3D Safari: Learning to Estimate Zebra Pose, Shape, and Texture from Images "In the Wild"

Silvia Zuffi, Angjoo Kanazawa,
Tanya Berger-Wolf, Michael J. Black

The Grevy's zebra

RED LIST

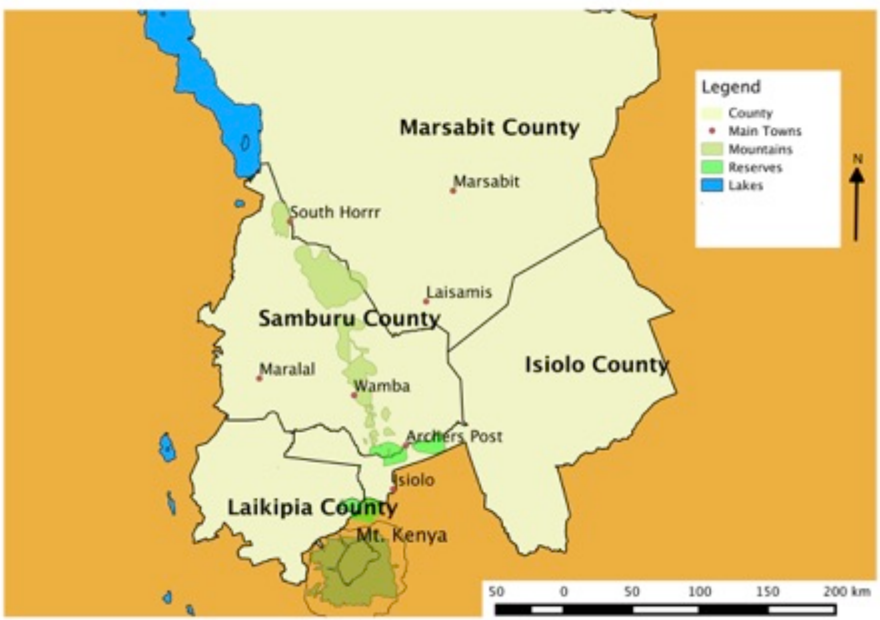
ENDANGERED

EN

The Grevy's zebra

https://zebra.wildbook.org/

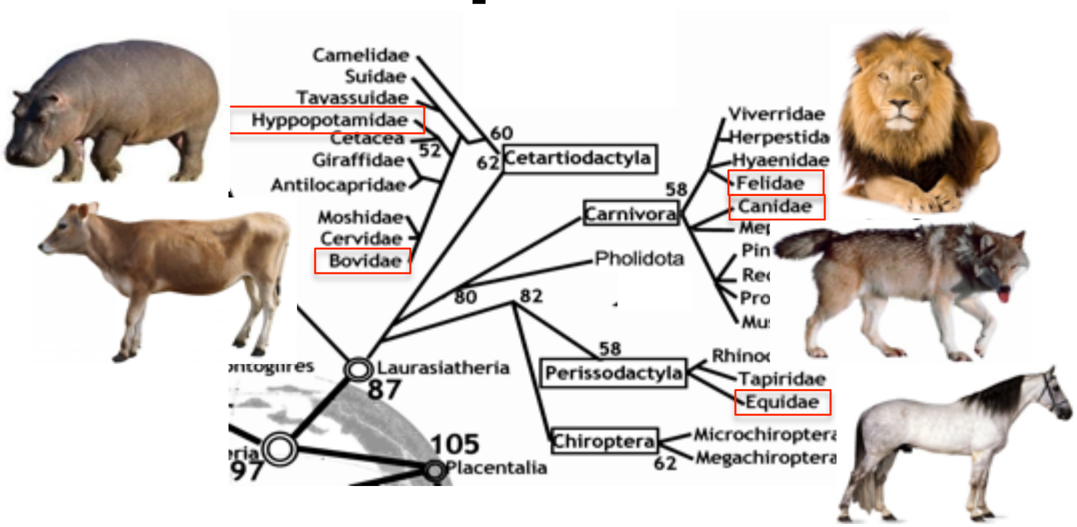
First census of the Grevy's zebra with photographs of ordinary citizens



THE GREAT GREVY'S RALLY

SMAL

- Skinned Multi-Animal Linear model
- A 3D shape model representing **articulation** and **shape variation** across different species



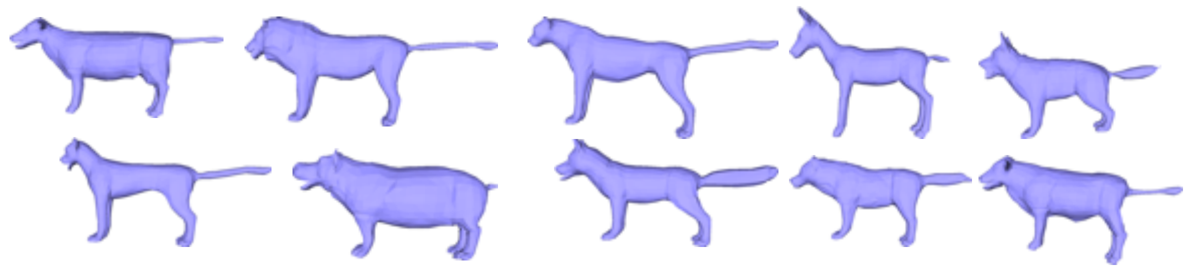
Examples from the training set

- **From 3D data**, fast to compute

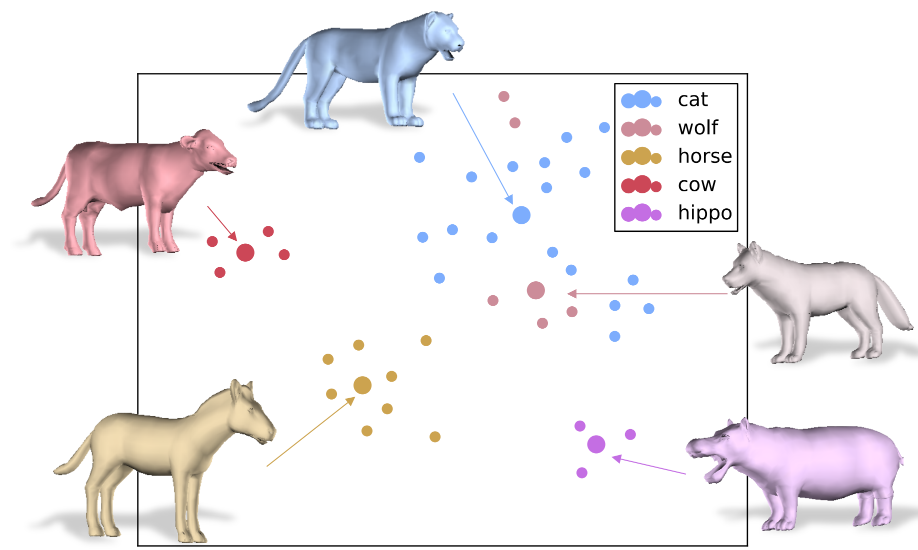
S. Zuffi, A. Kanazawa, D. Jacobs, M. J. Black, 3D Menagerie: Modeling the 3D Shape and Pose of Animals, CVPR 2017

SMAL shape space

Training set: Toys scans in correspondence and in reference pose

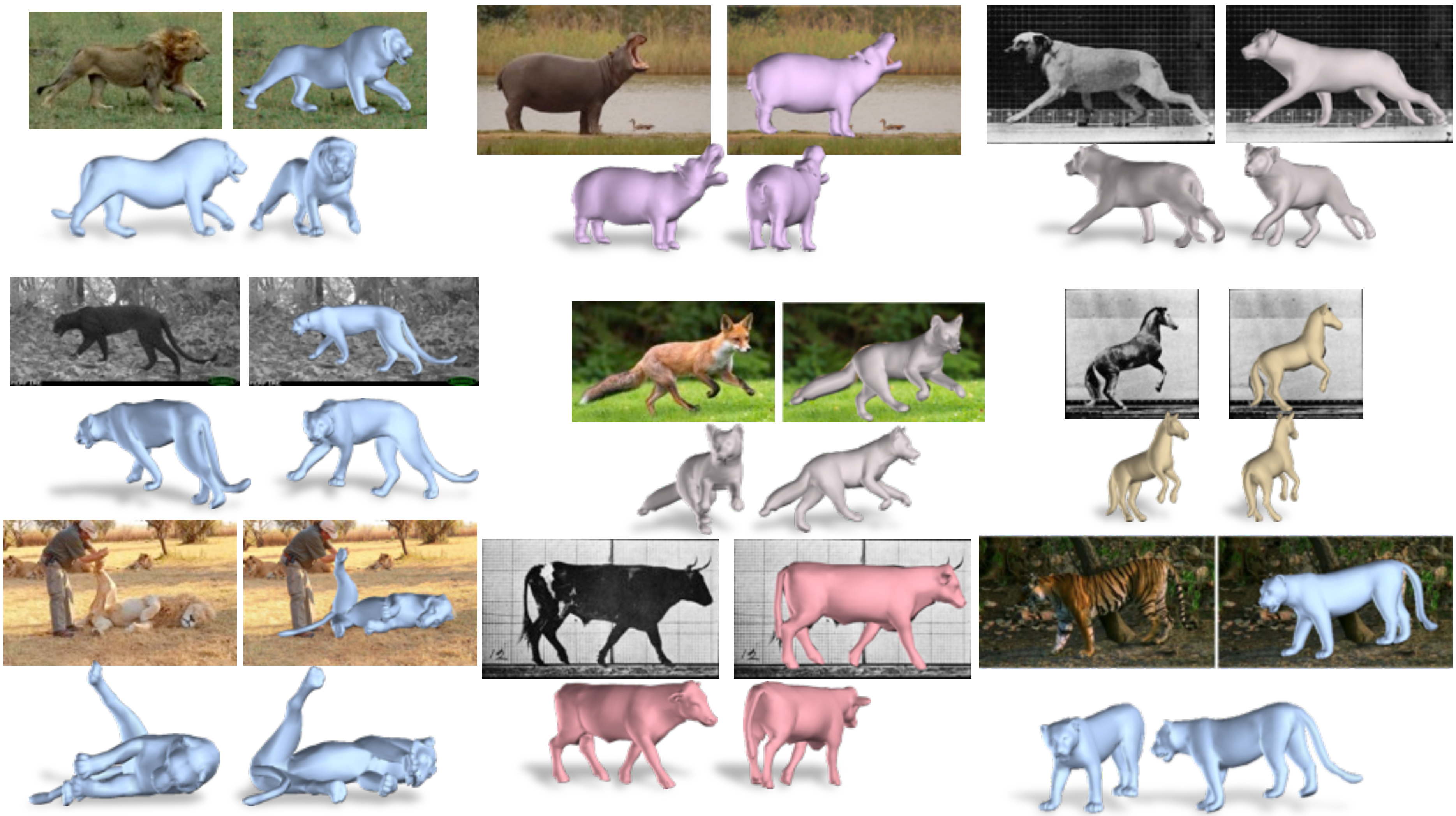


$$\mathbf{v}_{shape}(\beta) = \mathbf{v}_{template} + B_s\beta$$



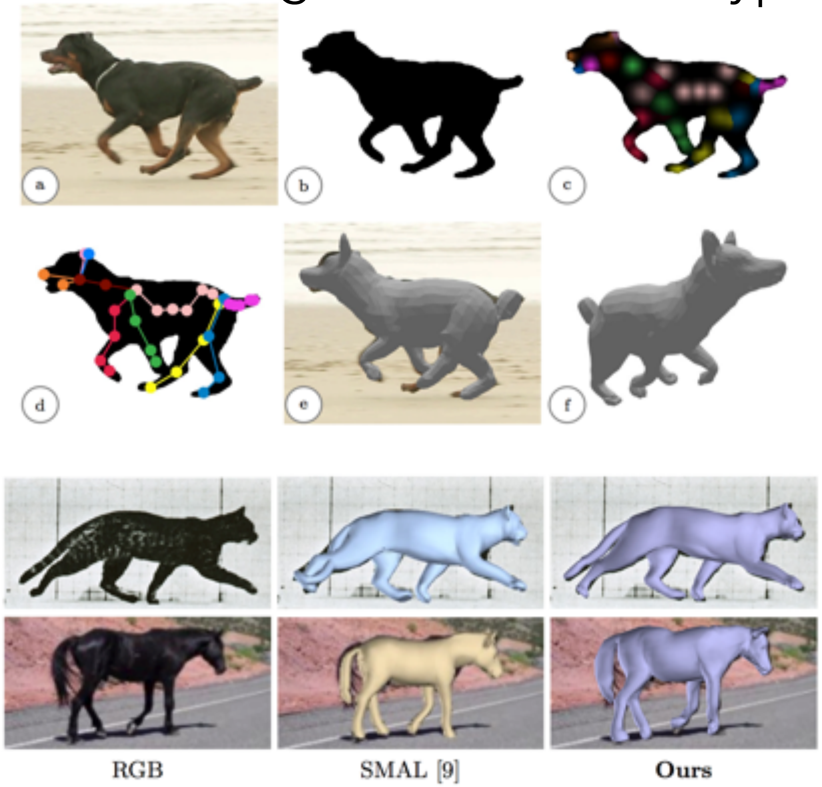
Applications of SMAL

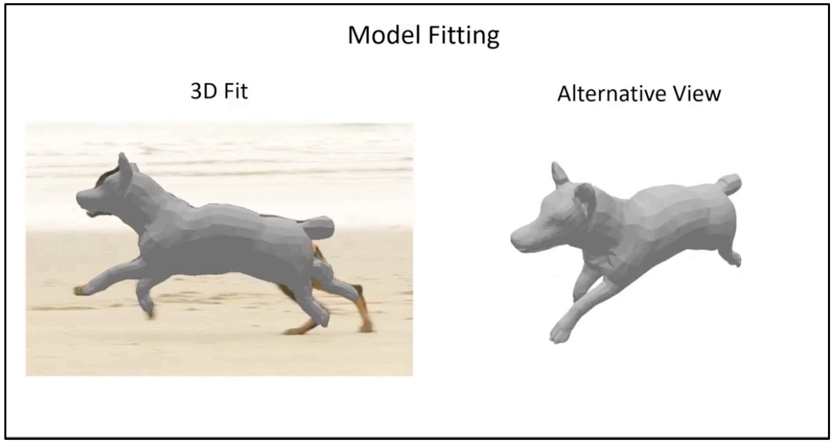
Manual segmentation and manually annotated keypoints



Applications of SMAL

Automatic segmentation and keypoints detection from silhouette





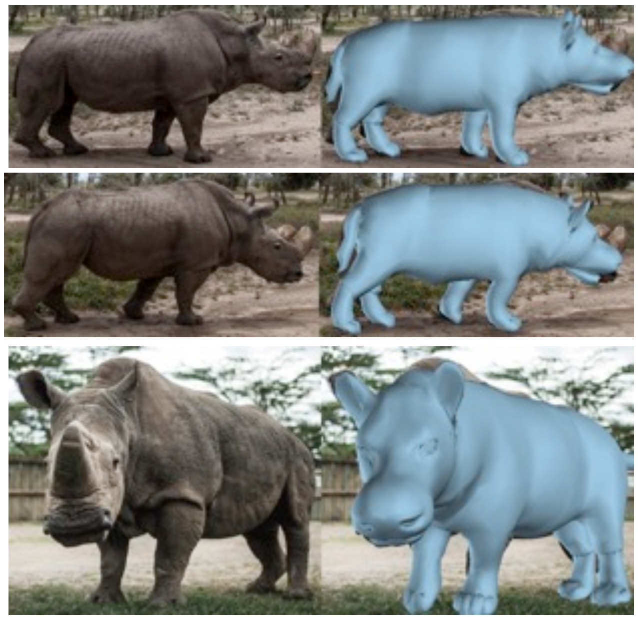
B. Biggs, T. Roddick, A. Fitzgibbon, R. Cipolla, Creatures great and SMAL: Recovering the shape and motion of animals from video, ACCV2019

Our work

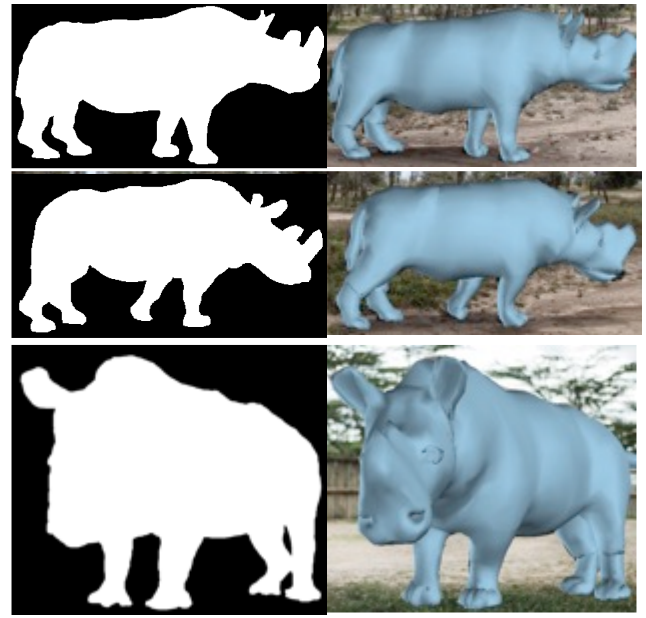
- **GOAL**: Estimate 3D shape and pose as a direct regression from RGB
- **APPROACH**: Supervised, training based only on synthetic data

SMAL with Refinement (SMALR)

1. SMAL model fitting



2. Model-free shape Refinement



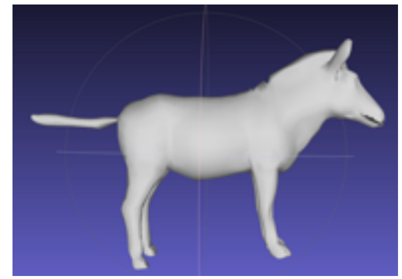
S. Zuffi, A. Kanazawa, M. J. Black, Lions and Tigers and Bears: Capturing Non-Rigid, 3D, Articulated Shape from Images, CVPR2018

Animals avatars with SMALR

S. Zuffi, A. Kanazawa, M. J. Black, Lions and Tigers and Bears:
Capturing Non-Rigid, 3D, Articulated Shape from Images, CVPR2018

Grevy's zebra avatars

Multiple images of the same subject

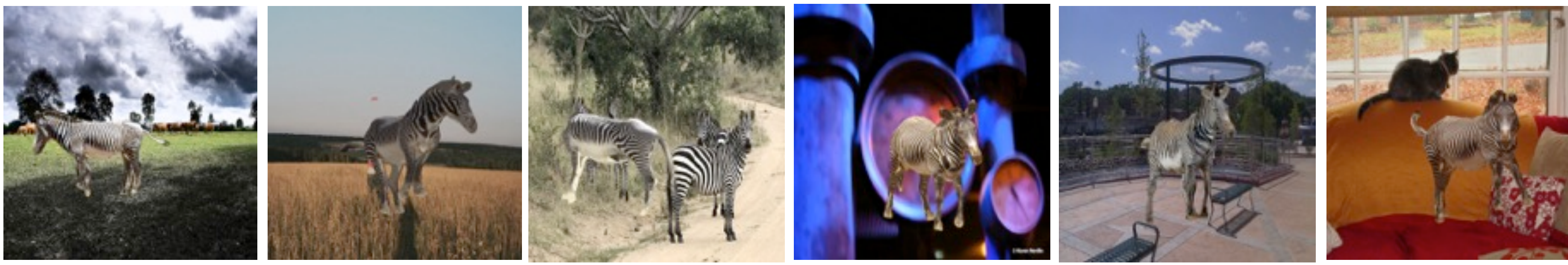


3D model

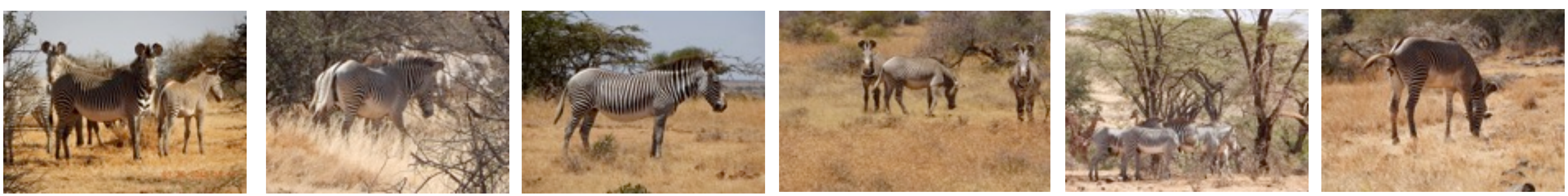
Texture map

Synthetic dataset from avatars

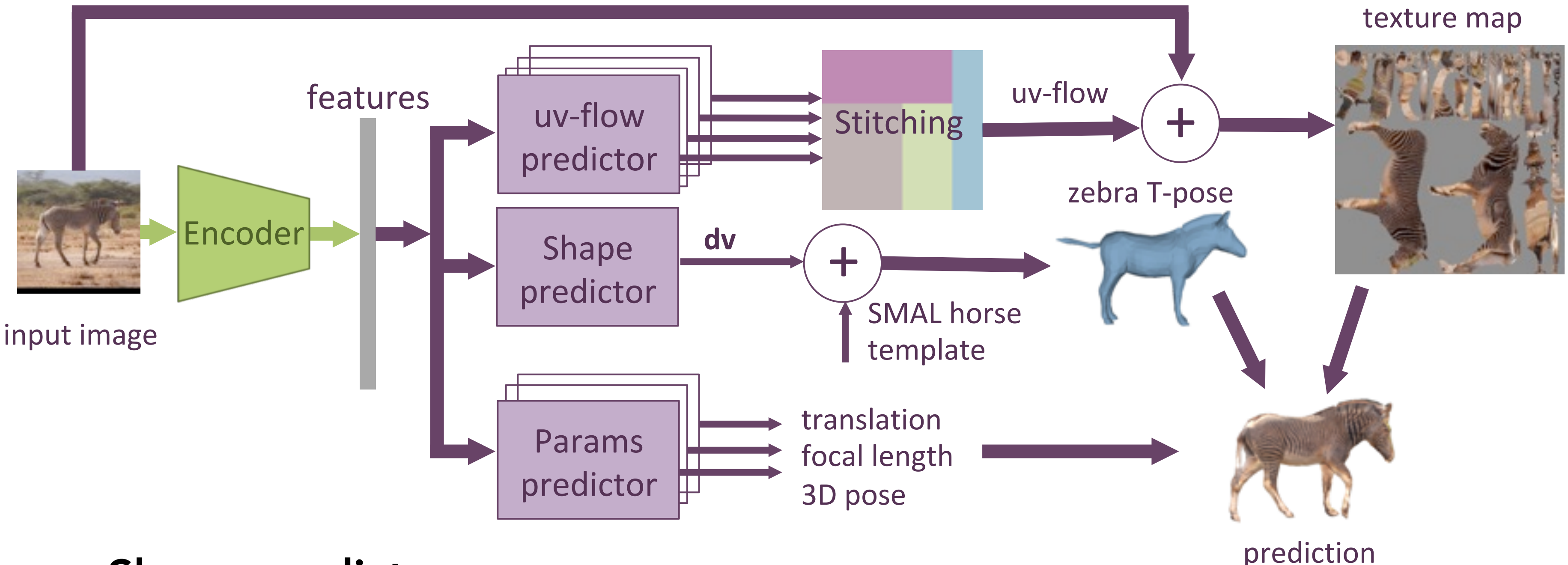
Synthetic



Real



Network



Shape predictor:

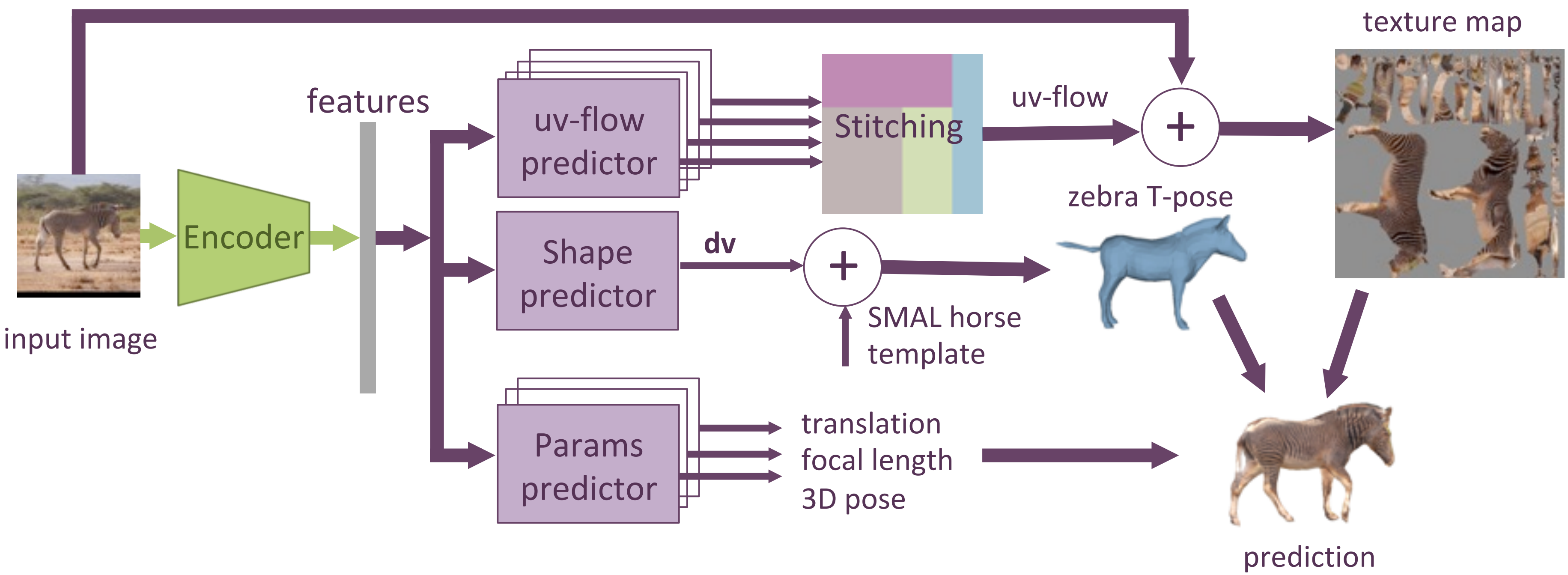
$$\mathbf{v}_{shape}(f_s) = \mathbf{v}_{template} + \mathbf{dv}$$

$$\mathbf{dv} = W f_s + b$$

SMAL model:

$$\mathbf{v}_{shape}(\beta) = \mathbf{v}_{template} + B_s \beta$$

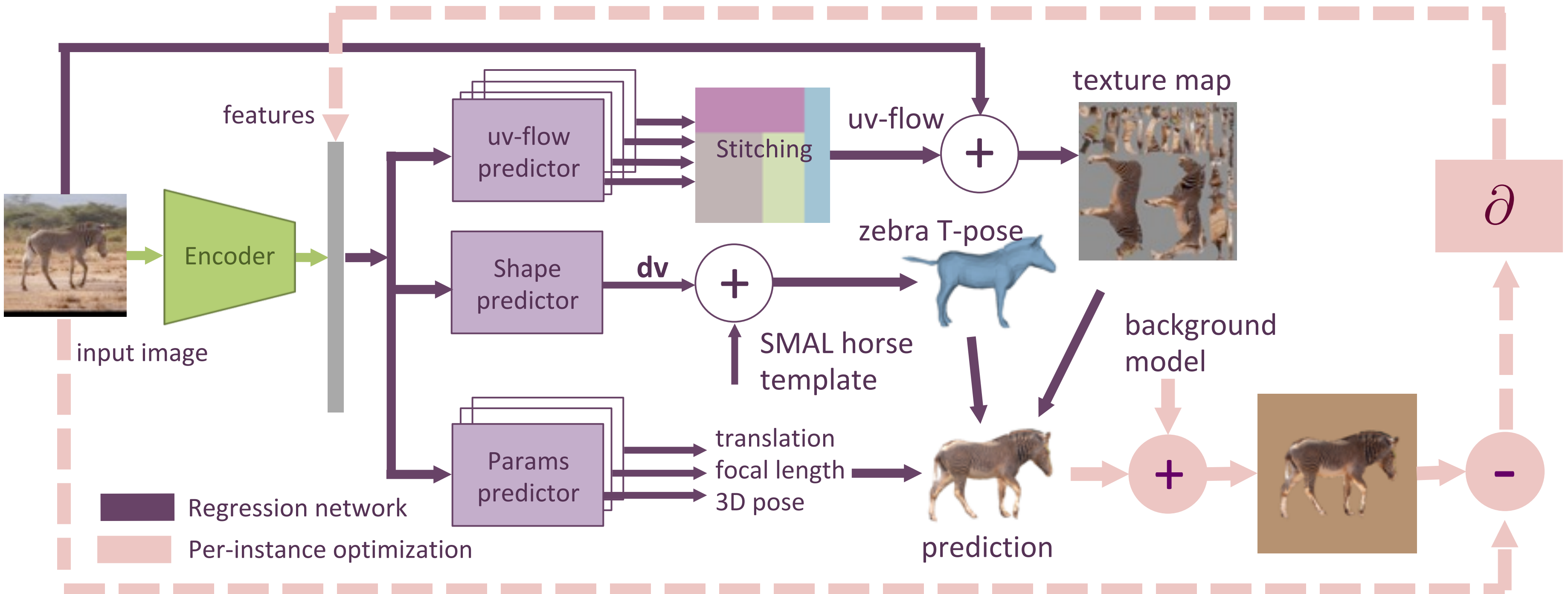
Network



$$L_{train} = L_{mask}(S_{gt}, S) + L_{kp_{2D}}(K_{2D,gt}, K_{2D}) + L_{cam}(f_{gt}, f) + L_{img}(I_{input}, I, S_{gt}) + L_{pose}(\theta_{gt}, \theta) + L_{trans}(\gamma_{gt}, \gamma) + L_{shape}(\mathbf{dv}_{gt}, \mathbf{dv}) + L_{uv}(\mathbf{uv}_{gt}, \mathbf{uv}) + L_{tex}(T_{gt}, T) + L_{dt}(\mathbf{uv}, S_{gt})$$

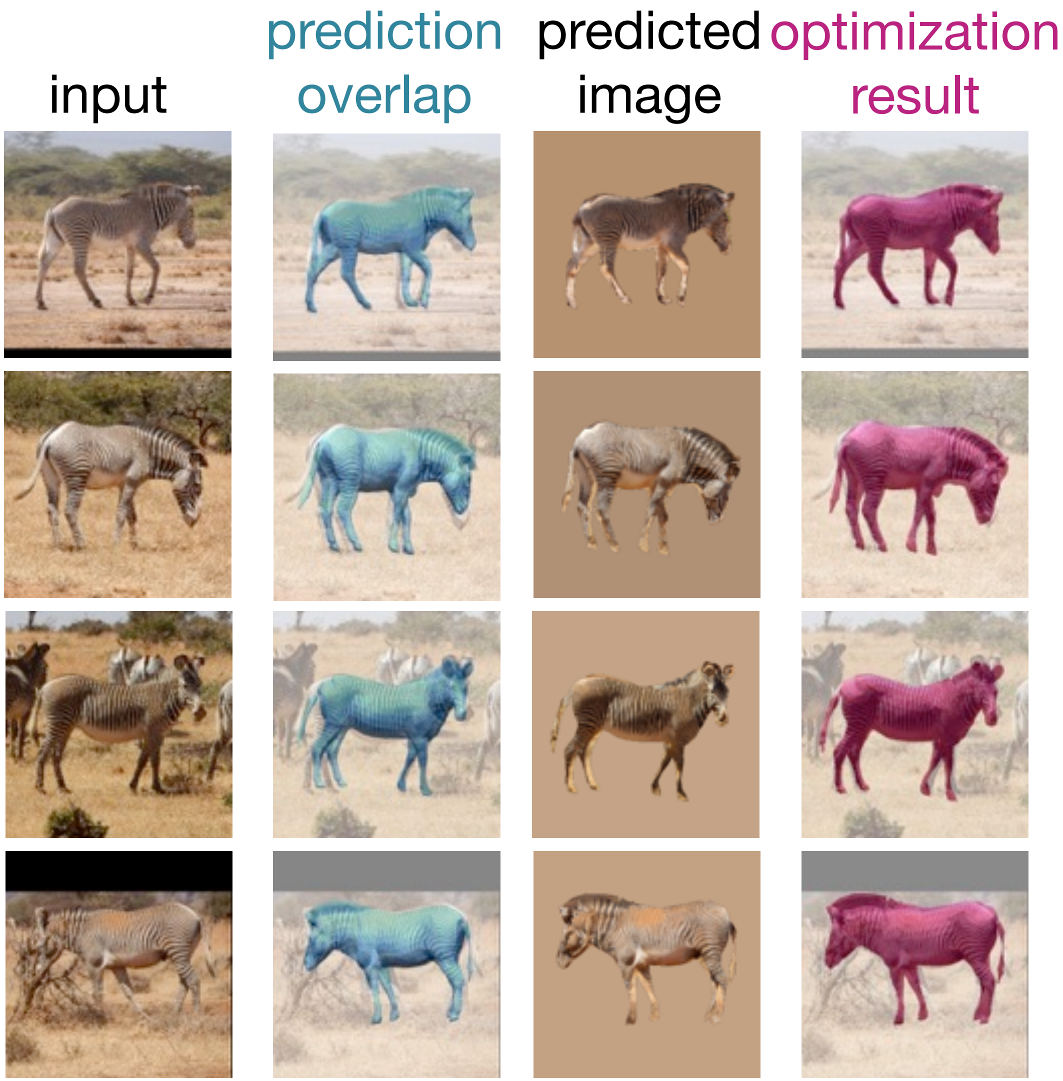
Results on test set

Unsupervised optimization



$$L_{opt} = L_{photo}(I_{input}, I) + L_{cam}(\hat{f}, f) + L_{trans}(\hat{\gamma}, \gamma)$$

Unsupervised optimization



Results

Method	PCK@0.05	PCK@0.1	IoU
(A) SMAL (gt kp and seg)	92.2	99.4	0.463
(B) feed-forward on synthetic	80.4	97.1	0.423
(C) opt features	**62.3**	**81.6**	**0.422**
(D) opt variables	59.2	80.6	0.418
(E) opt features bg img	59.7	80.5	0.416
(F) feed-forward pred.	59.5	80.3	0.416
(G) no texture	52.3	76.2	0.401
(H) noise bbox	58.7	79.9	0.415

Texture prediction helps!

Better to optimize over features

Poster n.93, 31st Oct 10:30

S. Zuffi, A. Kanazawa, T. Berger-Wolf, M.J. Black, 3D Safari: Learning to Estimate Zebra Pose, Shape, and Texture from Images "In the Wild", ICCV 2019