Learning face reconstruction in the wild, with 3DMM

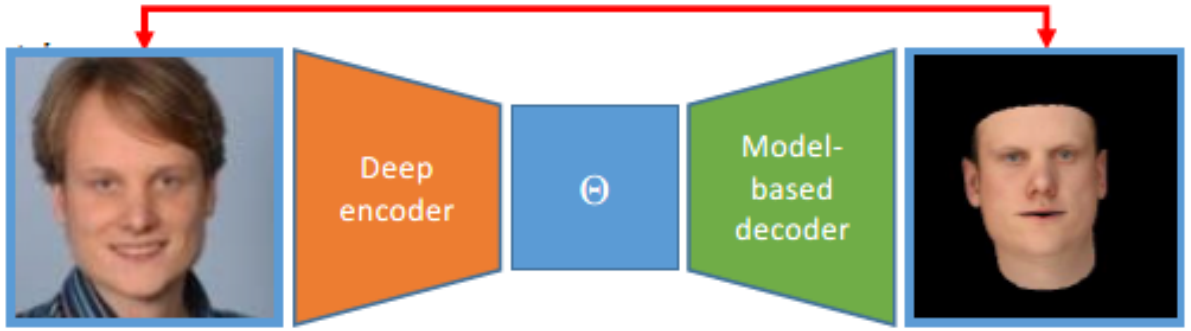
SEMINAR ON 9/23

YUDA QIU

Task

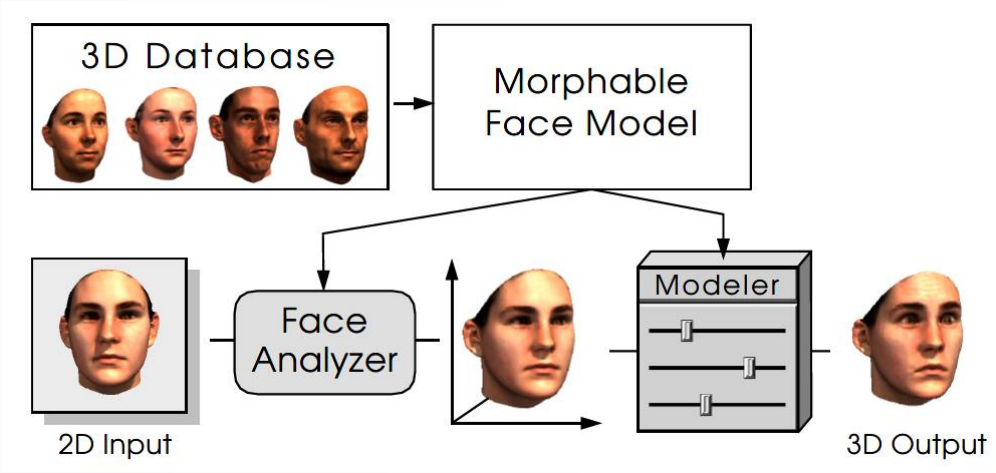
- Input: in-the-wild face image/ video/ image set
- Output: face geometry & face texture

- Lack of dense ground truth
- Ill-posed



Model-based self-supervision (model-based decoder)

3DMM representation



$$S_{model} = \overline{S} + \sum_{i=1}^{m-1} \alpha_i s_i \,, \quad T_{model} = \overline{T} + \sum_{i=1}^{m-1} \beta_i t_i \,, \qquad (1)$$

$\vec{\alpha}, \vec{\beta} \in \Re^{m-1}$. The probability for coefficients $\vec{\alpha}$ is given by

$$p(\vec{\alpha}) \sim exp[-\frac{1}{2} \sum_{i=1}^{m-1} (\alpha_i / \sigma_i)^2], \qquad (2)$$

- PCA to build statistics model
- Analysis-by-synthesis manner

A Morphable Model For The Synthesis Of 3D Faces

Volker Blanz Thomas Vetter

Max-Planck-Institut für biologische Kybernetik,
Tübingen, Germany*

$$\mathbf{x} = (\underbrace{\boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\beta}}_{\text{face}}, \underbrace{\mathbf{T}, \mathbf{t}, \boldsymbol{\gamma}}_{\text{scene}}) \,.$$

Paper lists

- (CVPR 2017) Regressing robust and discriminative 3d morphable models with a very deep neural network
- (ICCV 2017) MoFA: Model-based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction
- (CVPR 2018) Unsupervised Training for 3D Morphable Model Regression.
- (CVPR 2018) Self-supervised multi-level face model learning for monocular reconstruction at over 250 HZ
- **(CVPRW 2019, Best award) Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set.**
- **(CVPR 2019) GANFIT: Generative Adversarial Network Fitting for High Fidelity 3D Face Reconstruction.**
- **(CVPR 2019, oral) FML: Face Model Learning from Videos.**

Regressing Robust and Discriminative 3D Morphable Models with a very Deep Neural Network

Anh Tuấn Trần[1], Tal Hassner[2,3], Iacopo Masi[1], and Gérard Medioni[1]

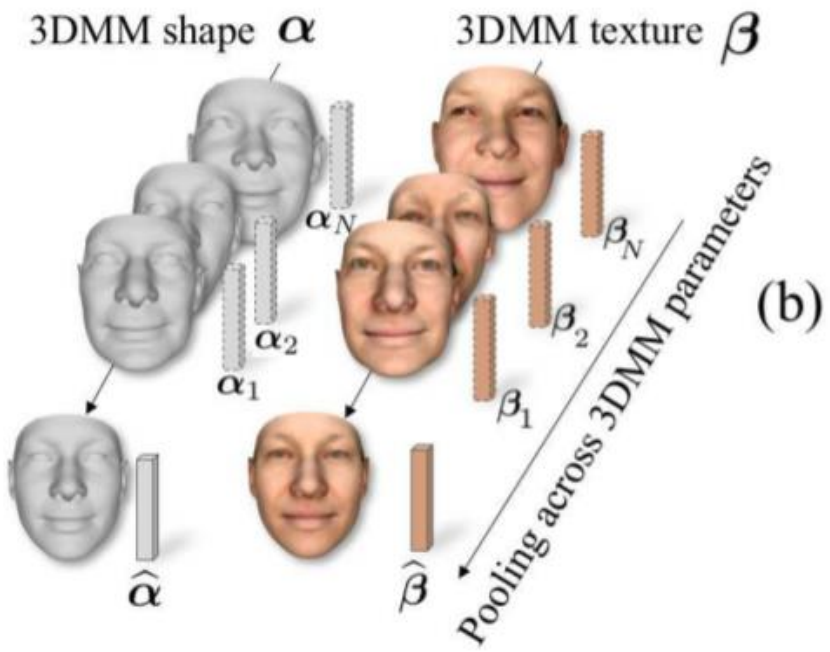
[1] Institute for Robotics and Intelligent Systems, USC, CA, USA

[2] Information Sciences Institute, USC, CA, USA

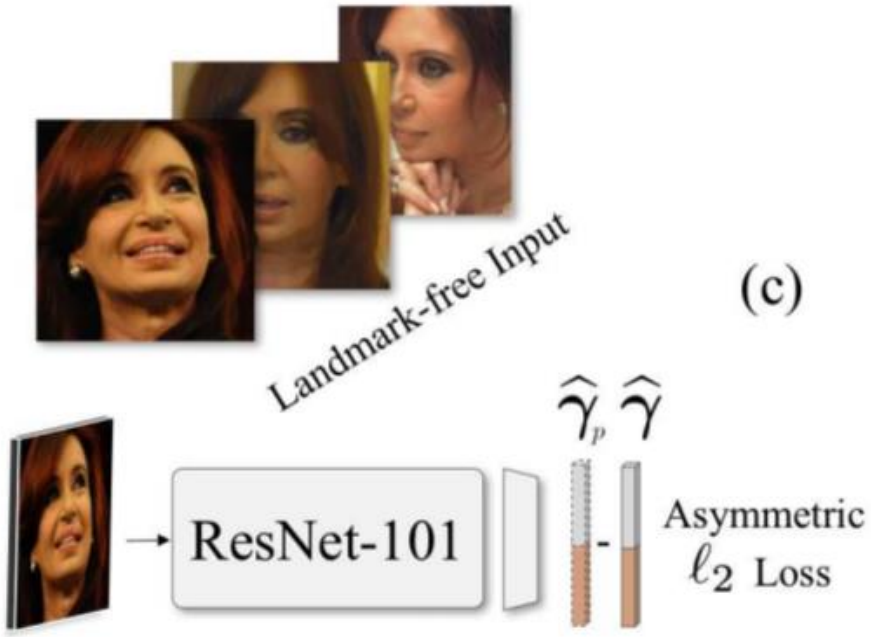
[3] The Open University of Israel, Israel

(a)

Faces in the wild
to train the system



(b)



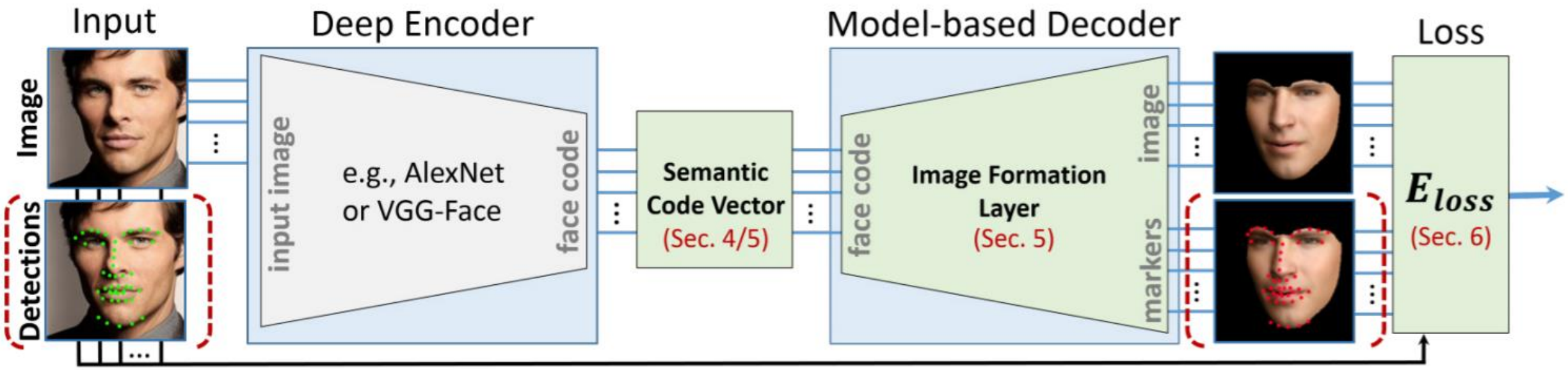
(c)

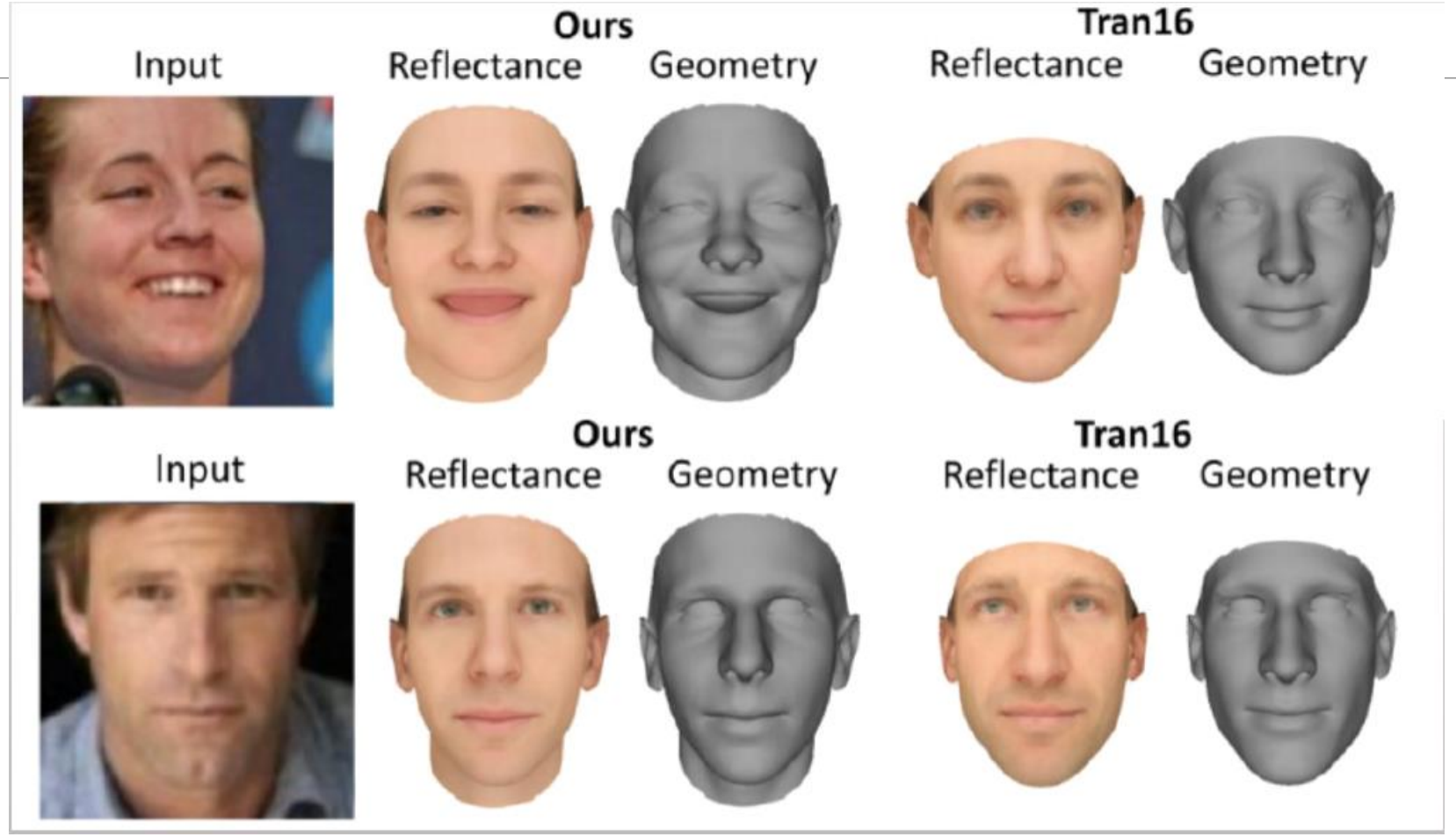
$$\hat{\gamma} = \sum_{i=1}^{N} w_i \cdot \gamma_i \quad \text{and} \quad \sum_{i=1}^{N} w_i = 1, \qquad (2)$$

where w_i are normalized per-image confidences provided by the CLNF facial landmark detector.

MoFA: Model-based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction

Ayush Tewari[1] Michael Zollhöfer[1] Hyeongwoo Kim[1] Pablo Garrido[1]
Florian Bernard[1,2] Patrick Pérez[3] Christian Theobalt[1]
[1]Max-Planck-Institute for Informatics [2] LCSB, University of Luxembourg [3]Technicolor


Input
Deep Encoder
Model-based Decoder
Loss
Image
Detections
input image
e.g., AlexNet
or VGG-Face
face code
Semantic
Code Vector
(Sec. 4/5)
face code
Image Formation
Layer
(Sec. 5)
image
markers
E_{loss}
(Sec. 6)

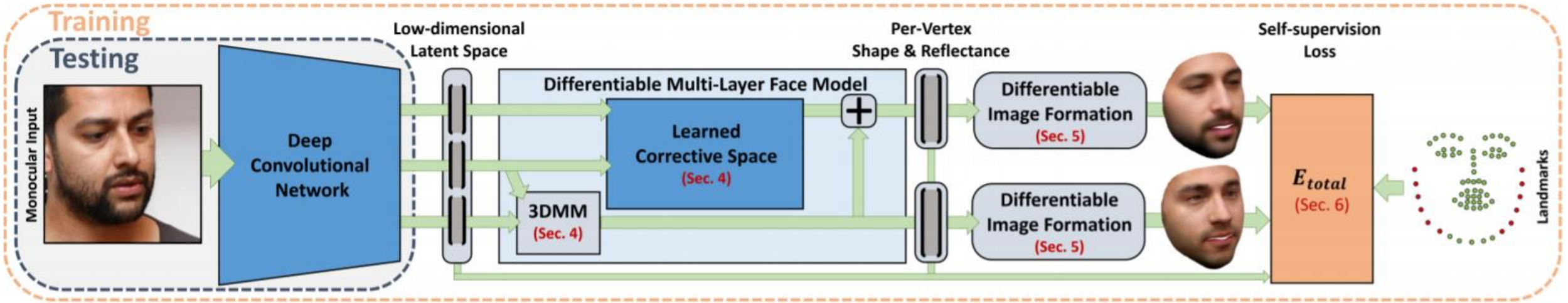

Input
Ours
Reflectance
Geometry
Tran16
Reflectance
Geometry
Input
Ours
Reflectance
Geometry
Tran16
Reflectance
Geometry

Self-supervised Multi-level Face Model Learning for Monocular Reconstruction at over 250 Hz

CVPR 2018 (Oral)

A. Tewari [1,2] M. Zollhöfer [1,2,3] F. Bernard [1,2] P. Garrido [1,2] H. Kim [1,2] P. Perez [4] C.Theobalt [1,2]

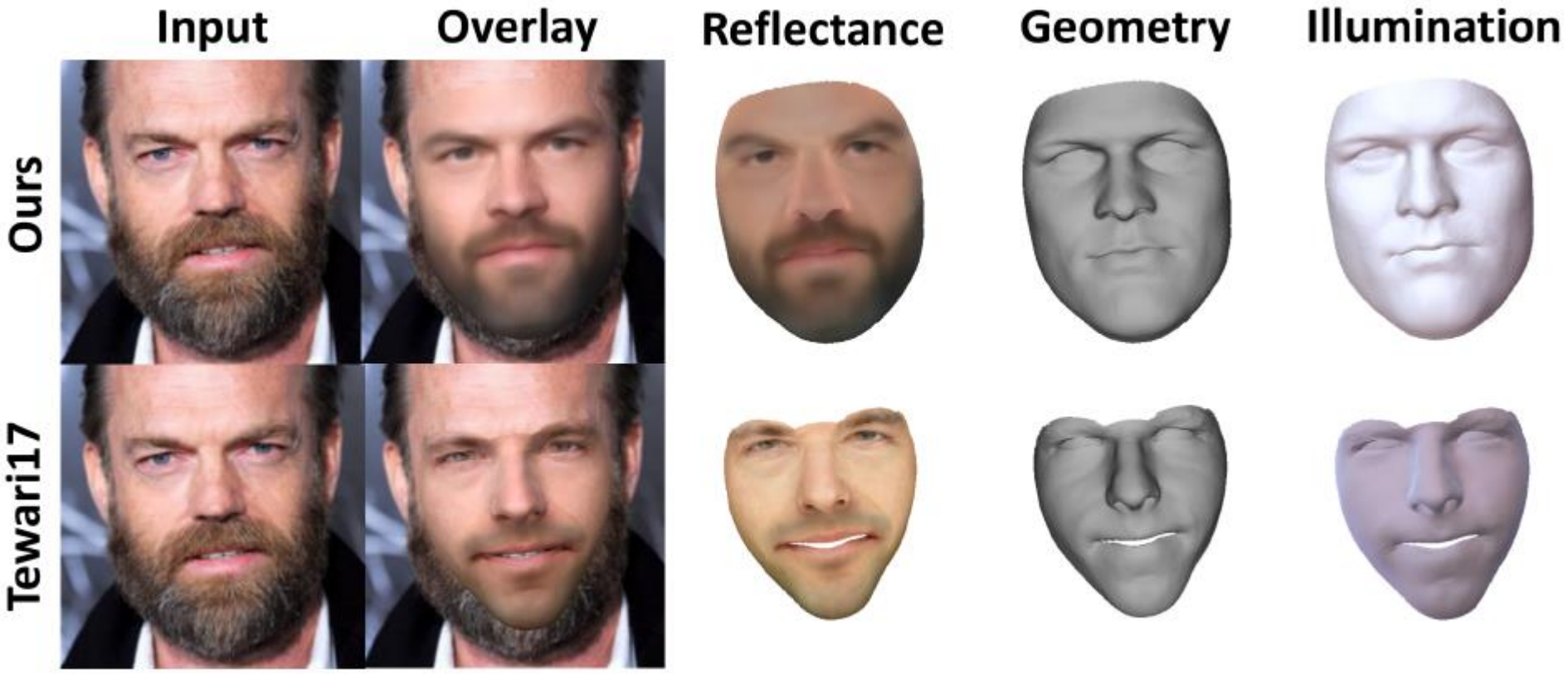
[1]MPI Informatics [2]Saarland Informatics Campus [3]Stanford University [4]Technicolor



$$\mathbf{v}^{\mathrm{f}}(\mathbf{x}_g) = \mathbf{v}^{\mathrm{b}}(\boldsymbol{\alpha}) + \mathcal{F}_g(\boldsymbol{\delta}_g|\Theta_g) \in \mathbb{R}^{3N}(\text{geometry}),$$

$$\mathbf{r}^{\mathrm{f}}(\mathbf{x}_r) = \mathbf{r}^{\mathrm{b}}(\boldsymbol{\beta}) + \mathcal{F}_r(\boldsymbol{\delta}_r|\Theta_r) \in \mathbb{R}^{3N}(\text{reflectance}),$$

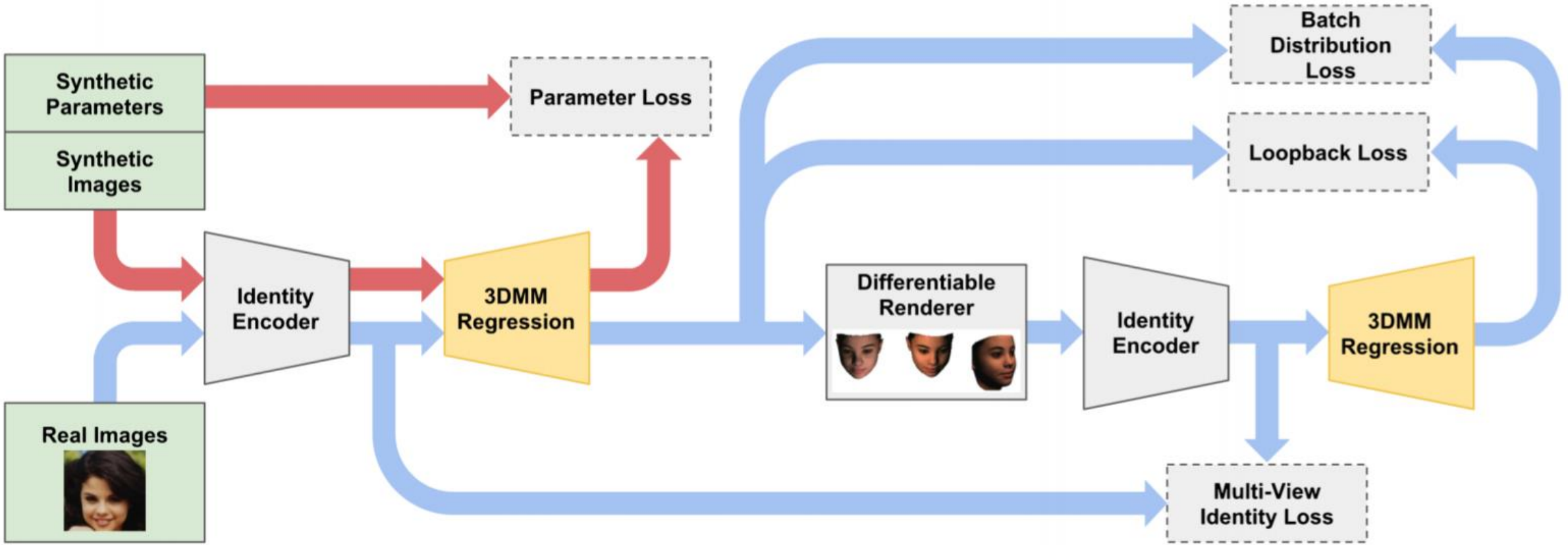
$$\mathbf{x} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}_g, \boldsymbol{\delta}_r, \mathbf{R}, \mathbf{t}, \boldsymbol{\gamma}, \Theta_g, \Theta_r) \in \mathbb{R}^{257+2C+|\Theta_g|+|\Theta_r|}.$$


Input
Overlay
Reflectance
Geometry
Illumination
Ours
Tewari17

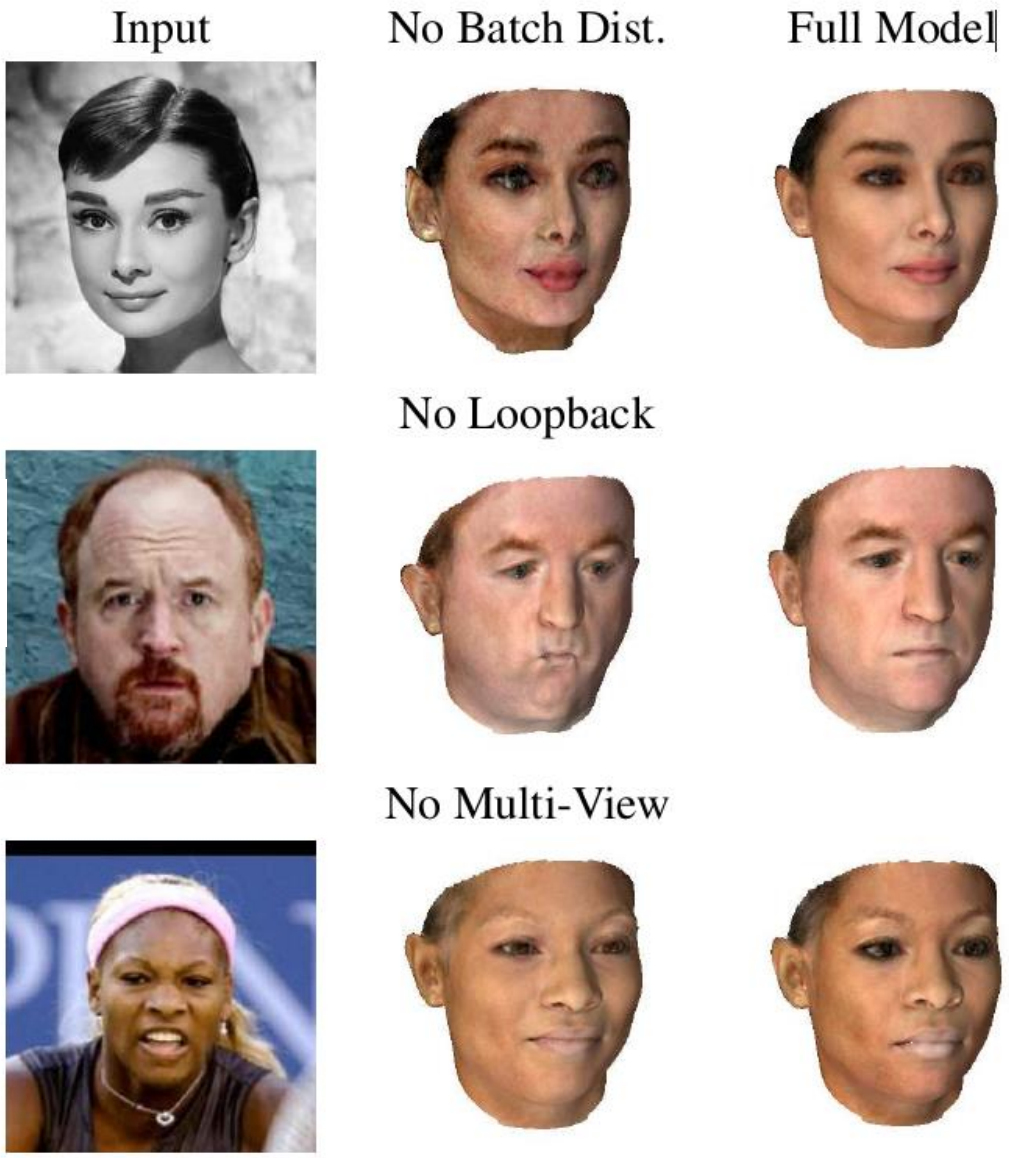
Unsupervised Training for 3D Morphable Model Regression

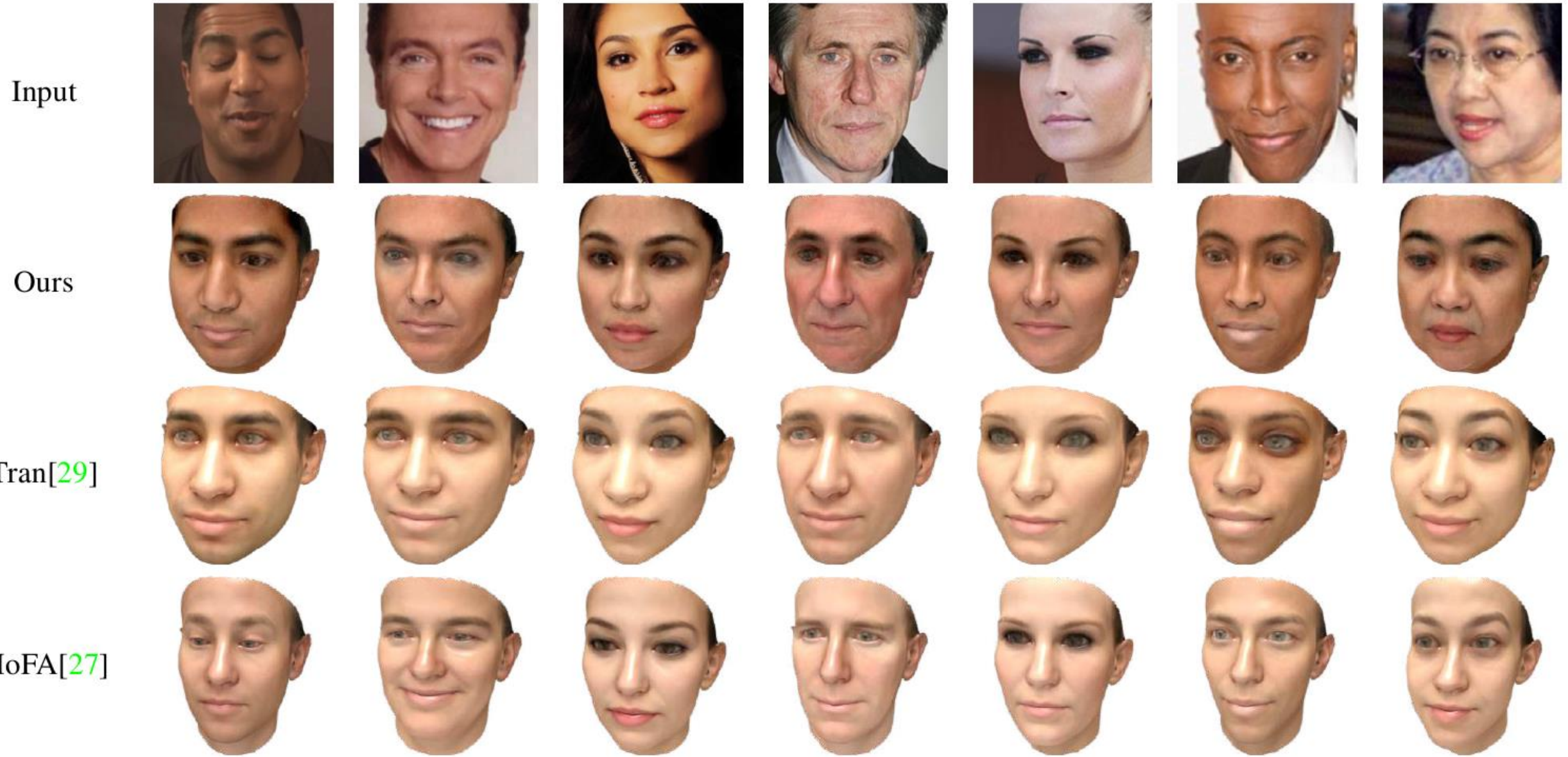
Kyle Genova[1,2] Forrester Cole[2] Aaron Maschinot[2] Aaron Sarna[2] Daniel Vlasic[2] William T. Freeman[2,3]

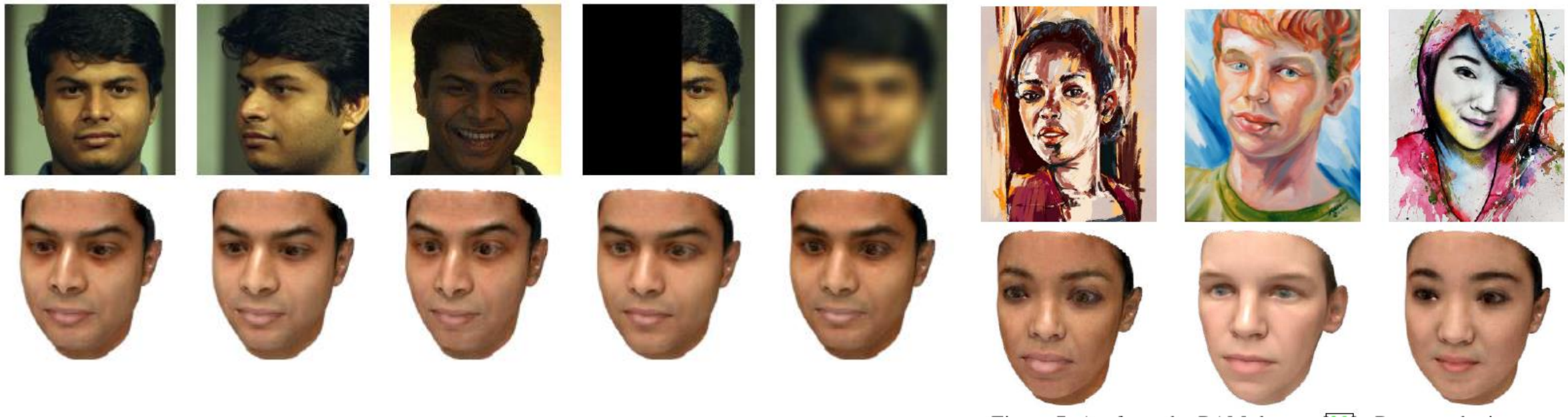
[1]Princeton University [2]Google Research [3]MIT CSAIL


Synthetic Parameters
Synthetic Images
Real Images
Parameter Loss
Identity Encoder
3DMM Regression
Differentiable Renderer
Batch Distribution Loss
Loopback Loss
Identity Encoder
3DMM Regression
Multi-View Identity Loss

$$L = L_{param} + L_{id} + \omega_{batch} L_{batch} + \omega_{loop} L_{loop}$$




Input
Ours
Tran[29]
MoFA[27]



Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set

Yu Deng*[1,2] Jiaolong Yang[2] Sicheng Xu[3,2] Dong Chen[2] Yunde Jia[3] Xin Tong[2]

[1]Tsinghua University [2]Microsoft Research Asia [3]Beijing Institute of Technology

{v-denyu,jiaoyan,doch,xtong}@microsoft.com, {xusicheng,jiayunde}@bit.edu.cn

(a) Our framework

R-Net

(Concat)

C-Net

Coefficients:

α - identity

β - expression

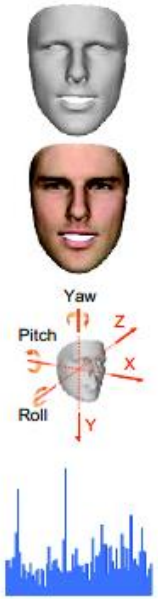
δ - texture

p - pose

γ - lighting

Confidence:

c - identity confidence

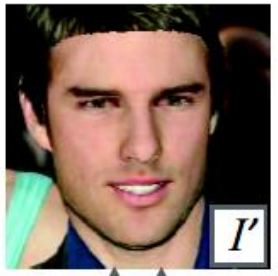


(b) Training pipeline for single image 3D face reconstruction

I

R-Net

Analytic image generation



I'

Skin Estimation

Landmark detection

A

$\mathbf{q}$

Image-level loss

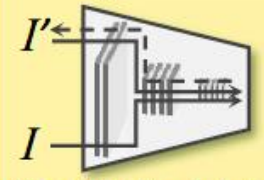
$(I' - I) \odot A$

Robust photometric loss

$\mathbf{q}' - \mathbf{q}$

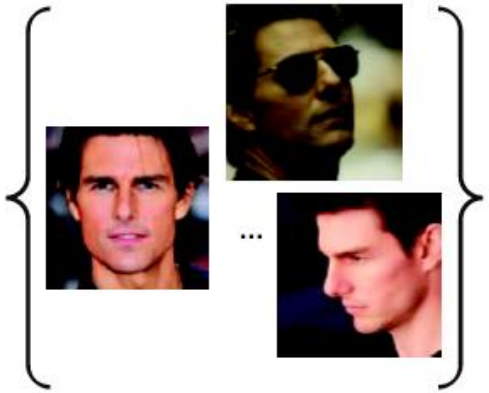
Landmark location loss

Perception-level loss



Deep identity feature loss

(c) Training pipeline for multi-image 3D face reconstruction with shape aggregation



R-Net

C-Net

$\alpha_i \mid c_i$

β_i

δ_i

p_i

γ_i

Analytic identity aggregation

α_{aggr}

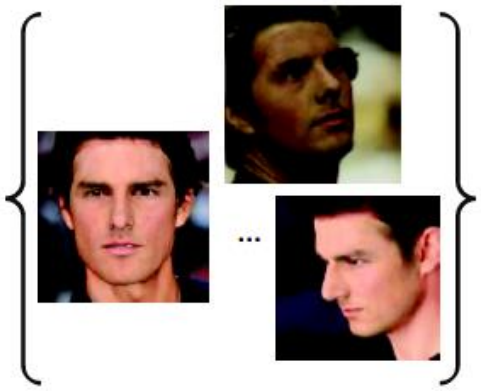
β_i

δ_i

p_i

γ_i

Analytic image generation



Losses in **(b)** (all images)


5.57
3.59
2.24
1.55
1.48
curtis
5.59
5.14
4.54
3.37
2.50
7.14
3.77
3.27
2.53
1.79

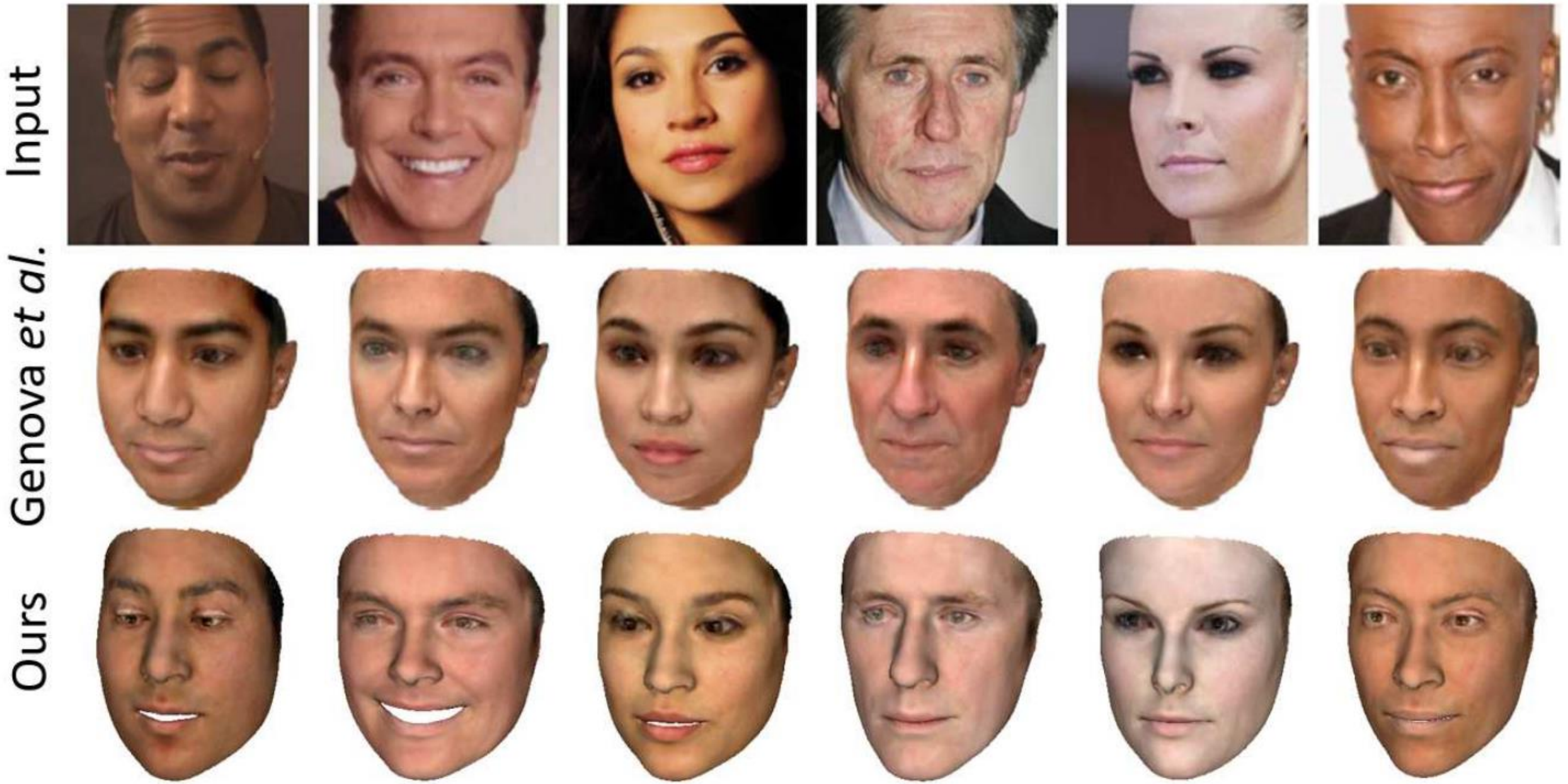
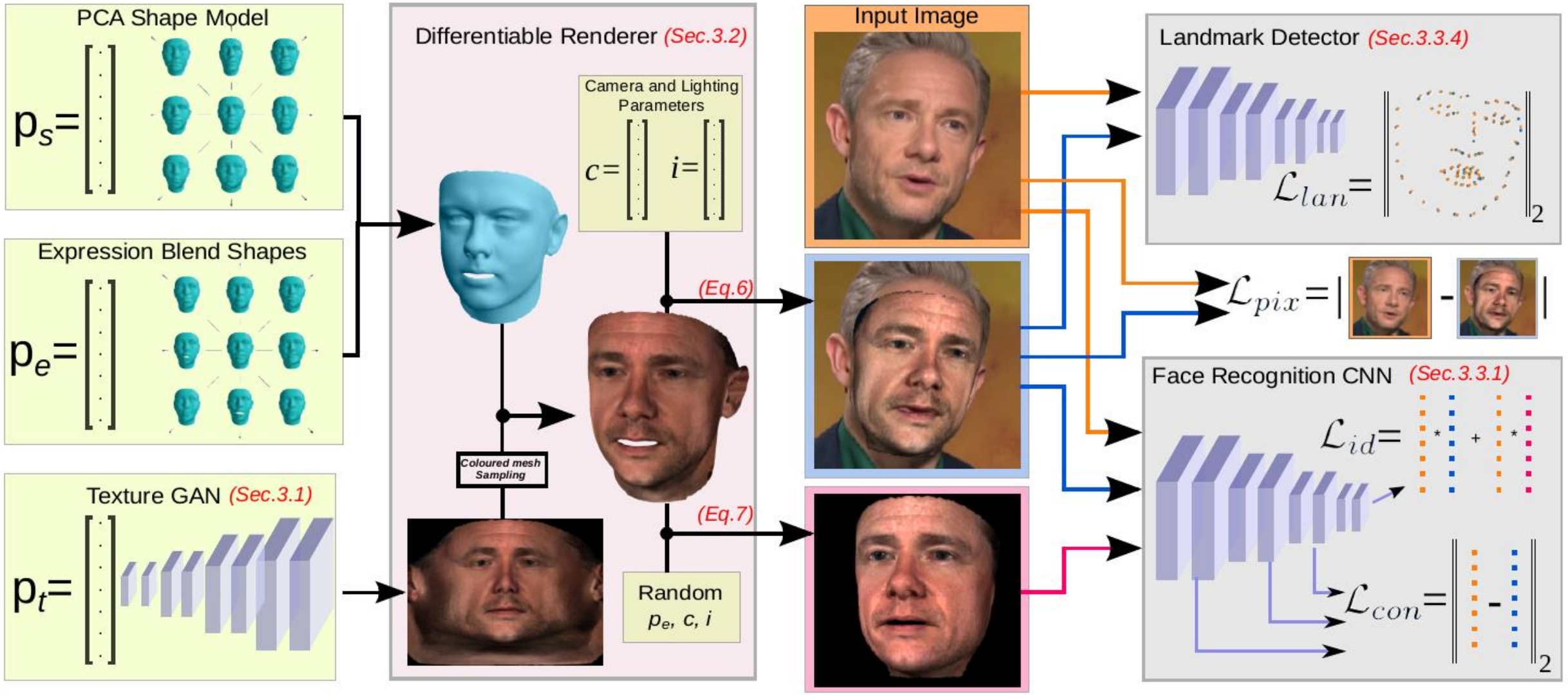

Input
Genova *et al.*
Ours

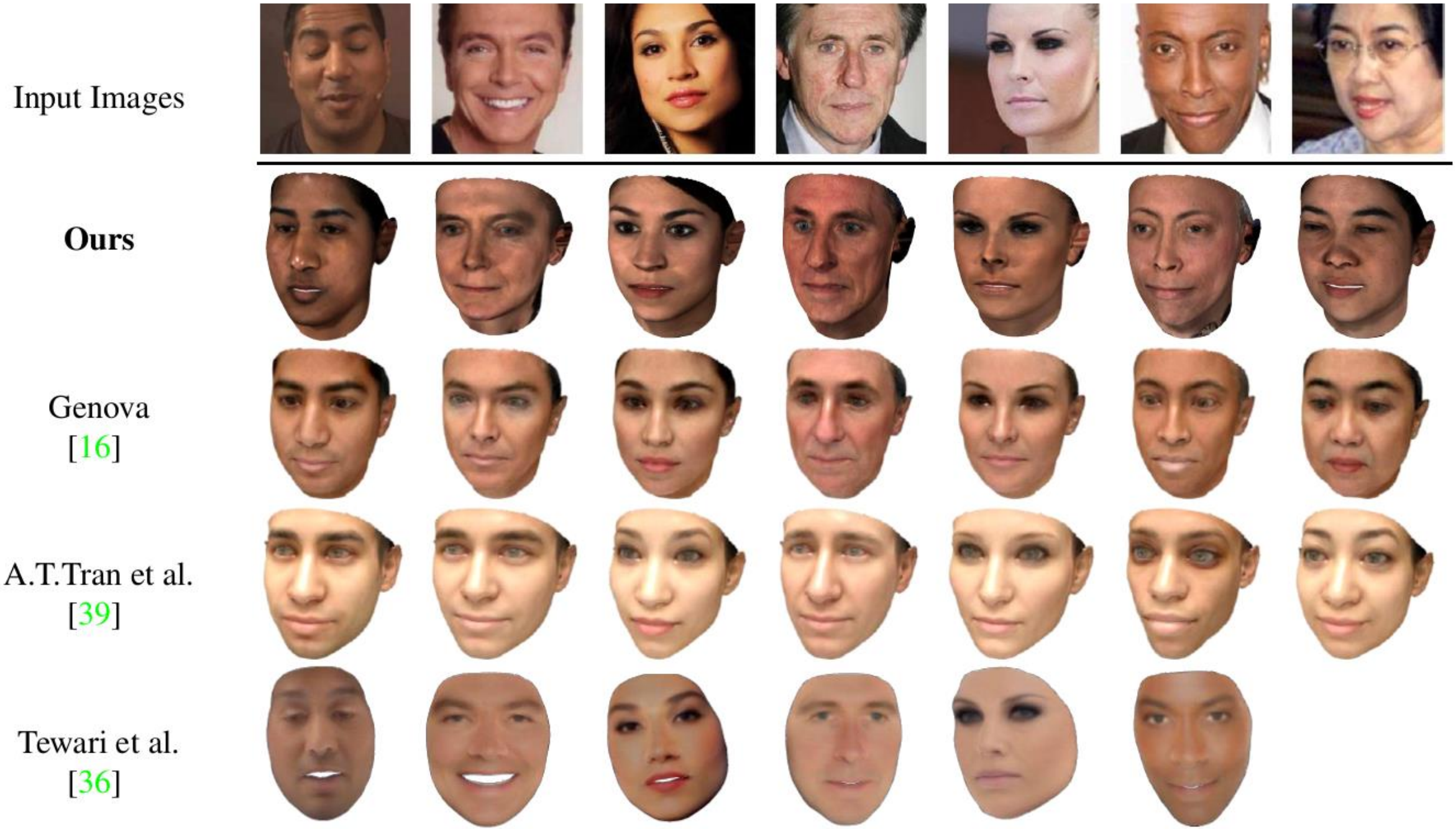
Table 4. Multi-image reconstruction errors on MICC rendered images with different aggregation strategies (see text for details).

Shape error mean	1.97±0.70
Shape averaging	1.78±0.59
Our S1: Global Aggr. with c^j	1.71±0.56
Our S2: Global Aggr. with $\sum_i \boldsymbol{c}_i^j$	1.70±0.55
Our S3: Max Conf. $j = \arg\max_j \sum \boldsymbol{c}_i^j$	1.71±0.50
Our S4: Elementwise Aggr. with $\boldsymbol{c}^j$	**1.67±0.54**

GANFIT: Generative Adversarial Network Fitting for High Fidelity 3D Face Reconstruction

Baris Gecer[1,2], Stylianos Ploumpis[1,2], Irene Kotsia[3], and Stefanos Zafeiriou[1,2]
[1]Imperial College London
[2]FaceSoft.io
[3]University of Middlesex

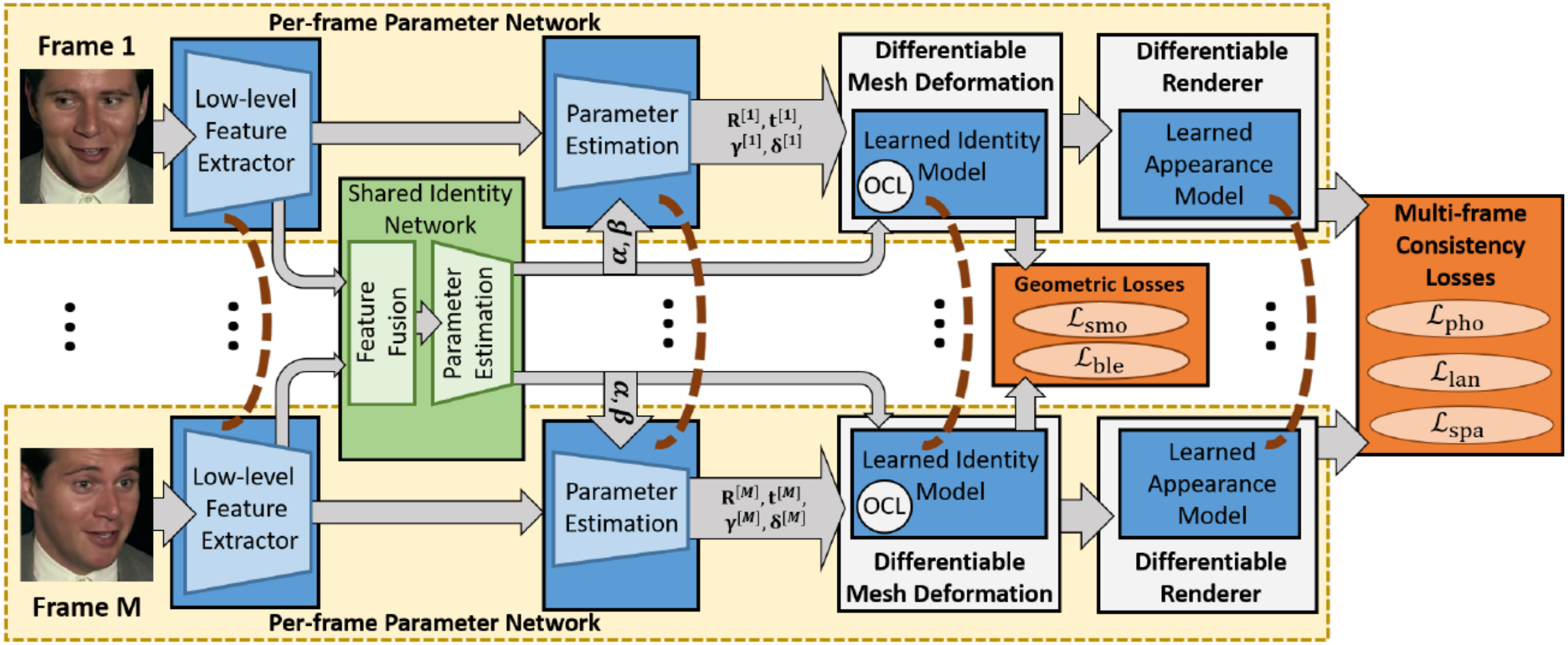

PCA Shape Model
$p_s=$
Expression Blend Shapes
$p_e=$
Texture GAN (Sec.3.1)
$p_t=$
Differentiable Renderer (Sec.3.2)
Camera and Lighting Parameters
$c=$ $i=$
(Eq.6)
Coloured mesh Sampling
(Eq.7)
Random
p_e, c, i
Input Image
Landmark Detector (Sec.3.3.4)
$\mathcal{L}_{lan}=$
$\mathcal{L}_{pix}=$
Face Recognition CNN (Sec.3.3.1)
$\mathcal{L}_{id}=$
$\mathcal{L}_{con}=$

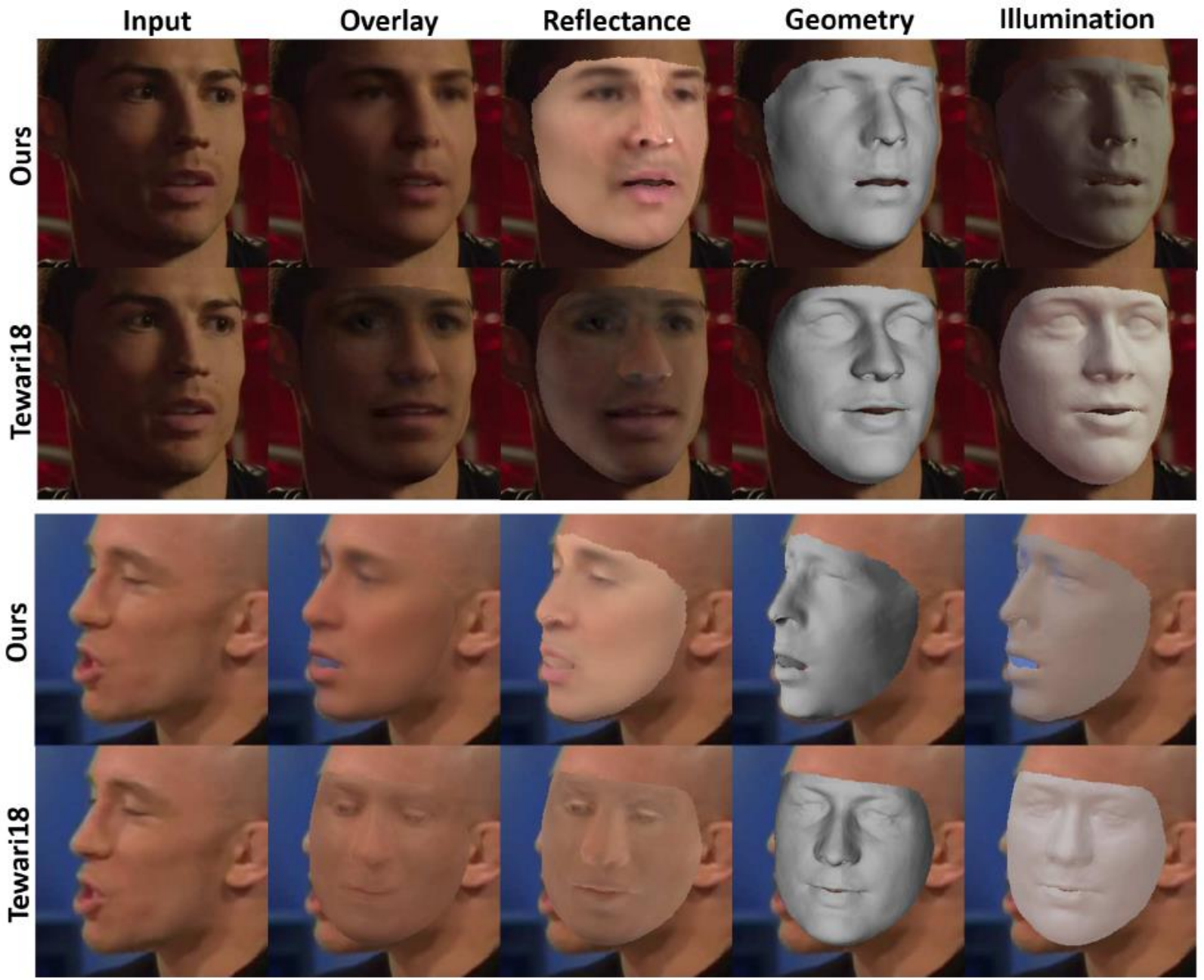

Input Images
Ours
Genova
[16]
A.T.Tran et al.
[39]
Tewari et al.
[36]

FML: Face Model Learning from Videos

Ayush Tewari[1] Florian Bernard[1] Pablo Garrido[2] Gaurav Bharaj[2] Mohamed Elgharib[1]
Hans-Peter Seidel[1] Patrick Pérez[3] Michael Zollhöfer[4] Christian Theobalt[1]

[1]MPI Informatics, Saarland Informatics Campus [2]Technicolor [3]Valeo.ai [4]Stanford University


Per-frame Parameter Network
Frame 1
Low-level Feature Extractor
Parameter Estimation
$\mathbf{R}^{[1]}, \mathbf{t}^{[1]}, \boldsymbol{\gamma}^{[1]}, \boldsymbol{\delta}^{[1]}$
Differentiable Mesh Deformation
Learned Identity Model
OCL
Differentiable Renderer
Learned Appearance Model
Shared Identity Network
Feature Fusion
Parameter Estimation
α, β
α', β'
Geometric Losses
$\mathcal{L}_{\text{smo}}$
$\mathcal{L}_{\text{ble}}$
Multi-frame Consistency Losses
$\mathcal{L}_{\text{pho}}$
$\mathcal{L}_{\text{lan}}$
$\mathcal{L}_{\text{spa}}$
Frame M
Low-level Feature Extractor
Parameter Estimation
$\mathbf{R}^{[M]}, \mathbf{t}^{[M]}, \boldsymbol{\gamma}^{[M]}, \boldsymbol{\delta}^{[M]}$
Learned Identity Model
OCL
Differentiable Mesh Deformation
Learned Appearance Model
Differentiable Renderer
Per-frame Parameter Network


Input
Overlay
Reflectance
Geometry
Illumination
Ours
Tewari18
Ours
Tewari18

Q&A