Perceiving Non-Rigid Objects in the 3D World

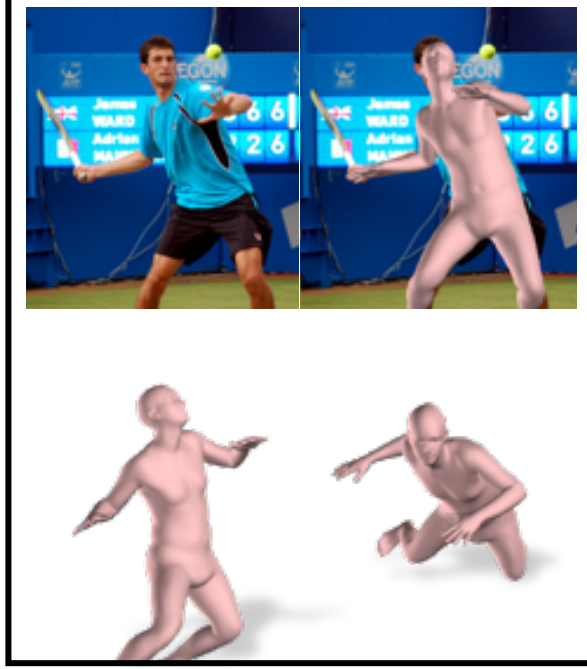
Angjoo Kanazawa

We live in a world that is 3D and dynamic.

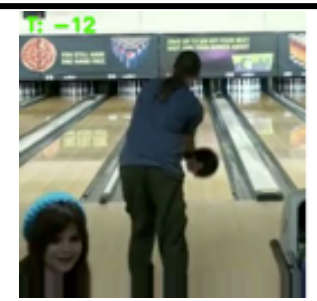
Goal: **Perceive** these embodied agents in the 3D world from visual inputs

What I won’t talk about: Humans

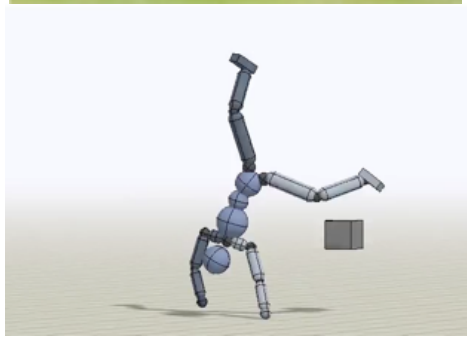
3D Humans from a Single Image and Video



3D Prediction

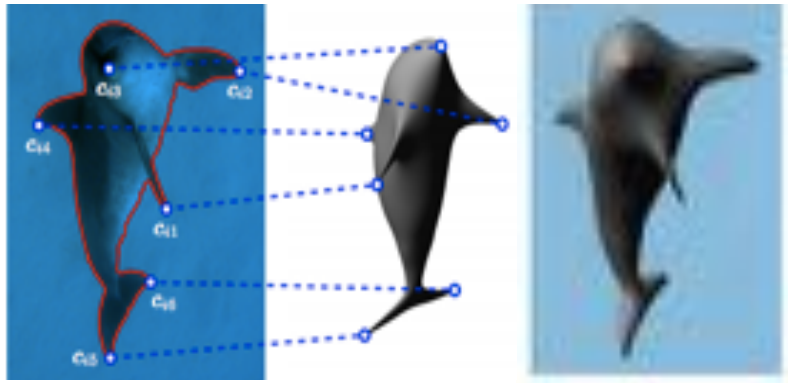


3D for Learning to Act



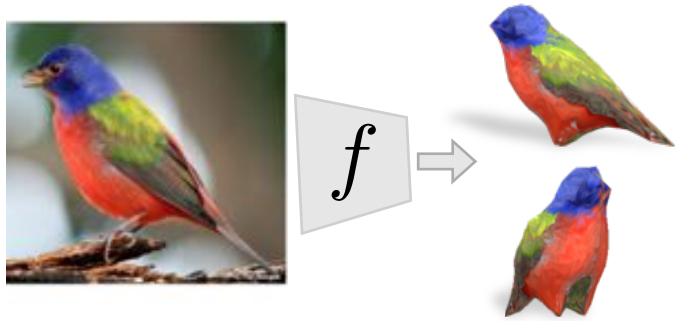
3D Human talk at DynaVis workshop @ 3:30pm

This talk: Articulated deformable objects



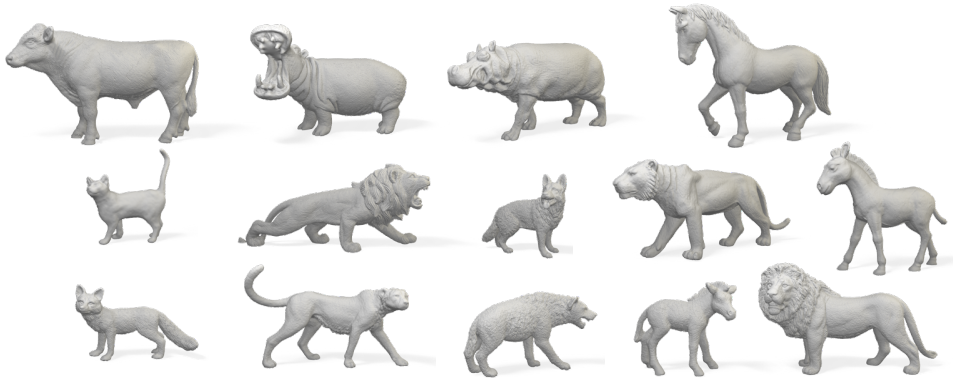
[Cashman and Fitzgibbon. 2012]

[Kanazawa, Kovalsky, Basri, Jacobs. EUROGRAPHICS 2016]



[Kanazawa*, Tulsiani*, Efros, Malik, ECCV 2018]

Deformable 3D model from limited 3D data



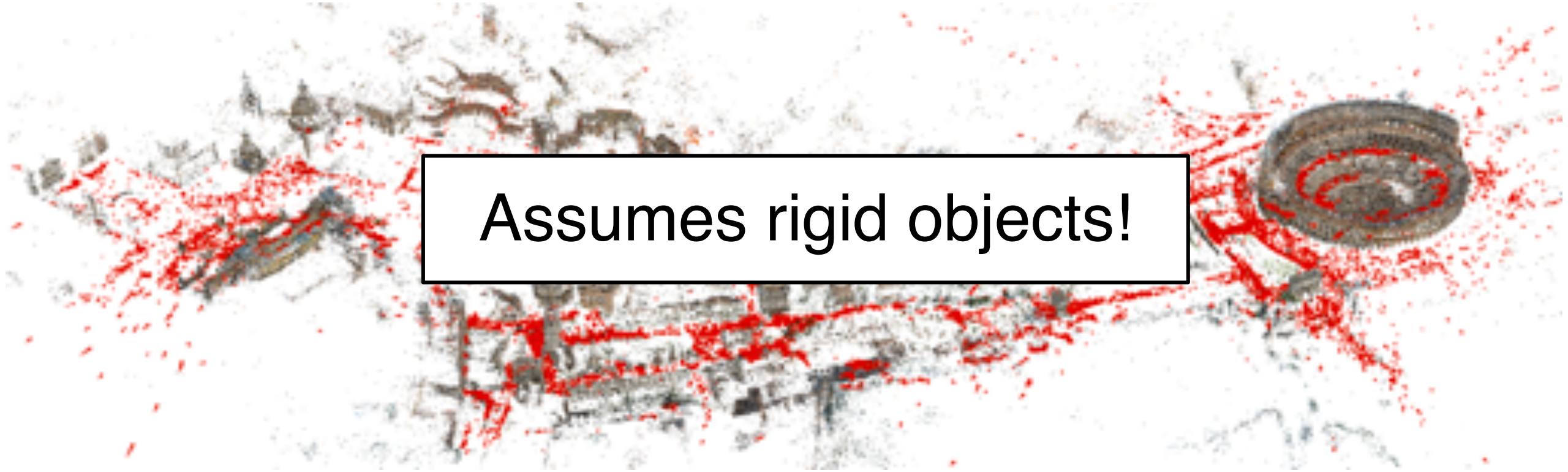
[Zuffi, Kanazawa, Black CVPR 2017]

Shape detail from Images

[Zuffi, Kanazawa, Black CVPR 2018]

What makes non-rigid 3D reconstruction hard?

Lots of success in Multi-view 3D Reconstruction

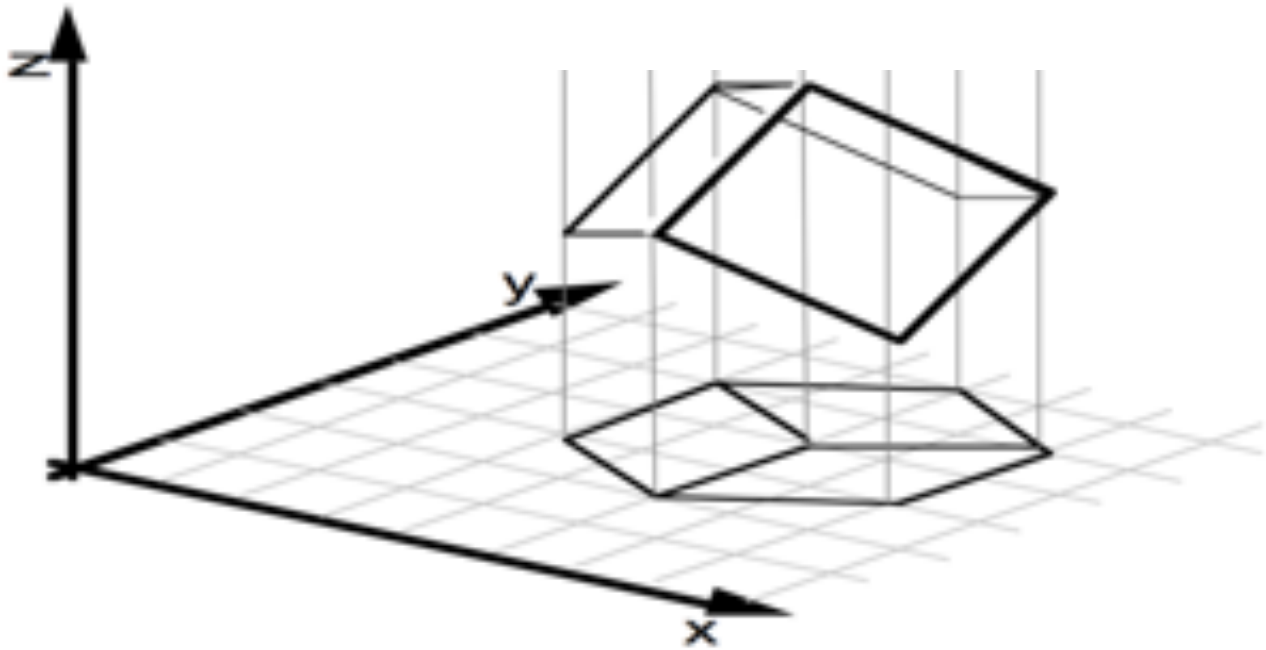


Furukawa and Ponce. PAMI ’10. Agarwal et al. ICCV ’09. Schönberger et al. CVPR ’16. …

Today's Non-rigid 3D Solution: Motion Capture

Desired: 3D perception from images

3D from 2D is inherently under-constrained



[Sinha and Adelson ‘93]

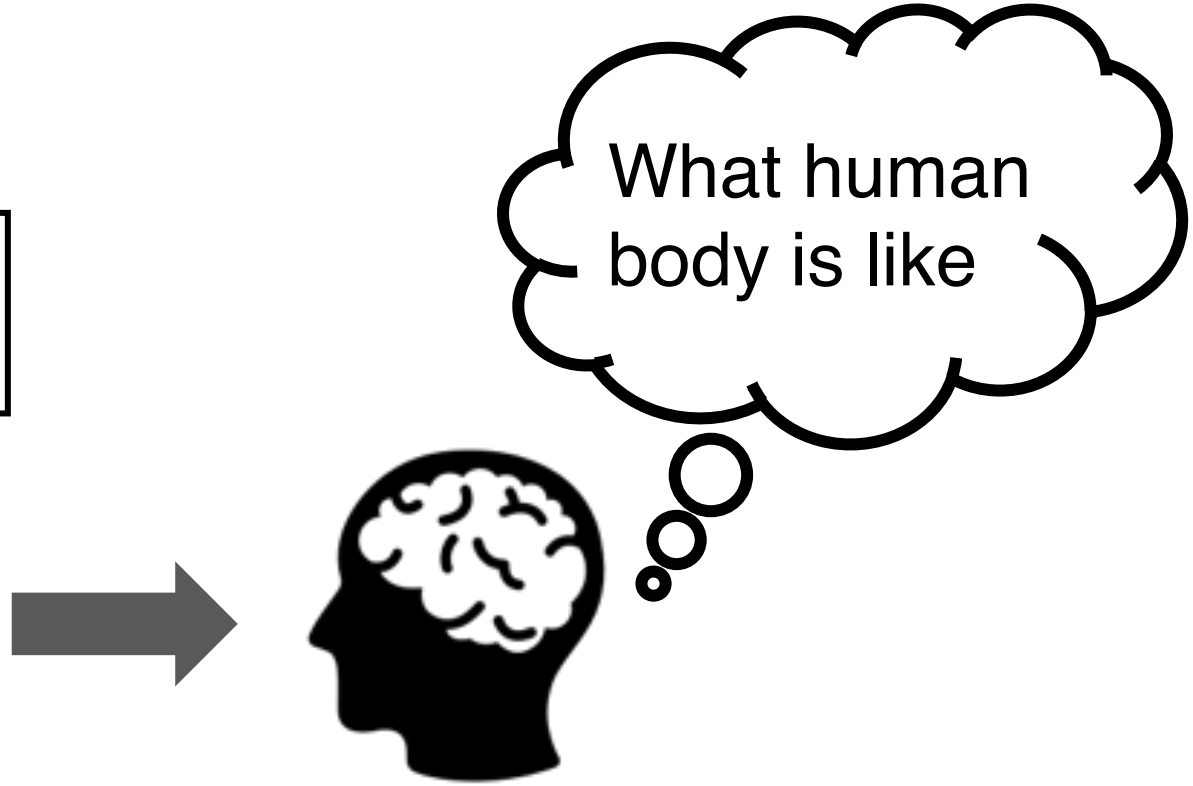
But we can perceive a lot of 3D structure from a single image.

How do we resolve this?

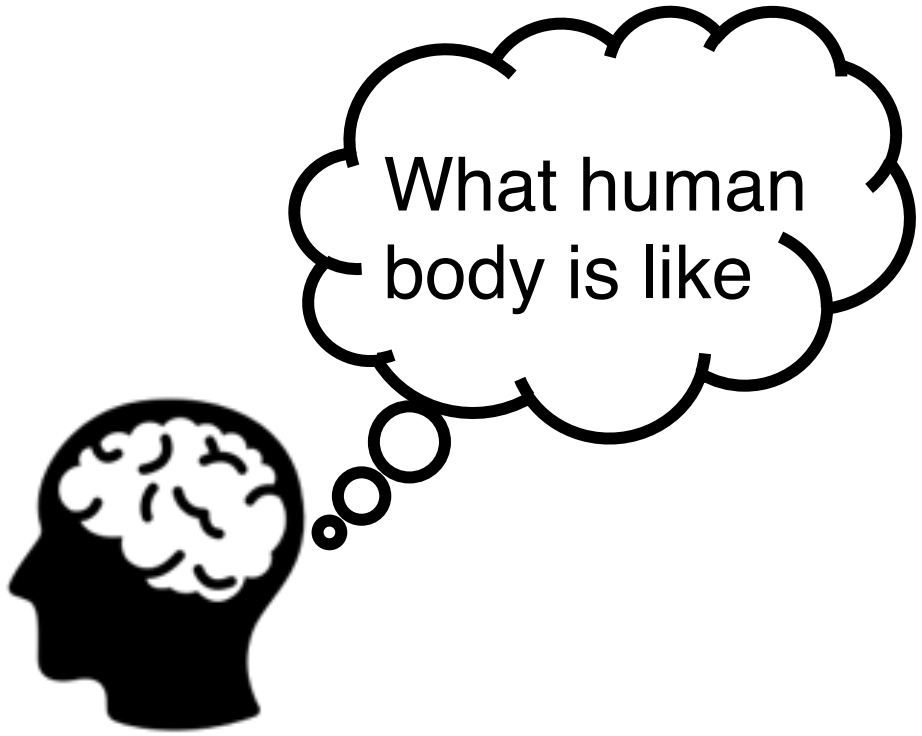
[Bogo and **Kanazawa** et al. ECCV '16]

How do we resolve this?

"Remembrance of Objects Past"



Key Question: How do we get this prior?



Humans are Special

Robinette et al., Civilian American and European Surface Anthropometry Resource (CAESAR) 2002.

Morphable Model of Human Bodies

SMPL [Loper et al. SIGGRAPH Asia '15]

Humans are Special

...

[LSP, MPII, MS COCO,…]

[Cao et al. CVPR'17]

Problem with Animals

Limited availability of 3D Data

- Deformable
- Hard to get 3D scans for training models
- Can’t bring into the lab
- Not cooperative! Won’t stay still

of 3D models on Turbosquid

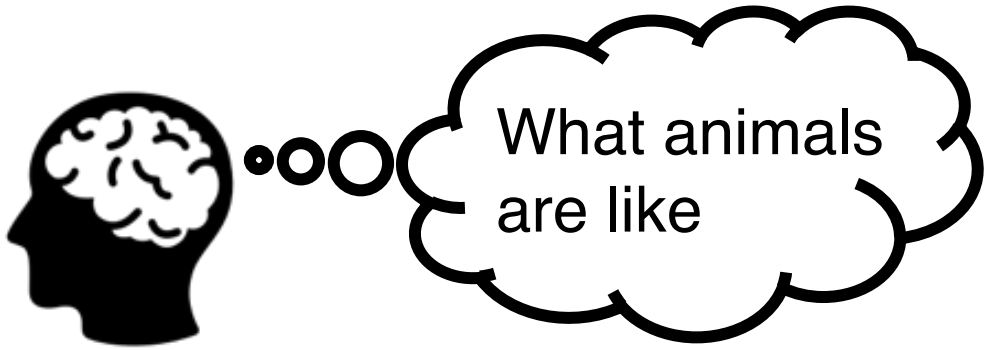
Cats ???

But the internet is full of cat pictures..

Motivation:

Non-rigid 3D Modeling

Correspondence across deformation

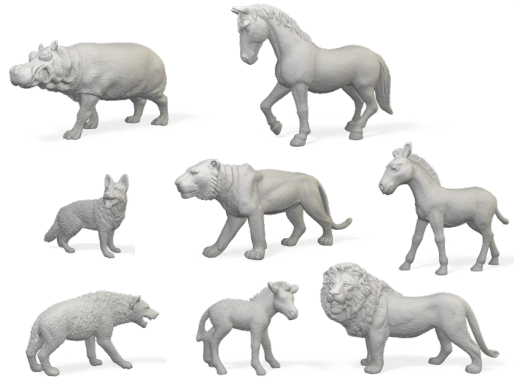


<u>Under Limited Supervision</u>

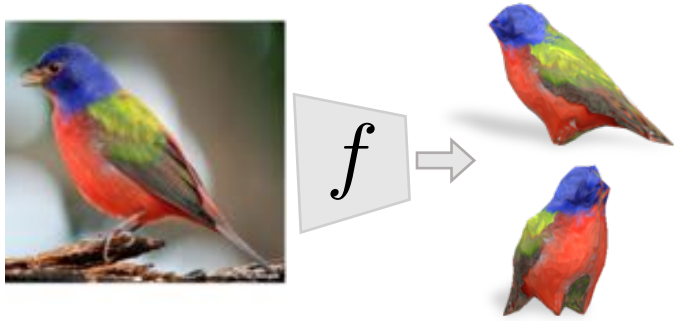
Overview of 3D Animal Reconstruction

1. Let's start with a template 3D model + images

2. What if we had some 3D data?



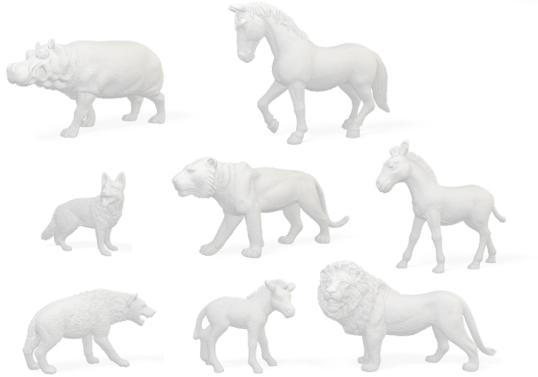
3. What if we don't have any 3D data?



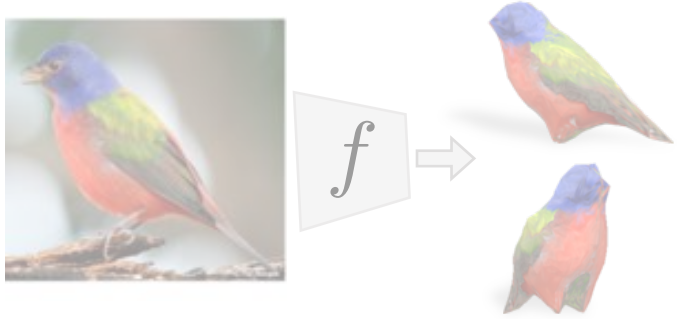
Overview of 3D Animal Reconstruction

1. Let's start with a template 3D model + images

2. What if we had some 3D data?

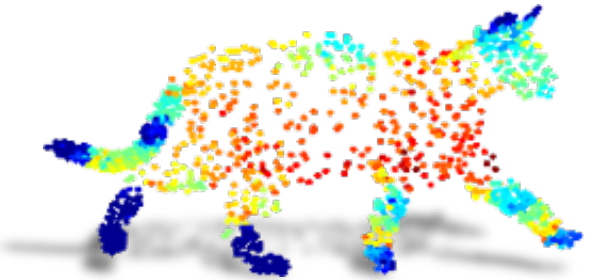


3. What if we don't have any 3D data?



Learning 3D Deformation of Animals From 2D Images

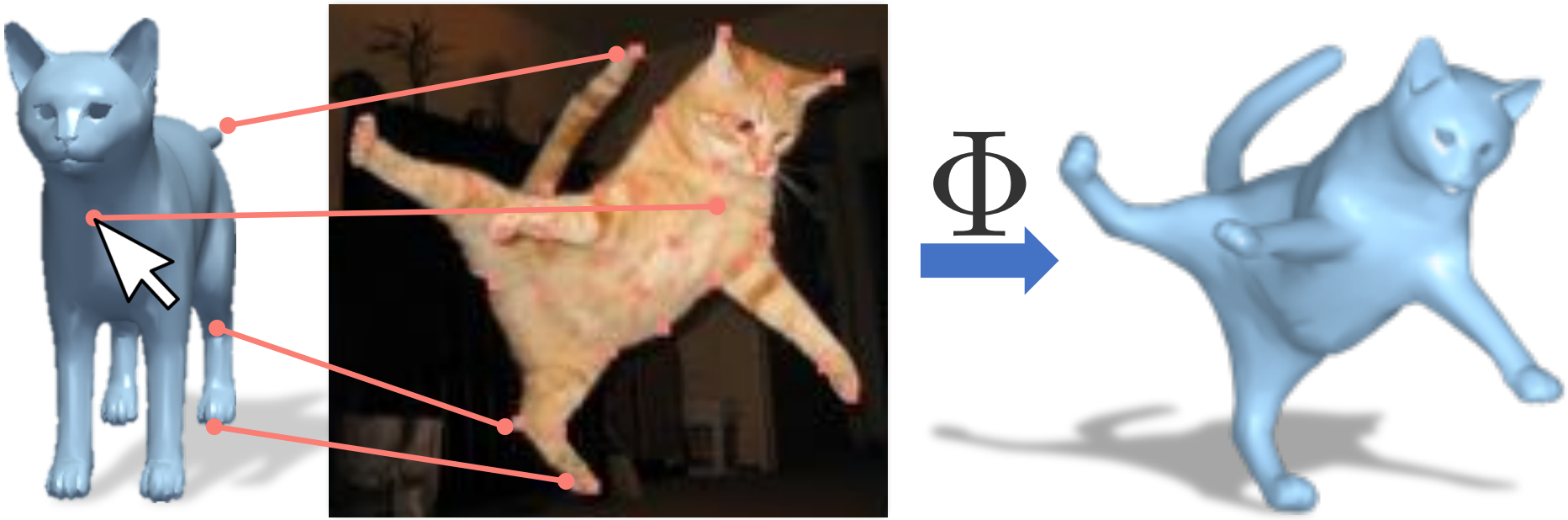
Angjoo Kanazawa, Shahar Kovalsky, Ronen Basri, David Jacobs



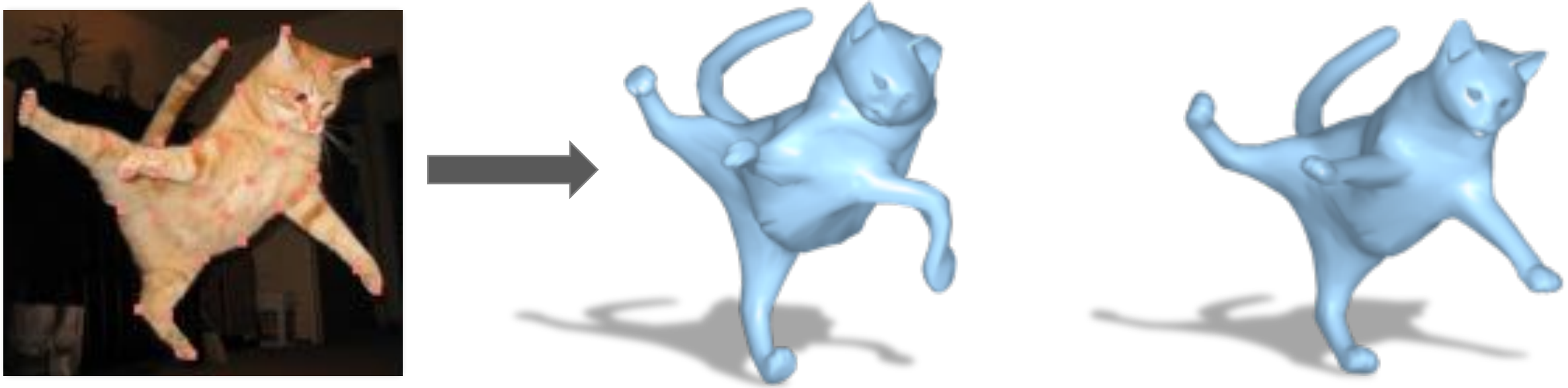
EUROGRAPHICS 2016 **BEST PAPER**

WEIZMANN INSTITUTE OF SCIENCE

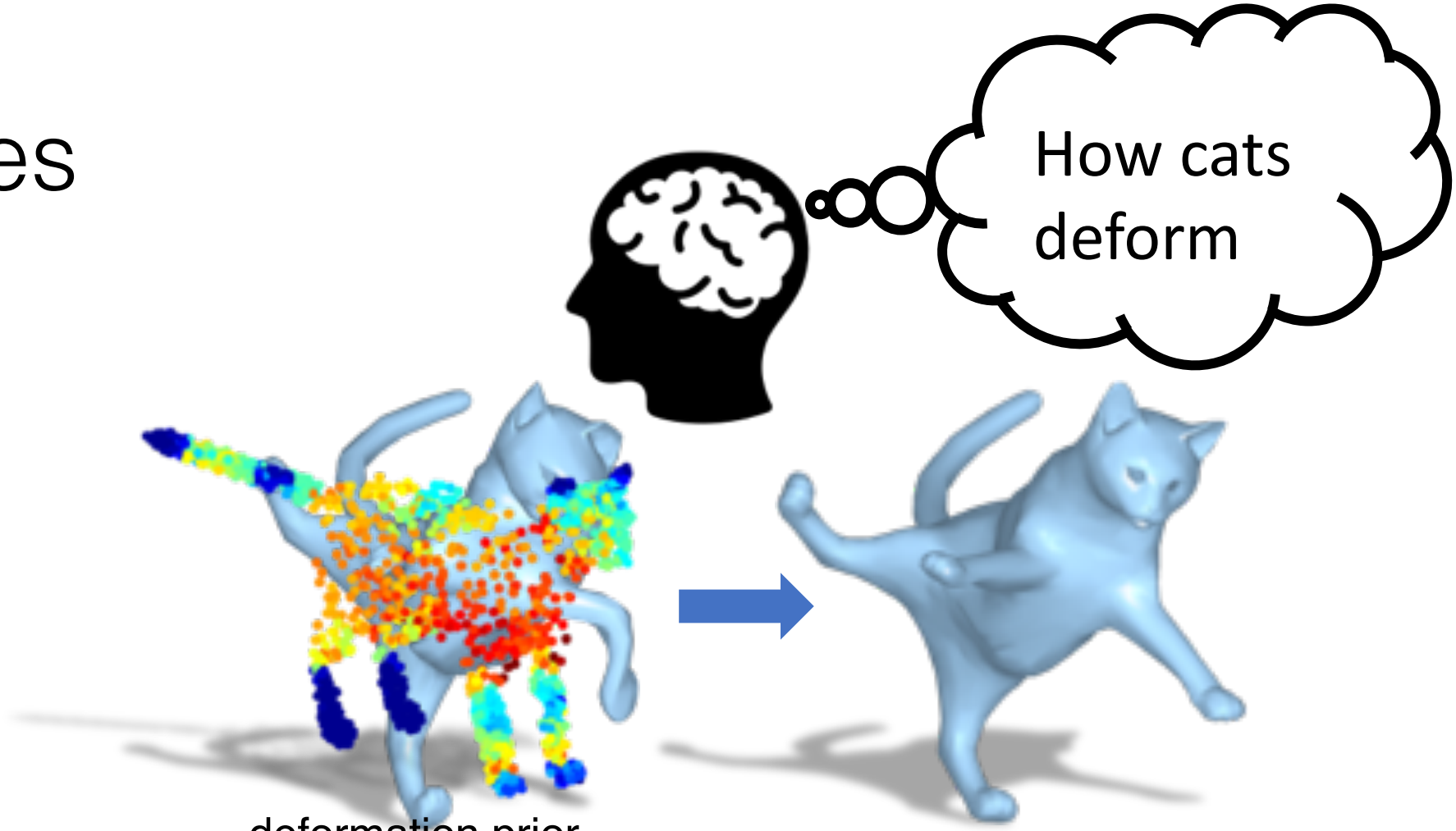
Idea: Top-down 3D modelling



But many solutions exist…



Ambiguities

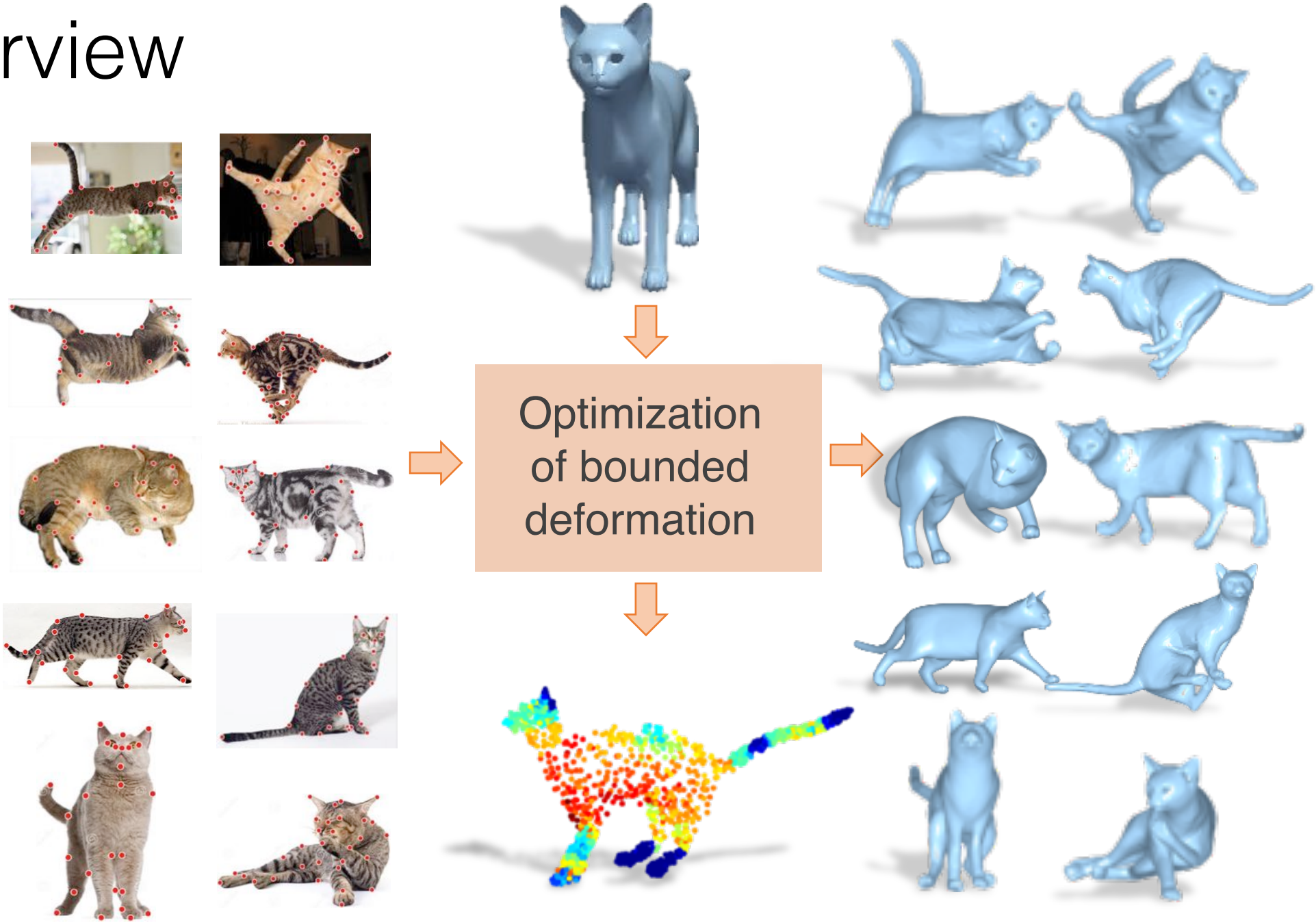


Q: Can we learn 3D deformation prior from 2D images?

Intuition

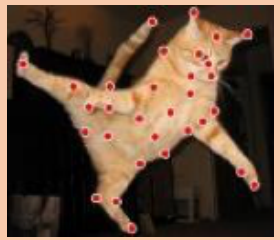
Highly deformable regions are **sparse** and **consistent** across multiple images of cats

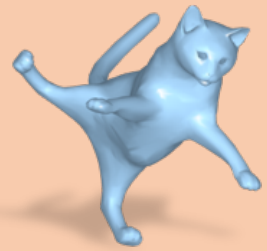
Overview



Problem Formulation

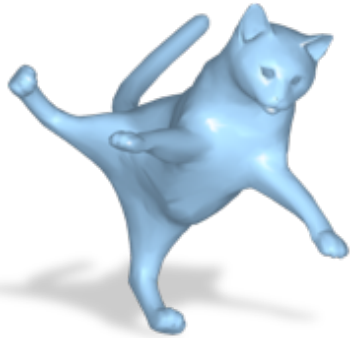
$$\sum_{\mathbf{cats}} \| \quad - \quad \|_2^2$$





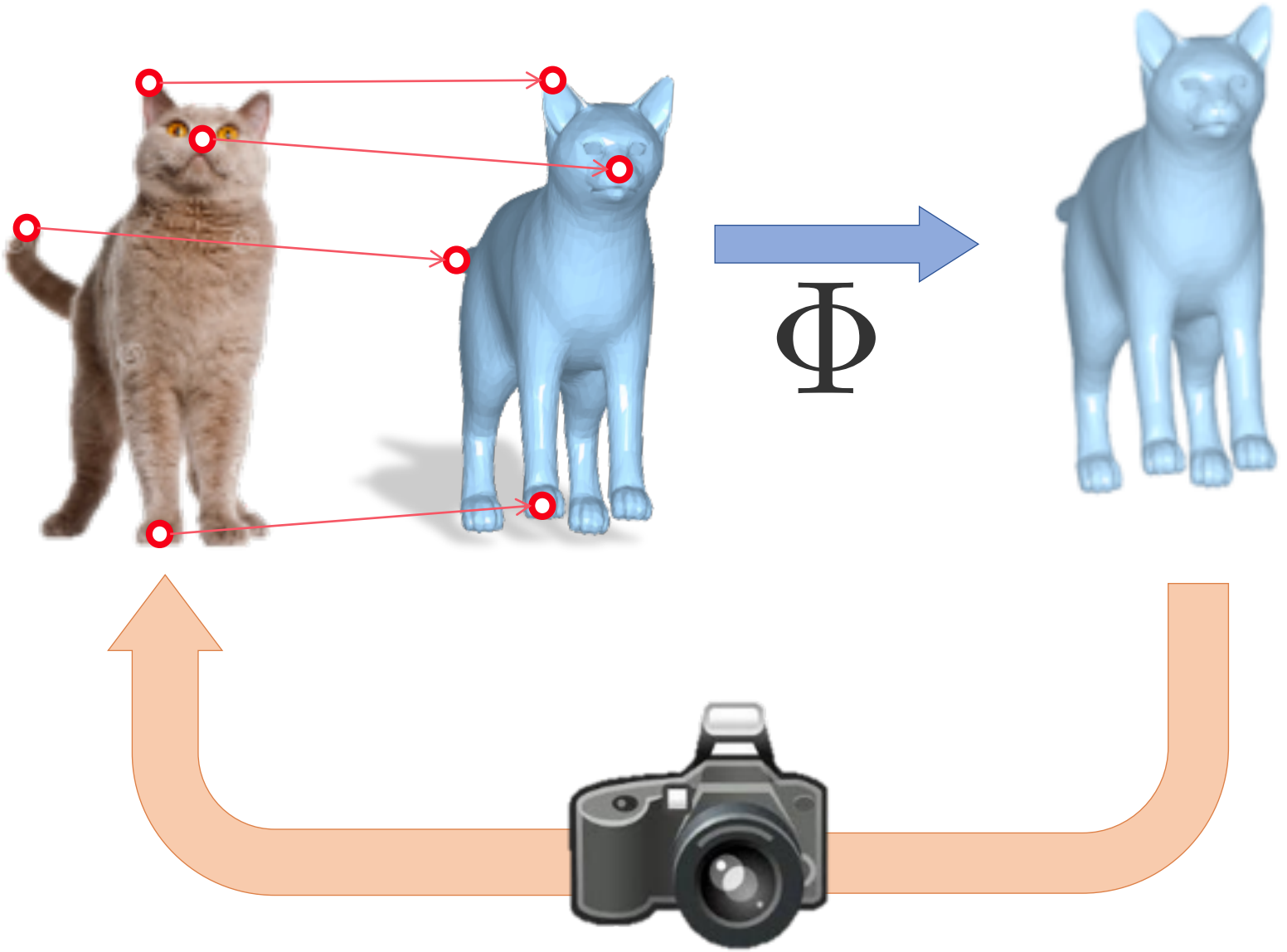
Data term

$s.t.$



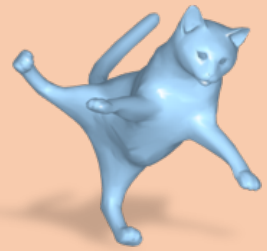
bounded by

Data term: 2D reprojection error



Problem Formulation

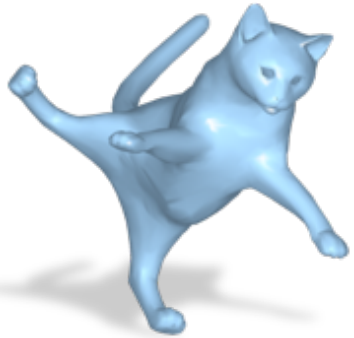
$$\sum_{\mathbf{cats}} \| \quad - \quad \|_2^2 + \lambda \| \quad \|_1$$



Data term

Stiffness

$s.t.$

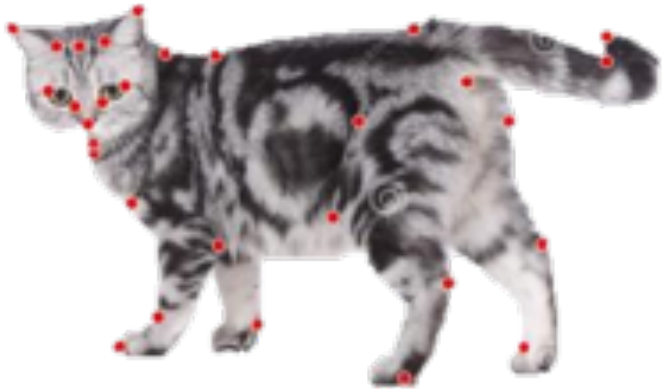


bounded by

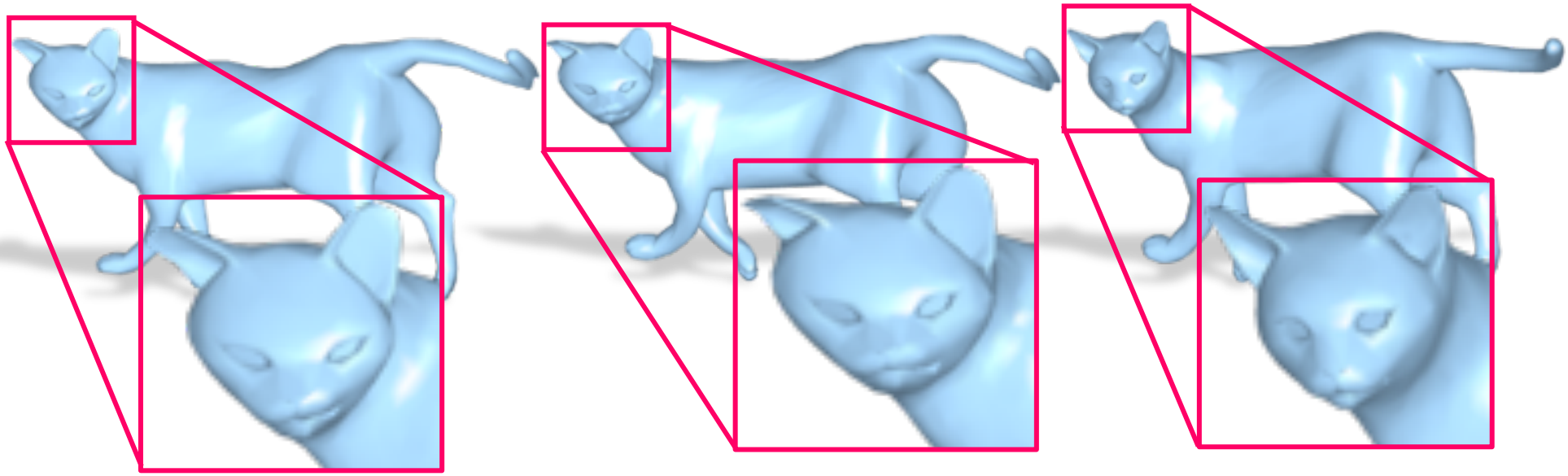
Results

No Prior	Uniform Prior	Learned Prior

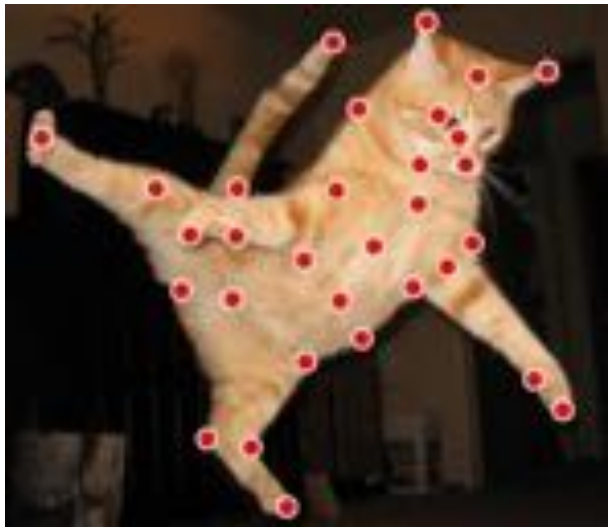
Results

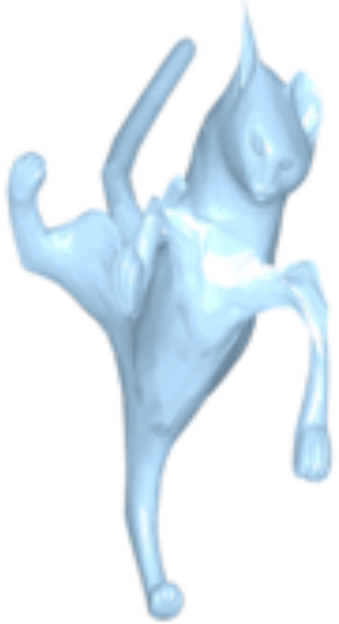
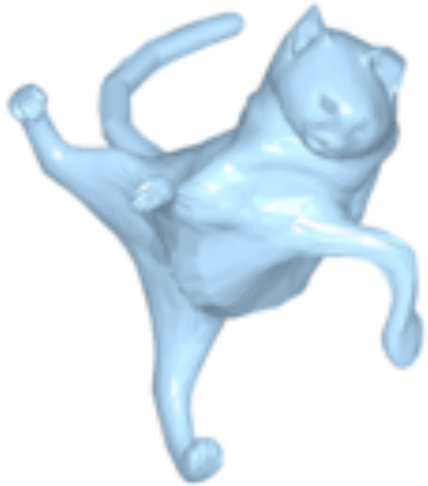


No Prior	Uniform Prior	Learned Prior



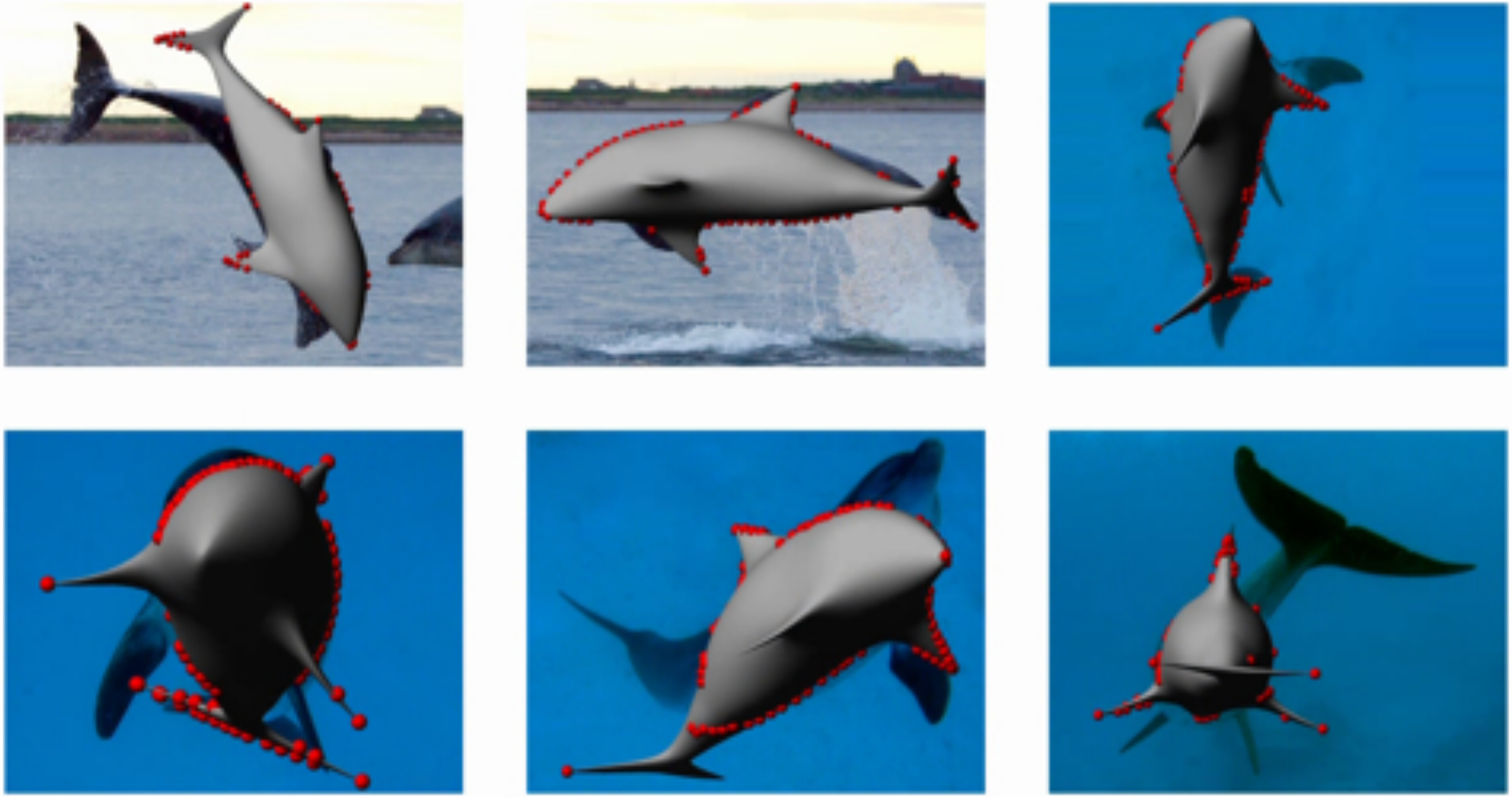
Results



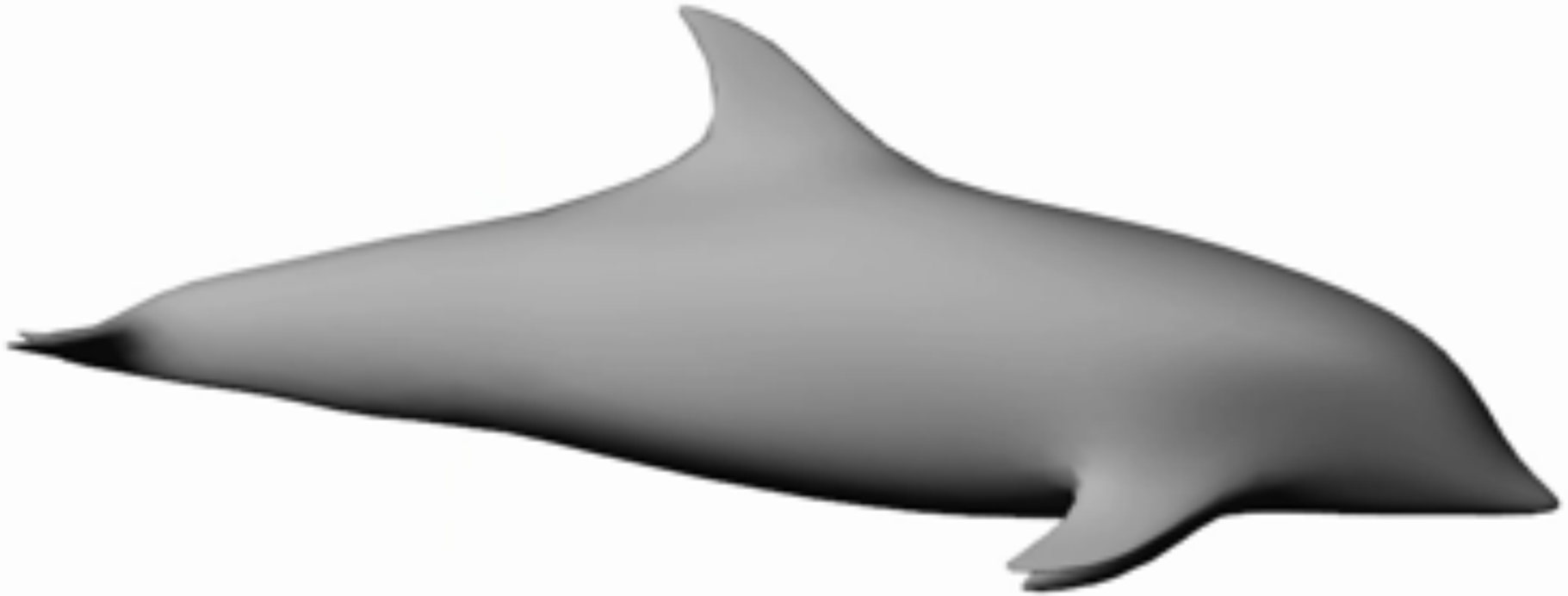
No Prior	Uniform Prior	Learned Prior
		

Learning shape deformation from images

What shapes are dolphins? [Cashman and Fitzgibbon 2012]

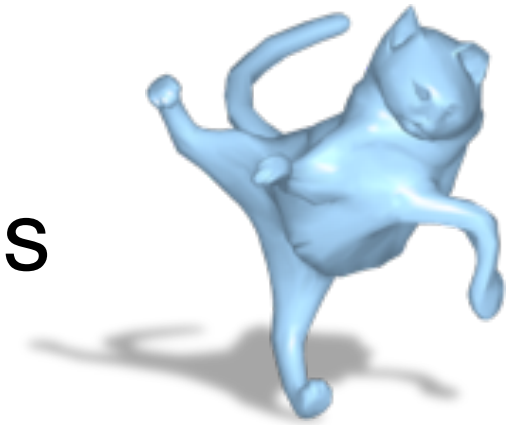


Morphable model parameters: 1



Key take away

- Fitting to single image is ambiguous

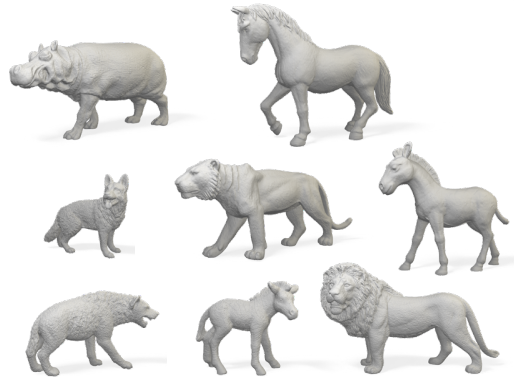


- but fitting a single model to many images allow learning priors that constrain the model

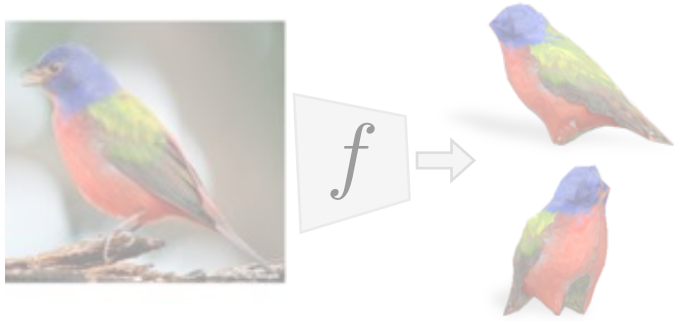
Overview of 3D Animal Reconstruction

1. Let's start with a template 3D model + images

2. What if we had some 3D data?



3. What if we don't have any 3D data?



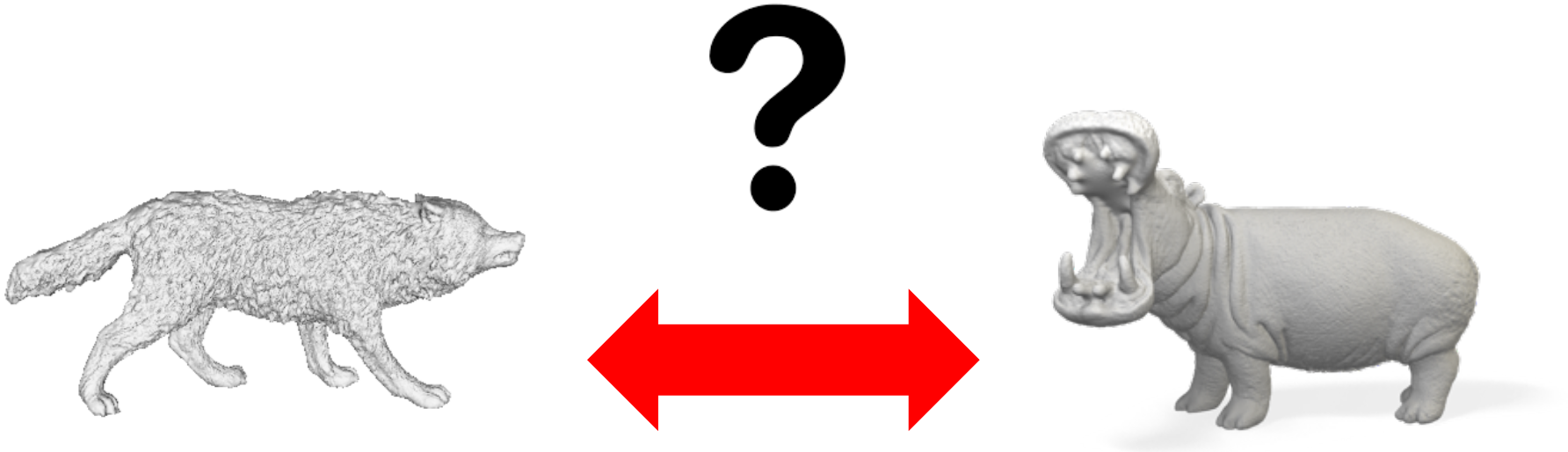
Why only use one 3D template?
Aren't there some 3D models around?

Skinned Multi-Animal Linear (SMAL) model

Learn from toy animal scans

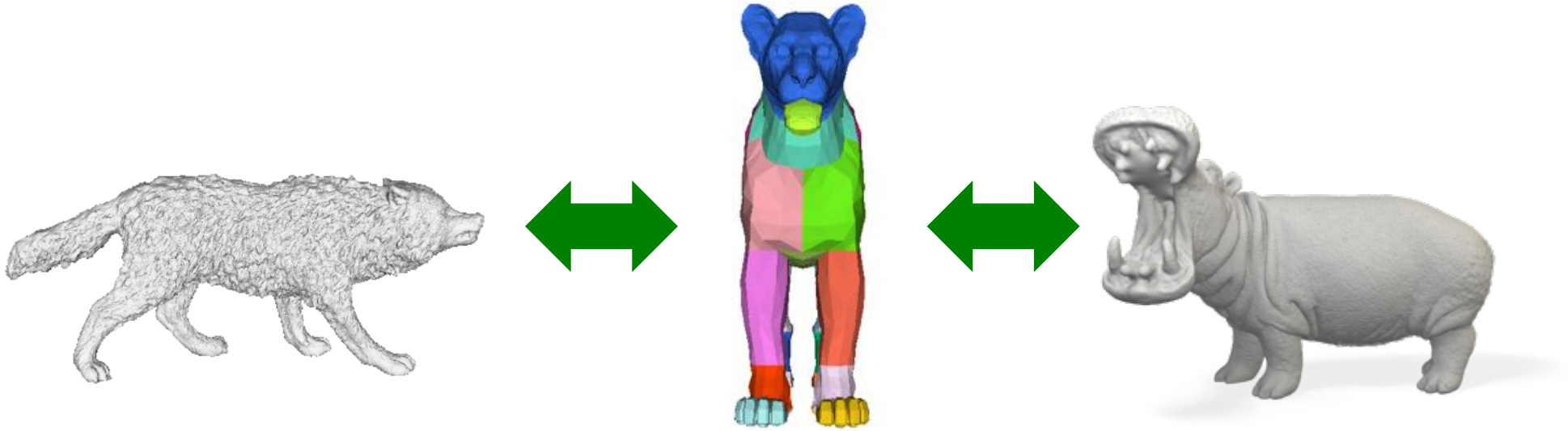
S. Zuffi, A. Kanazawa, D. Jacobs, M. J. Black, “3D Menagerie: Modeling the 3D Shape and Pose of Animals”, CVPR 2017

Challenge:
How to align a wolf to a hippo?



How to align a wolf to a hippo?

Approach: align both to a deformable lioness model



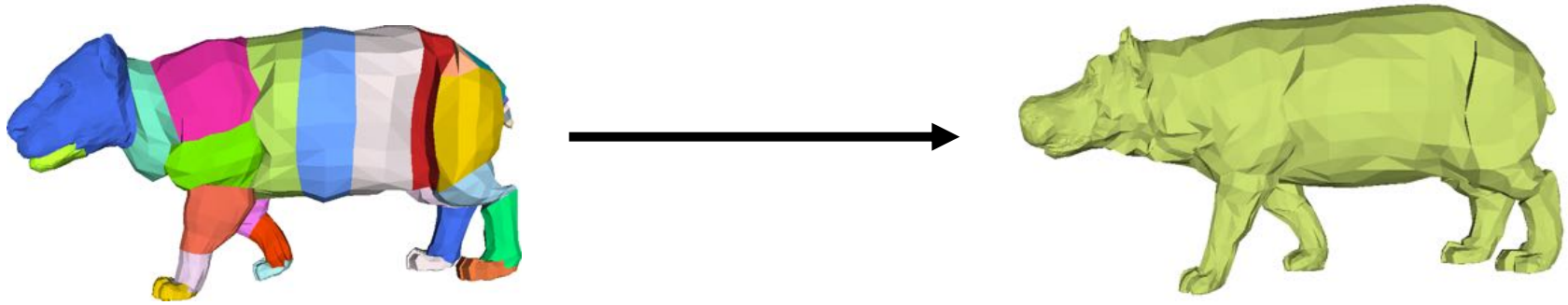
The queen of animals!

Align a coarse stitched parts to each scan

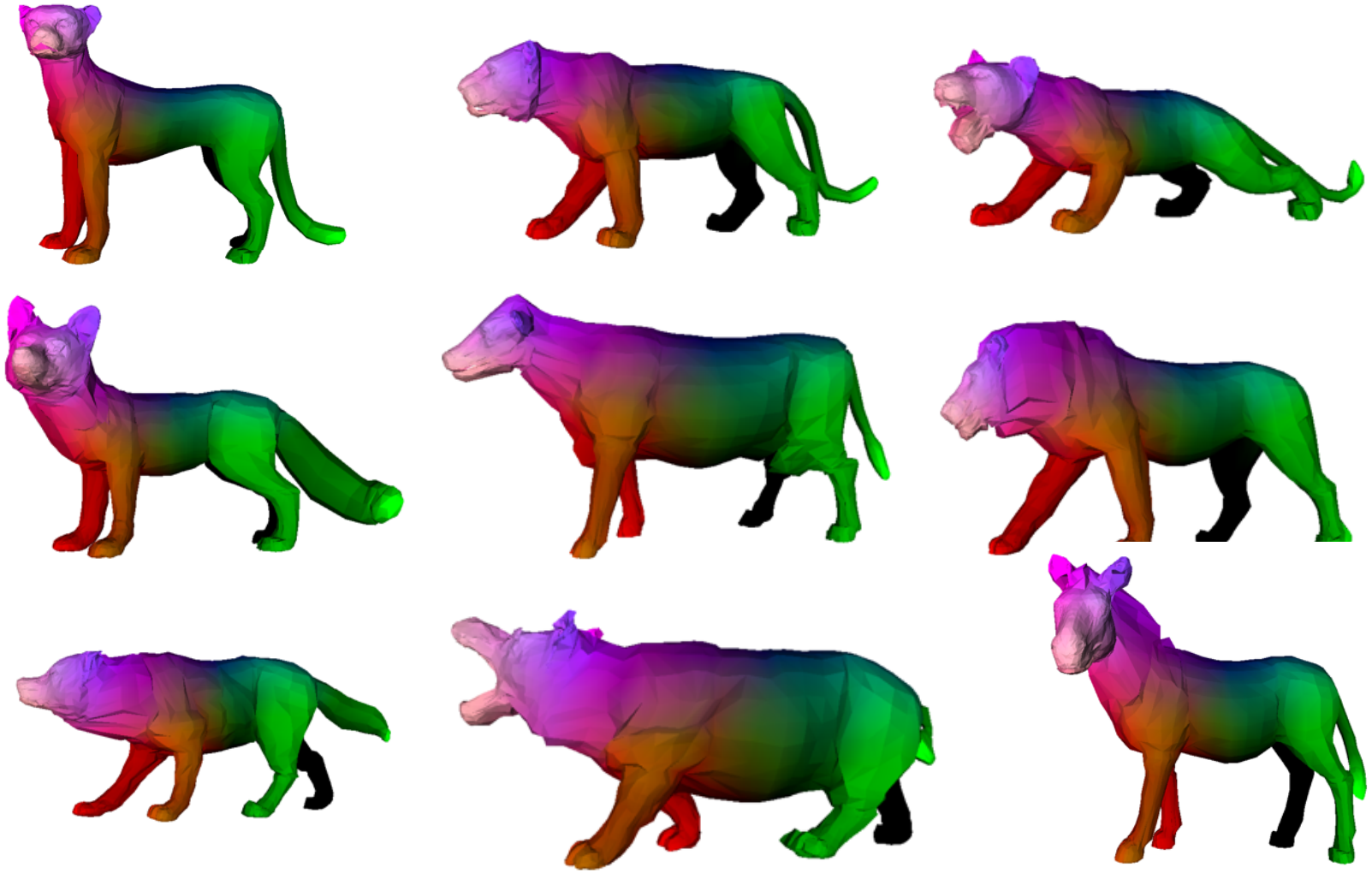
Progress of fitting a scan

Free form deformation refinement

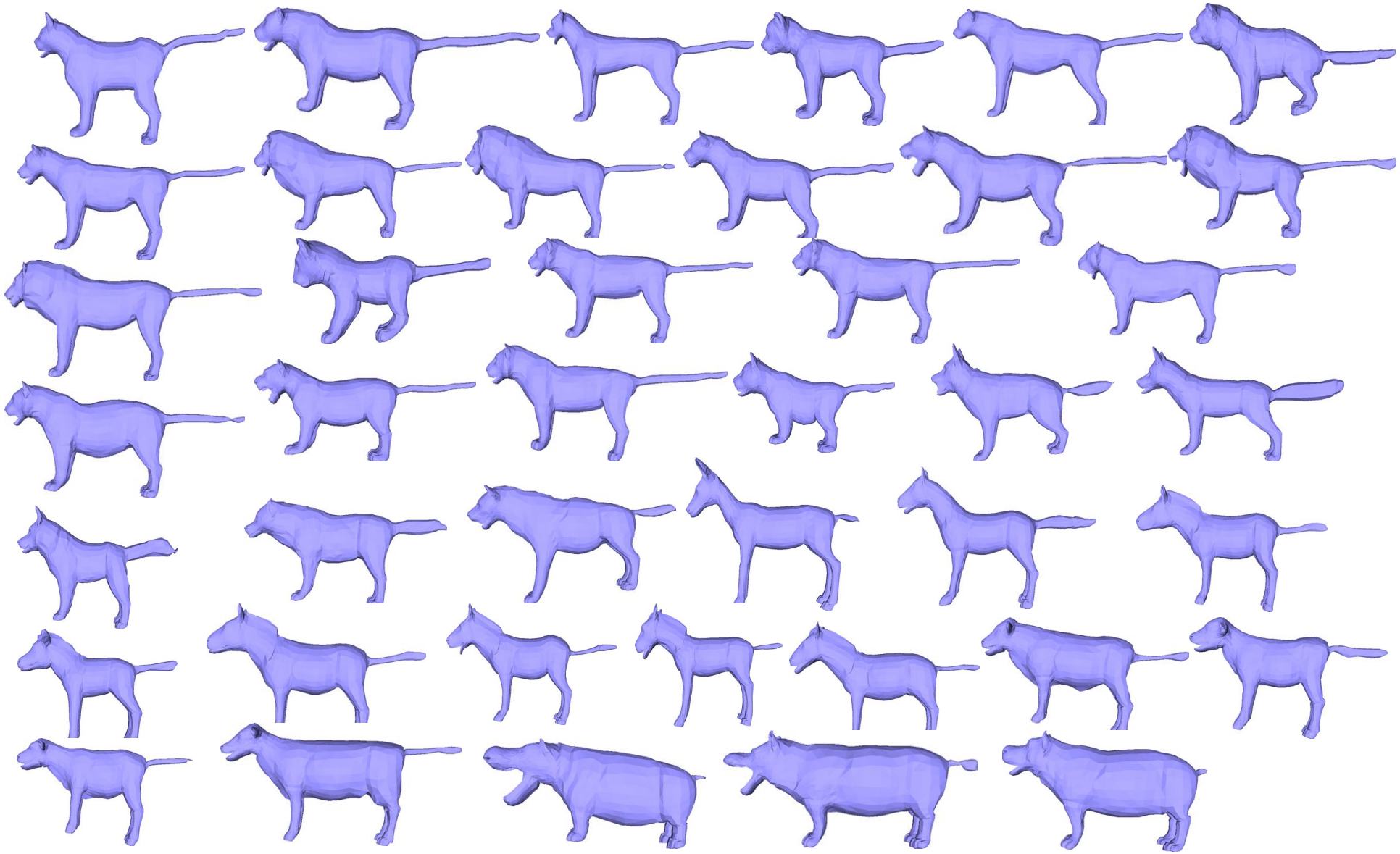
with As-Rigid-As-Possible reguralization



Model free refinement: All animals in correspondence

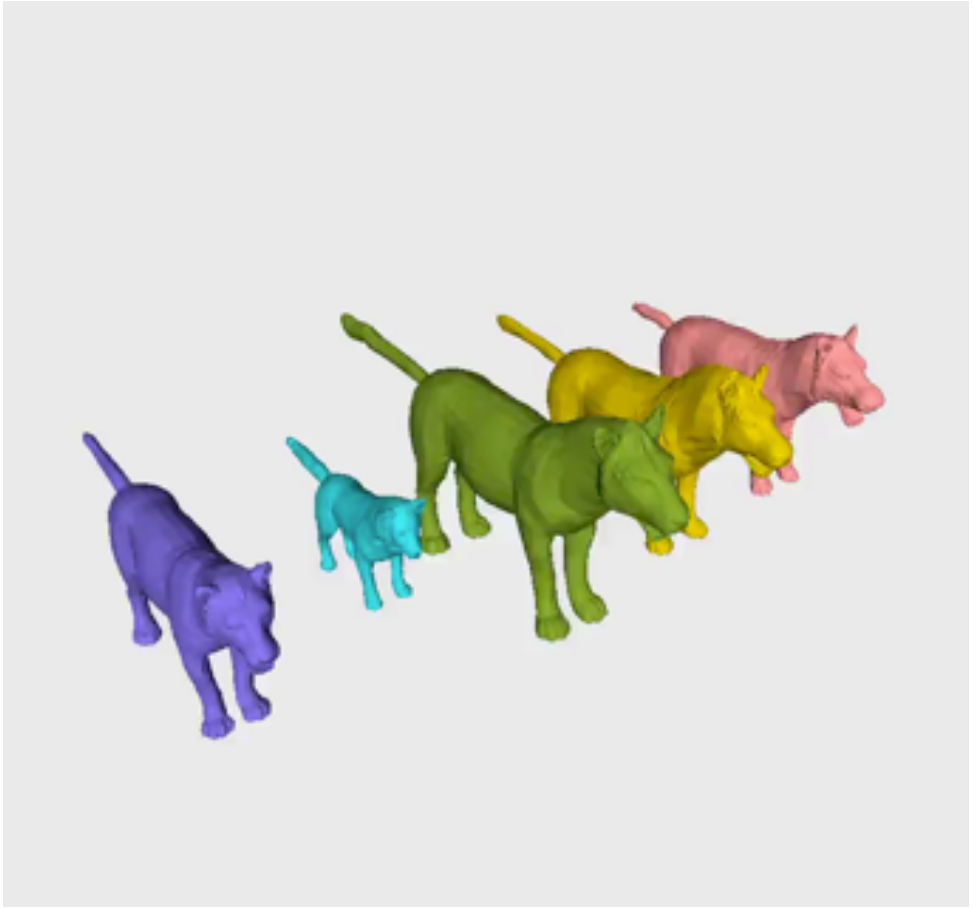


Animal shape dataset



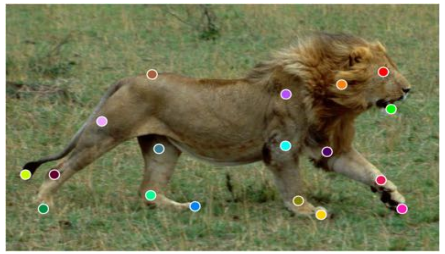
SMAL:
Skinned Multi-Animal Linear Model

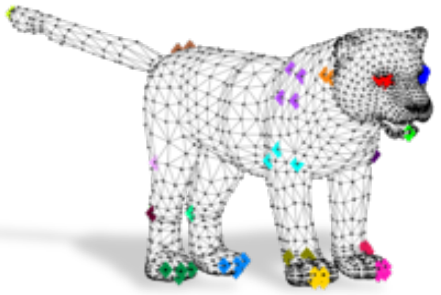
PCA Shape space



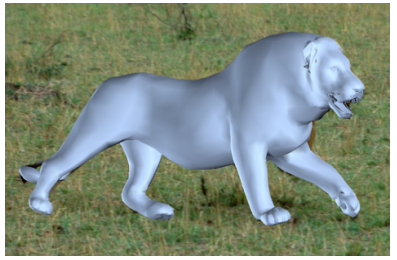
Now you can fit this model to images

With 2D keypoints and silhouettes



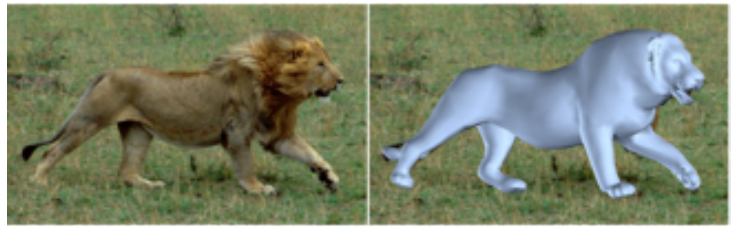


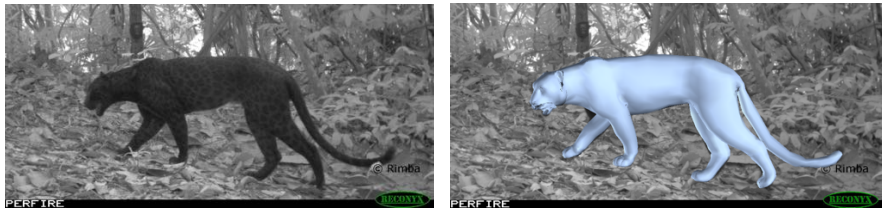
min

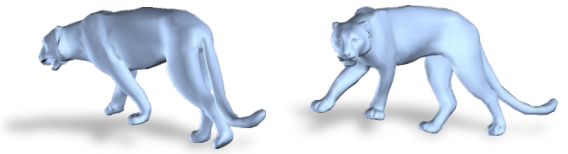
$$\left\| \quad \text{  } \quad - \quad \text{  } \quad \right\|$$

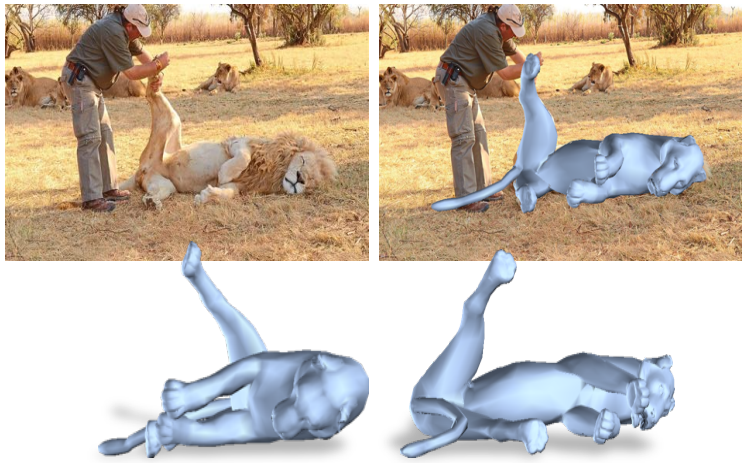
+ lots of priors

Results: Big Cats

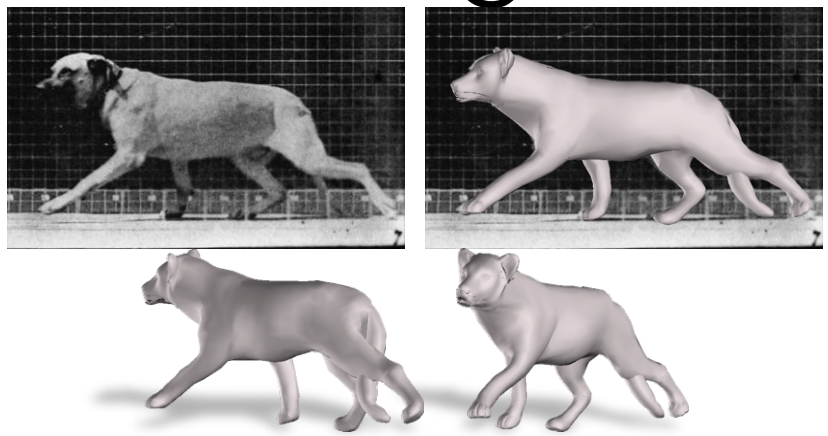


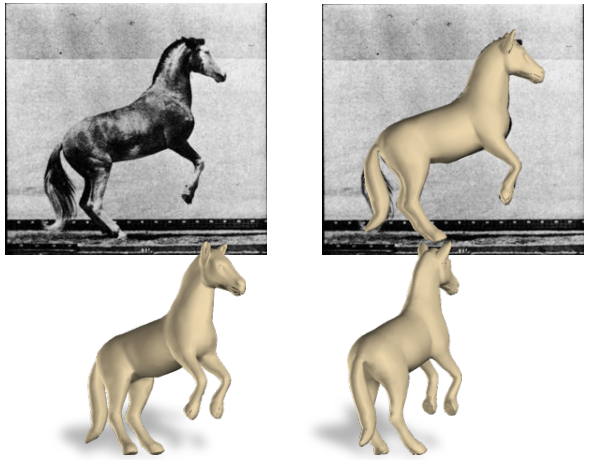


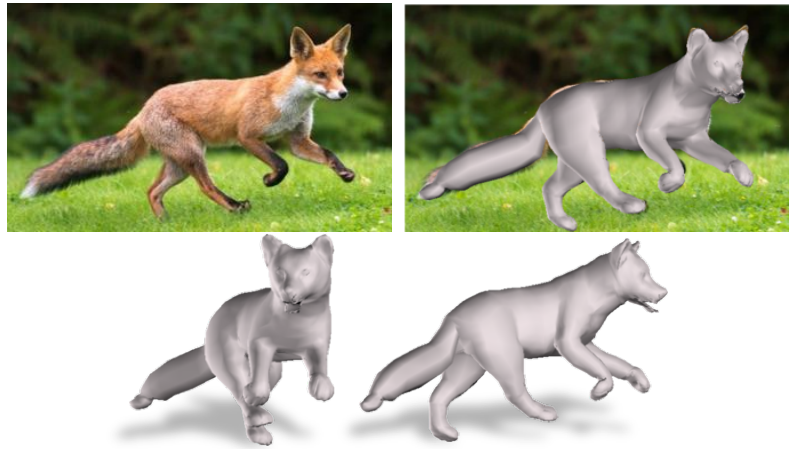


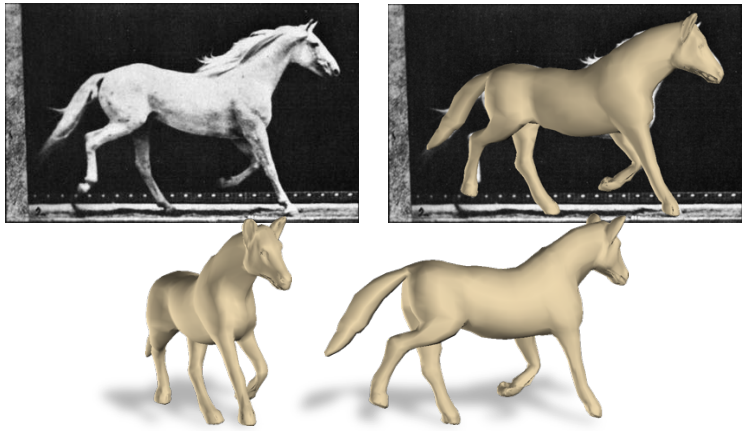


Results: Dogs and Horses

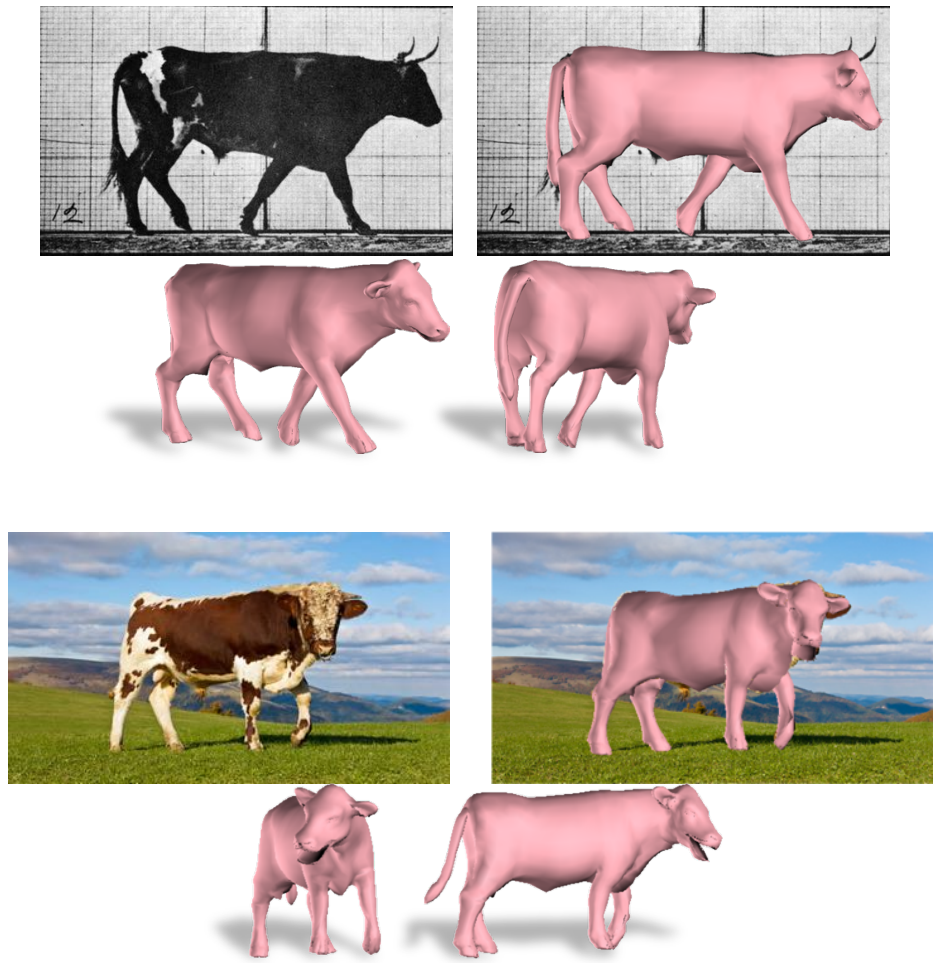


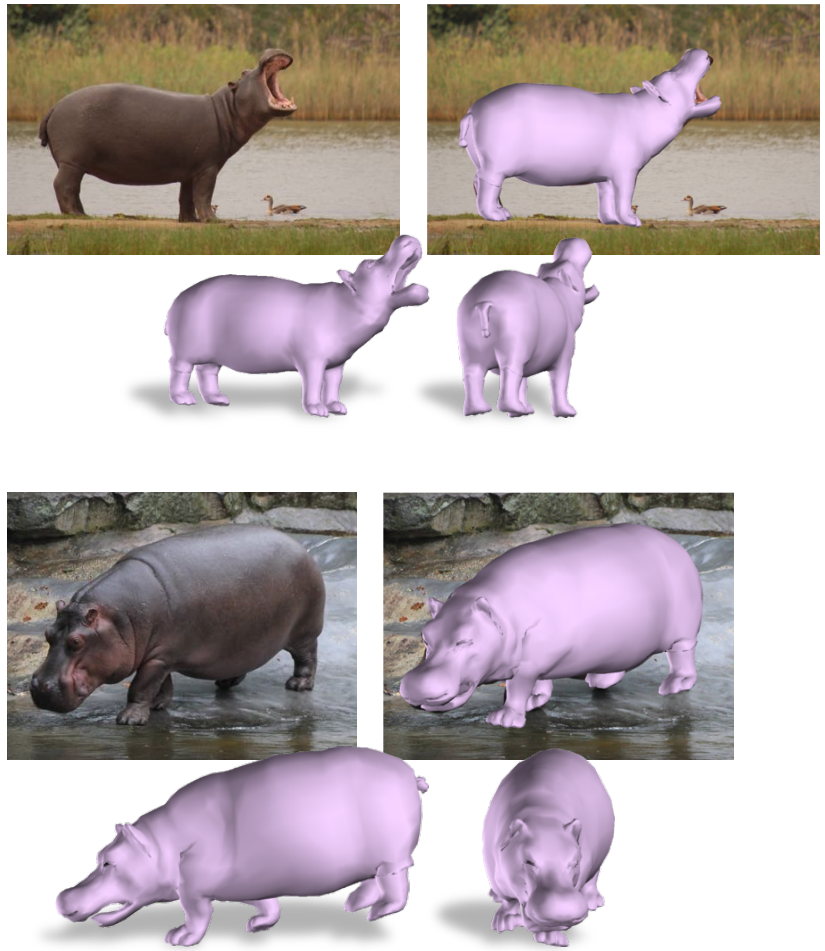




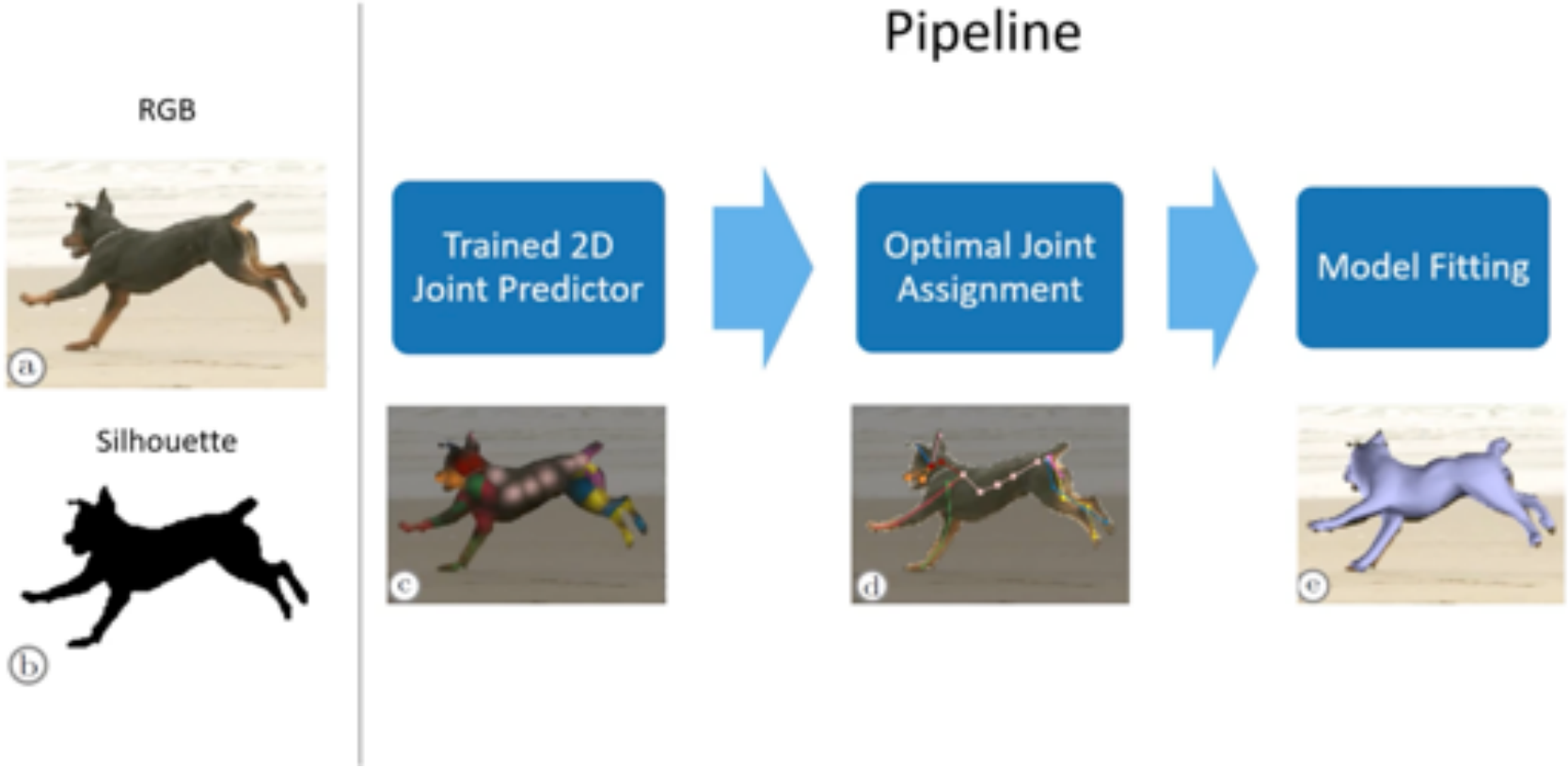


Results: Cows and Hippos





Follow up work:
Removes 2D keypoint annotations

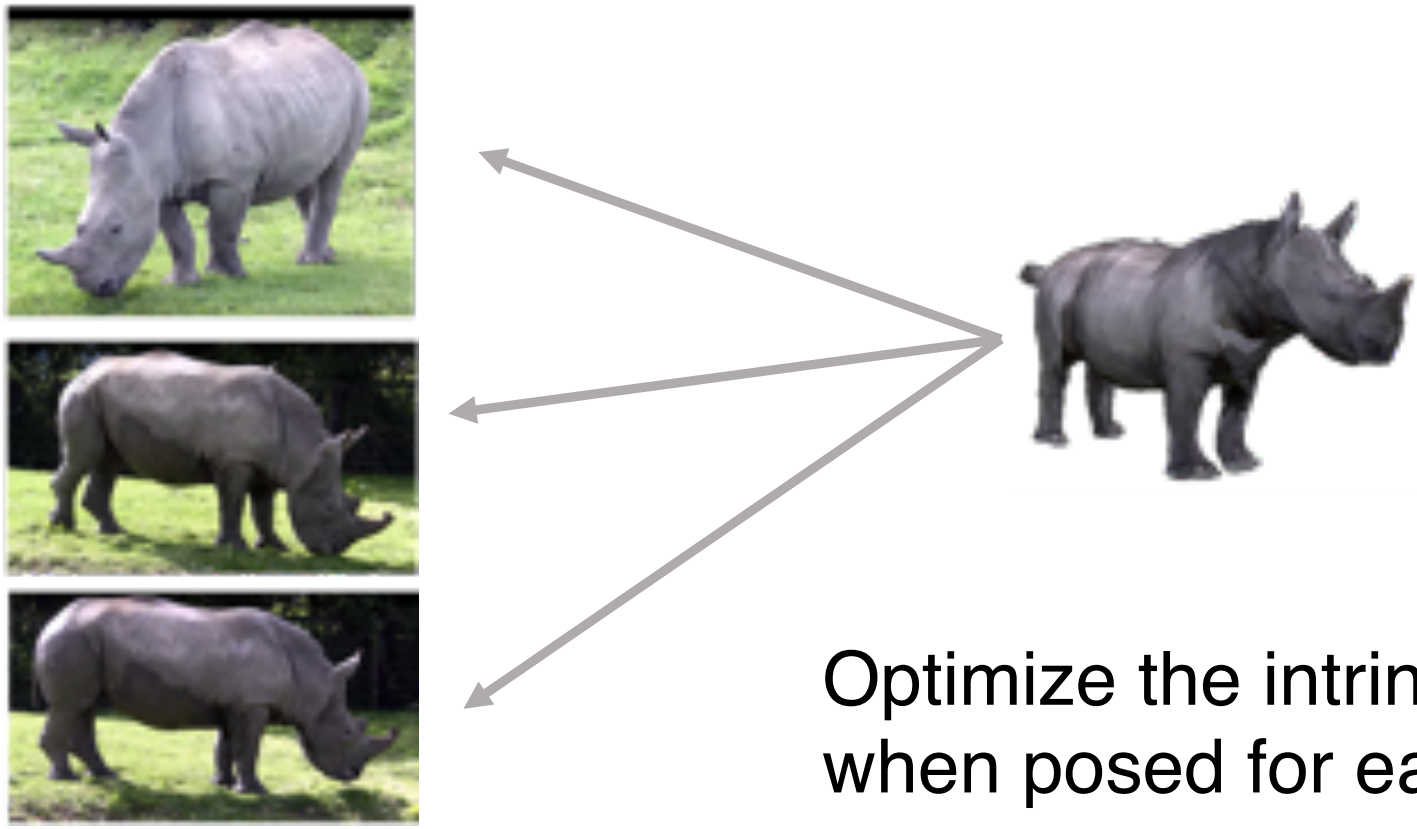


Benjamin Biggs, Thomas Roddick, Andrew Fitzgibbon, Roberto Cipolla ACCV 2018

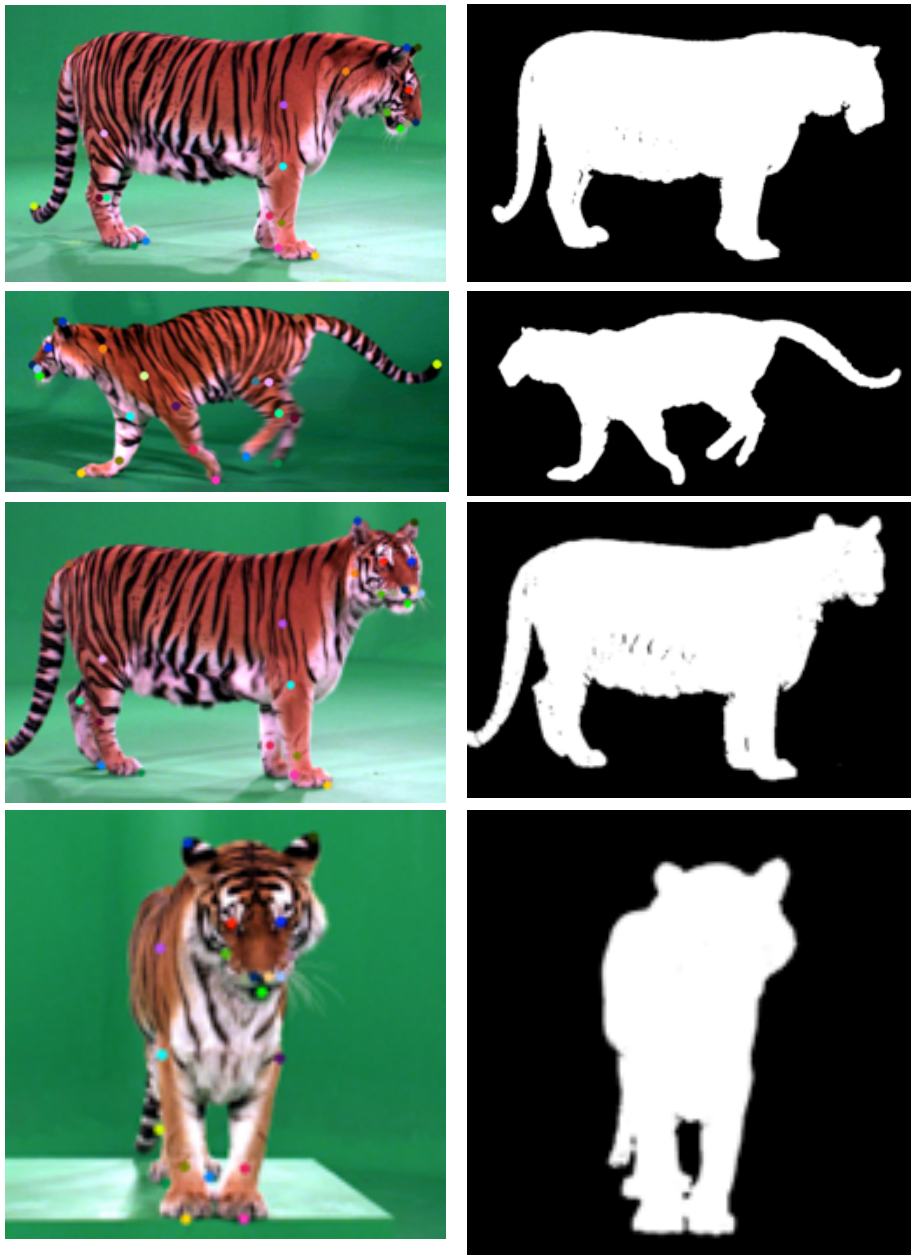
Problem: Not every animal is a toy!

- A very low-frequency approximation to quadrupeds
- How to get the horn of the rhinos?

Key idea: Animals deform, but they have a consistent underlying shape



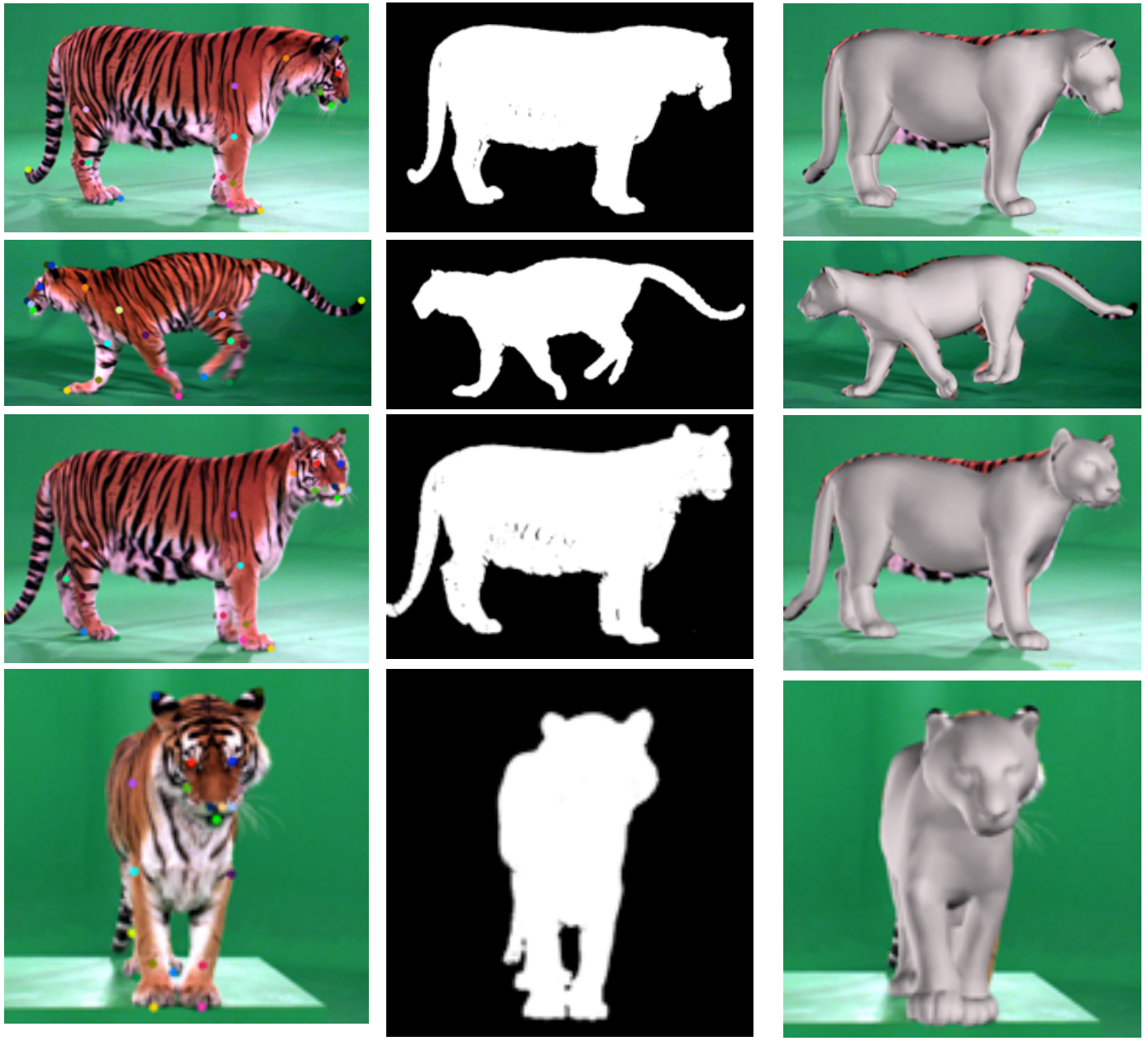
Optimize the intrinsic shape so that, when posed for each image, it explains all views.



Input: 2D key points and silhouettes

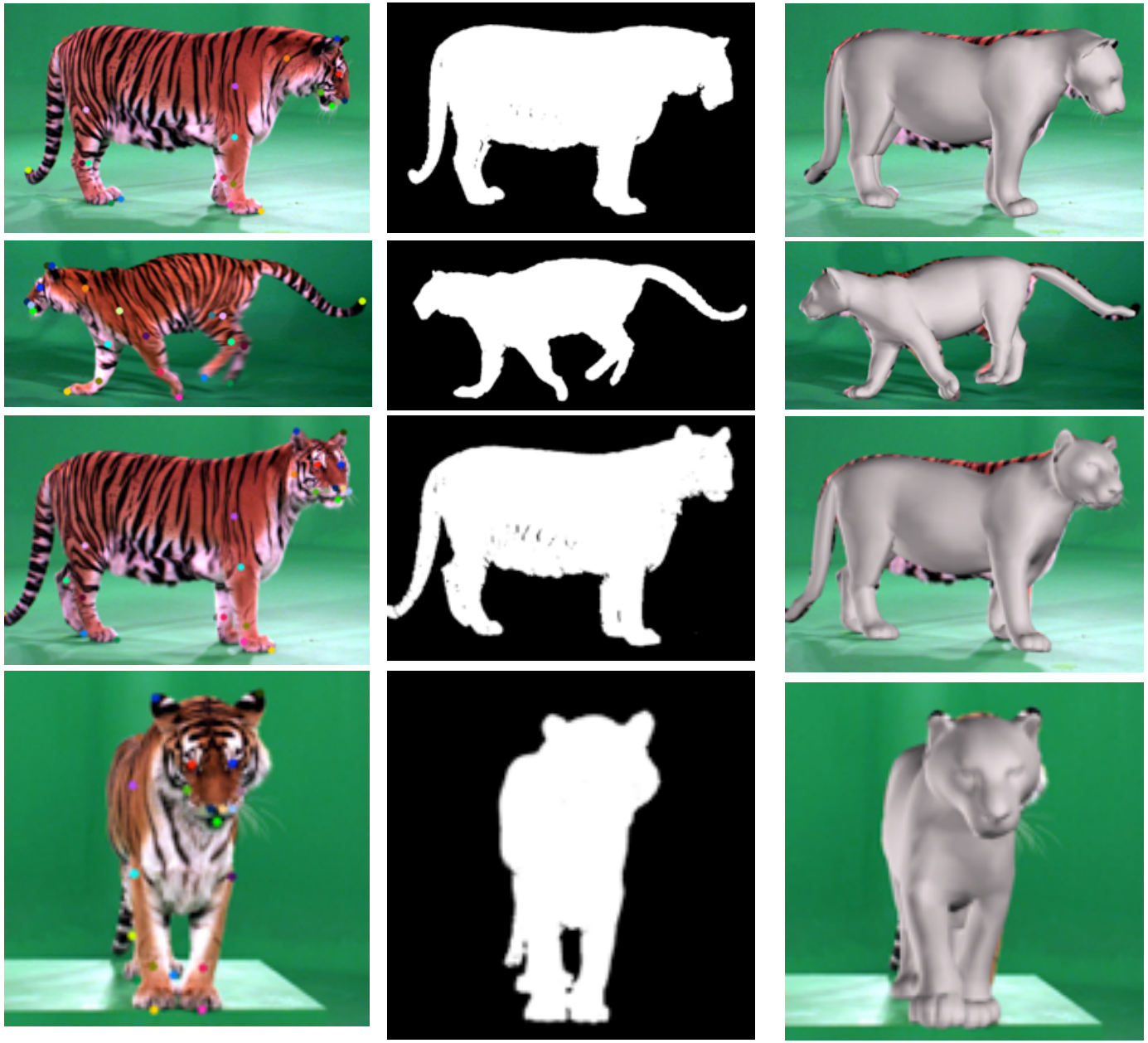
Note that camera view and body pose change.

Not classic multi-view capture



Fit SMAL to the data.

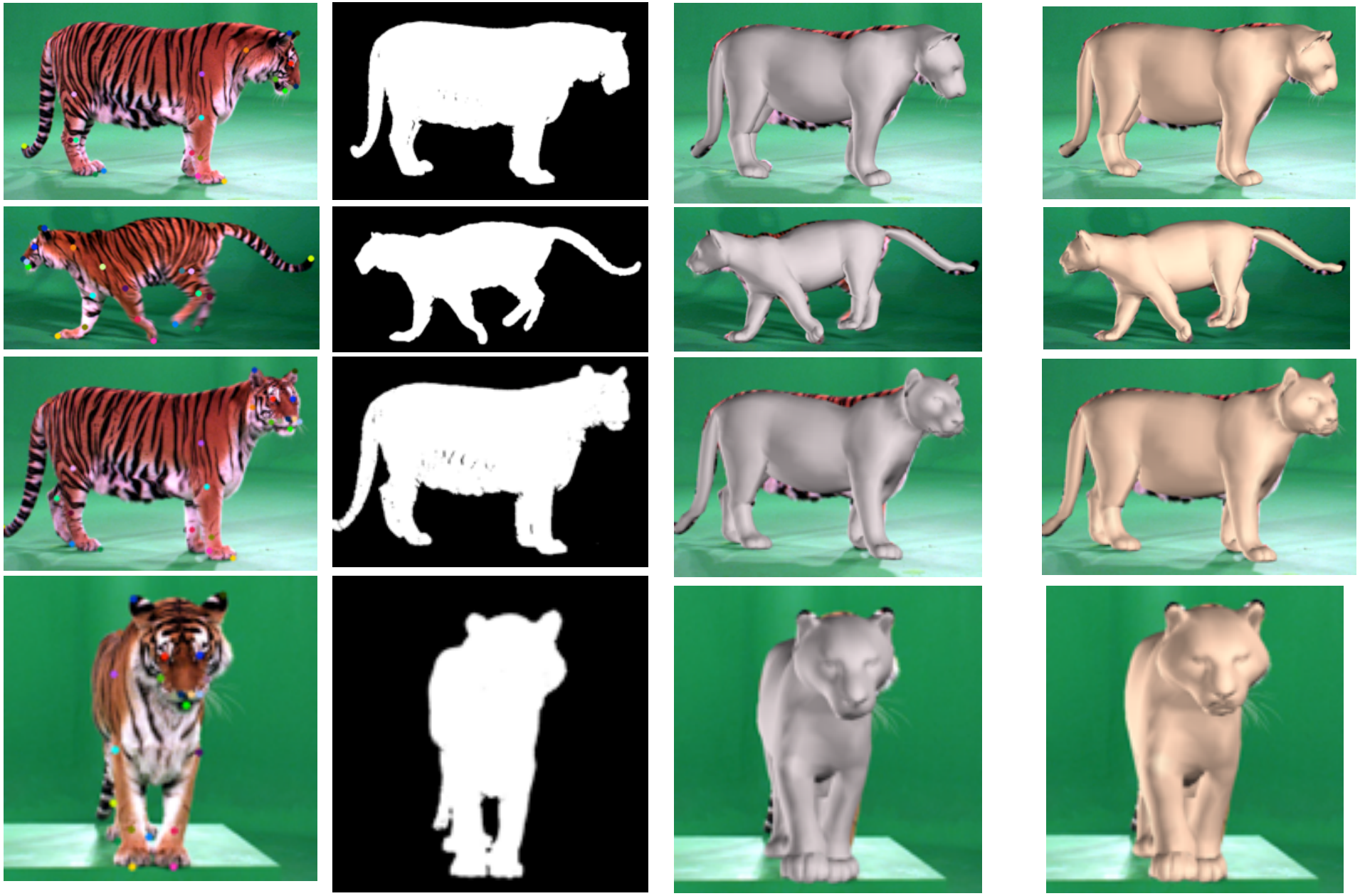
Result is generic and does not capture the individual detail.

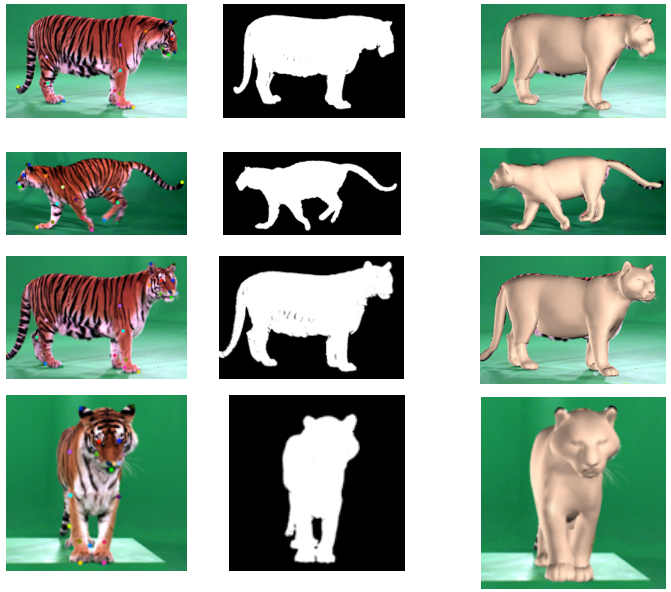


Idea: Optimize the intrinsic shape so that, when posed, it explains all views.

Key: **regularization.**

SMAL with Refinement = SMALR





UV texture map

Example

SMAL fit

SMALR

Texture sampled from SMALR shape










More results

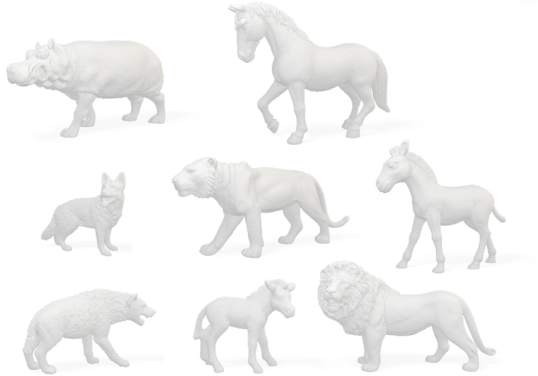
The same concept holds for humans, explored in [Alldieck et al. CVPR 2018]

3D prints

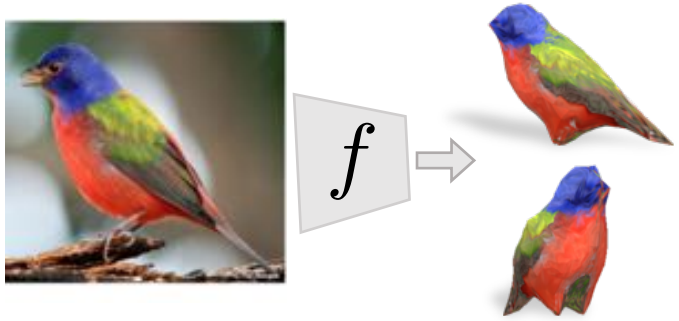
Overview of 3D Animal Reconstruction

1. Let's start with a template 3D model + images

2. What if we had some 3D data?



3. What if we don't have any 3D data?



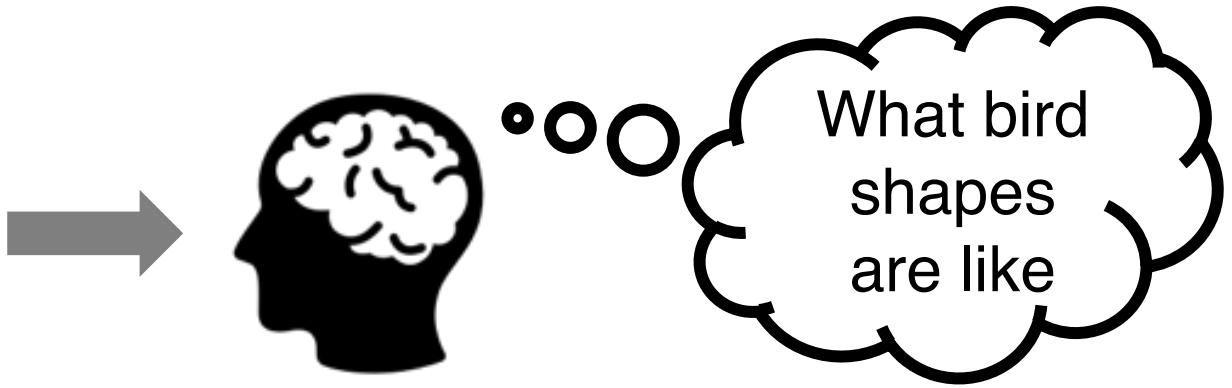
More generality: what if there's no scans?

Goal:

- **Learn a 3D deformable shape model from 2D information**
- **Without any ground truth 3D data**

[**Kanazawa** and Tulsiani et al. ECCV 2018]

Training



Testing

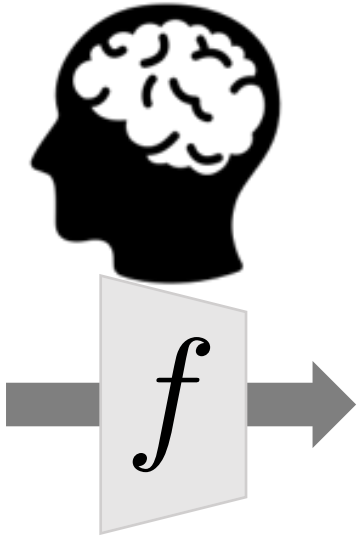


Image Formation

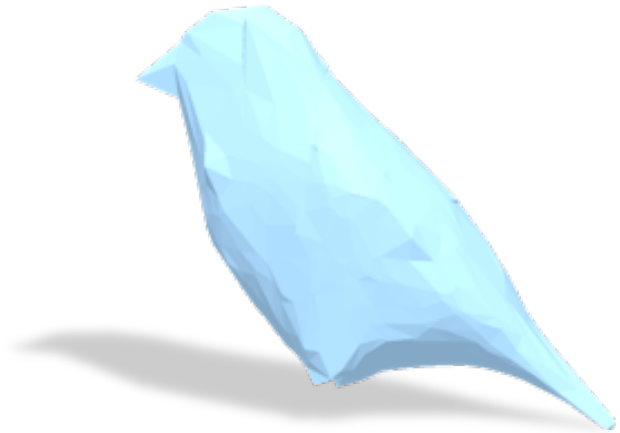
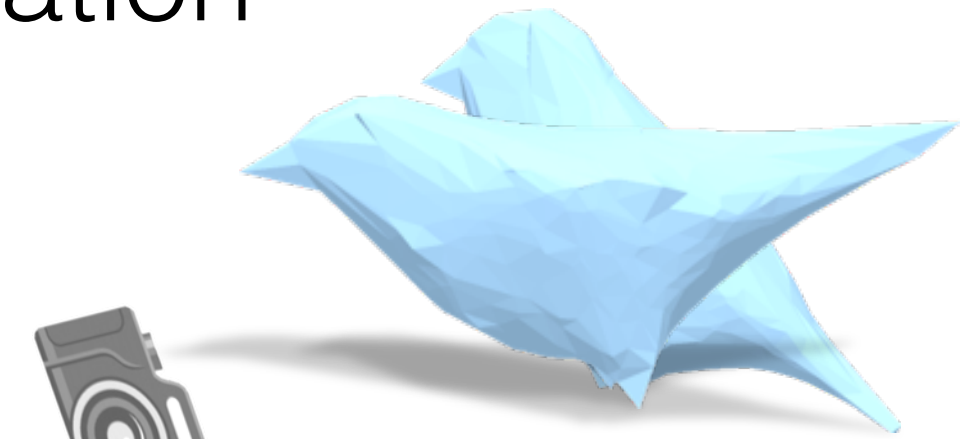
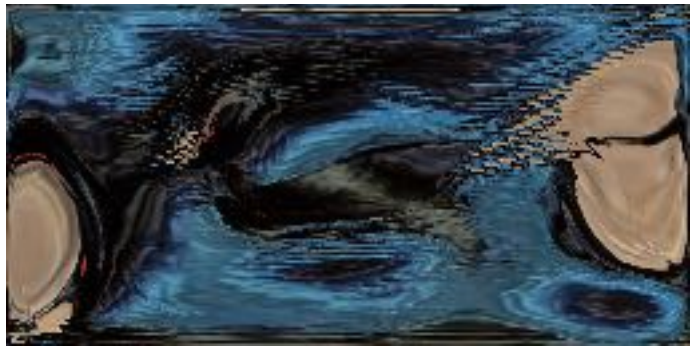


Image Formation

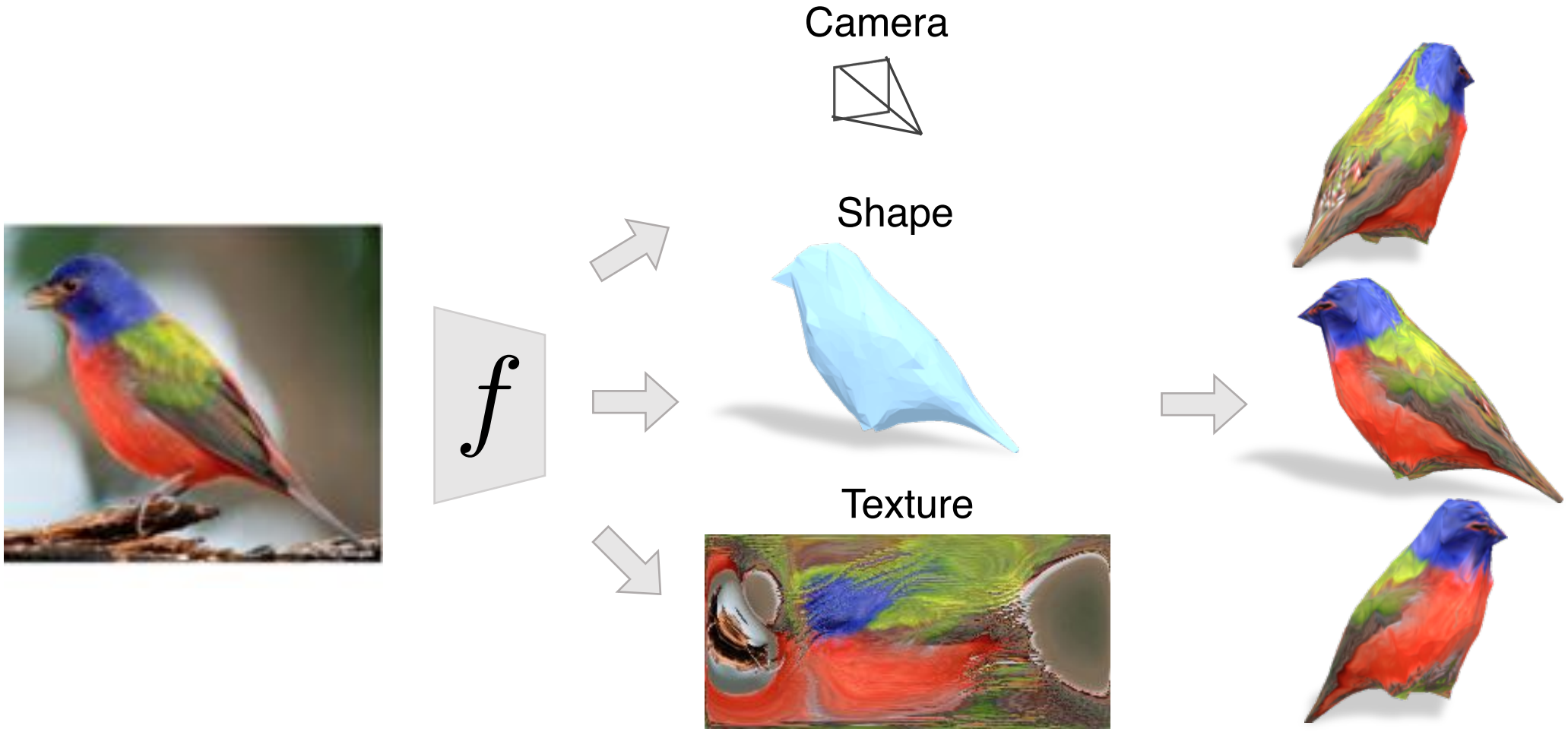
Image Formation



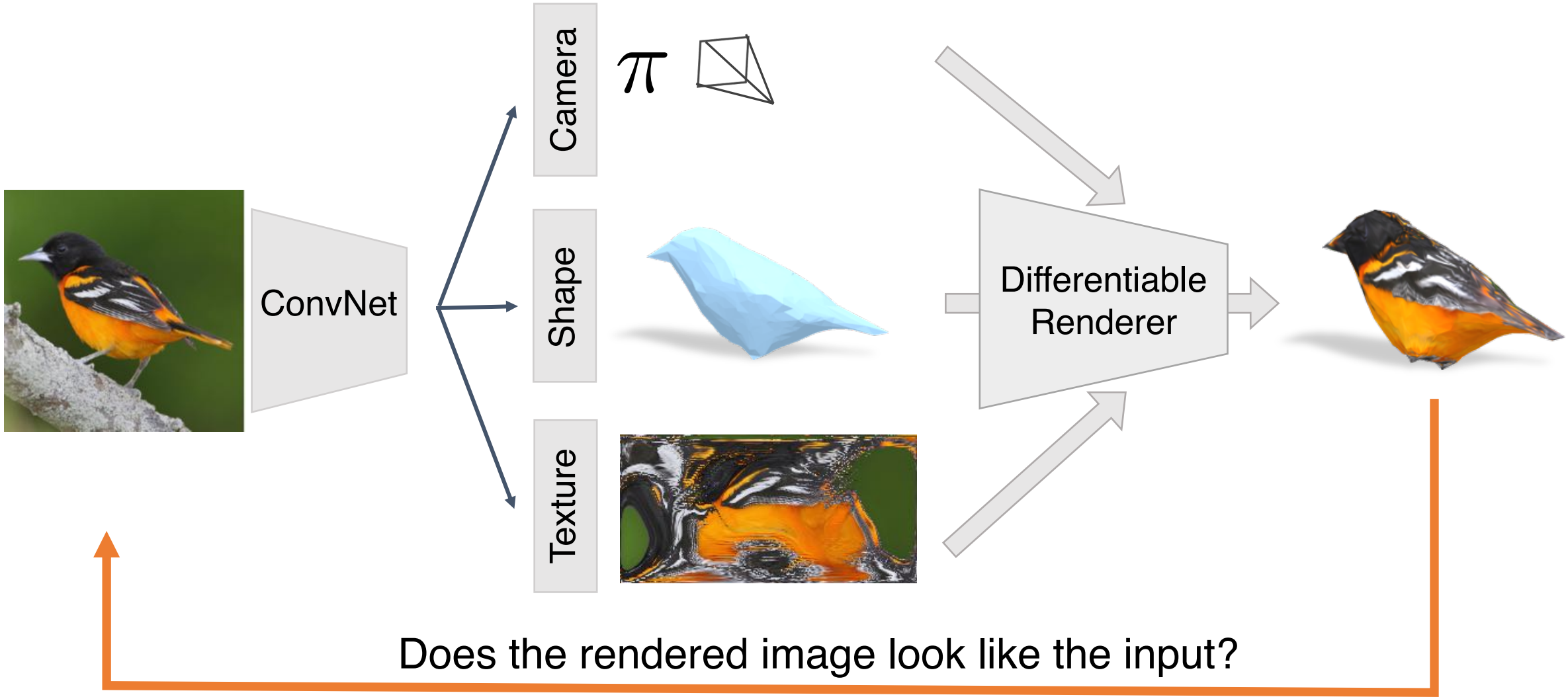


Analysis-by-synthesis

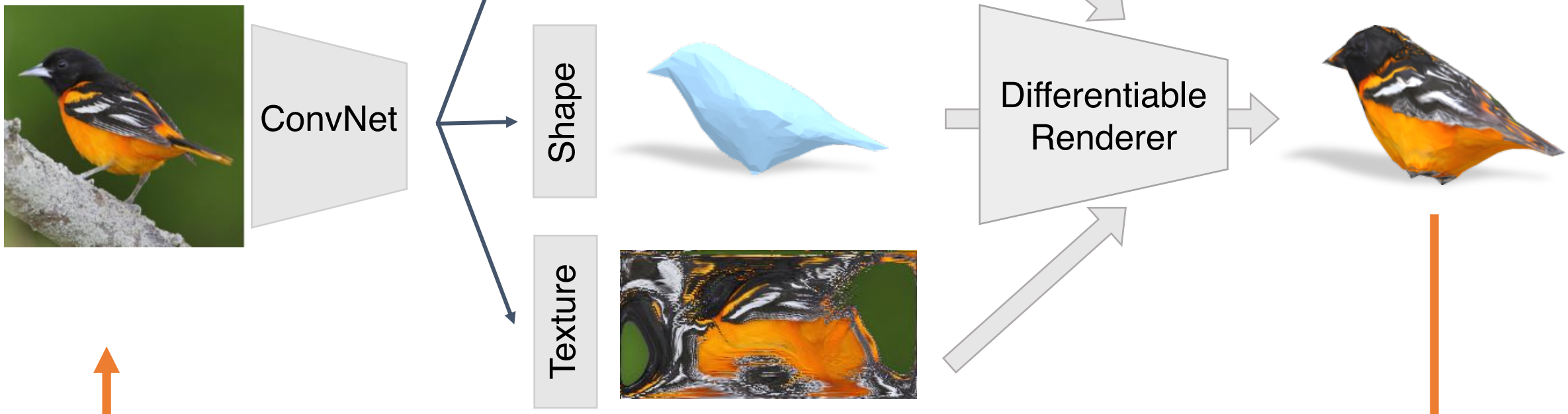
Idea: disentangle the image into 3D factors



Approach


ConvNet
Camera
Shape
Texture
Differentiable Renderer
Does the rendered image look like the input?

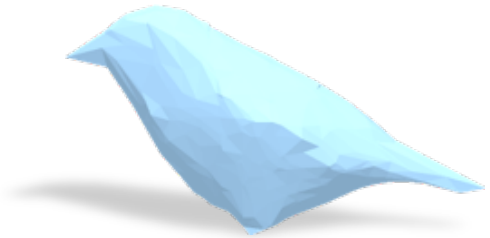
Approach



Does the rendered bird look like the input bird?

3D Morphable Model

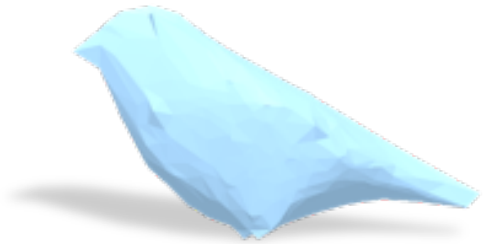
Predicted Shape



$V \in \mathbb{R}^{N \times 3}$

=

Learned Mean Shape

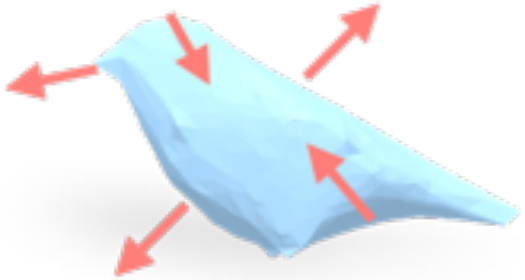


$\bar{V} \in \mathbb{R}^{N \times 3}$

Shared across-category

+

Shape Deformation



$\Delta V \in \mathbb{R}^{N \times 3}$

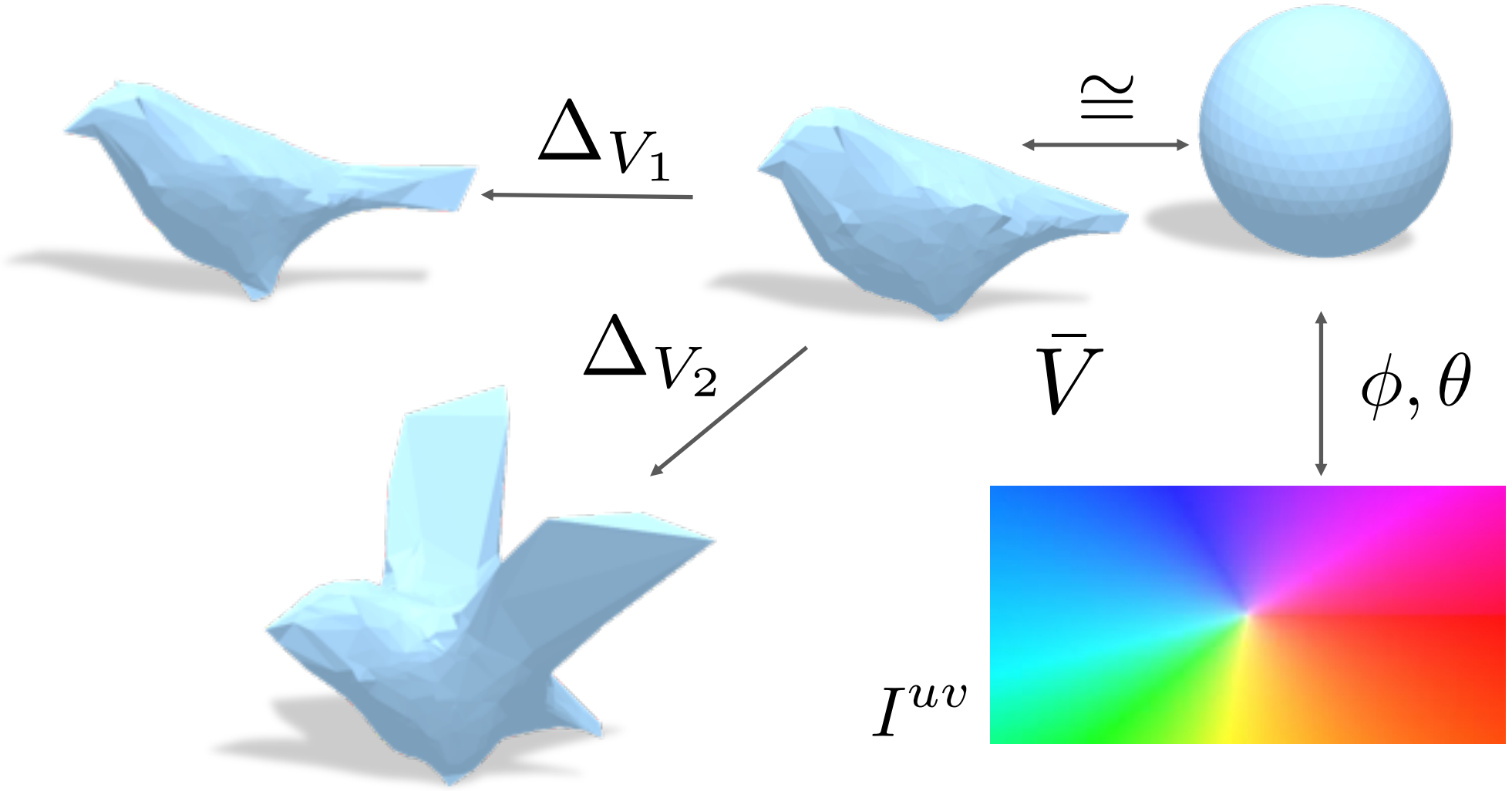
Instance specific

Reminiscent of early 3D morphable models [Blanz & Vetter ‘99], but learned without any 3D data

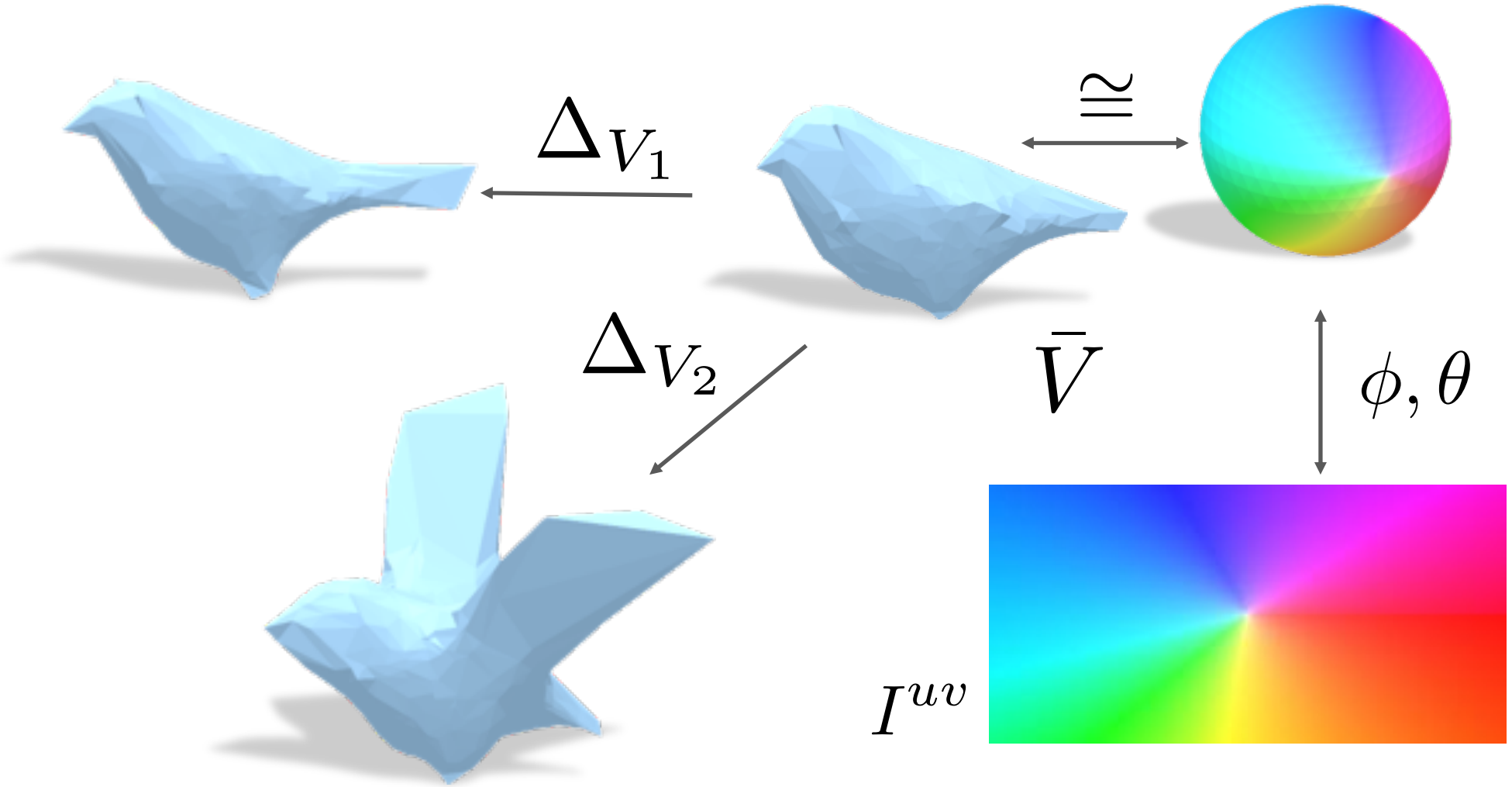
Texturing

UV Image

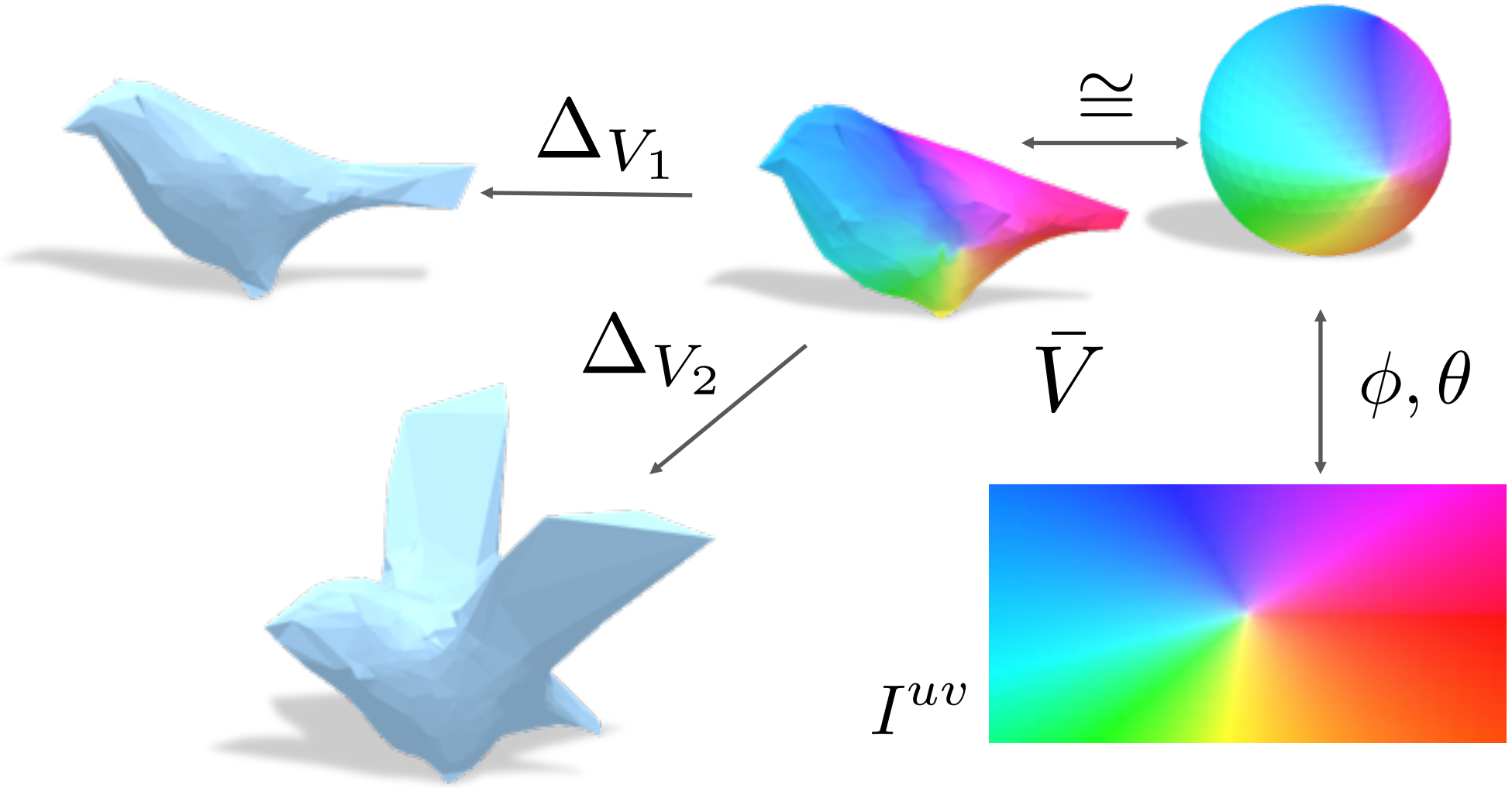
Texture Representation



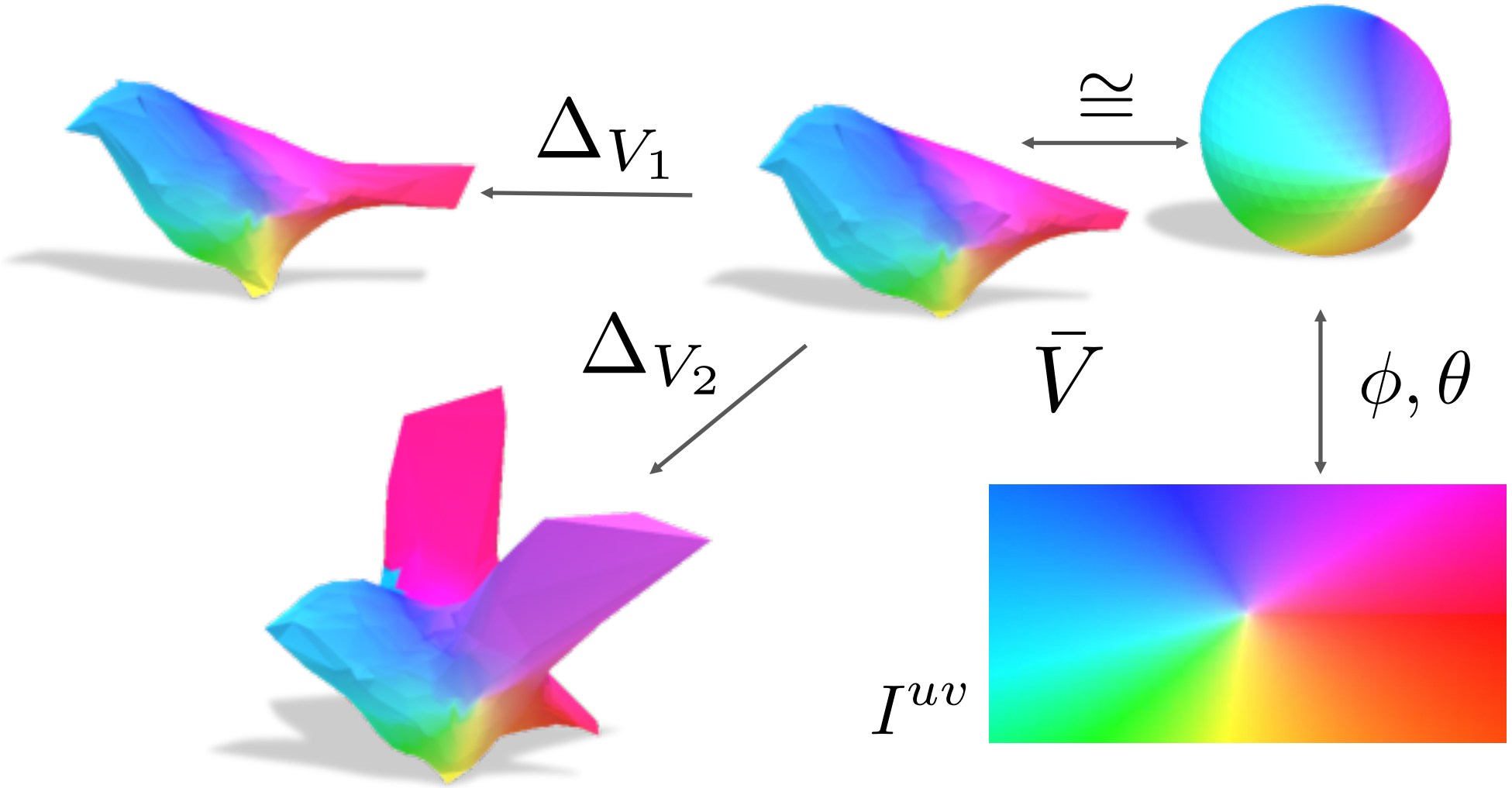
Texture Representation



Texture Representation



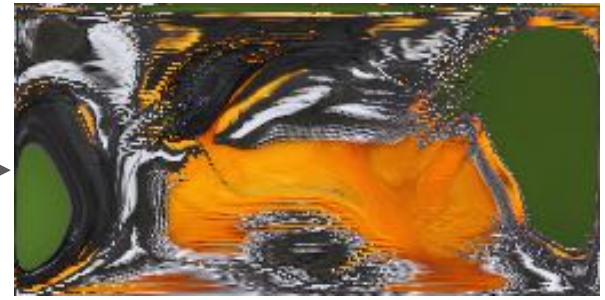
Texture Representation



Texture as UV Image Prediction

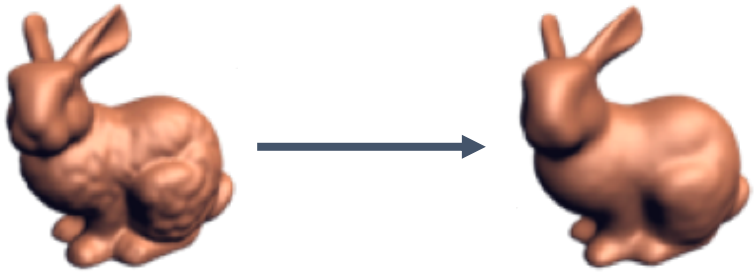
ConvNet

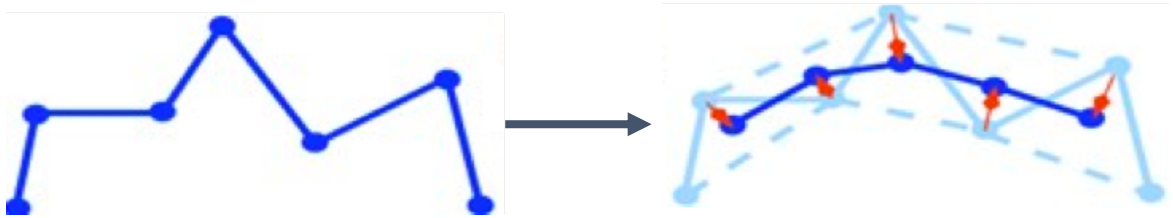
UV Image



Geometric priors

Laplacian Smoothness:

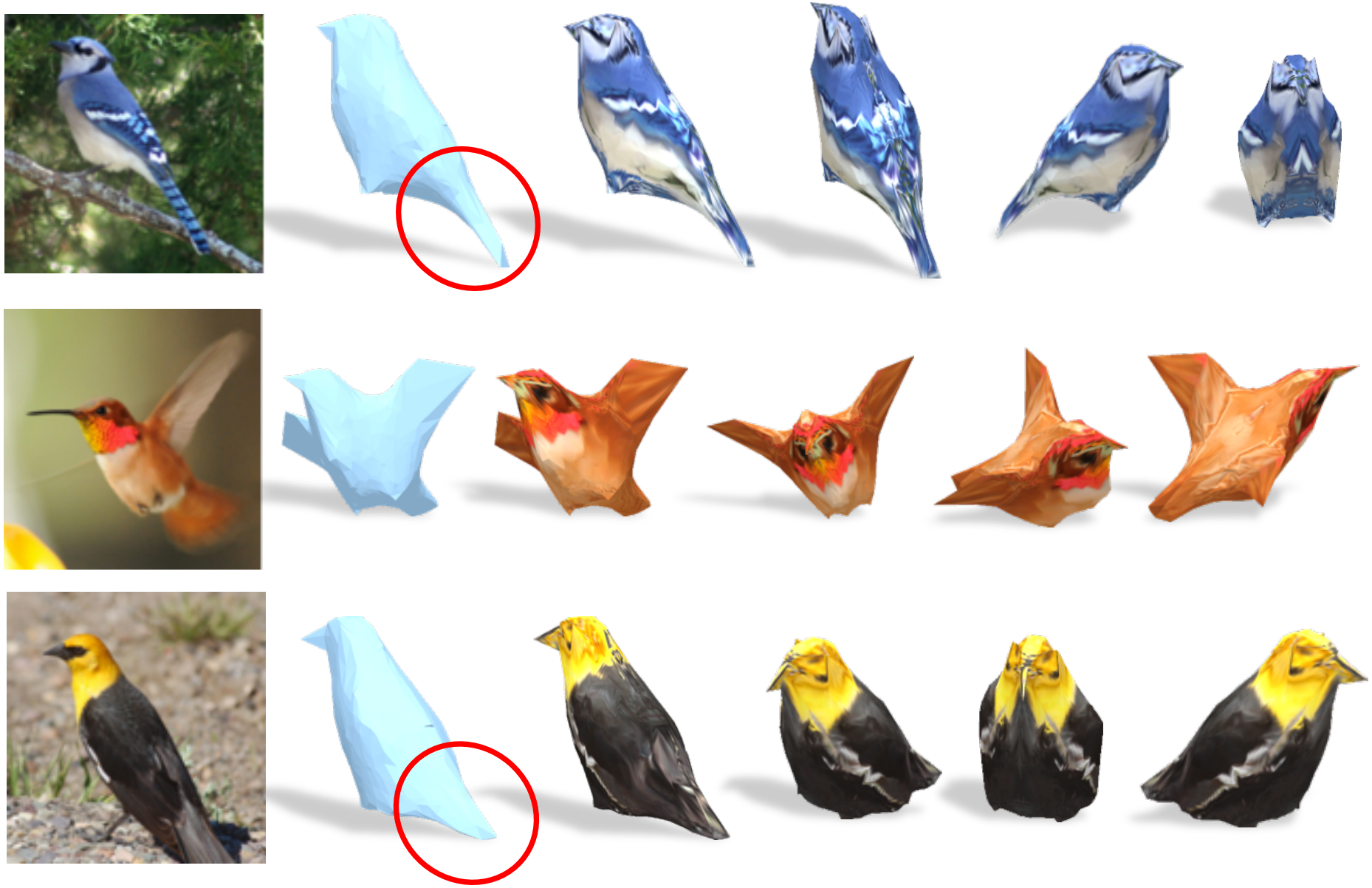




Minimize mean curvature

Bilateral symmetry in vertices & faces:

Results



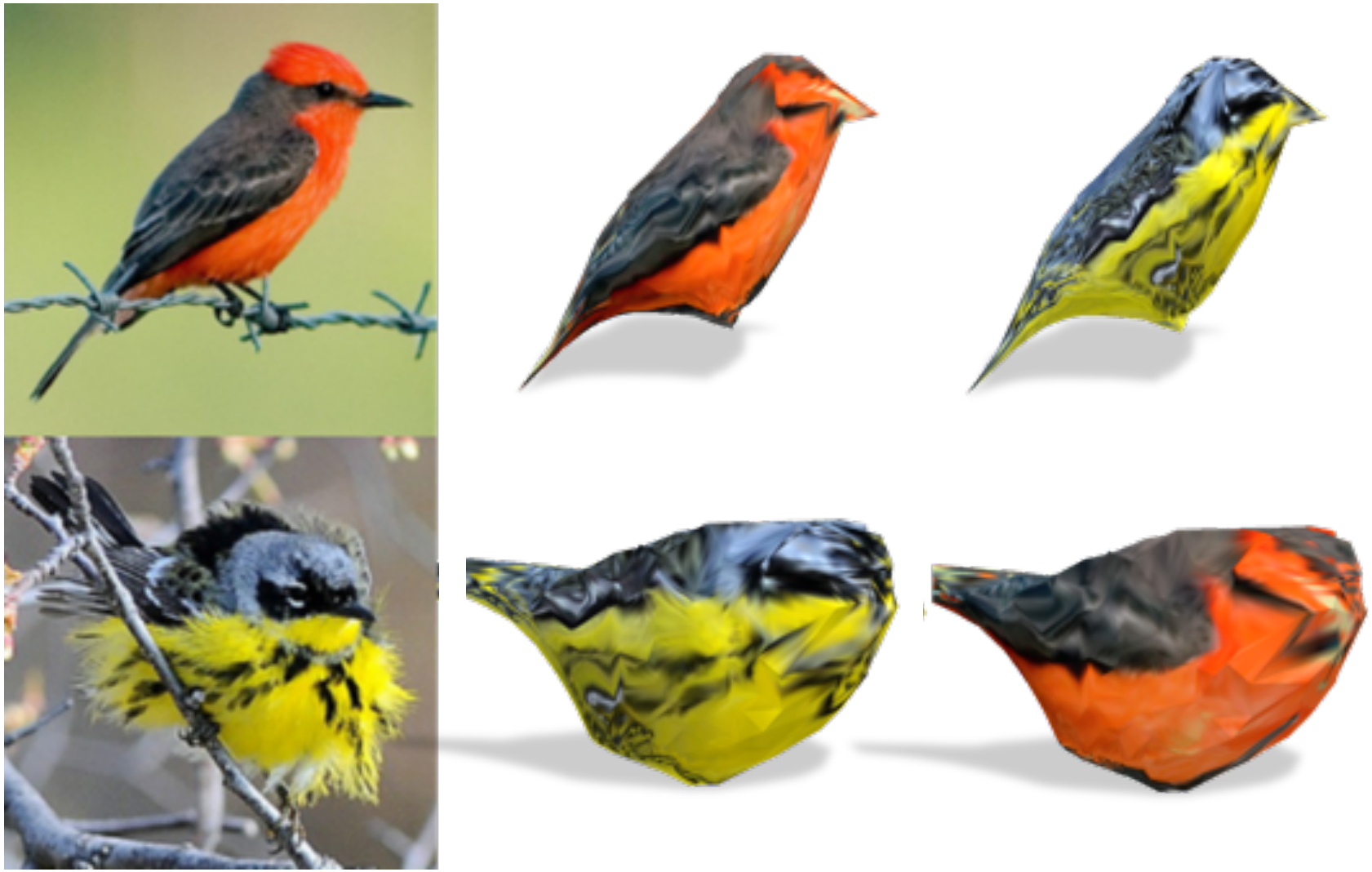
Other objects

Reconstruction evaluation on PASCAL 3D+ (IOU ↑)

Method	Aeroplane	Car
CSDM [11]	0.40	0.60
DRC [24]	0.42	0.67
Ours	0.46	0.64

Texture Transfer

Texture Transfer

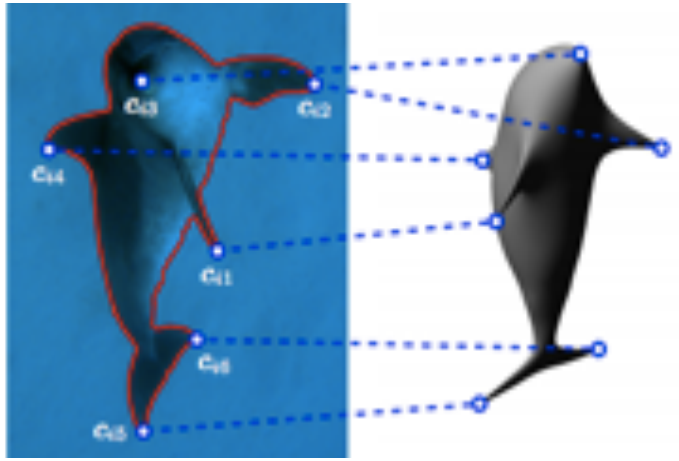


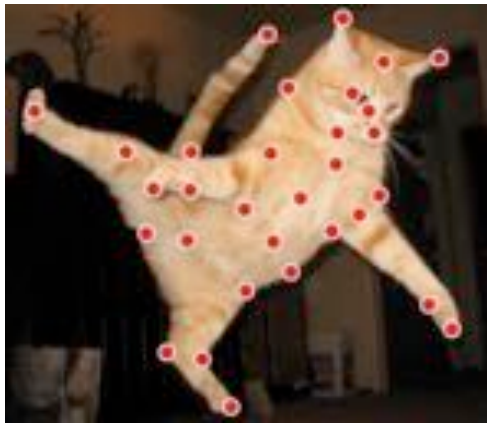
More results

A step towards scalable 3D learning from images

Conclusion:
Animals shed light to interesting problems

- How to model non-rigid objects?
- How to learn this model from limited supervision?





Biggest challenge: Unsupervised correspondence mining