JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Hi-LASSIE: High-Fidelity Articulated Shape and Skeleton Discovery from Sparse Image Ensemble

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 [4]Yonsei University

Poster: TUE-PM-068

Estimating articulated animal bodies from sparse images

Image ensemble




3D articulated shapes
(per-instance)

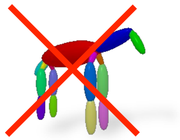



Estimating articulated animal bodies from sparse images

Problem setting

- **Inputs:** 20-30 in-the-wild images of an animal class
- **Outputs:** per-instance camera viewpoint, pose, and shape
- **Without** pre-defined shape/skeleton templates

Shape template



Skeleton template

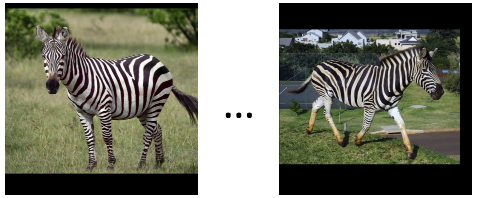
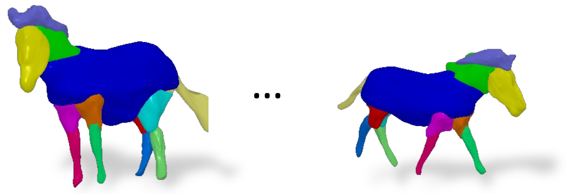


Image ensemble

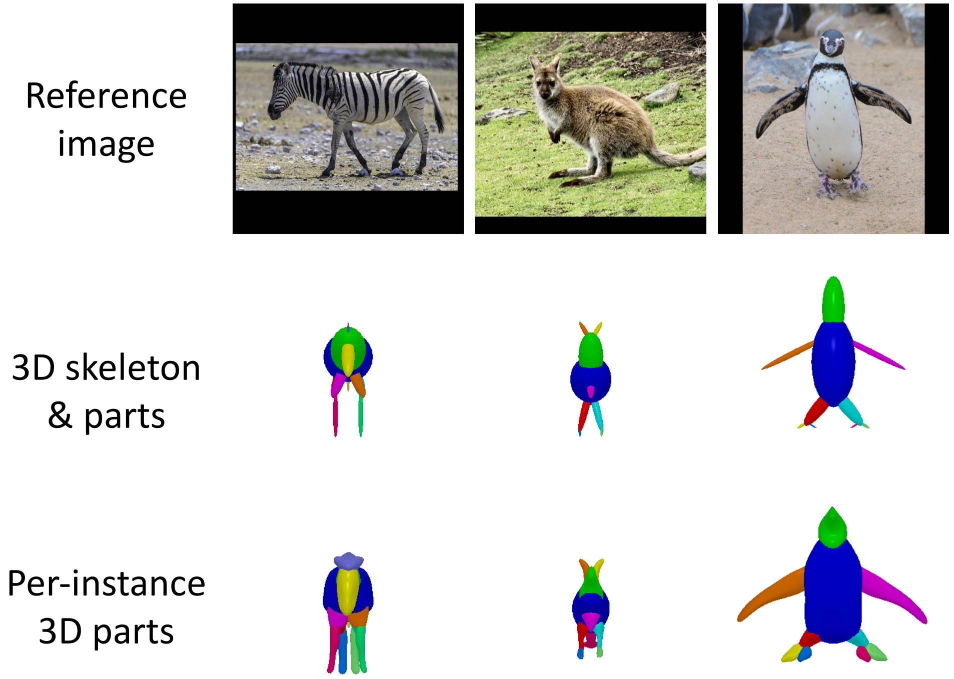


Per-Instance camera, pose, shape

Hi-LASSIE: High-Fidelity Articulated Shape and Skeleton Discovery from Sparse Image Ensemble

Discovering 3D skeleton and parts

- LASSIE [1]
  - human-annotated skeleton
  - shared part shapes
- Hi-LASSIE
  - skeleton discovery
  - per-instance part shapes



Key advantages

- In-the-wild images
- SOTA reconstruction accuracy
- Minimal user input (select reference image)

[1] Yao, Chun-Han, et al. " Lassie: Learning articulated shapes from sparse image ensemble via 3d part discovery." *NeurIPS*. 2022.

Estimating articulated animal bodies from sparse images

Problem setting

- **Inputs:** 20-30 in-the-wild images of an animal class
- **Outputs:** per-instance camera viewpoint, pose, and shape
- **Without** pre-defined shape/skeleton templates

Challenges:

- **In-the-wild images:** diverse background, lighting, camera, pose, and texture
- **No** image-level annotations (camera, keypoints, silhouettes)

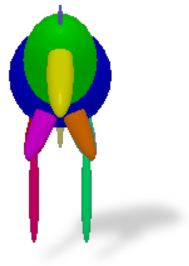
(Pascal-Part [1] horse images)

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

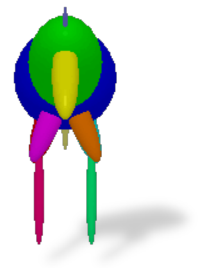
Discovering 3D skeleton and parts

3D part surfaces upon skeleton

- 3D parts: simple geometry, rigid motion, semantic consistency
- 3D skeleton: constrain articulation & part connectivity



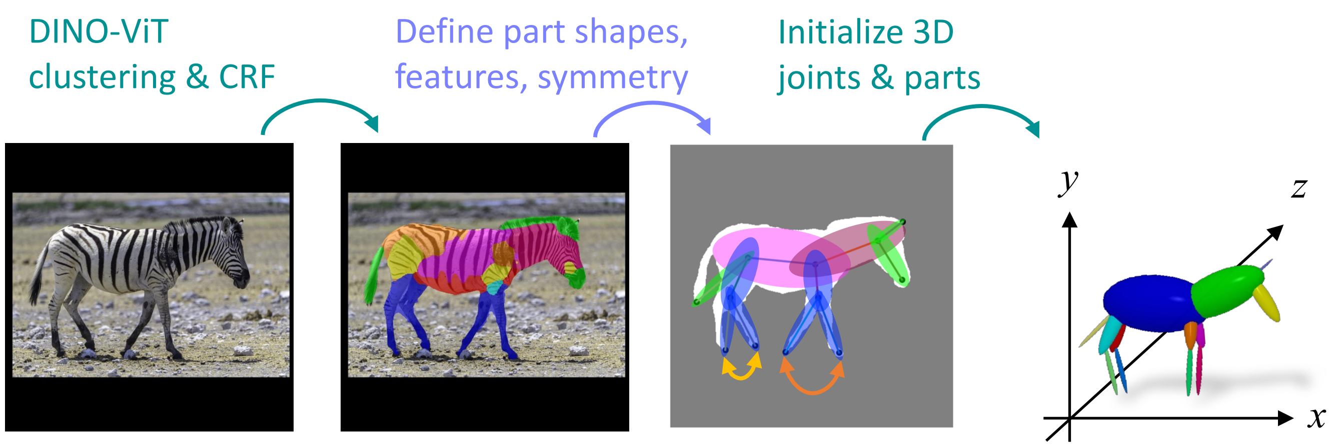
Discovering 3D skeleton and parts



3D part surfaces upon skeleton

- 3D parts: simple geometry, rigid motion, semantic consistency
- 3D skeleton: constrain articulation & part connectivity

3D skeleton discovery from reference image



Reference image

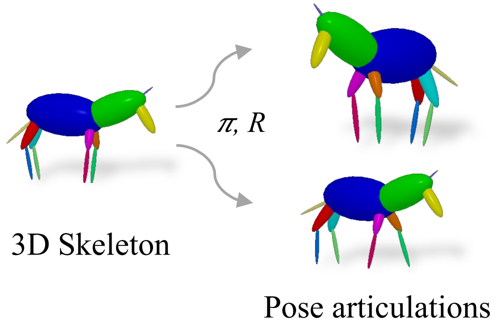
DINO [1] features & silhouette

Part features, shape, symmetry

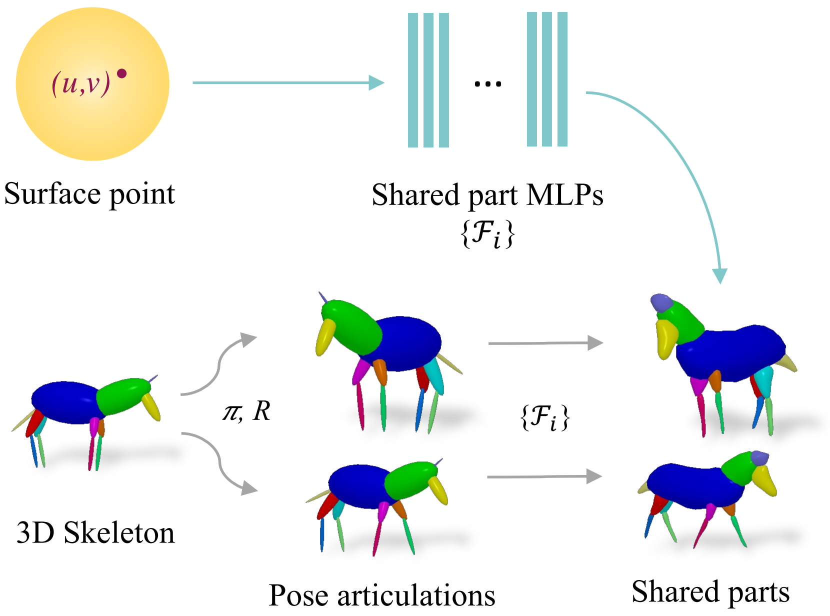
3D skeleton & parts

[1] Caron, Mathilde, et al. "Emerging properties in self-supervised vision transformers." *ICCV*. 2021.

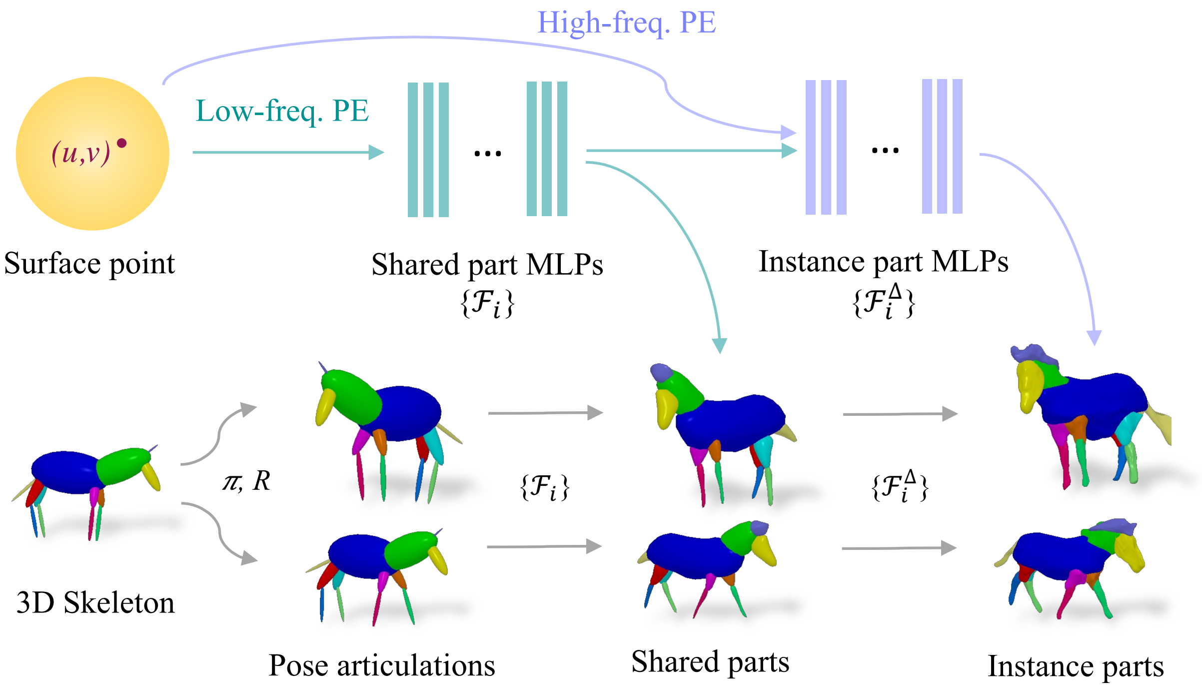
Hi-LASSIE optimization framework

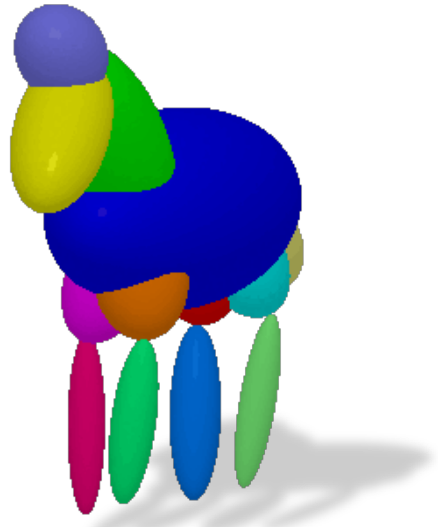


Hi-LASSIE optimization framework



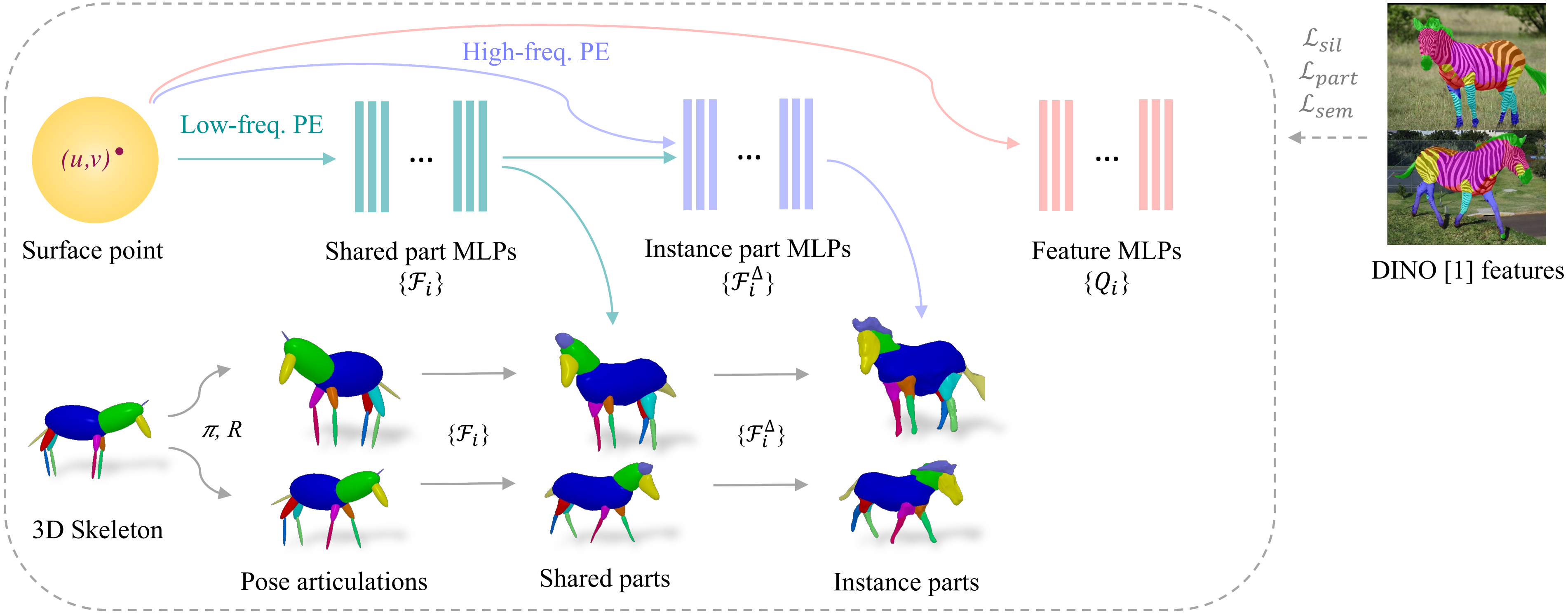
Hi-LASSIE optimization framework





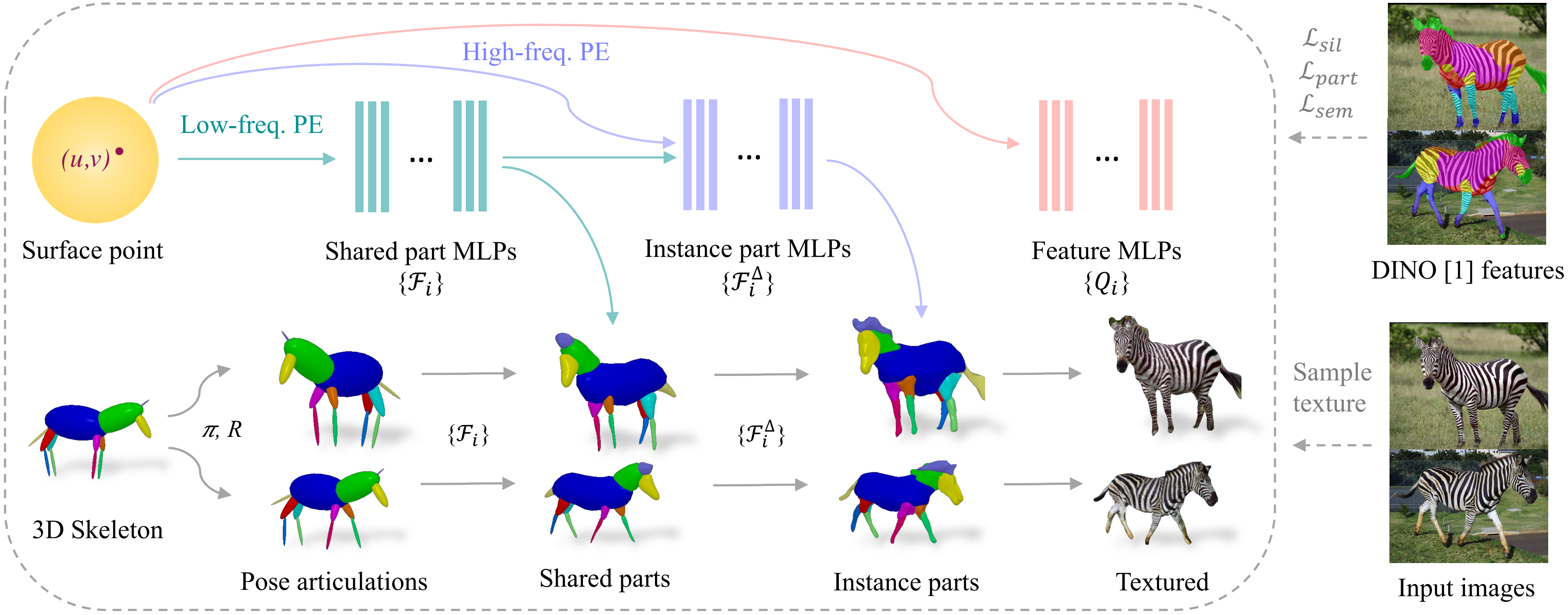
Frequency decomposition
(low-to-high frequency)

Hi-LASSIE optimization framework



[1] Caron, Mathilde, et al. "Emerging properties in self-supervised vision transformers." *ICCV*. 2021.

Hi-LASSIE optimization framework



[1] Caron, Mathilde, et al. "Emerging properties in self-supervised vision transformers." *ICCV*. 2021.

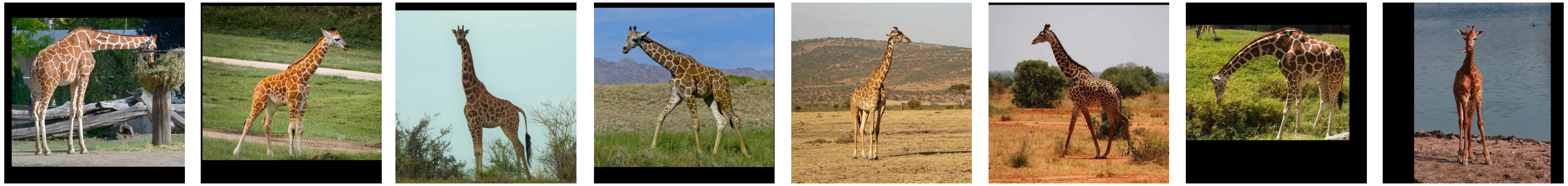
Datasets

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

Horse images

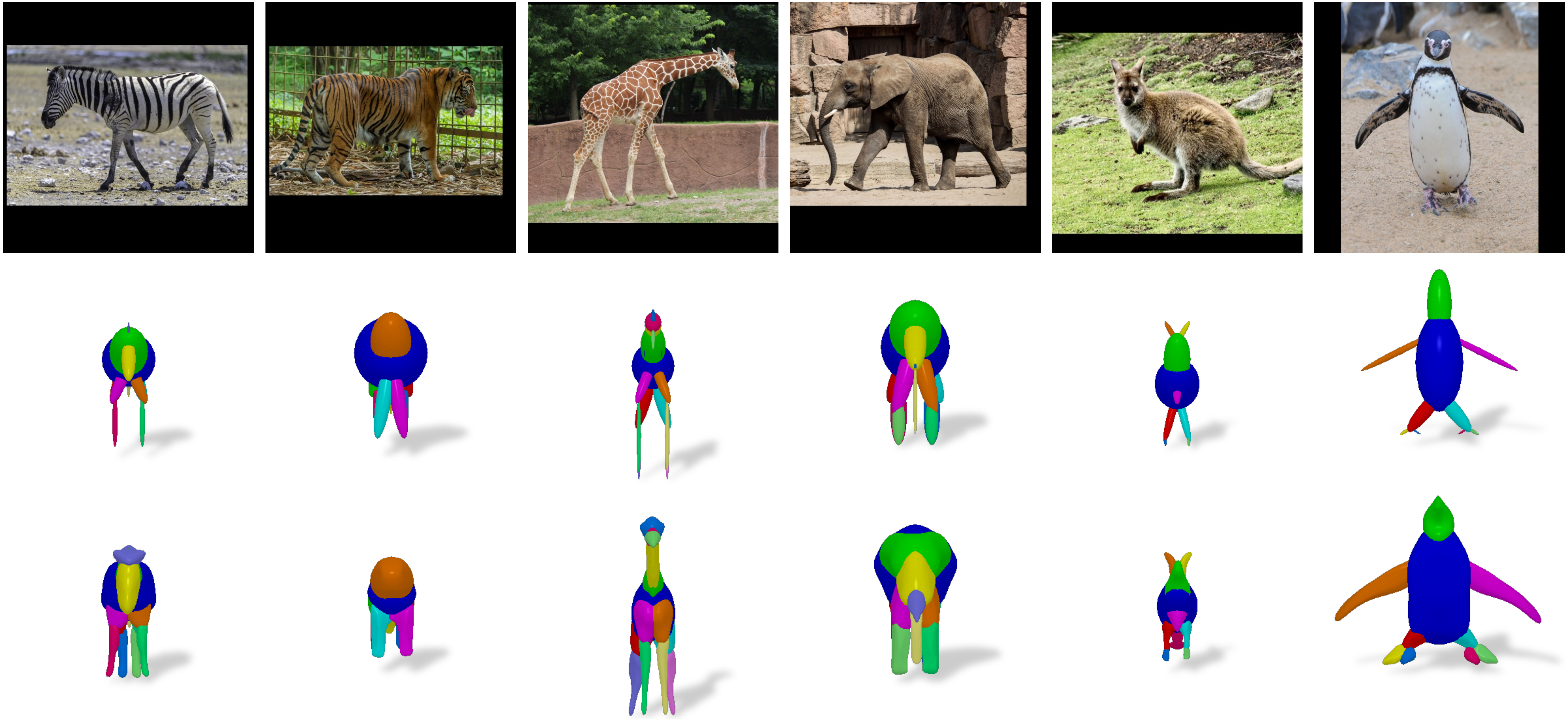
LASSIE [2]: zebra, tiger, giraffe, elephant, kangaroo, penguin

Giraffe images



[1] Chen, Xianjie, et al. "Detect what you can: Detecting and representing objects using holistic models and body parts." *CVPR*. 2014.
[2] Yao, Chun-Han, et al. " Lassie: Learning articulated shapes from sparse image ensemble via 3d part discovery." *NeurIPS*. 2022.

Class-level results: reference image, 3D skeleton, shared parts



Per-instance results: LASSIE v.s. Hi-LASSIE

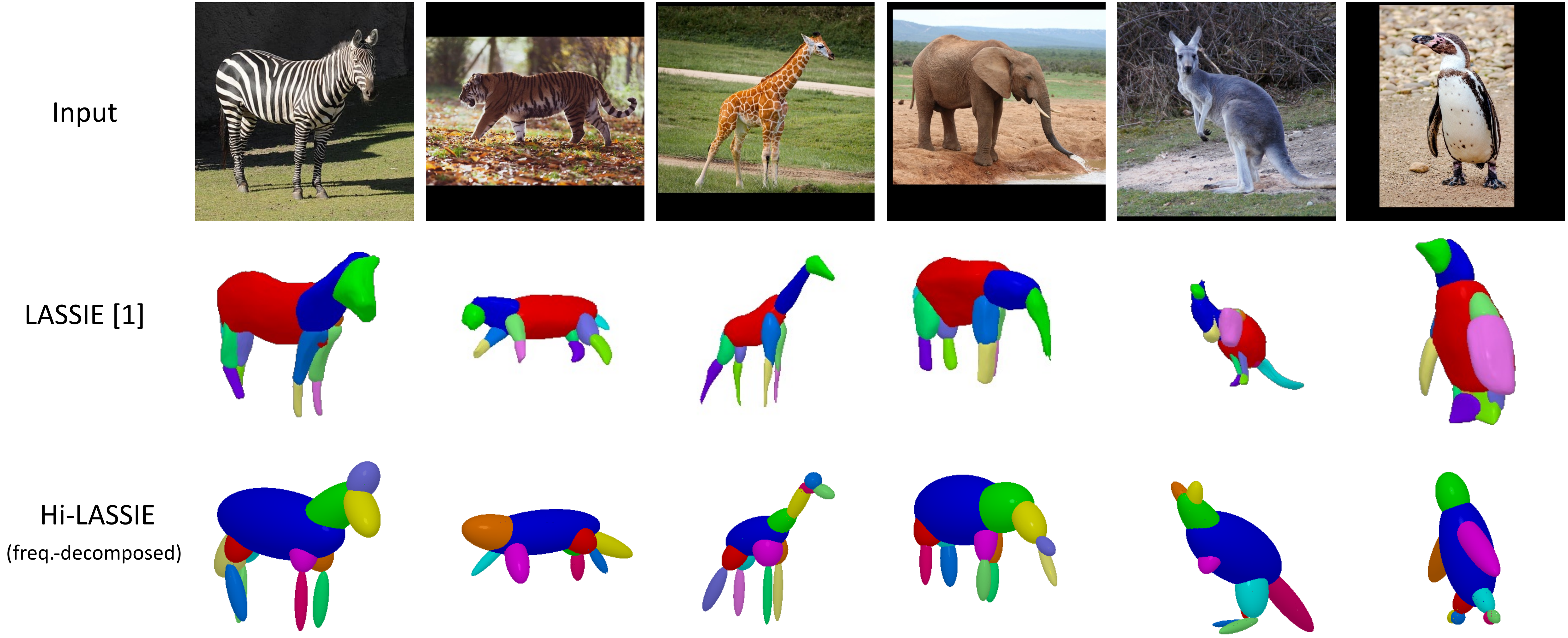
Input

LASSIE [1]

Hi-LASSIE

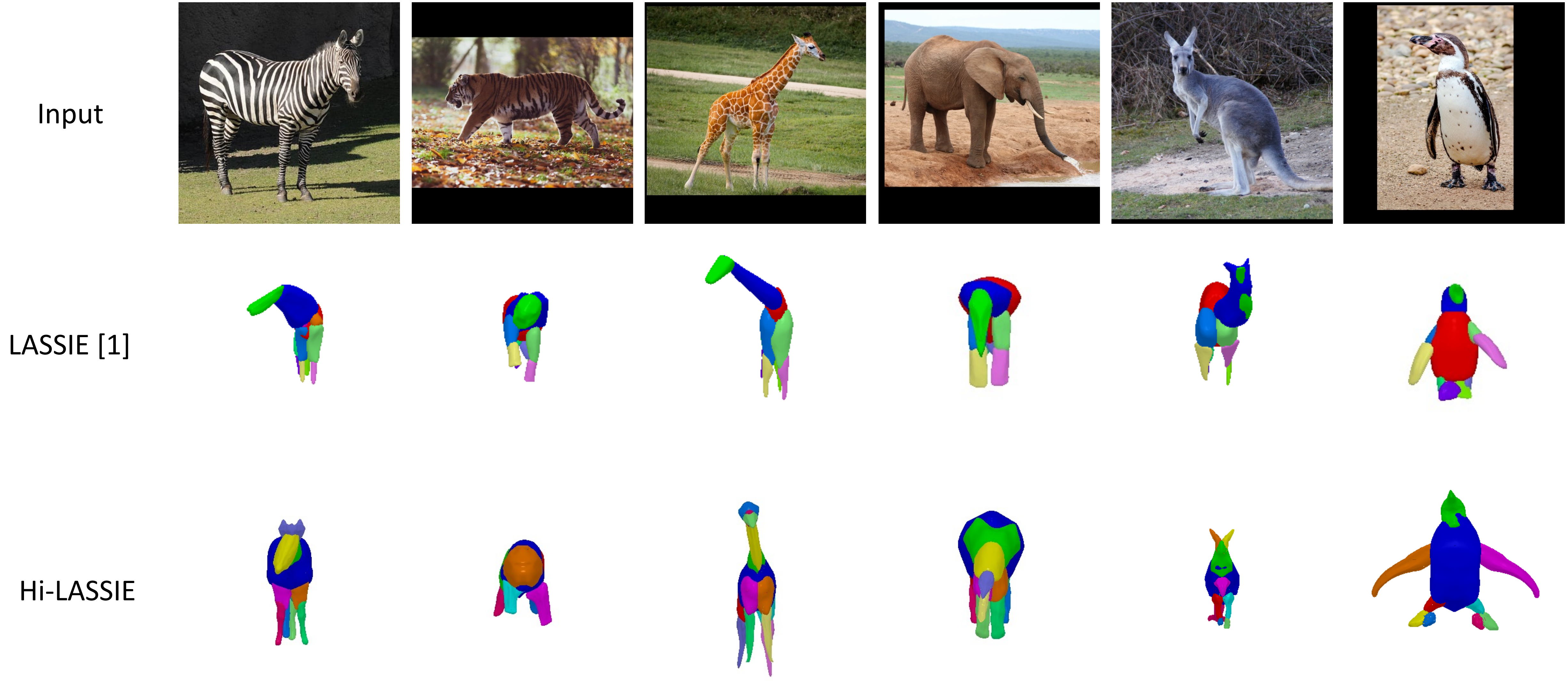
[1] Yao, Chun-Han, et al. " Lassie: Learning articulated shapes from sparse image ensemble via 3d part discovery." *NeurIPS*. 2022.

Per-instance results (3D parts): LASSIE v.s. Hi-LASSIE



[1] Yao, Chun-Han, et al. " Lassie: Learning articulated shapes from sparse image ensemble via 3d part discovery." *NeurIPS*. 2022.

Per-instance results (3D parts): LASSIE v.s. Hi-LASSIE



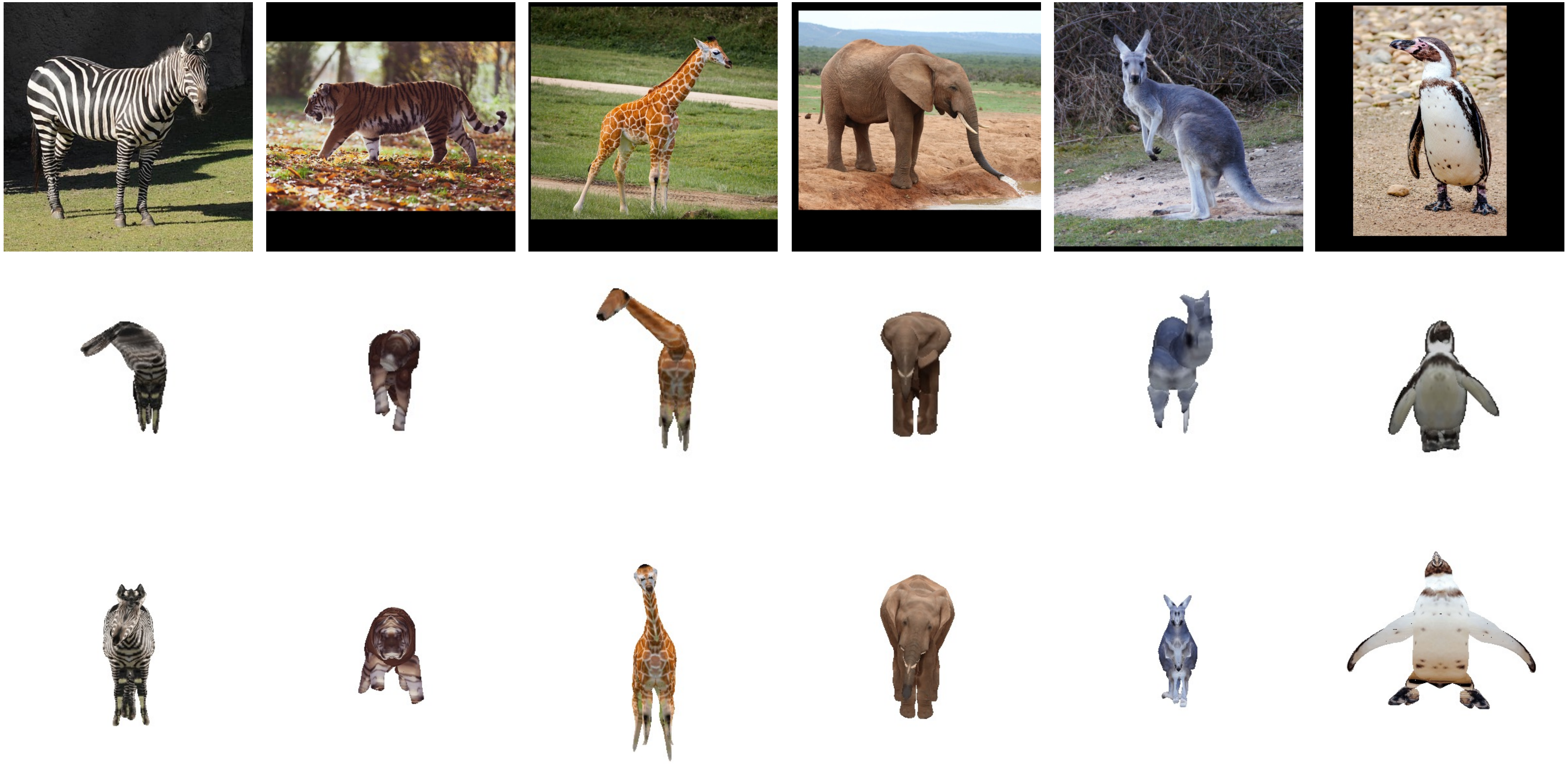
[1] Yao, Chun-Han, et al. " Lassie: Learning articulated shapes from sparse image ensemble via 3d part discovery." *NeurIPS*. 2022.

Per-instance results (texture): LASSIE v.s. Hi-LASSIE

Input

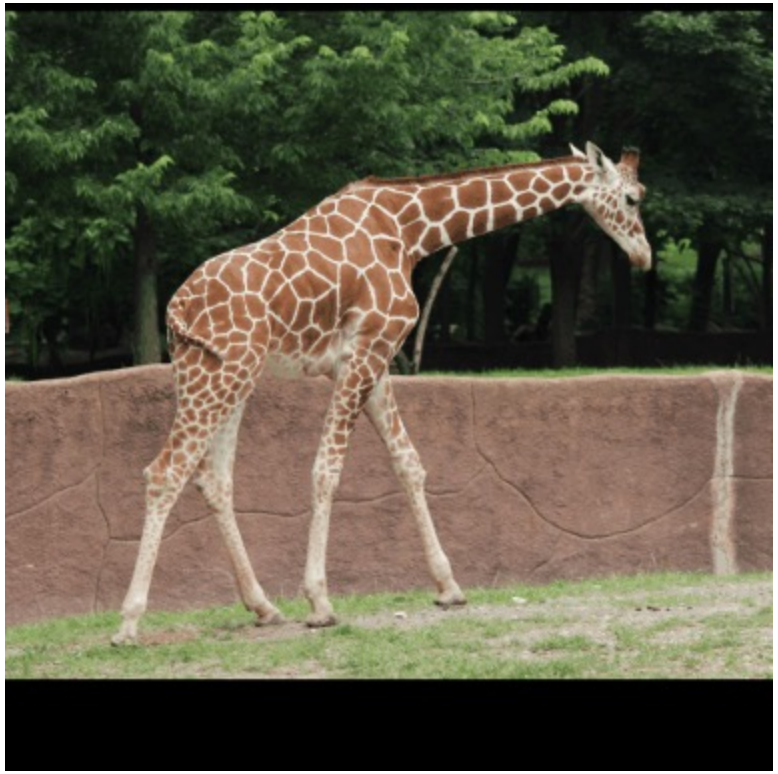
LASSIE [1]

Hi-LASSIE



[1] Yao, Chun-Han, et al. " Lassie: Learning articulated shapes from sparse image ensemble via 3d part discovery." *NeurIPS*. 2022.

Application: animation via pose interpolation

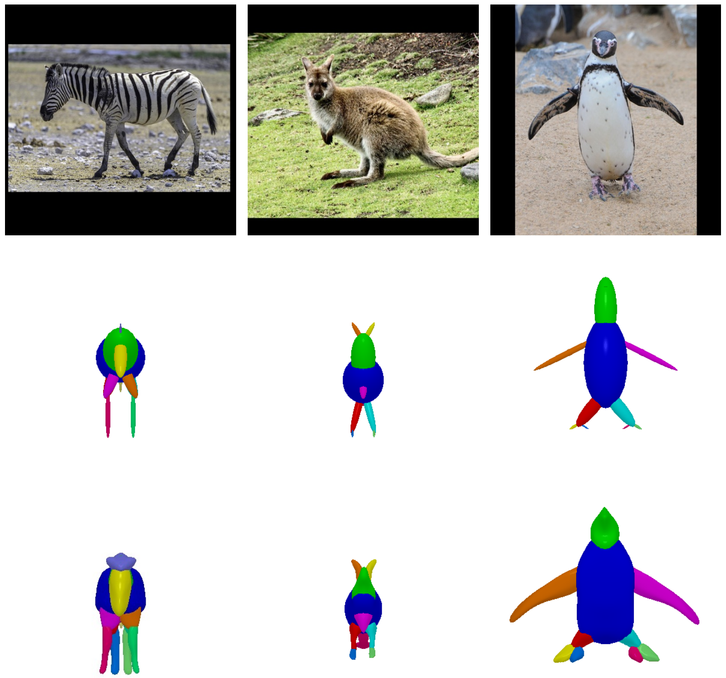


Hi-LASSIE: High-Fidelity Articulated Shape and Skeleton Discovery from Sparse Image Ensemble

First approach to reconstruct articulated shapes from sparse images in-the-wild without 3D shape/skeleton templates or per-image annotations

Key advantages

- In-the-wild images
- Minimal user input (select reference image)
- SOTA reconstruction accuracy



Main technical contributions

- 3D skeleton discovery
- Frequency-decomposed neural surfaces
- Zoomed-in part rendering and optimization
- Semantic feature MLPs