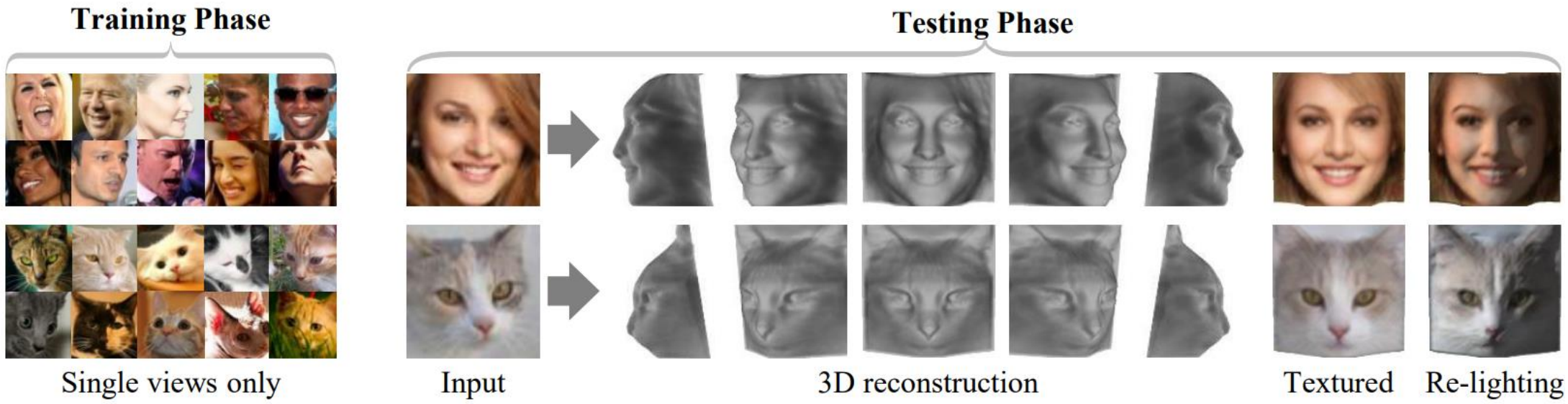
Unsupervised Learning of Probably Symmetric Deformable 3D Objects from Images in the Wild

210118_LabSeminar

강사무엘

Abstract

Abstract



The method is based on an **autoencoder** that factors each input image into depth, texture, viewpoint and illumination.

Author model objects that are probably, but not certainly, symmetric **by predicting a symmetry probability map**, learned end-to-end with the other components of the model.

Method

Method

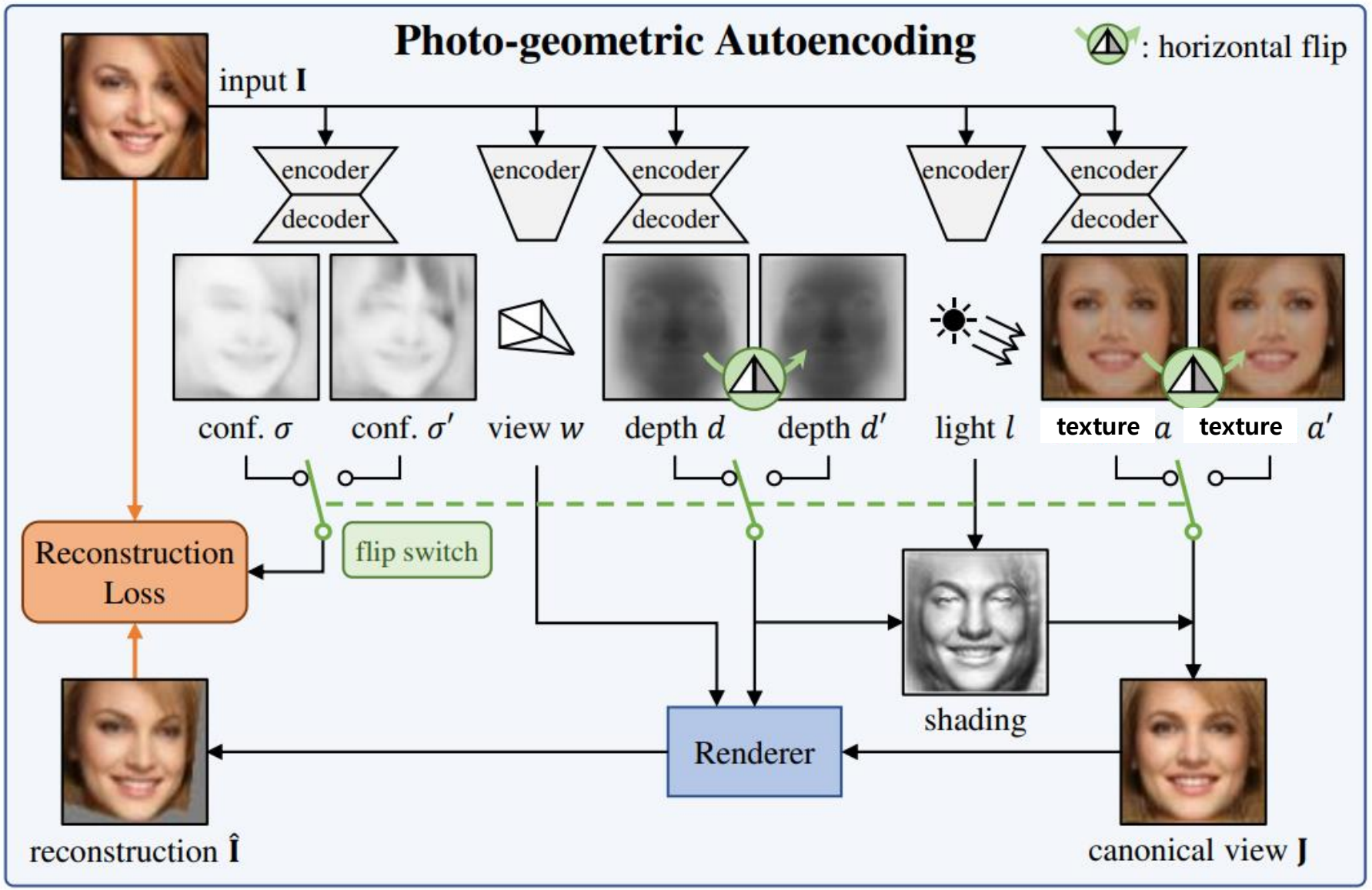
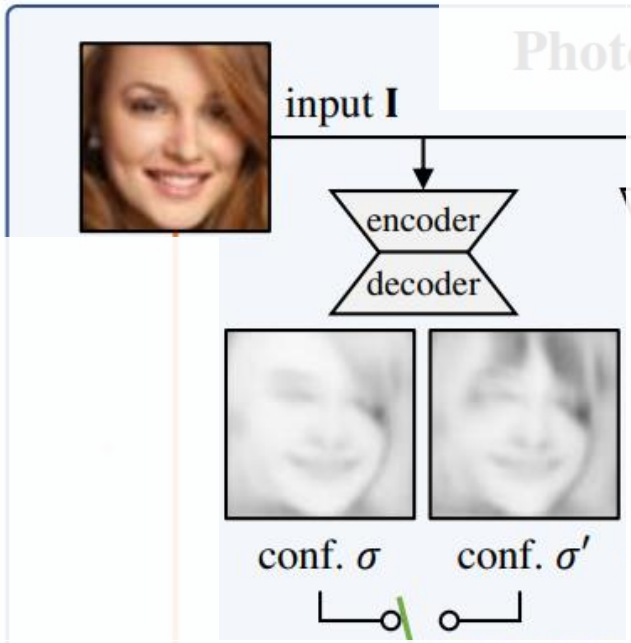
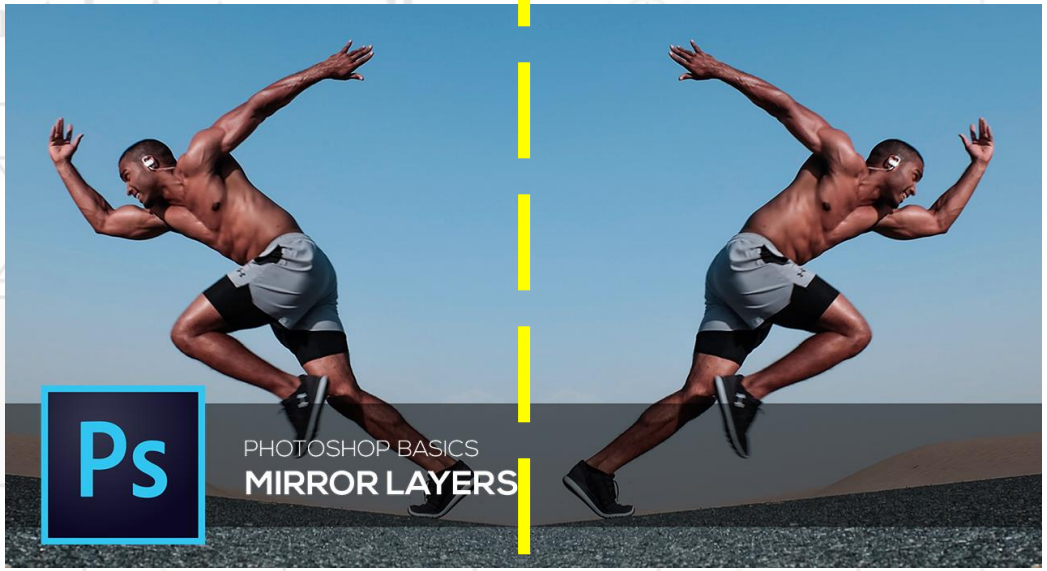


Photo-geometric Autoencoding
: horizontal flip
input I
encoder
decoder
encoder
encoder
decoder
encoder
encoder
decoder
conf. σ
conf. σ′
view w
depth d
depth d′
light l
texture a
texture a′
flip switch
Reconstruction Loss
shading
Renderer
reconstruction Î
canonical view J

Method





1. Author use the fact that many object categories are **bilaterally symmetric (양쪽 대칭)**

2. The appearance of object instances **is never perfectly symmetric.** Model also estimates, for each pixel in the input image, **a confidence score** that explains the probability of the pixel having a symmetric counterpart in the image (conf σ, σ′)

3. Choose high confidence score and use it

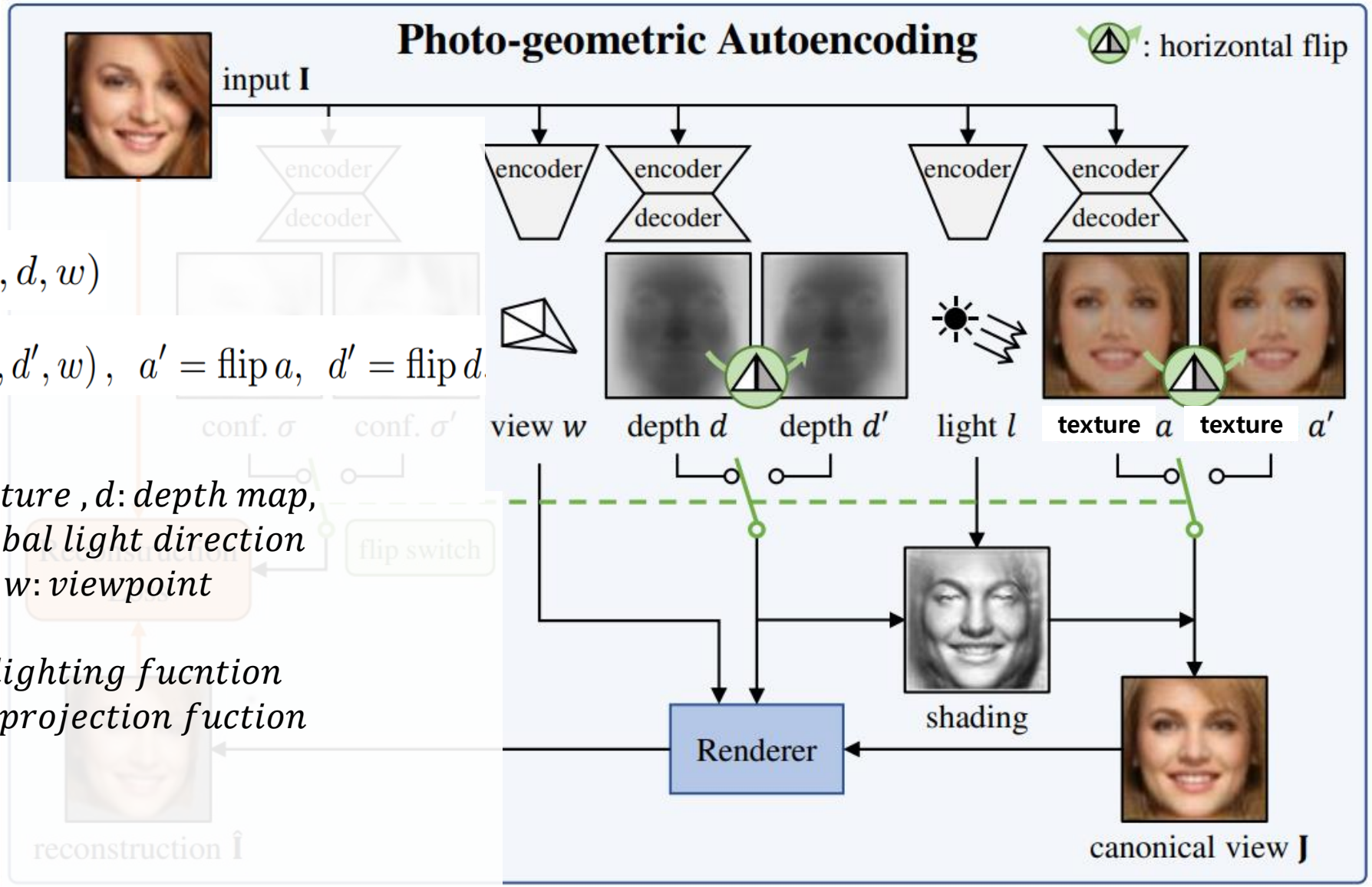
Method

$$\hat{\mathbf{I}} = \Pi\left(\Lambda(a, d, l), d, w\right)$$

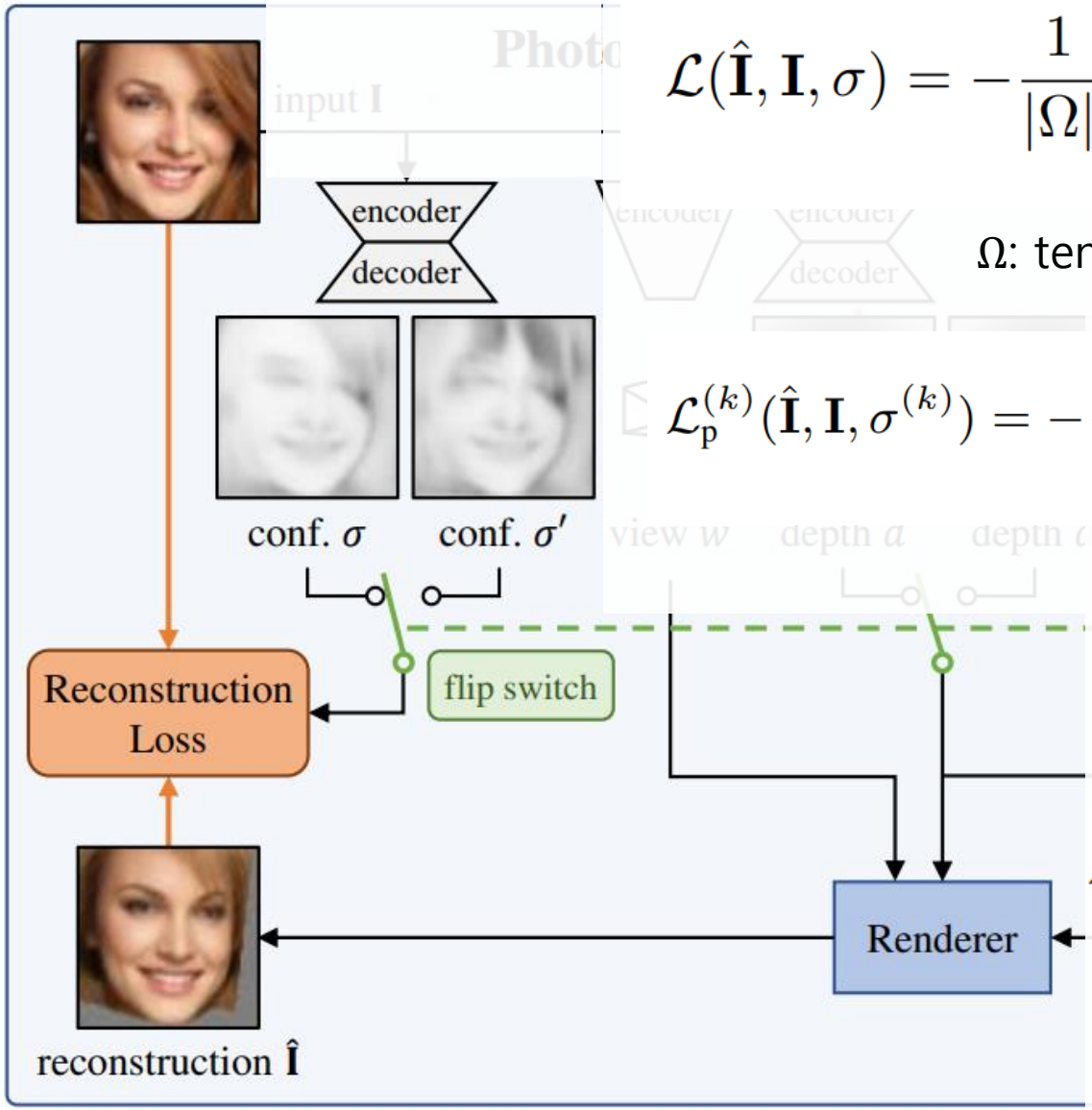
$$\hat{\mathbf{I}}' = \Pi\left(\Lambda(a', d', l), d', w\right), \quad a' = \text{flip}\, a, \quad d' = \text{flip}\, d$$

α : *texture* , d: *depth map*,
l : *global light direction*
w: *viewpoint*

Π : *lighting fucntion*
Λ : *reprojection fuction*



Method



$$\mathcal{L}(\hat{\mathbf{I}}, \mathbf{I}, \sigma) = -\frac{1}{|\Omega|} \sum_{uv \in \Omega} \ln \frac{1}{\sqrt{2}\sigma_{uv}} \exp -\frac{\sqrt{2}\ell_{1,uv}}{\sigma_{uv}},$$

Ω: tensor in $R^{3\times W\times H}$ $\quad \ell_{1,uv} = |\hat{\mathbf{I}}_{uv} - \mathbf{I}_{uv}|$

$$\mathcal{L}_{\mathrm{p}}^{(k)}(\hat{\mathbf{I}}, \mathbf{I}, \sigma^{(k)}) = -\frac{1}{|\Omega_k|} \sum_{uv \in \Omega_k} \ln \frac{1}{\sqrt{2\pi(\sigma_{uv}^{(k)})^2}} \exp -\frac{(\ell_{uv}^{(k)})^2}{2(\sigma_{uv}^{(k)})^2}$$

$$\mathcal{L} + \lambda_{\mathrm{p}}\mathcal{L}_{\mathrm{p}} \text{ with } \lambda_{\mathrm{p}} = 1$$

Experiment

Experiment

Training Data : CelebA, BFM

Test Data : 3DFAW

Metrics : Depth Correlation

Experiment

Encoder (5 layer) + Decoder(5 layer)	Depth map (# of outChannel :1, non-activation) Texture (# of outChannel :3)
Encoder (6 layer)	Viewpoint (# of outChannel :6) Lighting & Reprojection Function (# of outChannel :4)
Encoder (5 layer) + Decoder(5 layer)	Confidence Map (# of outChannel : 2)

Result

Experiment

	Depth Corr. ↑
Ground truth	66
AIGN [61] (**supervised**, from [40])	50.81
DepthNetGAN [40] (**supervised**, from [40])	58.68
MOFA [57] (**model-based**, from [40])	15.97
DepthNet [40] (from [40])	26.32
DepthNet [40] (from GitHub)	35.77
Ours	48.98
Ours (w/ CelebA pre-training)	54.65

Experiment

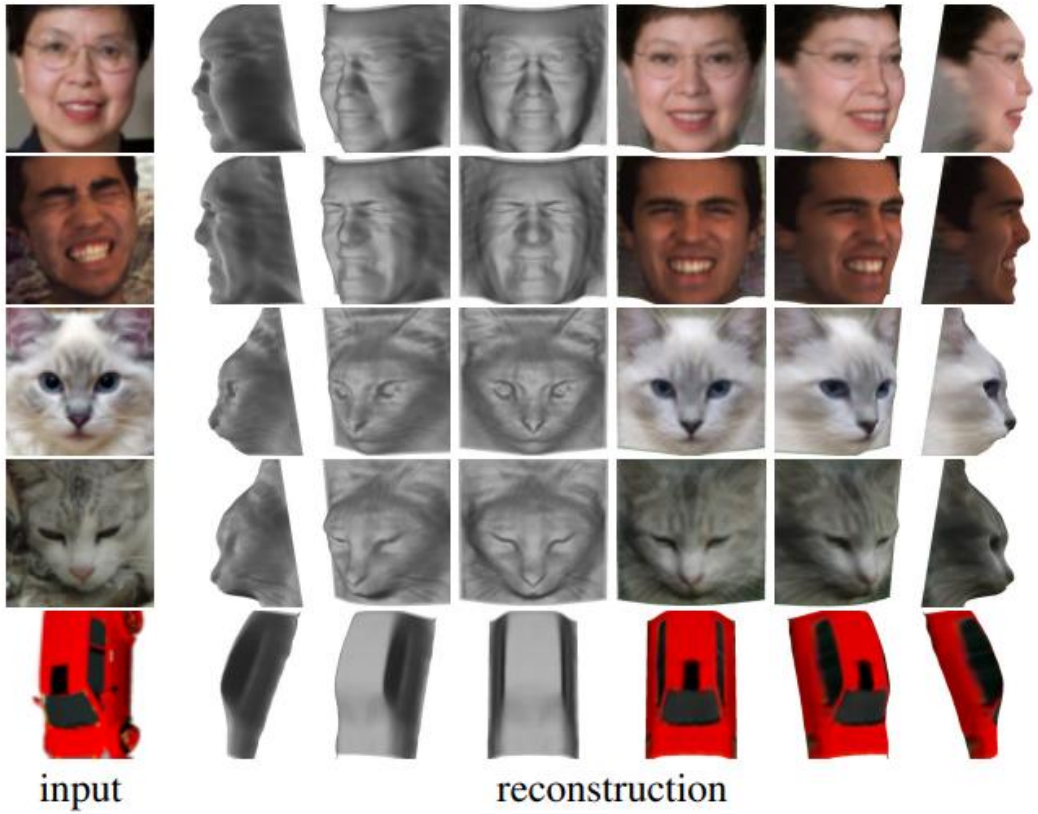
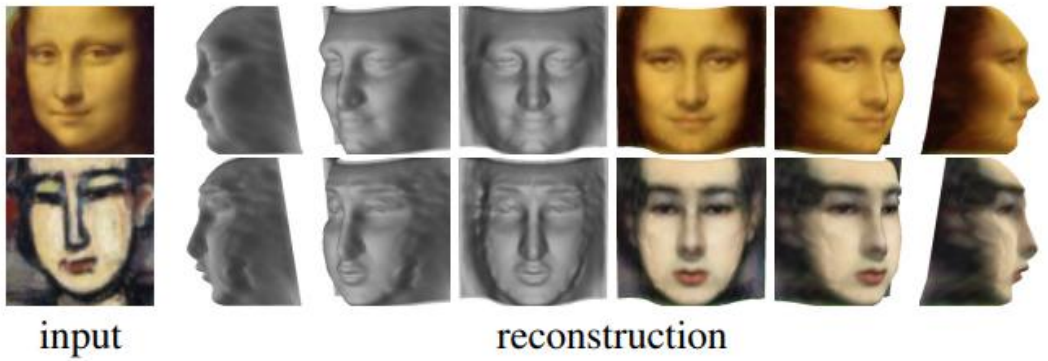
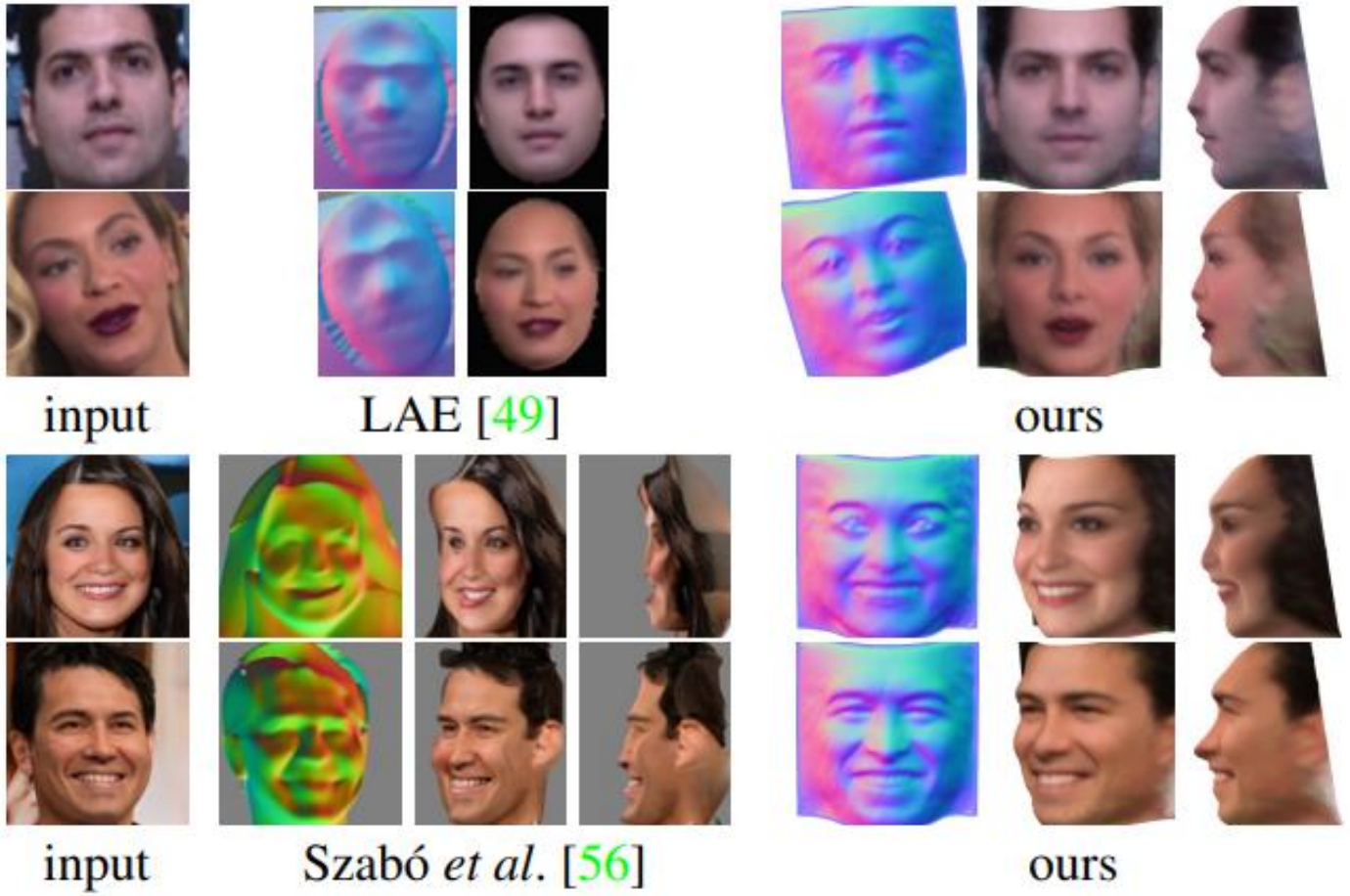


Figure 4: **Reconstruction of faces, cats and cars.**



Experiment



Conclusion

Conclusion

1. Novel Method for 3D reconstruction by using only 2D Image
2. Only using (close to) symmetric Image Data
3. Ignore shadows – leads to inaccurate reconstructions under extreme lighting conditions

Thank you