LASSIE: Learning Articulated Shape from Sparse Image Ensemble via 3D Part Discovery

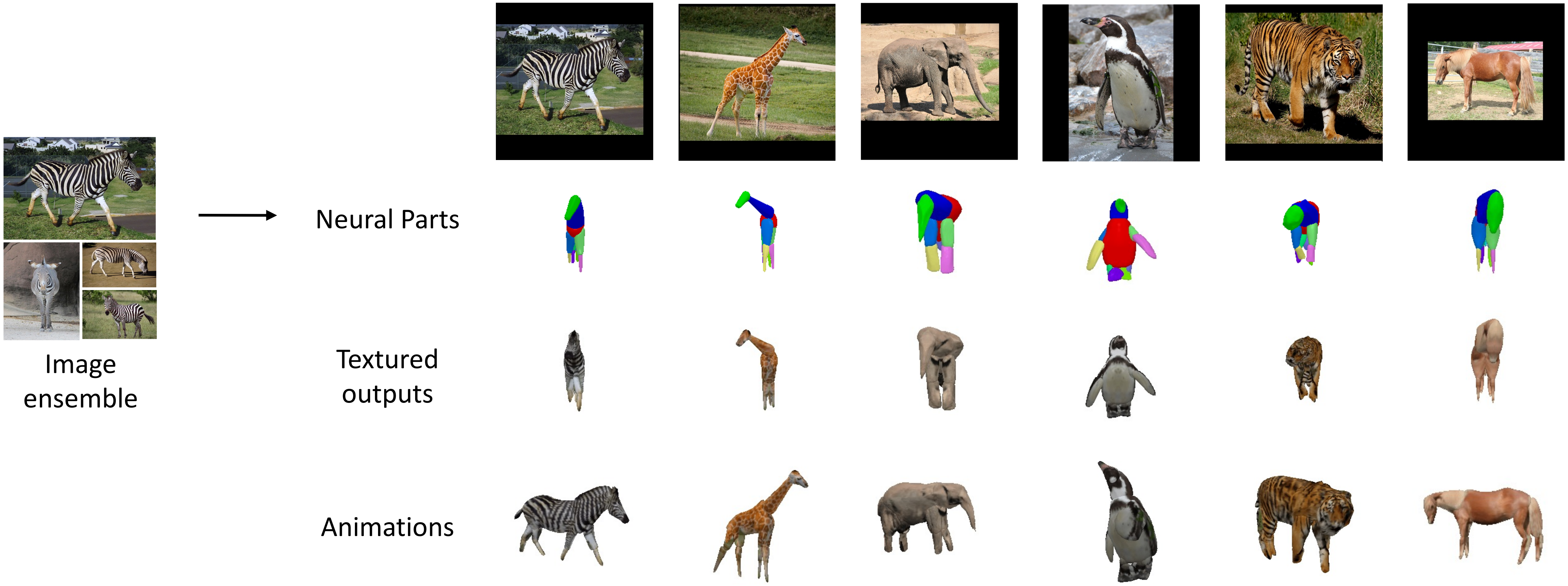


Image ensemble

Neural Parts

Textured outputs

Animations

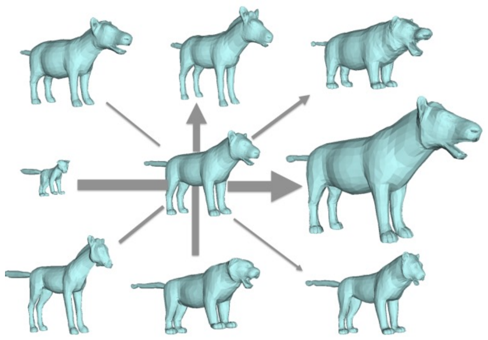
Chun-Han Yao[1], Wei-Chih Hung[2], Yuanzhen Li[3], Michael Rubinstein[3], Ming-Hsuan Yang[134], Varun Jampani[3]

[1]UC Merced, [2]Waymo, [3]Google Research, [4]Yonsei University

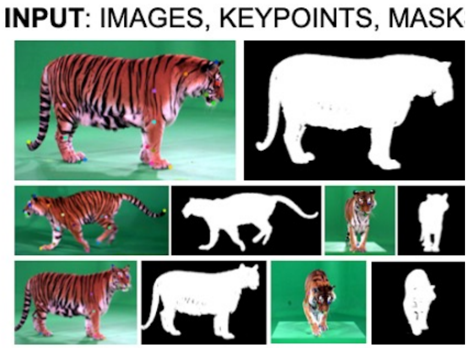
Optimizing Articulated Shapes from Sparse Images

Prior works

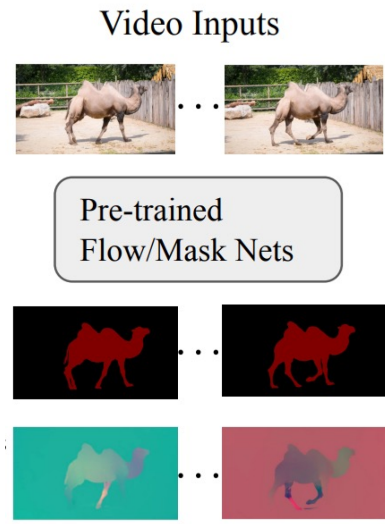
- Rely on statistical shape model, human annotations on images, or temporal correspondence in videos



SMAL [1]



SMALR [2]



LASR [3]

---

[1] Zuffi, Silvia, et al. "3D menagerie: Modeling the 3D shape and pose of animals." *CVPR*. 2017.

[2] Zuffi, Silvia, Angjoo Kanazawa, and Michael J. Black. "Lions and tigers and bears: Capturing non-rigid, 3d, articulated shape from images." *CVPR*. 2018.

[3] Yang, Gengshan, et al. "Lasr: Learning articulated shape reconstruction from a monocular video." *CVPR*. 2021.

Optimizing Articulated Shapes from Sparse Images

Prior works

- Rely on statistical shape model, human annotations on images, or temporal correspondence in videos

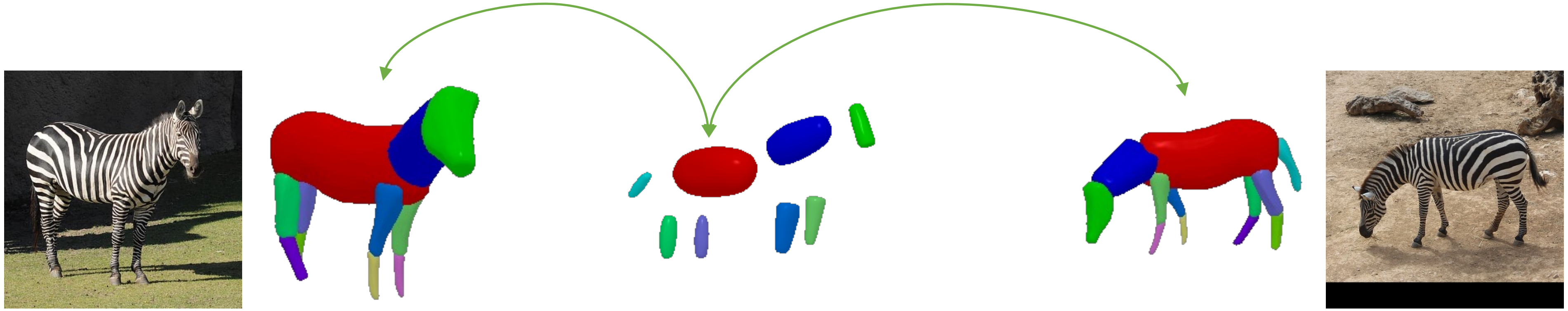
LASSIE

- **Inputs:** 10-30 in-the-wild images of an animal class
- **Outputs**: camera viewpoint, pose articulation, and deformable shape
- **No** pre-defined shape model, per-image annotations, or temporal information

Discovering 3D Neural Parts

Why reconstruct 3D parts?

- Simple geometry and rigid motion
- Semantic consistency across instances, articulation, viewpoints



Discovering 3D Neural Parts

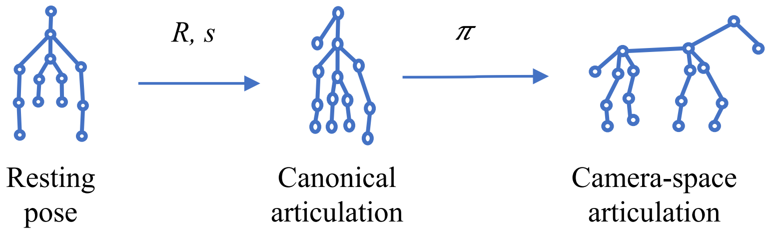
Why reconstruct 3D parts?

- Simple geometry and rigid motion
- Semantic consistency across instances, articulation, viewpoints

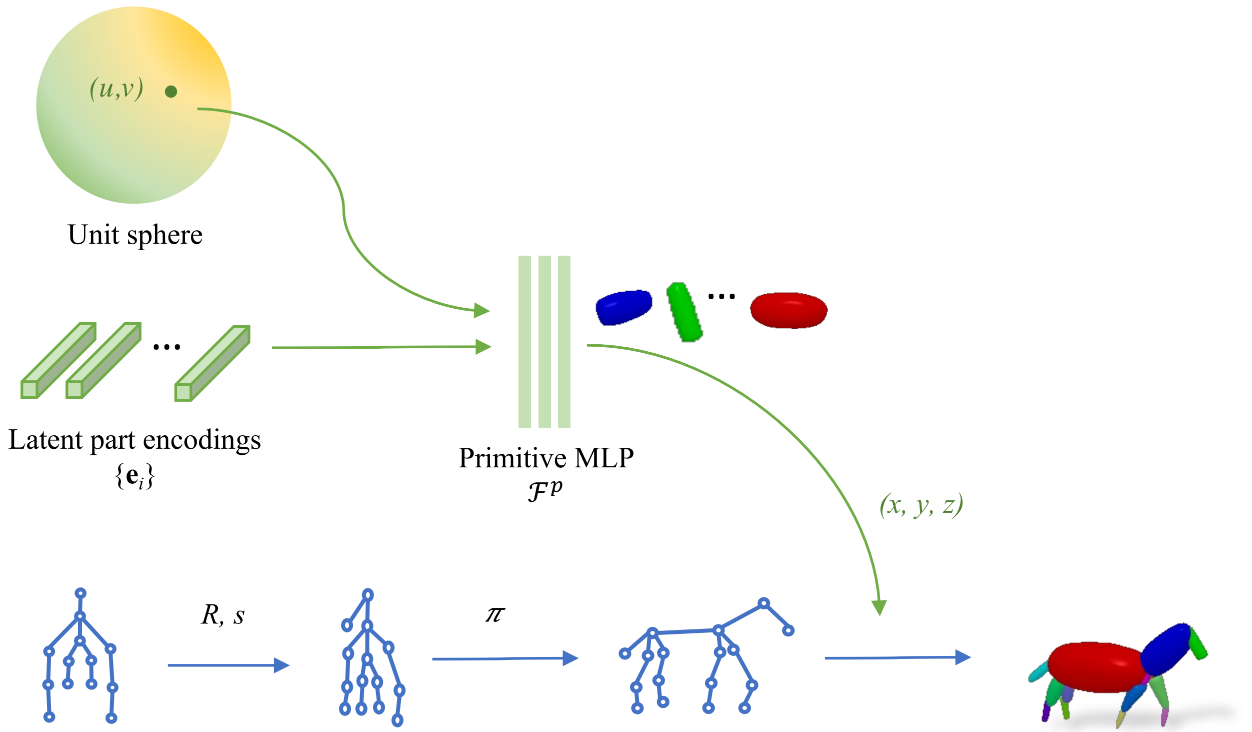
Self-supervised part discovery

- 3D shape and part annotations are hard to obtain
- Part discovery by learning to reconstruct

Representation: Skeleton-based Neural Part Surfaces



Representation: Skeleton-based Neural Part Surfaces



Unit sphere

Latent part encodings
$\{\mathbf{e}_i\}$

Primitive MLP
$\mathcal{F}^p$

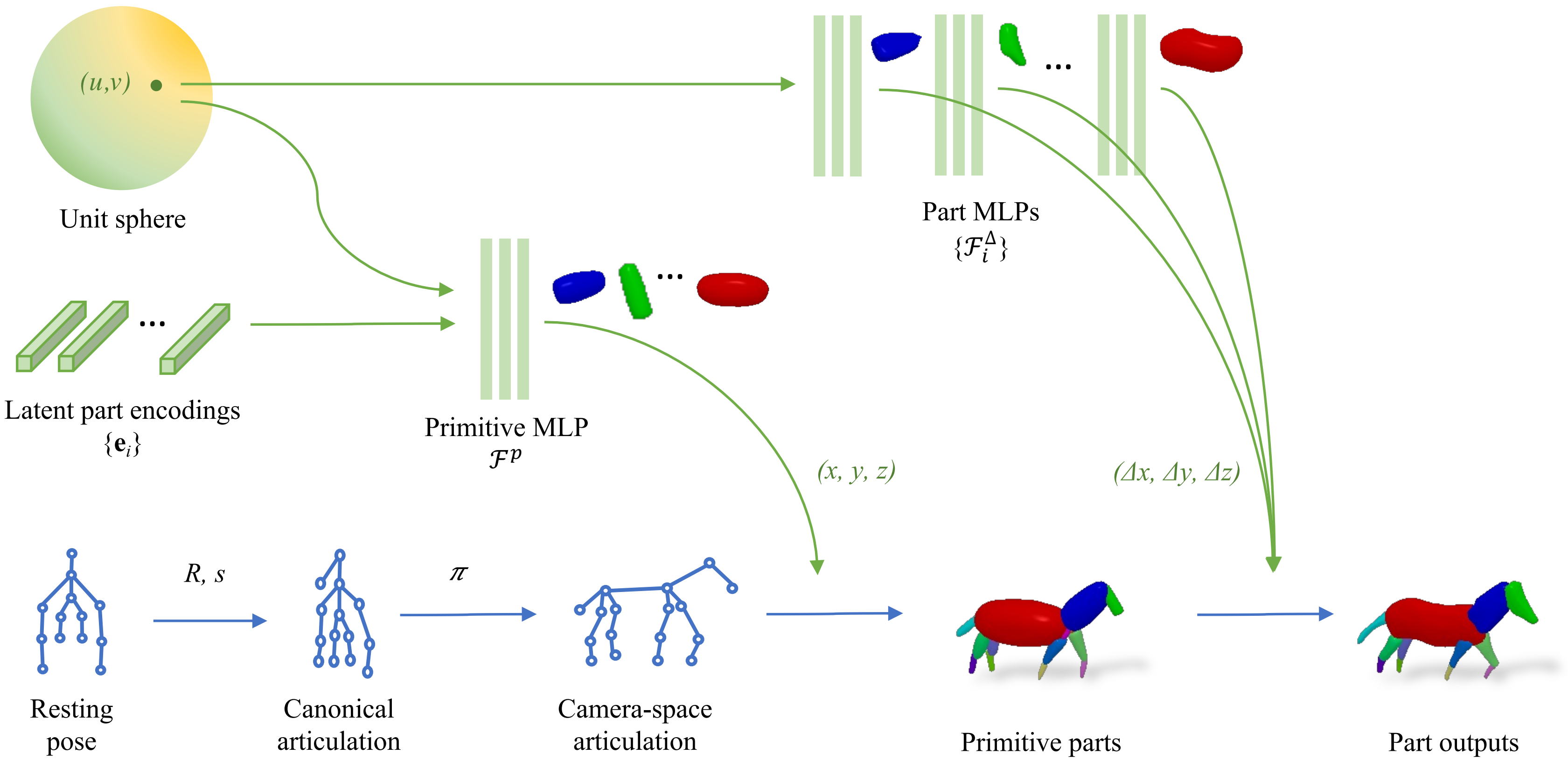
Resting pose

Canonical articulation

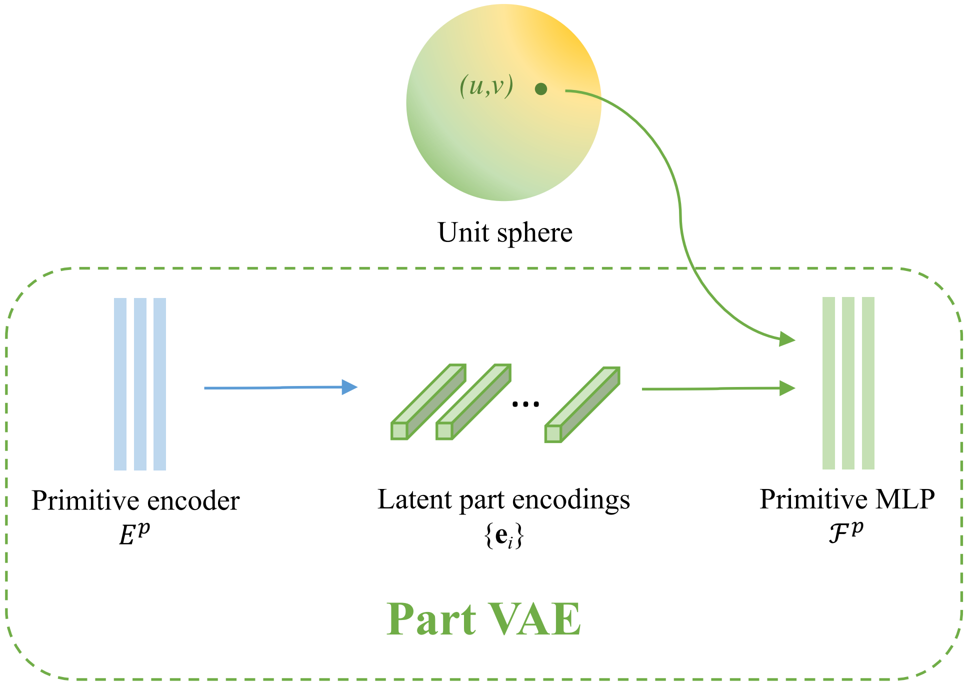
Camera-space articulation

Primitive parts

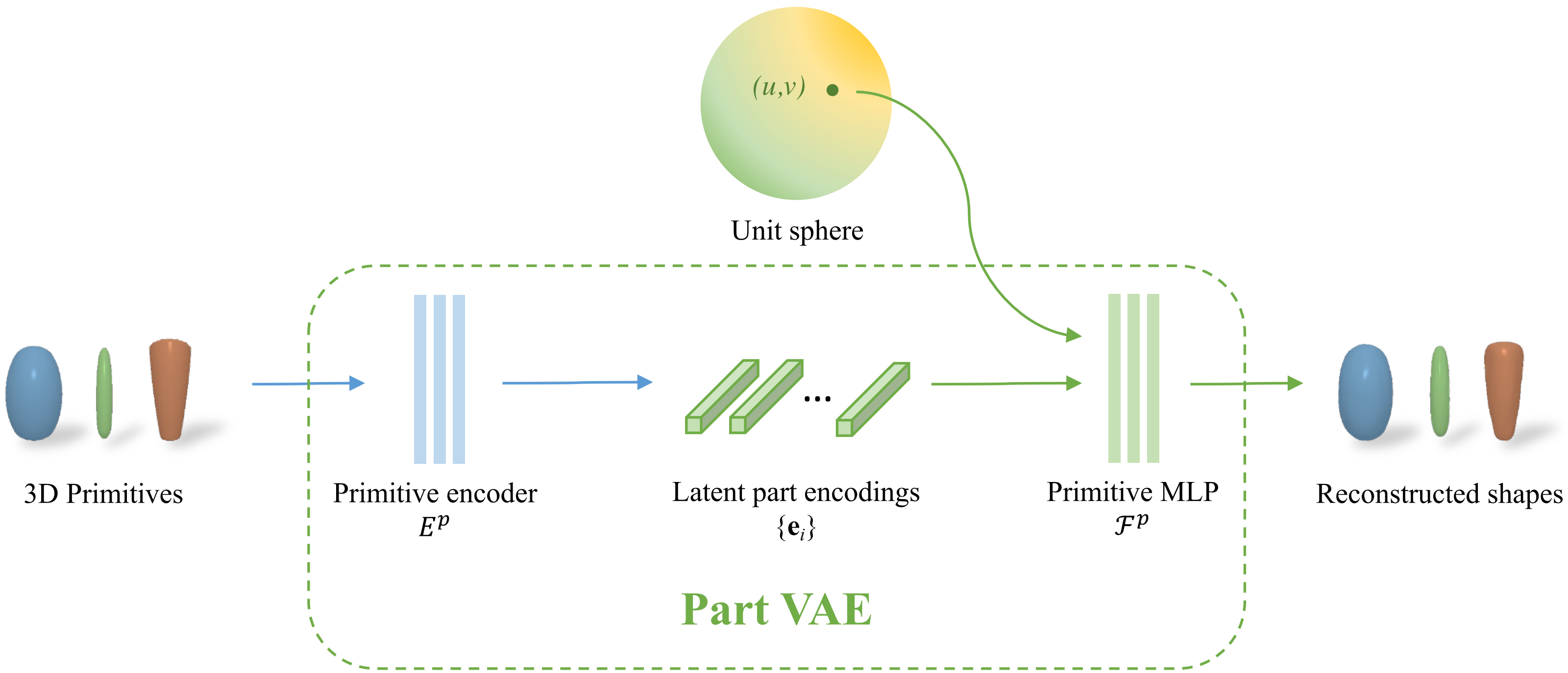
Representation: Skeleton-based Neural Part Surfaces



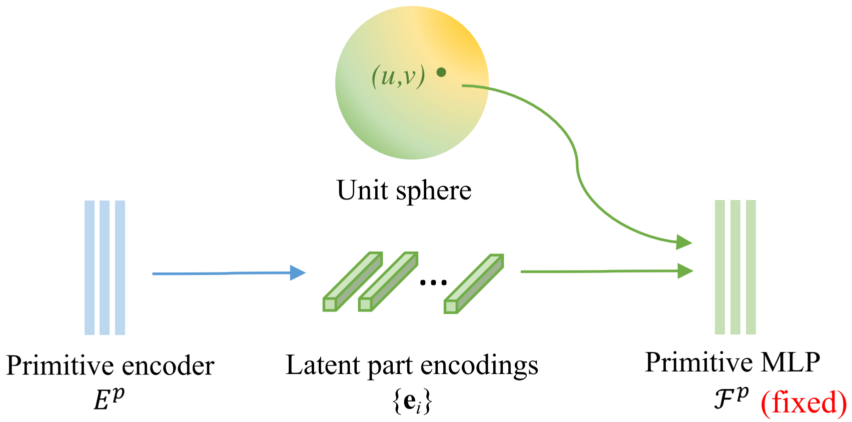
Learning Latent Part Prior with 3D Primitives



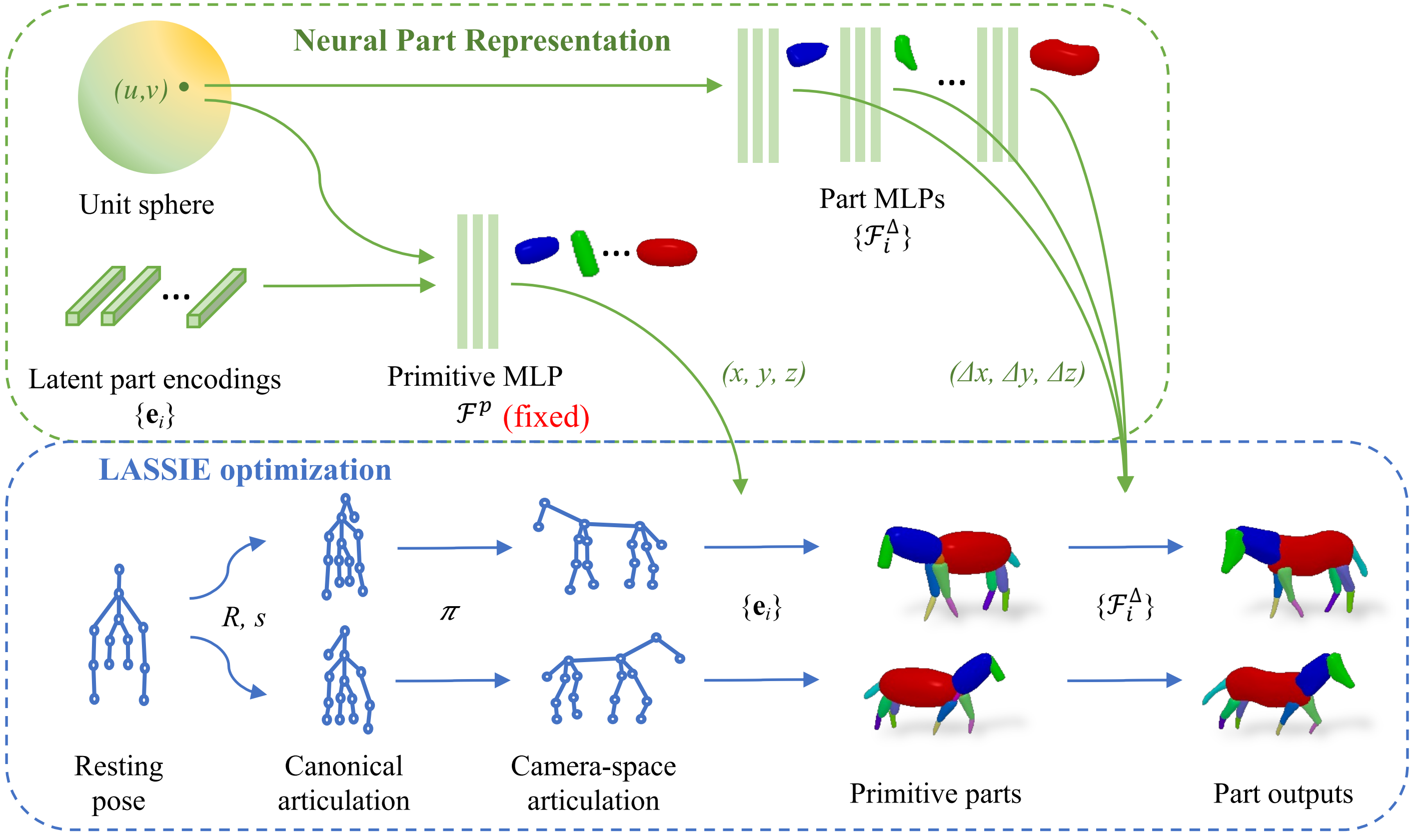
Learning Latent Part Prior with 3D Primitives



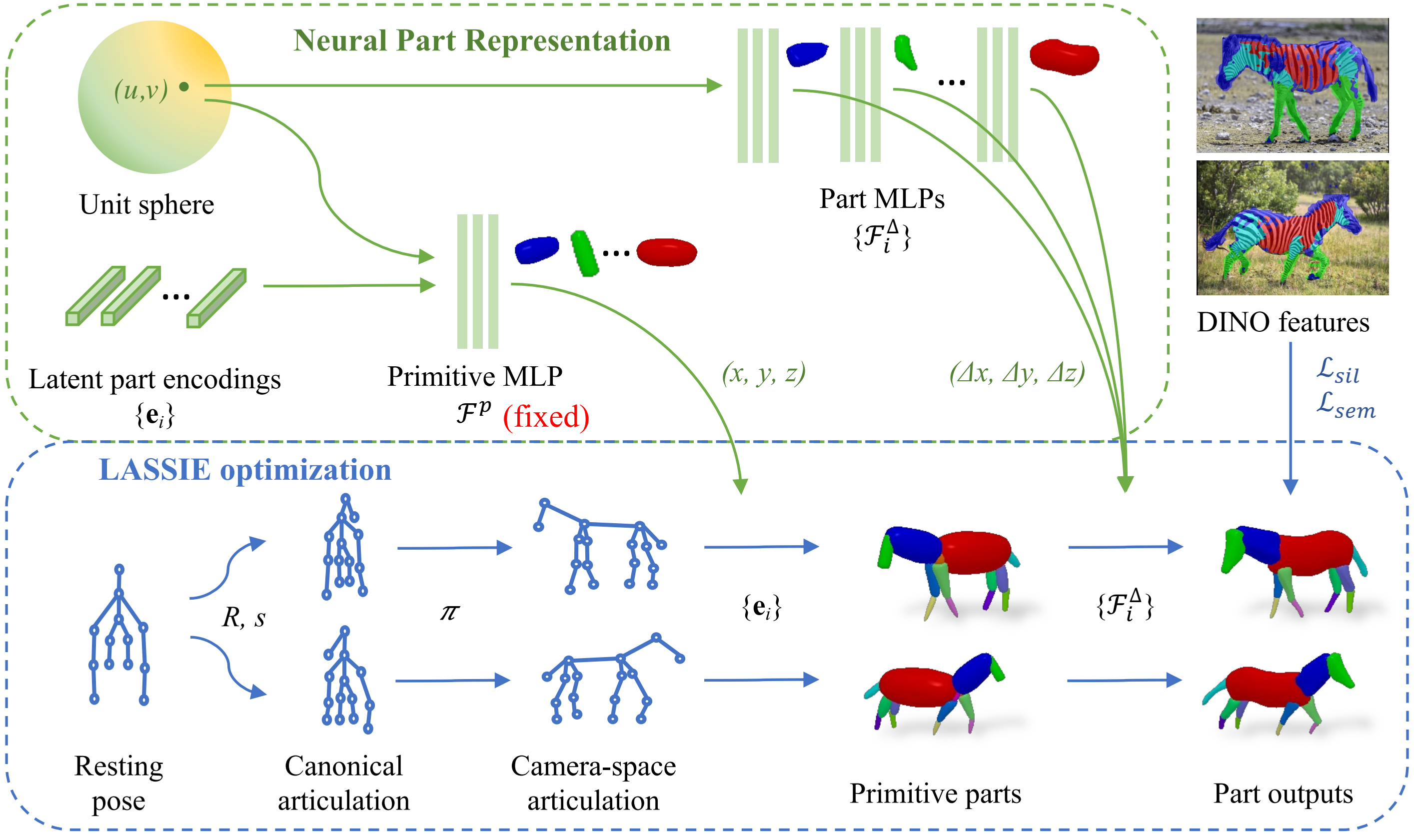
LASSIE Optimization Framework



LASSIE Optimization Framework



LASSIE Optimization Framework



Datasets

Pascal-part [1]: horse, cow, sheep

Images

Part masks

Our image ensemble: zebra, tiger, giraffe, elephant, kangaroo, penguin

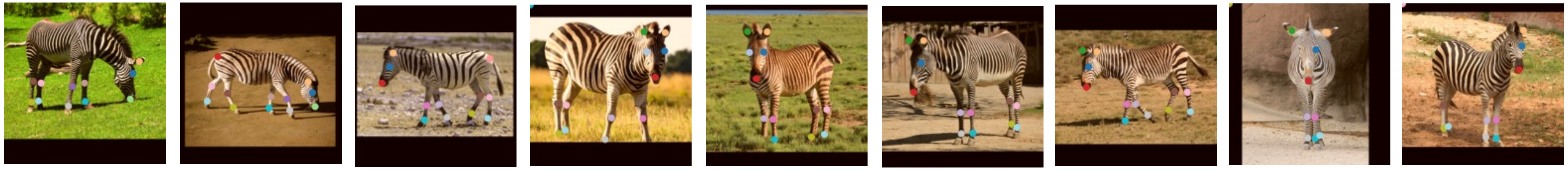
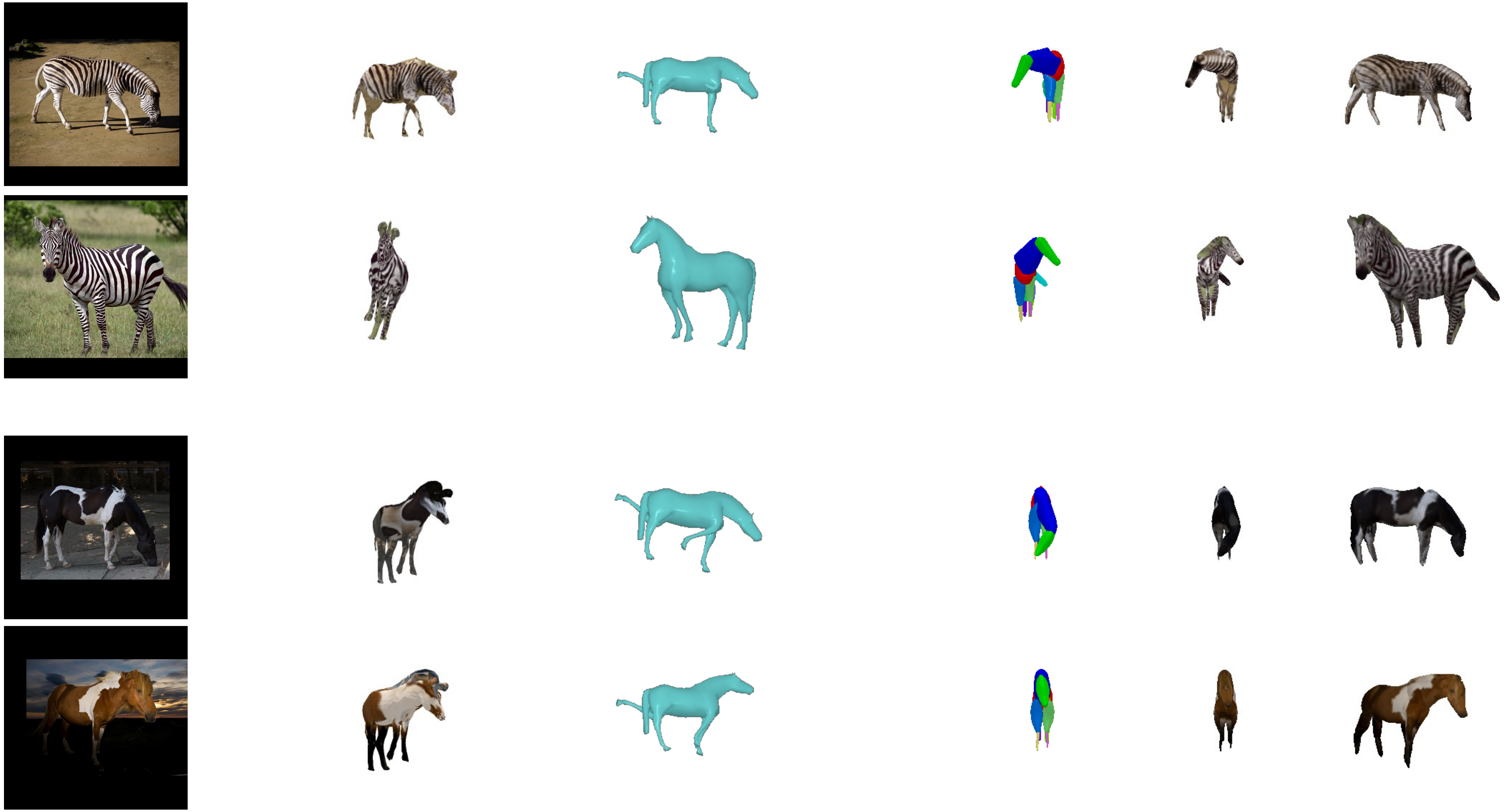


Image & keypoints

[1] Chen, Xianjie, et al. "Detect what you can: Detecting and representing objects using holistic models and body parts." *CVPR*. 2014.

Qualitative Comparisons – 3D reconstruction



Qualitative Comparisons – 2D (part) segmentation

Input

DINO clustering

3D Safari

A-CSM

LASSIE

Applications - animation

Applications – pose/motion transfer

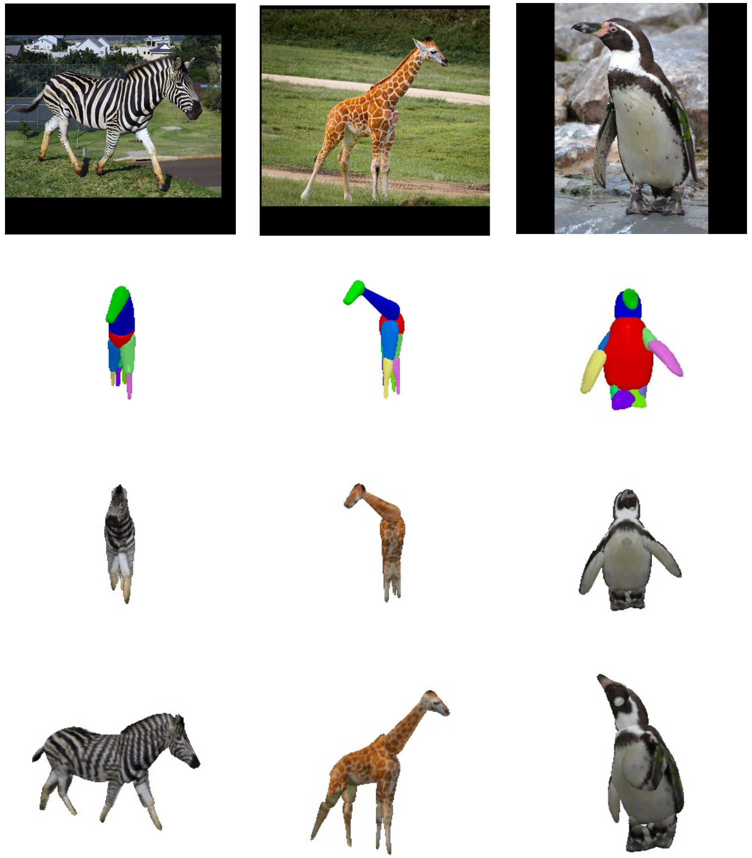
Source

Target
pose/motion

Result

LASSIE: Learning Articulated Shape from Sparse Image Ensemble via 3D Part Discovery

First approach for articulated shape reconstruction from sparse image ensemble in-the-wild



Key advantages

- In-the-wild images
- Self-supervised
- SOTA reconstruction accuracy
- Semantically consistent part discovery

Main technical contributions

- Skeleton-based neural part surfaces
- Latent part prior learning
- Semantic consistency loss based on self-supervised ViT features