with audio

MagicPony: Learning Articulated 3D Animals in the Wild

Shangzhe Wu* Ruining Li* Tomas Jakab* Christian Rupprecht Andrea Vedaldi

Visual Geometry Group, University of Oxford

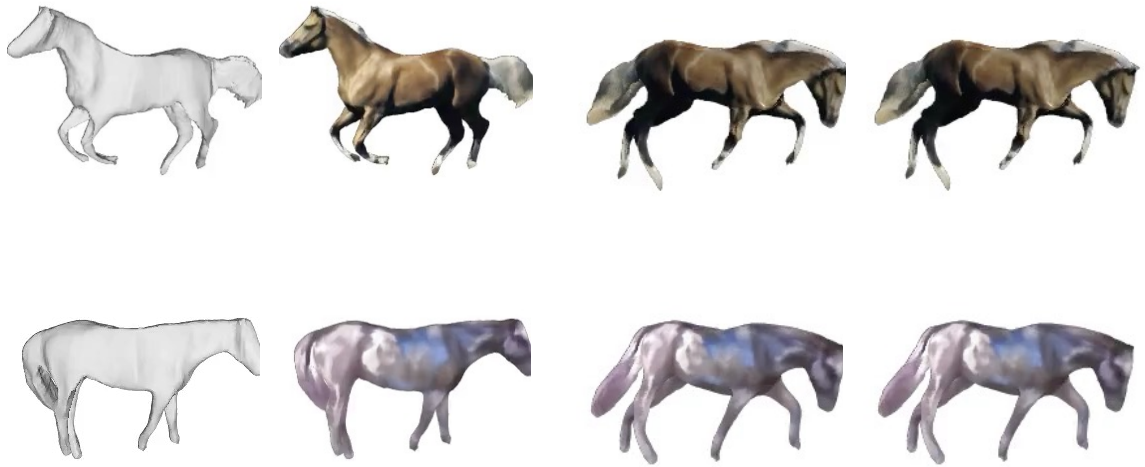
(* Equal Contribution)

Training

Single-Image Inference

Single-view Images

Test Image



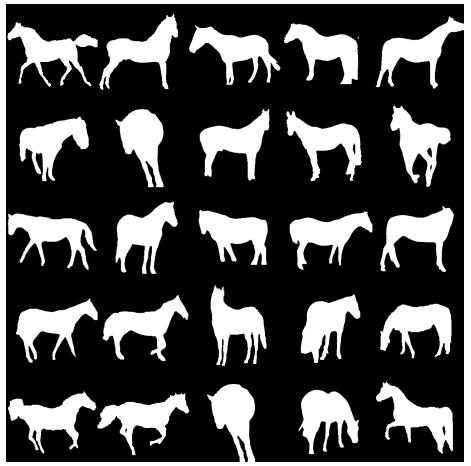
Articulated 3D Shape

Animation

Training Data

Single-view Images

Off-the-shelf PointRend [1]



Instance Masks

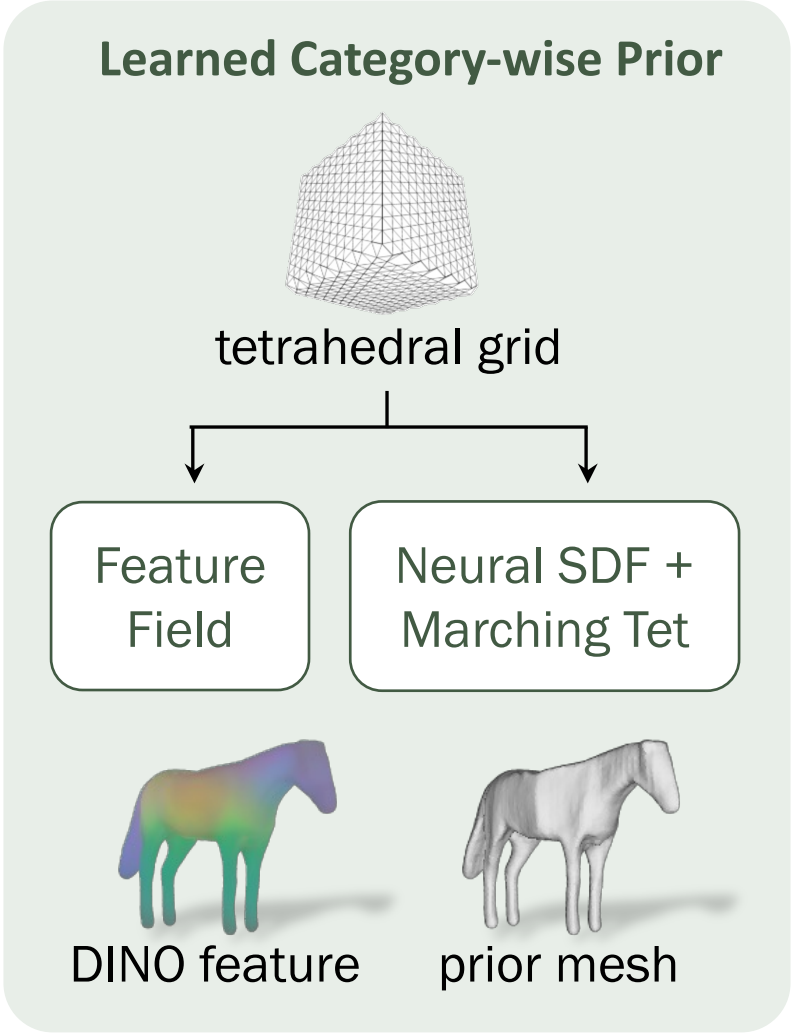
Off-the-shelf DINO-ViT [2]

Self-supervised Image Features

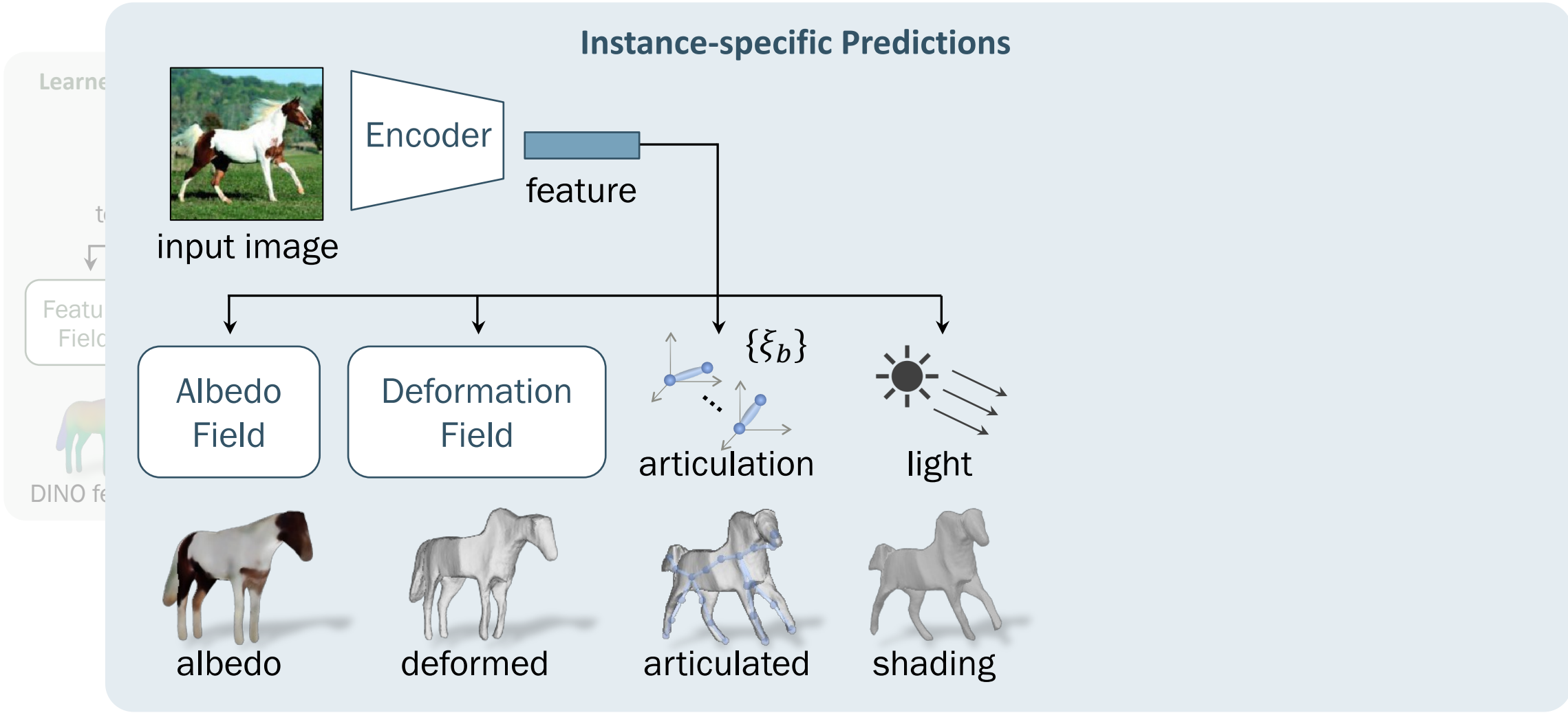
No keypoint or viewpoint supervision, nor template shapes

[1] PointRend: Image Segmentation as Rendering. Kirillov et. al. CVPR 2020. [2] Emerging Properties in Self-supervised Vision Transformers. Caron et. al. ICCV 2021.

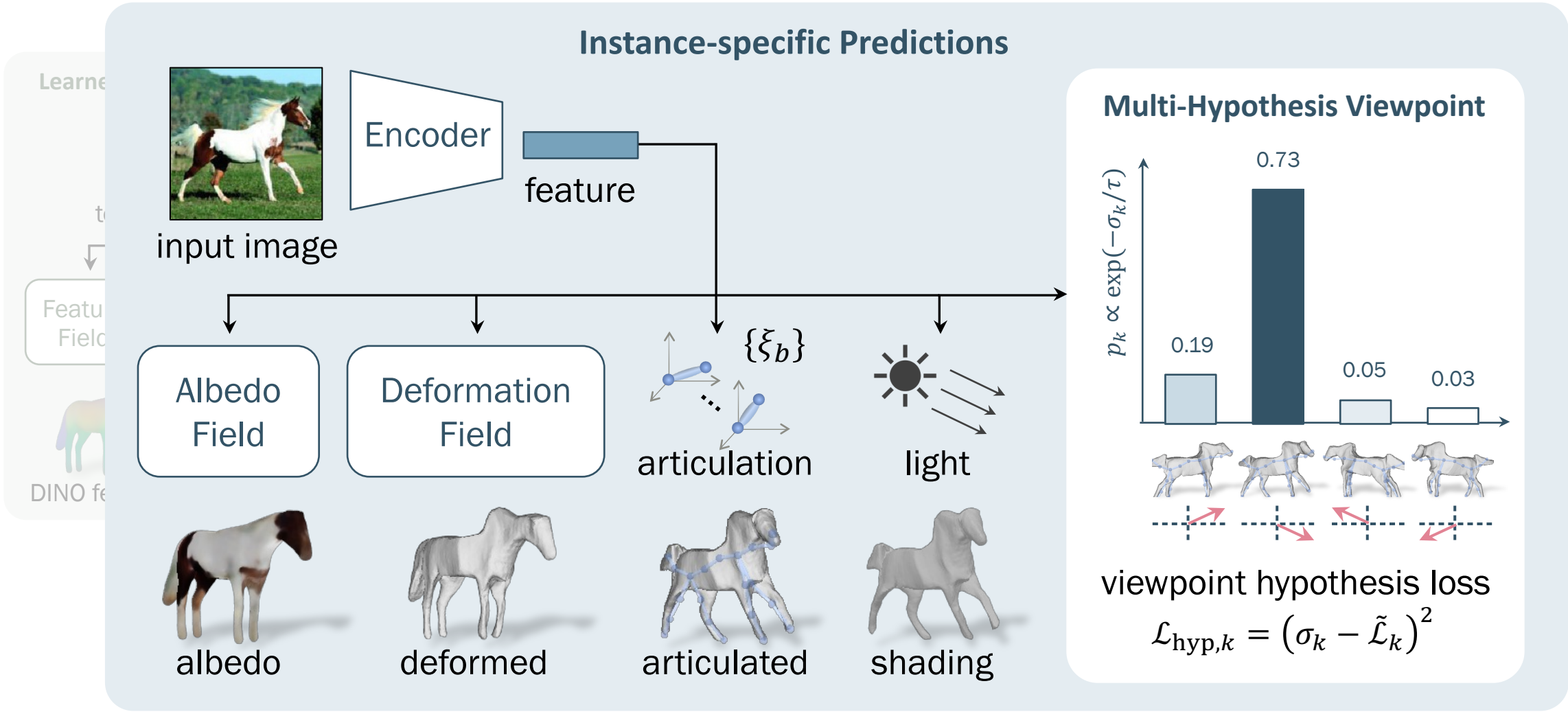
Implicit-Explicit 3D Representation



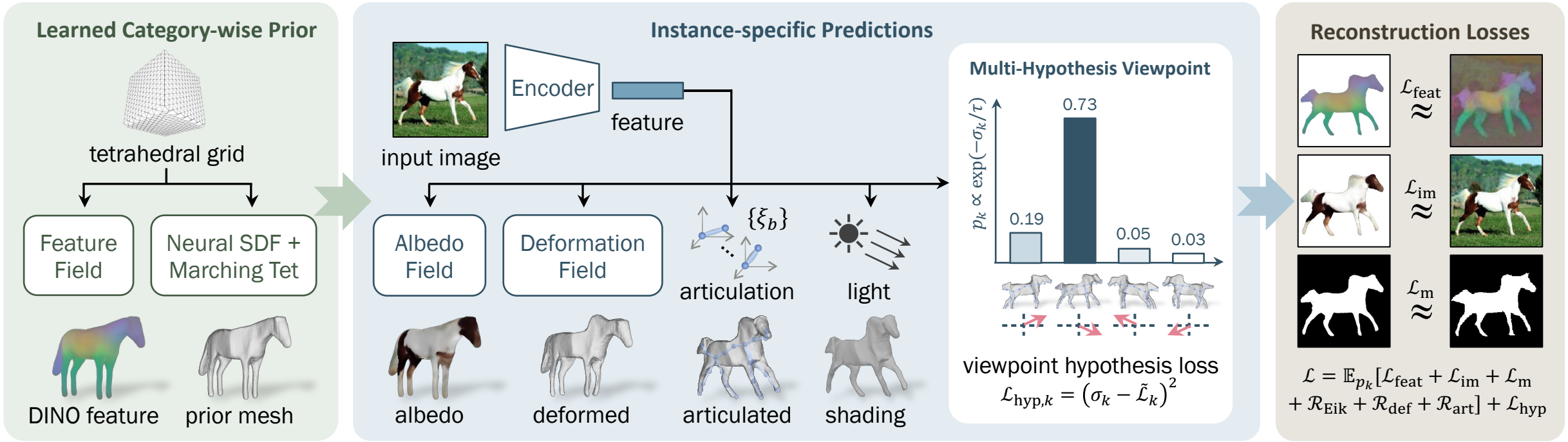
Hierarchical Shape Prediction



Hierarchical Shape Prediction



Implicit-Explicit 3D Representation

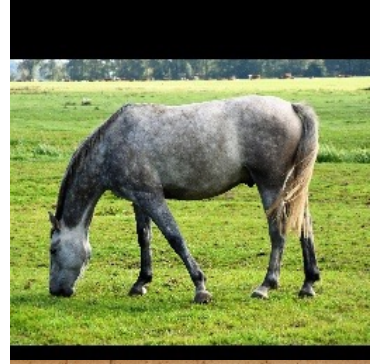
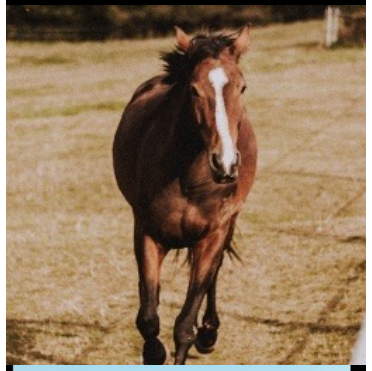


Entire pipeline trained end-to-end with reconstruction losses

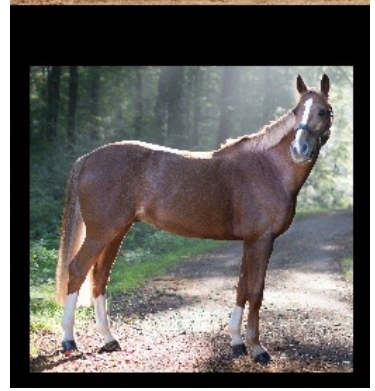
(except for frozen DINO-ViT [1] image encoder, pre-trained via self-supervision)

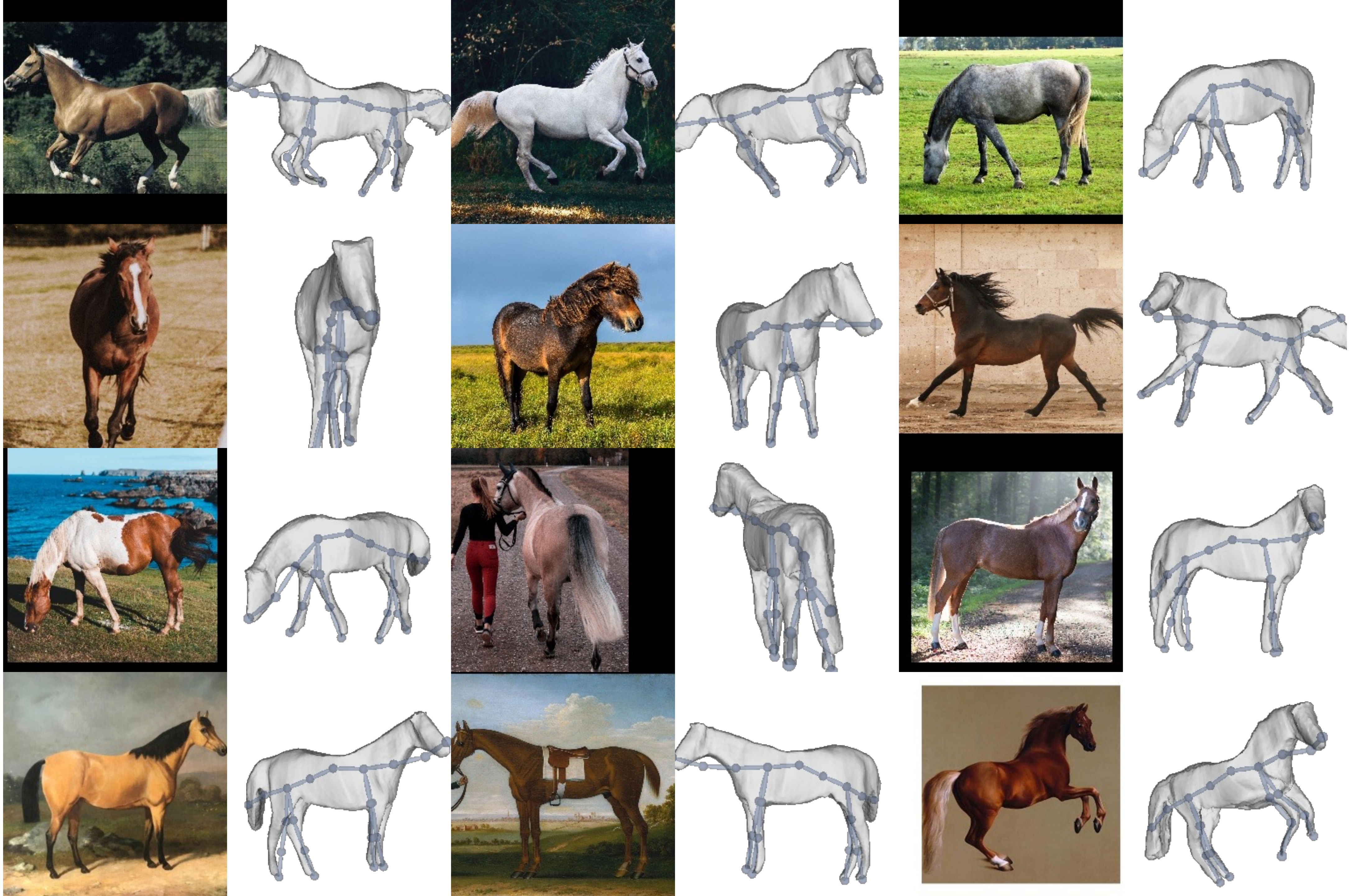
[1] Emerging Properties in Self-supervised Vision Transformers. Caron et. al. ICCV 2021.

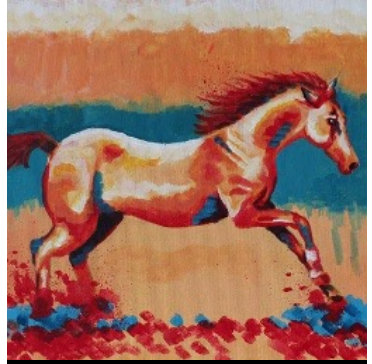


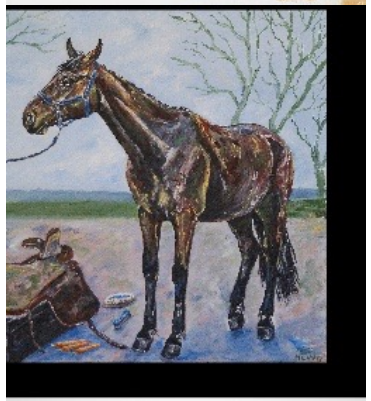



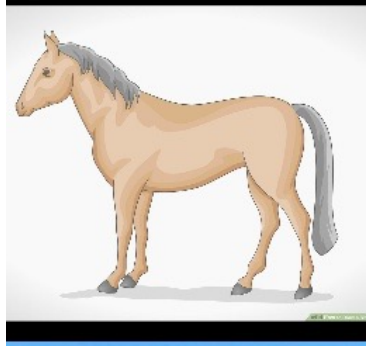


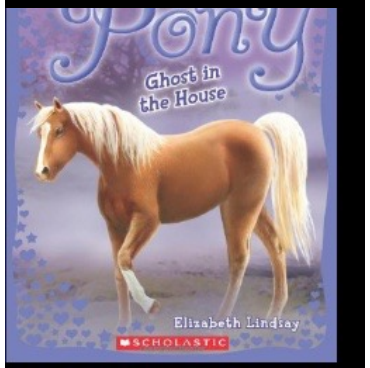



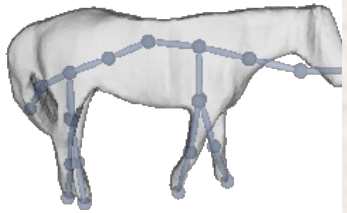



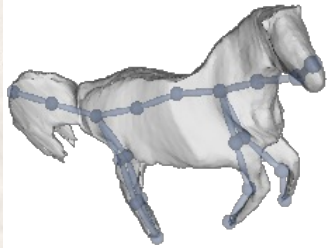
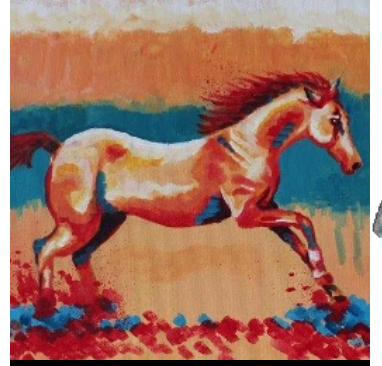
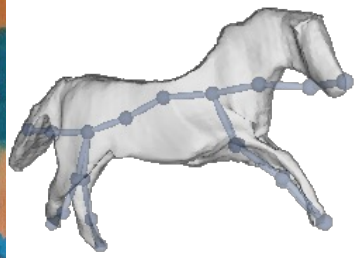
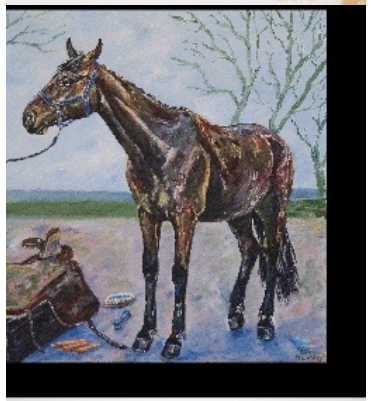
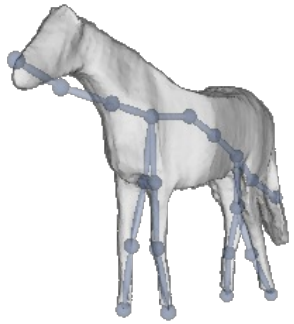
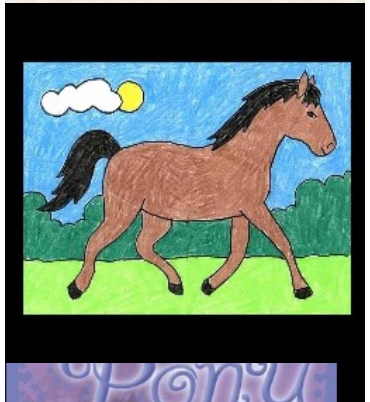
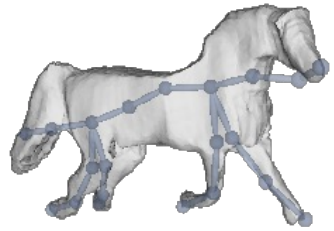


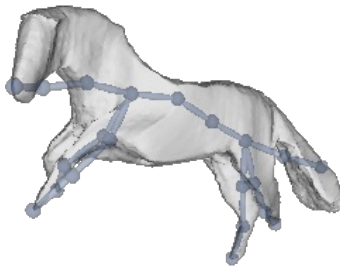
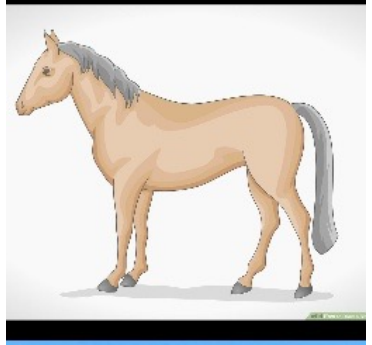
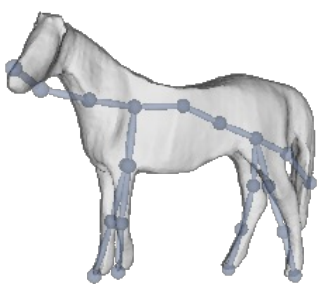
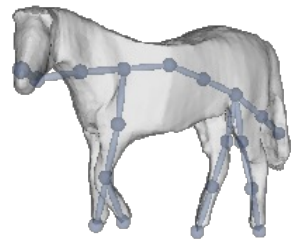


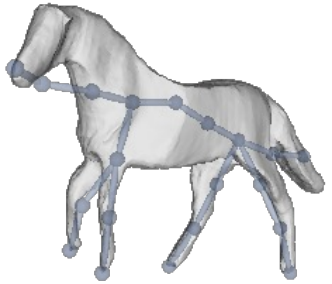


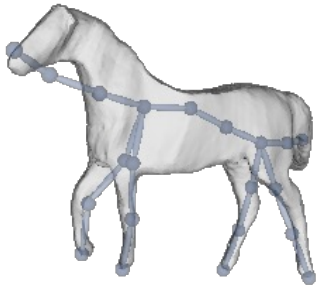

Pony
Ghost in the House
Elizabeth Lindsay
SCHOLASTIC

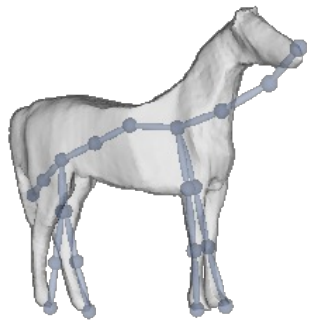



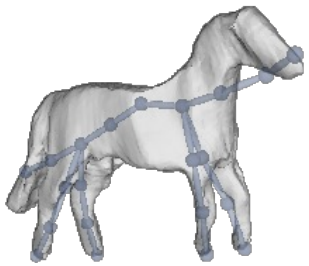






Pony
Ghost in the House
Elizabeth Lindsay
SCHOLASTIC




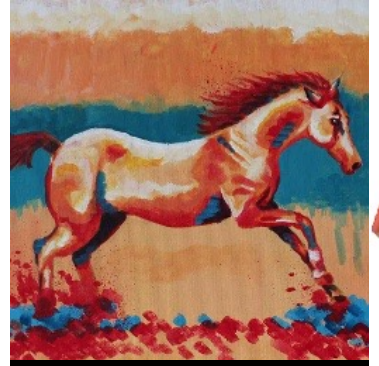


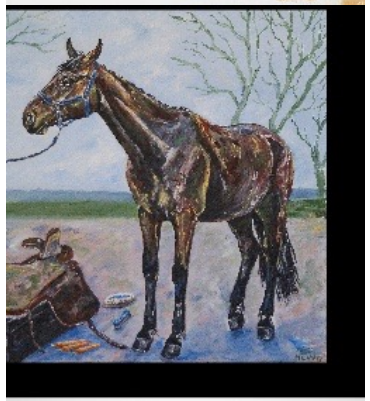






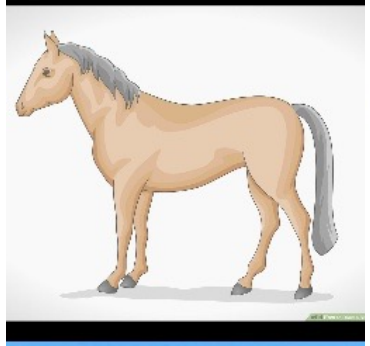


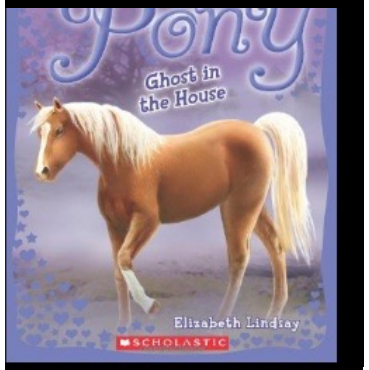











Pony
Ghost in the House
Elizabeth Lindsay
SCHOLASTIC

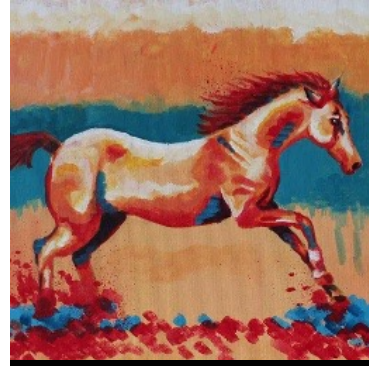


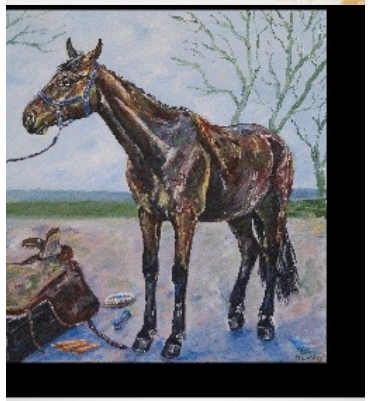






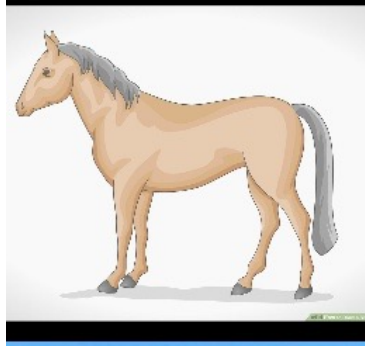


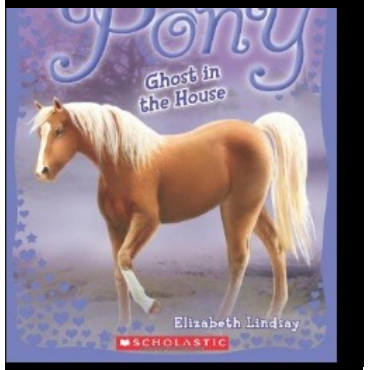









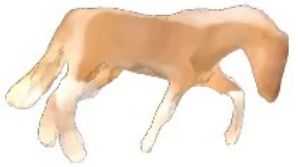


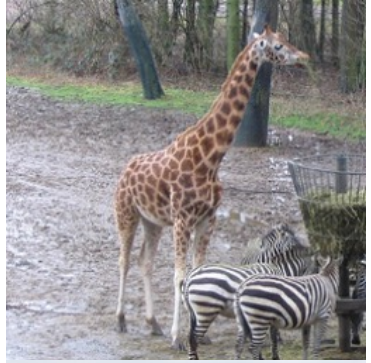

Pony
Ghost in the House
Elizabeth Lindsay
SCHOLASTIC

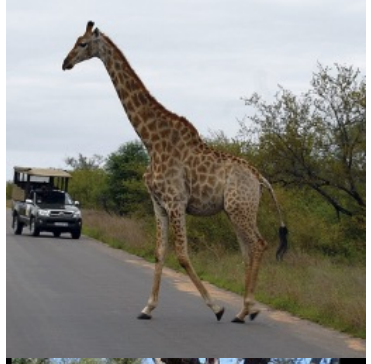
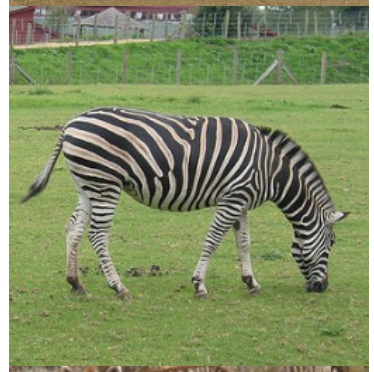
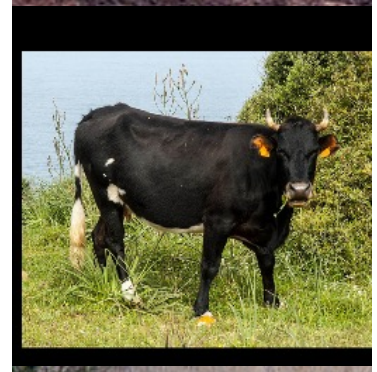




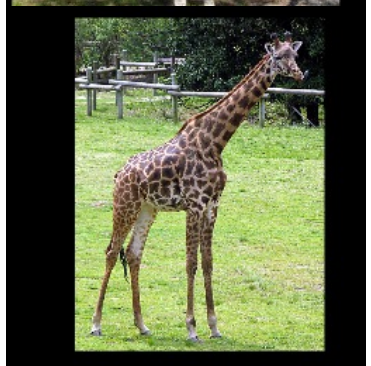
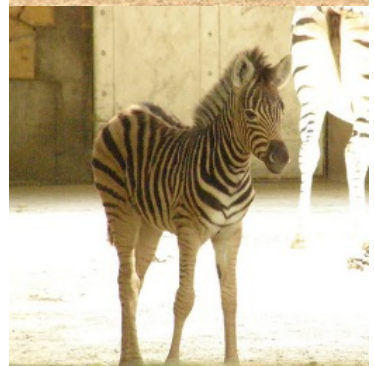


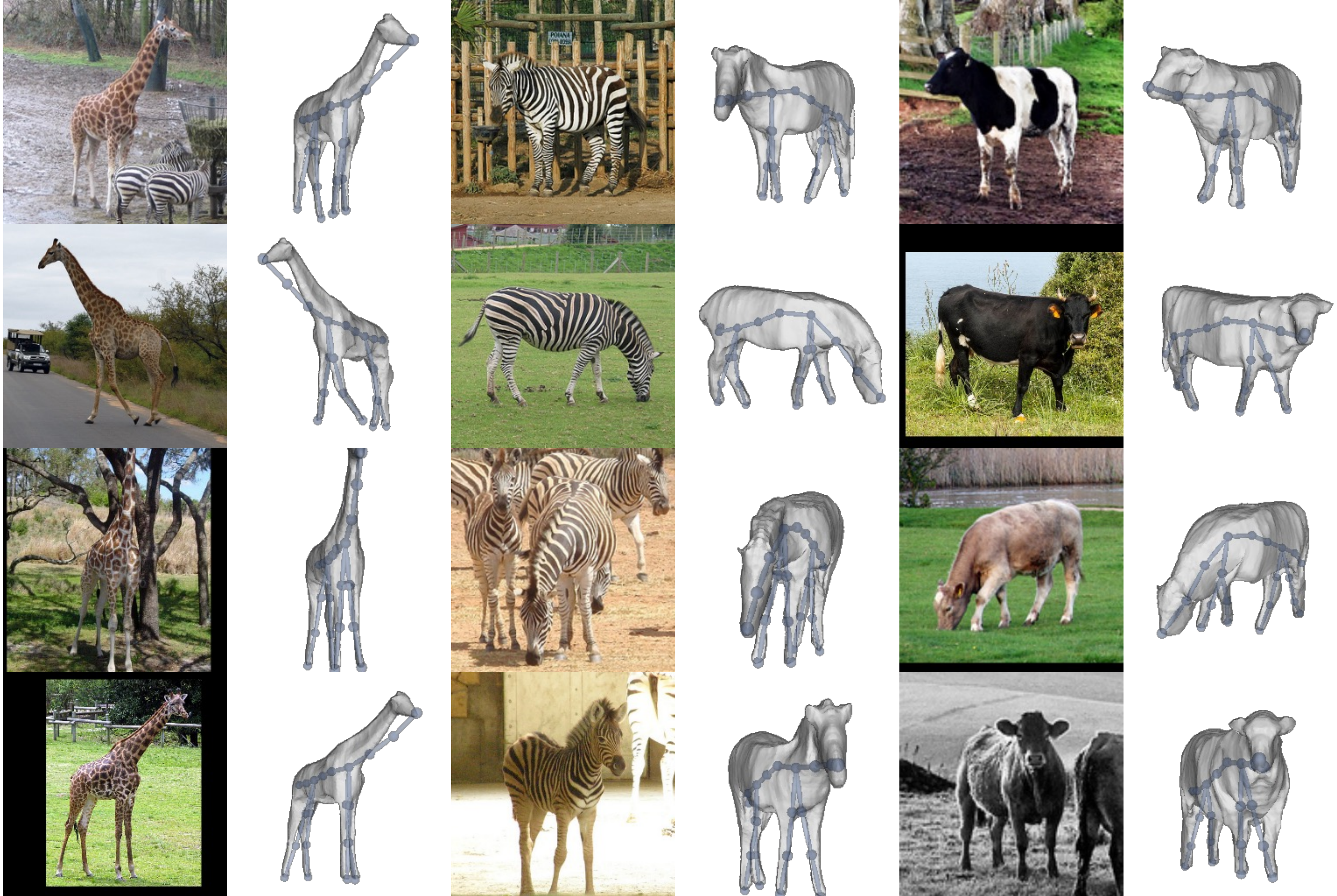


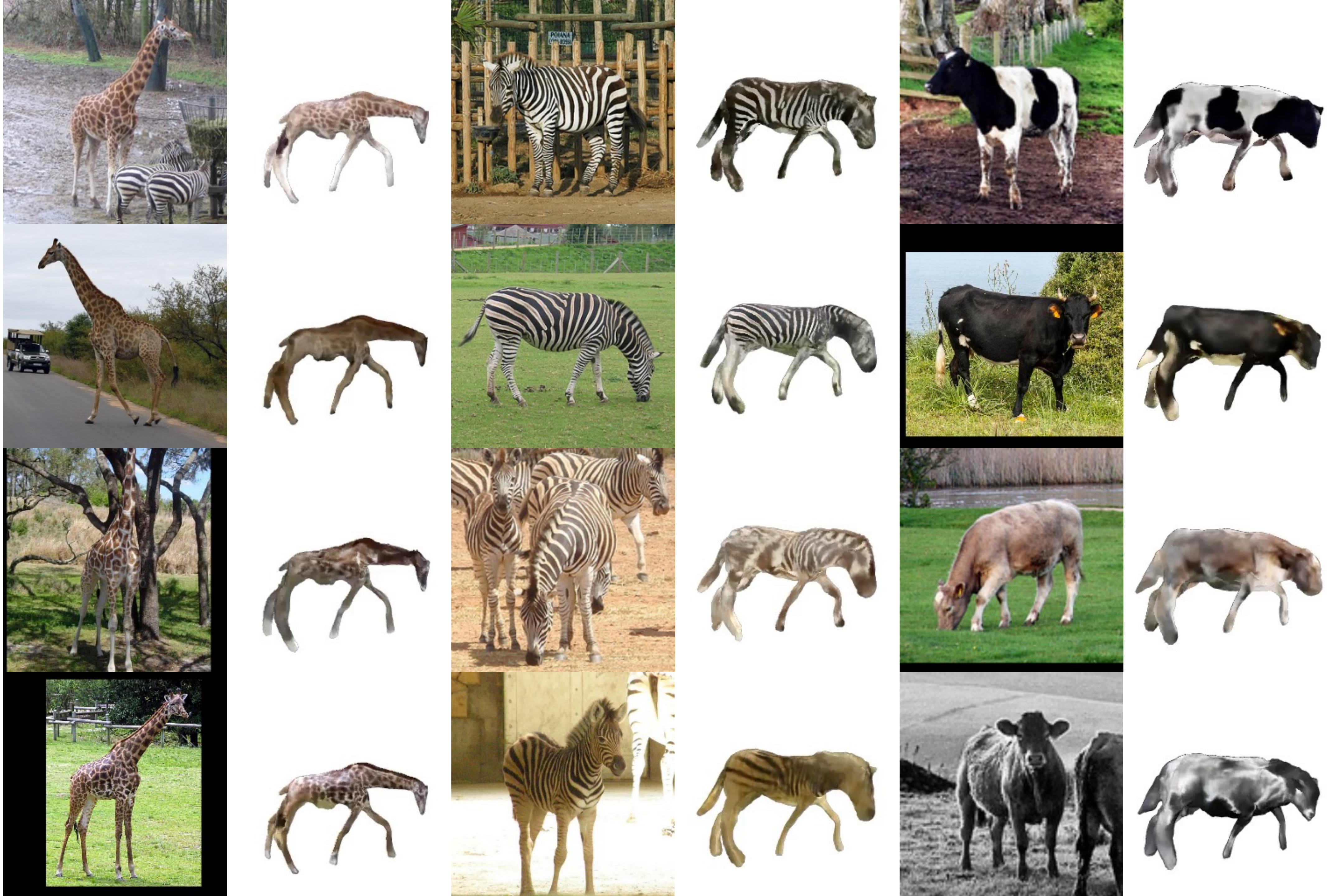


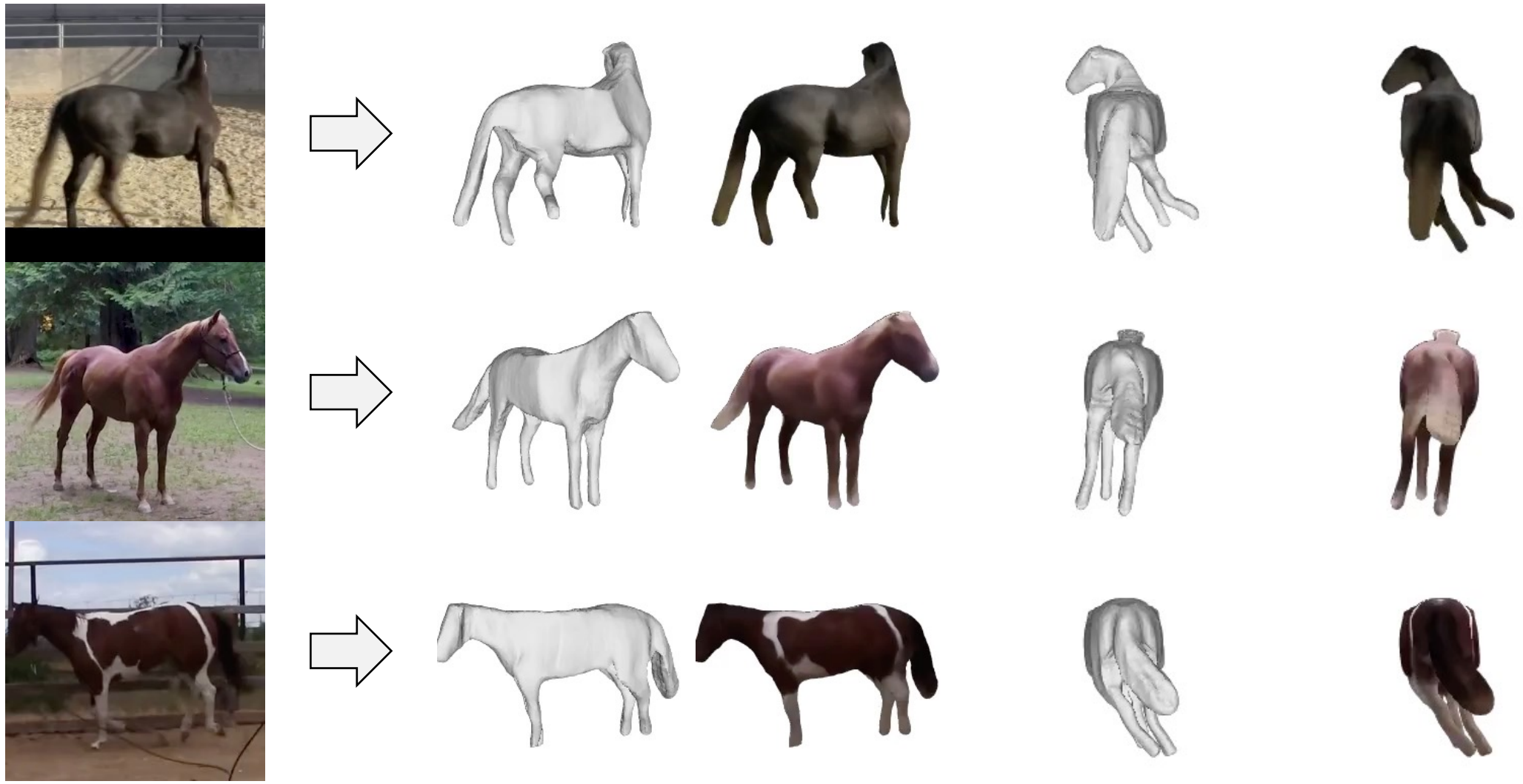




Frame-by-Frame Inference on Videos



Input Frames

Input View

360° Rotations

3D Printed Horse Reconstruction