Unsupervised Learning of Probably Symmetric Deformable 3D Objects from Images in the Wild

Shangzhe Wu Christian Rupprecht Andrea Vedaldi

VISUAL GEOMETRY GROUP, UNIVERSITY OF OXFORD

UNIVERSITY OF
OXFORD

Agenda

- Problem Introduction
- Method Overview
- Results
- Discussions
- Conclusions

What is 3D Reconstruction?

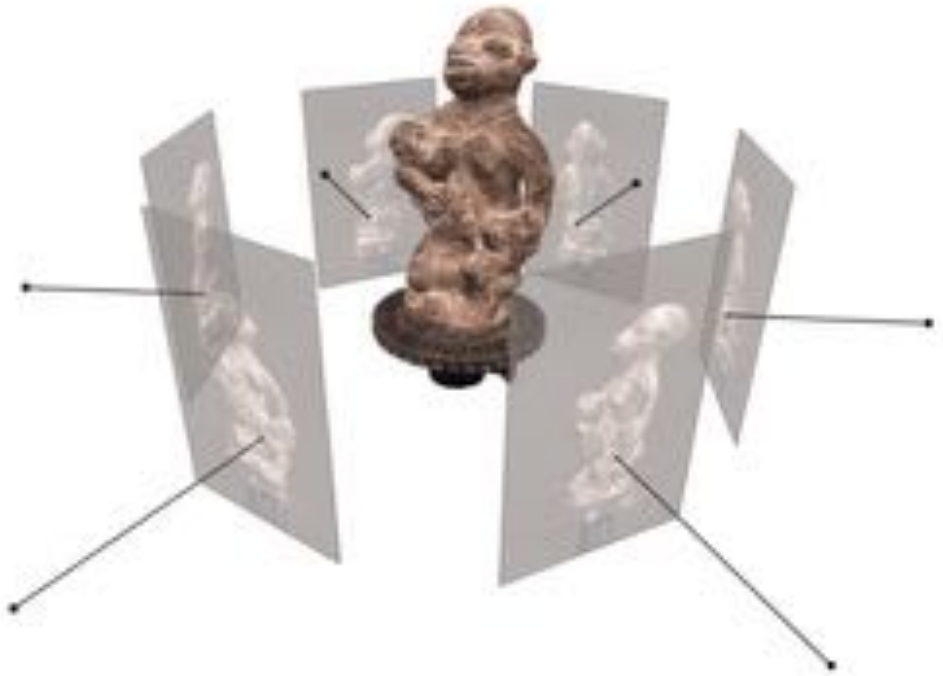
2D Observations

Vision

Reconstruction

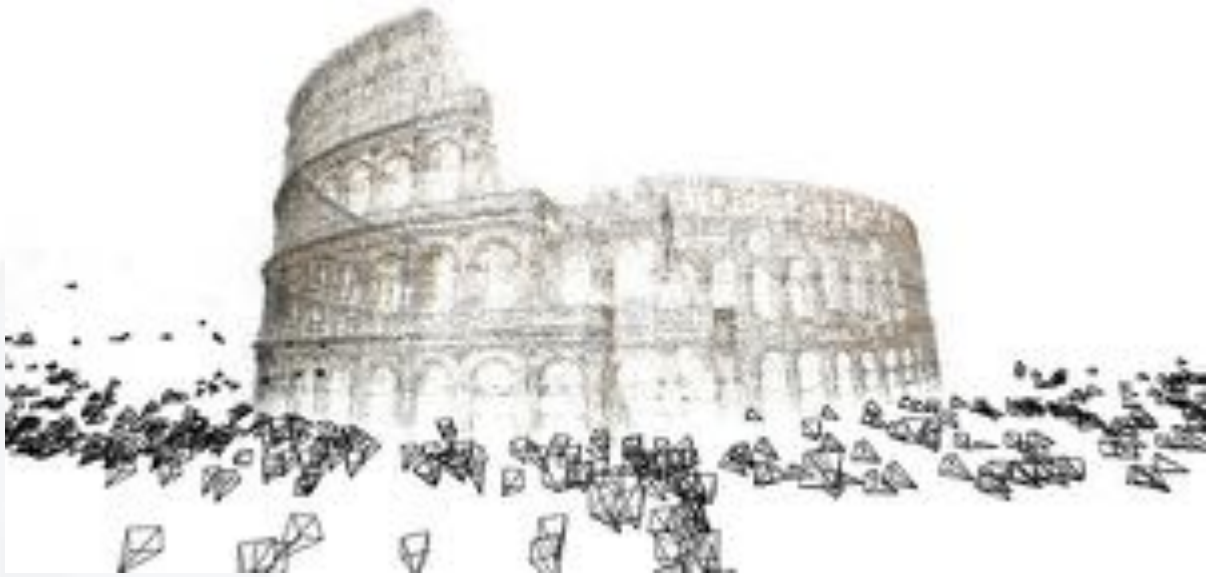
Rendering

Graphics



3D Representation

Multi-view 3D Reconstruction

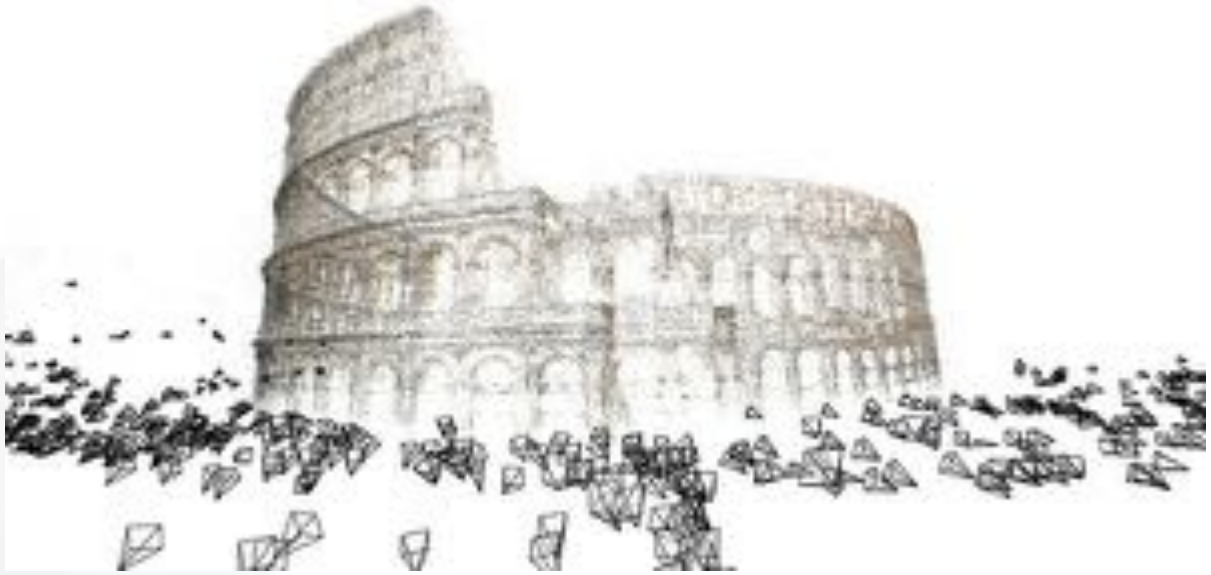


Building Rome in a Day.
Agarwal et al. ICCV'09

static scene

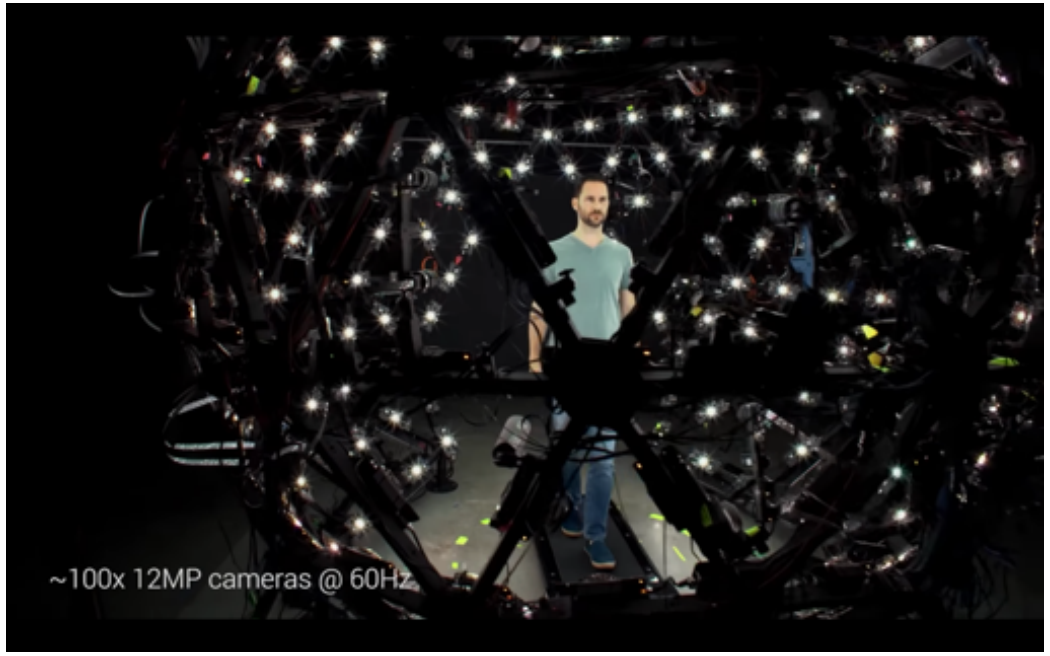
but.. the world is dynamic

Multi-view 3D Reconstruction



Building Rome in a Day.
Agarwal et al. ICCV'09

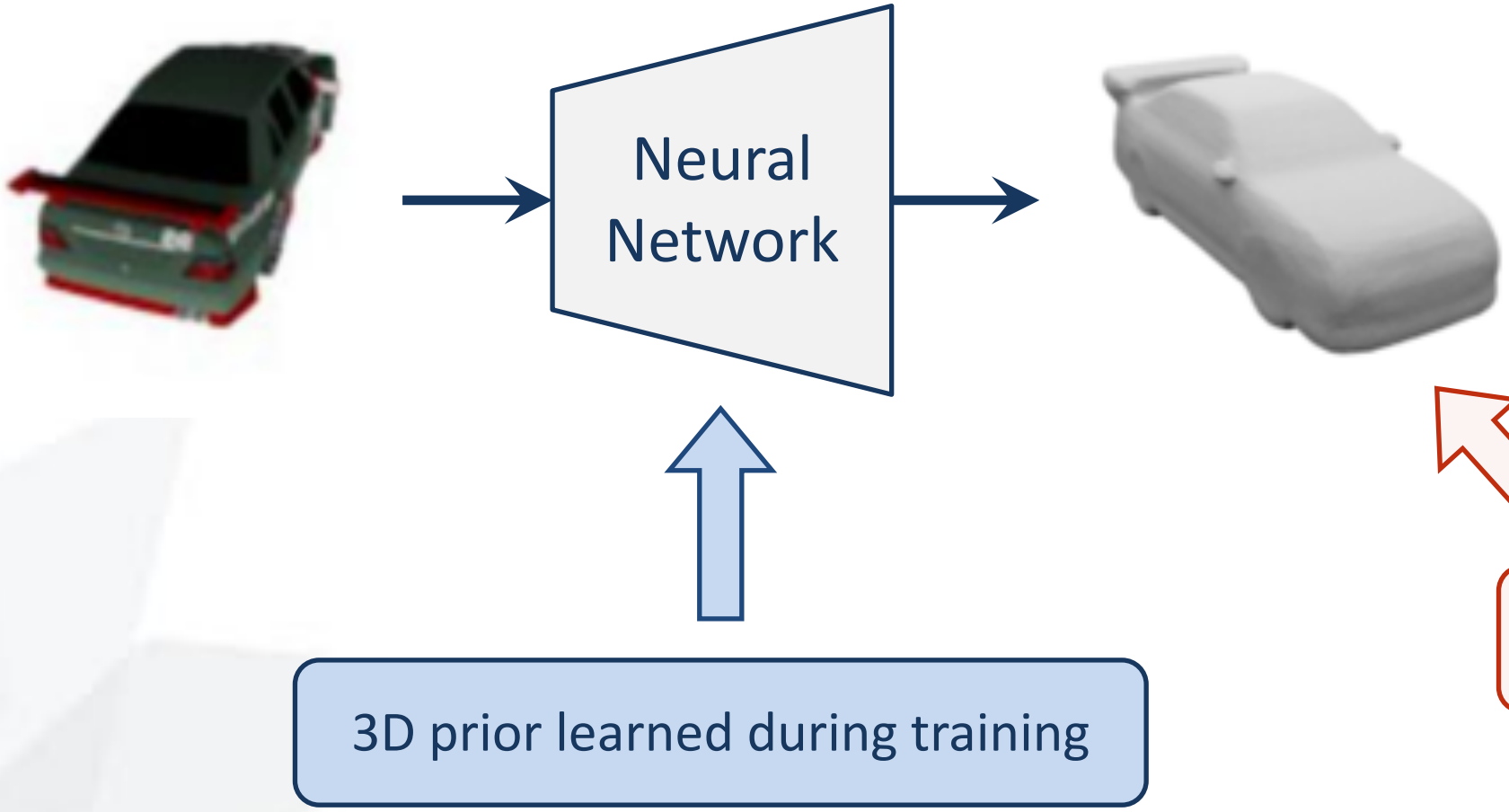
static scene



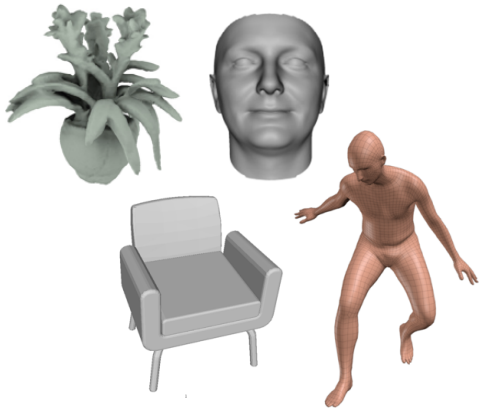
The Relightables: Volumetric Performance Capture of Humans with Realistic Relighting. Guo et al. SIGGRAPH Asia'19

100 cameras
too expensive for me :(

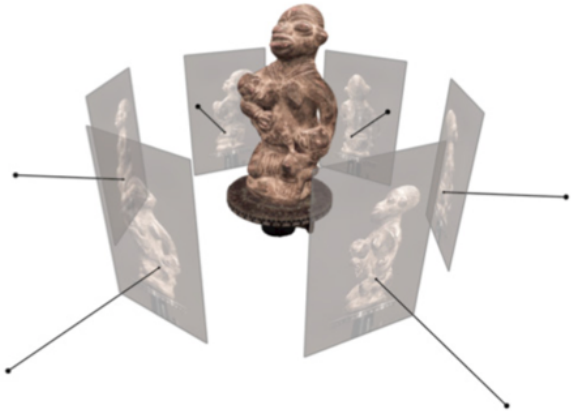
Learning-based Single-view 3D Reconstruction



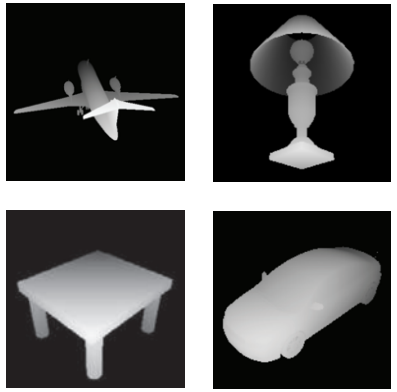
Supervision during Training



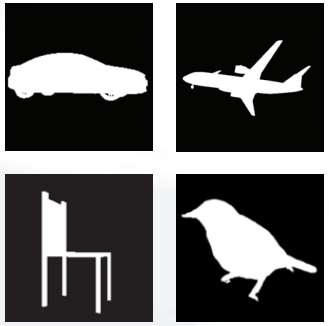
3D ground truth or shape models



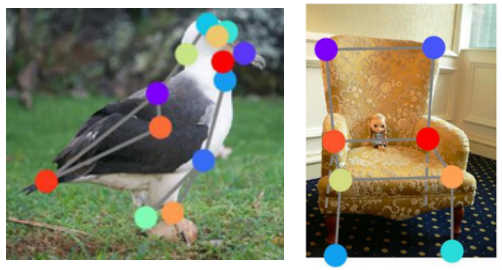
multi-views



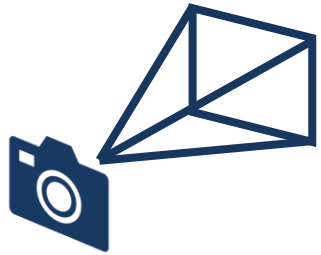
depth maps



silhouettes

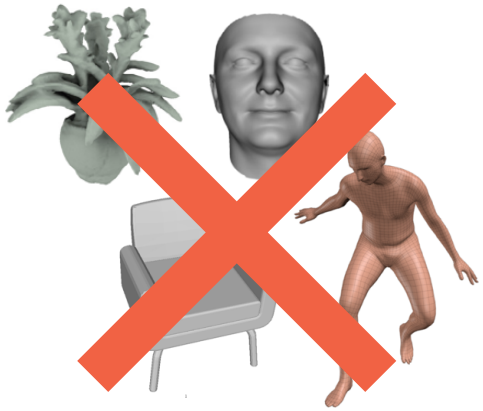


keypoints



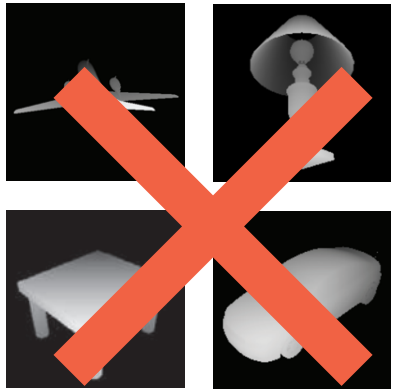
camera viewpoint

Unsupervised Learning of 3D Objects

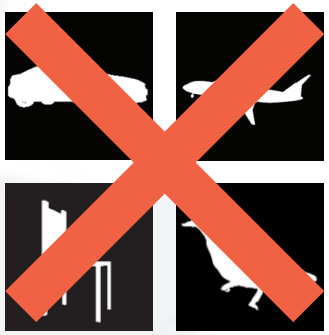


3D ground truth or shape models

multi-views



depth maps



silhouettes

keypoints

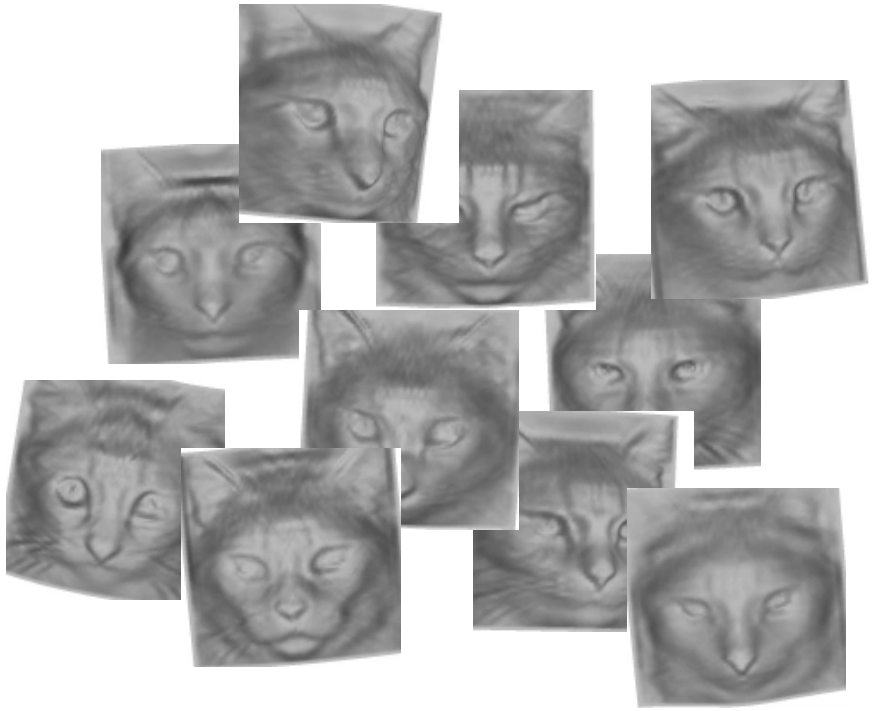
camera viewpoint

Unsupervised Learning of 3D Objects

Training Data

Unsup3D

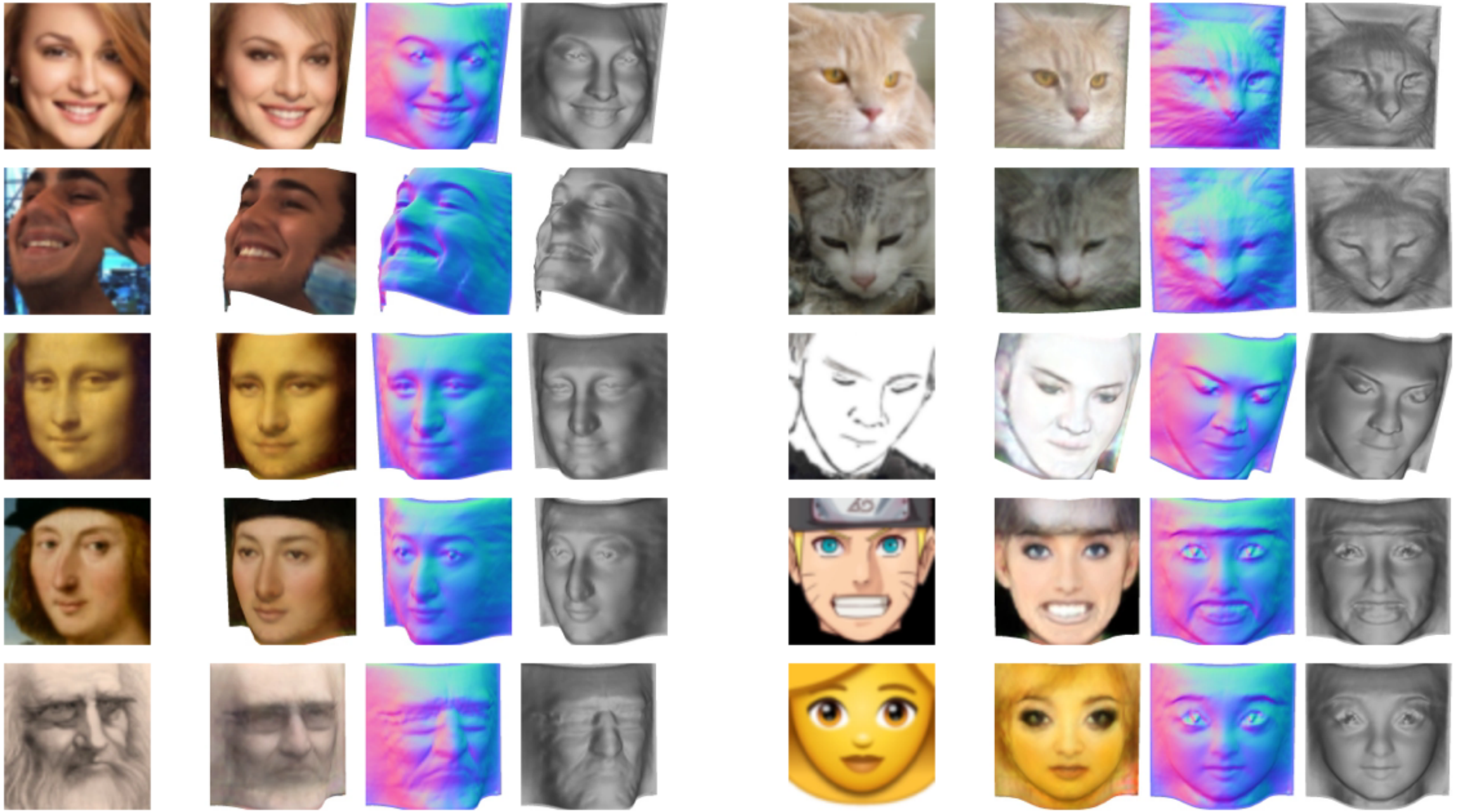
Output



single-view images of a category

NO other supervision!

instance-specific 3D shapes



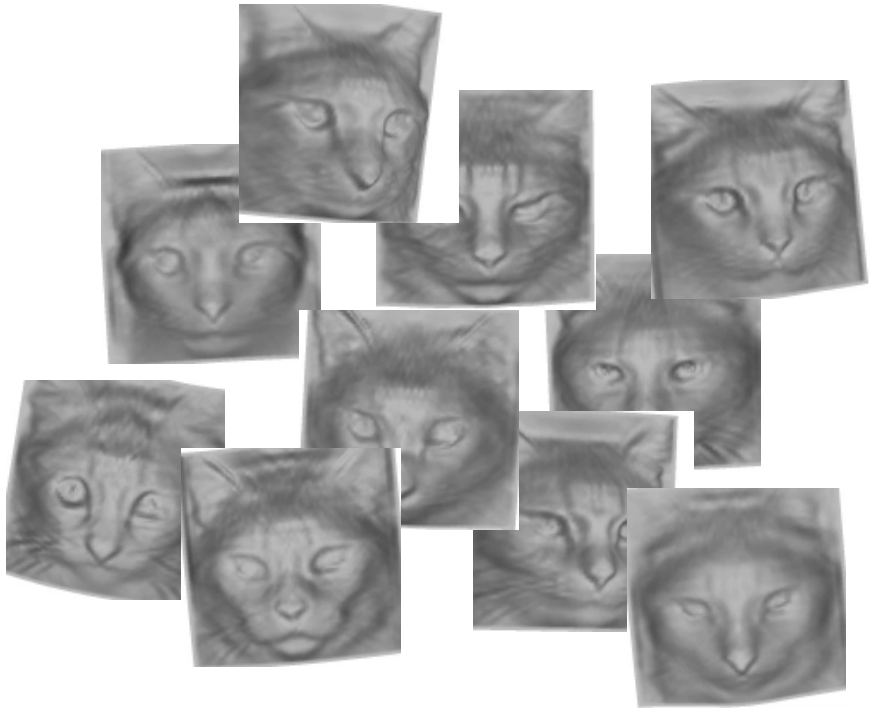
input reconstruction input reconstruction

Unsupervised Learning of 3D Objects

Training Data

Unsup3D

Output



single-view images of a category

NO other supervision!

instance-specific 3D shapes

Symmetries in the World

Training Pipeline:

Photo-Geometric Autoencoding

Photo-Geometric Autoencoding

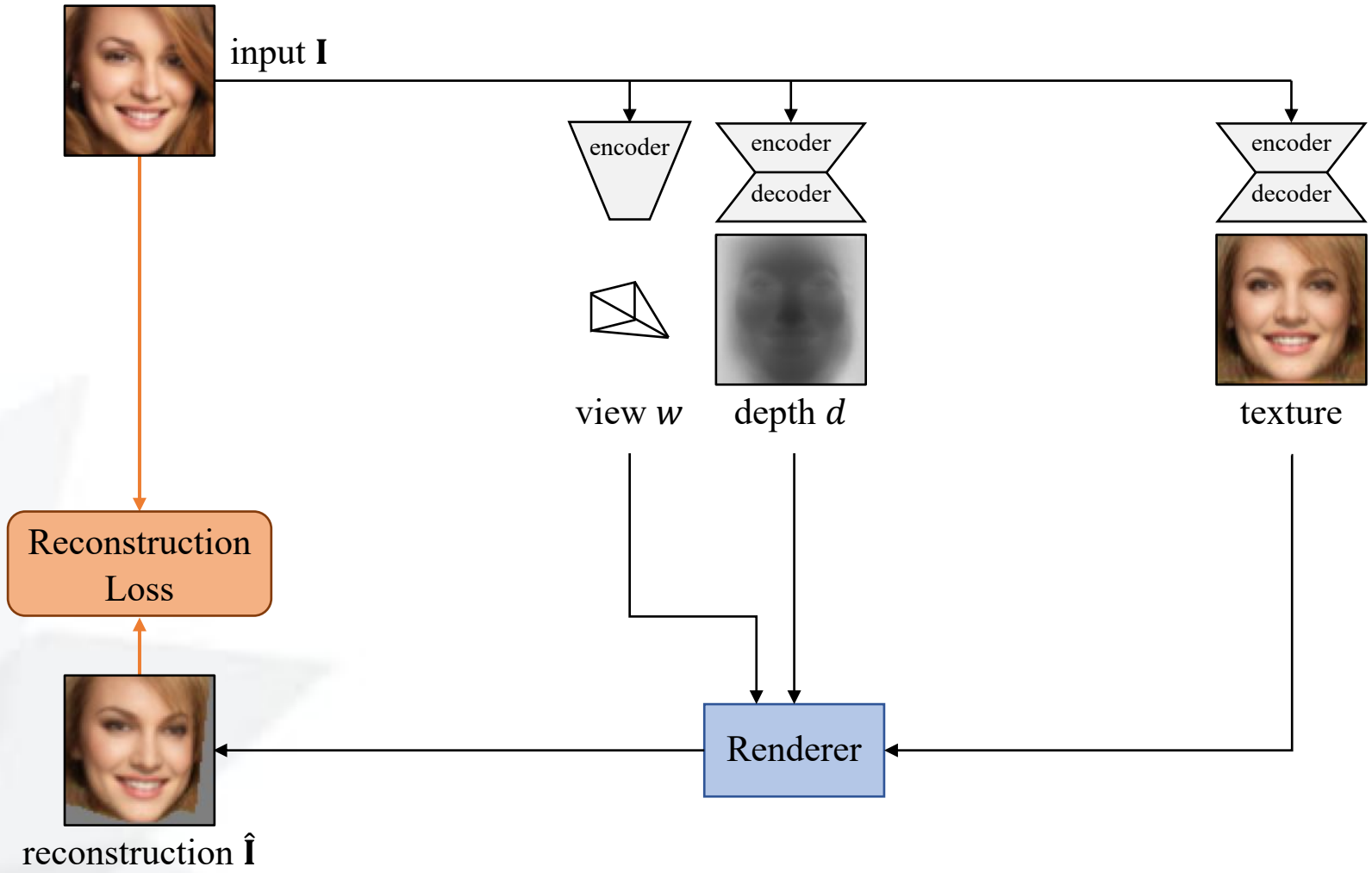


Photo-Geometric Autoencoding

Q1: How to avoid degenerate solutions?

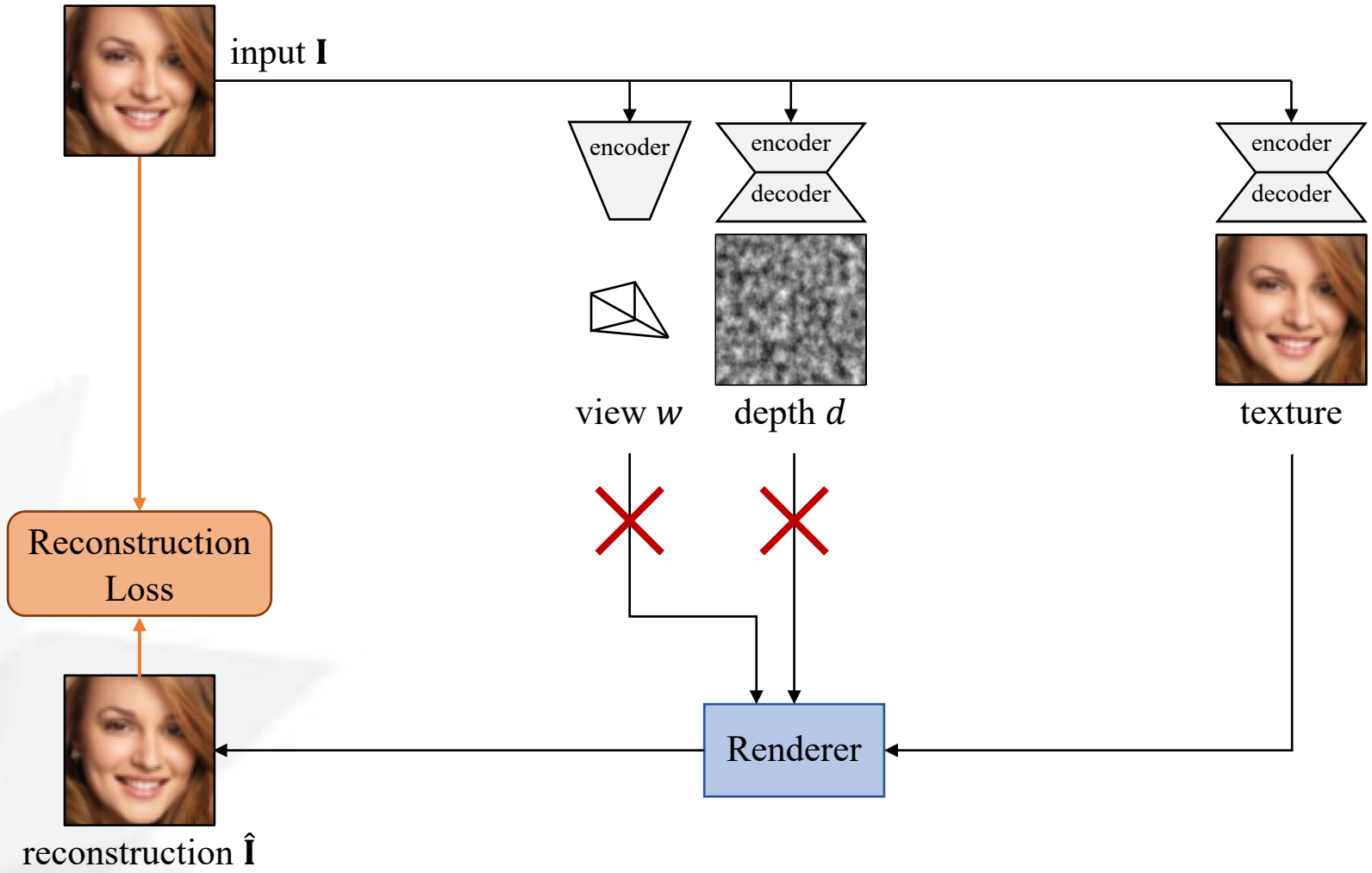


Photo-Geometric Autoencoding

Q1: How to avoid degenerate solutions?

A1: Enforce symmetry

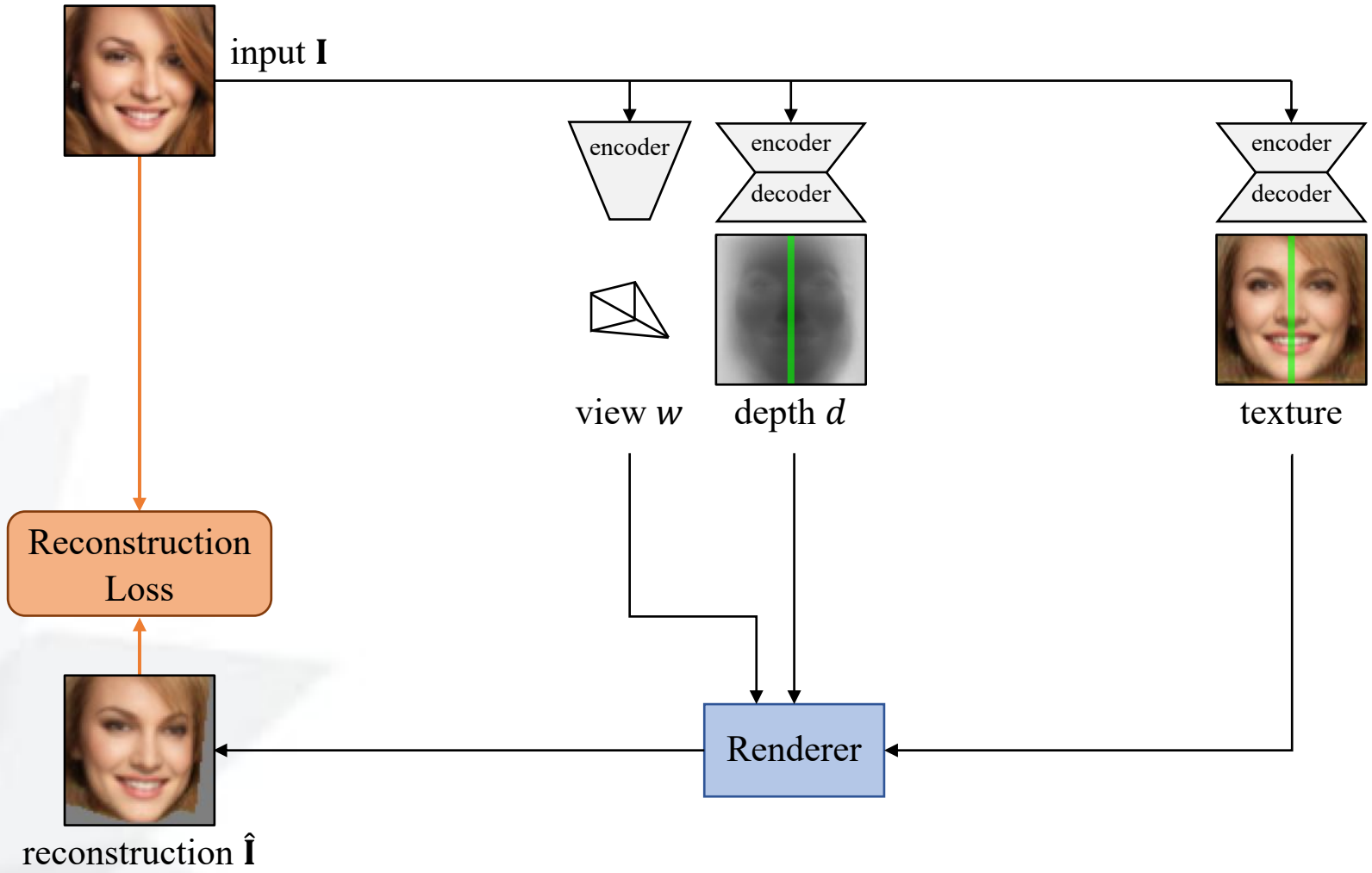


Photo-Geometric Autoencoding

Q1: How to avoid degenerate solutions?

A1: Enforce symmetry by flipping

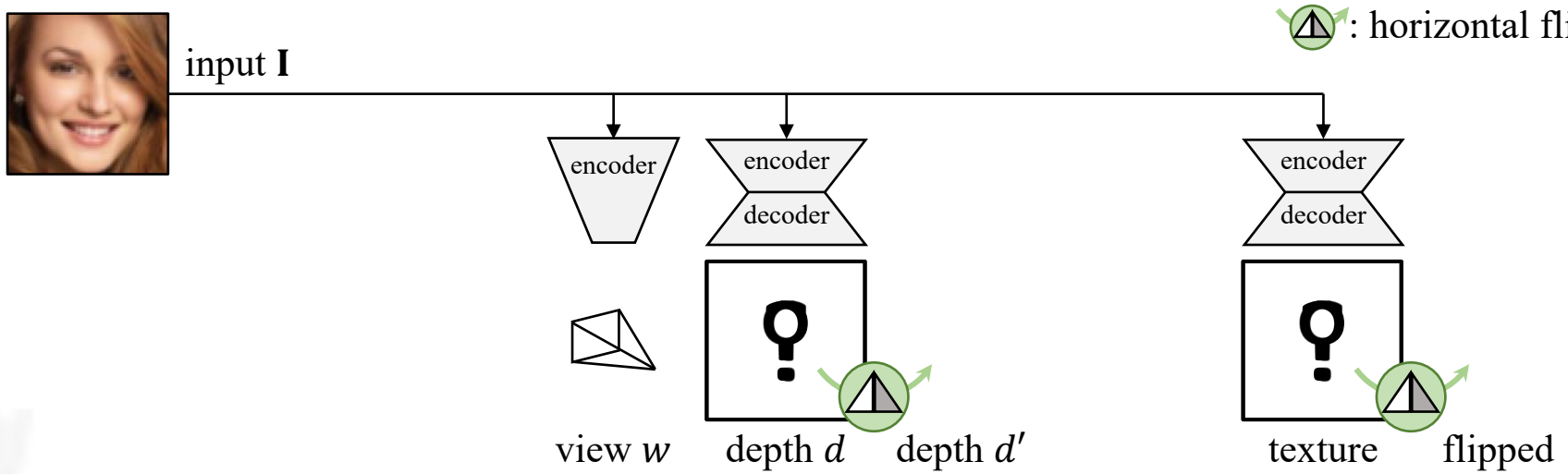


Photo-Geometric Autoencoding

Q1: How to avoid degenerate solutions?

A1: Enforce symmetry by flipping

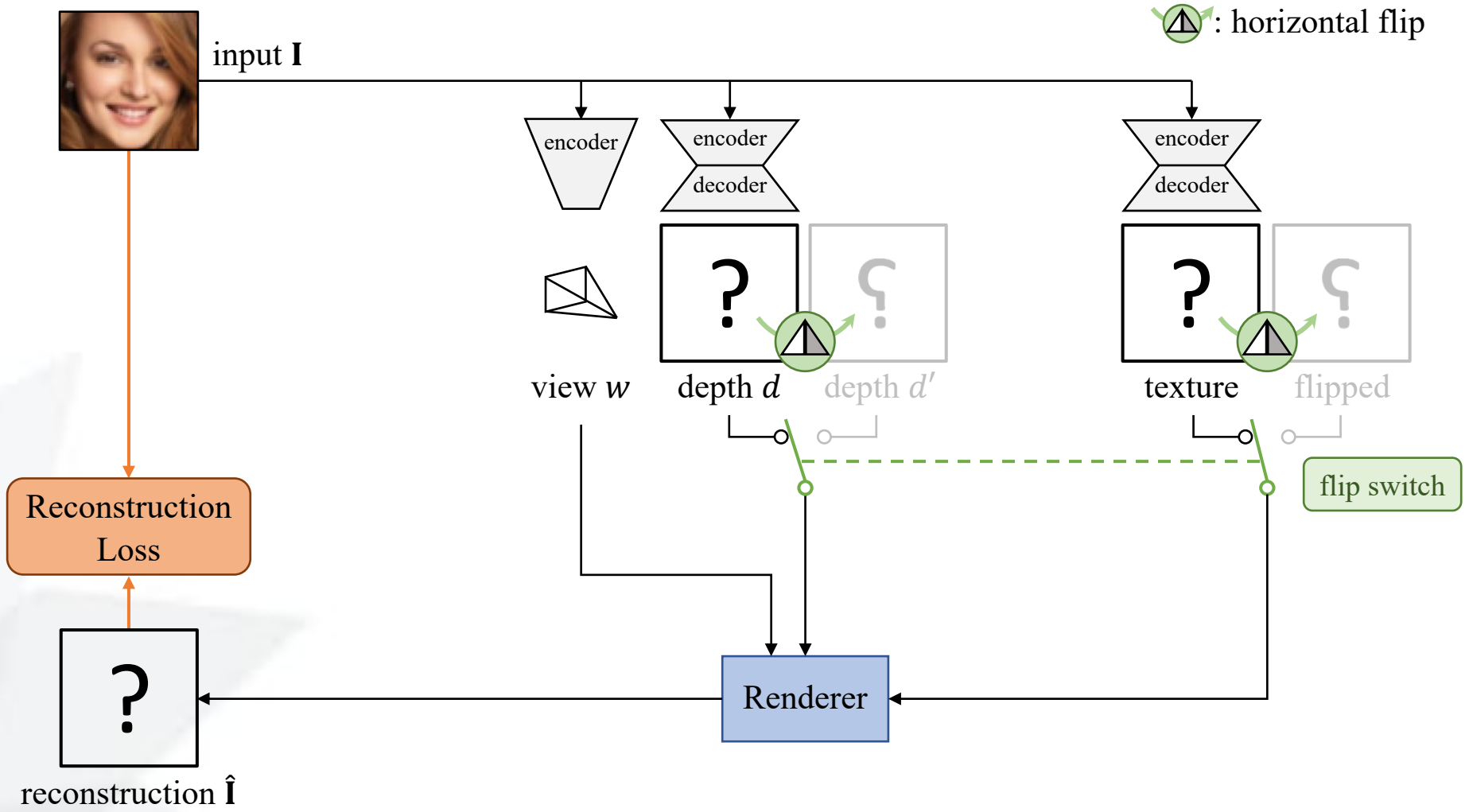


Photo-Geometric Autoencoding

Q1: How to avoid degenerate solutions?

A1: Enforce symmetry by flipping

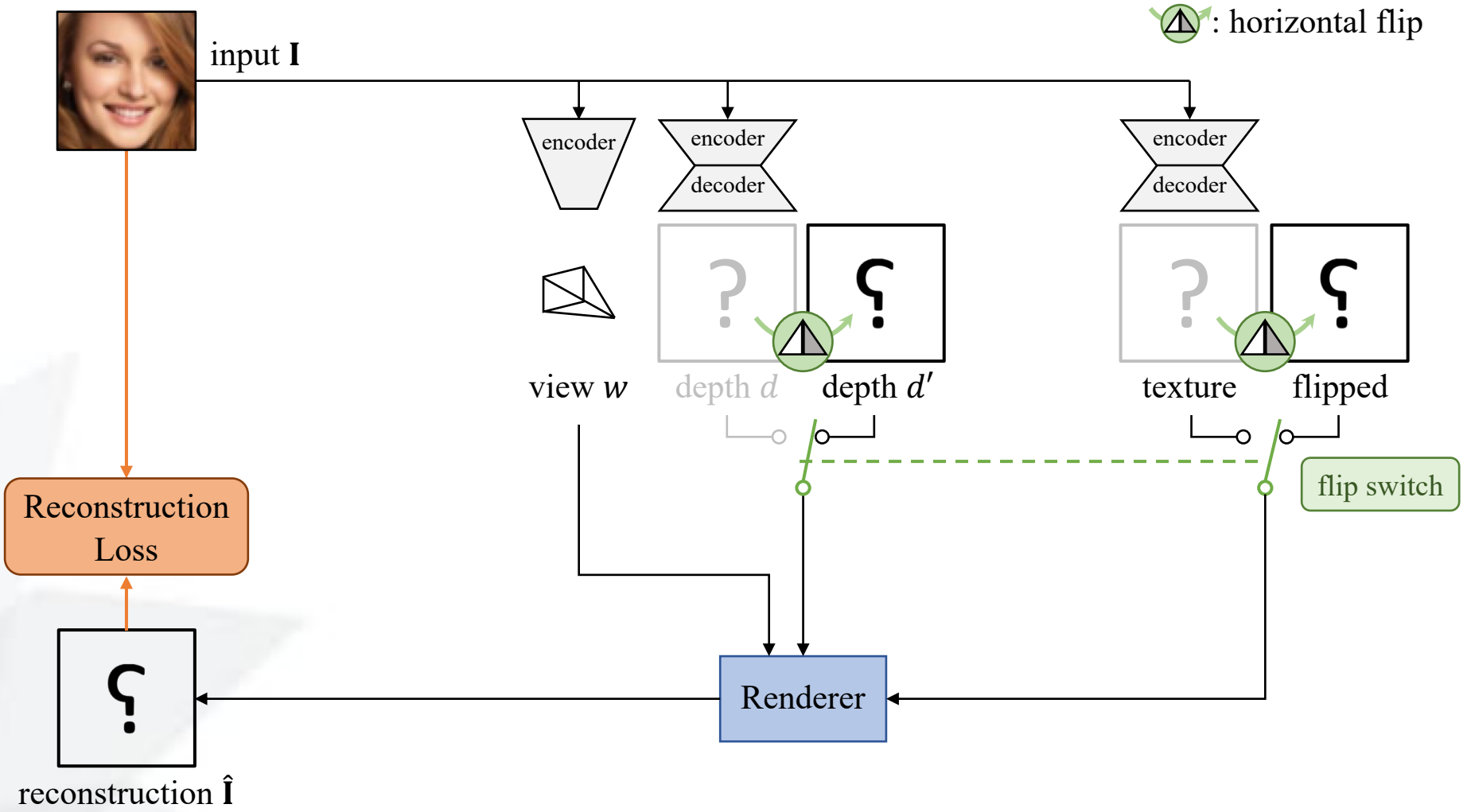


Photo-Geometric Autoencoding

Q1: How to avoid degenerate solutions?

A1: Enforce symmetry by flipping

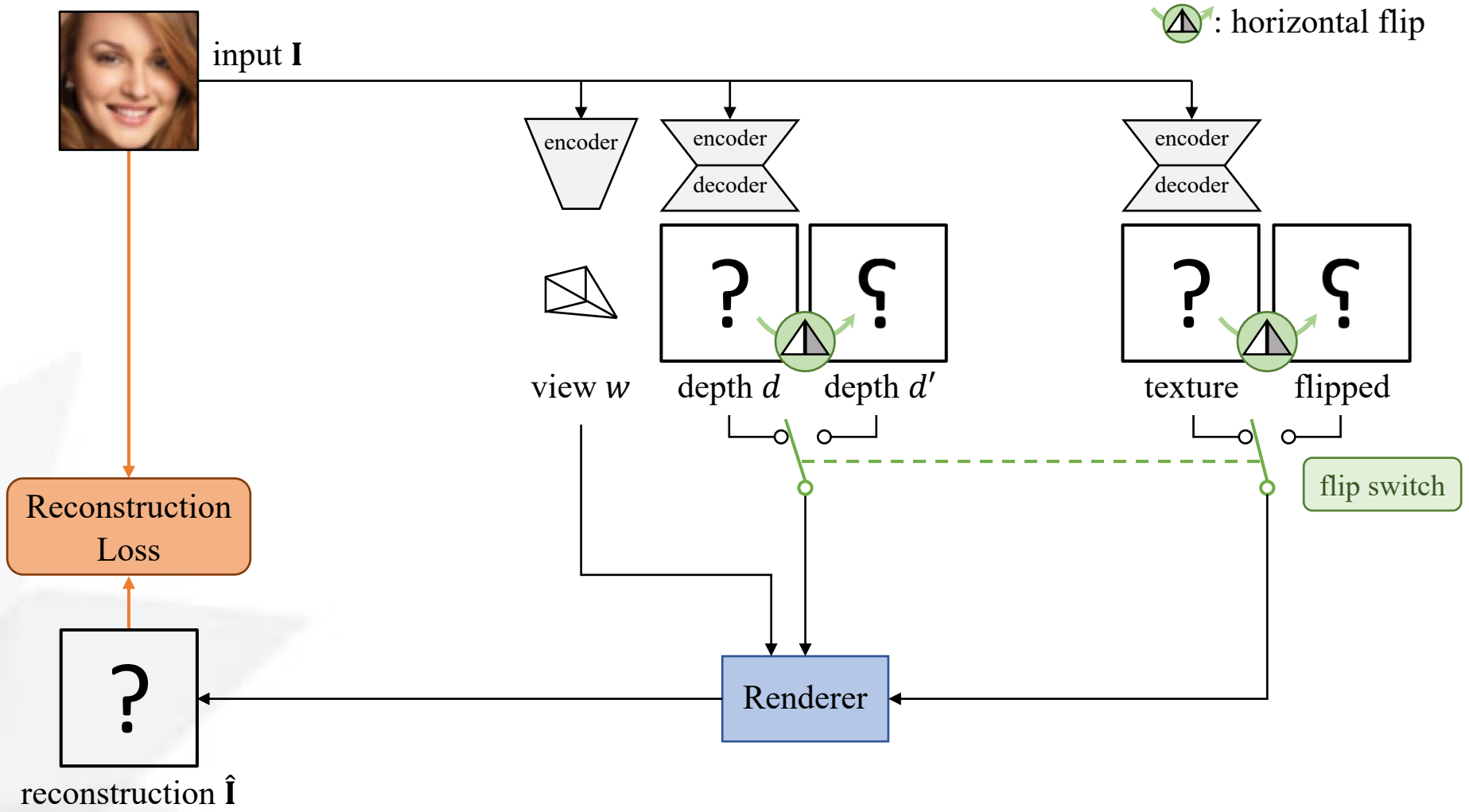


Photo-Geometric Autoencoding

Q1: How to avoid degenerate solutions?

A1: Enforce symmetry by flipping

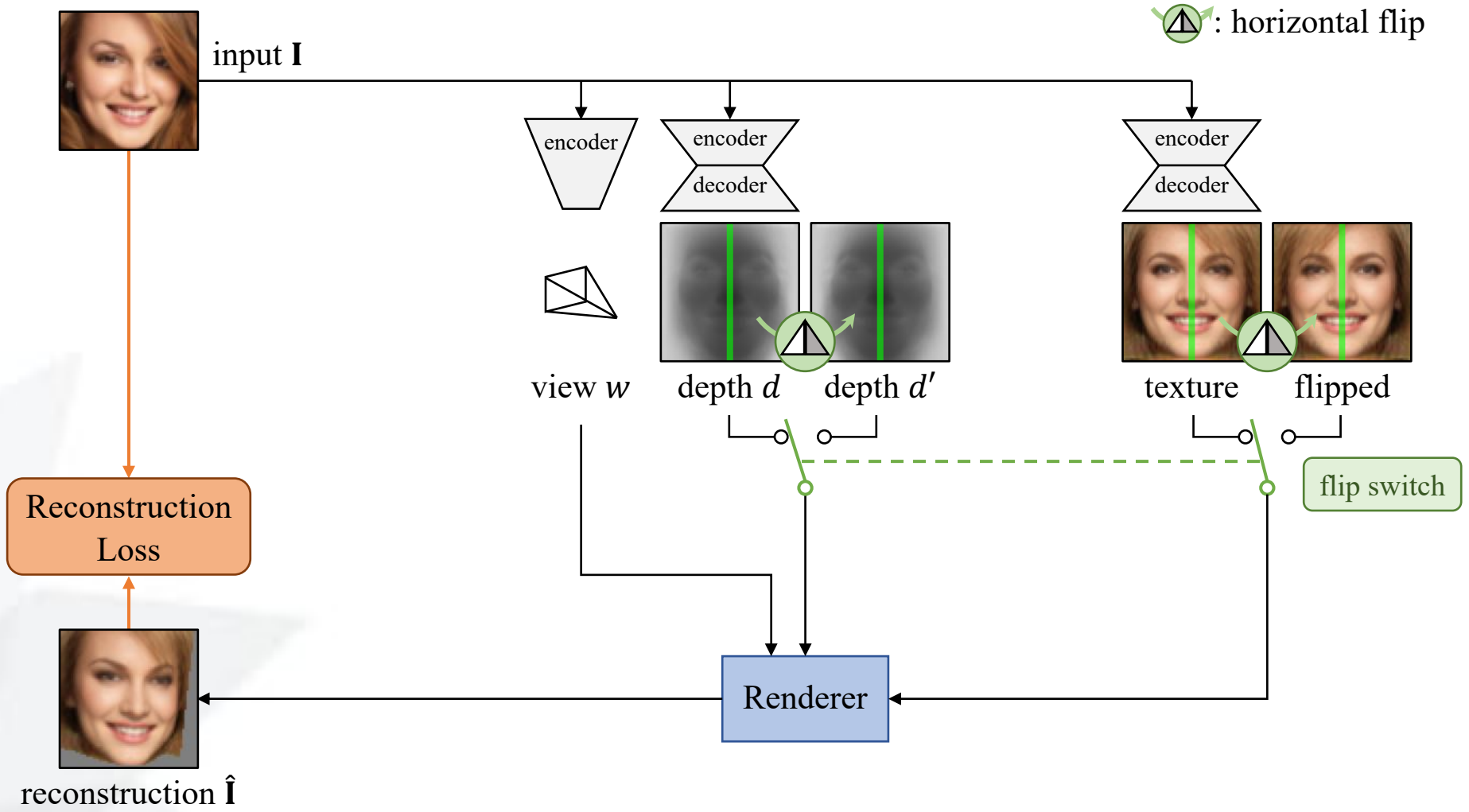


Photo-Geometric Autoencoding

Q2: What about non-symmetric lighting?

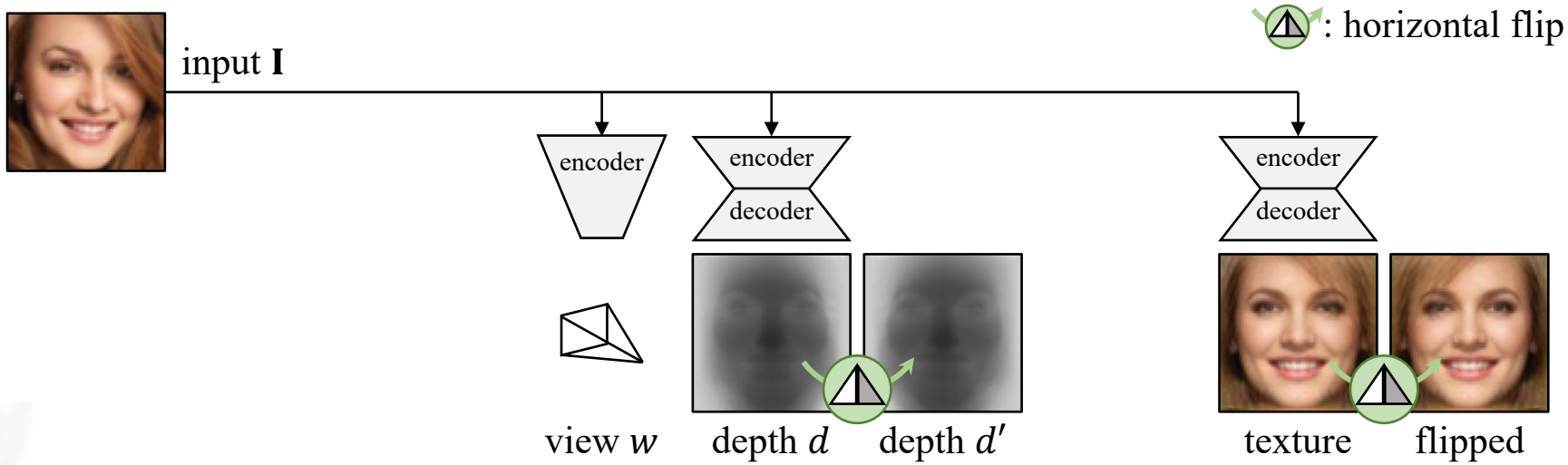


Photo-Geometric Autoencoding

Q2: What about non-symmetric lighting?

A2: Enforce symmetry on albedo

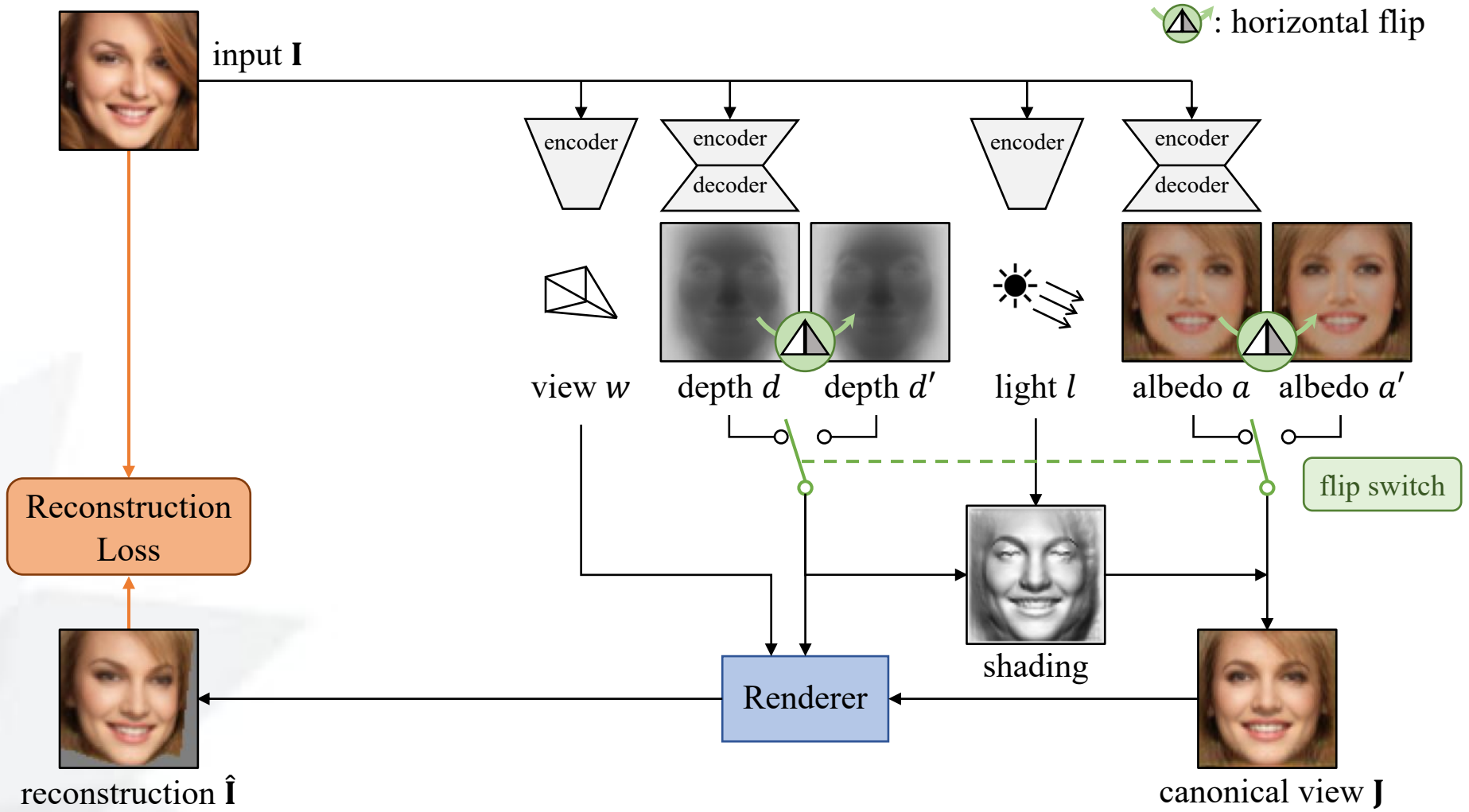


Photo-Geometric Autoencoding

Q3: Non-symmetric albedo, deformation, etc?

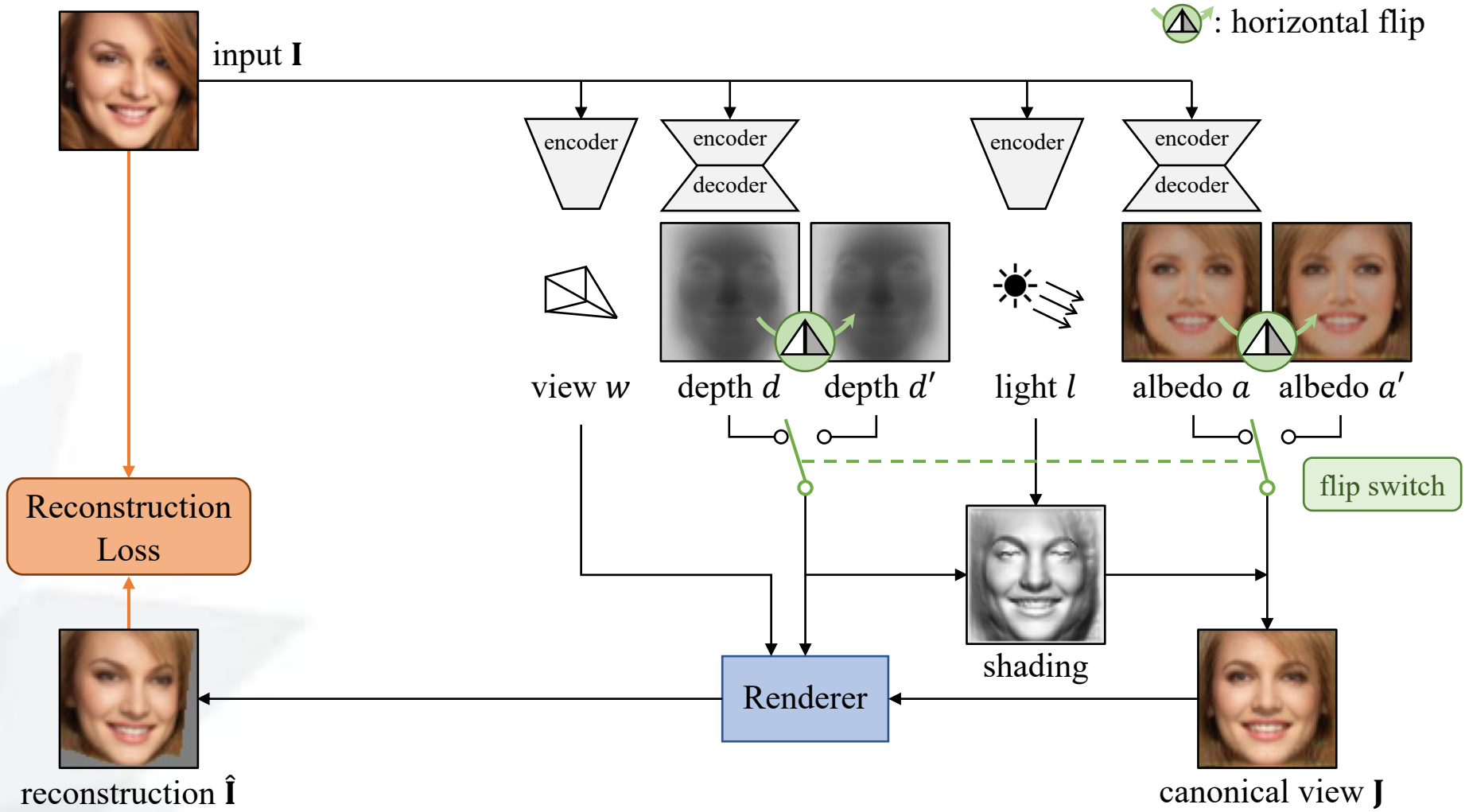


Photo-Geometric Autoencoding

Q3: Non-symmetric albedo, deformation, etc? **A3**: Predict uncertainty

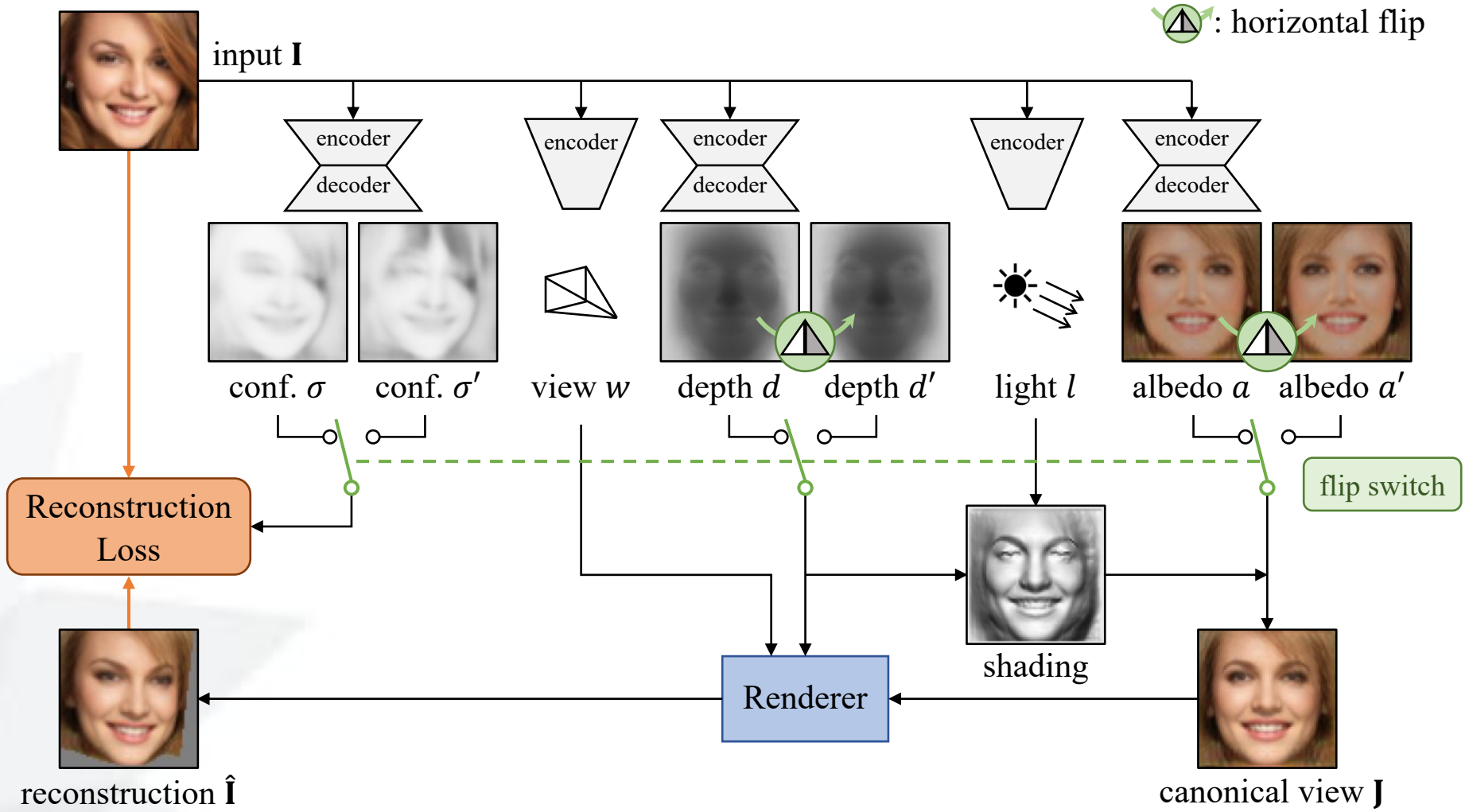


Photo-Geometric Autoencoding

Q3: Non-symmetric albedo, deformation, etc? **A3**: Predict uncertainty

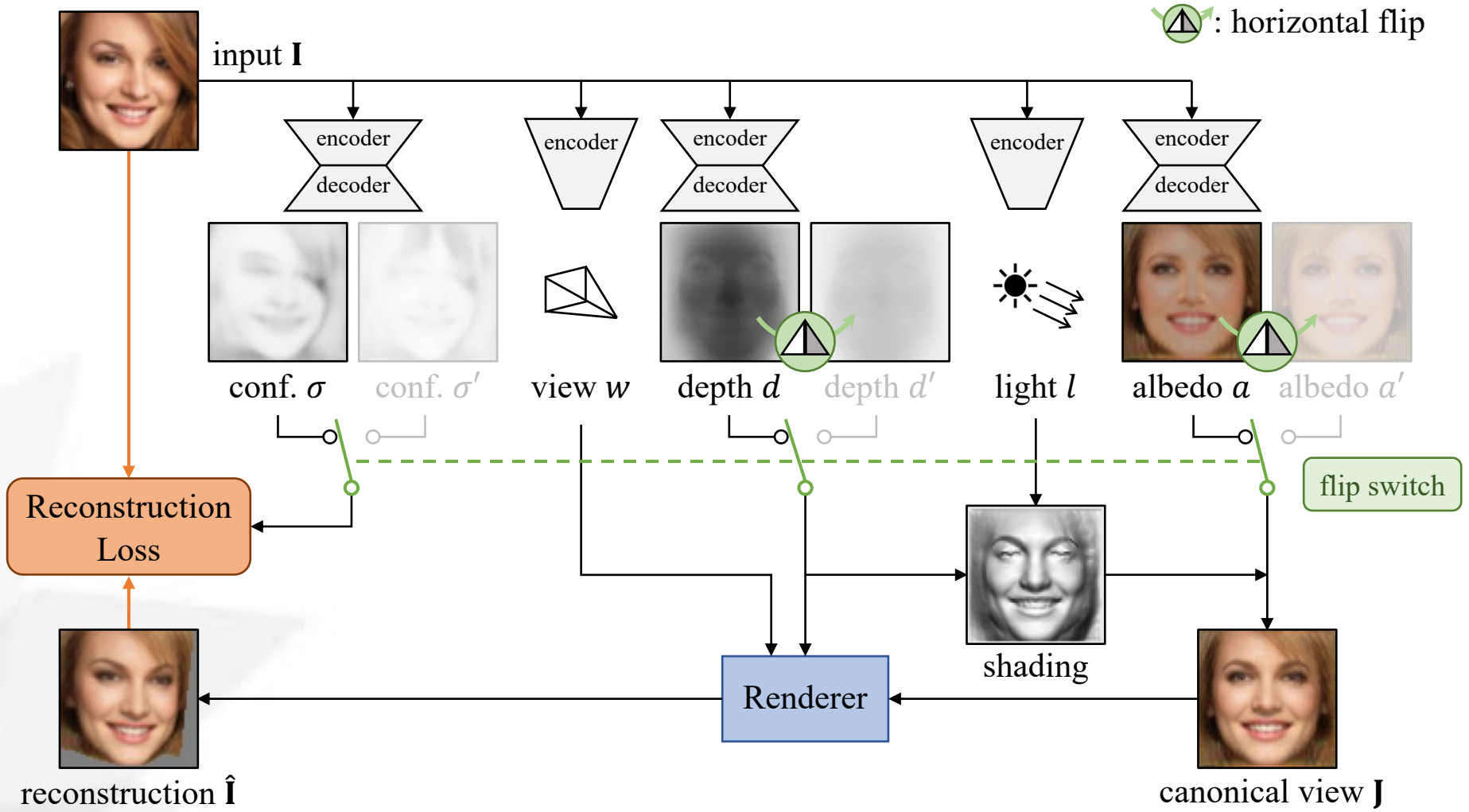


Photo-Geometric Autoencoding

Q3: Non-symmetric albedo, deformation, etc? **A3**: Predict uncertainty

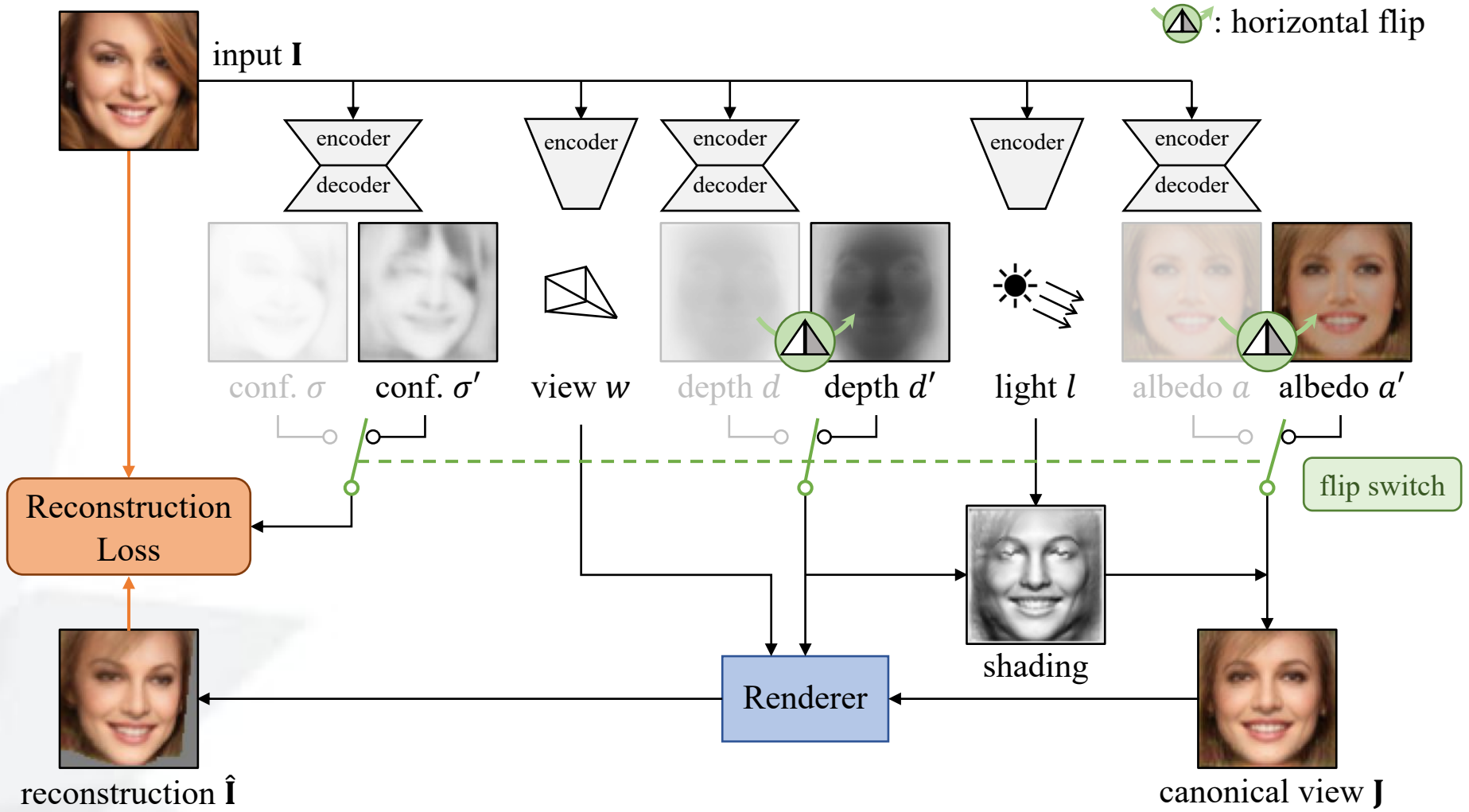


Photo-Geometric Autoencoding

Q3: Non-symmetric albedo, deformation, etc? **A3**: Predict uncertainty

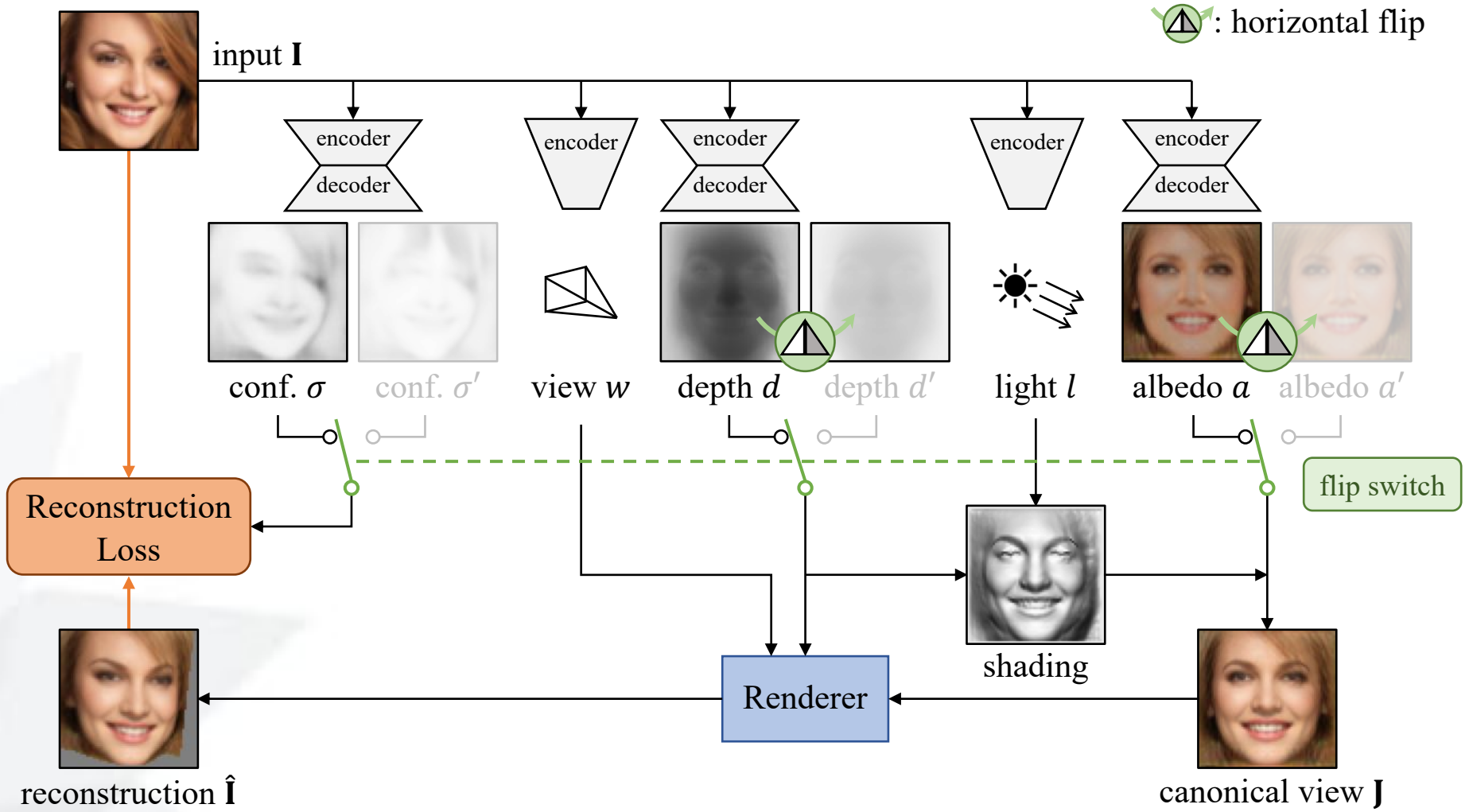


Photo-Geometric Autoencoding

Q3: Non-symmetric albedo, deformation, etc? **A3**: Predict uncertainty

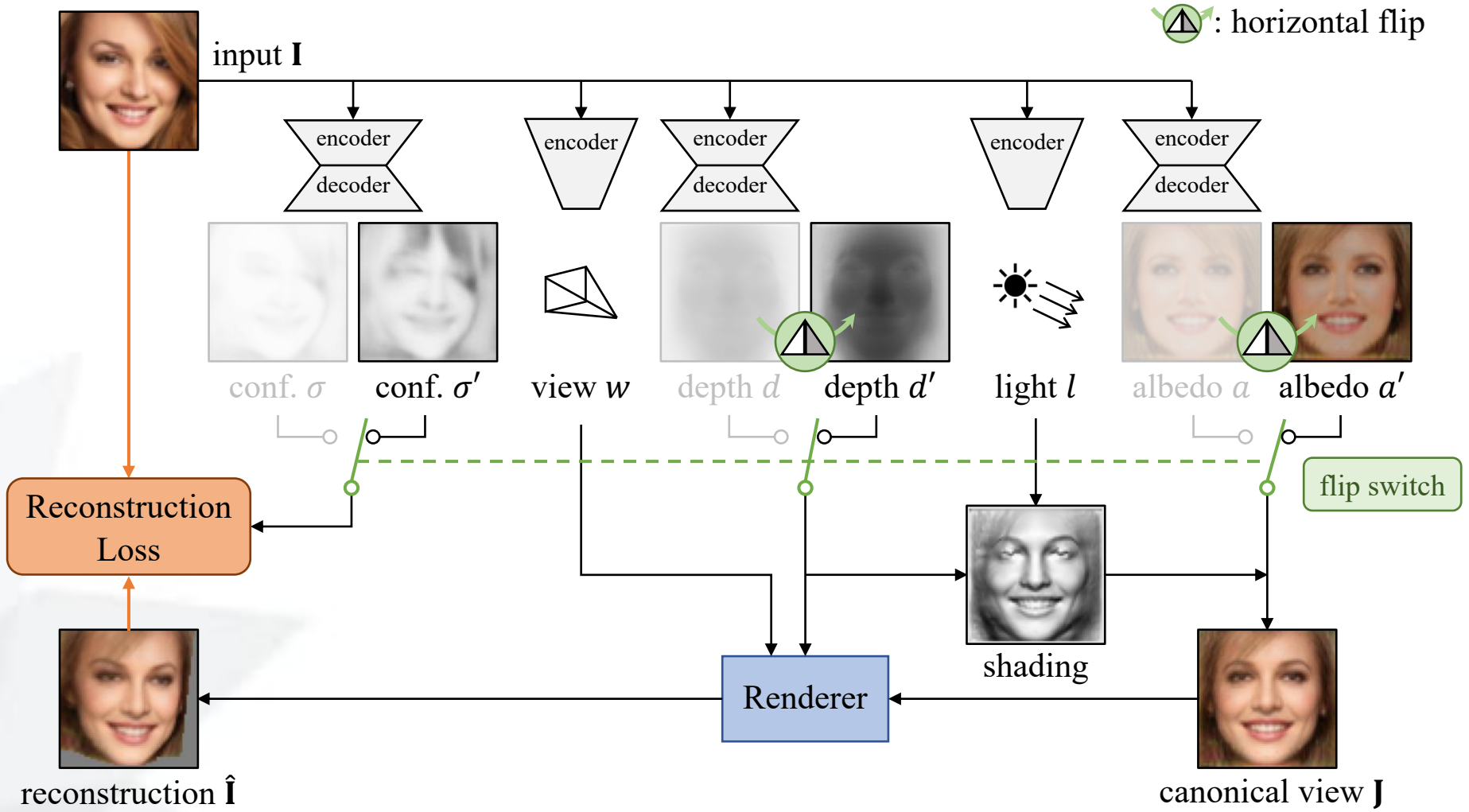


Photo-Geometric Autoencoding

Q3: Non-symmetric albedo, deformation, etc? **A3**: Predict uncertainty

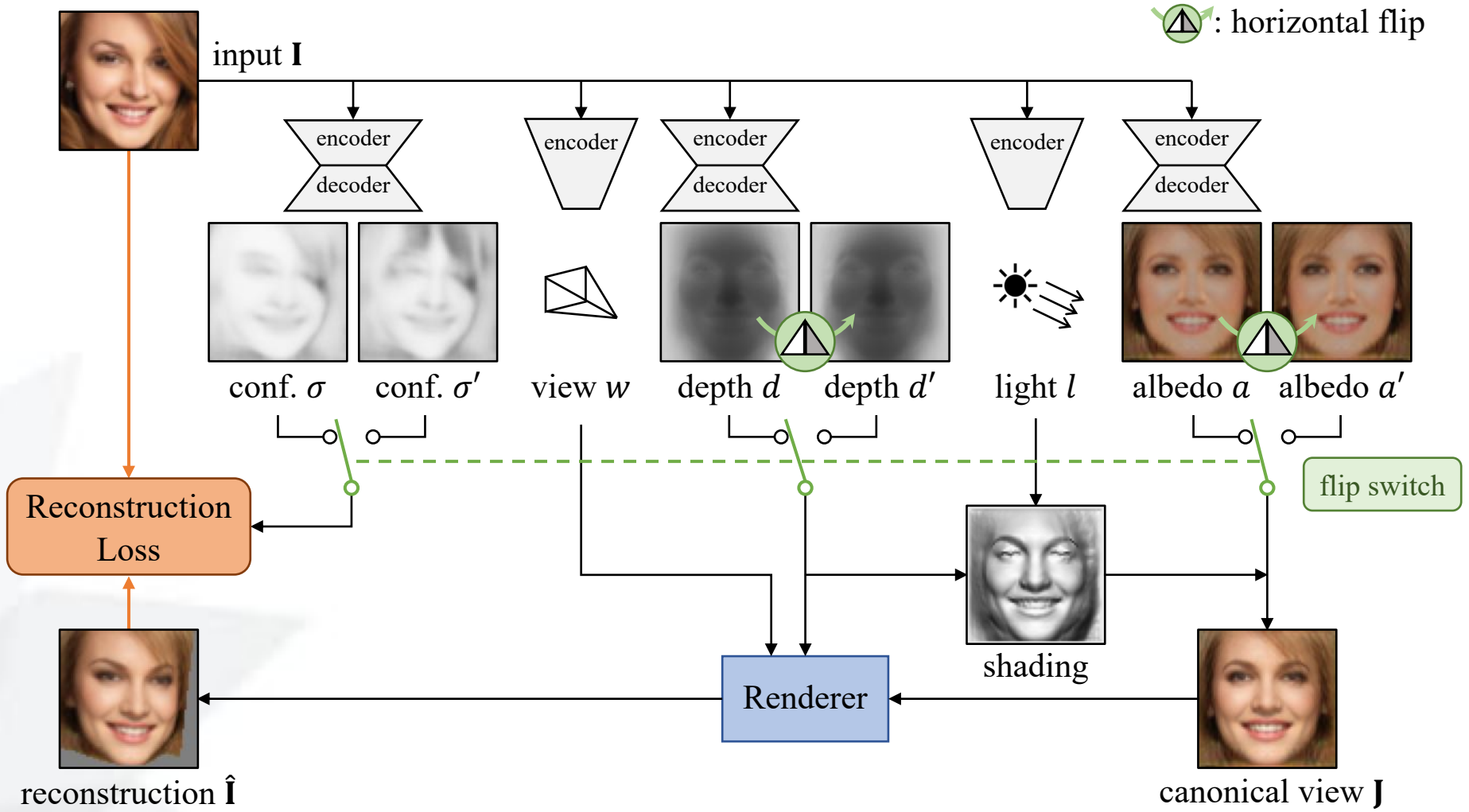
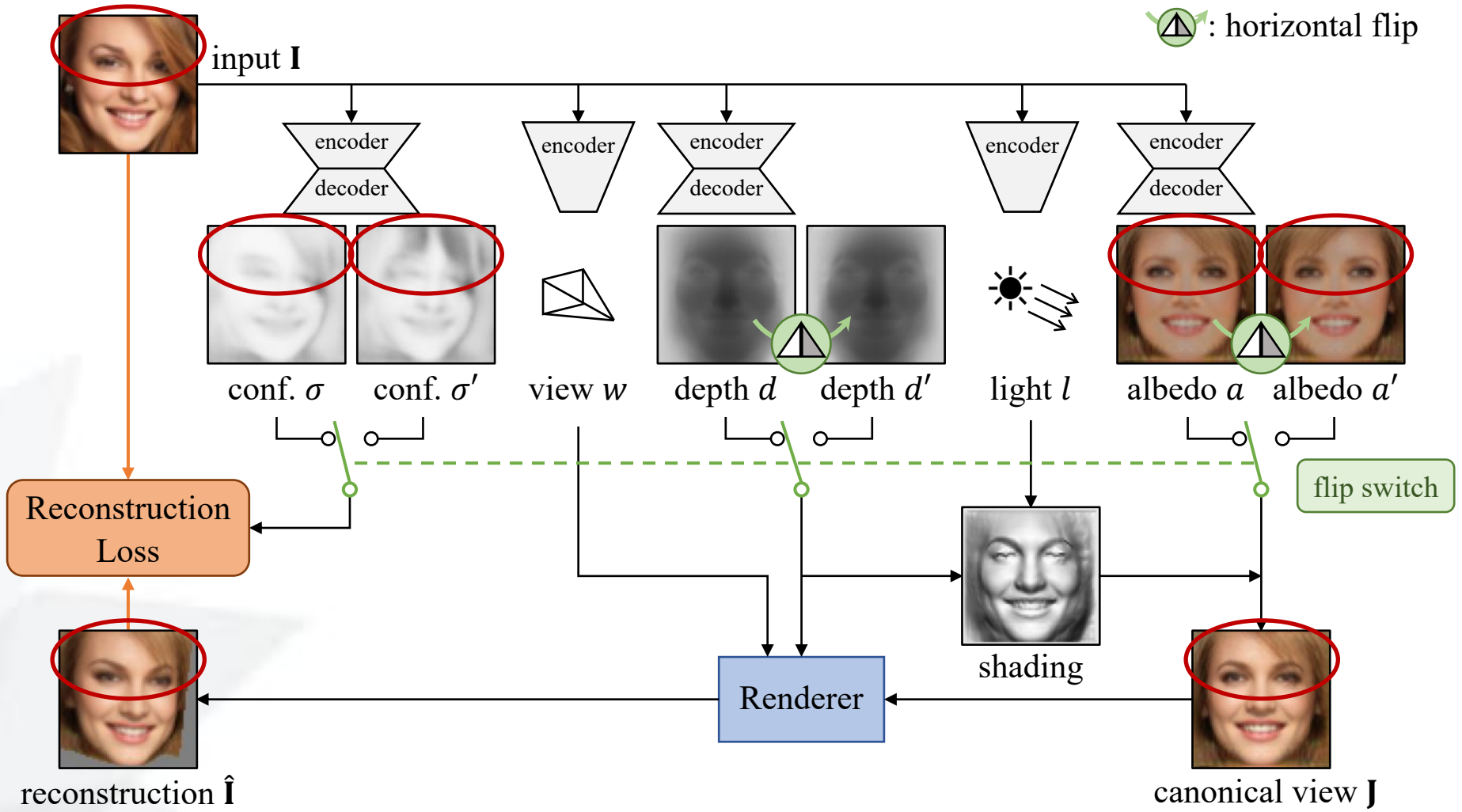


Photo-Geometric Autoencoding

Q3: Non-symmetric albedo, deformation, etc? **A3**: Predict uncertainty



Results on **human faces**

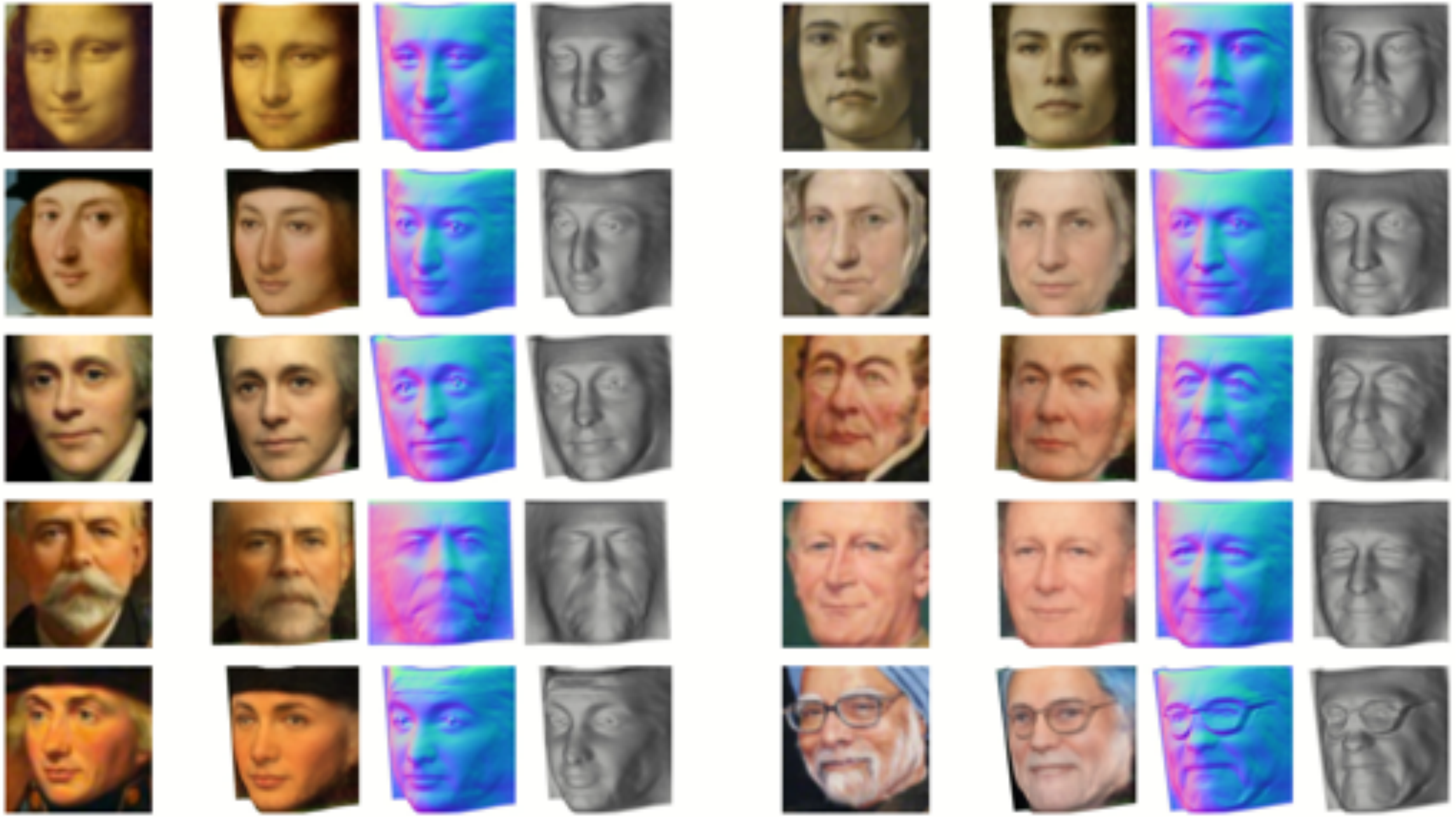
Images taken from CelebA, 3DFAW

input reconstruction input reconstruction

Results on **face paintings**

Images taken from [1]

[1] Elliot J. Crowley, Omkar M. Parkhi, and Andrew Zisserman. Face painting: querying art with photos. In Proc. BMVC, 2015.

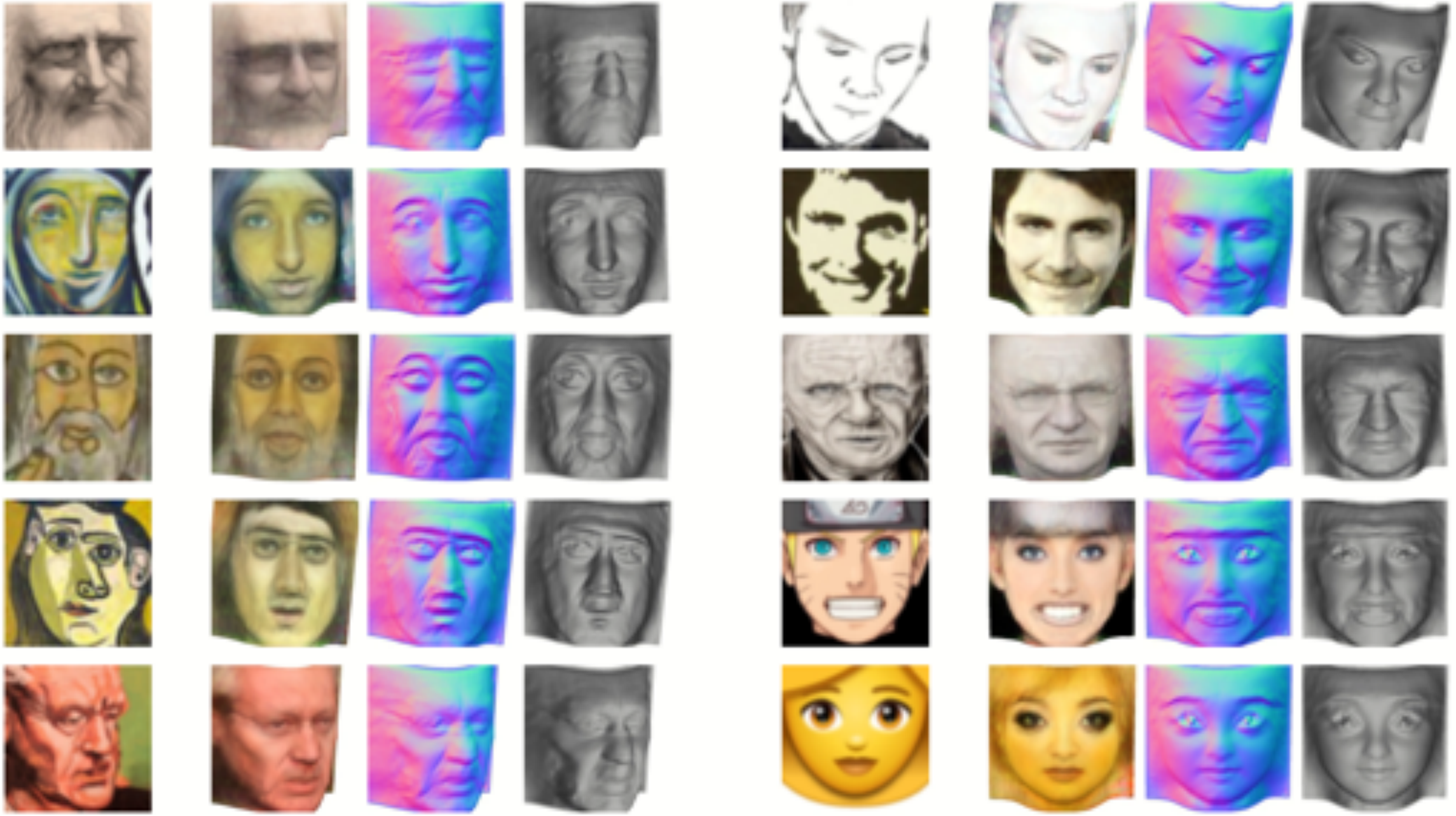


input reconstruction input reconstruction

Results on **abstract faces**

Images taken from [1] and the Internet

[1] Elliot J. Crowley, Omkar M. Parkhi, and Andrew Zisserman. Face painting: querying art with photos. In Proc. BMVC, 2015.

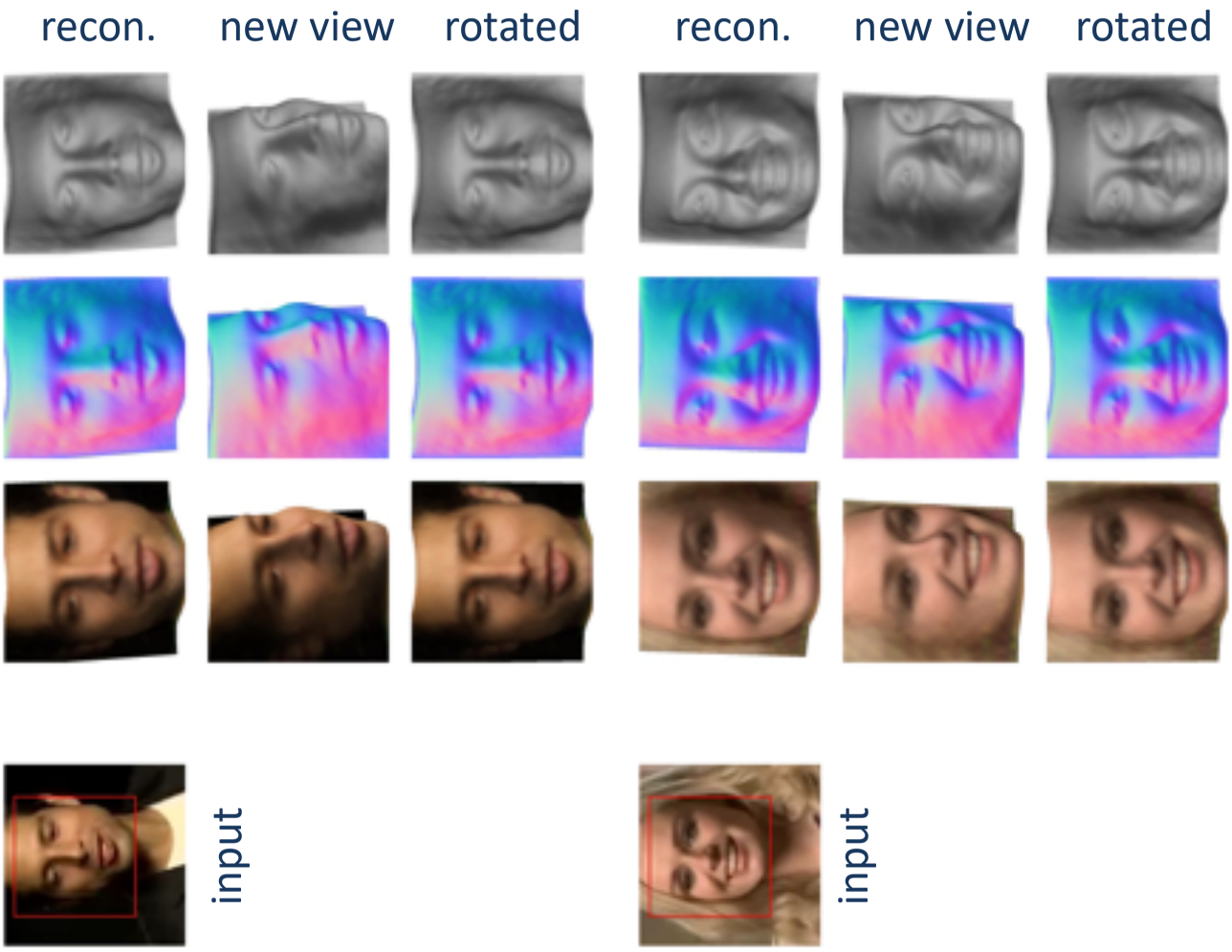


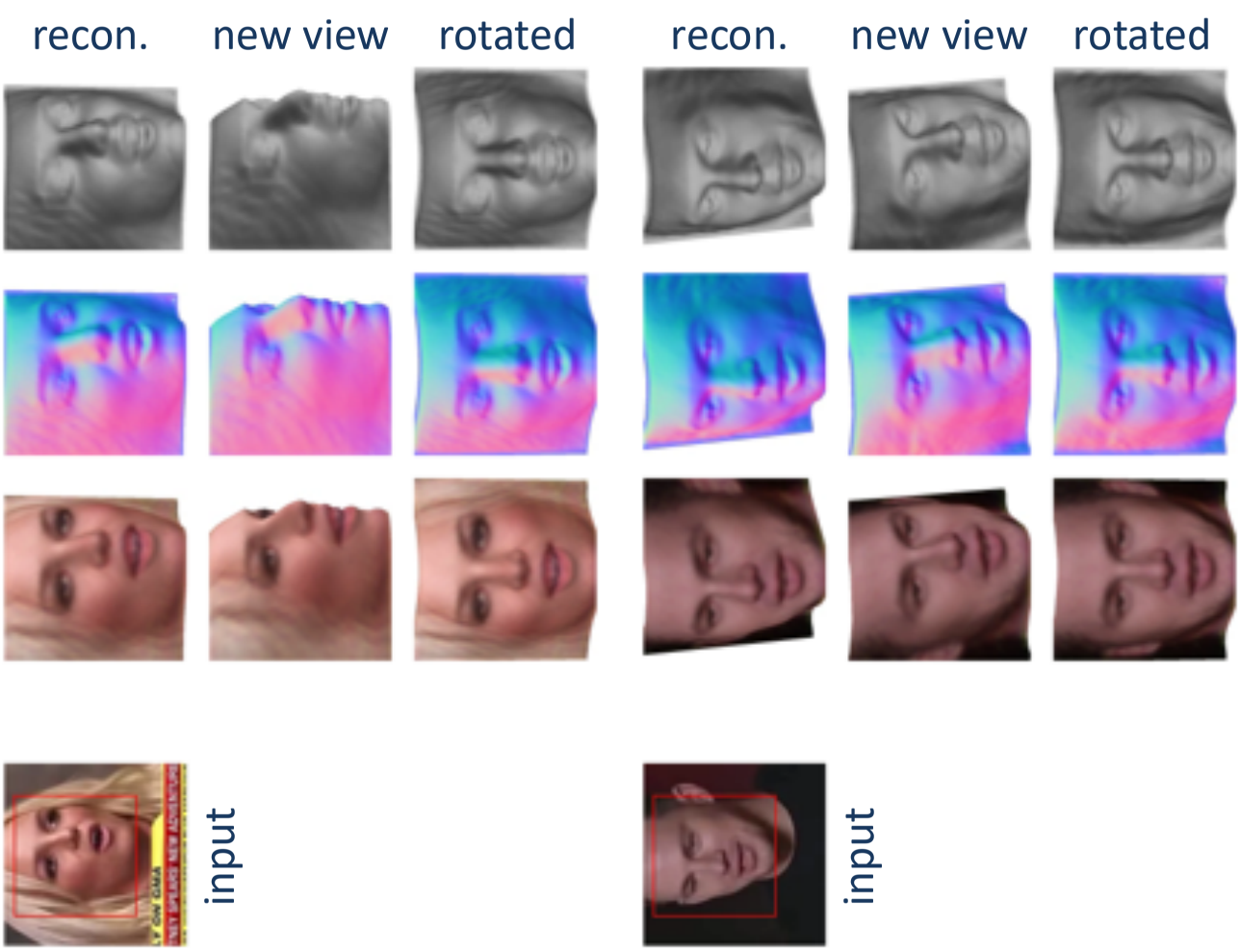
input reconstruction input reconstruction

Results on **video frames**

Video clips taken from VoxCeleb2

We do **not** use videos for training or fine-tuning. These results are obtained by applying our model trained on CelebA **frame by frame**.


recon.
new view
rotated
recon.
new view
rotated
input
input


recon.
new view
rotated
recon.
new view
rotated
input
input

Relighting effects

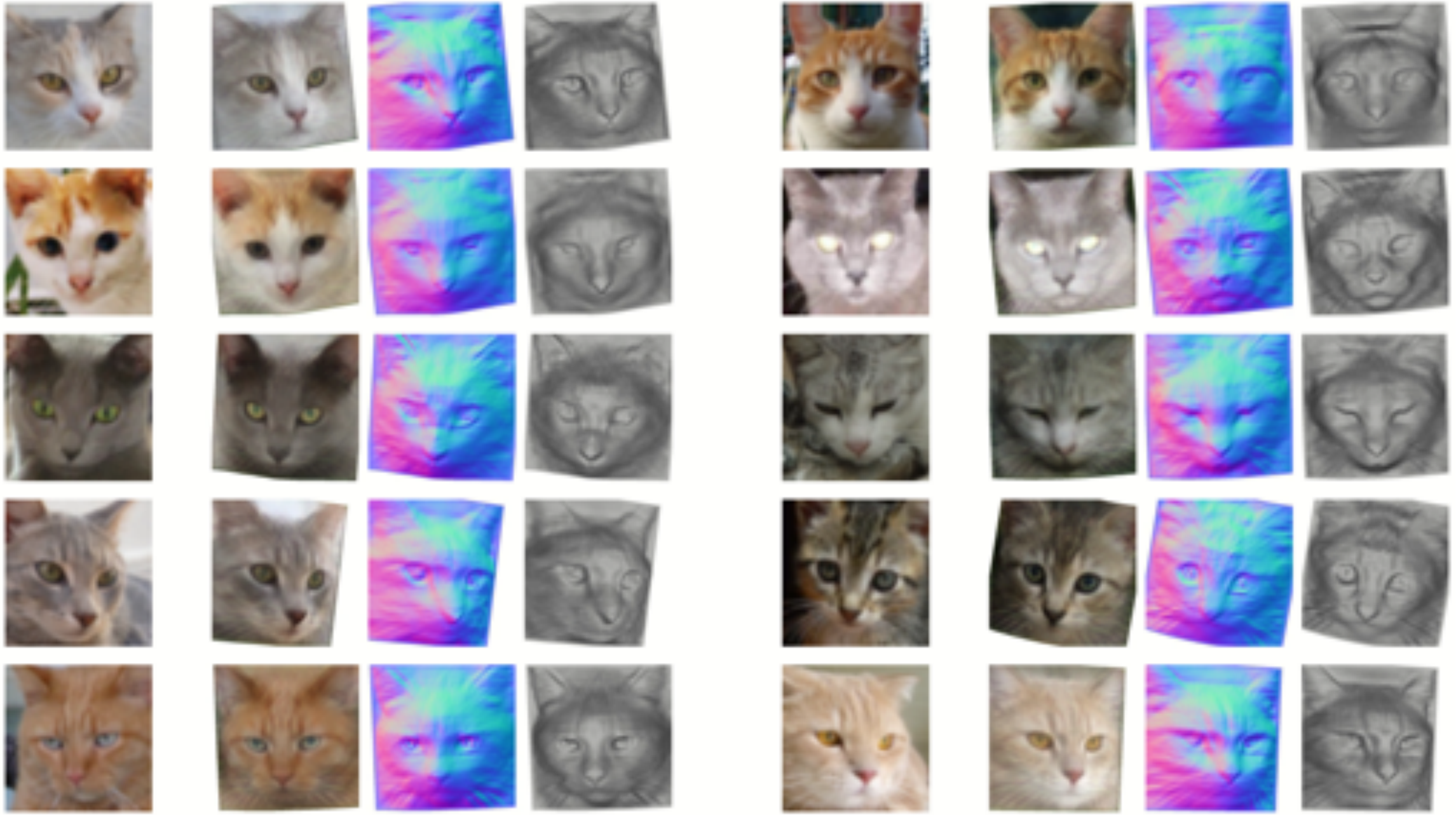
Images taken from CelebA

input reconstruction input reconstruction

Results on **cat faces**

Images taken from [2] and [3]

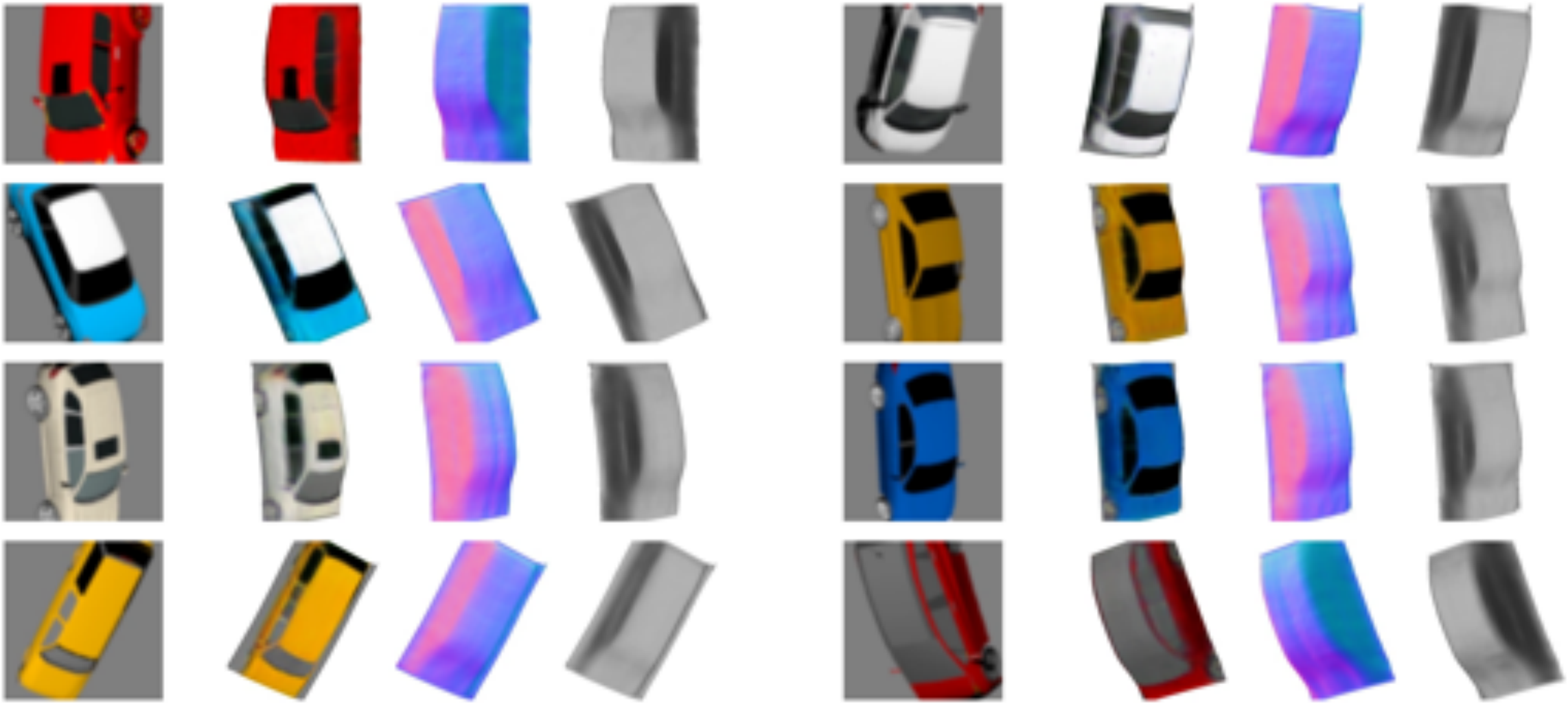
[2] Weiwei Zhang, Jian Sun, and Xiaoou Tang. Cat head detection - how to effectively exploit shape and texture features. In Proc. ECCV, 2008.
[3] Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman, and C. V. Jawahar. Cats and dogs. In Proc. CVPR, 2012.



input reconstruction input reconstruction

Results on **synthetic cars**

Images rendered using ShapeNet

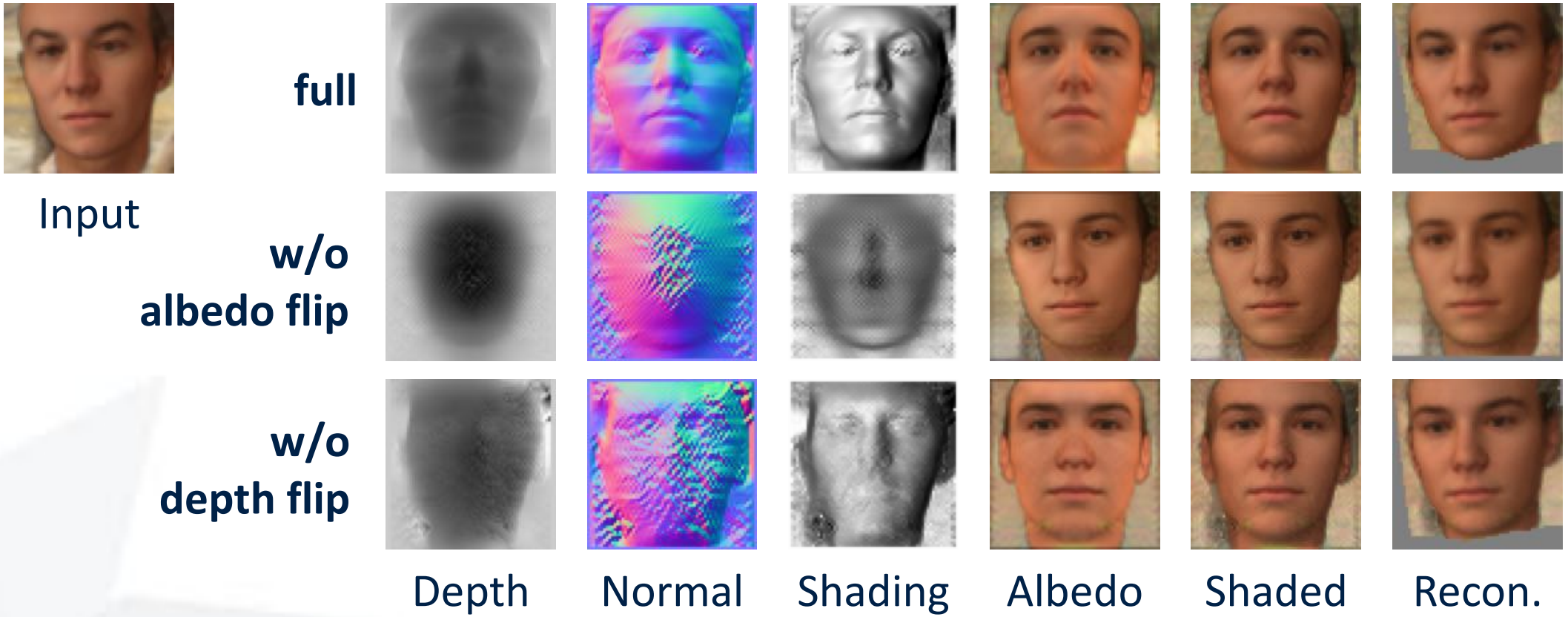

input
reconstruction
input
reconstruction

Symmetry Plane Visualization

Asymmetry Visualization

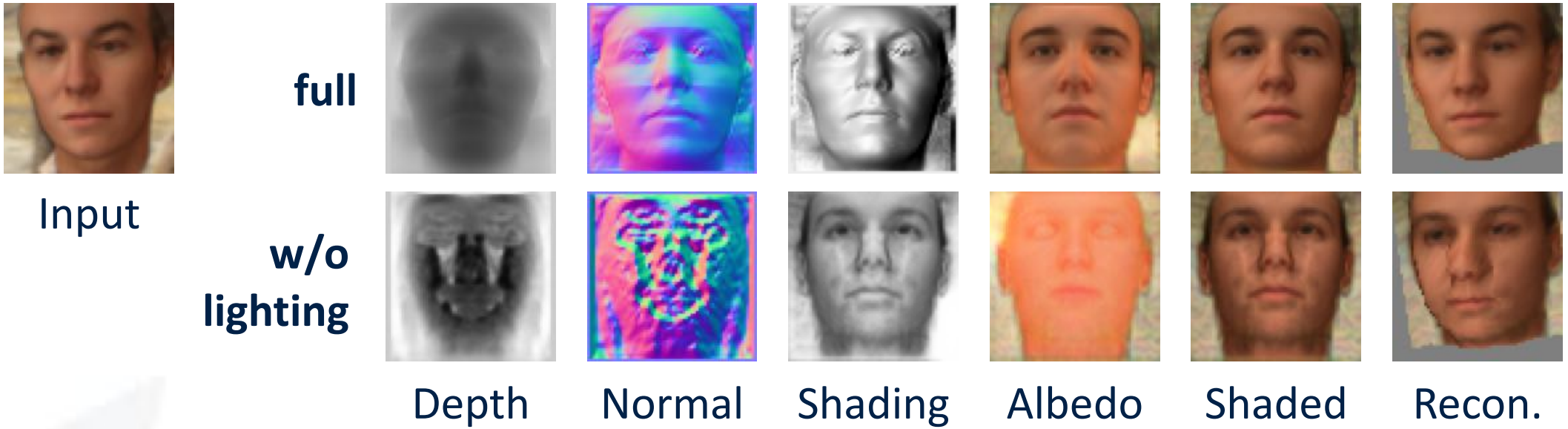
Discussion:
Ablation Studies

Ablation – Symmetry



Insight #1: **Symmetry** avoids degeneracy

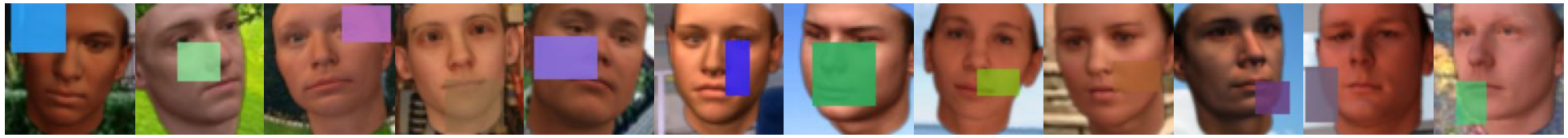
Ablation – Lighting (Shape from Shading)



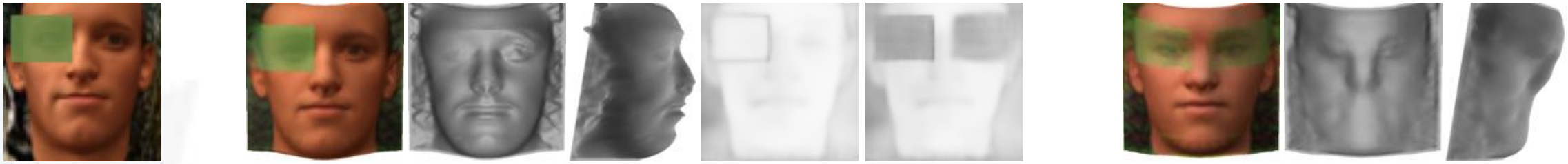
Insight #2: **Lighting** avoids bumpy shapes and provides cues for shape

Ablation – Confidence Maps

Asymmetry perturbation



conf. σ conf. σ'



input

recon. w/ conf.

recon. w/o conf.

Insight #3: **Confidence maps** allows for asymmetry modelling

Discussion:
Limitations

Limitation #1: Poor side reconstruction

input reconstructions input reconstructions

Why?

- Canonical depth map cannot represent the shape of the side

Limitation #2: Side pose input

input

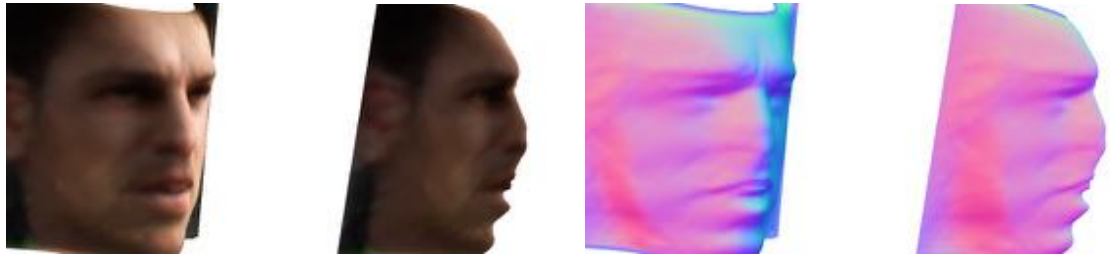
reconstructions

Why?

- There are no/few symmetric correspondences present in the image, which voids the symmetry regularization
- No enough side images in the training set

Limitation #3: Lambertian shading

input



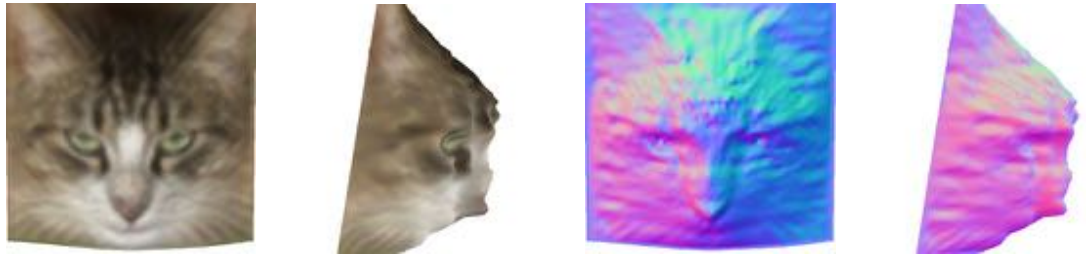
reconstructions

Why?

- We assume Lambertian shading with one dominant directional light, and do not model specularity and shadow

Limitation #4: Dark texture vs. dark shading

input



reconstructions

Why?

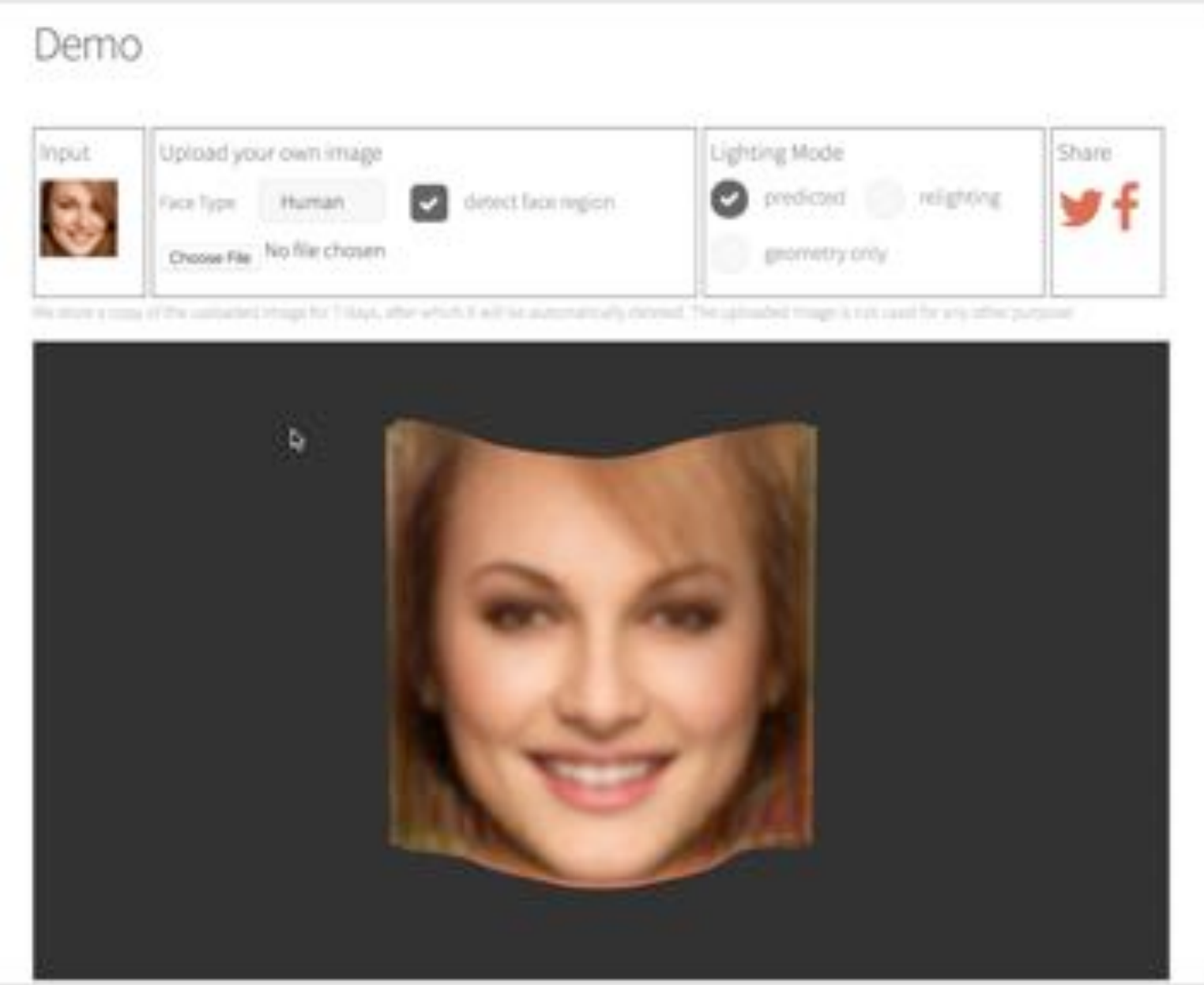
- Disentangling dark texture and dark shading is hard. The model may produce dark shading with spiky shape to reconstruct dark texture.

Conclusions

- We present an unsupervised method for learning deformable 3D objects from only raw single-view images
- Bilateral symmetry in common objects provides a powerful constraint for learning 3D shapes
- Shading provides important geometric cues and helps regularize shape
- By modeling symmetric albedo and non-symmetric shading separately, our model automatically learns intrinsic image decomposition without supervision
- Confidence maps can be used to model asymmetries

Key Takeaways

- 3D understanding is possible from only single-view 2D observations; image recognition should go beyond 2D
- Reducing supervision in training is important for 3D understanding in the wild
- Physical cues and visual patterns are useful, such as symmetry, shading, planes, repetitions, etc



Thank you!

Demo: bit.ly/2zBNjXx

Code: github.com/elliottwu/unsup3d

VGG
UNIVERSITY OF OXFORD

UNIVERSITY OF OXFORD