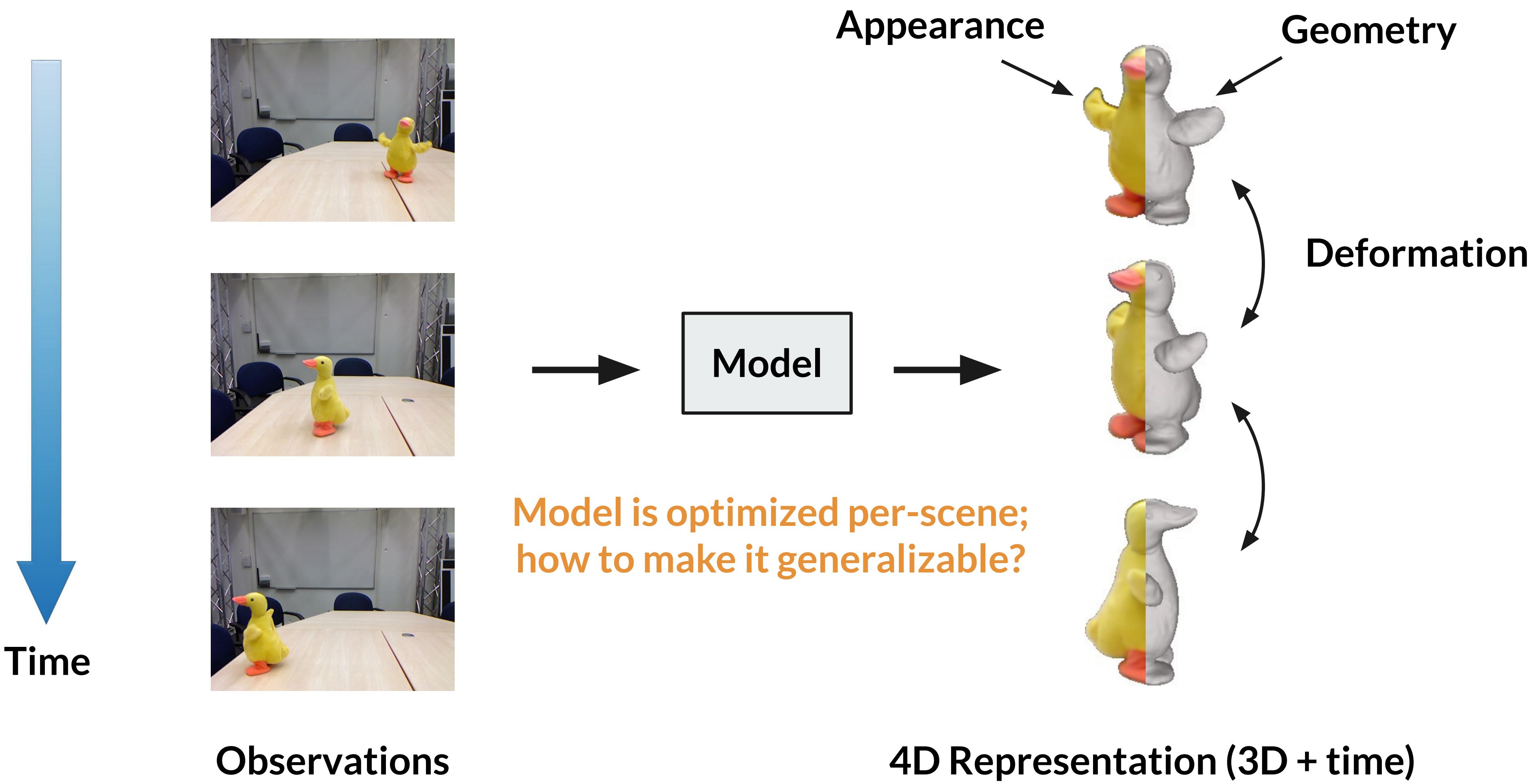
Task

Non-Rigid 3D Reconstruction and View Synthesis



Task

Non-Rigid 3D Reconstruction and View Synthesis

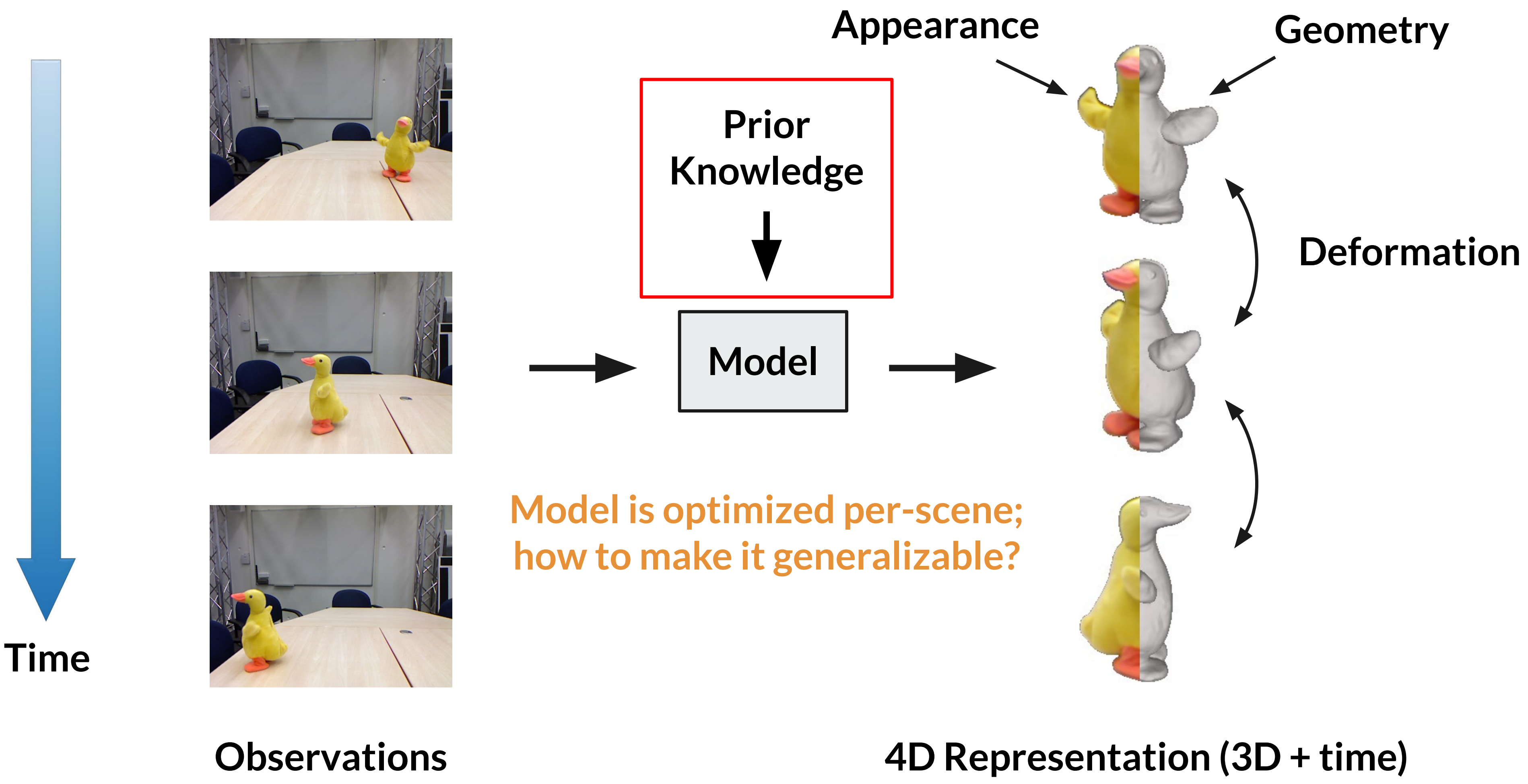


Model is optimized per-scene; how to make it generalizable?

Also, how to get a better reconstruction when observations are sparse?

Task

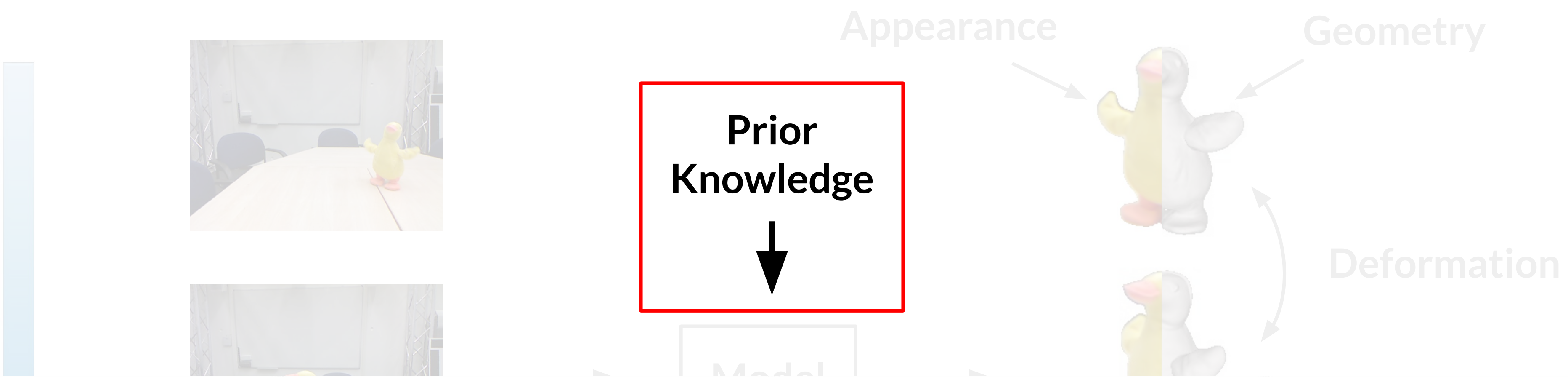
Non-Rigid 3D Reconstruction and View Synthesis



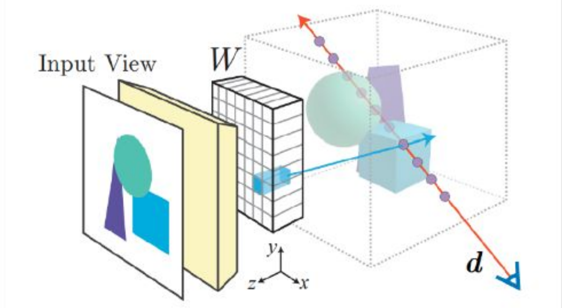
Model is optimized per-scene; how to make it generalizable?

Also, how to get a better reconstruction when observations are sparse?

Task
Non-Rigid 3D Reconstruction and View Synthesis

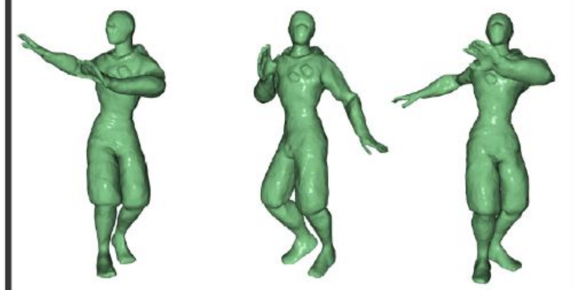


Prior Knowledge

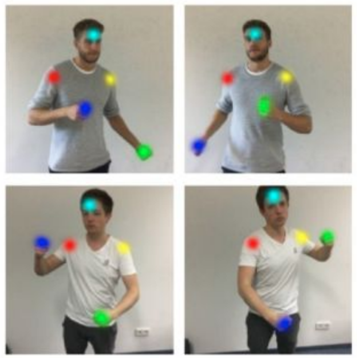


3D Shape/Appearance Priors

3D Deformation/Flow Priors



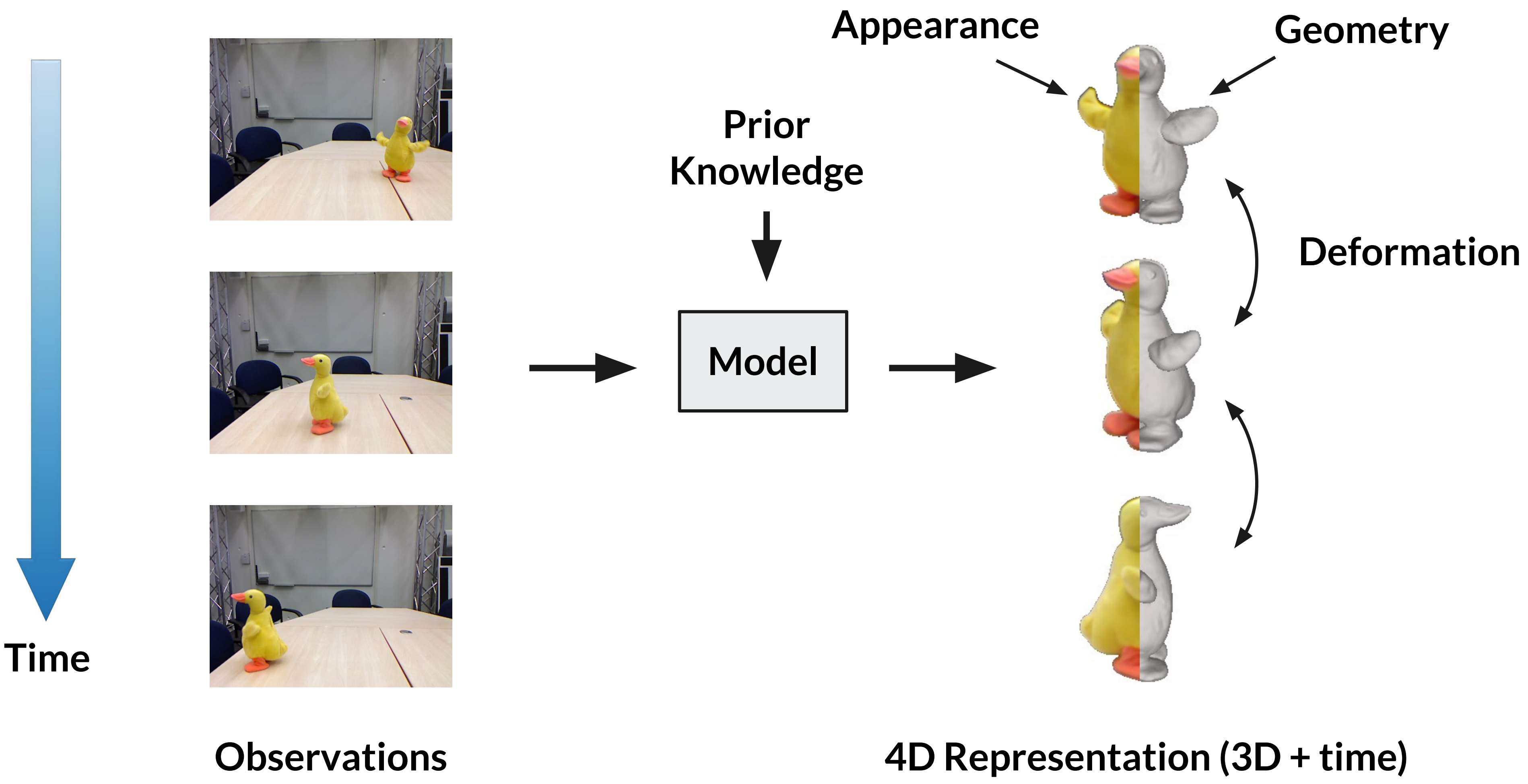
Articulation Priors



2D Correspondence Priors

Task

Non-Rigid 3D Reconstruction and View Synthesis



Task

Non-Rigid 3D Reconstruction and View Synthesis

Let’s look at the trends

Trends

1. Speed and Quality Advancements

2. Handling of Large Deformations / Long-Term 3D Correspondences

3. Modelling Articulated Motion for General Objects

Trends

1. Speed and Quality Advancements

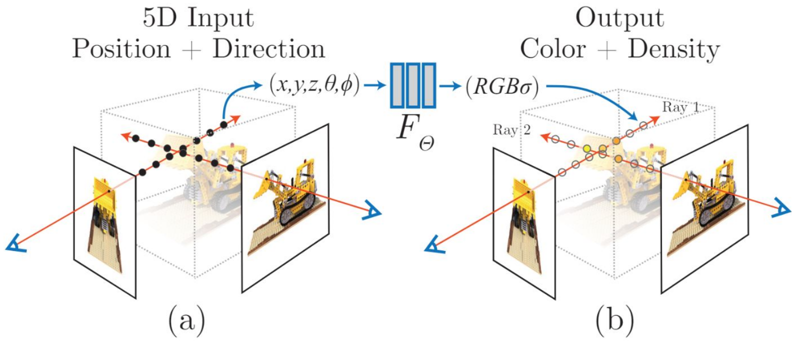
2. Handling of Large Deformations / Long-term 3D correspondences

3. Modelling Articulated Motion for General Objects

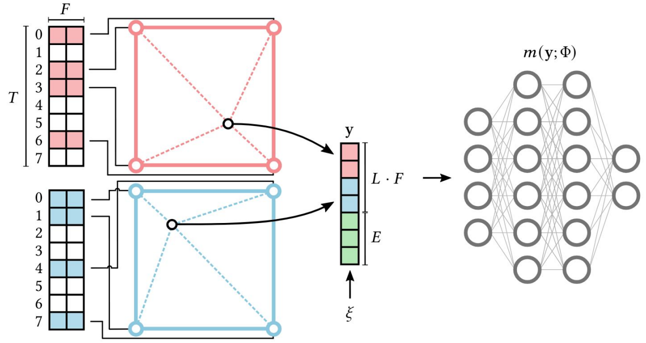
Speed and Quality Advancements

Seminal Works in 3D Rigid Reconstruction and View Synthesis

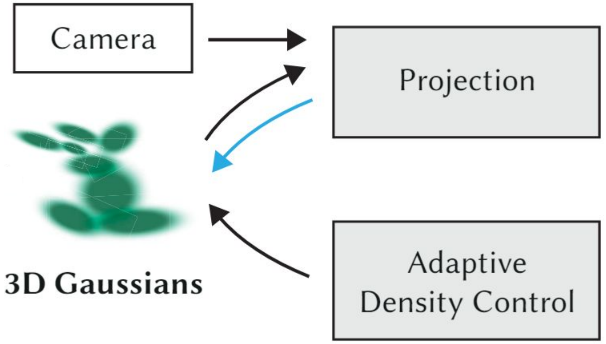
Quality or speed advancements in non-rigid setting follows the advancements in rigid setting:



NeRF



Instant-NGP

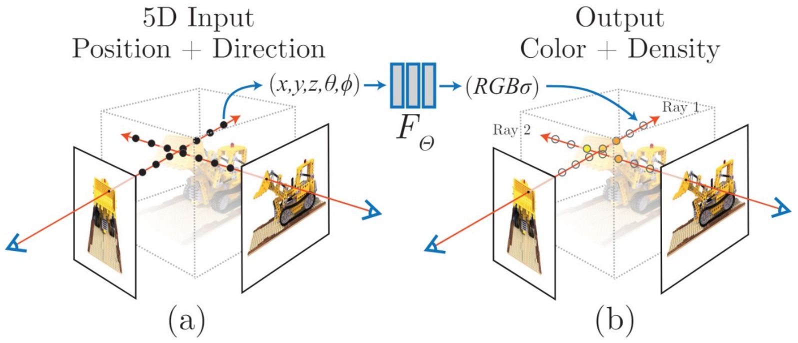


3D Gaussian Splatting

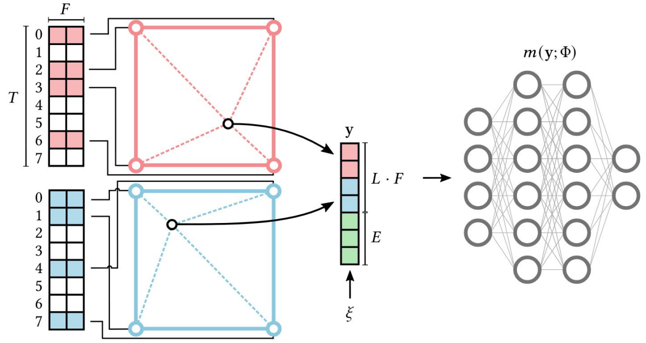
Speed and Quality Advancements

Seminal Works in 3D Rigid Reconstruction and View Synthesis

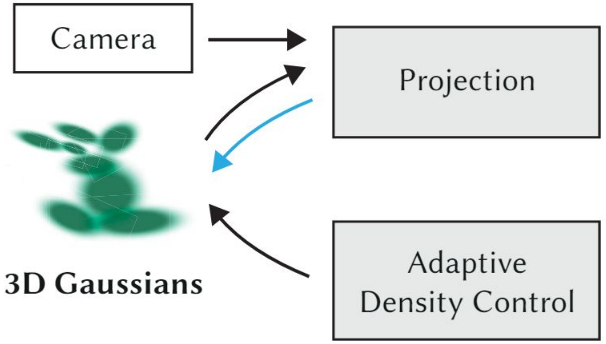
Quality or speed advancements in non-rigid setting follows the advancements in rigid setting:



NeRF



Instant-NGP

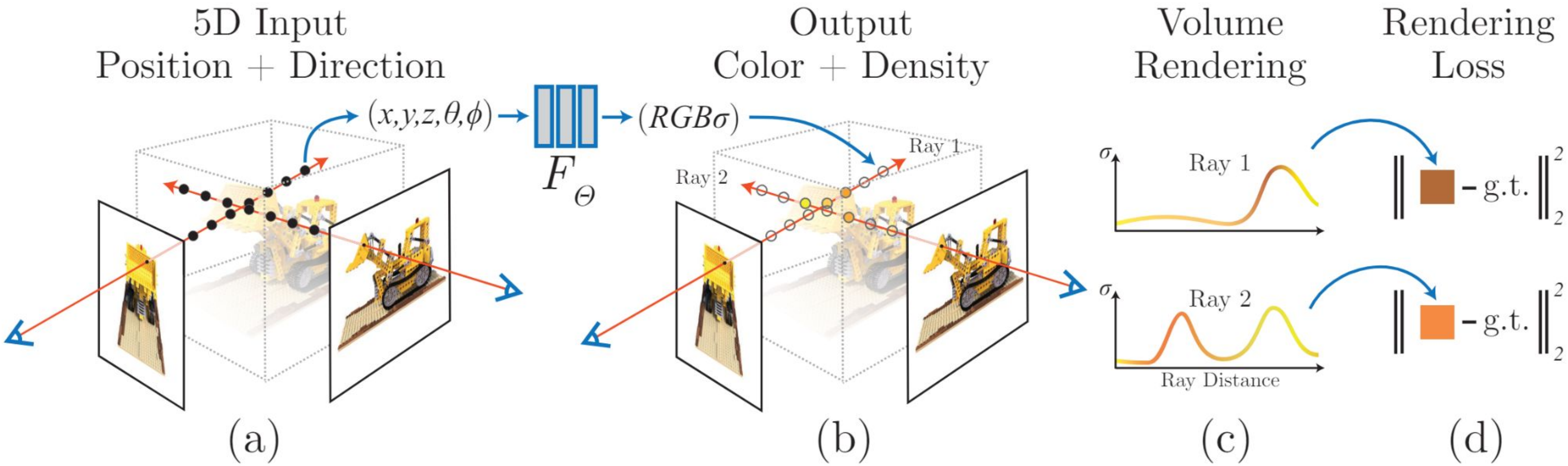


3D Gaussian Splatting

Let's see how these rigid setting advancements have been adapted to the non-rigid setting in recent years

Speed and Quality Advancements

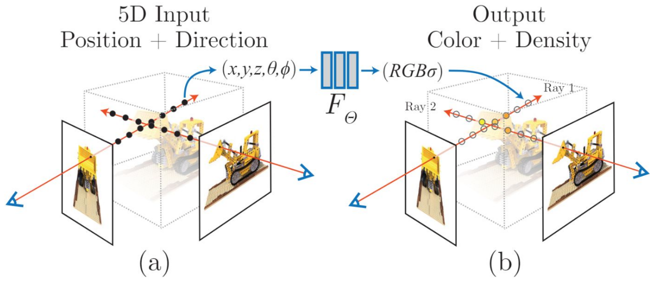
Photo-realistic View Synthesis: Neural Scene Representations



NeRF

Speed and Quality Advancements

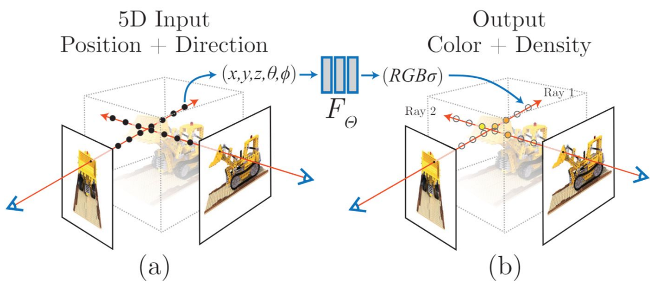
Photo-realistic View Synthesis: Neural Scene Representations



NeRF

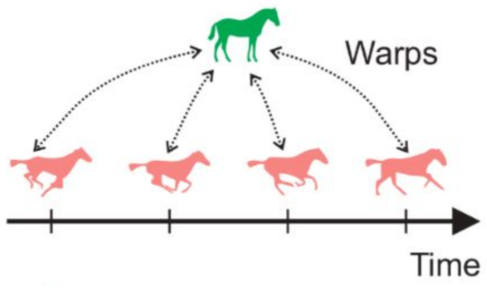
Speed and Quality Advancements

Photo-realistic View Synthesis: Neural Scene Representations



NeRF

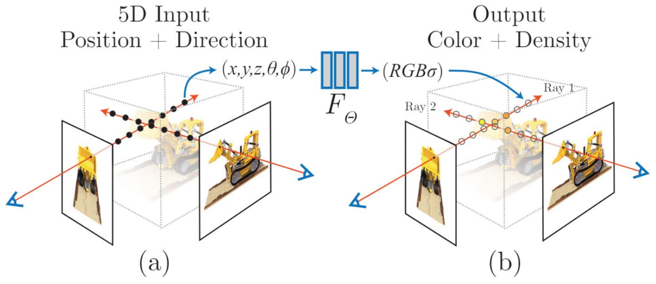
Deformable NeRF



Nerfies

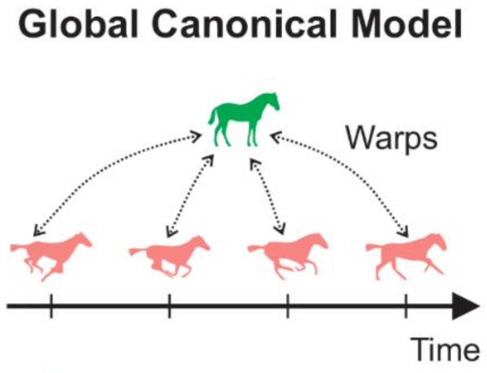
Speed and Quality Advancements

Photo-realistic View Synthesis: Neural Scene Representations

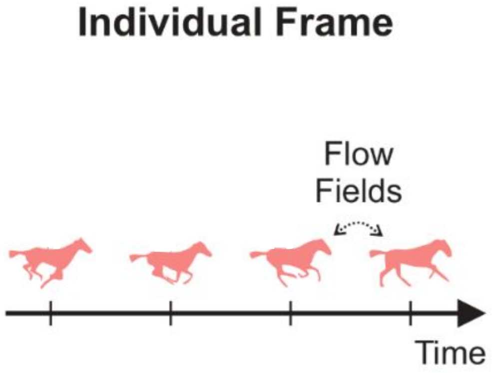


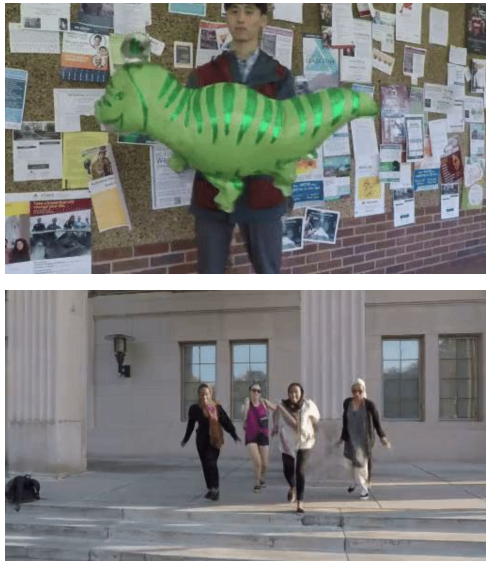
NeRF

Deformable NeRF



Space-Time NeRF



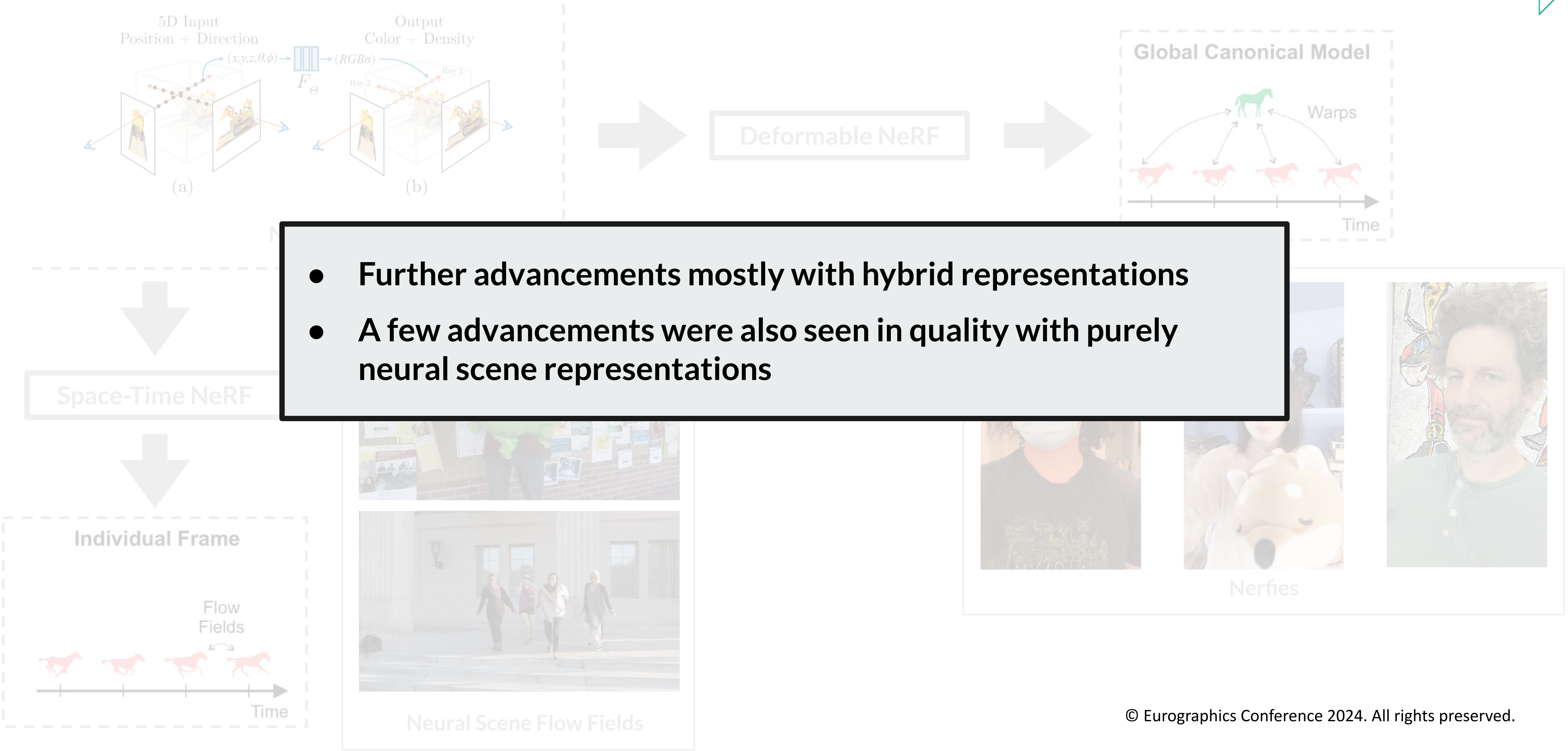


Neural Scene Flow Fields

Nerfies

Speed and Quality Advancements

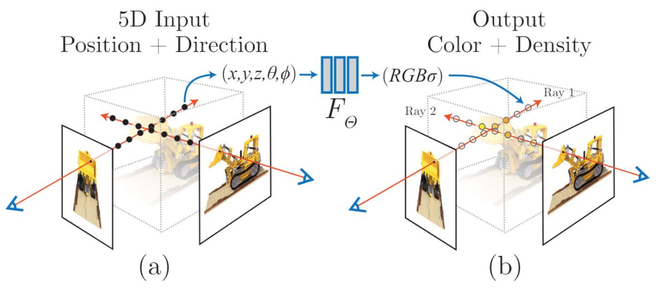
Photo-realistic View Synthesis: Neural Scene Representations



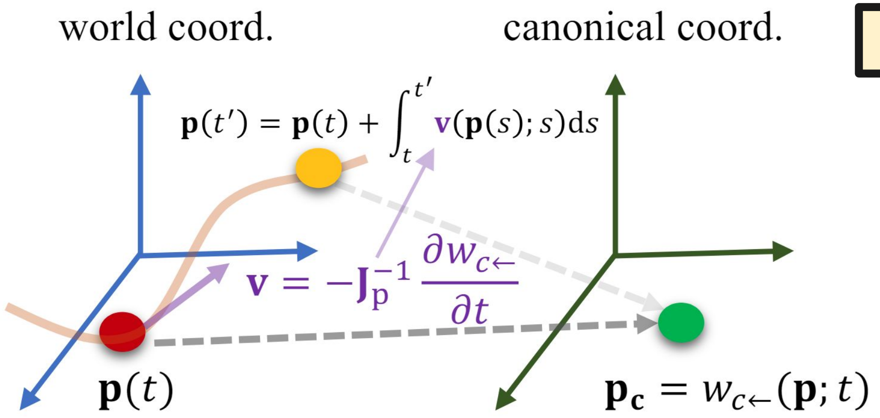
- **Further advancements mostly with hybrid representations**
- **A few advancements were also seen in quality with purely neural scene representations**

Speed and Quality Advancements

Photo-realistic View Synthesis: Neural Scene Representations

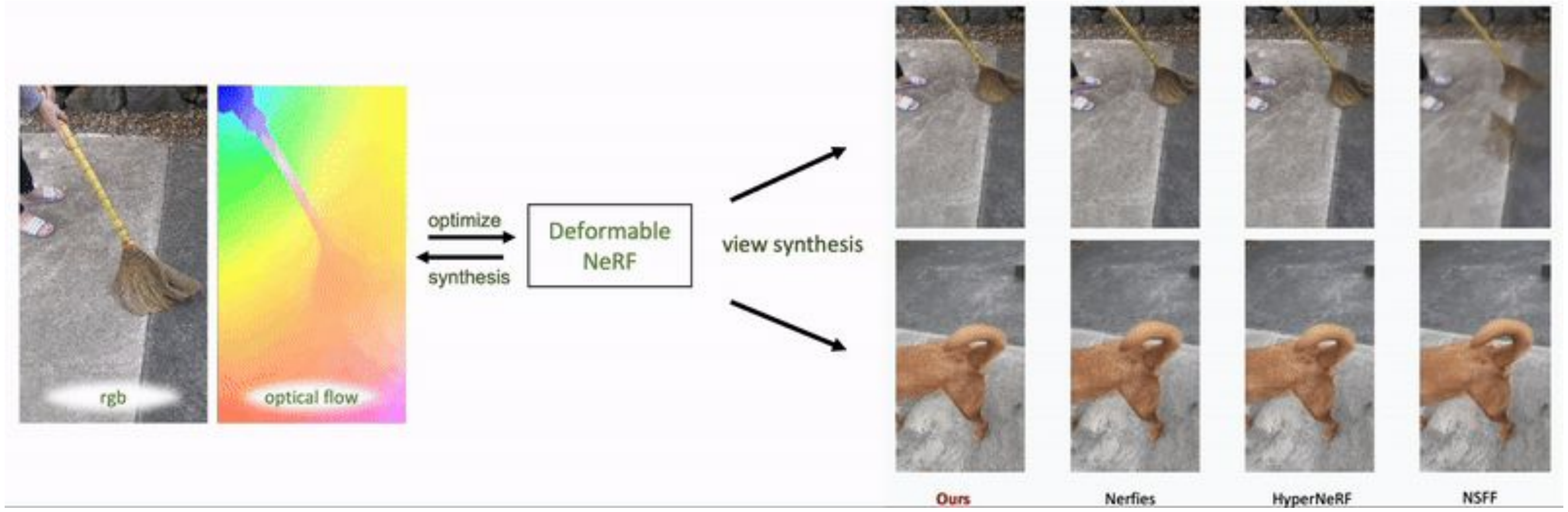


NeRF



Deformable NeRF

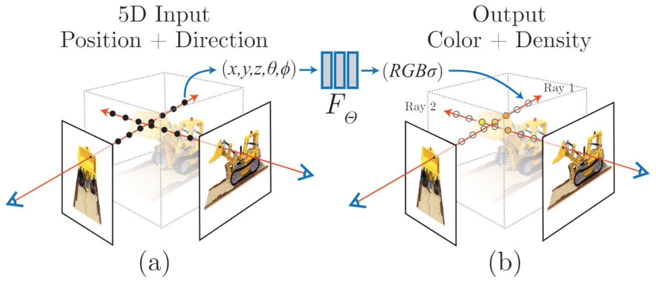
Optical Flow Supervision:



FSD-NeRF

Speed and Quality Advancements

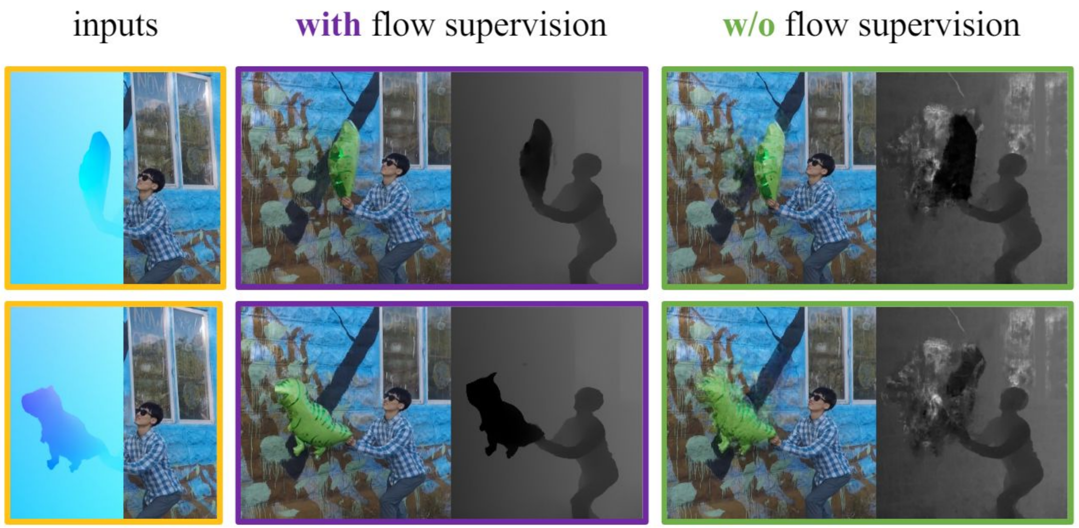
Photo-realistic View Synthesis: Neural Scene Representations



NeRF

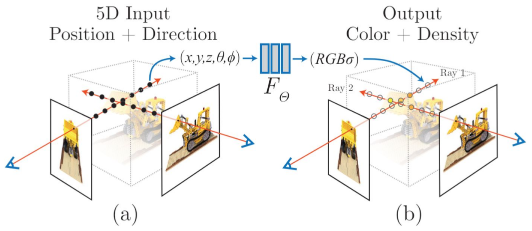
Deformable NeRF

Optical Flow Supervision:



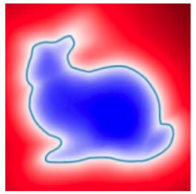
FSD-NeRF

Speed and Quality Advancements
High-fidelity Geometry: Neural Scene Representations



NeRF

+

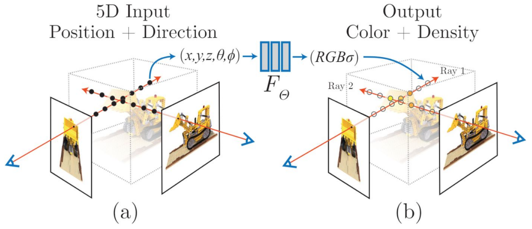


SDF

Dynamic Neural Surface Reconstruction

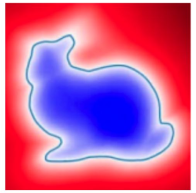
Speed and Quality Advancements

High-fidelity Geometry: Neural Scene Representations



NeRF

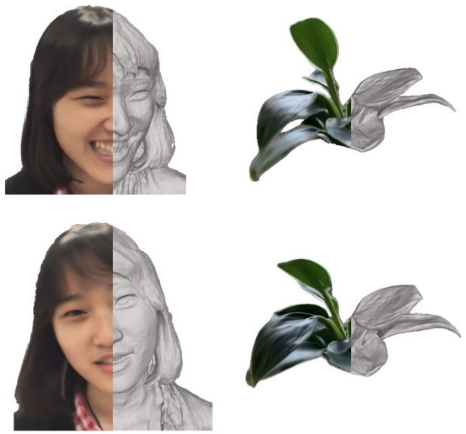
+



SDF

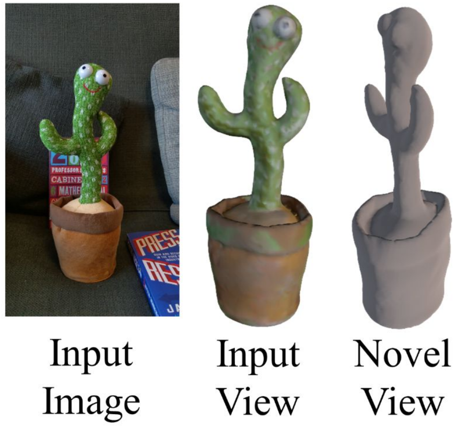
Dynamic Neural Surface Reconstruction

- RGB-D with mask



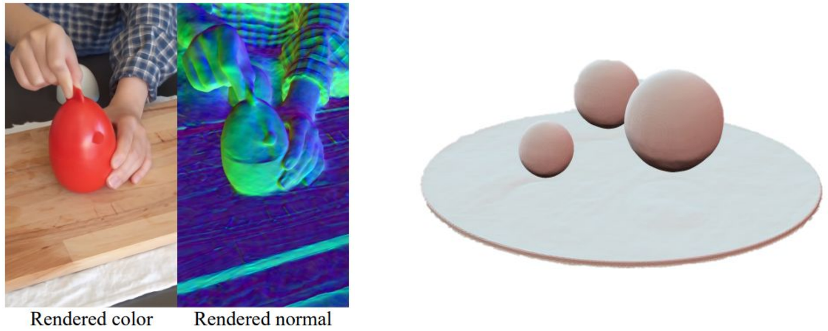
NDR

- RGB with mask and mesh proxy



Unbiased4D

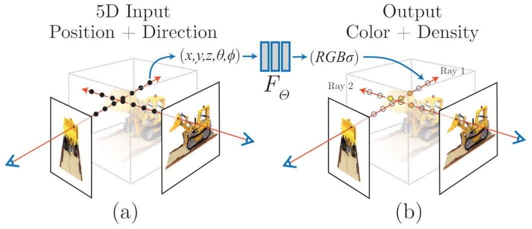
- RGB only



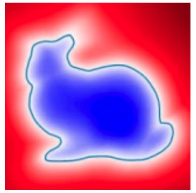
4DRegSDF

Speed and Quality Advancements

High-fidelity Geometry: Neural Scene Representations



+

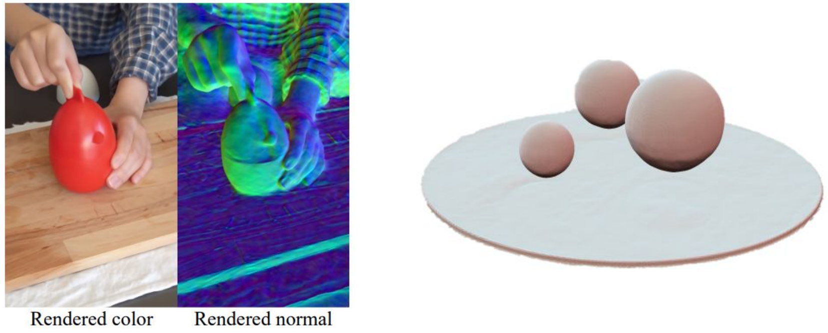


SDF

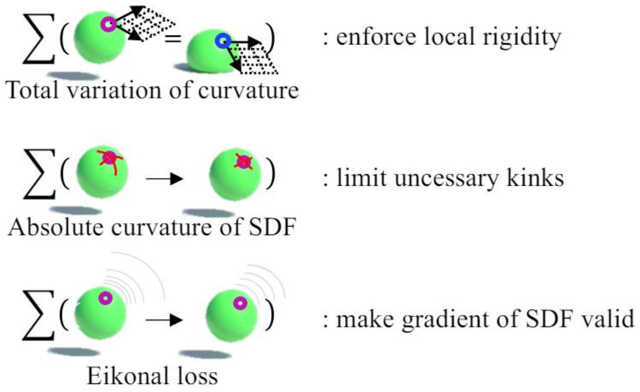
NeRF

Dynamic Neural Surface Reconstruction

- RGB only



4DRegSDF



Speed and Quality Advancements
High-fidelity Geometry: Neural Scene Representations

- Extensions regarding additional inputs, surface reconstruction, model improvements, etc. are usually seen with pure neural fields first

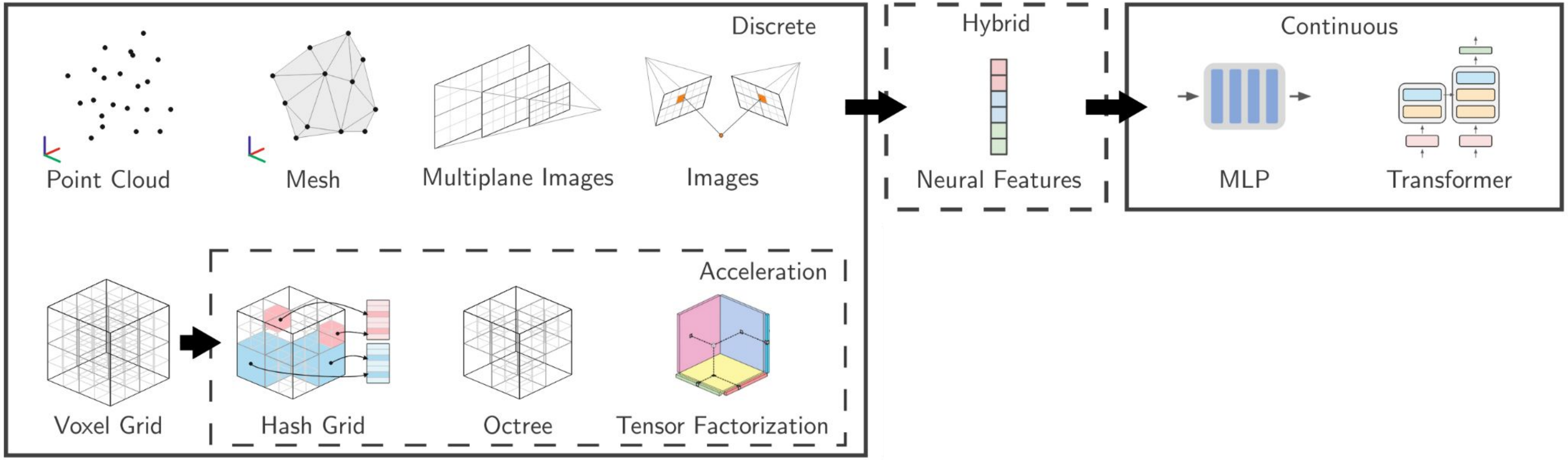

- **Adopts photorealistic reconstruction and view synthesis from rigid setting**
- **Also adopts the slow rendering and training speed**
- **Early methods cannot handle long sequences or novel views that are significantly different than training views**

- RGB-D with mask
- RGB with mask and mesh proxy
- RGB only

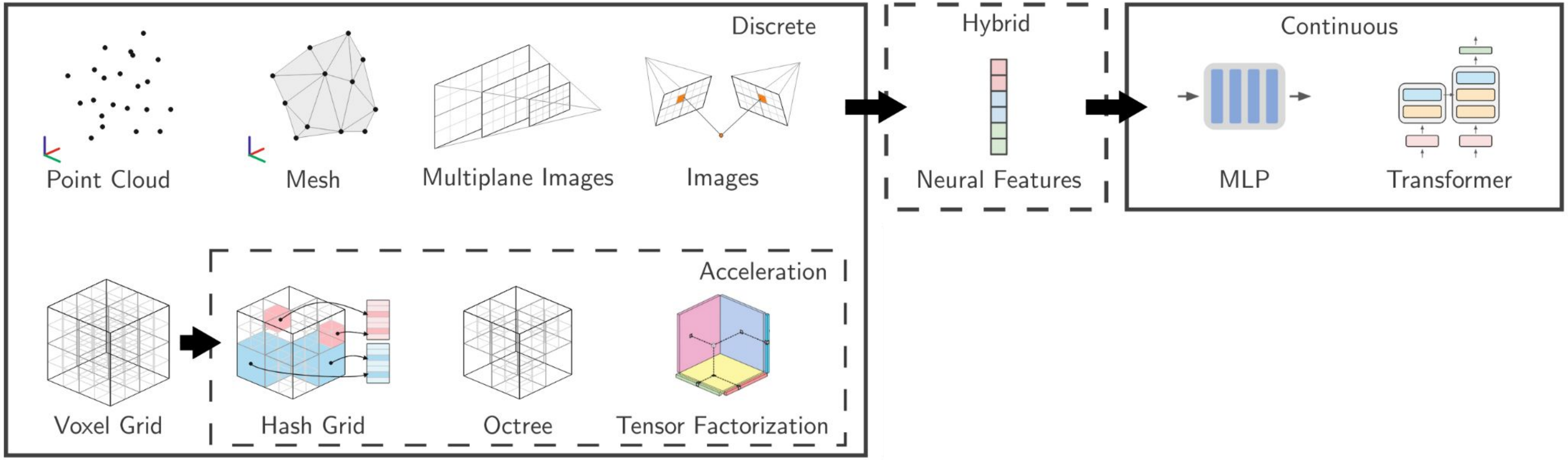
Surface Reconstruction

Speed and Quality Advancements

3D Scene Representations



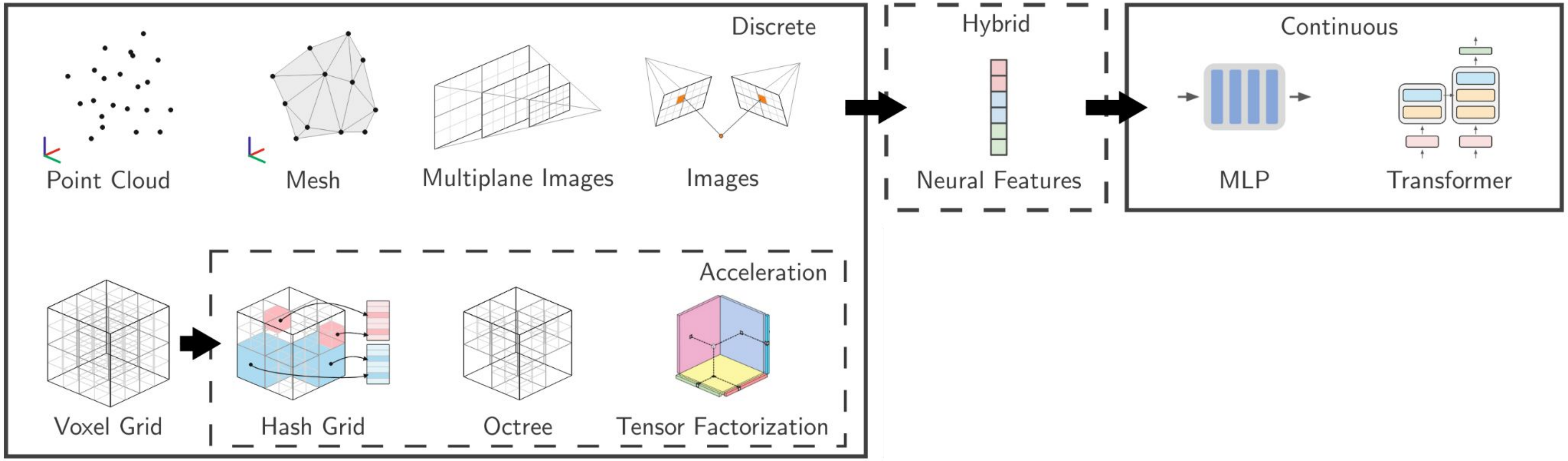
Speed and Quality Advancements
3D Scene Representations



$$\mathbf{y} = \rho(\mathbf{x}, \mathcal{H};\ \theta)\,,$$

Speed and Quality Advancements

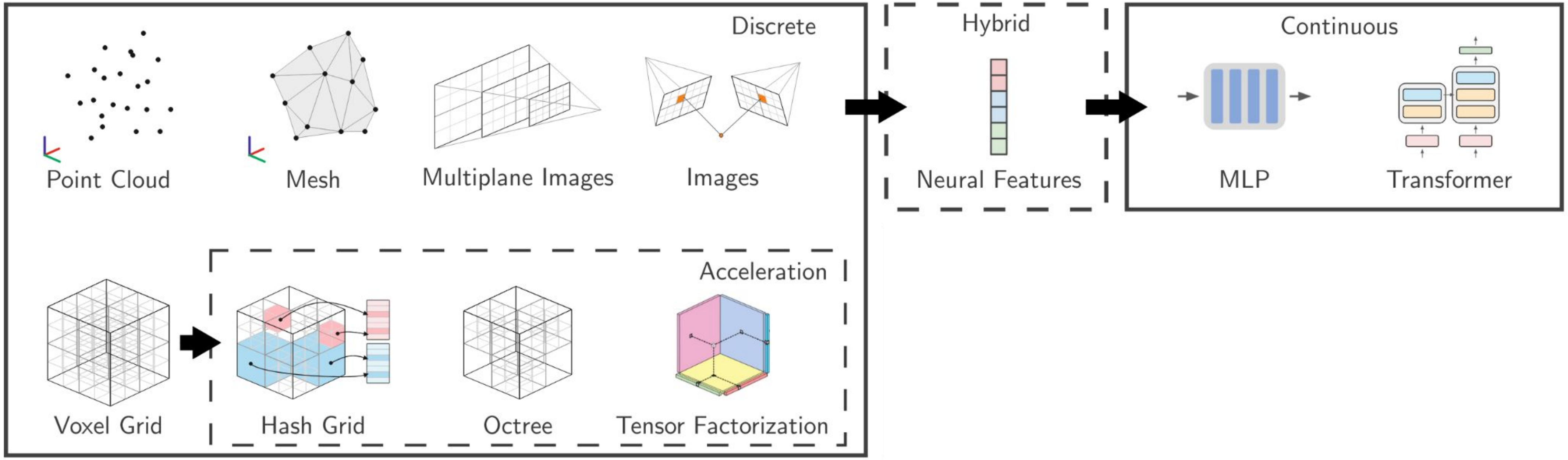
3D Scene Representations



$\mathbf{x} \in \mathbb{R}^3$ are the 3D coordinates

$$\mathbf{y} = \rho(\mathbf{x}, \mathcal{H};\ \theta)\,,$$

Speed and Quality Advancements
3D Scene Representations

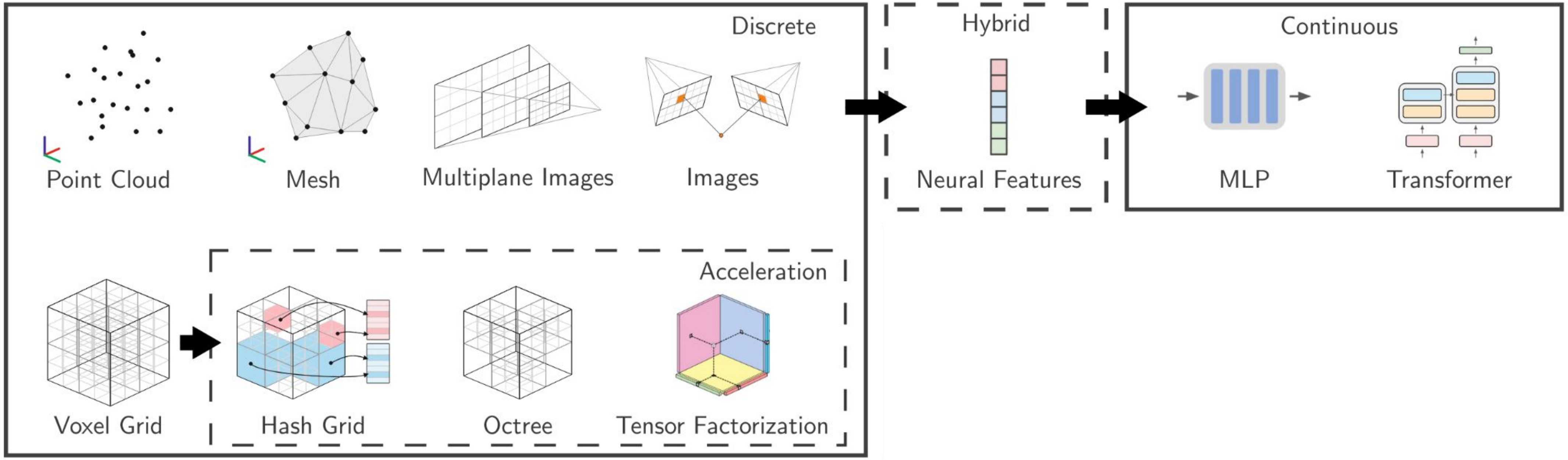


$$\mathbf{y} = \rho(\mathbf{x}, \mathcal{H};\ \theta)\,,$$

$\mathbf{x} \in \mathbb{R}^3$ are the 3D coordinates

$\mathcal{H}$ are optional additional inputs (e.g. view direction)

Speed and Quality Advancements
3D Scene Representations



$\mathbf{x} \in \mathbb{R}^3$ are the 3D coordinates

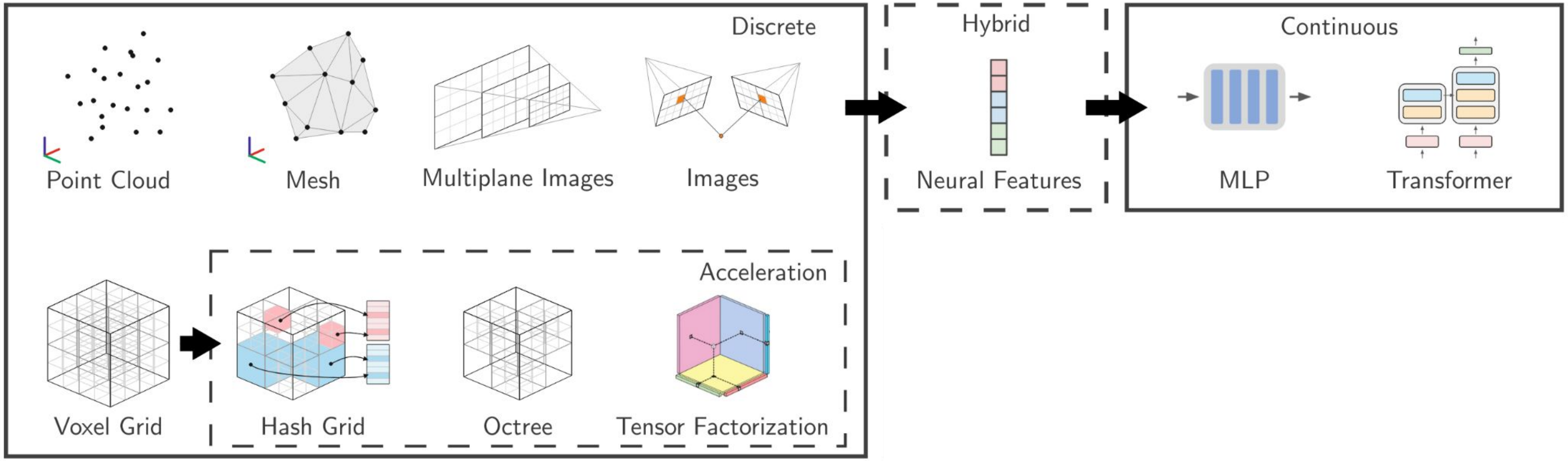
$$\mathbf{y} = \rho(\mathbf{x}, \mathcal{H};\ \theta)\,,$$

$\mathcal{H}$ are optional additional inputs (e.g. view direction)

θ stores the scene information

Speed and Quality Advancements

3D Scene Representations



$$\mathbf{y} = \rho(\mathbf{x}, \mathcal{H};\ \theta)\,,$$

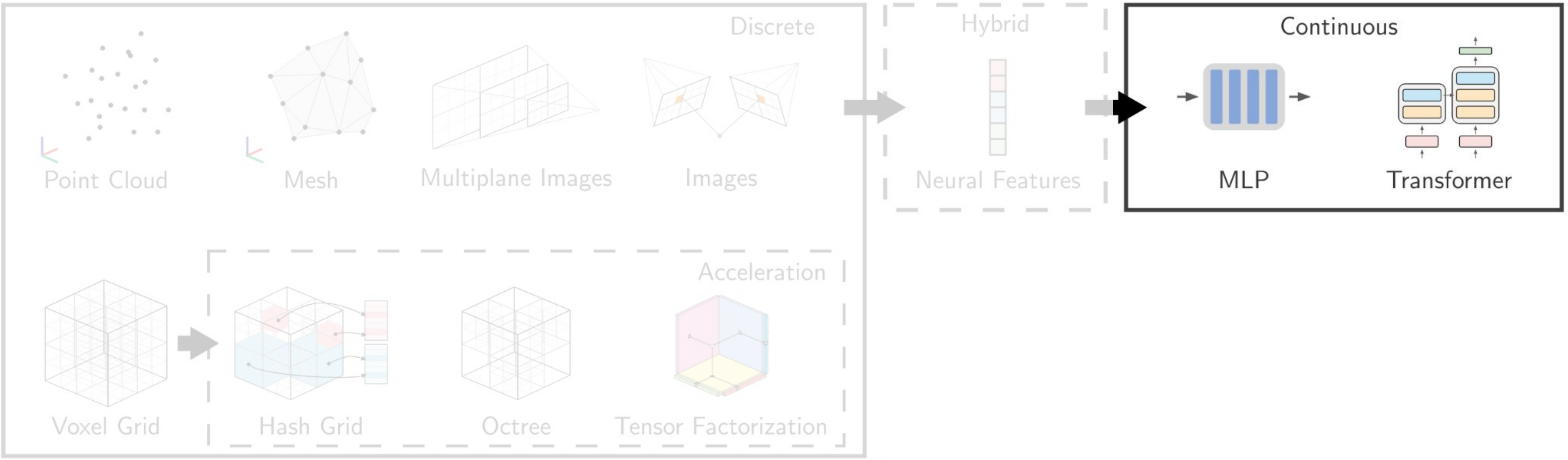
$\mathbf{x} \in \mathbb{R}^3$ are the 3D coordinates

$\mathcal{H}$ are optional additional inputs (e.g. view direction)

θ stores the scene information

$\mathbf{y}$ represents any scene property (e.g. geometry, colour, deformation, etc.)

Speed and Quality Advancements
3D Scene Representations



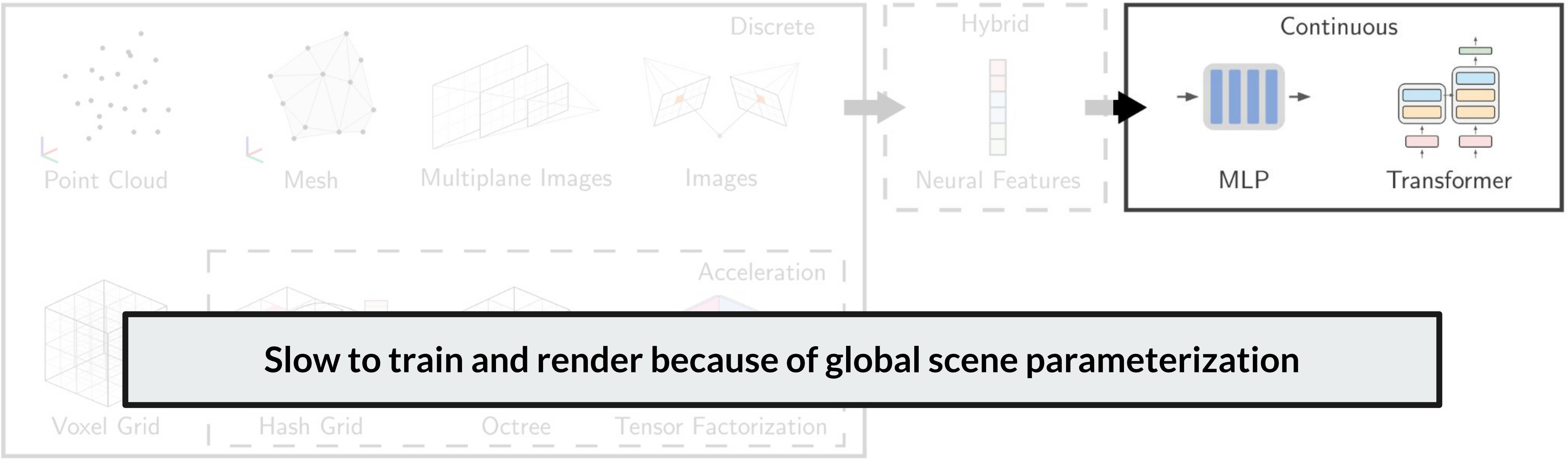
$$\mathbf{y} = \rho(\mathbf{x}, \mathcal{H};\ \theta)\,,$$

θ is stored in network parameters

ρ is an MLP or Transformer

Speed and Quality Advancements

3D Scene Representations

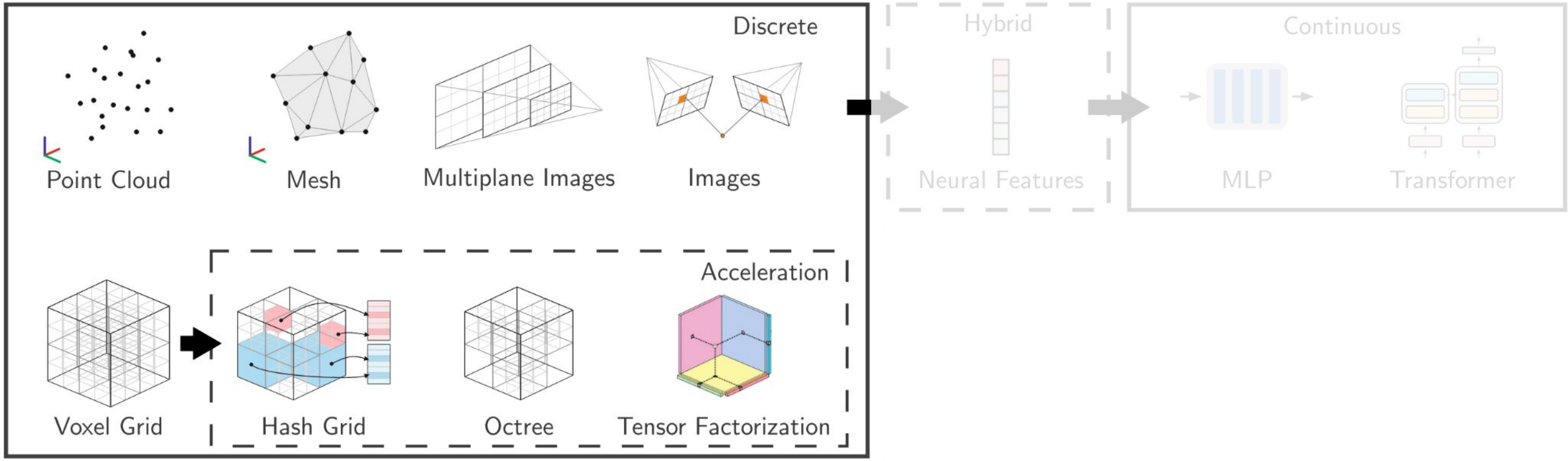


$$\mathbf{y} = \rho(\mathbf{x}, \mathcal{H};\ \theta)\,,$$

θ is stored in network parameters

ρ is an MLP or Transformer

Speed and Quality Advancements
3D Scene Representations

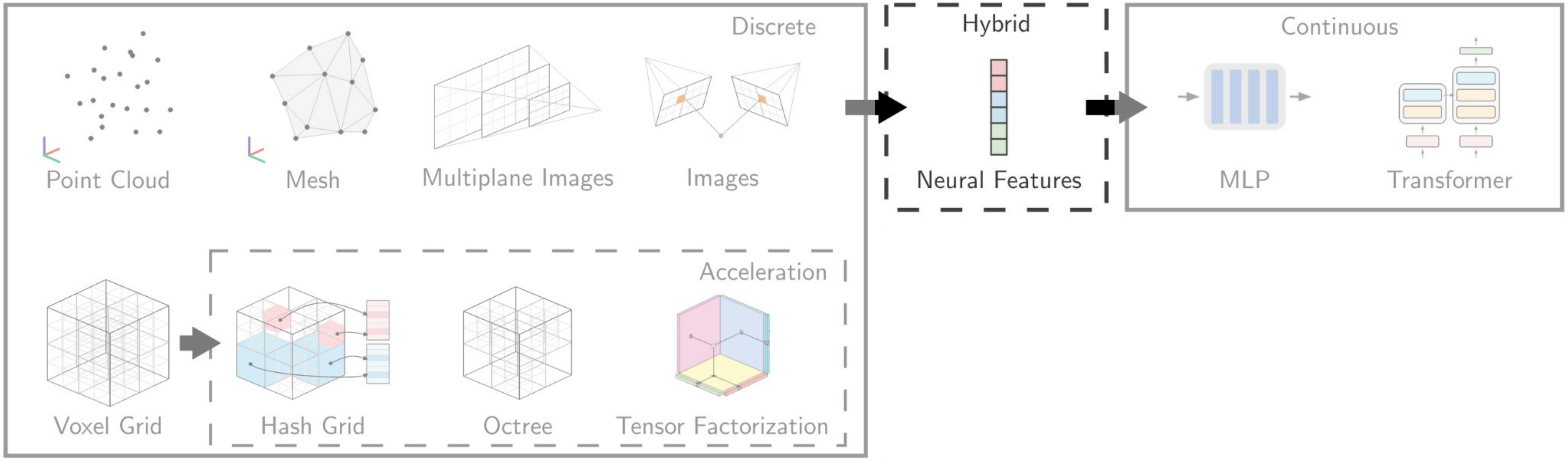


$$\mathbf{y} = \rho(\mathbf{x}, \mathcal{H};\ \theta)\,,$$

θ is stored at discretely defined nodes

ρ interpolates the scene information for any continuous 3D point

Speed and Quality Advancements
3D Scene Representations

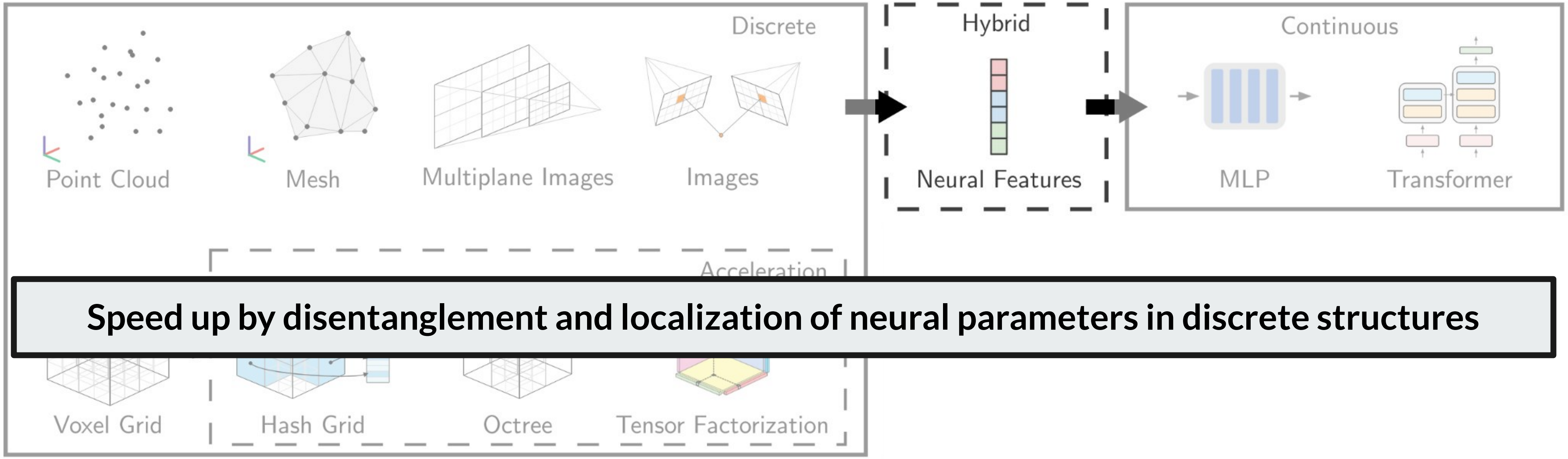


$$\mathbf{y} = \rho(\mathbf{x}, \mathcal{H};\ \theta)\,,$$

- θ are neural features stored in a discrete structure
- ρ defines interpolation of discrete information followed by network query

Speed and Quality Advancements

3D Scene Representations



Speed up by disentanglement and localization of neural parameters in discrete structures

$$\mathbf{y} = \rho(\mathbf{x}, \mathcal{H};\ \theta)\ ,$$

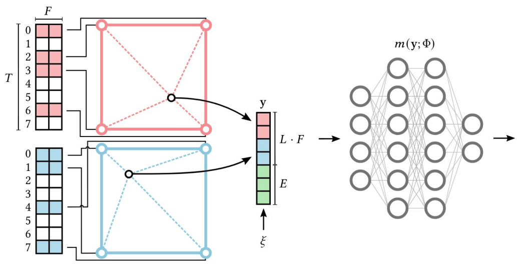
θ are neural features stored in a discrete structure

ρ defines interpolation of discrete information followed by network query

Speed and Quality Advancements

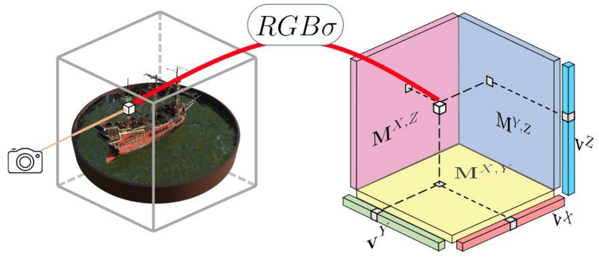
Seminal Hybrid Scene Representations for Rigid Setting

Voxel Grid



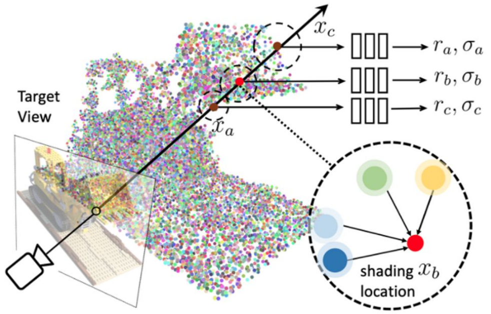
Instant-NGP

Planar Factorization



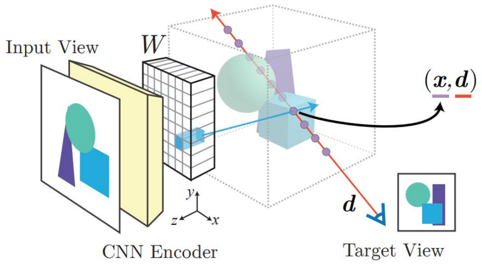
TensoRF

Points



Point-NeRF

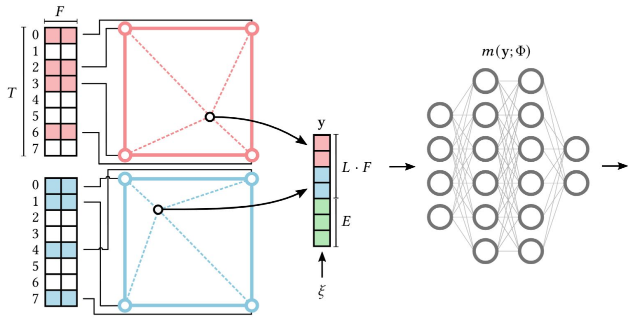
Image-based



pixelNeRF

Speed and Quality Advancements

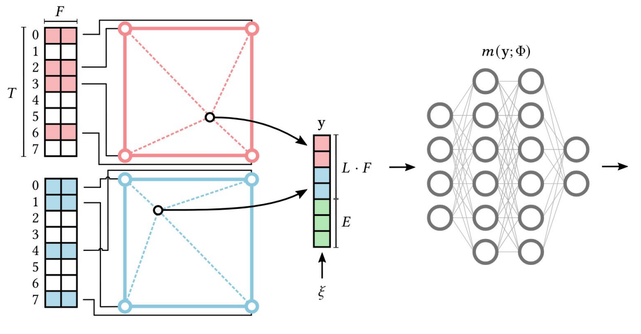
Faster Training and Rendering: Hybrid Neural Scene Representations



Voxel Grid (Instant-NGP)

Speed and Quality Advancements

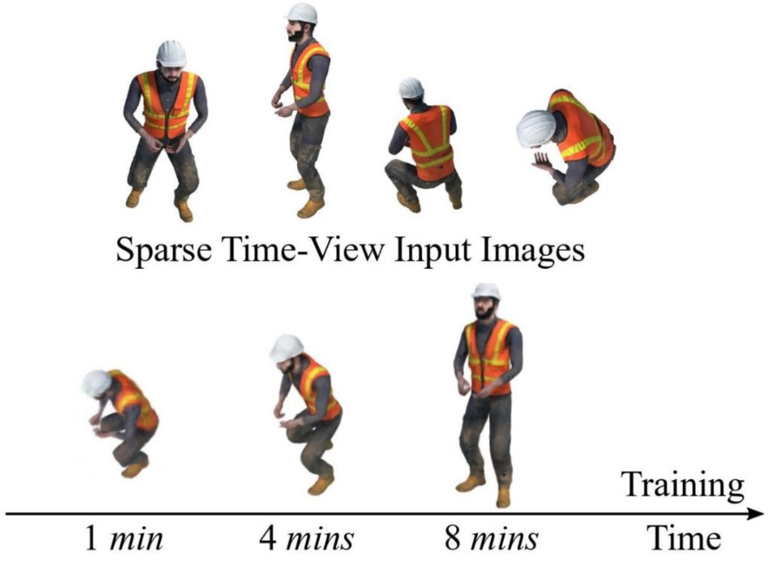
Faster Training and Rendering: Hybrid Neural Scene Representations

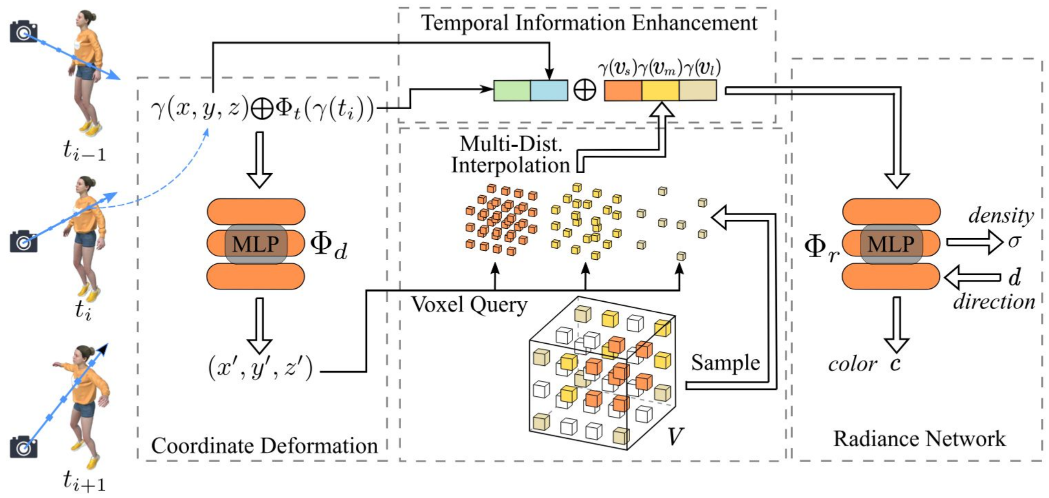


Voxel Grid (Instant-NGP)

Dynamic Voxel NeRF

- Both the deformation field and canonical space are parameterized by MLPs with voxel grids
- Very light deformation MLP for fast training
- Canonical radiance field enhanced through temporal embeddings

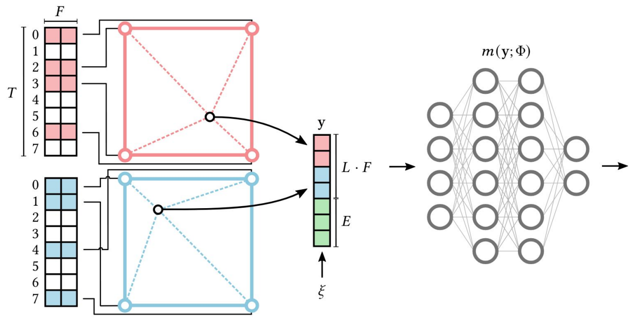




TiNeuVox

Speed and Quality Advancements

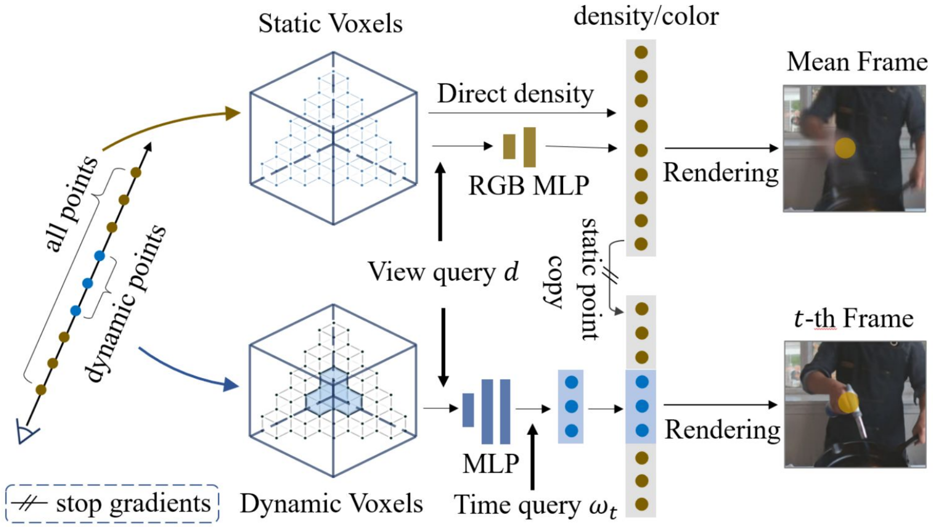
Faster Training and Rendering: Hybrid Neural Scene Representations



Voxel Grid (Instant-NGP)

Dynamic Voxel NeRF

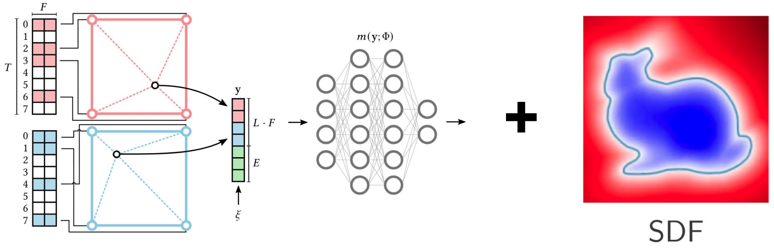
- **Uses static-dynamic decomposition for better scalability and motion modelling capacity**
- **Lightweight static model for fast training from multi-view videos and near real-time rendering**



MixVoxels

Speed and Quality Advancements

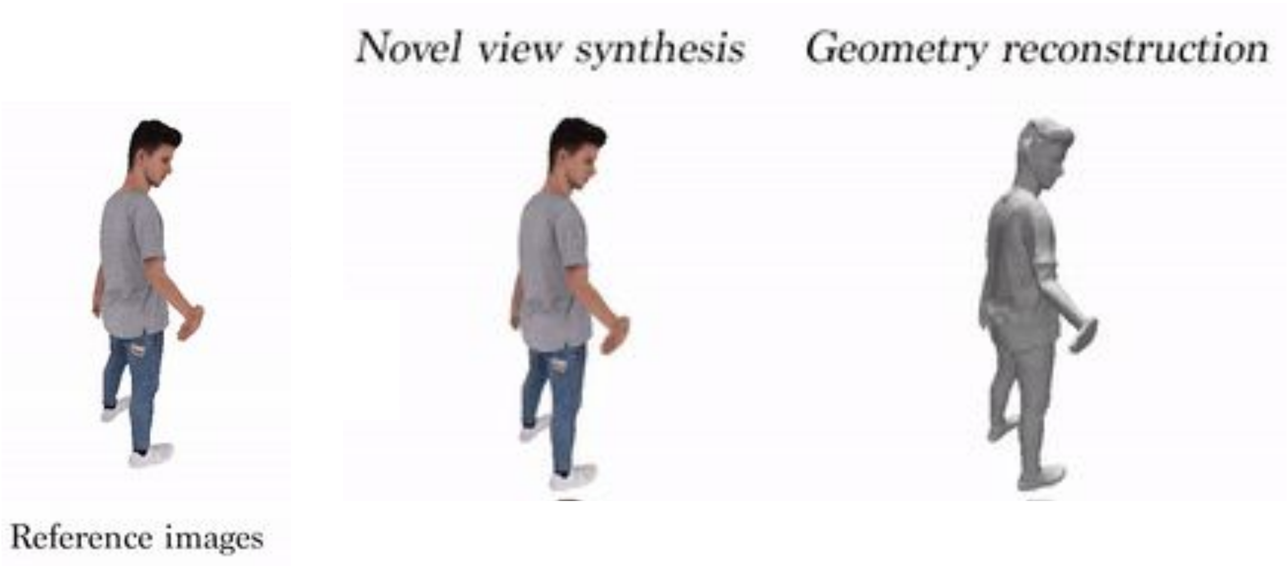
Faster Training and Rendering: Hybrid Neural Scene Representations



Voxel Grid (Instant-NGP)

Dynamic Neural Surface Reconstruction

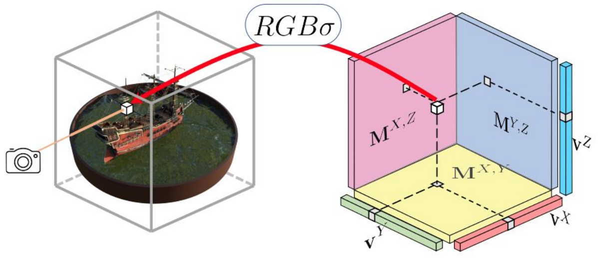
- **Online per-frame optimization from multiple views**
- **Speeds up surface reconstruction while retaining high-quality**



NeuS2

Speed and Quality Advancements

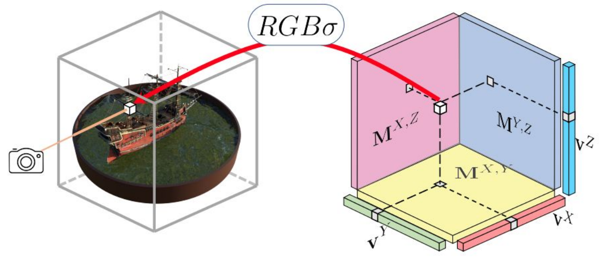
Faster Training and Rendering: Hybrid Neural Scene Representations



Planar Factorization (TensoRF)

Speed and Quality Advancements

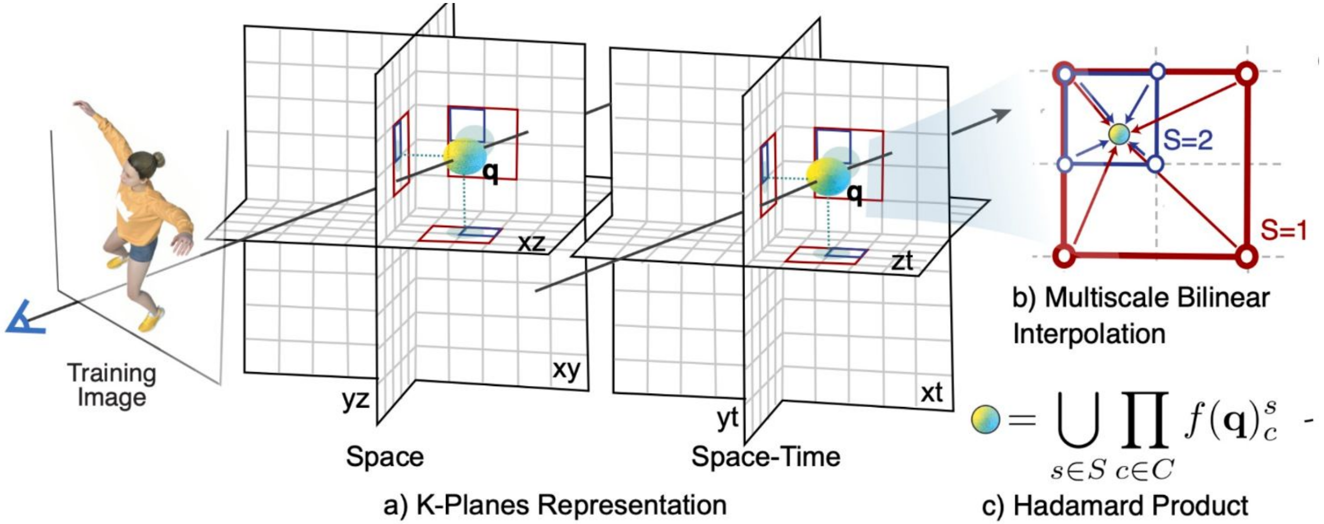
Faster Training and Rendering: Hybrid Neural Scene Representations



Planar Factorization (TensoRF)

Space-Time Planar NeRF

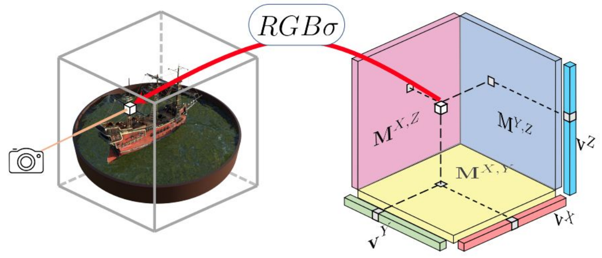
Compaction of voxel grid for memory efficiency



K-Planes

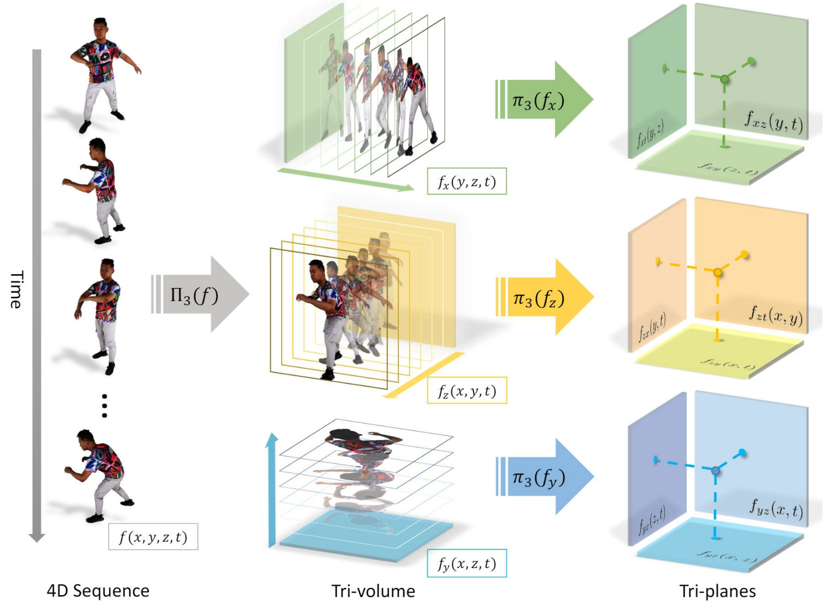
Speed and Quality Advancements

Faster Training and Rendering: Hybrid Neural Scene Representations



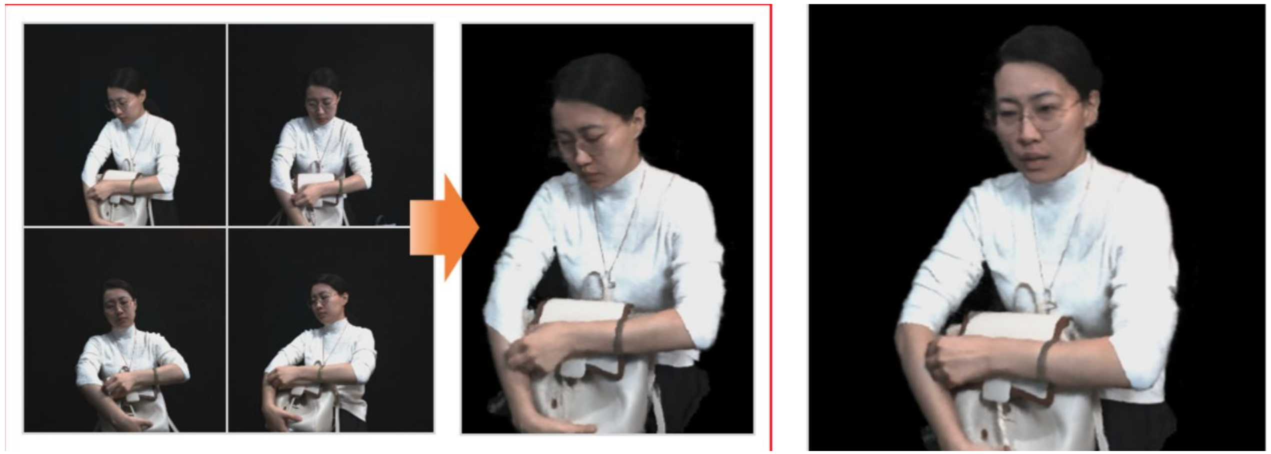
Planar Factorization (TensoRF)

Tensor4D



Space-Time Planar NeRF

- Coarse-to-fine hierarchical decomposition policy
- Results in 9 planes which can model finer details
- High-fidelity reconstruction from sparse multi-views



Speed and Quality Advancements

High-quality Rendering: Hybrid Neural Scene Representations

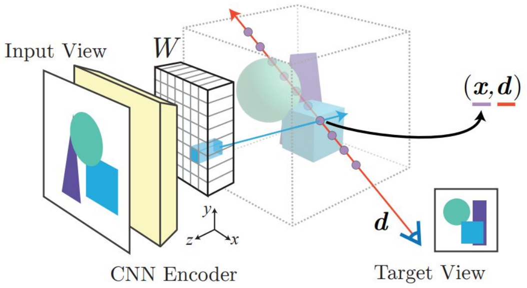


Image-based (pixelNeRF)

Speed and Quality Advancements

High-quality Rendering: Hybrid Neural Scene Representations

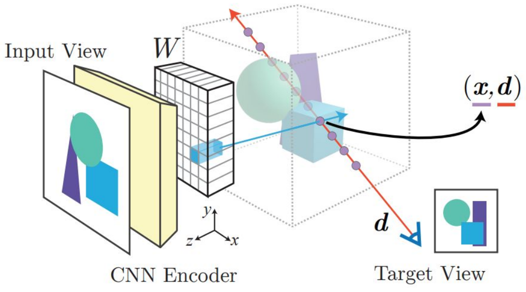
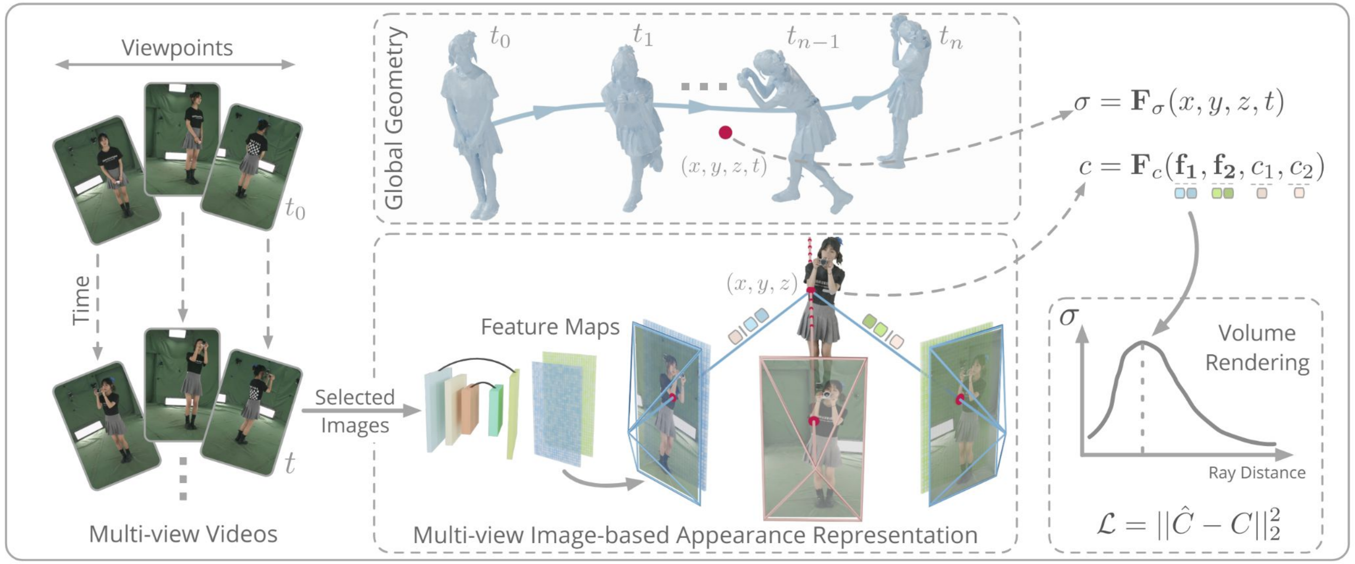


Image-based (pixelNeRF)

Image-based Dynamic NeRF

- **Image features for fine appearance details**
- **Aggregated from multiple views**
- **High-resolution rendering possible with high resolution training images, upto 4K!**



Im4D

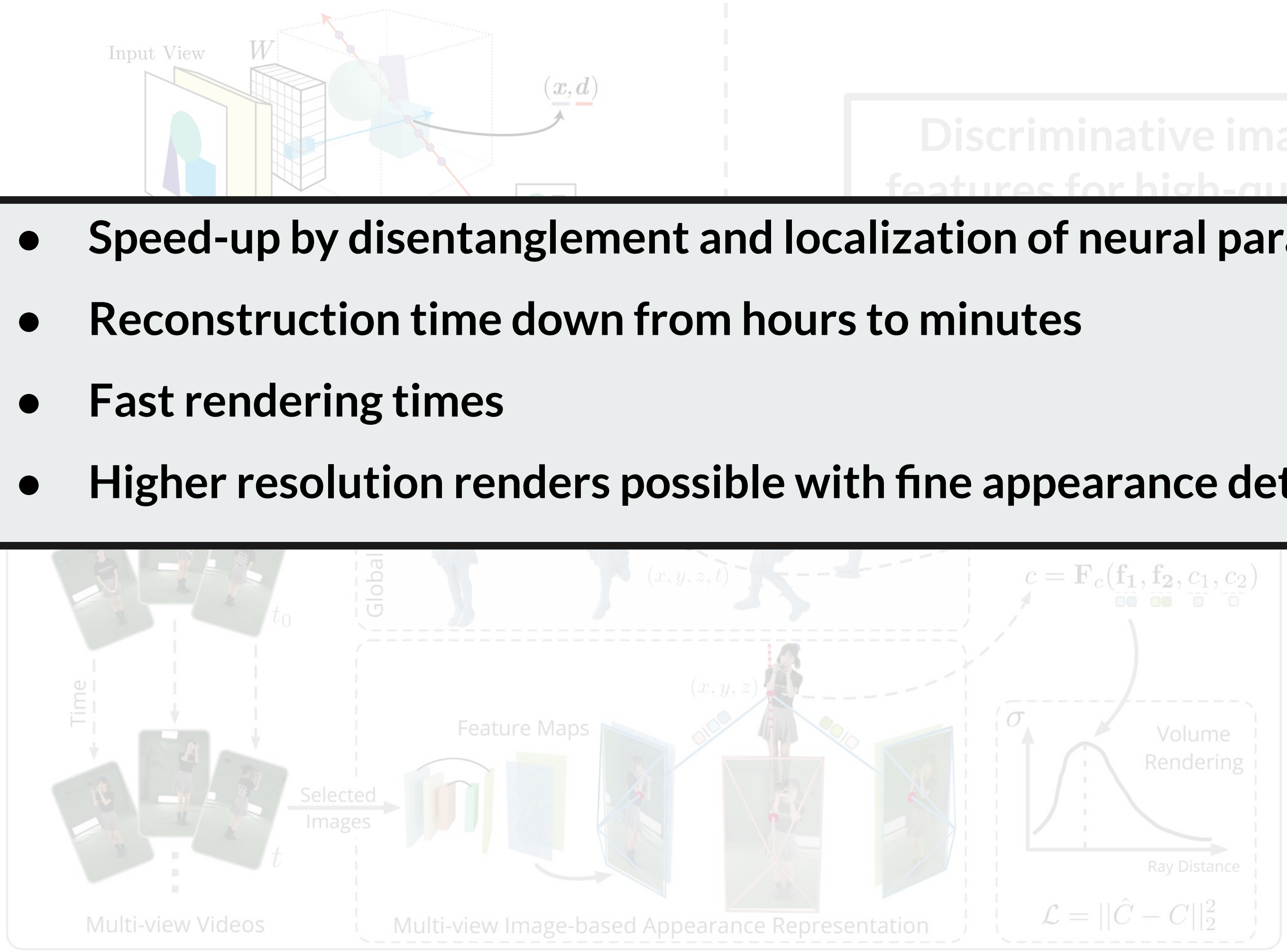
4K4D

Speed and Quality Advancements

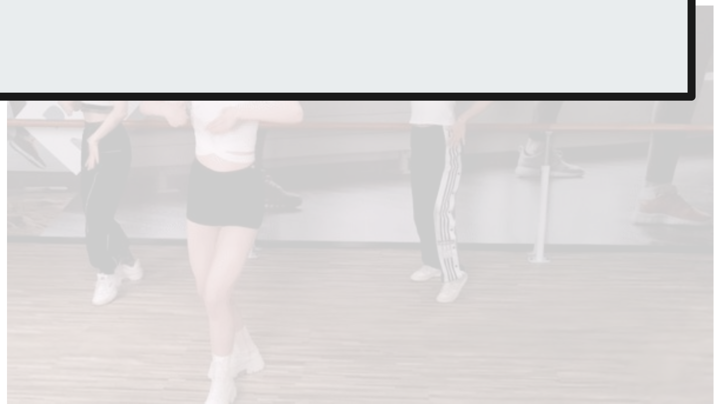
High-quality Rendering: Hybrid Neural Scene Representations

Image-based Dynamic NeRF

- Speed-up by disentanglement and localization of neural parameters in discrete structures
- Reconstruction time down from hours to minutes
- Fast rendering times
- Higher resolution renders possible with fine appearance details



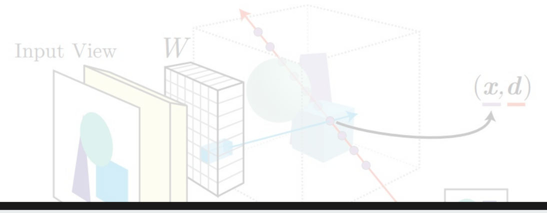
Im4D



4K4D

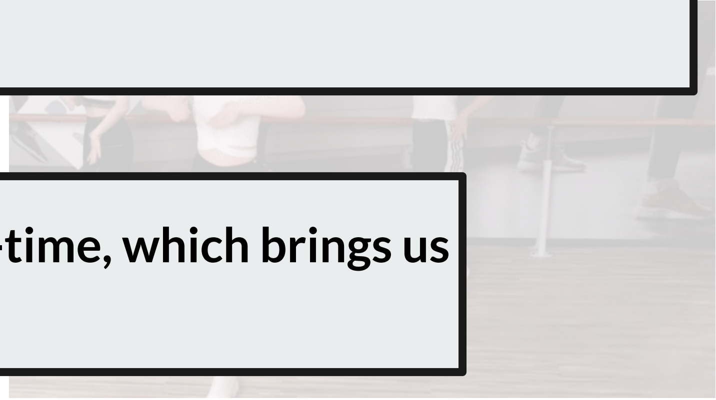
Speed and Quality Advancements

High-quality Rendering: Hybrid Neural Scene Representations



- Speed-up by disentanglement and localization of neural parameters in discrete structures
- Reconstruction time down from hours to minutes
- Fast rendering times
- Higher resolution renders possible with fine appearance details

Im4D

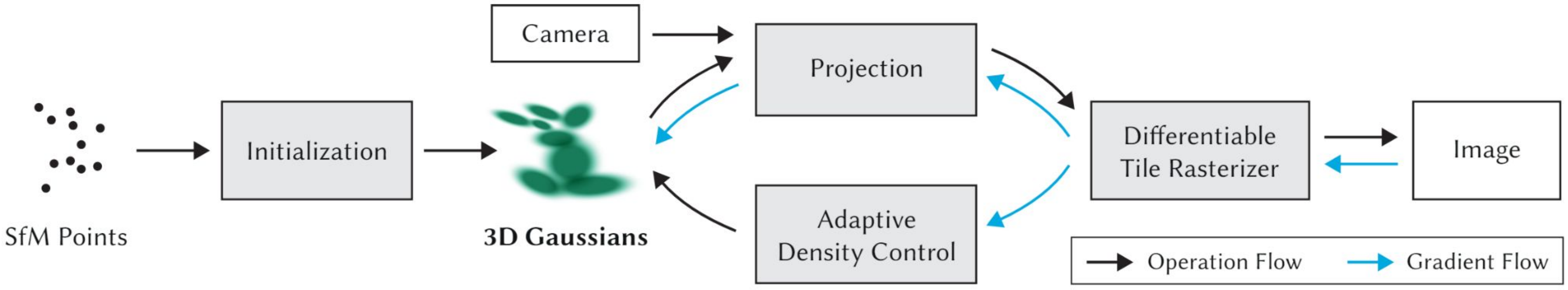


4K4D

Rendering is fast for hybrid representations but seldom real-time, which brings us to the next major breakthrough!

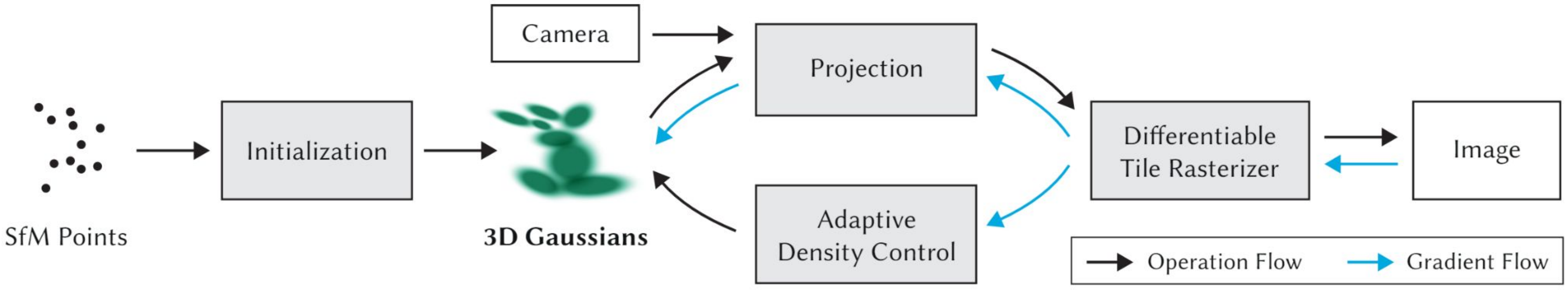
Speed and Quality Advancements

Real-time Rendering: 3D Gaussian Splatting



3D Gaussian Splatting

Speed and Quality Advancements
Real-time Rendering: 3D Gaussian Splatting

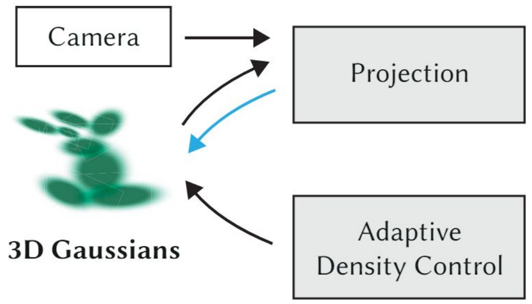


3D Gaussian Splatting

- Each 3D Gaussian stores position, rotation, scale and spherical harmonics coefficients, which are optimized from images using a fast tile-based rasterizer
- Much faster than volume rendering, enabling real-time performance

Speed and Quality Advancements

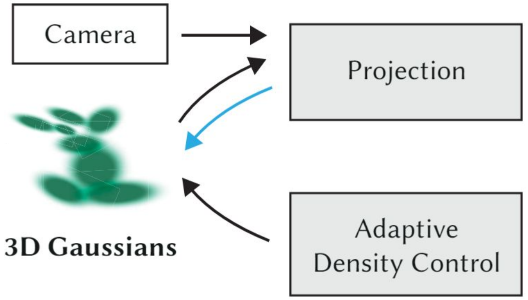
Real-time Rendering: 3D Gaussian Splatting



3D Gaussian Splatting

Speed and Quality Advancements

Real-time Rendering: 3D Gaussian Splatting

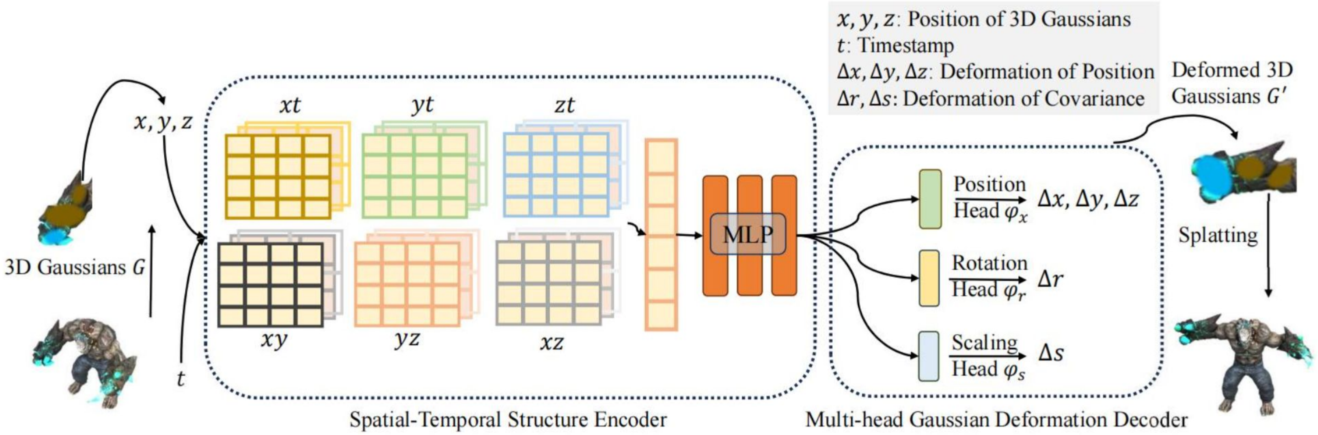


3D Gaussian Splatting

Deformable 3D Gaussian Splatting

- **Deform position, rotation and scale of canonical Gaussians to fit each time-step**
- **Upto 80 FPS rendering speed**

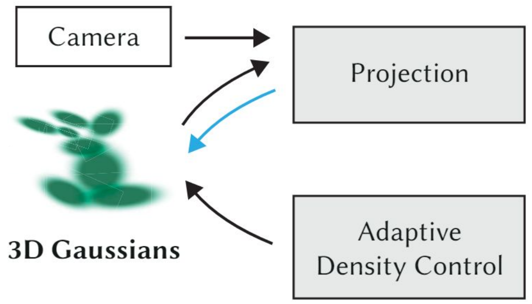
Deformable3DGS



4D-GS

Speed and Quality Advancements

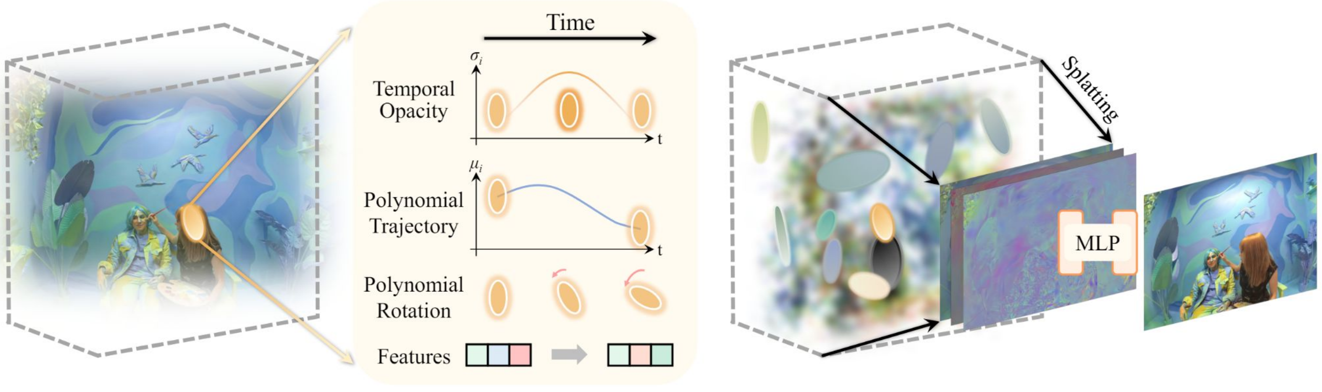
Real-time Rendering: 3D Gaussian Splatting



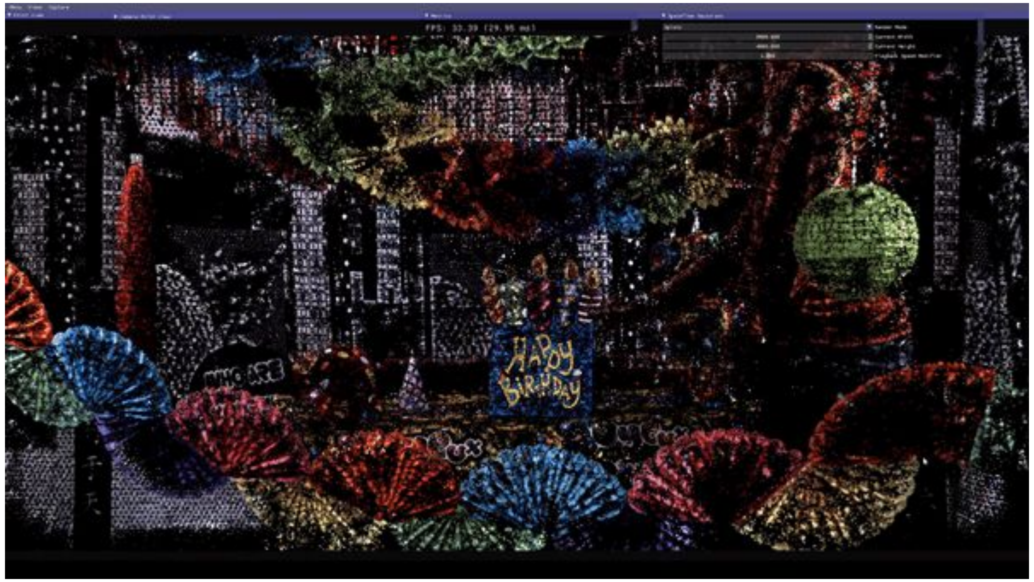
3D Gaussian Splatting

Space-time 3D Gaussian Splatting

- Extra 1D Gaussian added to 3D Gaussians
- Features instead of SHs, with an MLP to convert them into an RGB image after splatting
- 8K video rendering at 66 FPS!

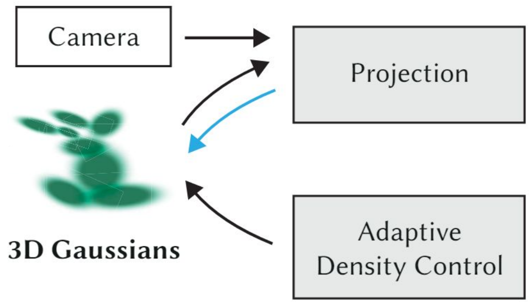


(a) Spacetime Gaussians (b) Feature Splatting and Rendering



SpacetimeGaussians

Speed and Quality Advancements
Real-time Rendering: 3D Gaussian Splatting

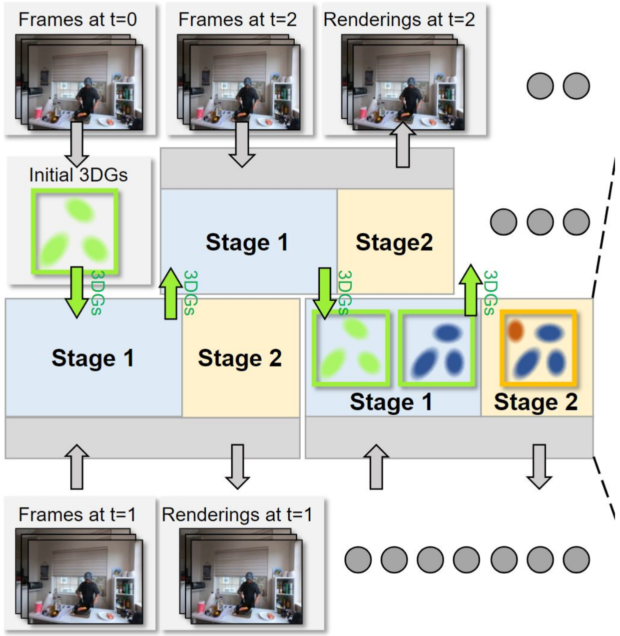


3D Gaussian Splatting

Streamable 3D Gaussian Splatting

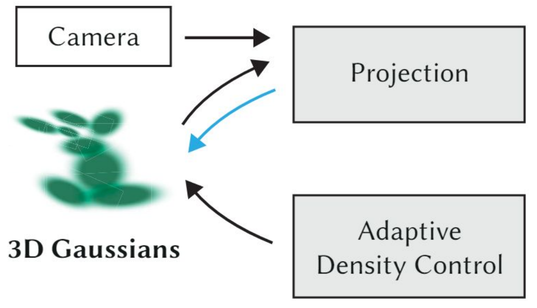
- **Online Reconstruction from multi-view videos**
- **Multi-resolution neural hash-grid as a cache for transformation**
- **Additional frame-specific Gaussian spawning**

3DGStream



Speed and Quality Advancements

Real-time Rendering: 3D Gaussian Splatting

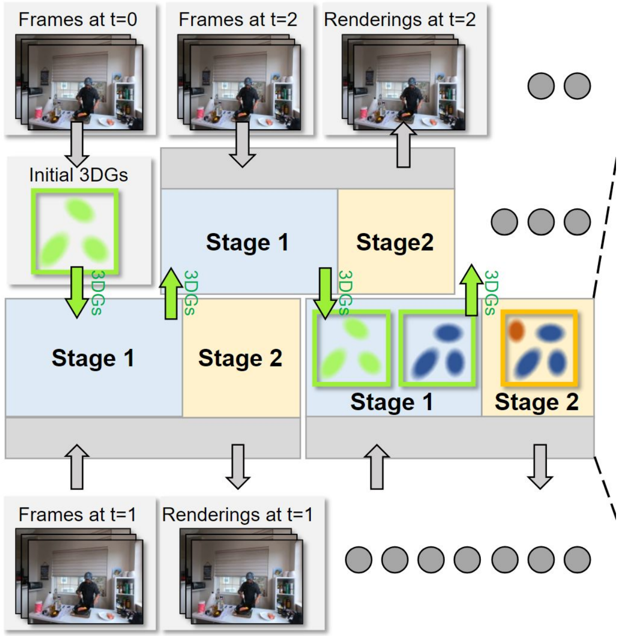


3D Gaussian Splatting

Streamable 3D Gaussian Splatting

- Online Reconstruction from multi-view videos
- Multi-resolution neural hash-grid as a cache for transformation
- Additional frame-specific Gaussian spawning

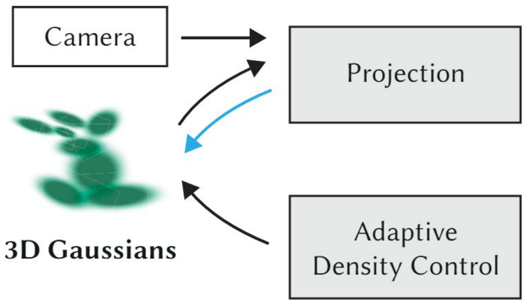
3DGStream



Reconstruction in 12 seconds with up to 200 FPS rendering speed!

Speed and Quality Advancements

Real-time Rendering: 3D Gaussian Splatting



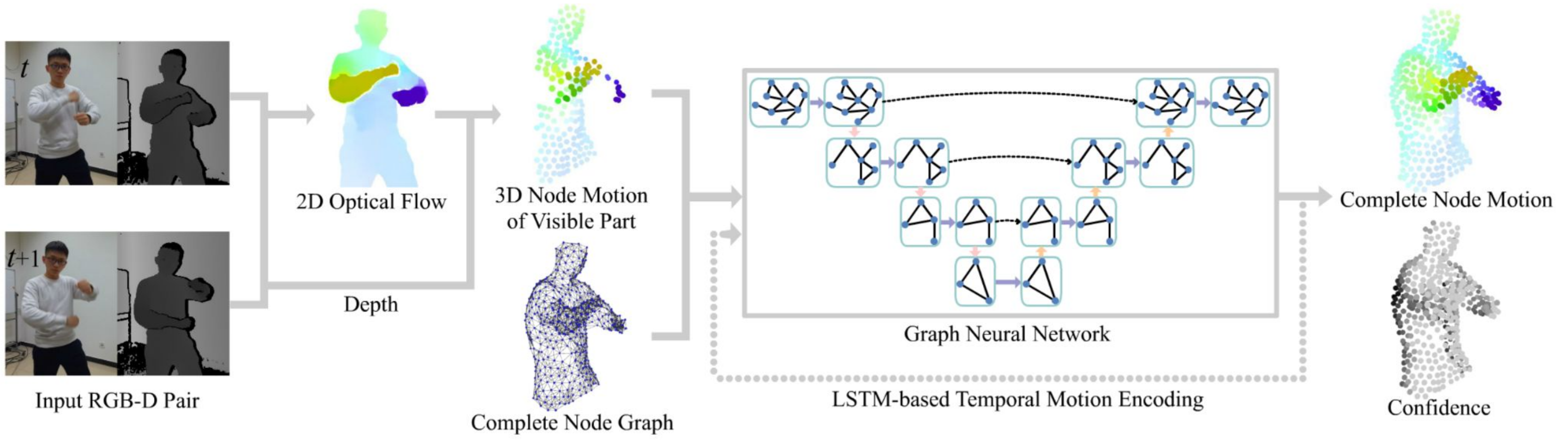
3D Gaussian Splatting

Even though it is getting close, real-time reconstruction is still only possible using classical representations

Speed and Quality Advancements

Real-time Reconstruction: Classical Representations

- Registers RGB-D frames into a canonical TSDF grid
- Uses a mesh-based deformation graph to track deformation of canonical frame to each timestep
- Pre-trains a GNN to predict motion of occluded regions from the visible motion
- Geometry only!



Live Demo

Occlusion Fusion

Trends

1. Speed and Quality Advancements

2. Handling of Large Defor

5 minute break!

-term 3D correspondences

3. Modelling Articulated Motion for General Objects

Trends

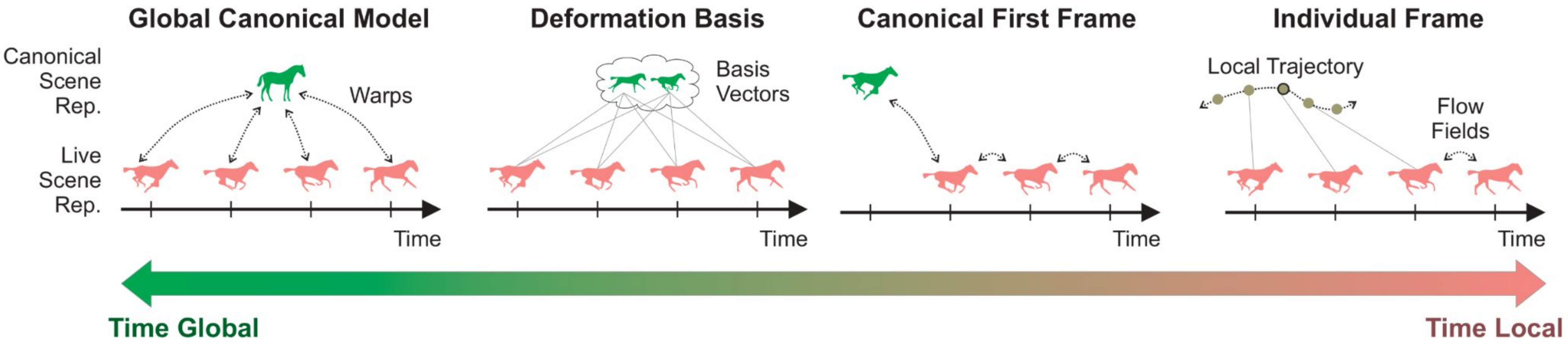
1. Speed and Quality Advancements

2. Handling of Large Deformations / Long-Term 3D Correspondences

3. Modelling Articulated Motion for General Objects

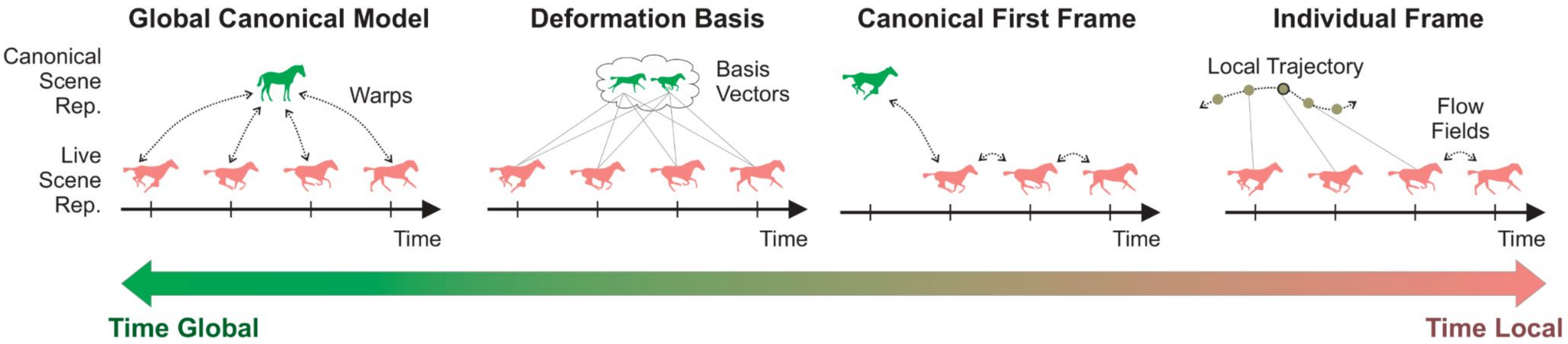
Large Motion vs. Long-term 3D Correspondences

Spatio-Temporal Modelling



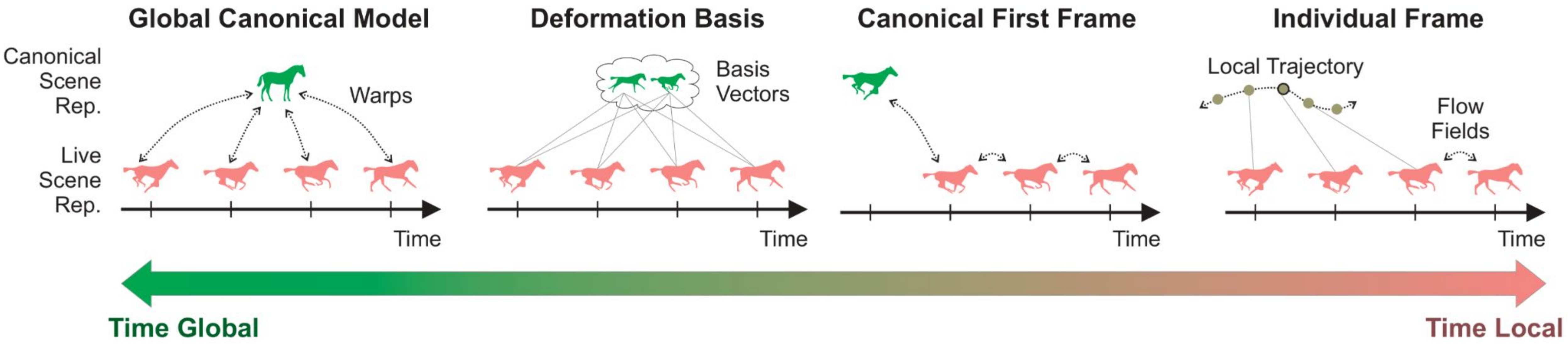
Design choice determines the trade-off between time consistency and large motion modelling!

Large Motion vs. Long-term 3D Correspondences
Spatio-Temporal Modelling



Time consistency enables applications like 3D editing and virtual asset creation

Large Motion vs. Long-term 3D Correspondences
Spatio-Temporal Modelling

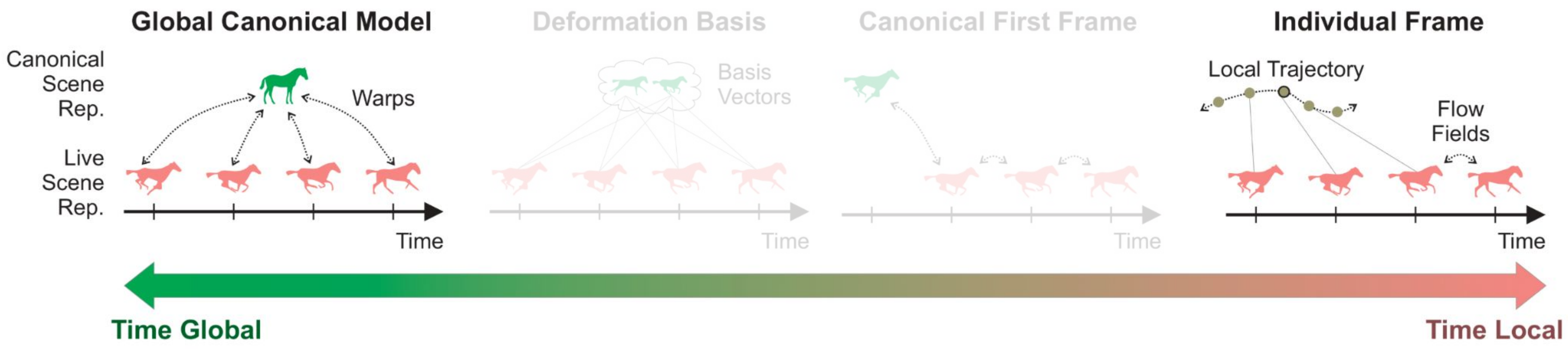


Time consistency enables applications like 3D editing and virtual asset creation

But we don't want to compromise on motion modelling

Large Motion vs. Long-term 3D Correspondences

Spatio-Temporal Modelling



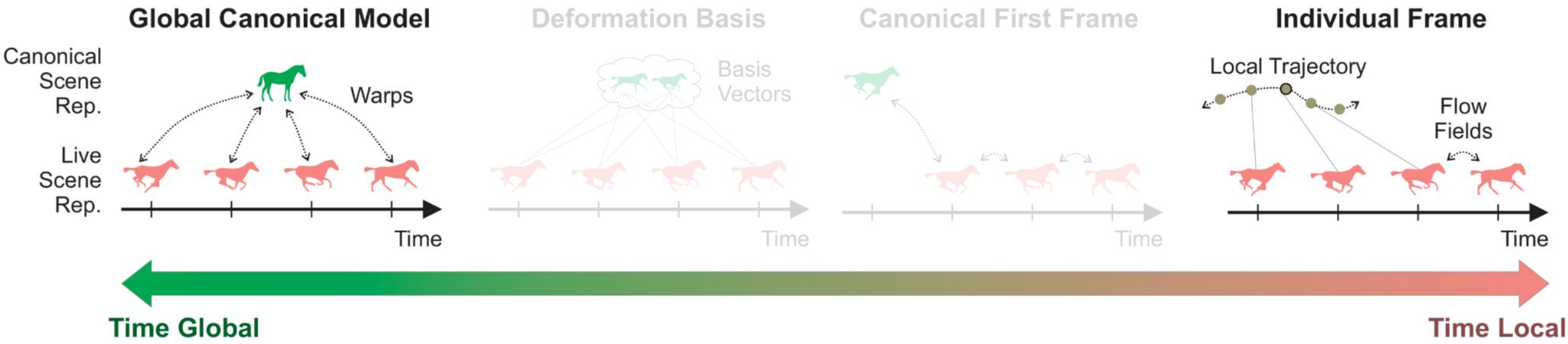
- Global

Temporal Correspondences

- Optional and local

Large Motion vs. Long-term 3D Correspondences

Spatio-Temporal Modelling



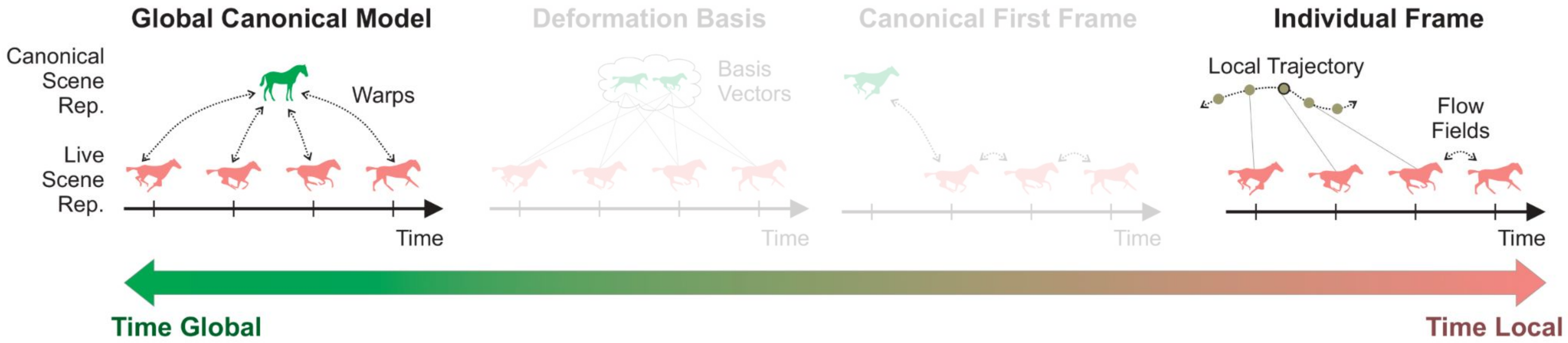
- Global
- Restricted

Temporal Correspondences

Motion Modelling

- Optional and local
- Empirically larger

Large Motion vs. Long-term 3D Correspondences
Spatio-Temporal Modelling



- Global
- Restricted
- Difficult; cannot handle discontinuities

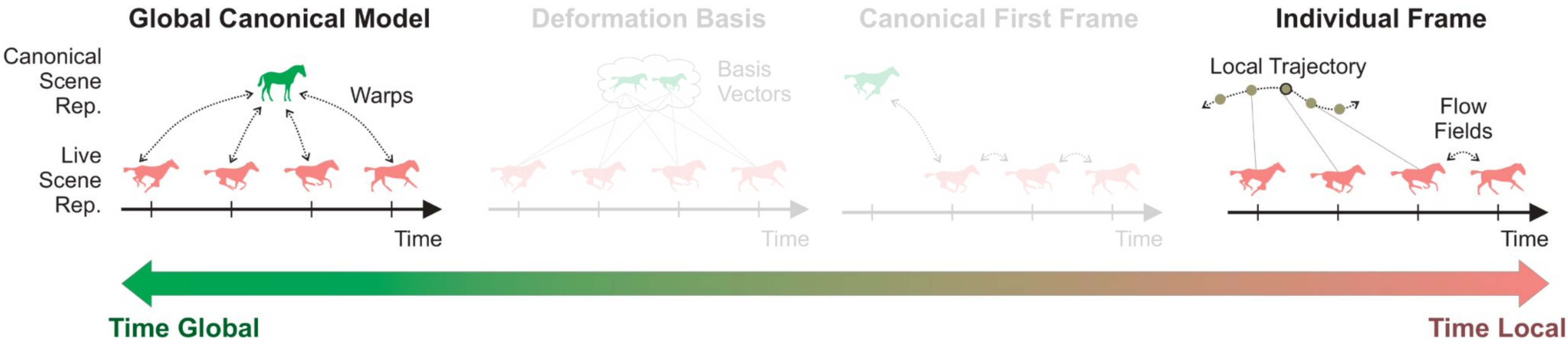
Temporal Correspondences

Motion Modelling

Topology Changes

- Optional and local
- Empirically larger
- Straightforward

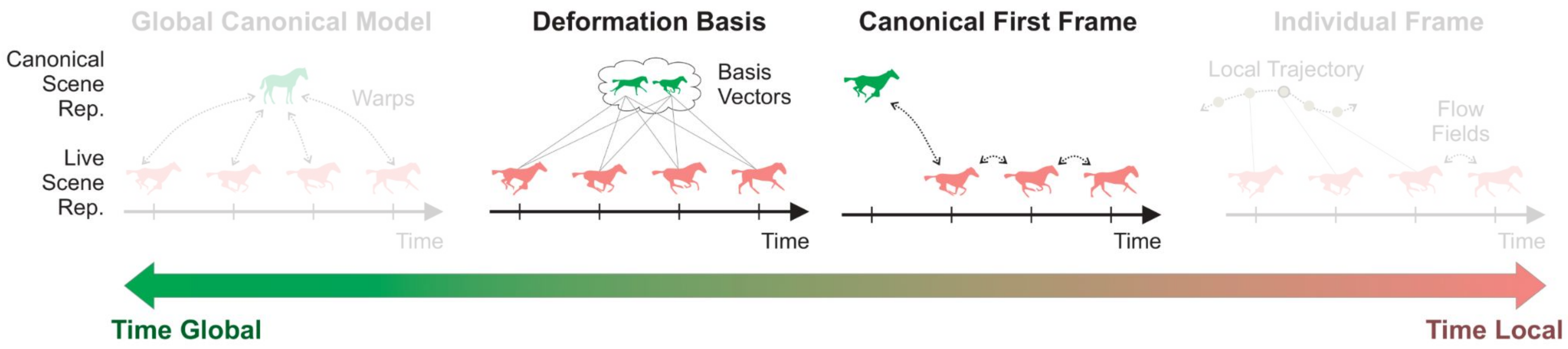
Large Motion vs. Long-term 3D Correspondences
Spatio-Temporal Modelling



Time Global		Time Local
• Global	**Temporal Correspondences**	• Optional and local
• Restricted	**Motion Modelling**	• Empirically larger
• Difficult; cannot handle discontinuities	**Topology Changes**	• Straightforward
• Difficult; canonical model is time-independent	**Appearance Changes**	• Straightforward

Large Motion vs. Long-term 3D Correspondences

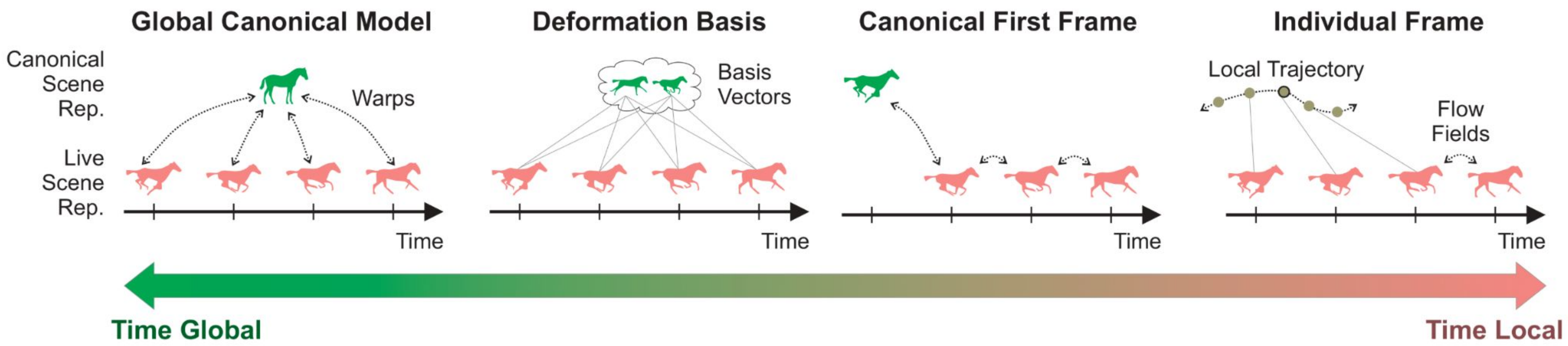
Spatio-Temporal Modelling



Trade-offs to balance the best of both worlds!

Large Motion vs. Long-term 3D Correspondences

Spatio-Temporal Modelling

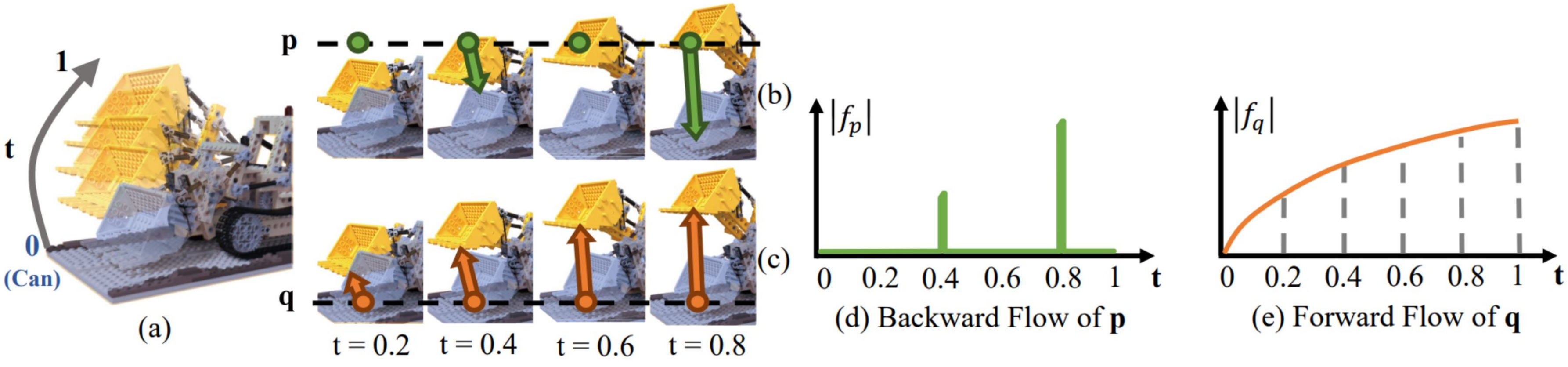


Let's look at some improvements for each type of modelling in recent years

Large Motion vs. Long-term 3D Correspondences
Spatio-Temporal Modelling

Forward Flow Modelling:

- Deformations are modelled from canonical to live frame for smooth and continuous motion model learning
- Enabled by a voxel-based canonical field for discrete forward warping
- Give point trajectories for each point in canonical space



ForwardFlowDNeRF

Time Global

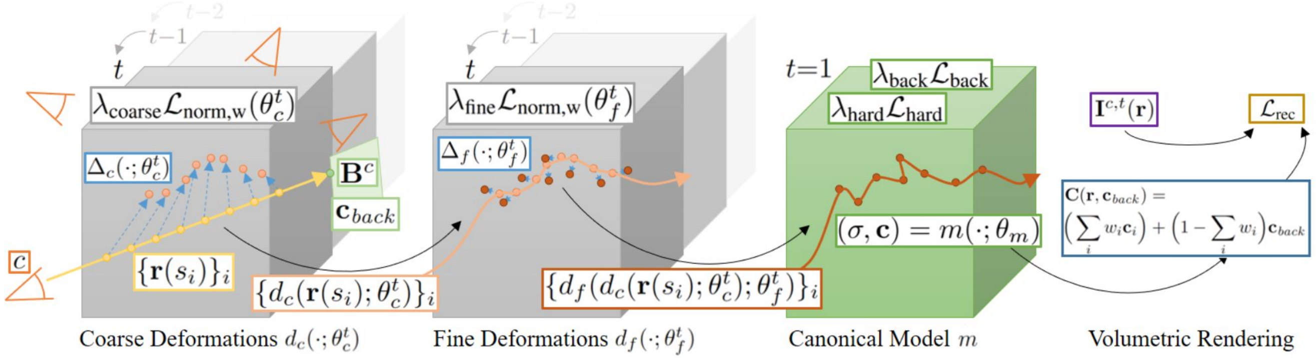
Global Canonical Model

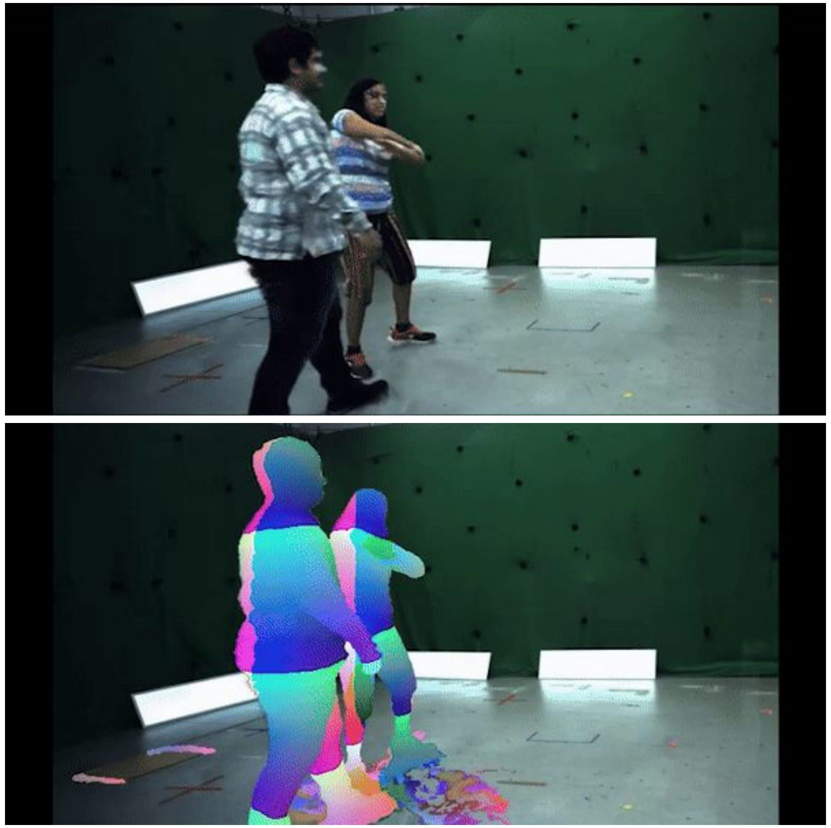
Time Local

Large Motion vs. Long-term 3D Correspondences
Spatio-Temporal Modelling

Time-Consistent Canonical Modelling:

- Builds a canonical model from the first frame of multiview videos and fixes it
- Online reconstruction of next timesteps
- Hard constraint on time-consistency of canonical model, thus improving temporal correspondences while handling large motion through coarse-to-fine deformations





SceNerFlow

Time Global

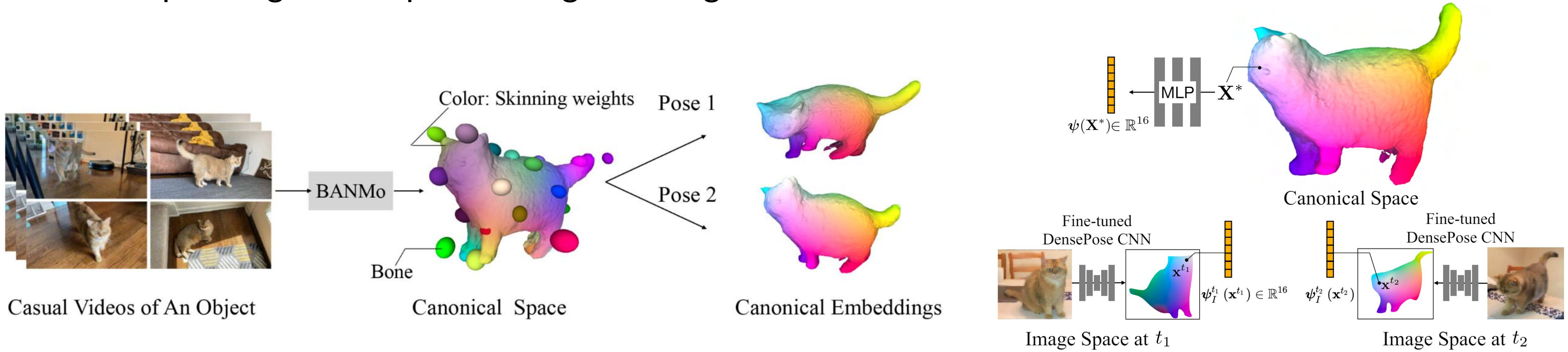
Global Canonical Model

Time Local

Large Motion vs. Long-term 3D Correspondences
Spatio-Temporal Modelling

Canonical Feature Embeddings:

- Shares canonical space over multiple videos of an object
- 2D DensePose features are distilled into the 3D canonical model as embeddings
- Enforcement of 3D canonical embeddings to match 2D DensePose features in each corresponding view improves long-term registration



BANMO

Global Canonical Model

Large Motion vs. Long-term 3D Correspondences

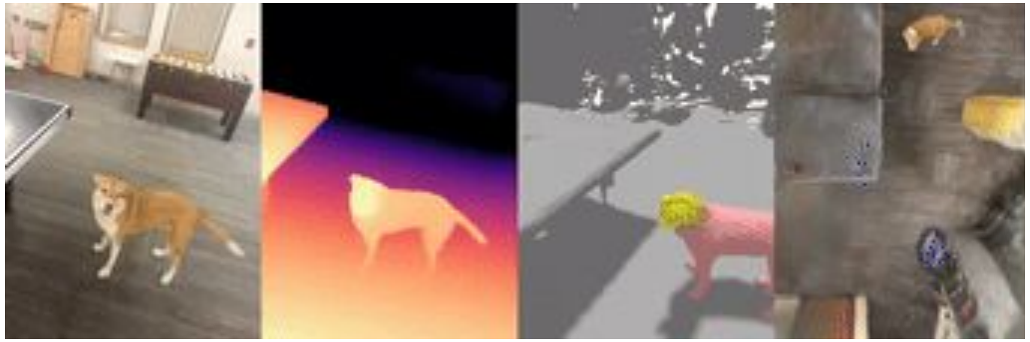
Spatio-Temporal Modelling

Decomposed Motion Modelling:

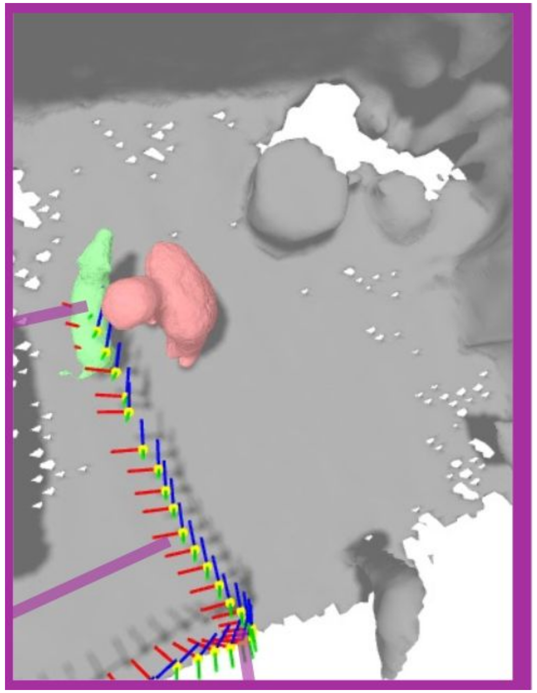
- Decomposes object motion into root pose and residual motion
- Simpler motion modelling allows it to scale to longer scenes
- Takes RGB-D input and needs root-pose initialization with PoseNet

Input RGBD Video

3D Scene Reconstruction



Total-Recon



Time Global

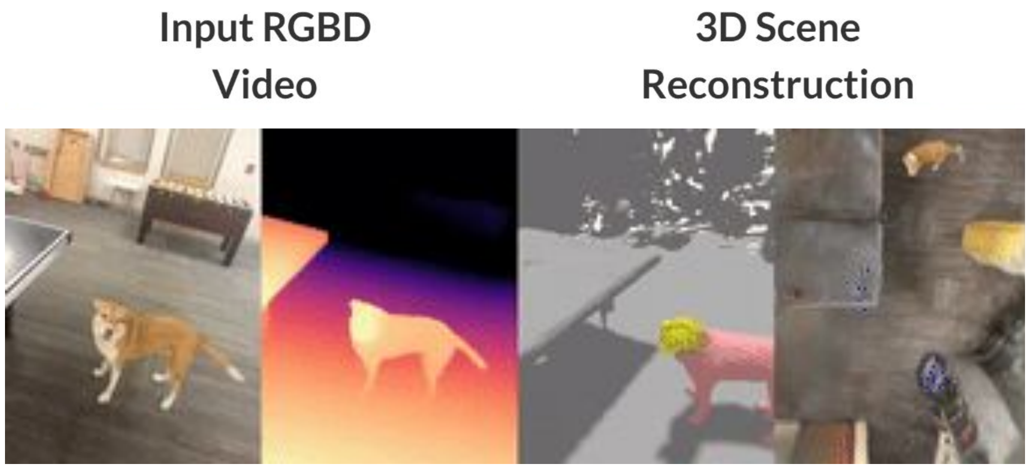
Global Canonical Model

Time Local

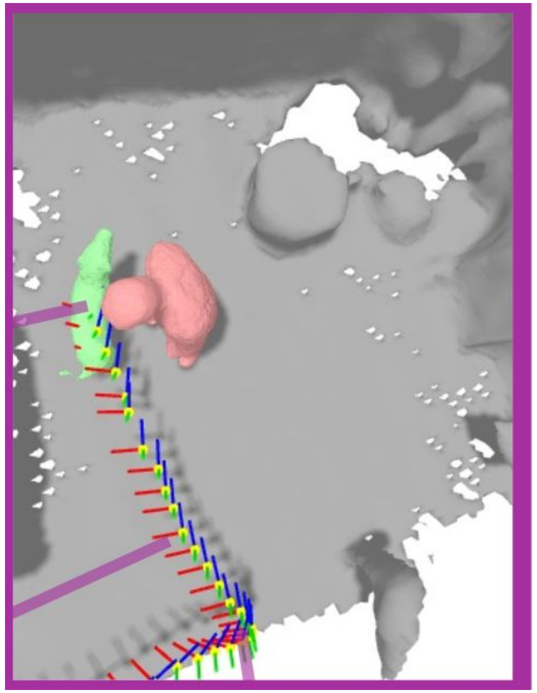
Large Motion vs. Long-term 3D Correspondences
Spatio-Temporal Modelling

Decomposed Motion Modelling:

- Decomposes object motion into root pose and residual motion
- Simpler motion modelling allows it to scale to longer scenes
- Takes RGB-D input and needs root-pose initialization with PoseNet



Total-Recon

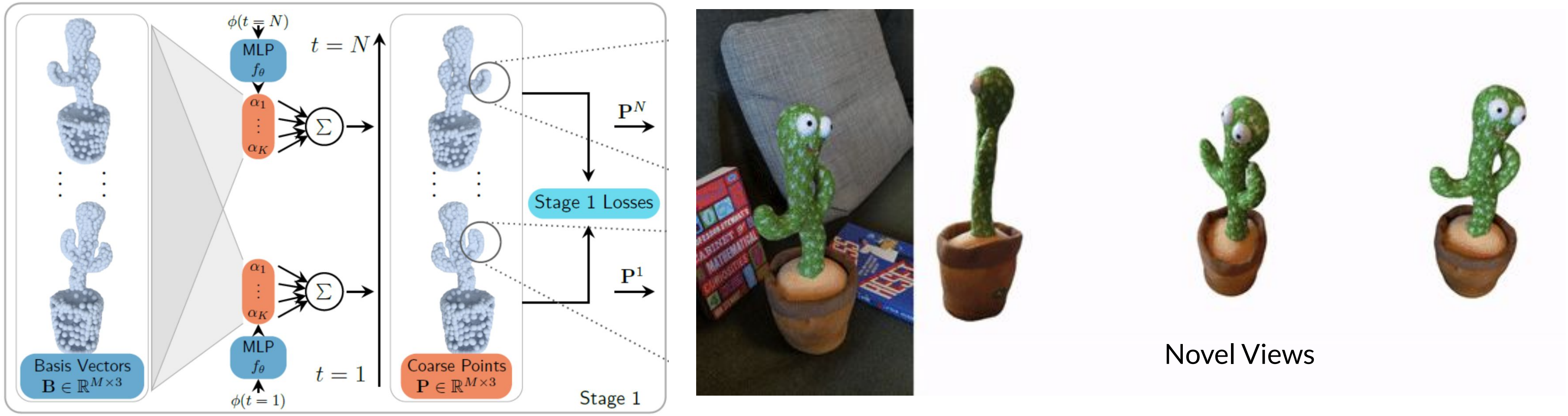


Scales up to minute-long RGB-D videos with large motion!

Large Motion vs. Long-term 3D Correspondences
Spatio-Temporal Modelling

Low-rank Deformation Template:

- Shared point template for each frame, automatically giving temporal correspondences
- Generated by low-rank basis, thus forcing information sharing
- Models complex motion while providing regularization for challenging novel views



Neural Parametric Gaussians

Time Global

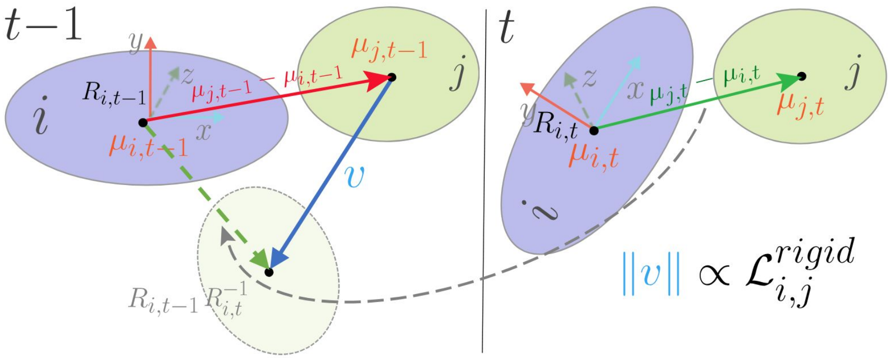
Deformation Basis

Time Local

Large Motion vs. Long-term 3D Correspondences
Spatio-Temporal Modelling

Per-frame Canonical Model Optimization:

- Rotation and position of canonical Gaussians are optimized for each timestep from last timestep, giving dense 6-DOF trajectories
- Models long-range motion but trajectories can drift over time
- Multi-view supervision required and surface rigidity losses introduced to tackle this



Dyn3DGS

Time Global

Canonical First Frame

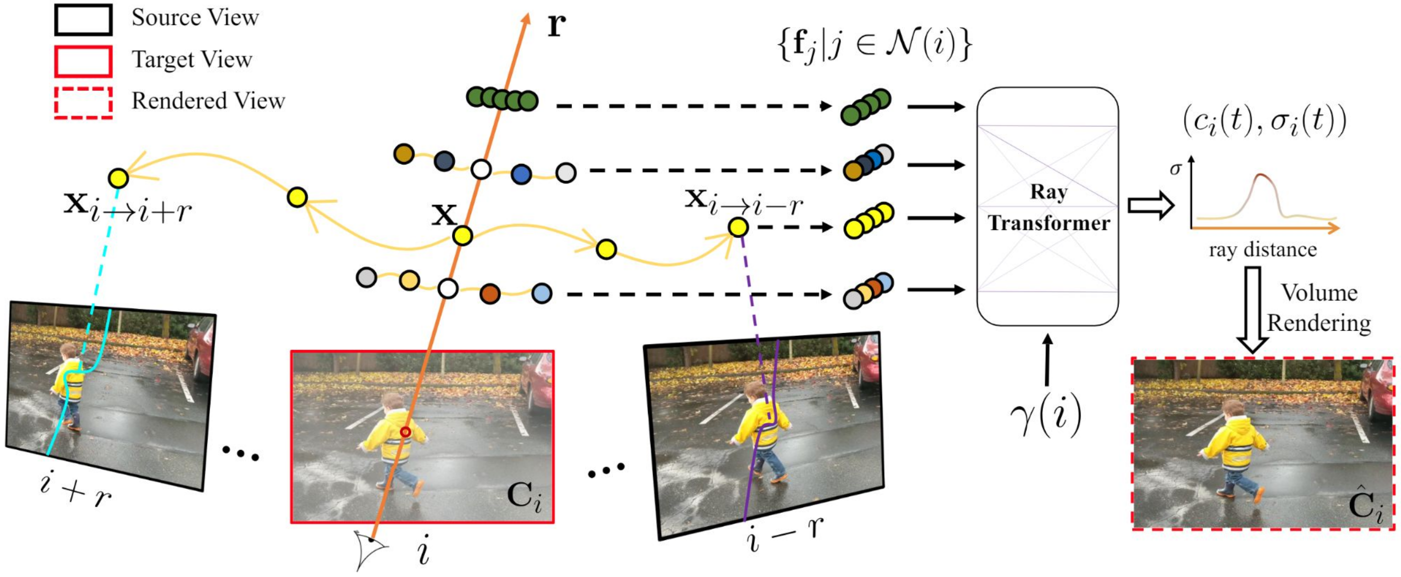
Time Local

Large Motion vs. Long-term 3D Correspondences

Spatio-Temporal Modelling

Motion Trajectory Modelling:

- Per-frame hybrid representation which takes in image features aggregated over time
- Motion trajectories allows information aggregation from a greater temporal neighbourhood



DynIBaR

Individual Frame

Large Motion vs. Long-term 3D Correspondences
Spatio-Temporal Modelling

Motion Trajectory Modelling:

- Per-frame hybrid representation which takes in image features aggregated over time
- Motion trajectories allows information aggregation from a greater temporal neighbourhood
- Improves time consistency while modelling free-form motion

DynIBaR

Individual Frame

Trends

1. Speed and Quality Advancements
2. Handling of Large Deformations / Long-Term 3D Correspondences
3. Modelling Articulated Motion for General Objects

Trends

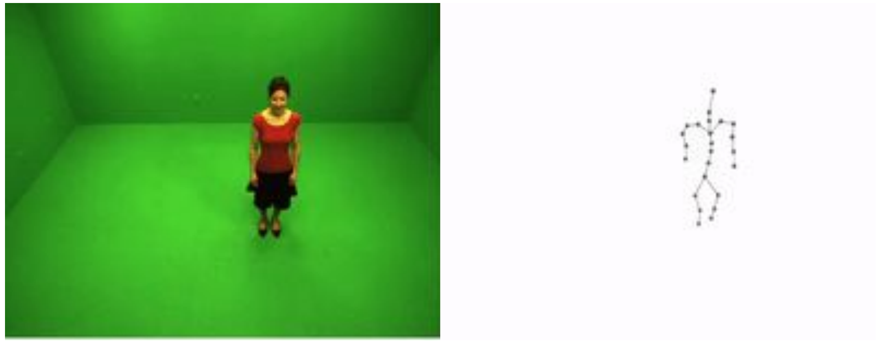
1. Speed and Quality Advancements

2. Handling of Large Deformations / Long-term 3D correspondences

3. Modelling Articulated Motion for General Objects

Modelling General Articulated Motion

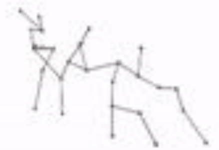
Deformations of humans, animals, and many other articulated objects can be represented and controlled by an underlying skeleton:



Modelling General Articulated Motion

Deformations of humans, animals, and many other articulated objects can be represented and controlled by an underlying skeleton:

- Skeletons allow reposing of objects to unseen poses



Modelling General Articulated Motion

Deformations of humans, animals, and many other articulated objects can be represented and controlled by an underlying skeleton:

- Skeletons allow reposing of objects to unseen poses
- If we know how an object category articulates, we can use that information as prior to estimate motion of new sequences of that object

Modelling General Articulated Motion

Deformations of humans, animals, and many other articulated objects can be represented and controlled by an underlying skeleton:

- Skeletons allow reposing of objects to unseen poses
- If we know how an object category articulates, we can use that information as prior to estimate motion of new sequences of that object
- Category-level object templates are available mostly for human categories (e.g. SMPL) which are obtained from expensive 3D data.

Modelling General Articulated Motion

Deformations of humans, animals, and many other articulated objects can be represented and controlled by an underlying skeleton:

- Skeletons allow reposing of objects to unseen poses
- If we know how an object category articulates, we can use that information as prior to estimate motion of new sequences of that object
- Category-level object templates are available mostly for human categories (e.g. SMPL) which are obtained from expensive 3D data.
- How can we obtain them for general object categories where such data is not available?

Modelling General Articulated Motion

Deformations of humans, animals, and many other articulated objects can be represented and controlled by an underlying skeleton:

- Skeletons allow reposing of objects to unseen poses
- If we know how an object category articulates, we can use that information as prior to estimate motion of new sequences of that object
- Category-level object templates are available mostly for human categories (e.g. SMPL) which are obtained from expensive 3D data.
- How can we obtain them for general object categories where such data is not available?

From RGB videos!

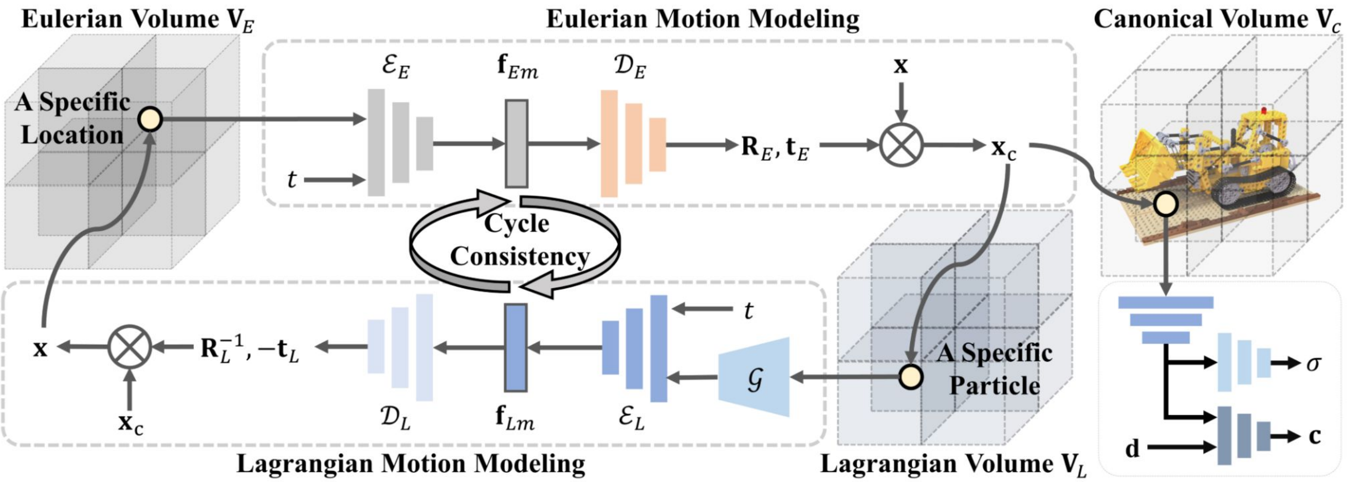
Modelling General Articulated Motion

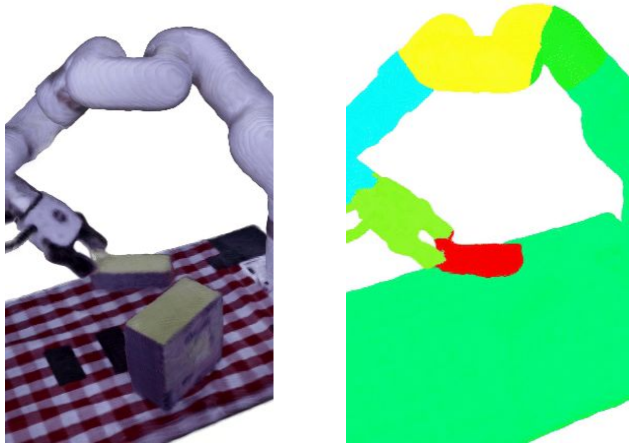
Self-Supervised Part Discovery for Reposing

Modelling General Articulated Motion
Self-Supervised Part Discovery for Reposing

Motion-based grouping:

- Models both backward and forward motion with feature grids
- Features from forward motion are grouped into slots using an attention mechanism
- Similar motion ➡ same slot ➡ same part
- Discovered parts can be skeletonized and reposed



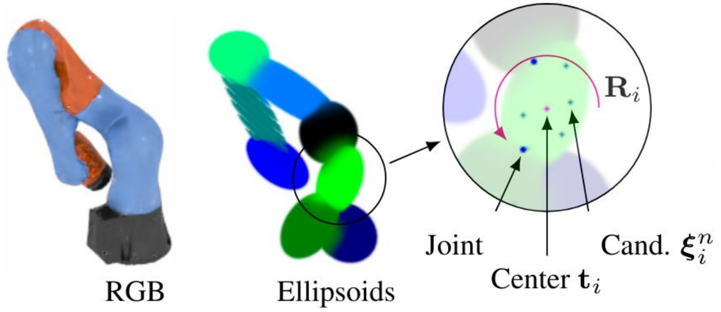


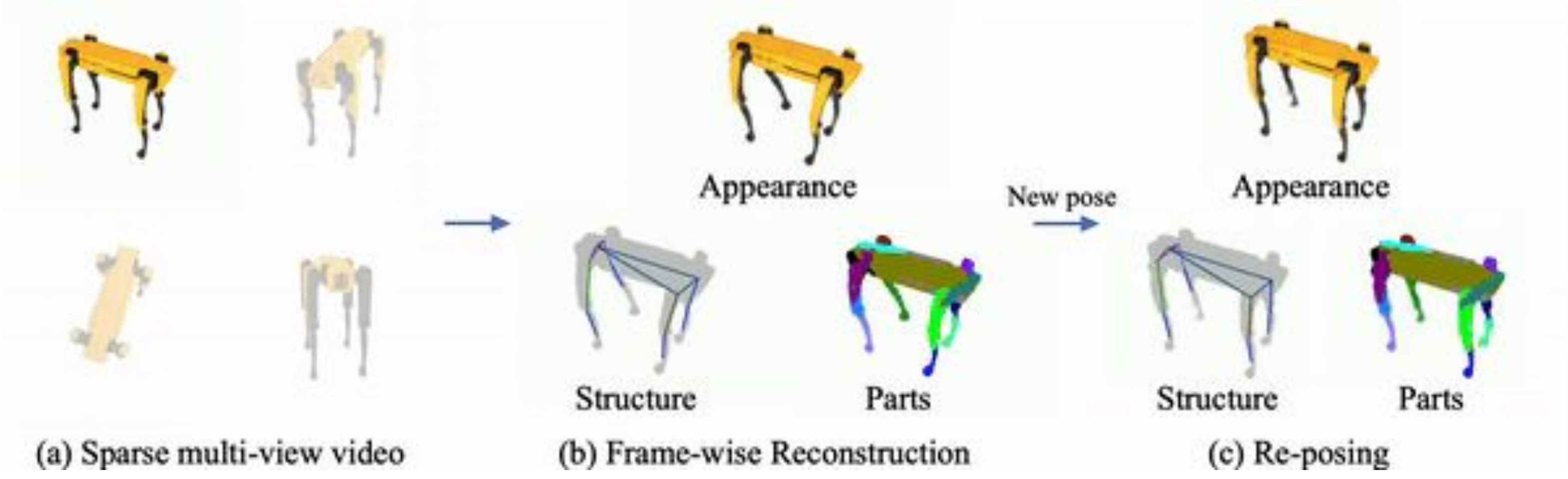
MovingParts

Modelling General Articulated Motion
Self-Supervised Part Discovery for Reposing

Unsupervised Part Prediction:

- Represent parts by ellipsoids in 3D
- Each ellipsoid has a rotation and a position
- Optimize the per-frame ellipsoids prediction MLP from multi-view videos
- Repose using discovered ellipsoids





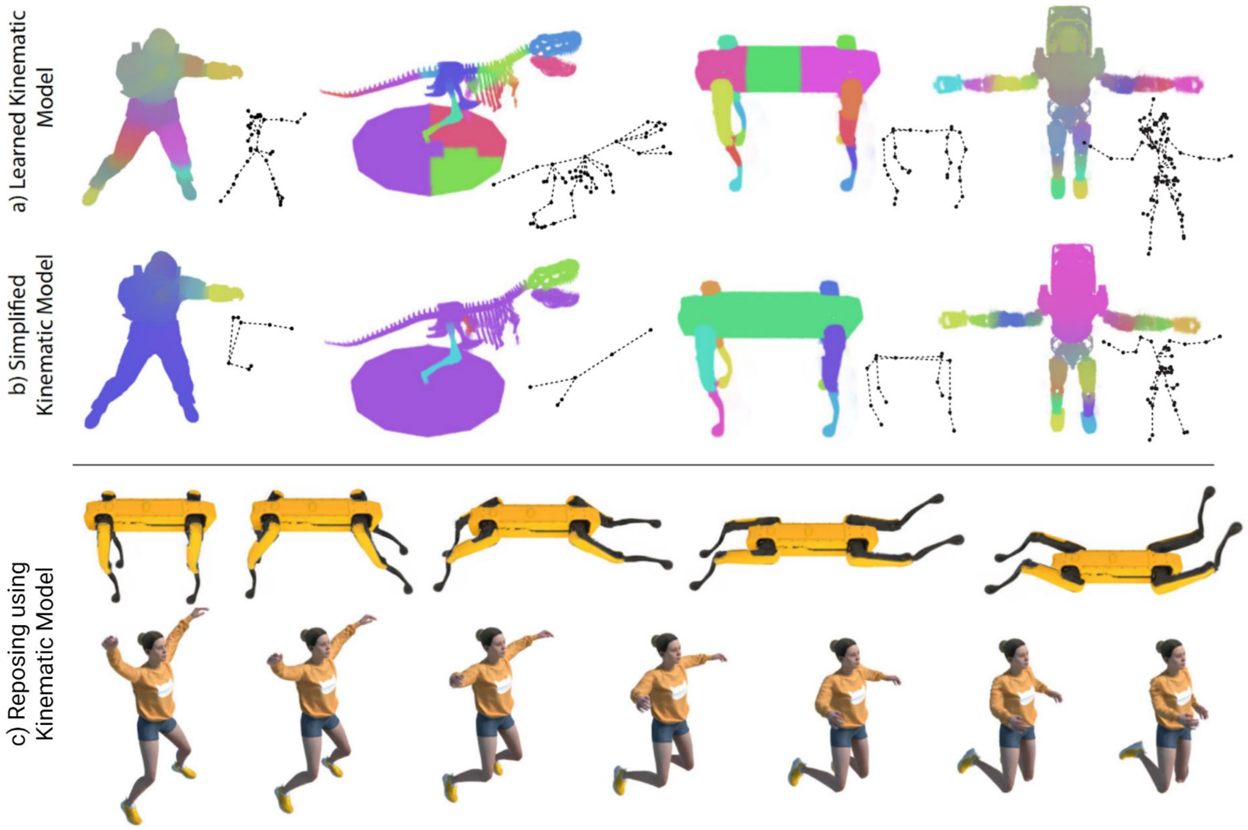
Watch-it-Move

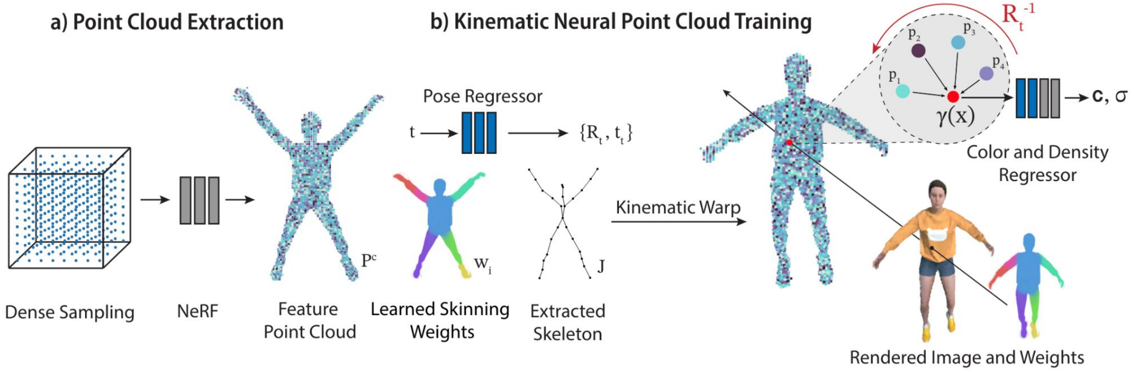
Modelling General Articulated Motion

Skeleton Discovery for Reposing

Morphological Operations:

- Point-based canonical representation extracted from a dynamic NeRF backbone
- Medial Axis Transform used to extract skeleton from canonical points
- Linear blend skinning-based model to learn forward dynamics from observations
- Repose using the learnt template
- Also fast because of the point-based hybrid representation





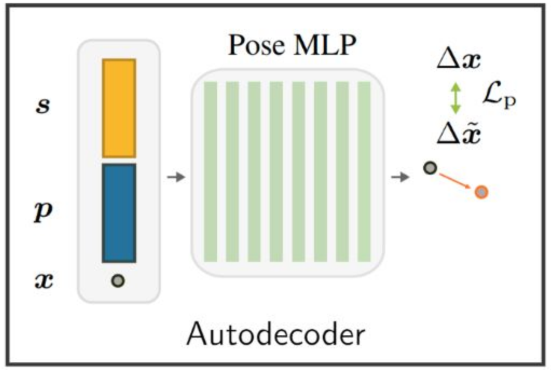
Uzolas *et al.*

Modelling General Articulated Motion

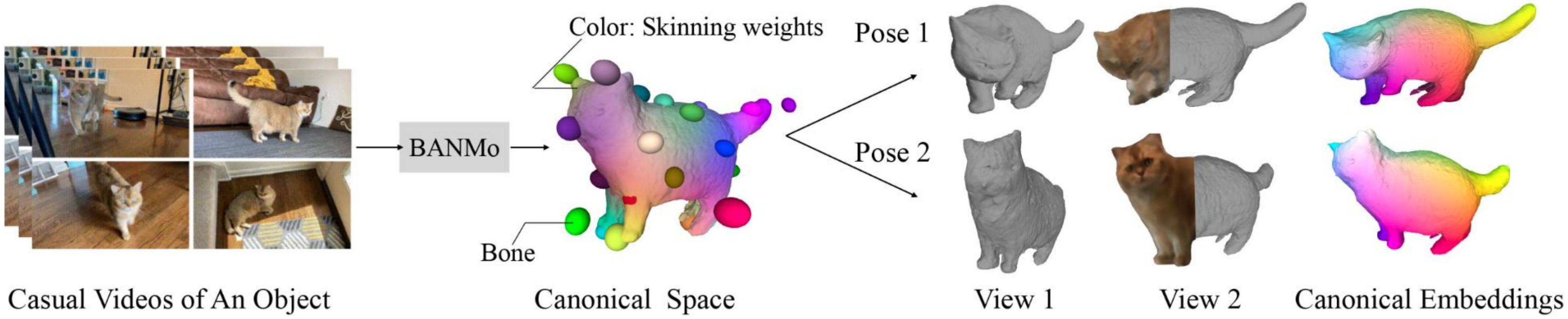
- Previous methods are trained on a single video sequence
- Can we utilize multiple videos of the same object to build an instance-level model?

Modelling General Articulated Motion
Modelling Articulations with Neural Bones

- Canonical space is shared between videos
- Bone positions and transforms are estimated per-frame using an auto-decoded MLP
- Articulated using volumetric skinning



Model captures the articulations across videos, providing better regularization



BANMO

Modelling General Articulated Motion
Modelling Articulations with Neural Bones

- Use optimized pose embeddings from a driving video for another structurally similar geometry model for motion retargeting!

Driving Sequence

Target Geometry

BANMO

Modelling General Articulated Motion

- A category of objects articulates in the same way (e.g. different breeds of cats)
- Can we learn category-level templates from videos to regularize motion even further and use it as a prior for instances?

Modelling General Articulated Motion

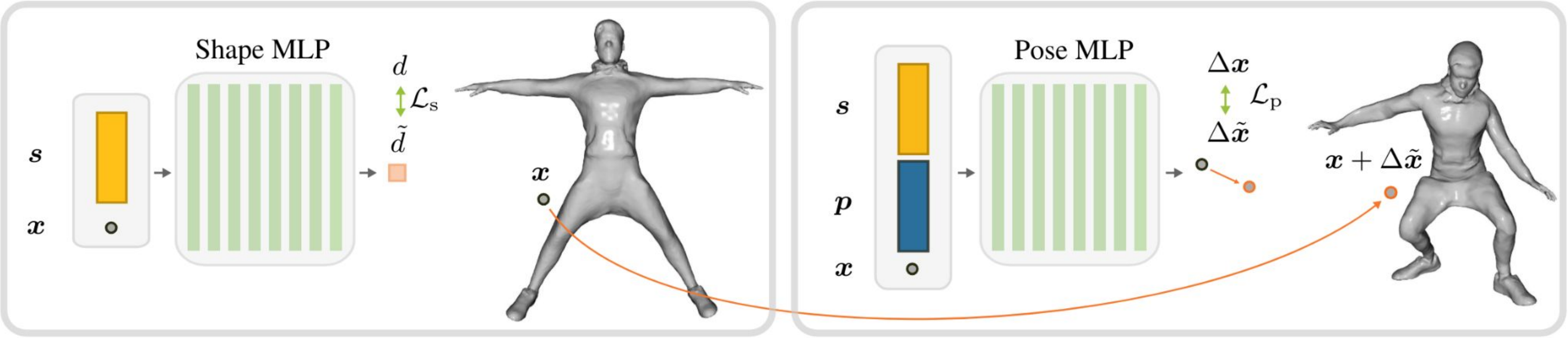
- A category of objects articulates in the same way (e.g. different breeds of cats)
- Can we learn category-level templates from videos to regularize motion even further and use it as a prior for instances?

Yes, but we need to capture shape variations between category instances as well!

Modelling General Articulated Motion

Category-level Modelling from Depth Videos

- Use auto-decoders to model both shape and pose variations
- Shape embeddings can capture category-level variations while pose embeddings capture instance articulations
- Optimize shape and then pose at test-time



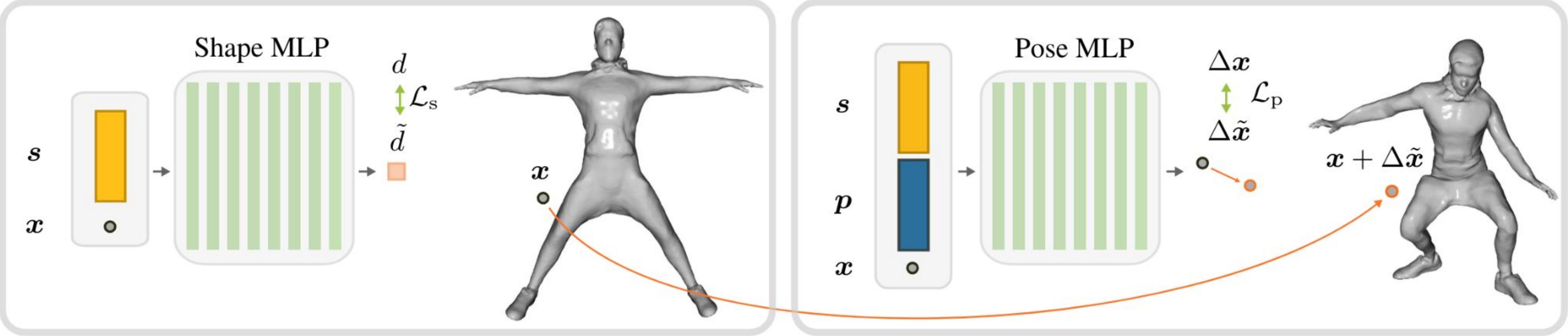
NPMs

Modelling General Articulated Motion

Category-level Modelling from Depth Videos

- Use auto-decoders to model both shape and pose variations
- Shape embeddings can capture category-level variations while pose embeddings capture instance articulations
- Optimize shape and then pose at test-time

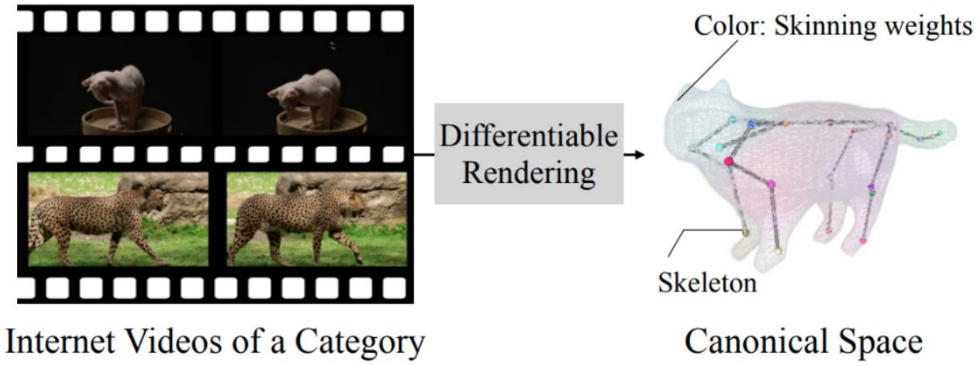
Learned from depth sequences. Can we do it from RGB videos?

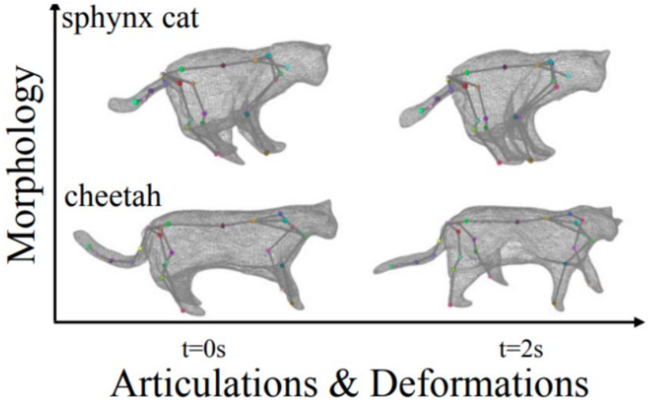


NPMs

Modelling General Articulated Motion
Category-level Modelling from RGB Videos

- Learn category-level shape and skeleton model from internet videos of a category
- Predict the instance-level bone locations for category skeleton using an auto-decoded MLP, similar to BANMO
- Capture instance-level articulations using BANMO





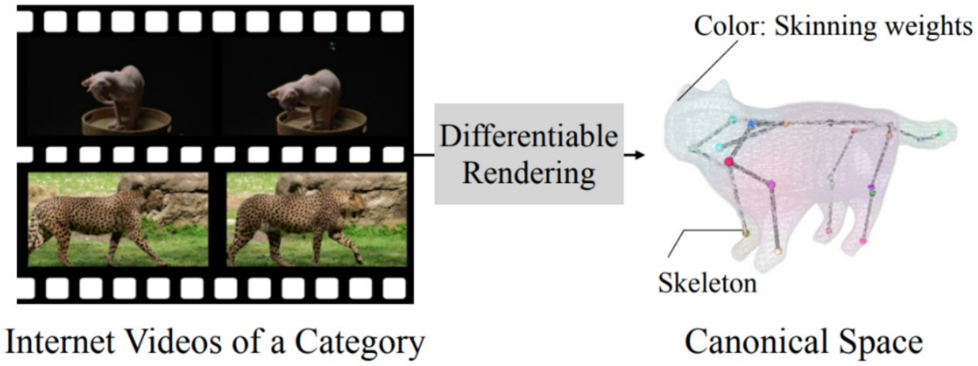
RAC

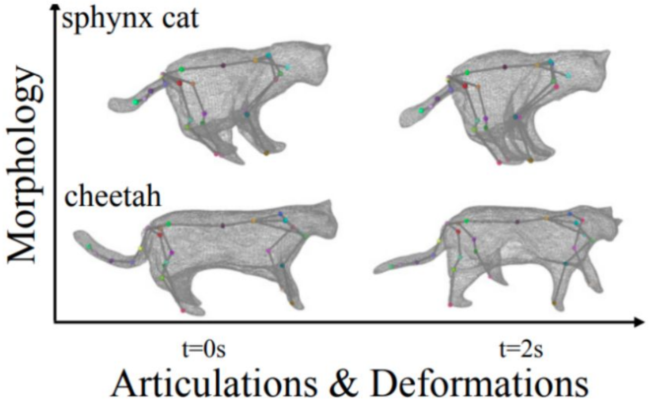
Modelling General Articulated Motion

Category-level Modelling from RGB Videos

- Learn category-level shape and skeleton model from internet videos of a category
- Predict the instance-level bone locations for category skeleton using an auto-decoded MLP, similar to BANMO
- Capture instance-level articulations using BANMO

Can we do it from image collections, which are more commonly available for general categories than videos?



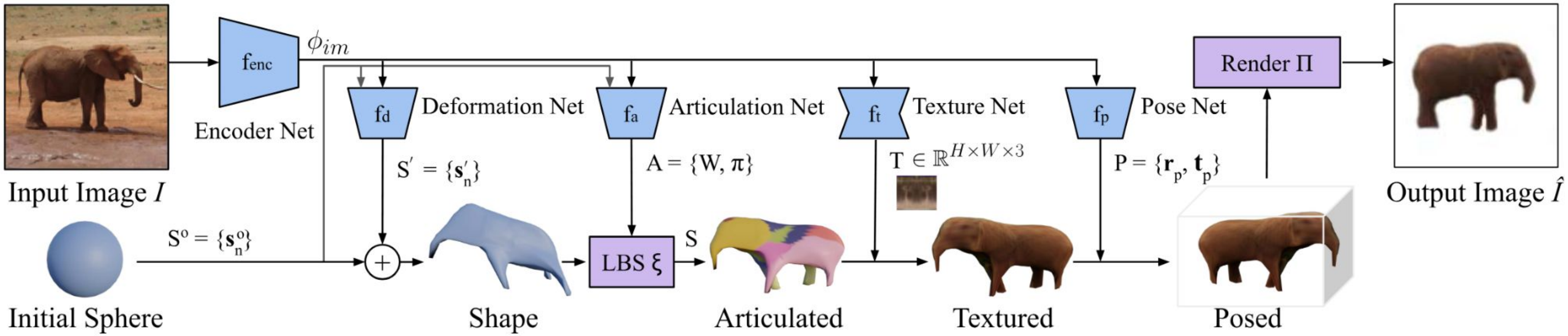


RAC

Modelling General Articulated Motion

Category-level Modelling from Image Collections

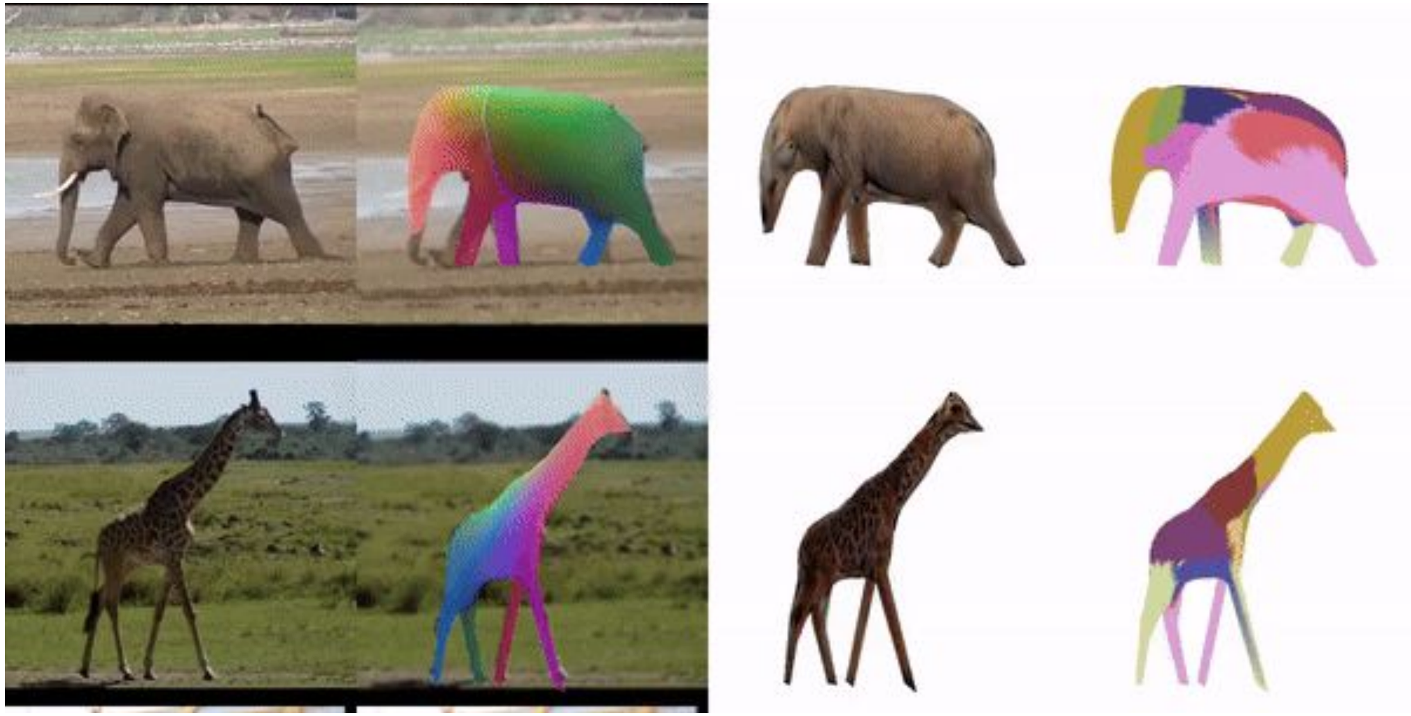
- Shape, articulation, pose and texture are directly predicted with separate decoders from an encoded image
- Category-level prior learned by shape and articulation decoders
- Enables prediction from single image at test-time



SAOR

Modelling General Articulated Motion
Category-level Modelling from Image Collections

- Per-frame video reconstruction



SAOR

Trends

1. Speed and Quality Advancements

2. Handling of Large Deformations / Long-term 3D correspondences

3. Modelling Articulated Motion for General Objects

Trends

1. Speed and Quality Advancements

2. Ha

Non-rigid 3D reconstruction is far from solved!

ndences

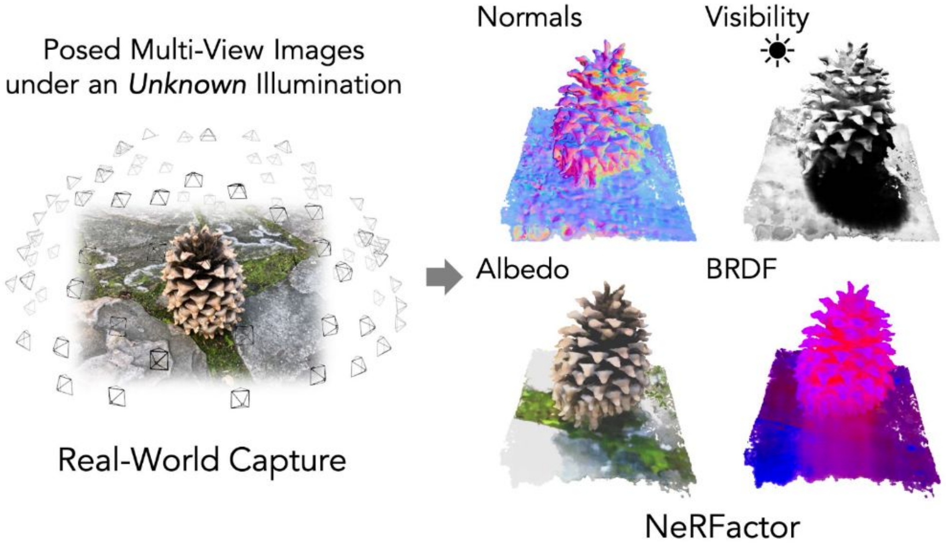
3. Modelling Articulated Motion for General Objects

Remaining Challenges and Future Directions

Remaining Challenges and Future Directions

Intrinsic Decomposition and Relighting

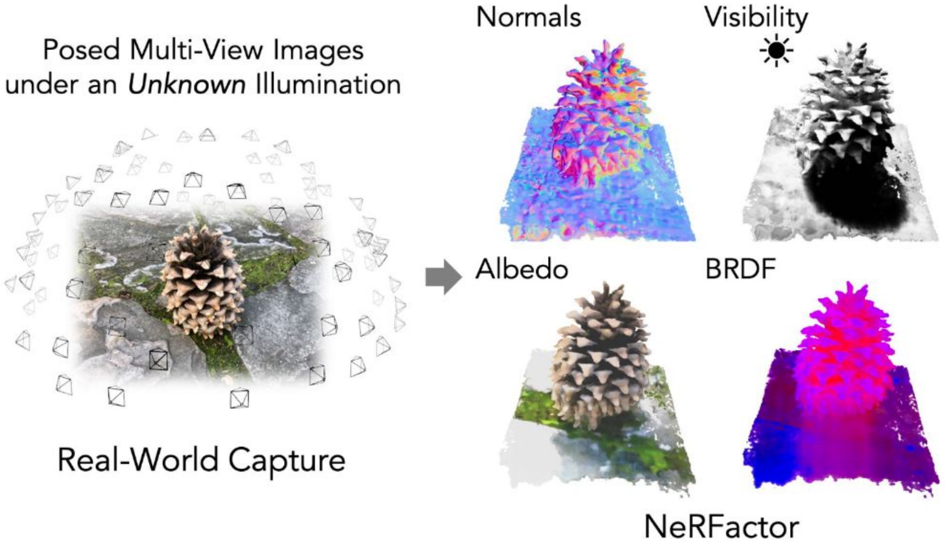
- Current methods for general scenes do not estimate materials and lighting
- Required to correctly relight objects in new environments



Remaining Challenges and Future Directions

Intrinsic Decomposition and Relighting

- Current methods for general scenes do not estimate materials and lighting
- Required to correctly relight objects in new environments



Faster Scene Representations

- Gaussian Splatting has introduced real-time rendering with photorealistic appearance
- Photorealistic reconstruction still requires offline training

Remaining Challenges and Future Directions

Reliable Camera Pose Estimation

- Current view synthesis methods rely on static Structure-from-Motion for camera poses
- Noisy when large and complex motions are present

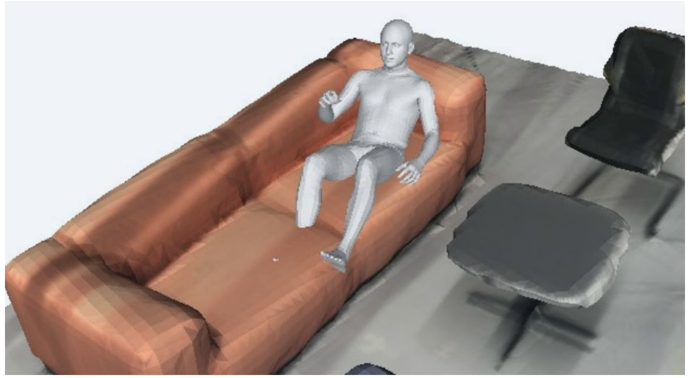
Remaining Challenges and Future Directions

Reliable Camera Pose Estimation

- Current view synthesis methods rely on static Structure-from-Motion for camera poses
- Noisy when large and complex motions are present

Multi-Object Interaction

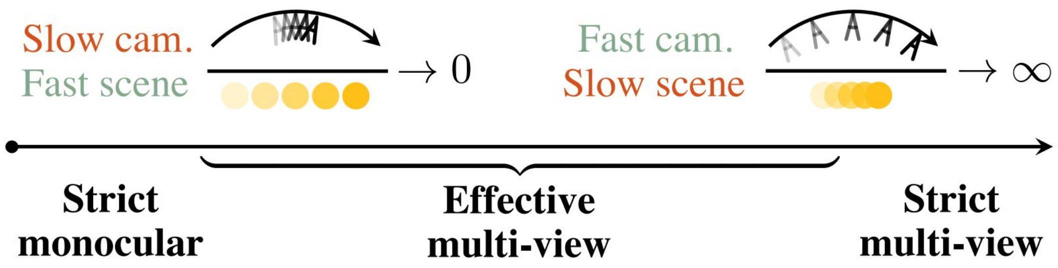
- Interaction between objects is not explicitly modelled by current methods for general objects
- Useful to enforce the correct dynamics and constraints



Remaining Challenges and Future Directions

Reconstruction from Sparse Casual Captures

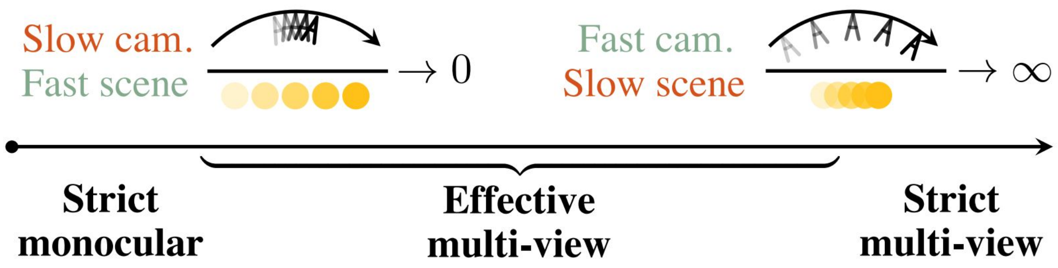
- Most methods evaluate on data with multi-view cues
- Reconstruction quality degrades for sparse, realistic monocular captures



Remaining Challenges and Future Directions

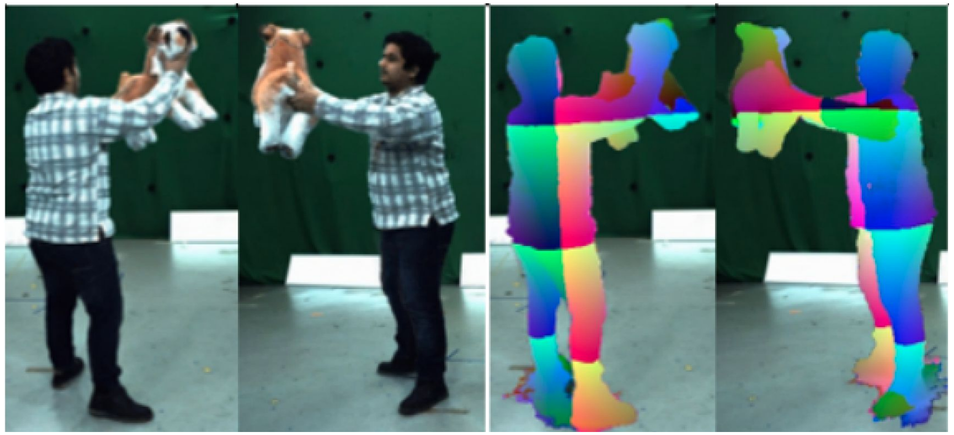
Reconstruction from Sparse Casual Captures

- Most methods evaluate on data with multi-view cues
- Reconstruction quality degrades for sparse, realistic monocular captures



Long-Term Dense Correspondences

- Recent works allow establishing 3D correspondences over time on lab-captured data
- Results not satisfactory for general real scenes with large and complex motion



Remaining Challenges and Future Directions

Generalizable Modeling and Generative Priors

- Text-to-image and text-to-video 2D diffusion models have been used as priors for 3D non-rigid scene generators
- We can see these powerful generative models being utilized for the non-rigid reconstruction task as well

Input Synthesized Novel Views

EUROGRAPHICS
CONFERENCE 2024

Thank you.

More Information:

https://razayunus.github.io/non-rigid-star

Contact Information:

https://razayunus.github.io

Thanks to all authors for their contribution to the STAR!