

# Learning Dynamic 3D Objects in the Wild

*Elliott / Shangzhe Wu*

Postdoc at Stanford SVL



**Stanford**  
University



UNIVERSITY OF  
**OXFORD**

# ☰ Stable Diffusion 2.1 Demo

Stable Diffusion 2.1 is the latest text-to-image model from StabilityAI. [Access Stable Diffusion 1 Space here](#)

For faster generation and API access you can try [DreamStudio Beta](#).

horse

**Generate image**

Enter a negative prompt





# Stable Diffusion 2.1 Demo

Stable Diffusion 2.1 is the latest text-to-image model from StabilityAI. [Access Stable Diffusion 1 Space here](#)

For faster generation and API access you can try [DreamStudio Beta](#).

horse

Generate image

Enter a negative prompt





# What is an object?

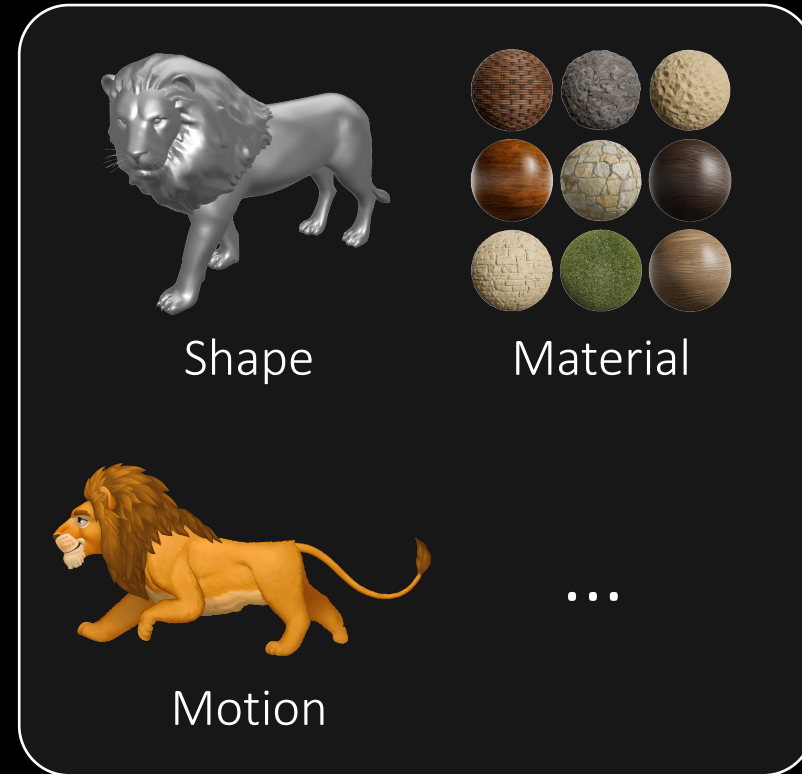




# Perceiving Physical Objects beyond 2D Pixels



A "View" of an Object



Shape

Material

Motion

...

3D Object Priors

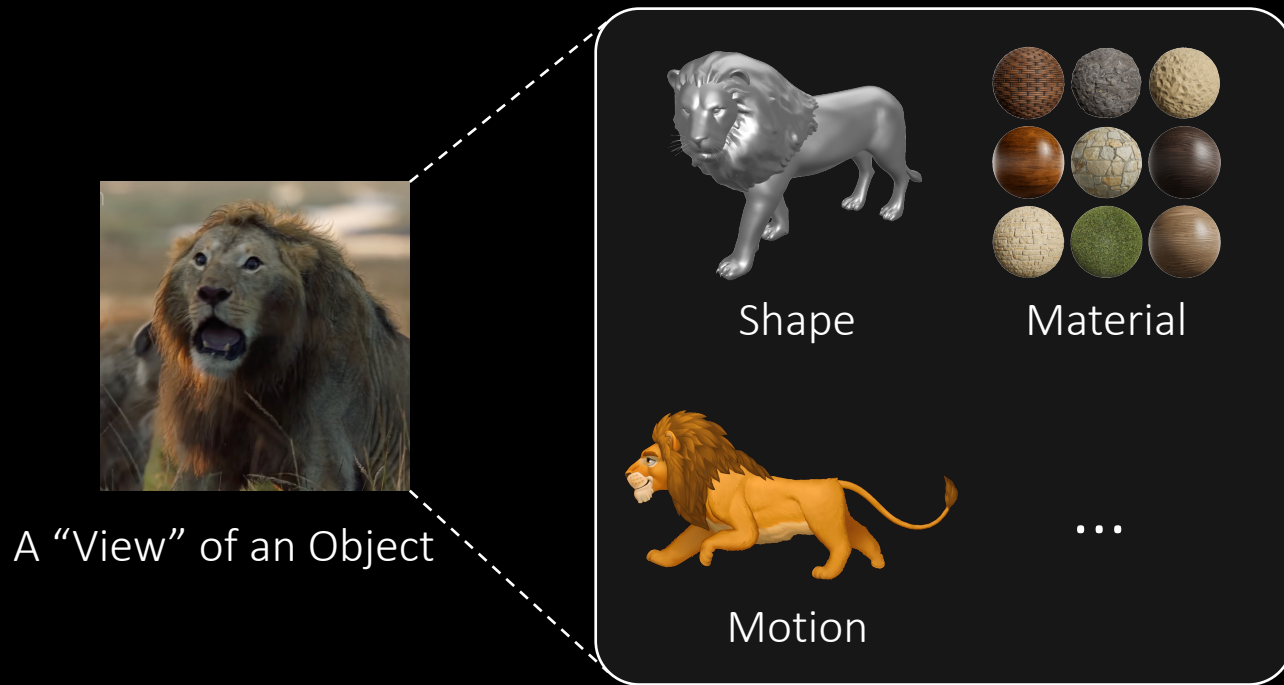
# Geometric Annotations by Humans



COCO  
Common Objects in Context



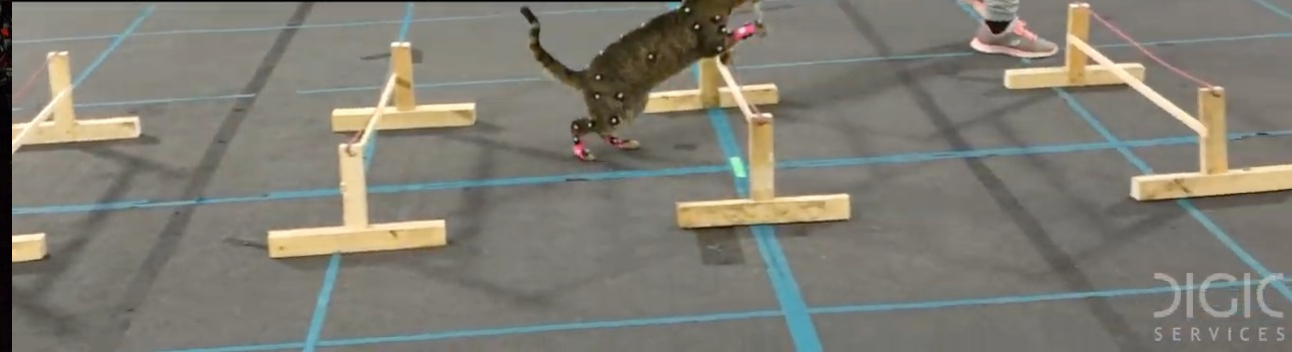
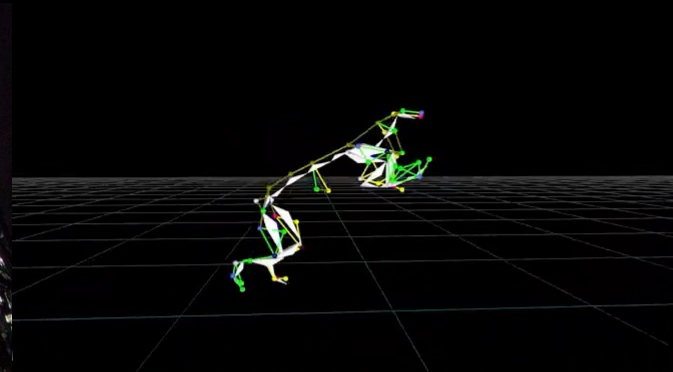
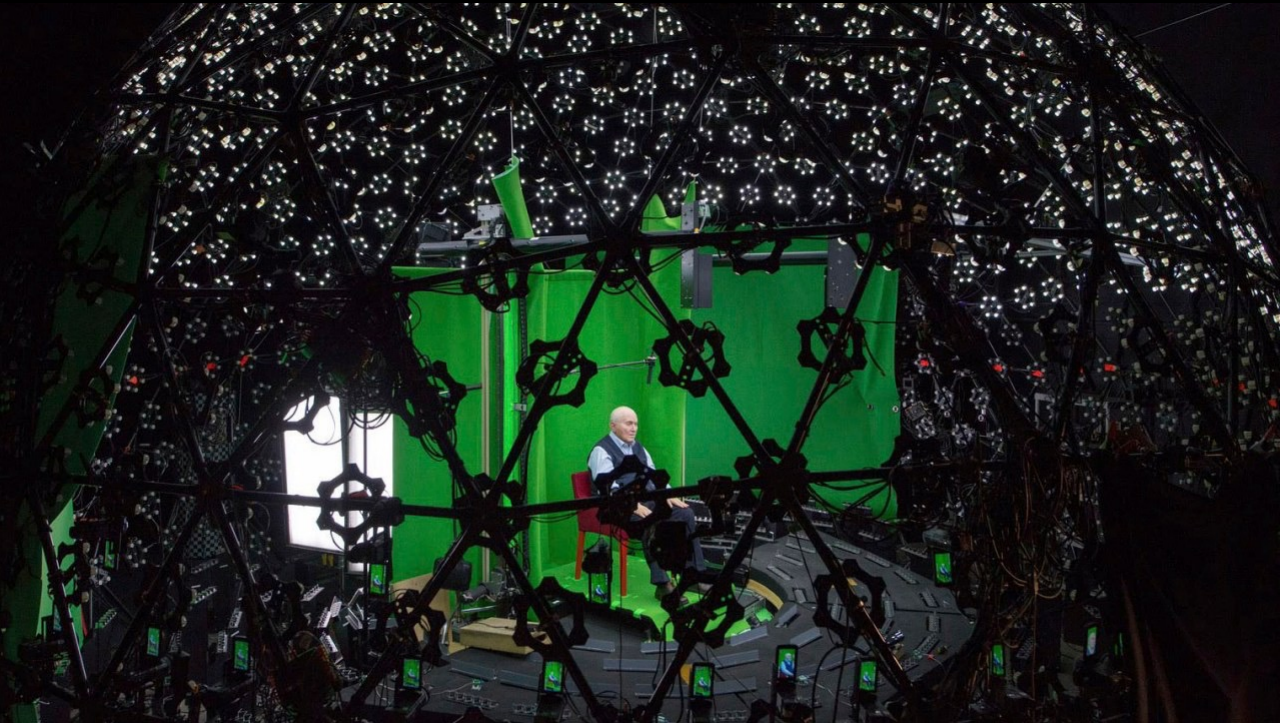
# Annotation beyond 2D is hard!



Physically-grounded 3D Representations

- 3D surfaces, normals?
- Materials (BRDFs)?
- Environment lighting?
- Physics: force, torque, mass, friction, velocity, acceleration...?

# Special Capturing Devices



Hard to scale up to all kinds of objects

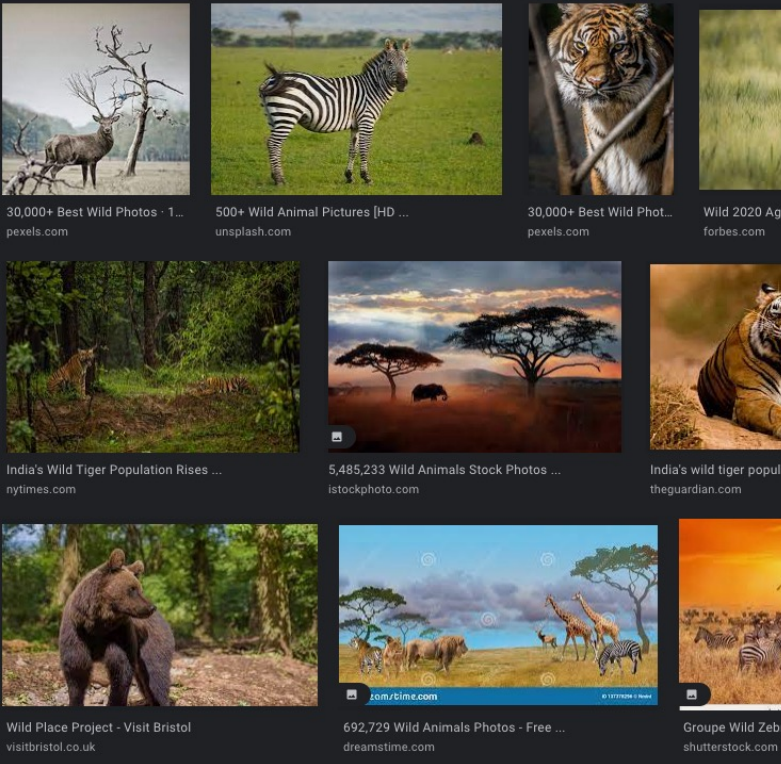


# Can we simply learn from “the wild”?

Google

All Images Videos News Shopping More

beautiful tiger lion nature woodstock



30,000+ Best Wild Photos · 1... pexels.com

500+ Wild Animal Pictures [HD ... unsplash.com

30,000+ Best Wild Phot... pexels.com

Wild 2020 Ag... forbes.com

India's Wild Tiger Population Rises ... nytimes.com

5,485,233 Wild Animals Stock Photos ... istockphoto.com

India's wild tiger popul... theguardian.com

Wild Place Project - Visit Bristol visitbristol.co.uk

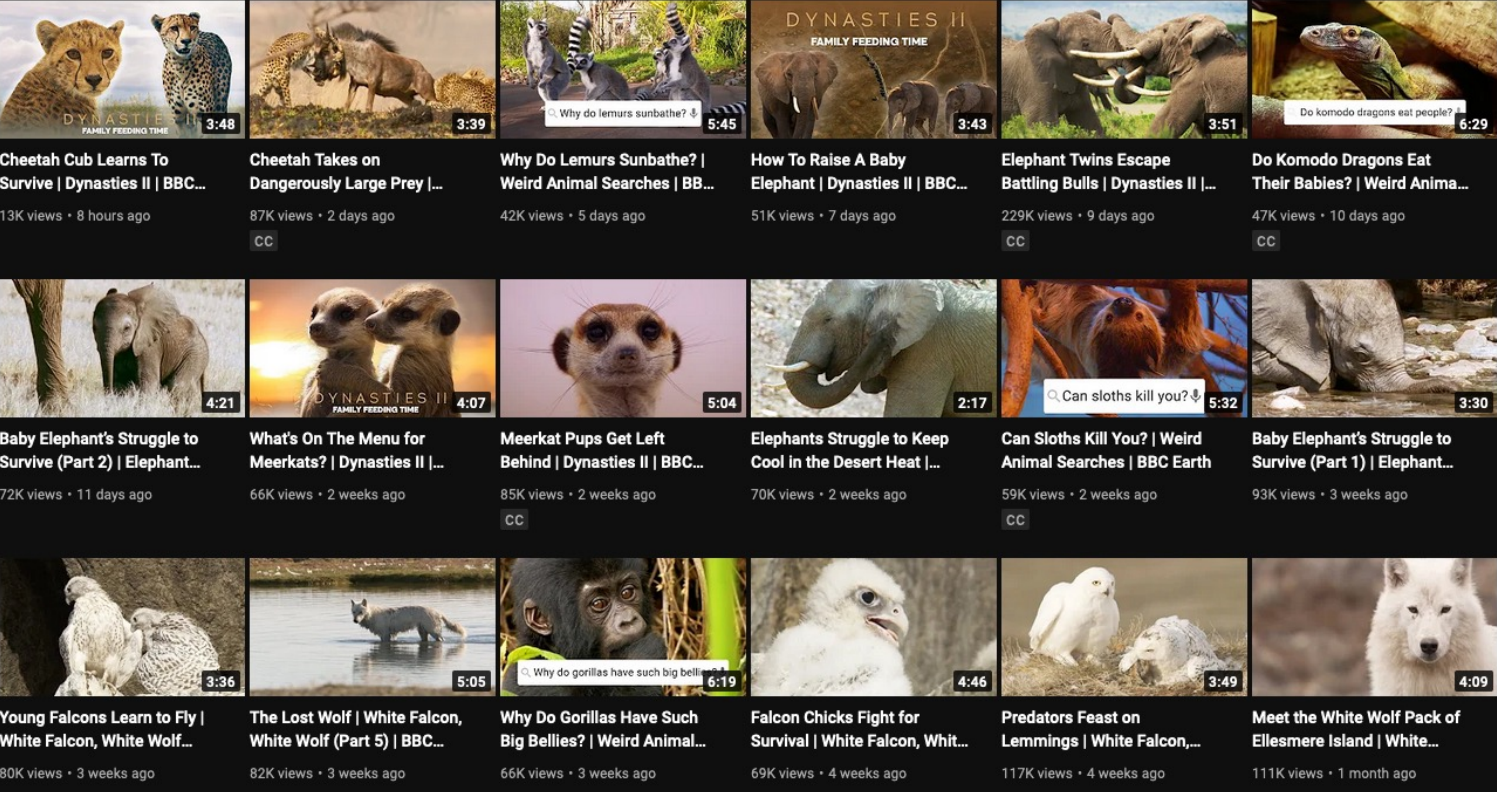
692,729 Wild Animals Photos - Free ... dreamstime.com

Gruppe Wild Zebr... shutterstock.com

**BBC Earth** 11.2M subscribers [SUBSCRIBE](#)

HOME **VIDEOS** PLAYLISTS COMMUNITY CHANNELS ABOUT

Uploads ▼ SORT BY



**Cheetah Cub Learns To Survive | Dynasties II | BBC...**  
13K views · 8 hours ago

**Cheetah Takes on Dangerously Large Prey |...**  
87K views · 2 days ago

**Why Do Lemurs Sunbathe? | Weird Animal Searches | BB...**  
42K views · 5 days ago

**How To Raise A Baby Elephant | Dynasties II | BBC...**  
51K views · 7 days ago

**Elephant Twins Escape Battling Bulls | Dynasties II |...**  
229K views · 9 days ago

**Do Komodo Dragons Eat Their Babies? | Weird Anima...**  
47K views · 10 days ago

**Baby Elephant's Struggle to Survive (Part 2) | Elephant...**  
72K views · 11 days ago

**What's On The Menu for Meerkats? | Dynasties II |...**  
66K views · 2 weeks ago

**Meerkat Pups Get Left Behind | Dynasties II | BBC...**  
85K views · 2 weeks ago

**Elephants Struggle to Keep Cool in the Desert Heat |...**  
70K views · 2 weeks ago

**Can Sloths Kill You? | Weird Animal Searches | BBC Earth**  
59K views · 2 weeks ago

**Baby Elephant's Struggle to Survive (Part 1) | Elephant...**  
93K views · 3 weeks ago

**Young Falcons Learn to Fly | White Falcon, White Wolf...**  
80K views · 3 weeks ago

**The Lost Wolf | White Falcon, White Wolf (Part 5) | BBC...**  
82K views · 3 weeks ago

**Why Do Gorillas Have Such Big Bellies? | Weird Animal...**  
66K views · 3 weeks ago

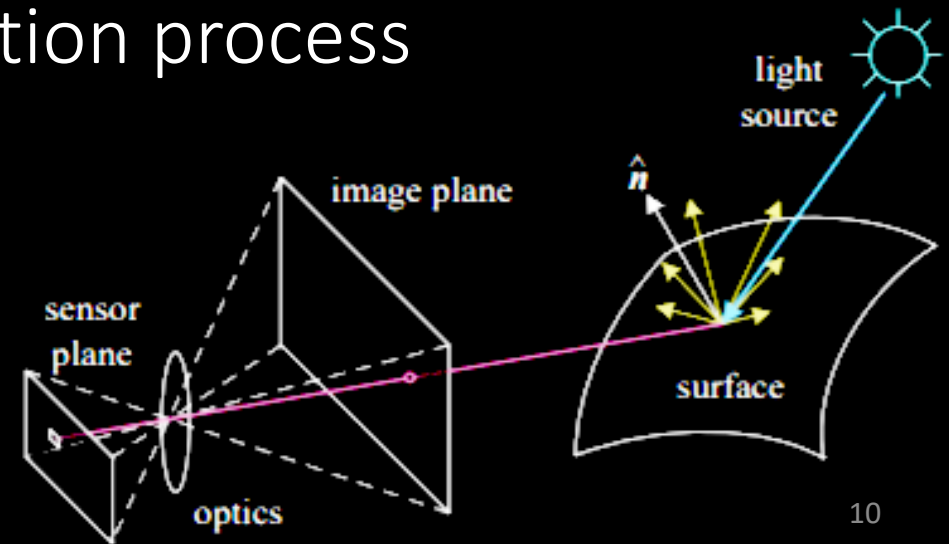
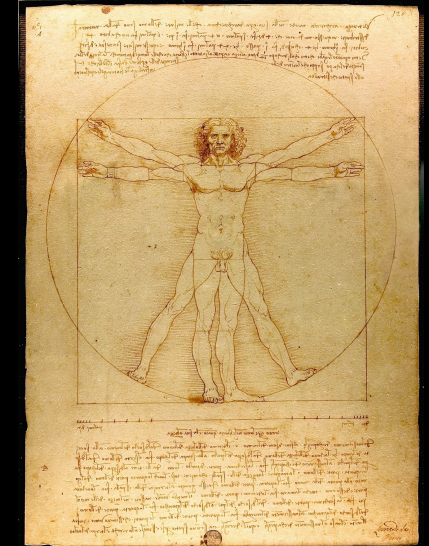
**Falcon Chicks Fight for Survival | White Falcon, Whit...**  
69K views · 4 weeks ago

**Predators Feast on Lemmings | White Falcon,...**  
117K views · 4 weeks ago

**Meet the White Wolf Pack of Ellesmere Island | White...**  
111K views · 1 month ago

# Luckily, we know how the world works (at least kind of...)

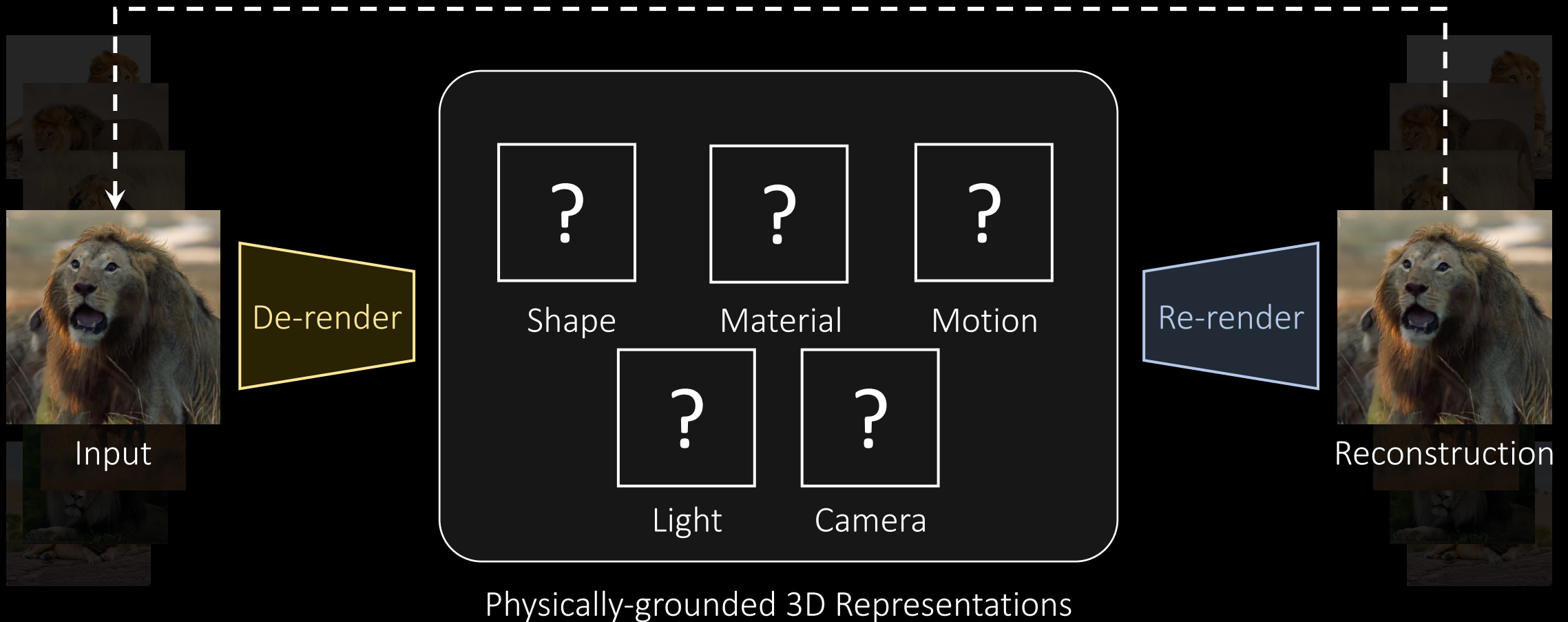
- It's a physical 3D world
- Lots of symmetries / regularities
- We can simulate the image formation process
- ...





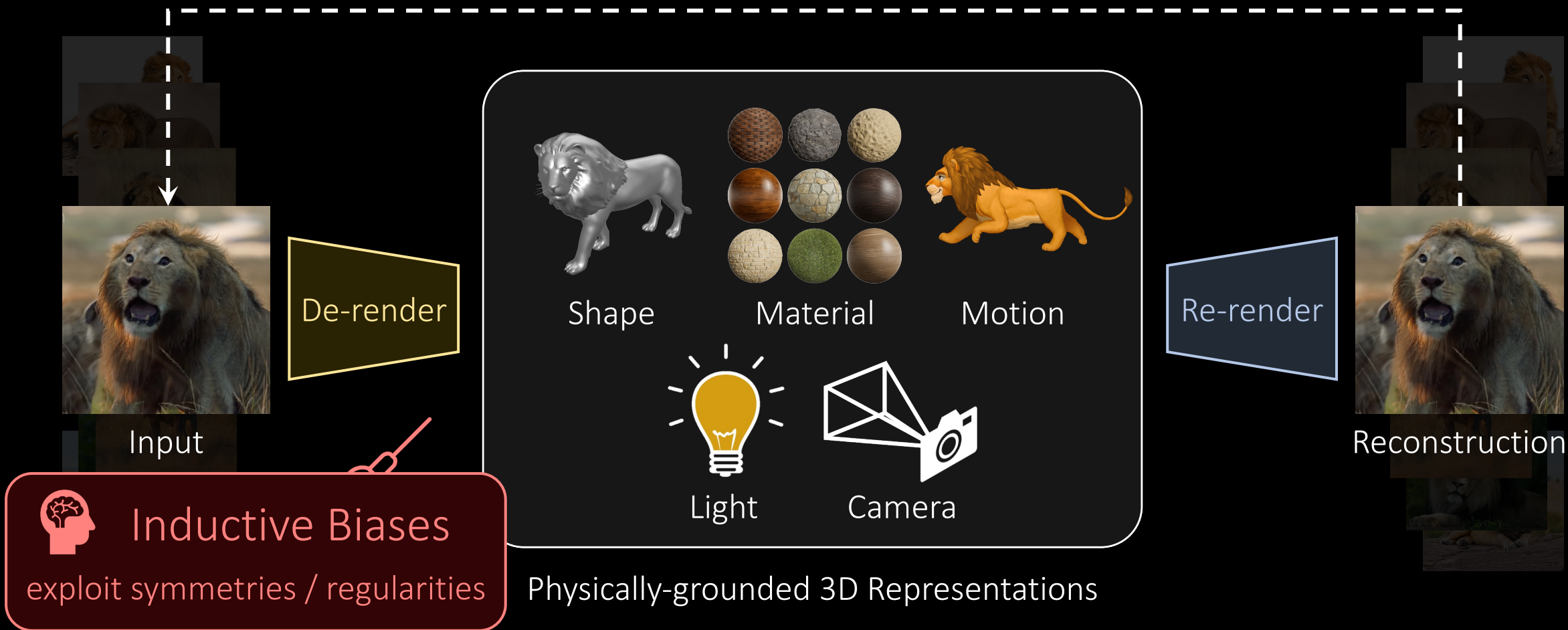
# Photo-Geometric Autoencoding

Minimize Reconstruction Error



# Photo-Geometric Autoencoding

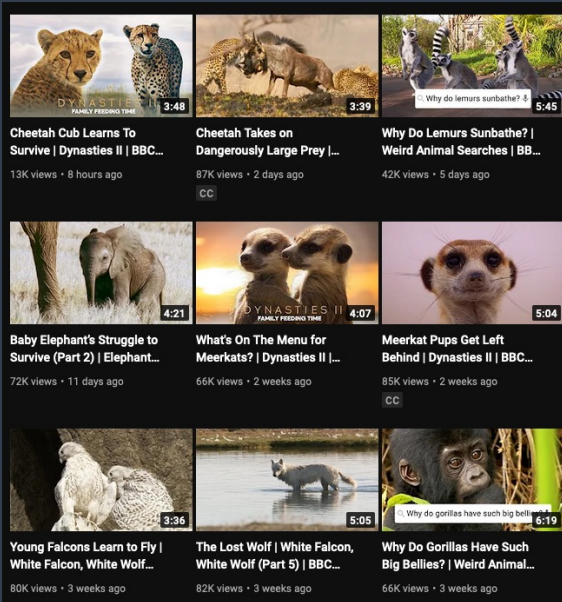
Minimize Reconstruction Error





# Learning Physical 3D Objects in the Wild

## Training



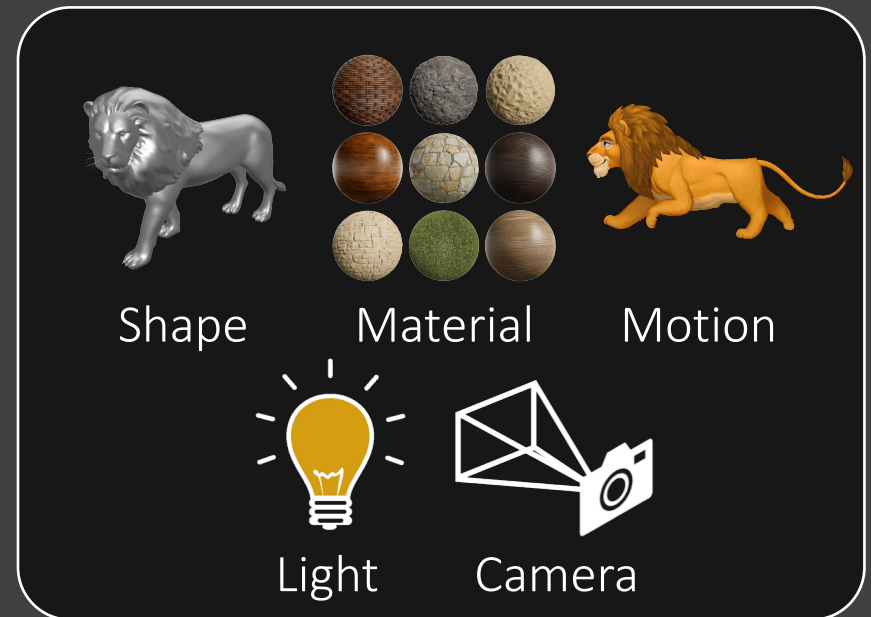
“In-the-Wild” Data



## Inference – Single Image De-rendering



De-render



Physics offers a path for learning compact, generalizable object representations.

# Unsupervised 3D Learning in the Wild

- 3D annotations are expensive and often infeasible at scale.
- Towards first principles in vision:
  - What are the minimal assumptions for 3D perception?
- Learning through inverse rendering gives rise to:
  - Physical interpretability and verifiability
  - Better generalization
  - Controllable generation



# Unsupervised Learning of Probably Symmetric Deformable 3D Objects from Images in the Wild

*CVPR 2020*

Shangzhe Wu

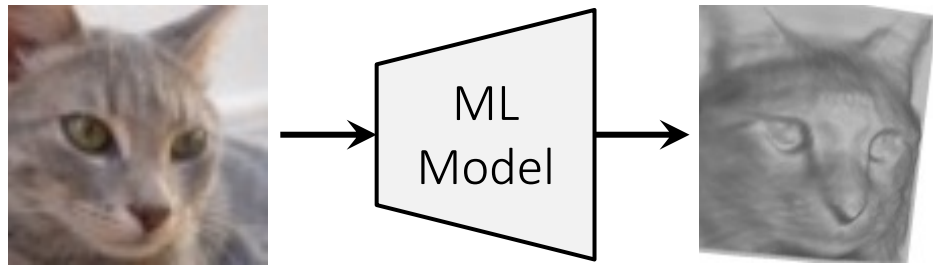
Christian Rupprecht

Andrea Vedaldi

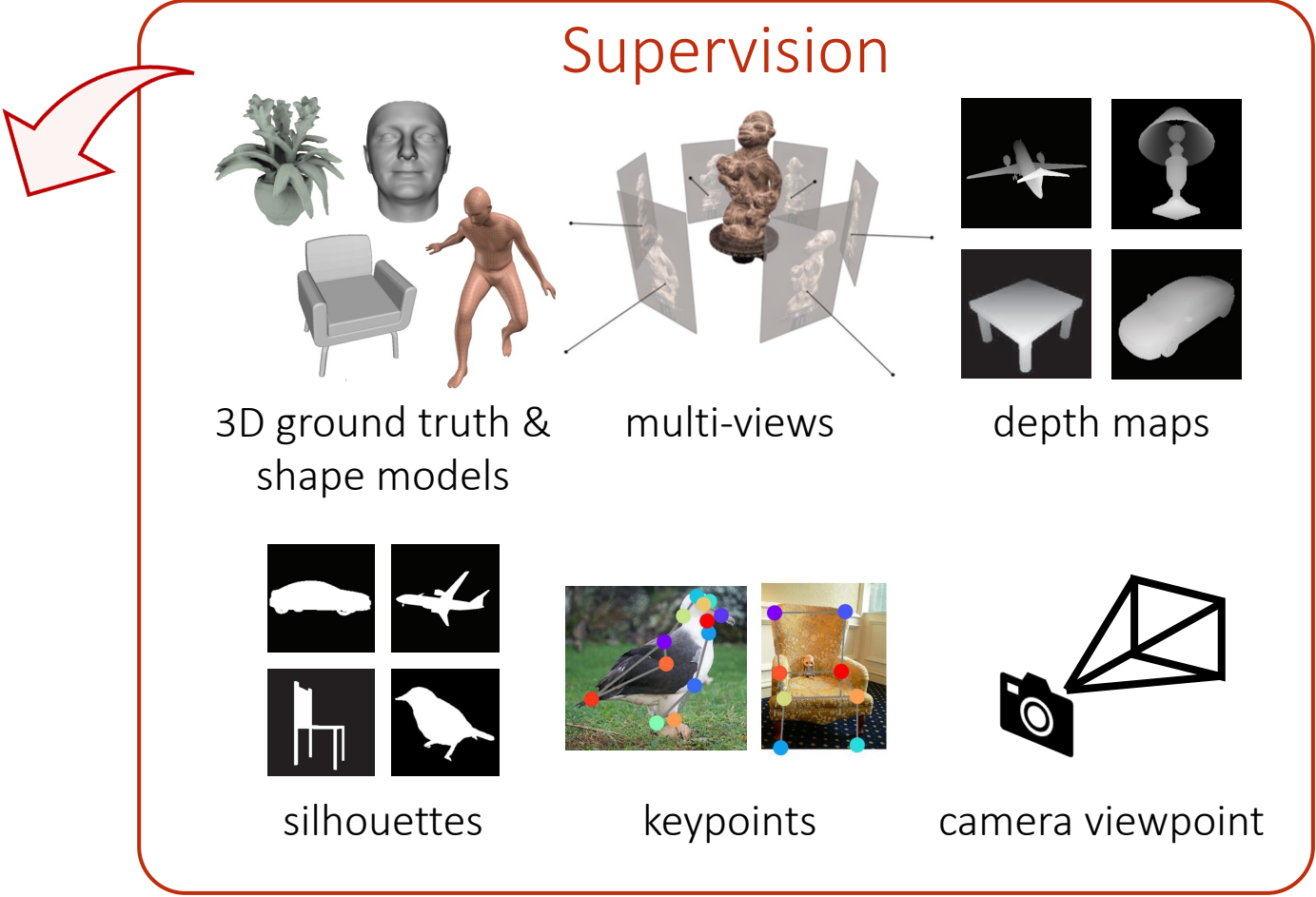


UNIVERSITY OF  
**OXFORD**

# Learning-based Single-view 3D Reconstruction



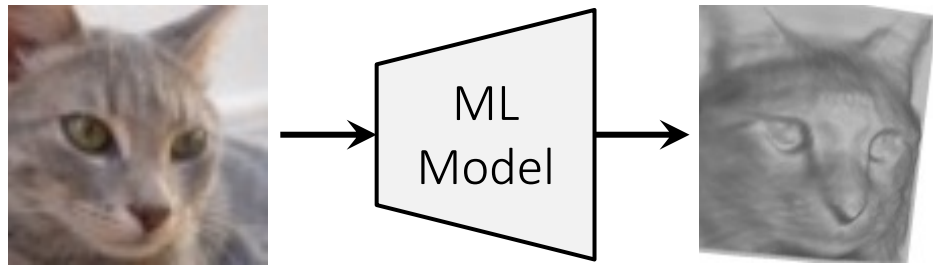
3D priors learned during training



3D annotations are expensive!



# Unsupervised Single-view 3D Reconstruction



3D priors learned during training

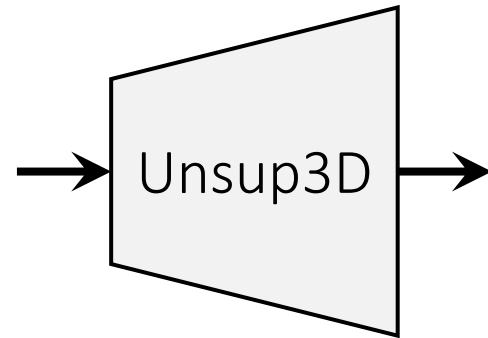


# Unsupervised Learning of Symmetric 3D Objects

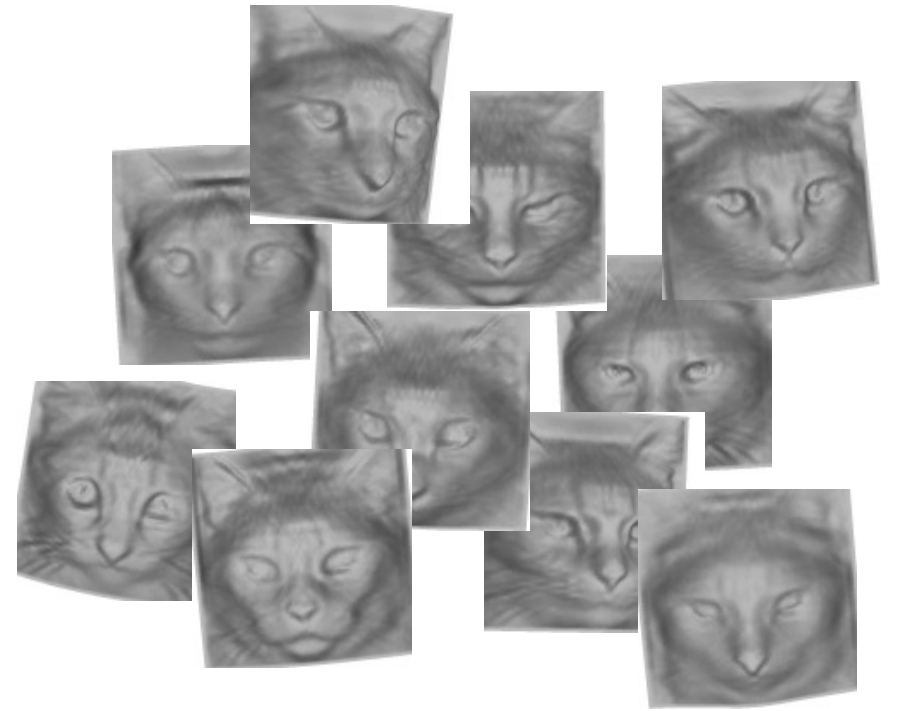
Training Data



single-view images of a category  
NO other supervision!

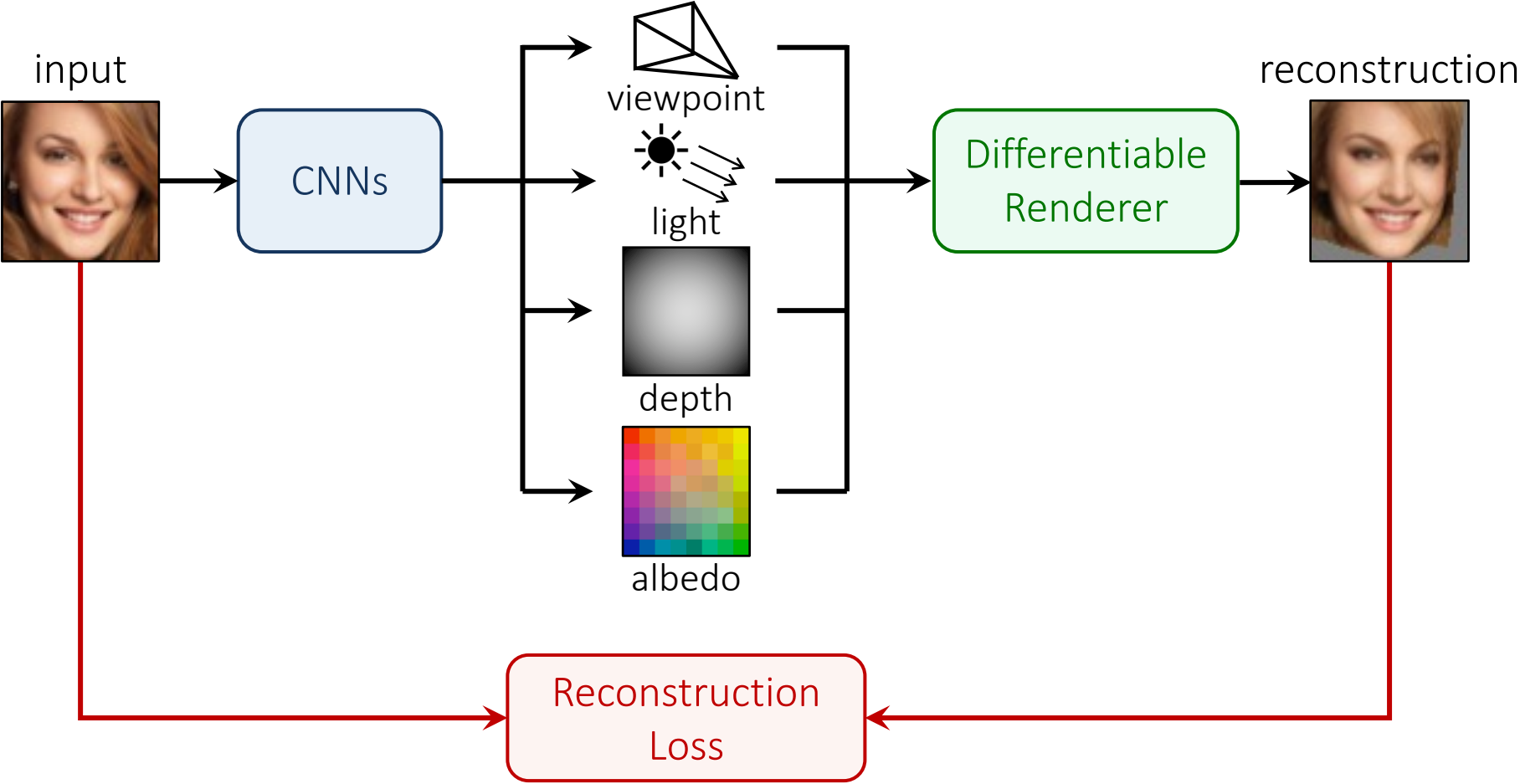


Output



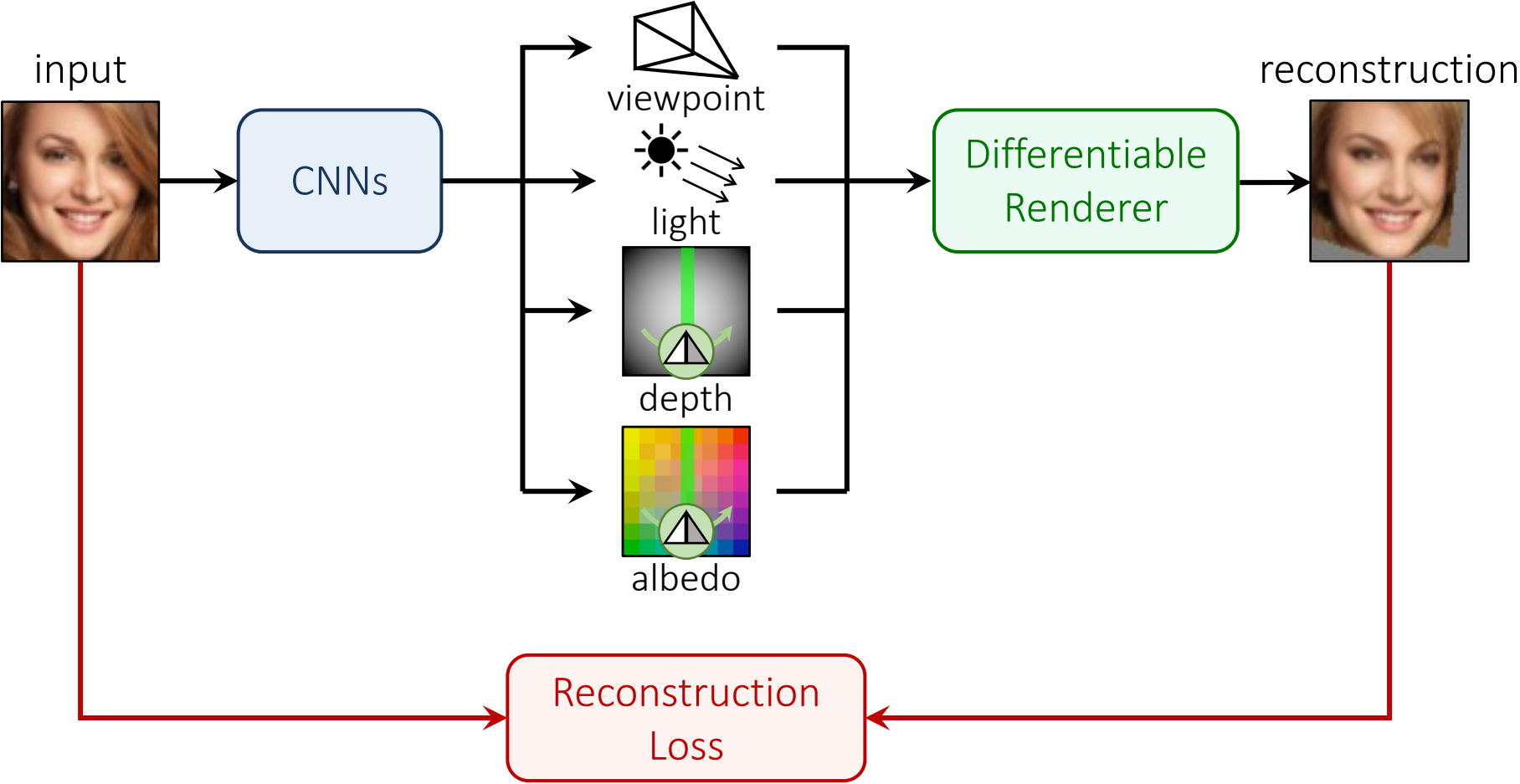
single image 3D reconstruction

# Photo-Geometric Autoencoding

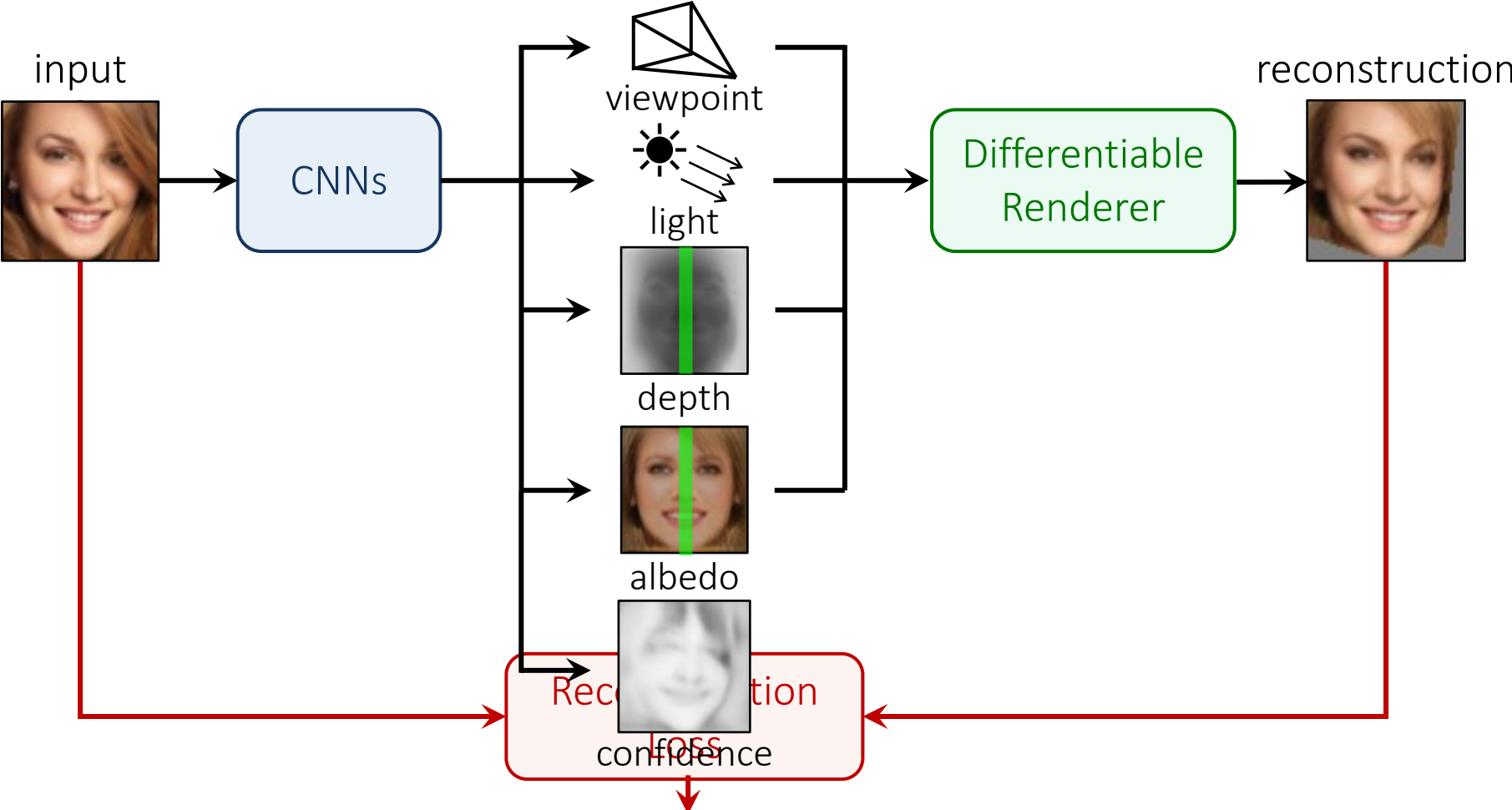


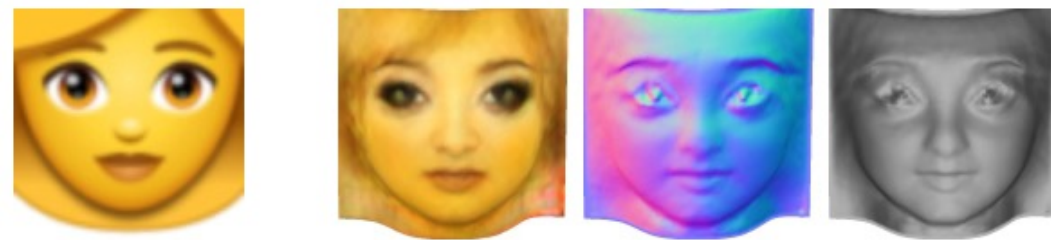
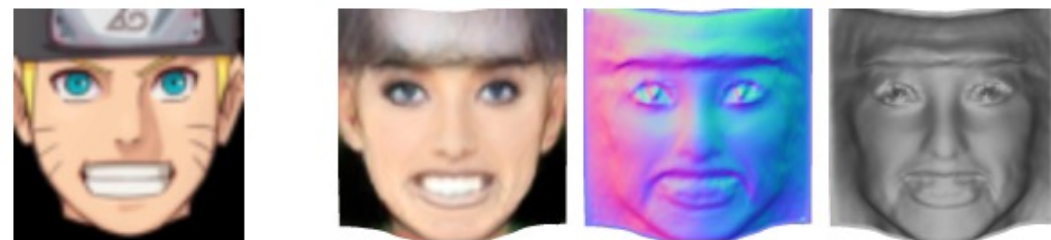
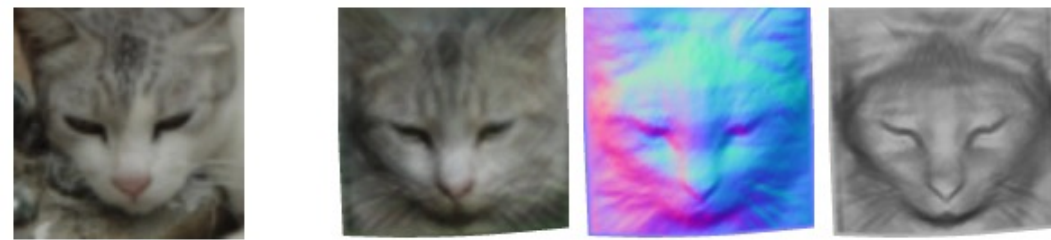


# Photo-Geometric Autoencoding



# Photo-Geometric Autoencoding with Symmetry





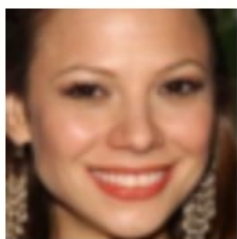
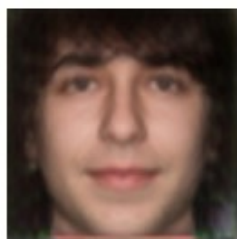
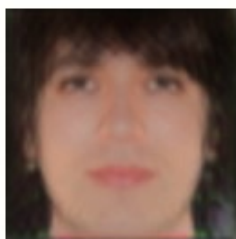
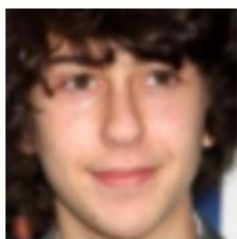
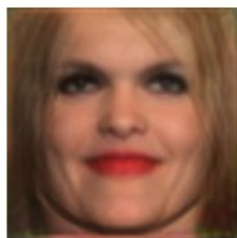
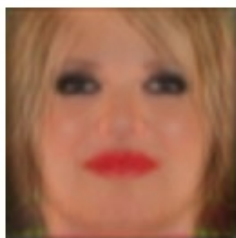
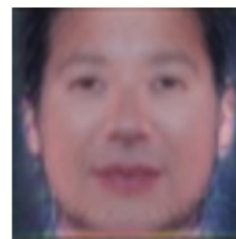
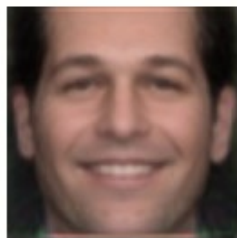
input

reconstruction

input

reconstruction





input

decompose & relight

input

decompose & relight<sup>23</sup>

# MagicPony: Learning Articulated 3D Animals in the Wild

Shangzhe Wu\* Ruining Li\* Tomas Jakob\* Christian Rupprecht Andrea Vedaldi

Visual Geometry Group, University of Oxford

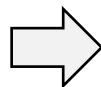
(\* Equal Contribution)

CVPR 2023

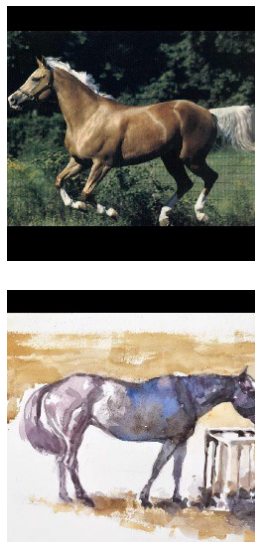
## Training



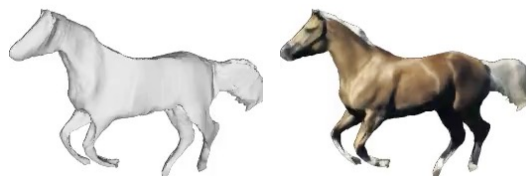
Single-view Images



## Single-Image Inference



Test Image



Articulated 3D Shape



Animation

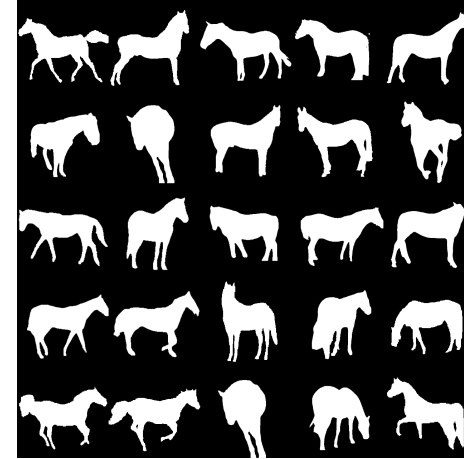


# Training Data

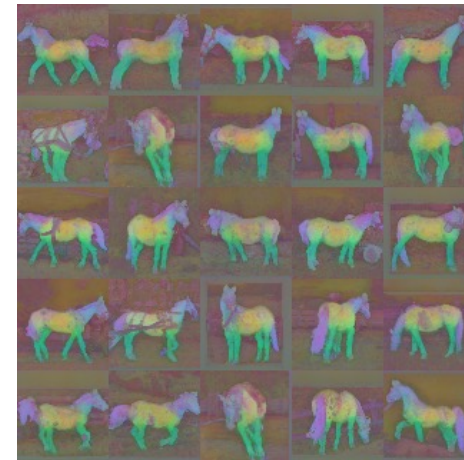


Single-view Images

No keypoint or viewpoint supervision,  
nor template shapes



Instance Masks



Self-supervised Image Features



# Correspondences from Self-supervised DINO Features



Self-supervised Image Features

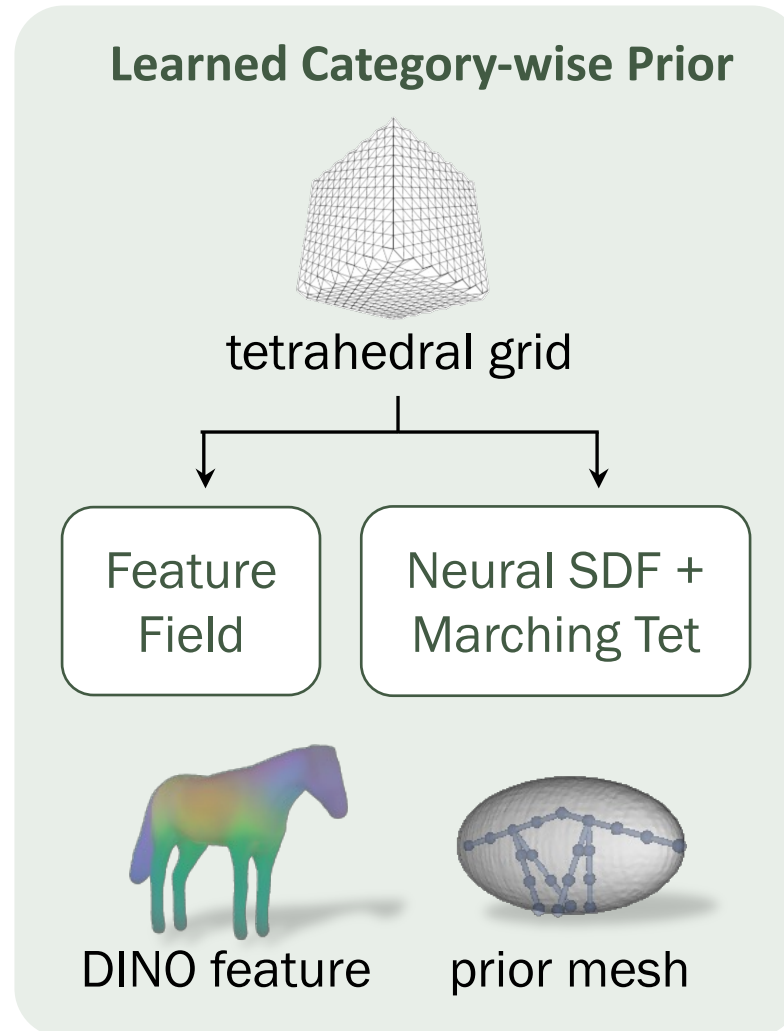
# Correspondences from Self-supervised DINO Features

Learned Category-wise Prior



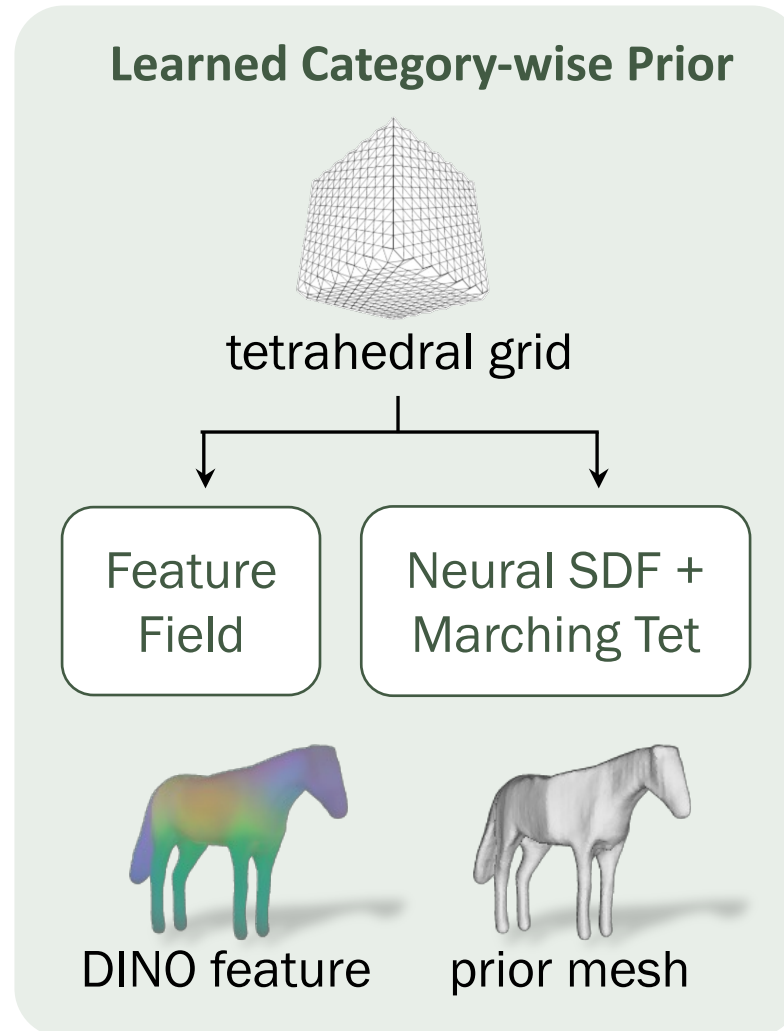
learned canonical DINO feature

# Implicit-Explicit 3D Representation





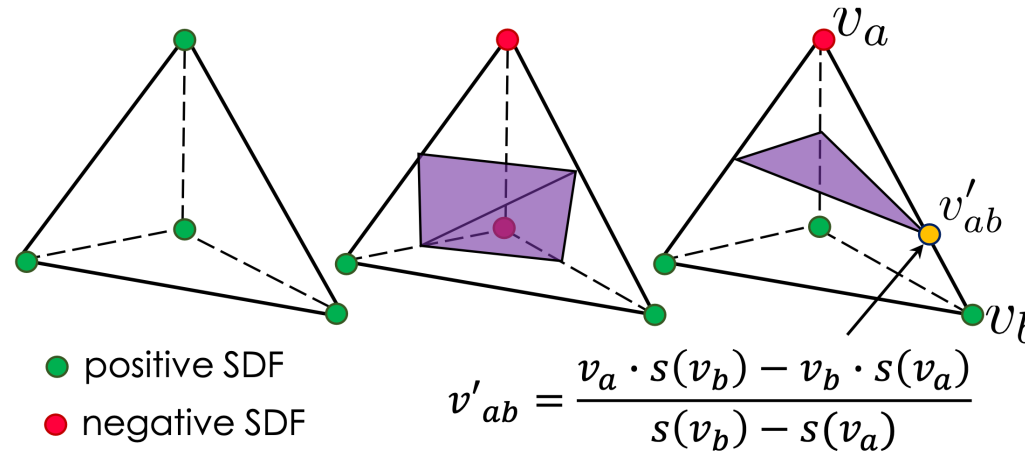
# Implicit-Explicit 3D Representation



# Implicit-Explicit 3D Representation

## Deep Marching Tetrahedra (DMTet)

Triangular meshes from Signed Distance Function (SDF)  $s(\cdot)$



**SDF**

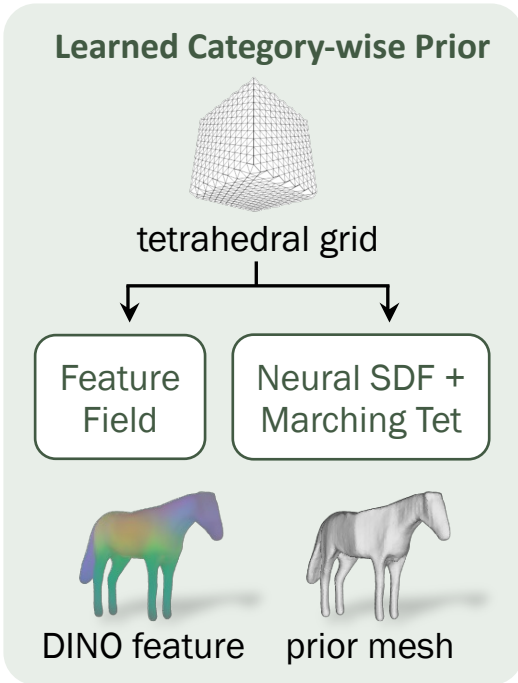
- ✓ Flexible topology
- ✓ Smooth gradients

**Mesh**

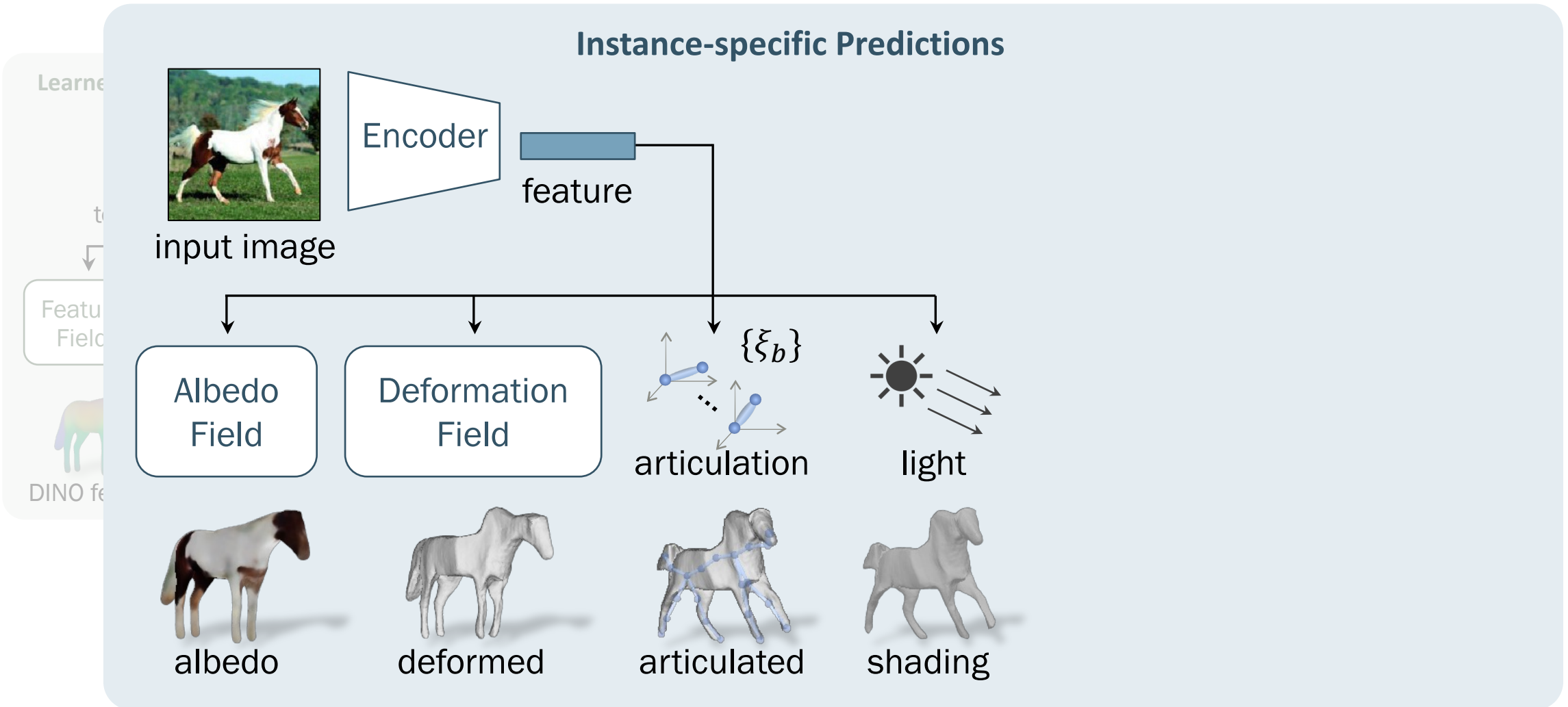
- ✓ Easy to render
- ✓ Easy to articulate

**DMTet**

- ✓ Differentiable
- ✓ Regular (no self-intersection)



# Hierarchical Shape Prediction



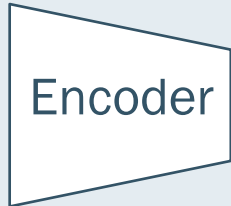


# Hierarchical Shape Prediction

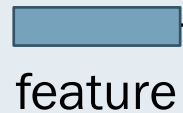
## Instance-specific Predictions



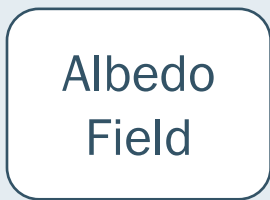
input image



Encoder



feature



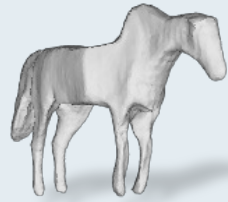
Albedo  
Field



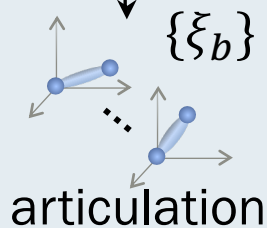
albedo



Deformation  
Field

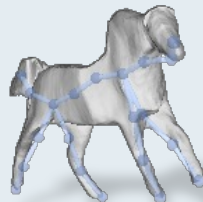


deformed

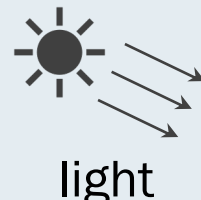


$\{\xi_b\}$

articulation



articulated

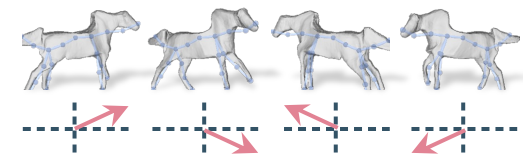
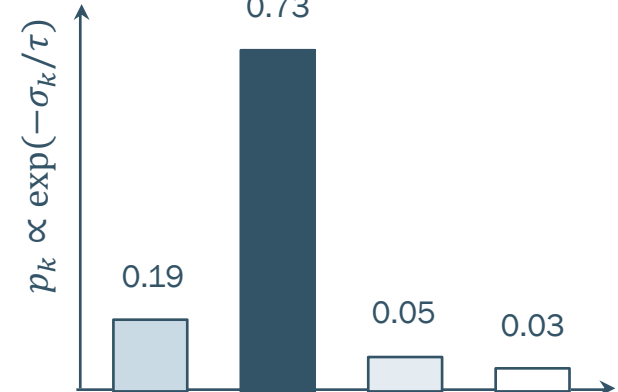


light



shading

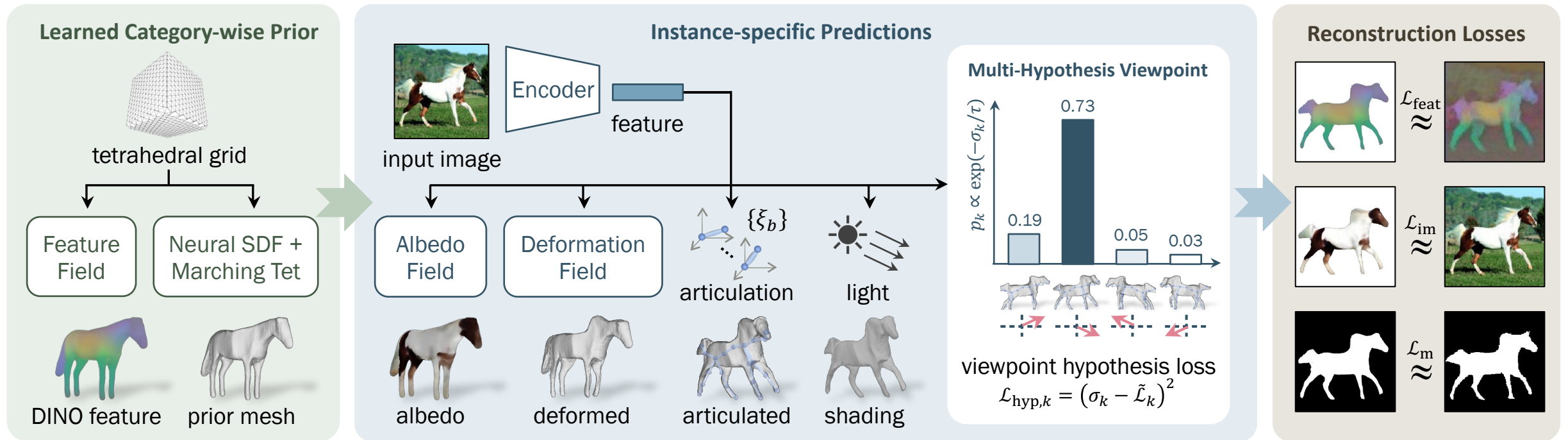
## Multi-Hypothesis Viewpoint



viewpoint hypothesis loss

$$\mathcal{L}_{\text{hyp},k} = (\sigma_k - \tilde{\mathcal{L}}_k)^2$$

# End-to-End Training with Image Rendering Losses

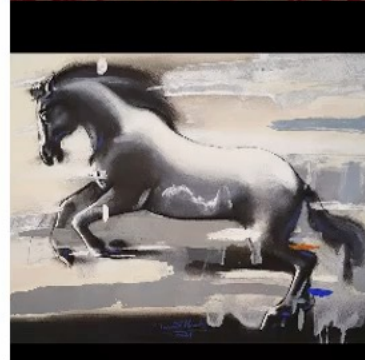
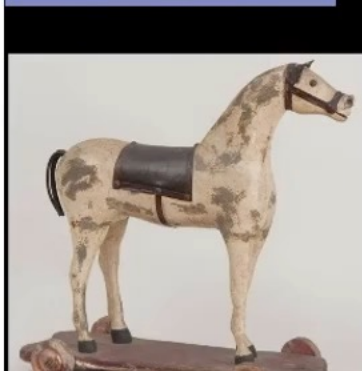
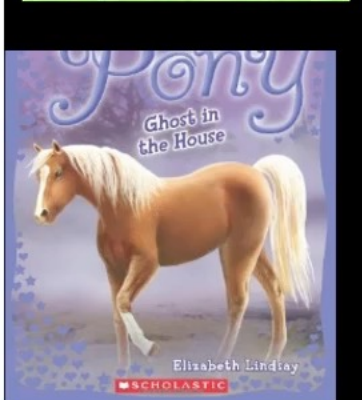
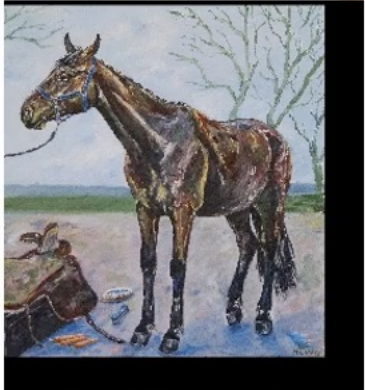
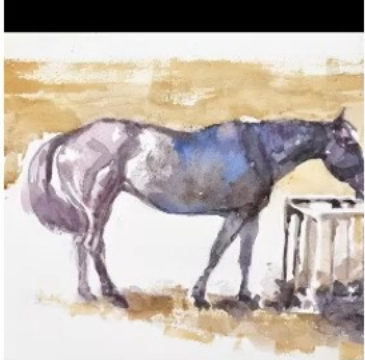


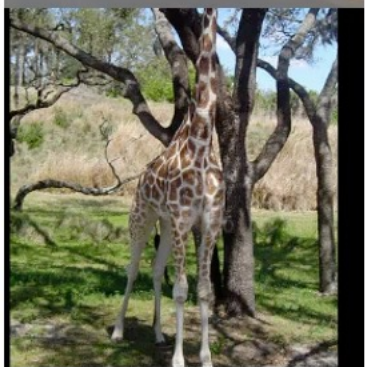
[1] Deep Marching Tetrahedra: a Hybrid Representation for High-Resolution 3D Shape Synthesis. Shen et. al. NeurIPS 2021.

[2] Emerging Properties in Self-supervised Vision Transformers. Caron et. al. ICCV 2021.







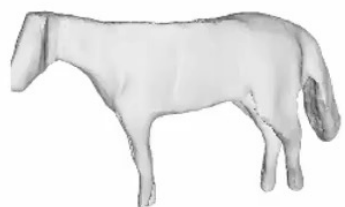
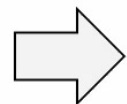
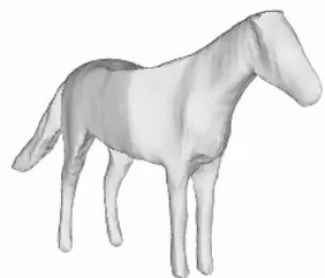
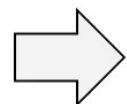
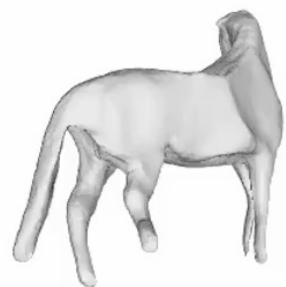
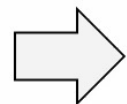




# Frame-by-Frame Inference on Videos

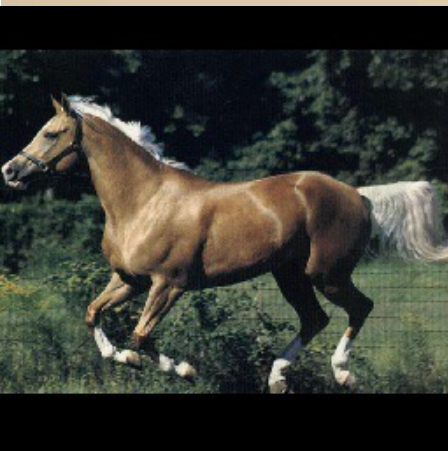


Input Frames



Input View

360° Rotations

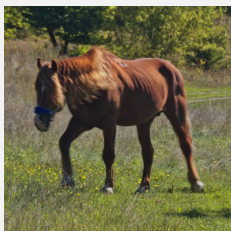
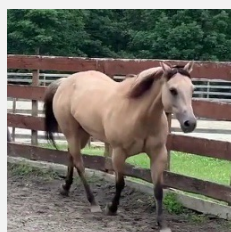
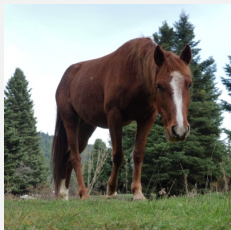


3D Printed Horse Reconstruction



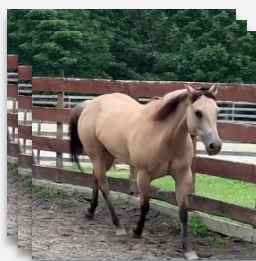
# Learning Articulated 3D Motion Prior

## Training Images

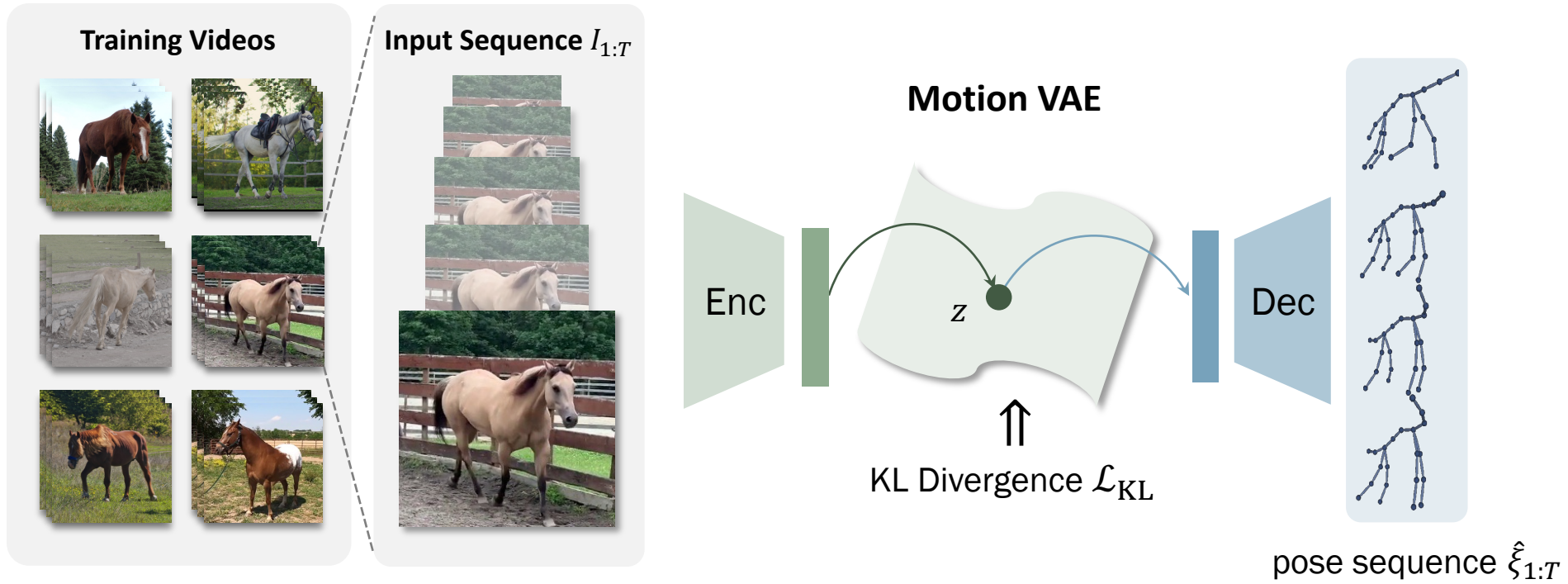


# Learning Articulated 3D Motion Prior

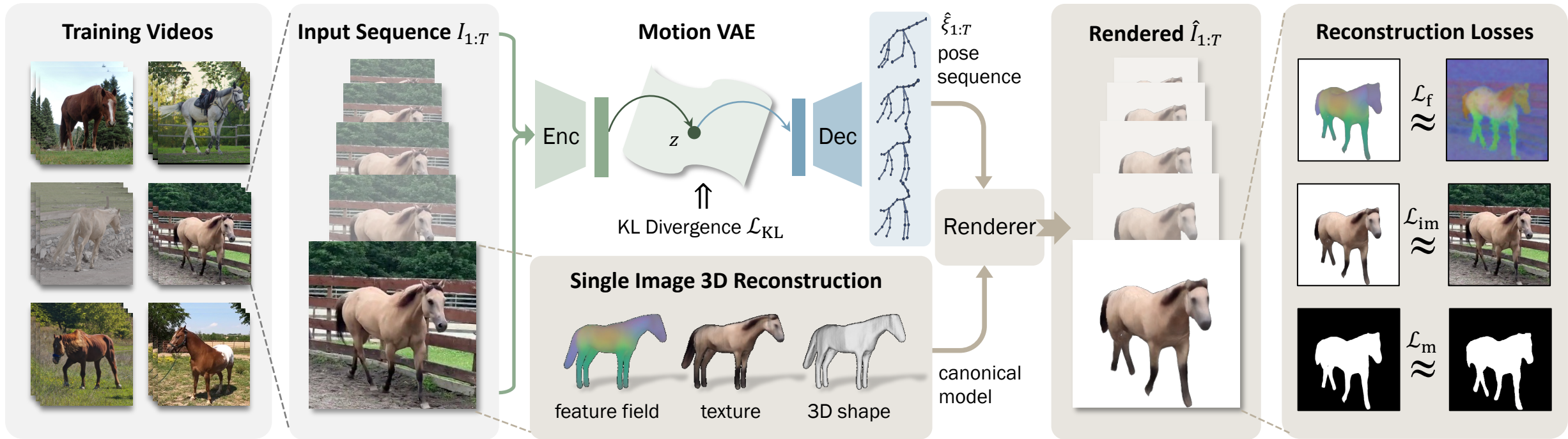
## Training Videos



# Learning Articulated 3D Motion Prior



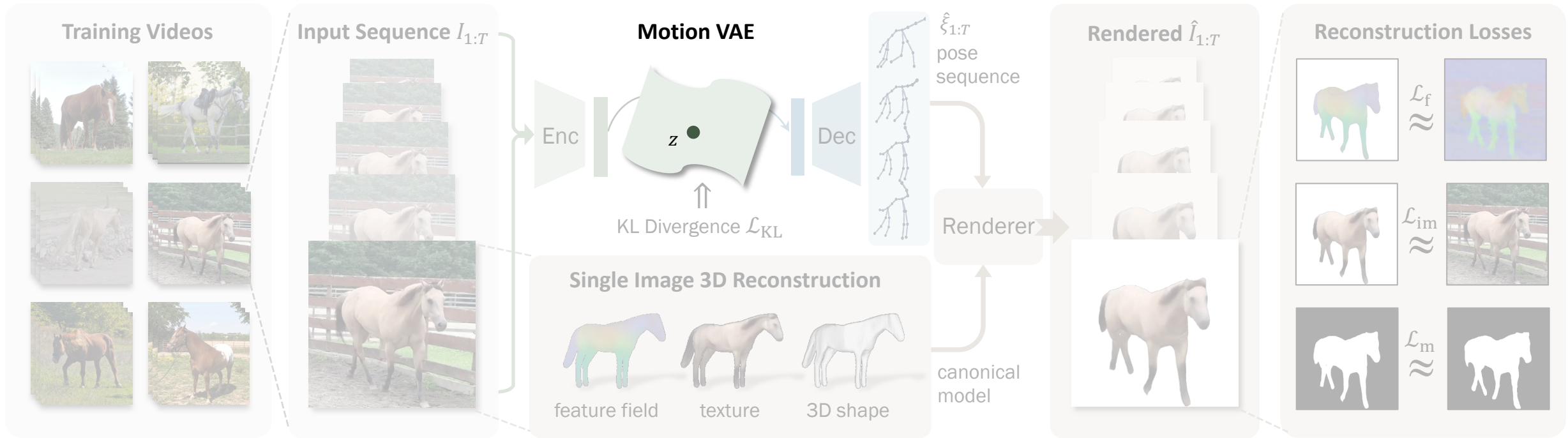
# Learning Articulated 3D Motion Prior



Trained with 2D reconstruction losses only without any pose annotations!



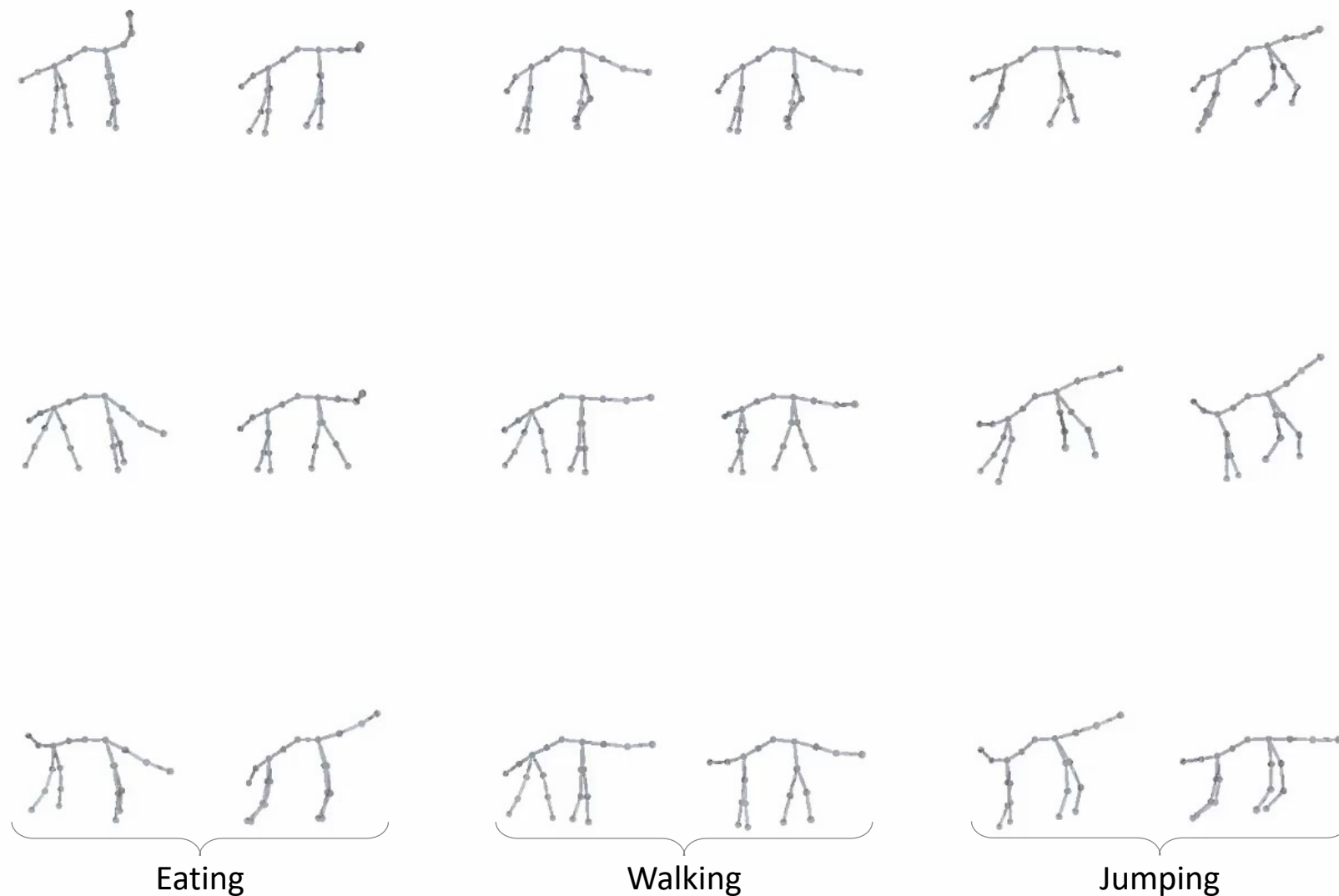
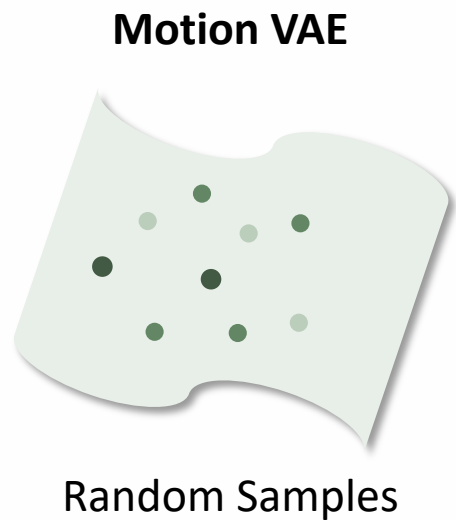
# Learning Articulated 3D Motion Prior



Trained with 2D reconstruction losses only without any pose annotations!

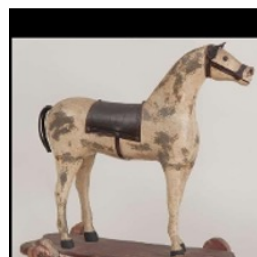
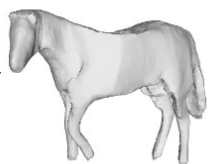
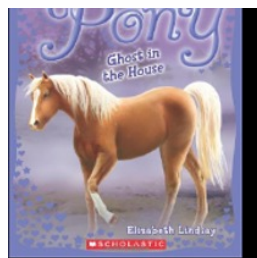
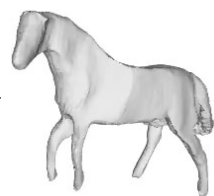
# Learning Articulated 3D Motion Prior

## Generated 3D Motion Sequences



**Input Image Reconstruction**

**Generated 3D Motion Sequences**



Eating

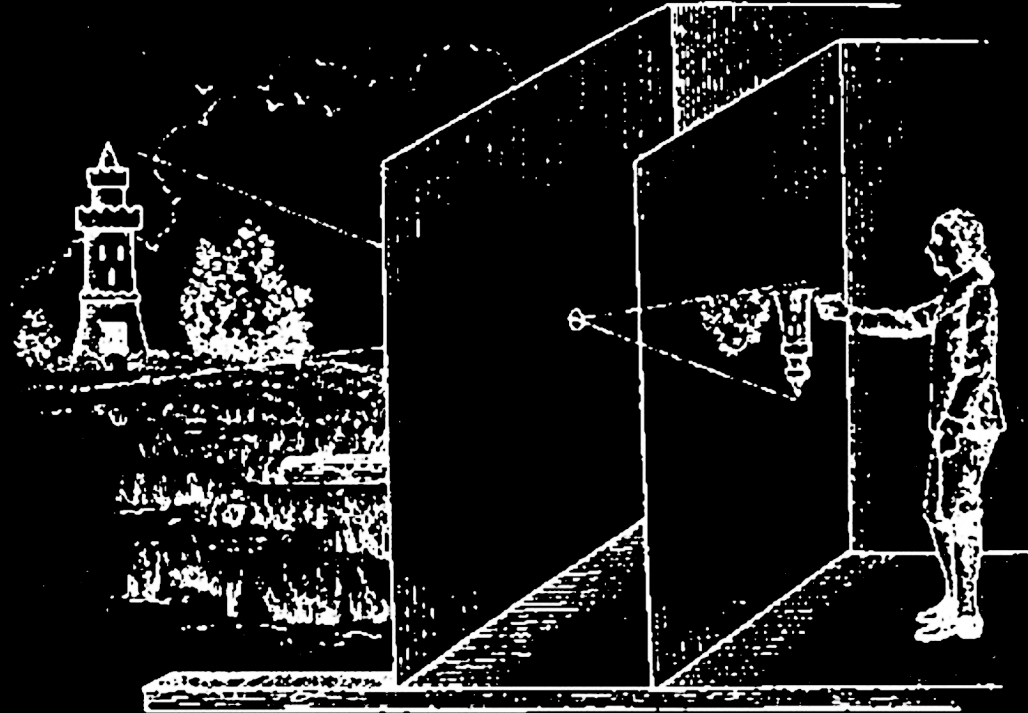


Walking



Jumping

# It's a ~~3D~~ World, After All Physical



Physics is the key to interpretability and generality!



# Learning Dynamic 3D Objects in the Wild



Stanford  
University

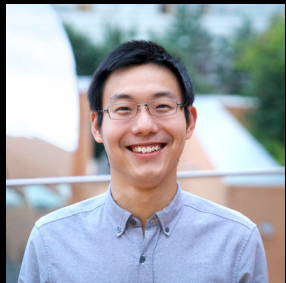
*Shangzhe Wu*

PostDoc at Stanford SVL



UNIVERSITY OF  
OXFORD

## Amazing Advisors & Collaborators



Jiajun Wu



Andrea Vedaldi



Christian Rupprecht



Noah Snavely



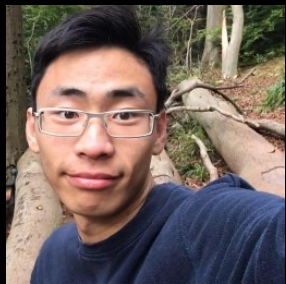
Yunzhi Zhang



Tomas Jakab



Ruining Li



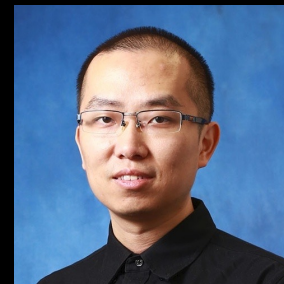
Zirui Wang



Felix Wimbauer



Keqiang Sun



Hongsheng Li



Ameesh Makadia



Richard Tucker



Angjoo Kanazawa