JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

StyleIPSB: Identity-Preserving Semantic Basis of StyleGAN for High Fidelity Face Swapping

Diqiong Jiang[1] , Dan Song[2*], Ruofeng Tong[1*], Min Tang[1]

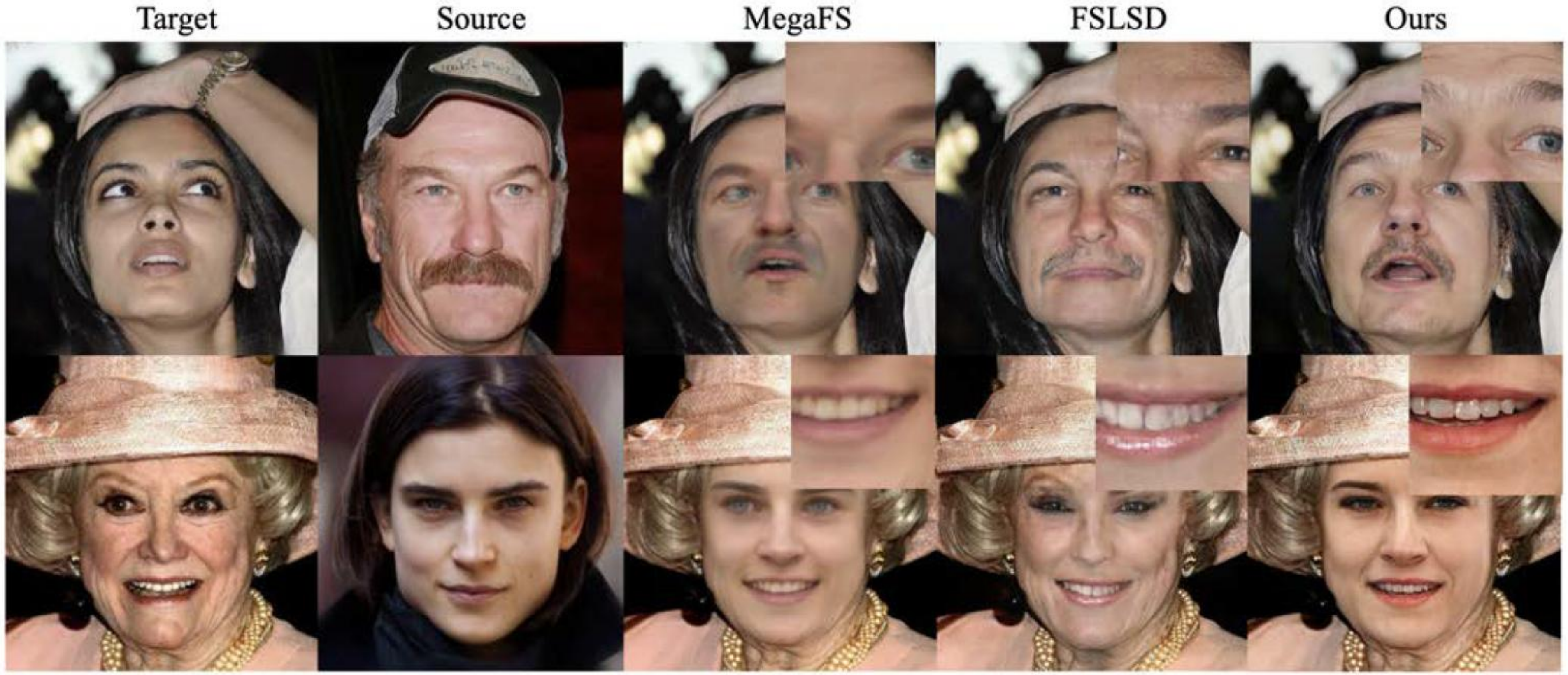
[1]Zhejiang University, China [2]Tianjin University, China

https://github.com/a686432/StyleIPSB

Paper Tag: TUE-AM-034

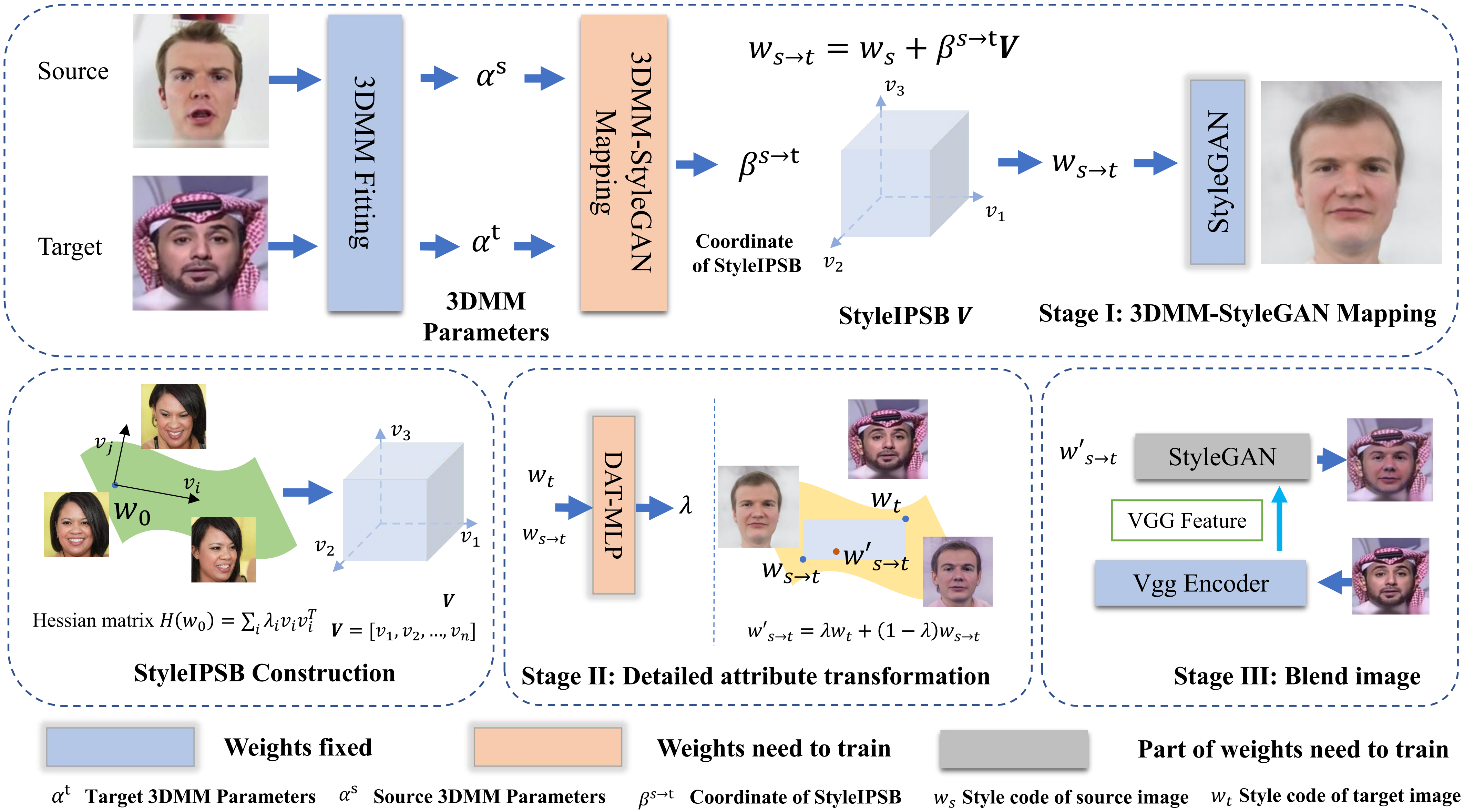
Quick Preview

StyleIPSB: A New Linear Space within StyleGAN



- ✓ Pore-level Details
- ✓ Preserved Identity
- ✓ Powerful Representation

Quick Preview



Quick Preview

Quantitative experiments on FaceForensics++ dataset

Method	ID Retri.(%) ↑	Exp Err. ↓	Pose Err. ↓
FaceSwap [3]	72.69	2.89	2.58
Deepfakes [1]	88.39	3.33	4.64
FaceShifter [22]	90.68	2.82	2.55
MegaFS [51]	90.83	2.92	2.64
FSLDS [47]	90.05	2.79	**2.46**
Ours	**95.05**	**2.23**	3.58

Quantitative experiments on CelebAHQ dataset

	FID ↓	Exp ↓	Pose ↓	ID similarity ↑
MageFS [51]	22.03	2.85	**0.043**	0.4837
RAFS [46]	13.25	3.15	-	0.5232
FSLDS [47]	10.01	2.99	0.053	0.4761
Ours	**9.37**	**2.75**	0.078	**0.5378**

Quick Preview

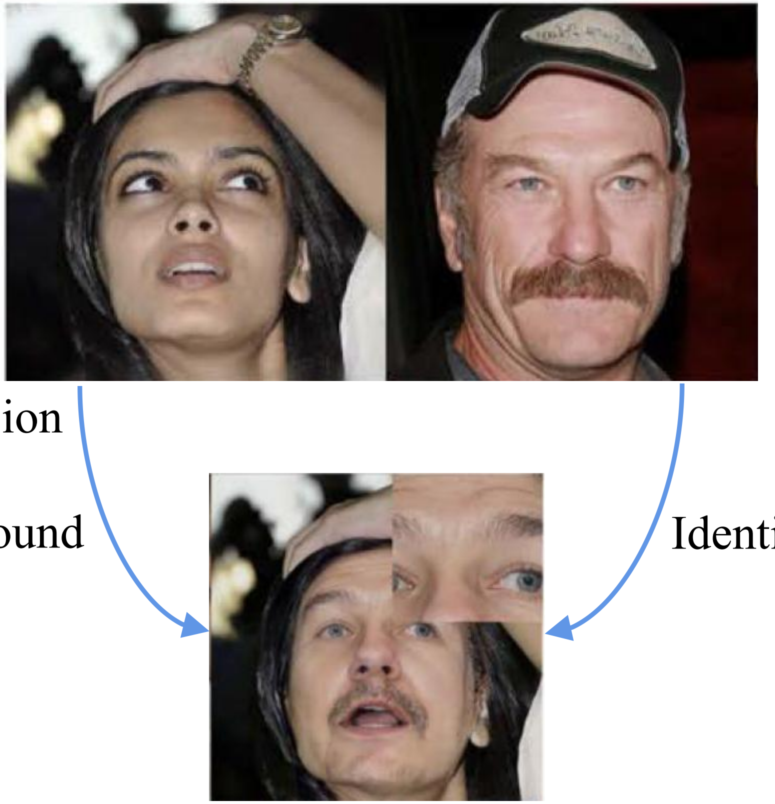
✓ StyleIPSB can represent various poses, expressions, and illuminations while preserving identity.

✓ StyleIPSB can generate pore-level details.

Motivation

✓ Face swapping

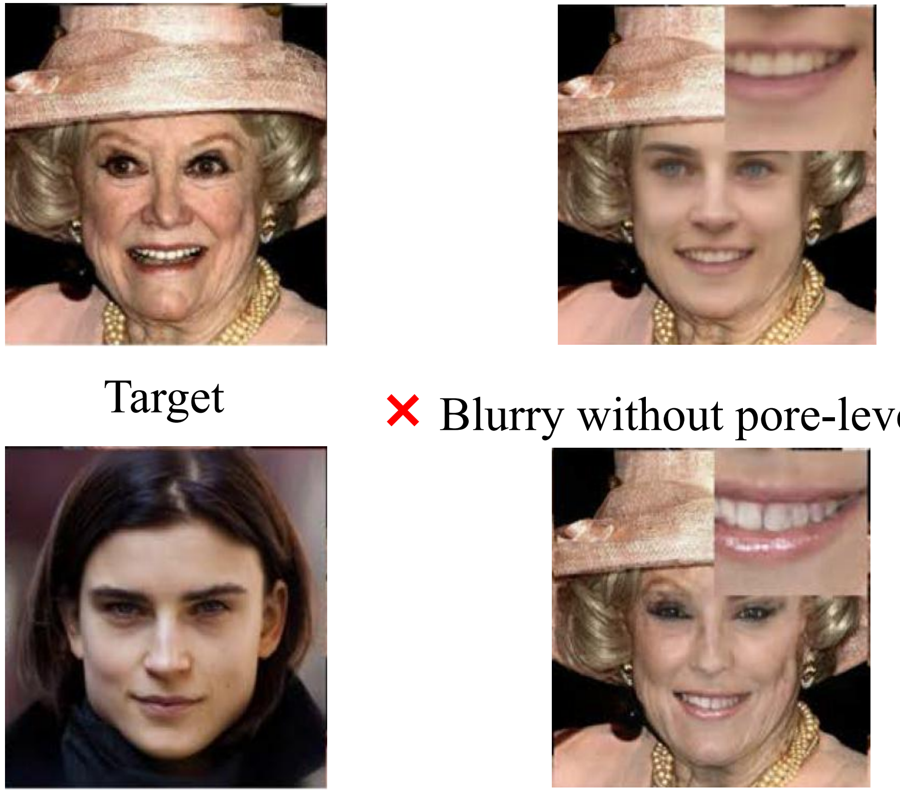
Target Source



Expression
Pose
Background
Hair
...

Identity

✓ Challenges



Target

✗ Blurry without pore-level details

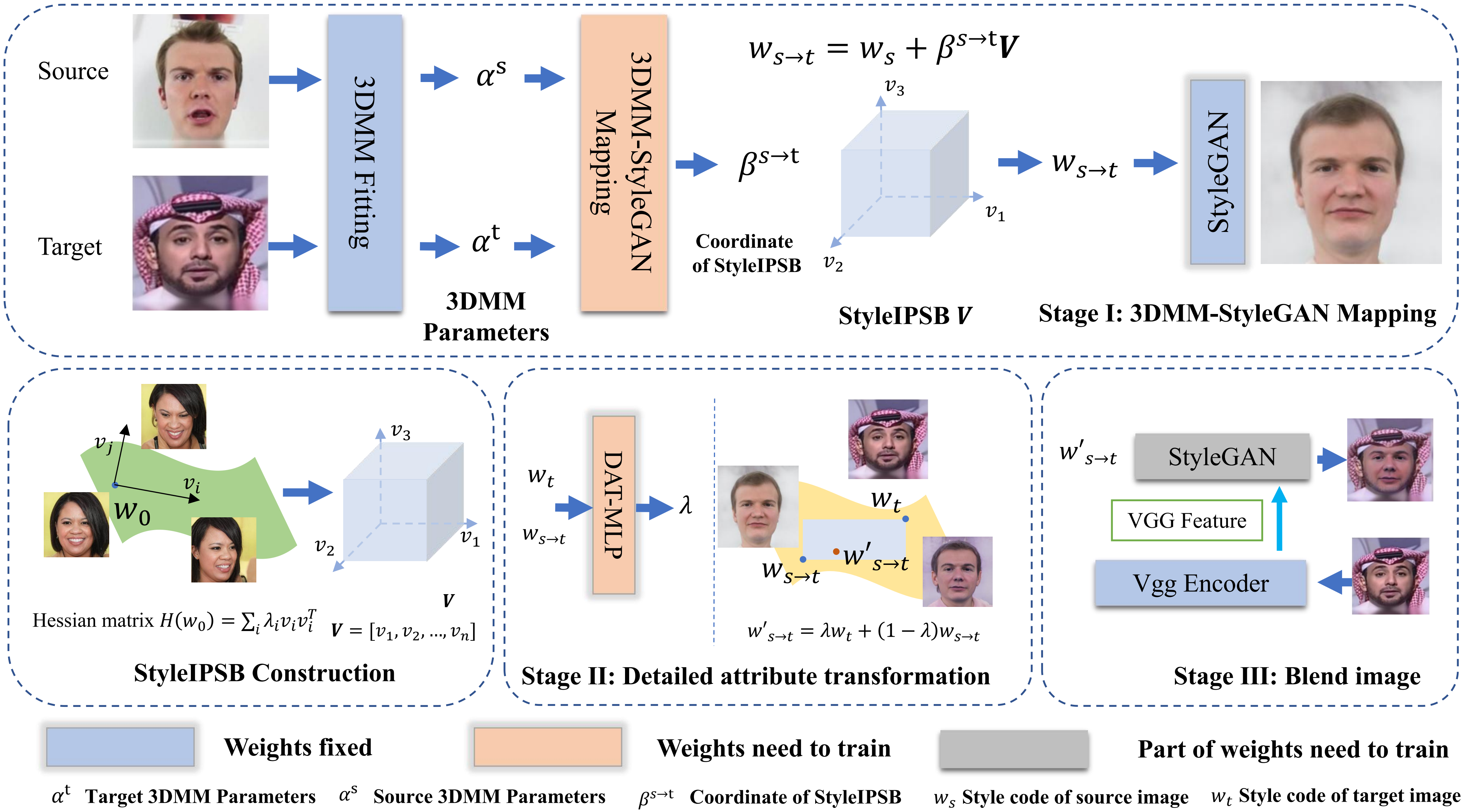
Source

✗ Fail to preserve identity

Contribution

- ✓ We propose a novel method of establishing identity preserving semantic bases of StyleGAN called StyleIPSB. The face image, generated by the linear space of StyleIPSB, remains pore-level details and identity-preserving.
- ✓ The proposed StyleGAN-3DMM mapping network serves as the bridge to narrow the gap between 3DMM and StyleIPSB, which can take advantage of the prominent semantic variance of 3DMM and the identity preserving and high-fidelity of styleIPSB.
- ✓ We propose the face swapping framework based on StyleIPSB and StyleGAN-3DMM mapping network. Extensive results show our method outperforms others in detail-preserving and identity-preserving.

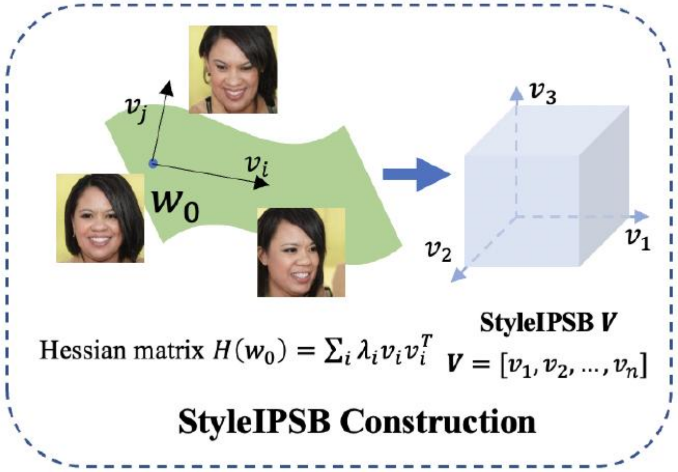
Method


Source
Target
3DMM Fitting
α^s
α^t
3DMM Parameters
3DMM-StyleGAN Mapping
$w_{s\to t} = w_s + \beta^{s\to t} V$
$\beta^{s\to t}$
Coordinate of StyleIPSB
v_1
v_2
v_3
StyleIPSB V
$w_{s\to t}$
StyleGAN
Stage I: 3DMM-StyleGAN Mapping
v_j
v_i
w_0
V
Hessian matrix $H(w_0) = \sum_i \lambda_i v_i v_i^T$
$V = [v_1, v_2, ..., v_n]$
StyleIPSB Construction
w_t
$w_{s\to t}$
DAT-MLP
λ
$w'_{s\to t}$
$w'_{s\to t} = \lambda w_t + (1-\lambda) w_{s\to t}$
Stage II: Detailed attribute transformation
StyleGAN
VGG Feature
Vgg Encoder
Stage III: Blend image
Weights fixed
Weights need to train
Part of weights need to train
α^t Target 3DMM Parameters
α^s Source 3DMM Parameters
$\beta^{s\to t}$ Coordinate of StyleIPSB
w_s Style code of source image
w_t Style code of target image

Method-StyleIPSB Construction

The properties of the proposed StyleIPSB:

- ✓ By ensuring the regressed style code within the W+ space of StyleGAN, we can more easily generate images with pore-level details.
- ✓ When changing the coordinates of the StyleIPSB, the identity remains preserved as much as possible.
- ✓ StyleIPSB can represent various poses, expressions, and illuminations.



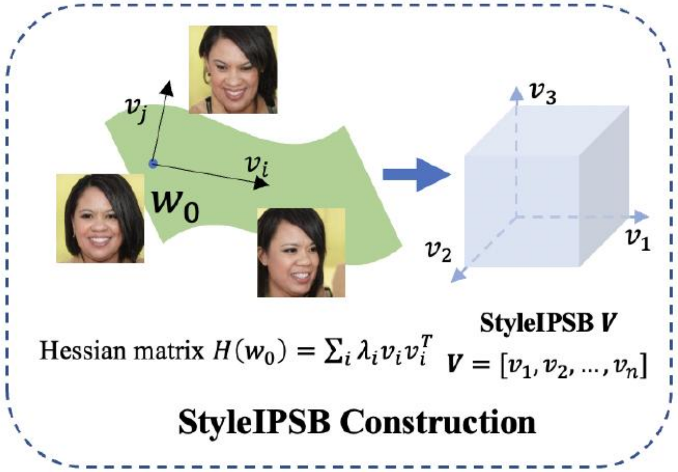
Method-StyleIPSB Construction

✓ The proposed identity-preserving distance metric:

$$D_p(w_1, w_2) = \frac{||M(G(w_1))_p - M(G(w_2))_p||^2}{L_{id}(G(w_1), G(w_2))}$$

$$D_e(w_1, w_2) = \frac{||M(G(w_1))_e - M(G(w_2))_e||^2}{L_{id}(G(w_1), G(w_2))}$$

$$D_i(w_1, w_2) = \frac{||M(G(w_1))_i - M(G(w_2))_i||^2}{L_{id}(G(w_1), G(w_2))}$$



StyleGAN network: $G : R^n \longrightarrow R^{H\times W\times 3}, w \mapsto I$ 3DMM Fitting network: $M : R^{H\times W\times 3} \longrightarrow R^{n'}, I \mapsto p$

✓ We decompose the Hessian matrix to find the direction with the fastest distance metric change in the W+ space.

$$D^2(w_0, w_0 + \delta w) \approx ||\delta w||_H^2 = \delta_w^T H(w_0)\delta w$$

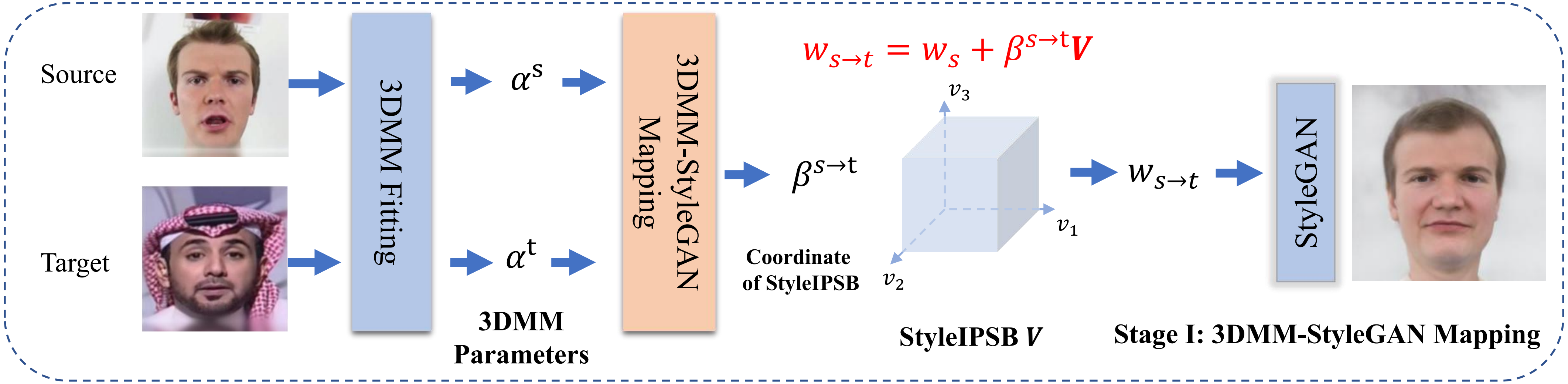
w_0 is randomly sampled, $w_0 + \delta w$ is the point near w_0.

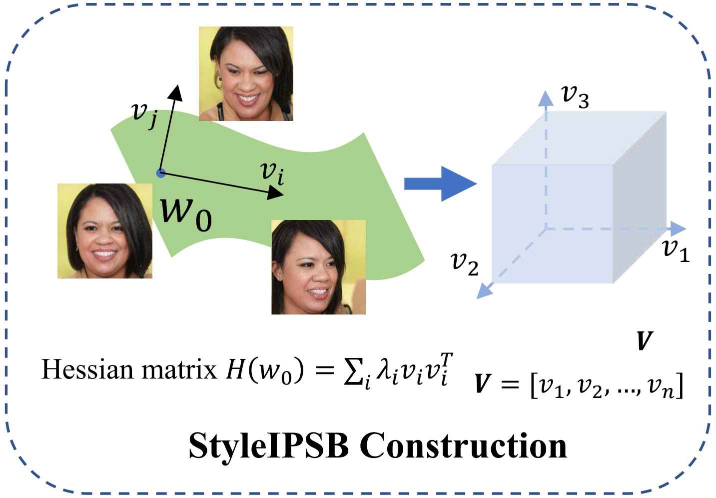
✓ The attributes change fast but identity changes slowly along the found direction.

Method-Stage 1: 3DMM-StyleGAN Mapping

Weights fixed **Weights need to train** **Part of weights need to train**

α^t **Target 3DMM Parameters** α^s **Source 3DMM Parameters** $\beta^{s\to t}$ **Coordinate of StyleIPSB** w_s **Style code of source image** w_t **Style code of target image**





✓ Stage 1 learns the mapping from 3D Morphable Model (3DMM) parameters, which capture the prominent semantic variance, to the coordinates of StyleIPSB that show higher identity preserving and fidelity.

Method-Stage 1: 3DMM-StyleGAN Mapping

✓ The loss function for the first stage:

$$L = L_{id} + \varepsilon_{attr} L_{attr}$$

✓ We reconstruct the 3D face from the transferred face and the target face, and then compare their difference in 3D face geometries and rendered images.

$$L_{attr}(\alpha^s, \alpha^t, \alpha^{s\to t}) = L_{geo}(\alpha^s, \alpha^{s\to t}) + L_{render}(\alpha^s, \alpha^t, \alpha^{s\to t})$$

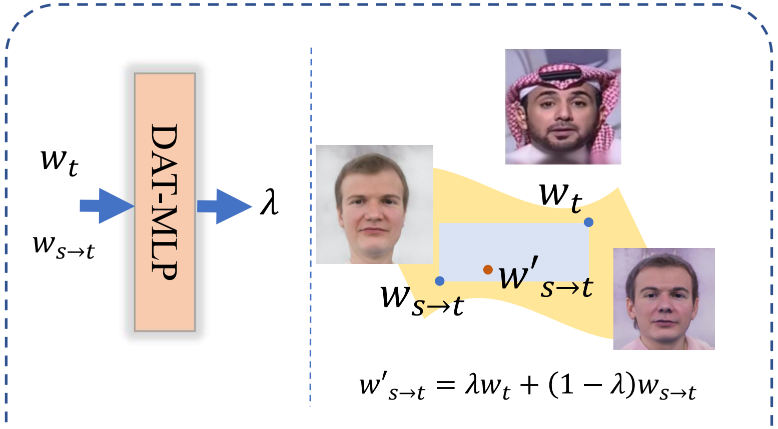
- The geometric term L_{geo} uses the L_2 loss between two face meshes:

$$L_{geo}(\alpha^s, \alpha^{s\to t}) = \frac{1}{N} \left\| G_{3DMM}(\alpha_s^s, \alpha_e^t, \alpha_p^t) - G_{3DMM}(\alpha_s^{s\to t}, \alpha_e^{s\to t}, \alpha_p^{s\to t}) \right\|_2$$

- The render term L_{render} uses the L_1 loss between two rendered images:

$$L_{render}(\alpha^s, \alpha^t, \alpha^{s\to t}) = \left\| R(\alpha_s^s, \alpha_e^t, \alpha_a^t, \alpha_i^t, \alpha_p^t) - R(\alpha_s^{s\to t}, \alpha_e^{s\to t}, \alpha_a^t, \alpha_i^{s\to t}, \alpha_p^{s\to t}) \right\|_1$$

Method-Stage 2: Detailed Attribute Transformation



Stage II: Detailed attribute transformation

- ✓ Detailed attribute transformation aims to transfer attributes beyond the 3DMM expressive capabilities.
- ✓ It contains **non-identity** attributes of the target image.

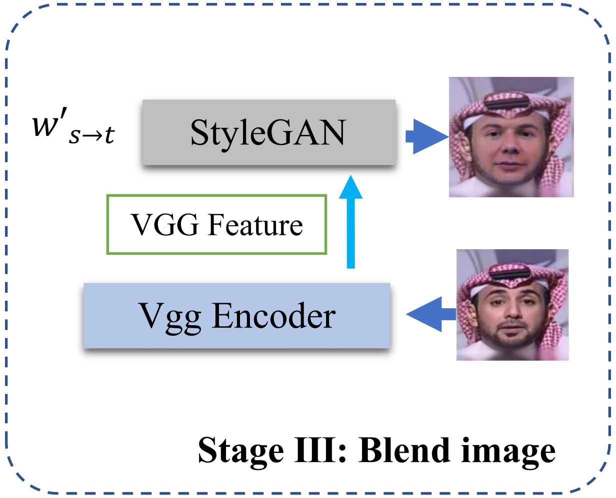
$$w'_{s\to t} = \lambda w_{s\to t} + (1-\lambda) w_t$$

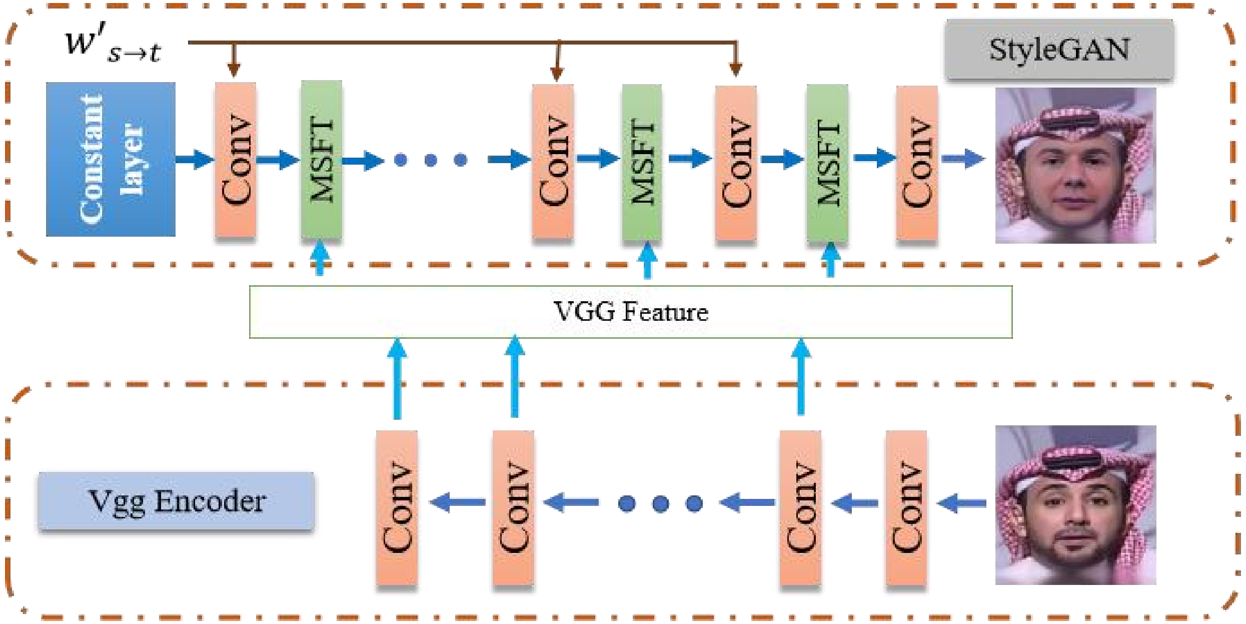
- ✓ The loss function for the second stage：

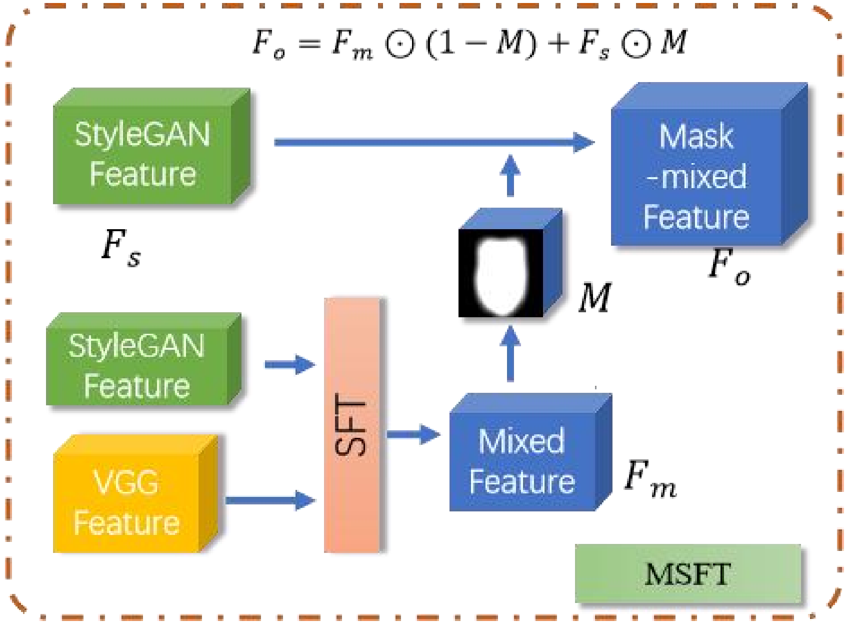
$$L = L_{id} + \varepsilon_p L_p + \varepsilon_{attr} L_{attr}(\alpha^s, \alpha^t, M(G(w'_{s\to t})))$$

Method-Stage 3: Blend Image

- ✓ Masked Spatial Feature Transform (MSFT) module aims to fuse the feature of the masked regions for image blending.
- ✓ The background loss and the perceptual loss make the swapped image have the same background as the target image: $L = L_b + \varepsilon_p L_p$ $\quad L_b = \|M \odot (I_r - I_t)\|$







Experiments

- ✓ Evaluating the properties of StyleIPSB.
- ✓ Evaluating the performance of 3DMM controlling facial attributes with StyleIPSB.
- ✓ Comparison of face swapping results with other methods.
- ✓ Ablation study.

Experiments

- ✓ Evaluating the properties of StyleIPSB.
- ✓ Evaluating the performance of 3DMM controlling facial attributes with StyleIPSB.
- ✓ Comparison of face swapping results with other methods.
- ✓ Ablation study.

Frontal

Profile

Pose - Pitch

Glasses

Light

Pose - Yaw

Frontal

Profile Glasses

Macro Expression - Smile

Frontal

Profile

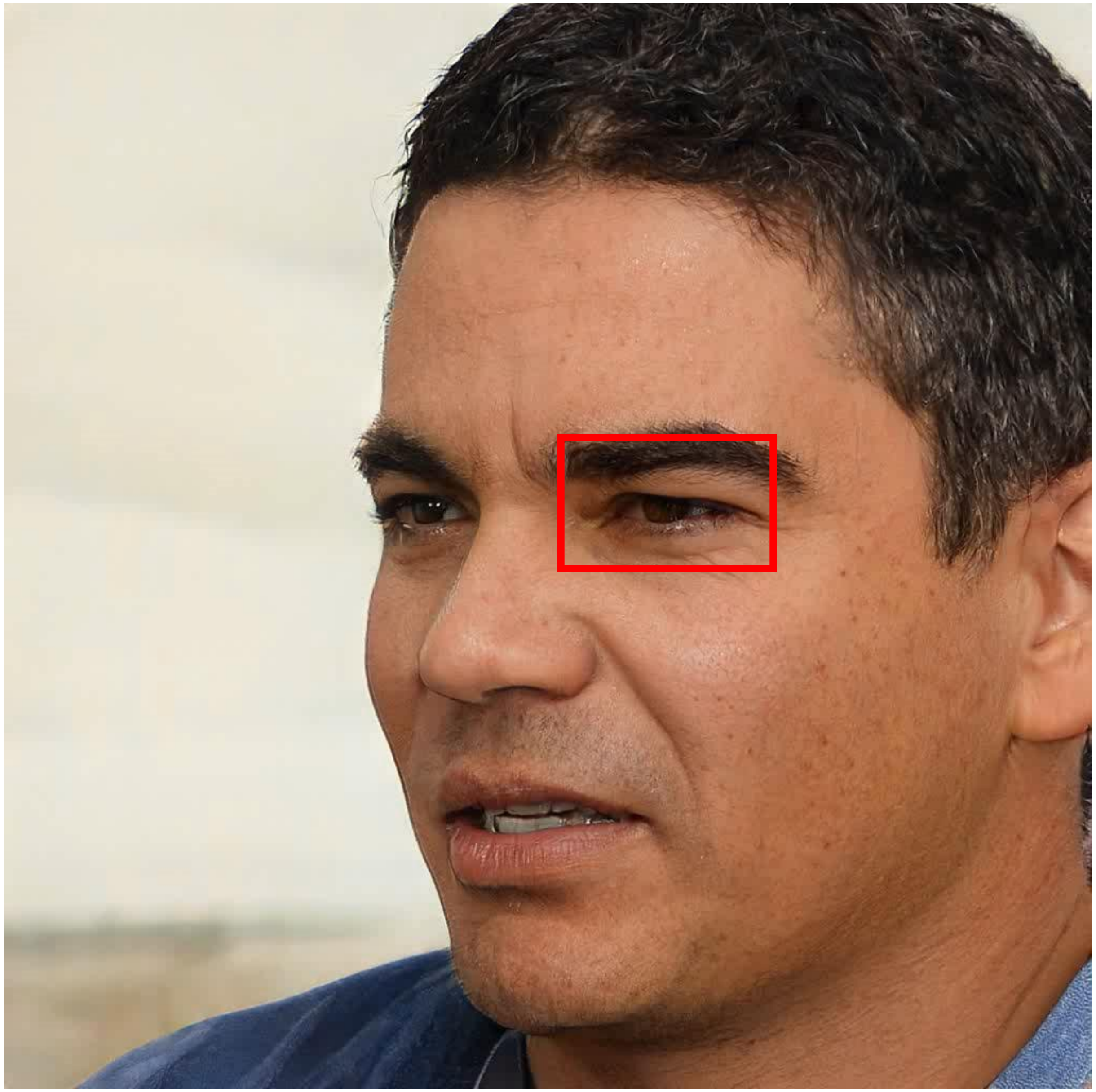
Macro Expression - Open Eyes

Profile

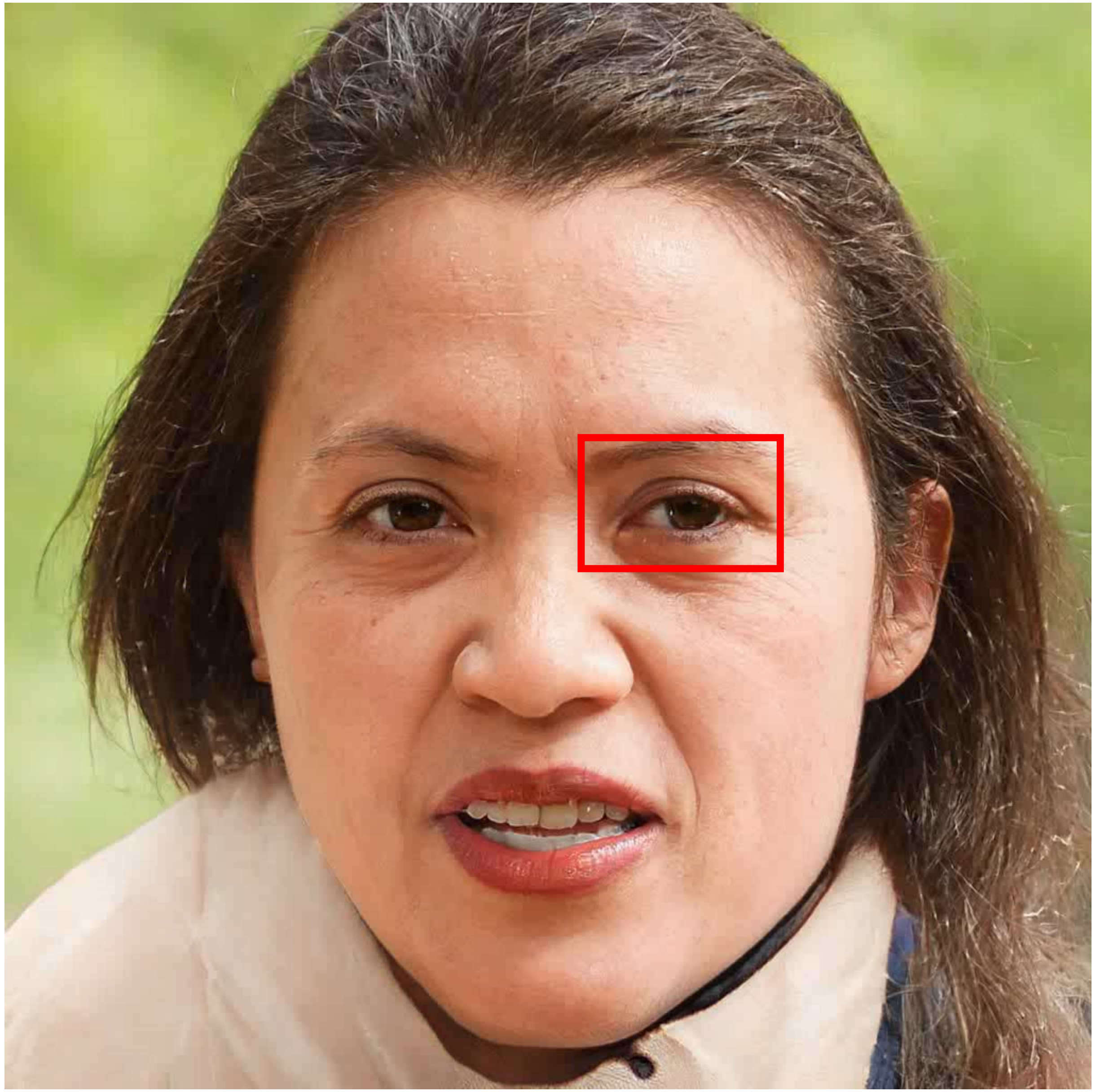
Light

Micro Expression - Raise Eyebrow

Profile



Light



Micro Expressions- Rotate Eyeball

Glasses

Profile

Light- Intensity

Frontal

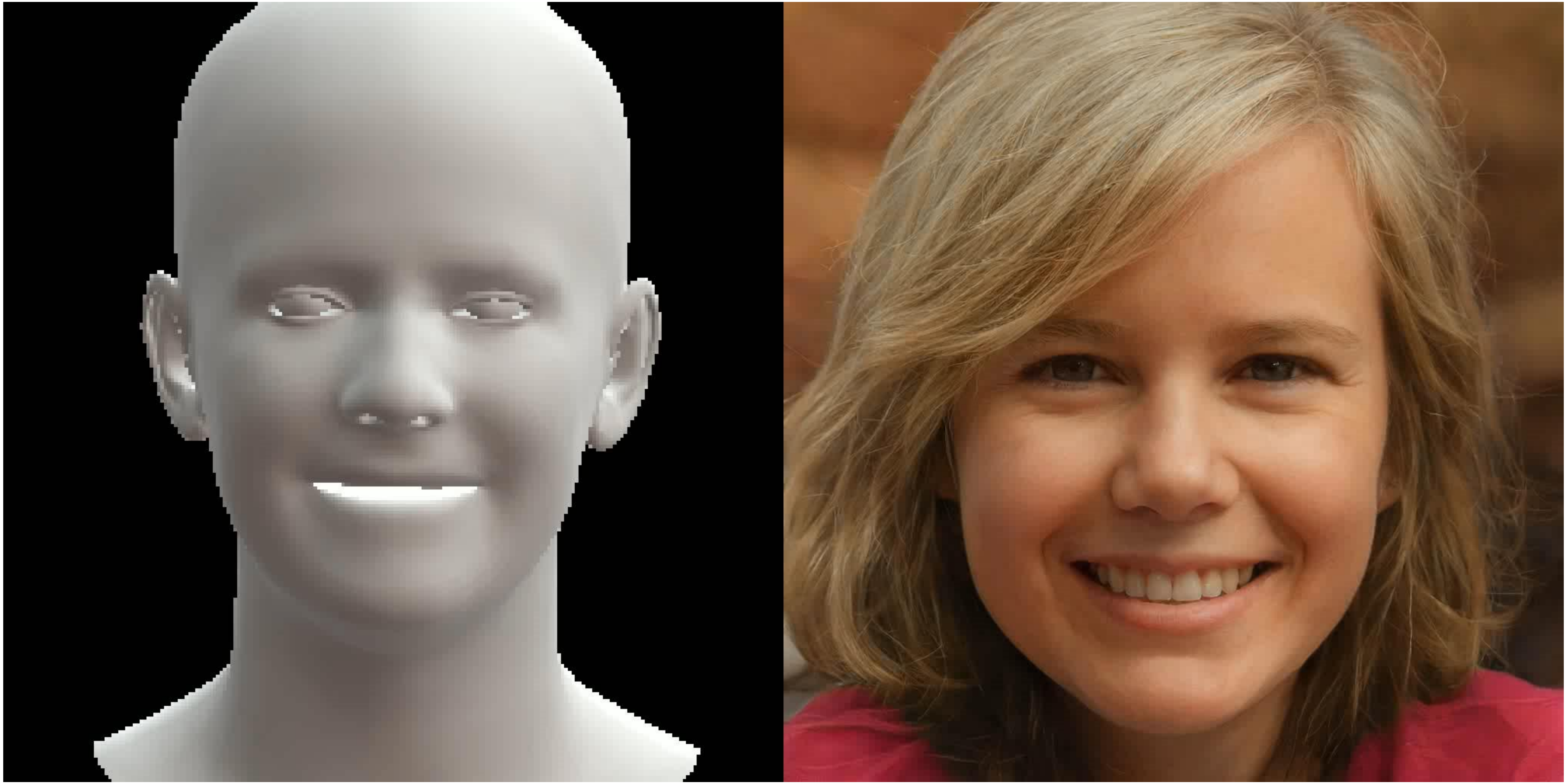
Glasses

Light- Direction

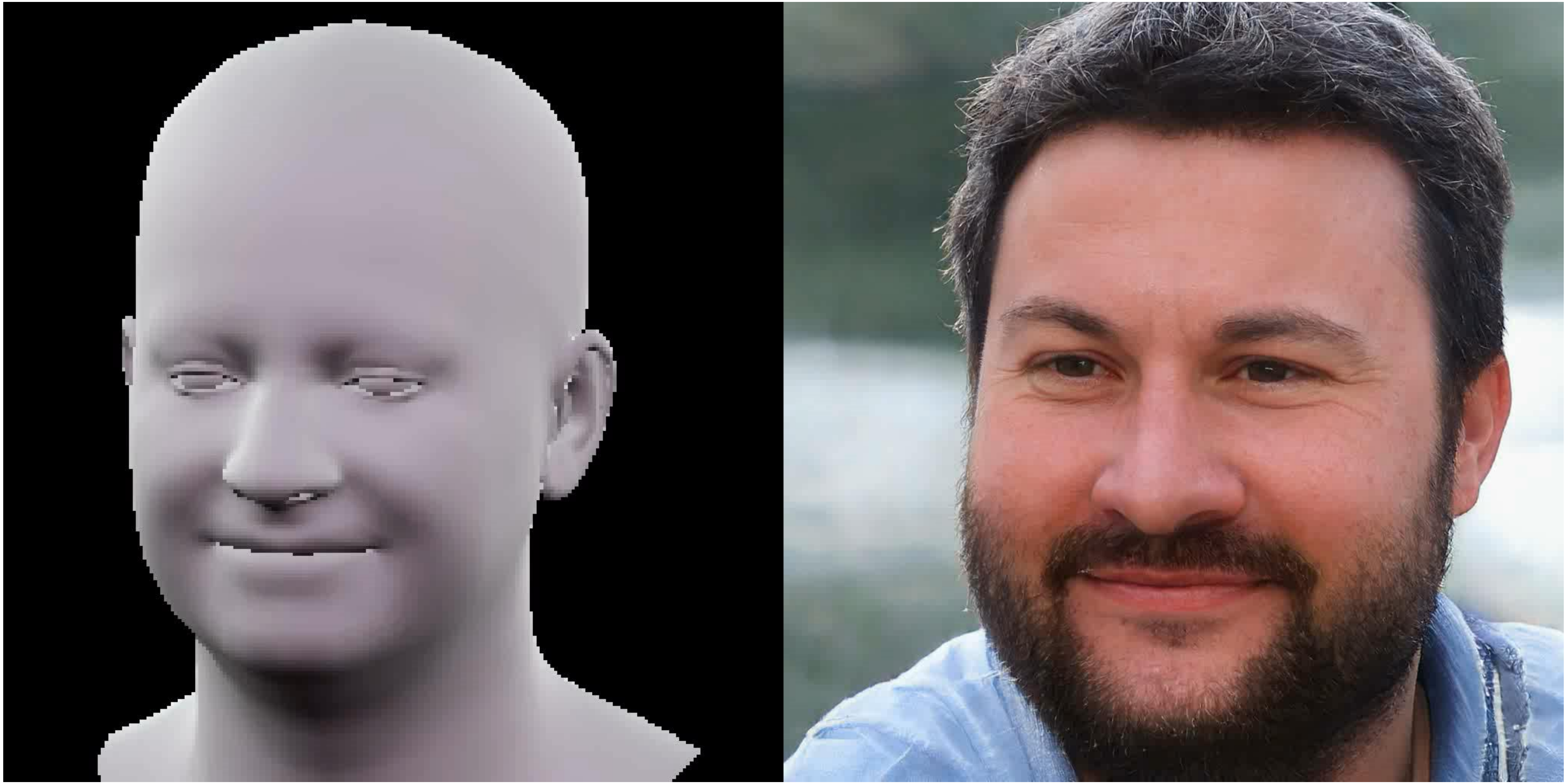
Light- Color Temperature

Experiments

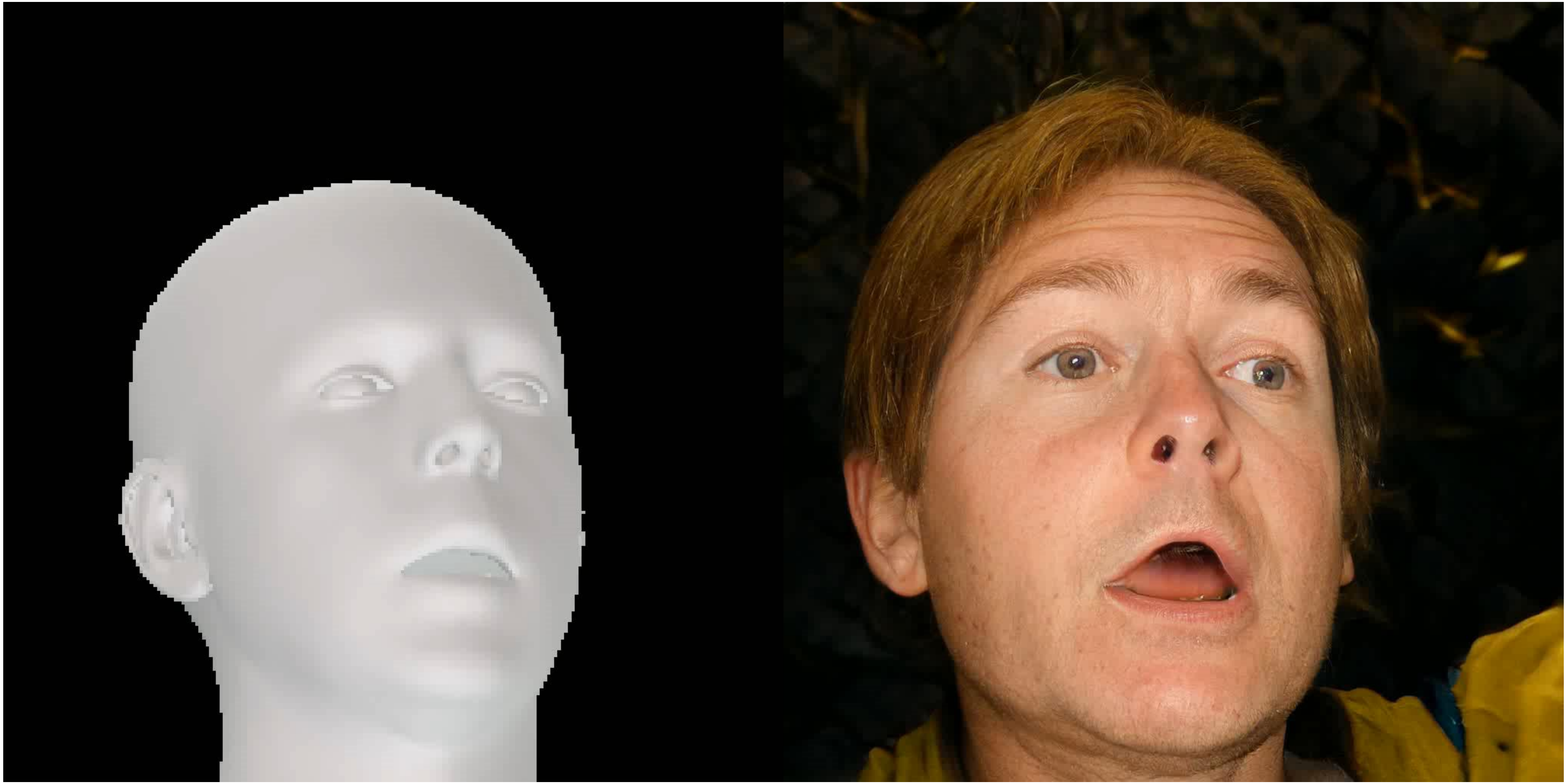
- ✓ Evaluating the properties of StyleIPSB.
- ✓ Evaluating the performance of 3DMM controlling facial attributes with StyleIPSB.
- ✓ Comparison of face swapping results with other methods.
- ✓ Ablation study.



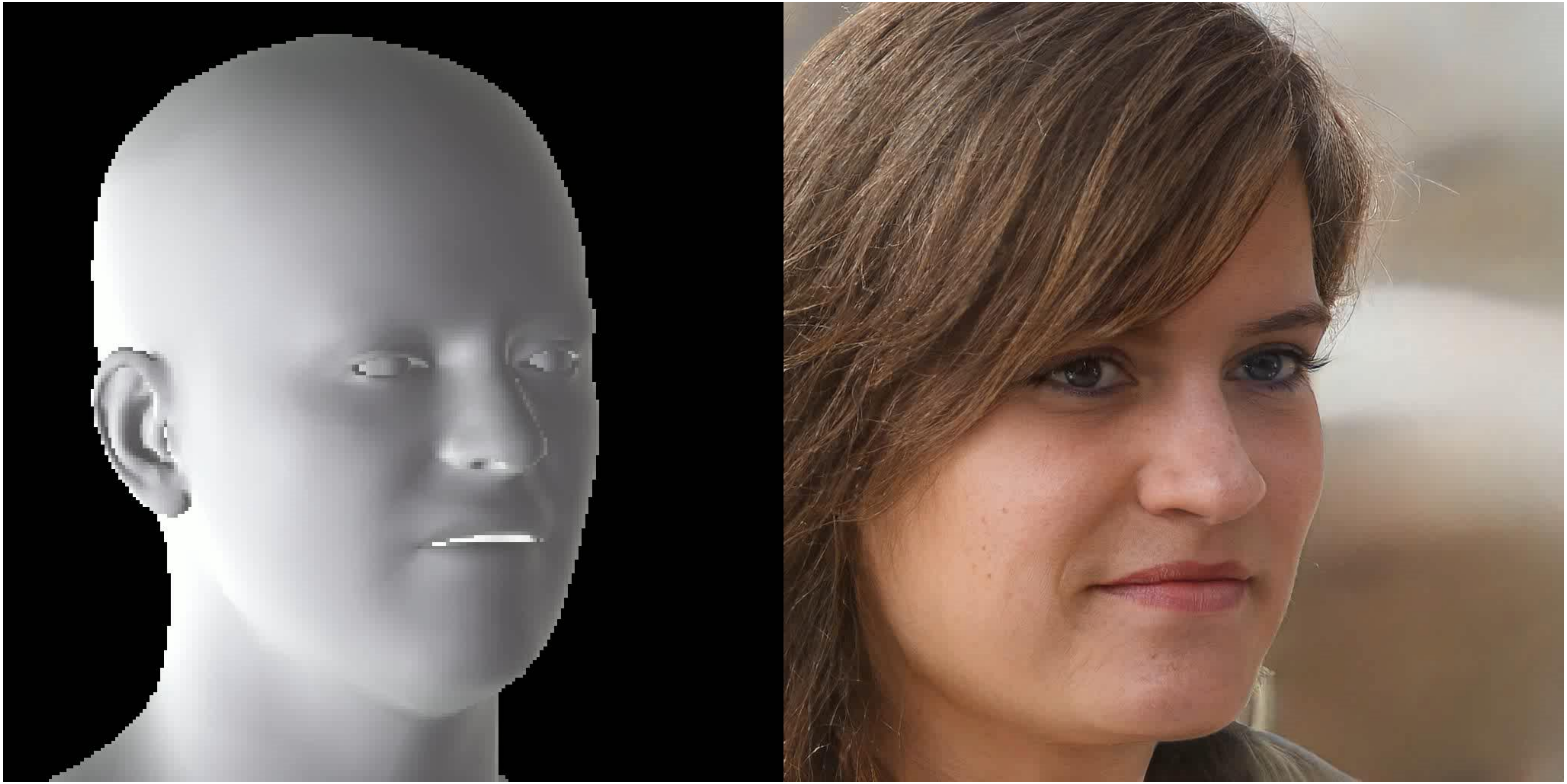
Pose - Pitch



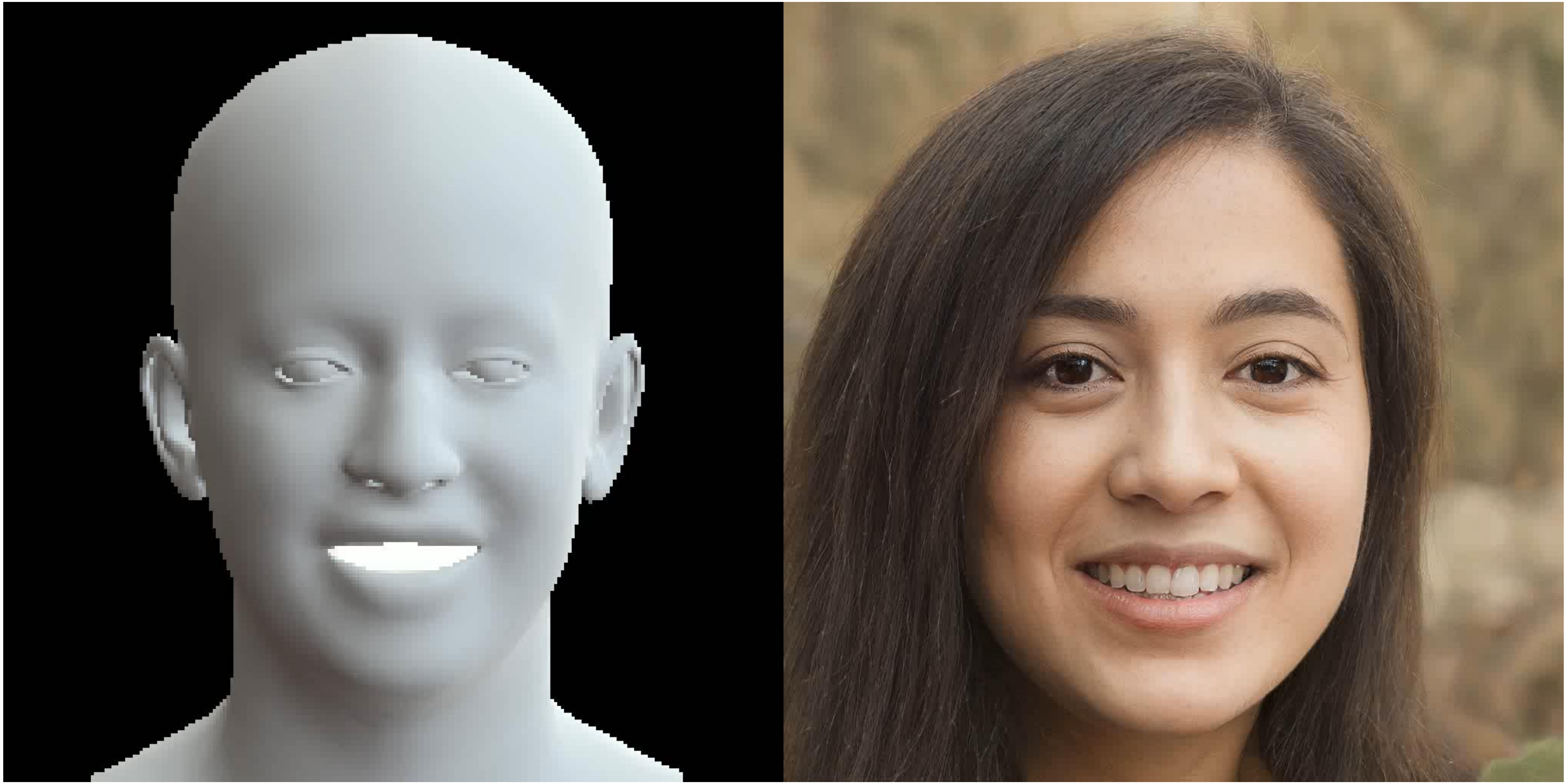
Pose - Yaw



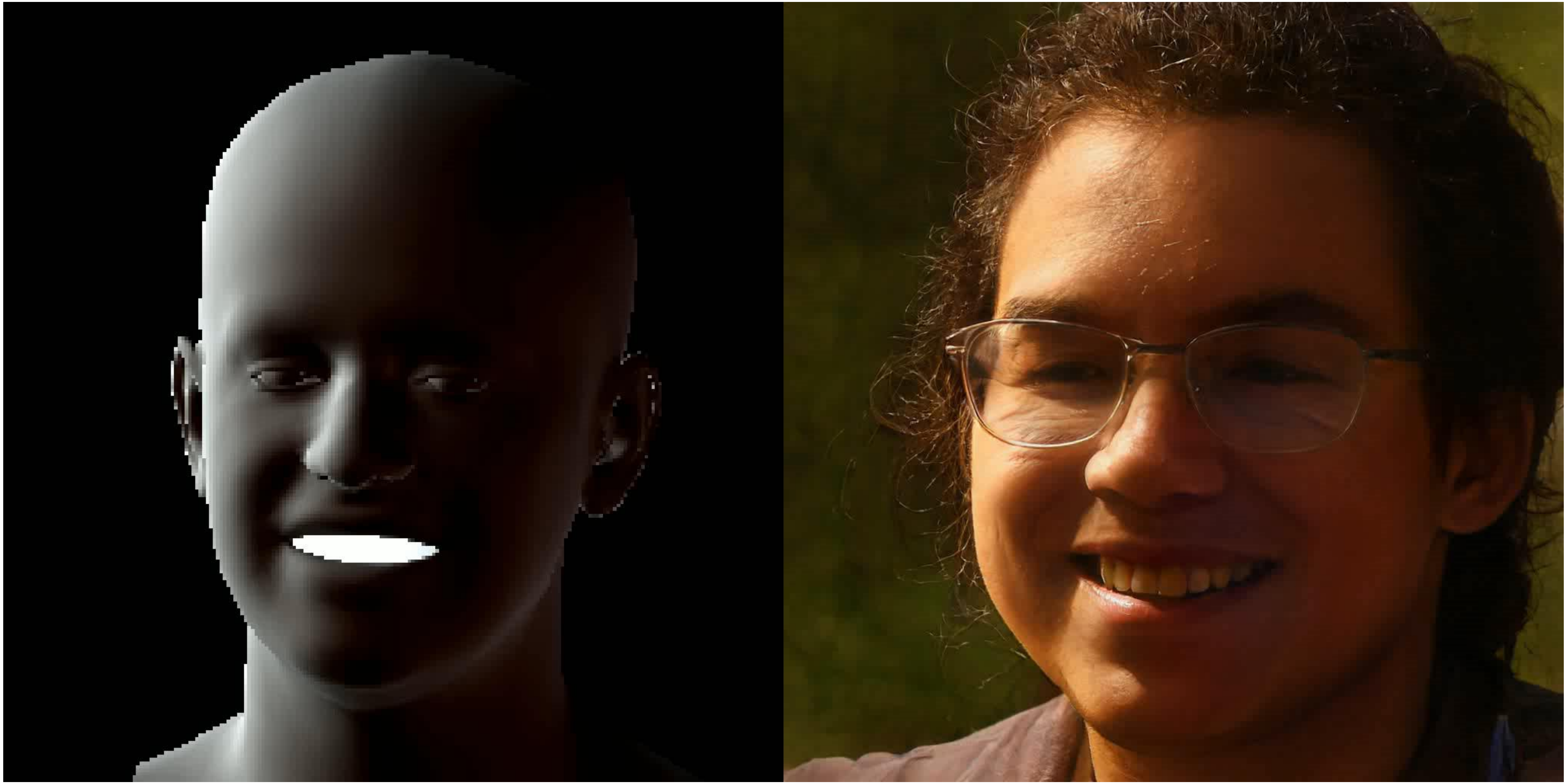
Expression - Smile



Expression - Disgust



Expression - Surprise



Light

Experiments

- ✓ Evaluating the properties of StyleIPSB.
- ✓ Evaluating the performance of 3DMM controlling facial attributes with StyleIPSB.
- ✓ Comparison of face swapping results with other methods.
- ✓ Ablation study.

Experiments

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Quantitative experiments on FaceForensics++ dataset

Method	ID Retri.(%) ↑	Exp Err. ↓	Pose Err. ↓
FaceSwap [3]	72.69	2.89	2.58
Deepfakes [1]	88.39	3.33	4.64
FaceShifter [22]	90.68	2.82	2.55
MegaFS [51]	90.83	2.92	2.64
FSLDS [47]	90.05	2.79	**2.46**
Ours	**95.05**	**2.23**	3.58

Quantitative experiments on CelebAHQ dataset

	FID ↓	Exp ↓	Pose ↓	ID similarity ↑
MageFS [51]	22.03	2.85	**0.043**	0.4837
RAFS [46]	13.25	3.15	-	0.5232
FSLDS [47]	10.01	2.99	0.053	0.4761
Ours	**9.37**	**2.75**	0.078	**0.5378**

Ours

GANSpace

Ours

InterfaceGAN

Ours

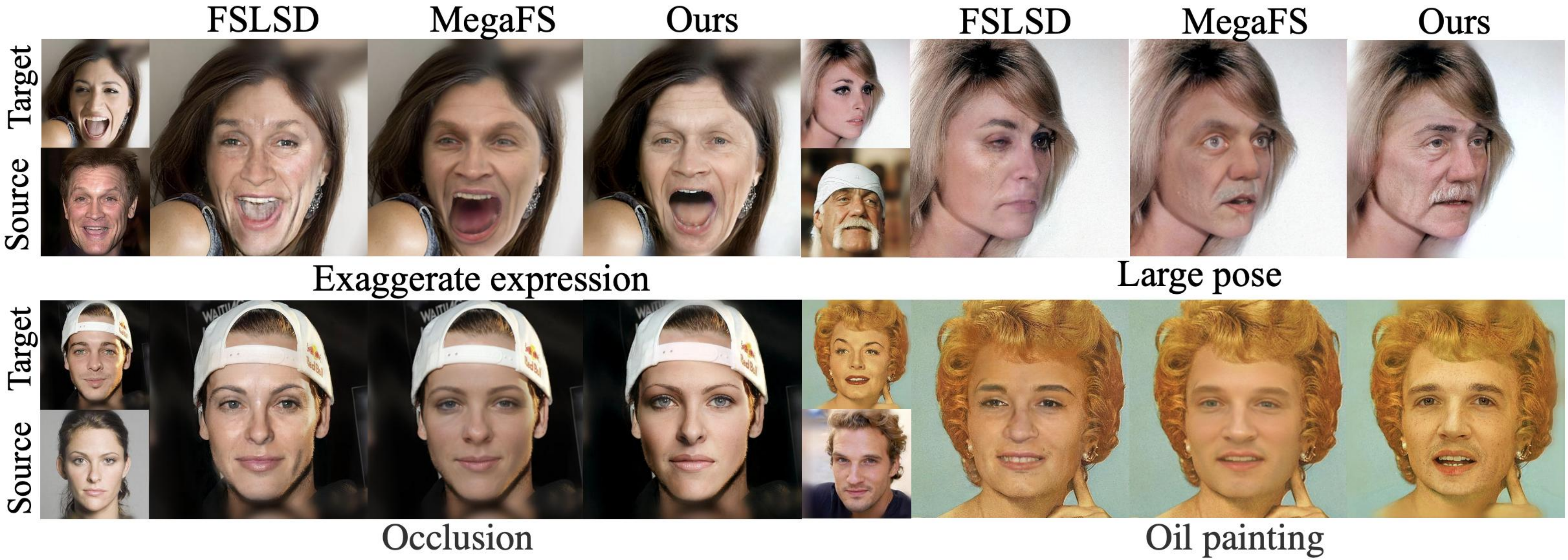
GANSpace

Ours

InterfaceGAN

Experiments

✓ Challenging conditions



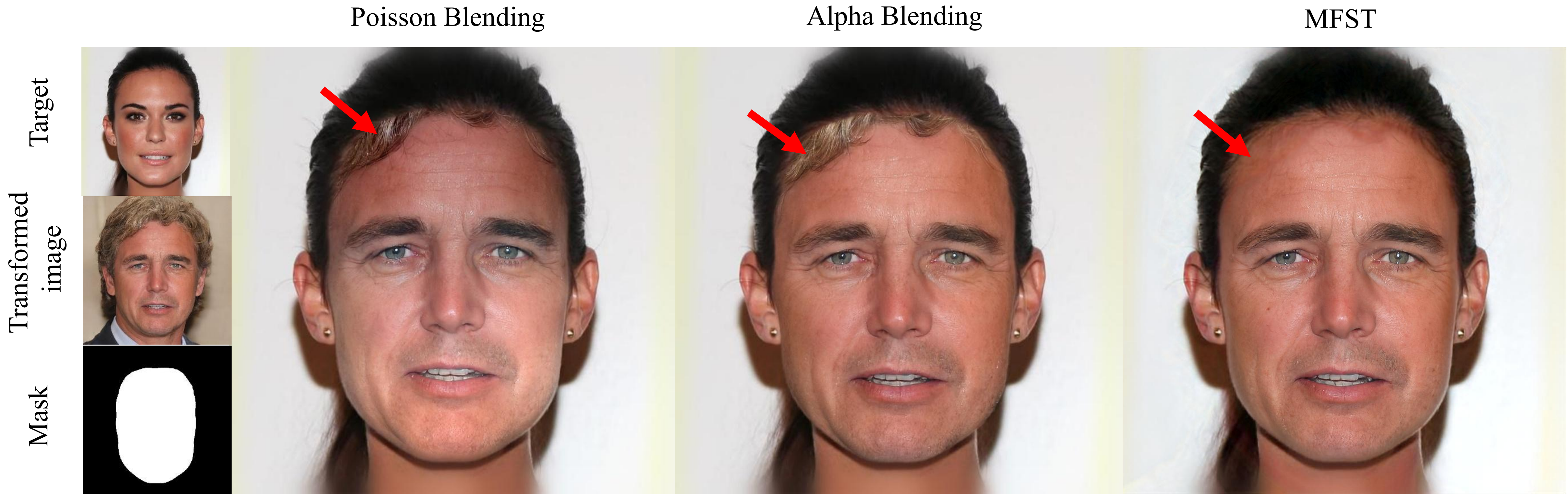
Experiments

- ✓ Evaluating the properties of StyleIPSB.
- ✓ Evaluating the performance of 3DMM controlling facial attributes with StyleIPSB.
- ✓ Comparison of face swapping results with other methods.
- ✓ Ablation study.

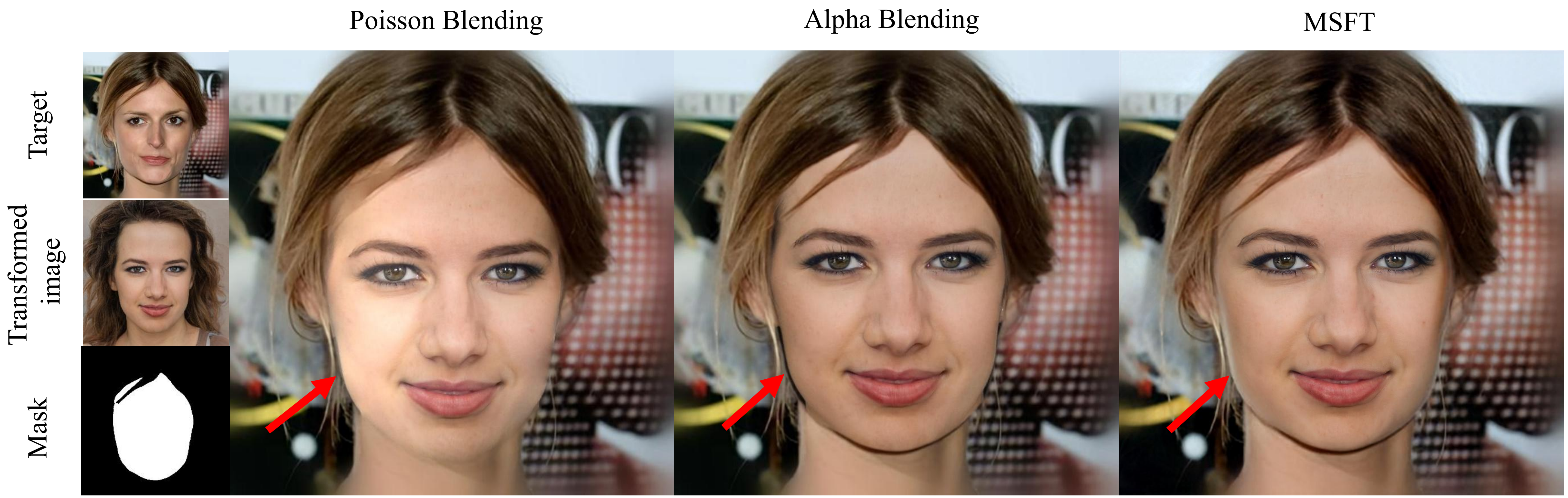
Ablation Study

	FID ↓	Exp ↓	Pose ↓	ID similarity ↑
No Basis	26.06	3.73	**0.074**	0.65
With basis	**22.15**	**3.37**	0.078	**0.67**

Ablation Study



Ablation Study



Conclusion and Limitation

- ✓ We have developed a new semantic basis for face swapping, called StyleIPSB, that is specifically designed to preserve identity and pore-level details. Our experiments have demonstrated that StyleIPSB outperforms other state-of-the-art methods.
- ✓ (1) Occlusion is limited by the mask. (2) The glasses in the source image cannot be removed. (3) Light and shadow cannot be perfectly restored in the case of complex illumination.

Thank You!