Controllable Face Aging

Haien Zeng, Hanjiang Lai

Speaker: Haien Zeng

Motivation

- People are aging differently and there are many possibilities for the aging process. Hence, facial attributes become important to synthesize the authentic and rational face images.

For the problem of looking for the missing children, when a child grows up, she/he may be fatter or thinner. It's difficult to identify this grown-up child is the missing child we are looking for without the aid of synthesize images.

Related Works

- Aging mechanisms & pattern modelling (Suo et al. 2012; Shlizerman & Suwajanakorn, 2014; Goodfellow et al. 2014)
- Conditional GANs (Zhang et al. 2017; Antipov et al. 2017)
- Pyramid architecture or recurrent module (Wang et al. 2016; Yang et al. 2018)
- GAN-based style transfer (Karras et al. 2019; Viazovetskyi et al. 2020; Karras et al. 2019; Isola et al. 2017)

However, they didn’t consider the face aging with controllable attributes.

Challenges & Contributions

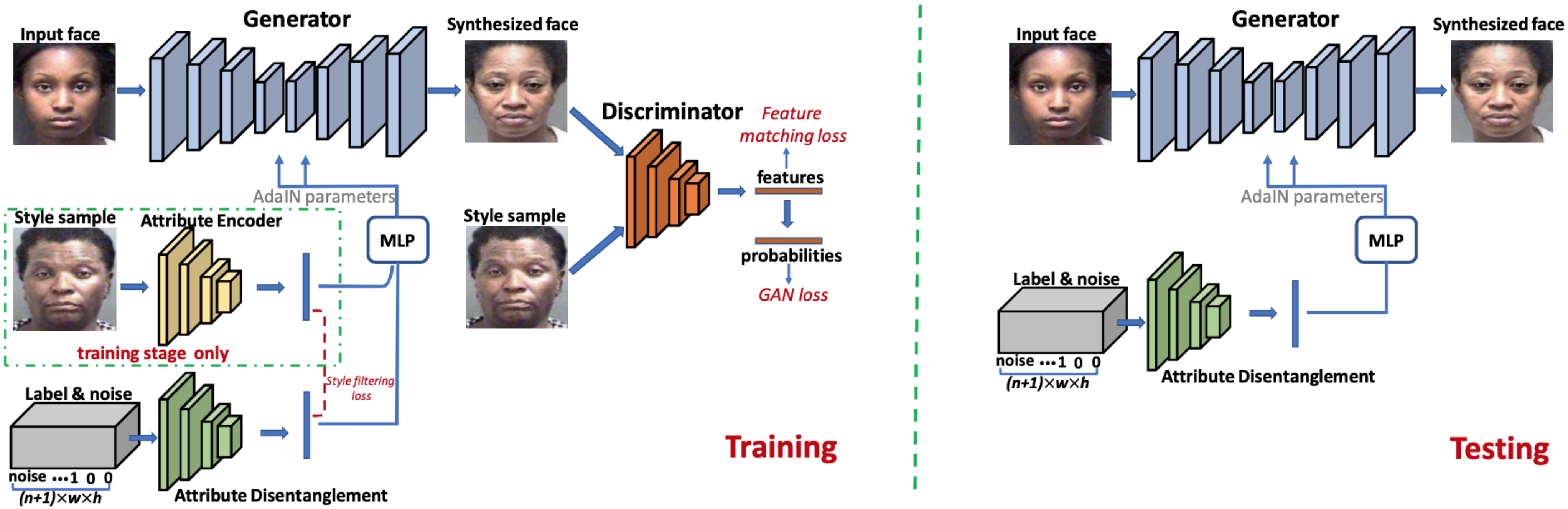
- People are aging differently under different conditions for changeable facial attributes, and it needs to keep some unchanged facial attributes during the aging process.

- We model an attribute disentanglement GAN for controllable face aging.
  - We introduce a controllable face aging to meet the need of the practical applications, which can offers more control over the attributes of the generated face images at different levels.
  - To remove the effect of other unwanted attributes, we propose an attribute disentanglement method to obtain more reliable common age pattern.

Methodology-Overview

- Our model has two inputs: an input face image and desired attributes. Let $S = \{age, gender, race\}$ to represent the label of the desired attributes. Denote the face dataset as $[I_i = (X_i, S_i)]_{i=1}^{M_i}$, where X_i is the $\mathrm{i-th}$ input image and S_i is the corresponding label, M_i is the number of training face images.
- In training, we introduce an individual attribute encoder E and a discriminator D for training the G and F.
  - The E is used to obtain the individual age pattern while F learns the common age pattern.
  - The discriminator D is used to distinguish the generated face images from the real face images.

Methodology-Module



- AdaIN Transformation for Generator
- Attribute Encoder & Attribute Disentanglement
- Multiple Binary Classification Discriminators

Methodology- Training

We utilize the two-stage training method to train our modules.

- In the first stage, we train G, E, D via individual attribute translation. The loss function is denote as:

$$\max_{D} \min_{G,E} \ell_{GAN}(D,E,G) + \lambda_1 \ell_R(G,E) + \lambda_2 \ell_{FM}(G,E),$$

Where

$$\ell_{GAN}(D,E,G) = \mathbb{E}_{\{X_i\}}[\log D^{S_i}(X_i)] + \mathbb{E}_{\{X_i,X_t\}}[\log(1 - D^{S_t}(G(X_i, E(X_t))))]$$

$$\ell_R(G,E) = \mathbb{E}_{\{X_i\}}[||X_i - G(X_i, E(X_i))||_1^1]$$

$$\ell_{FM}(G,E) = \mathbb{E}_{\{X_i,X_t\}}[||D_f(\hat{X}) - D_f(X_t)||_1^1]$$

Methodology- Training

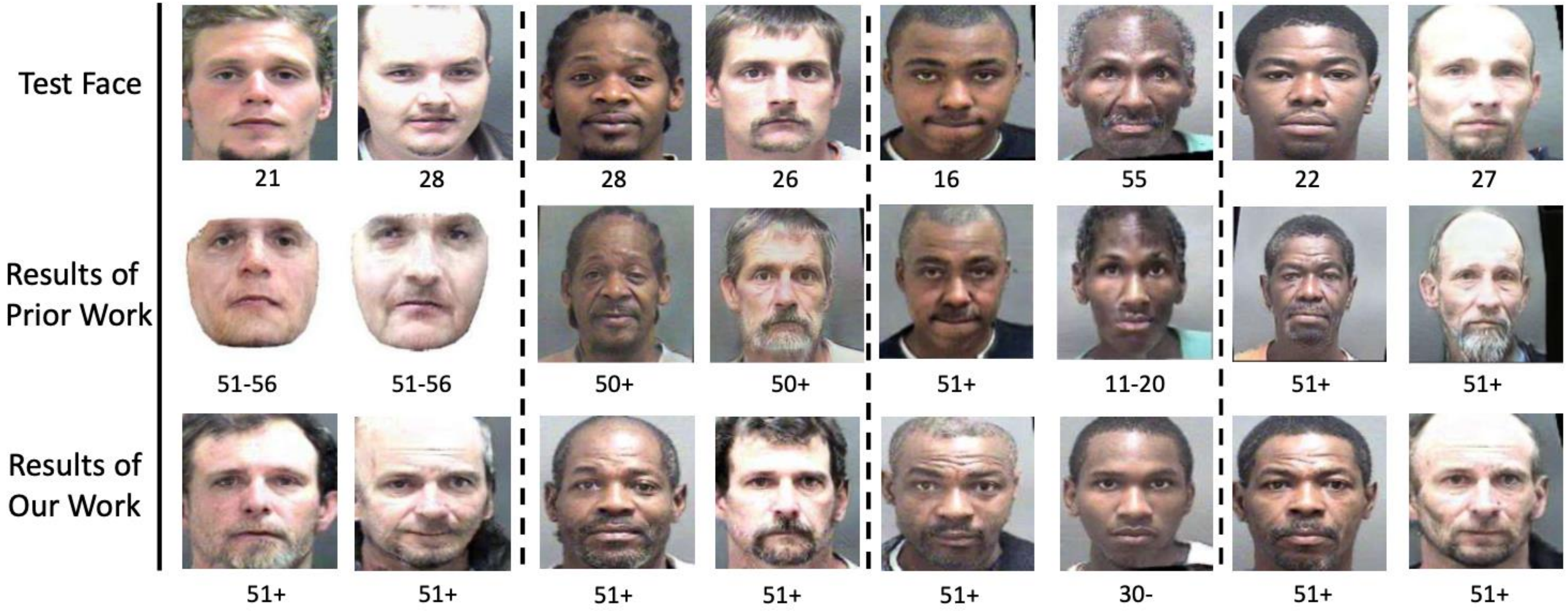
We utilize the two-stage training method to train our modules.

- In the second stage, we update F via common attribute translation. The loss function is denote as:

$$\min_{F} \quad \ell_{GAN}(D, F, G) + \lambda_1 \ell_R(G, F) + \lambda_2 \ell_{FM}(G, F) \\ + \lambda_3 \ell_{DIS}(E, F),$$

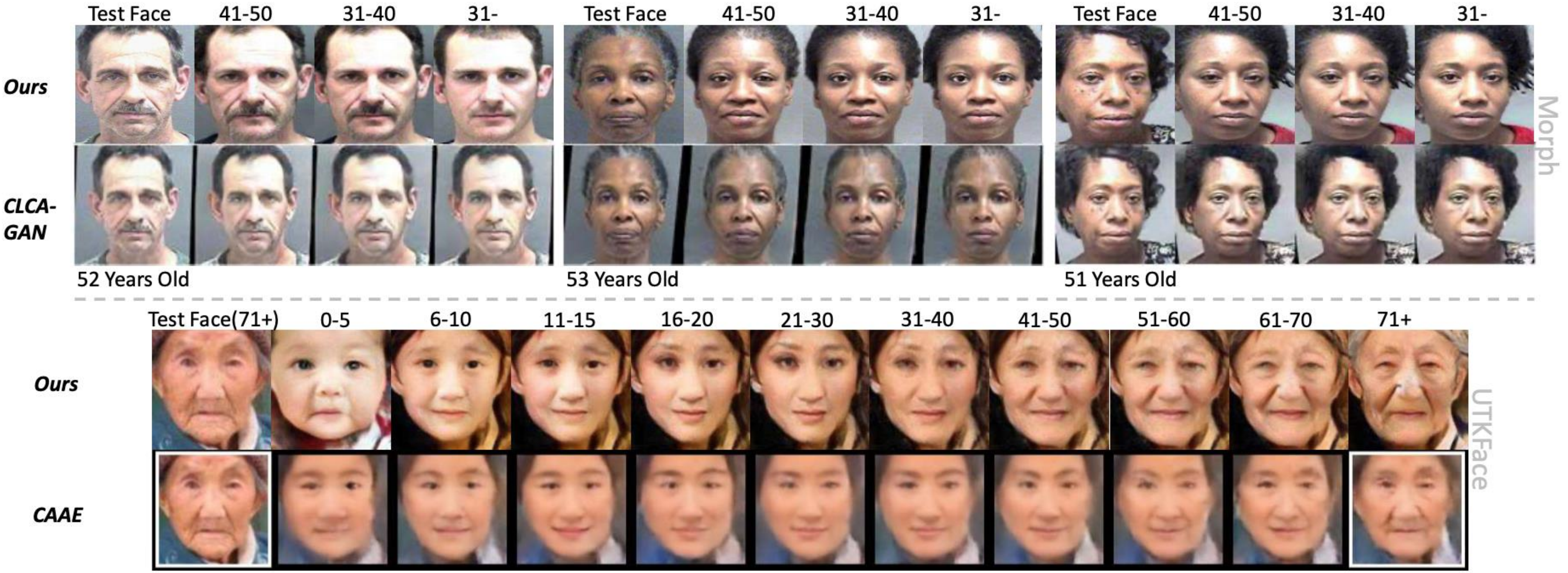
Where $\ell_{DIS}(E, F) = ||\hat{X}_E - \hat{X}_F||_1^1 + \beta ||E(X_t) - F(S_t)||_1^1$

Experiments



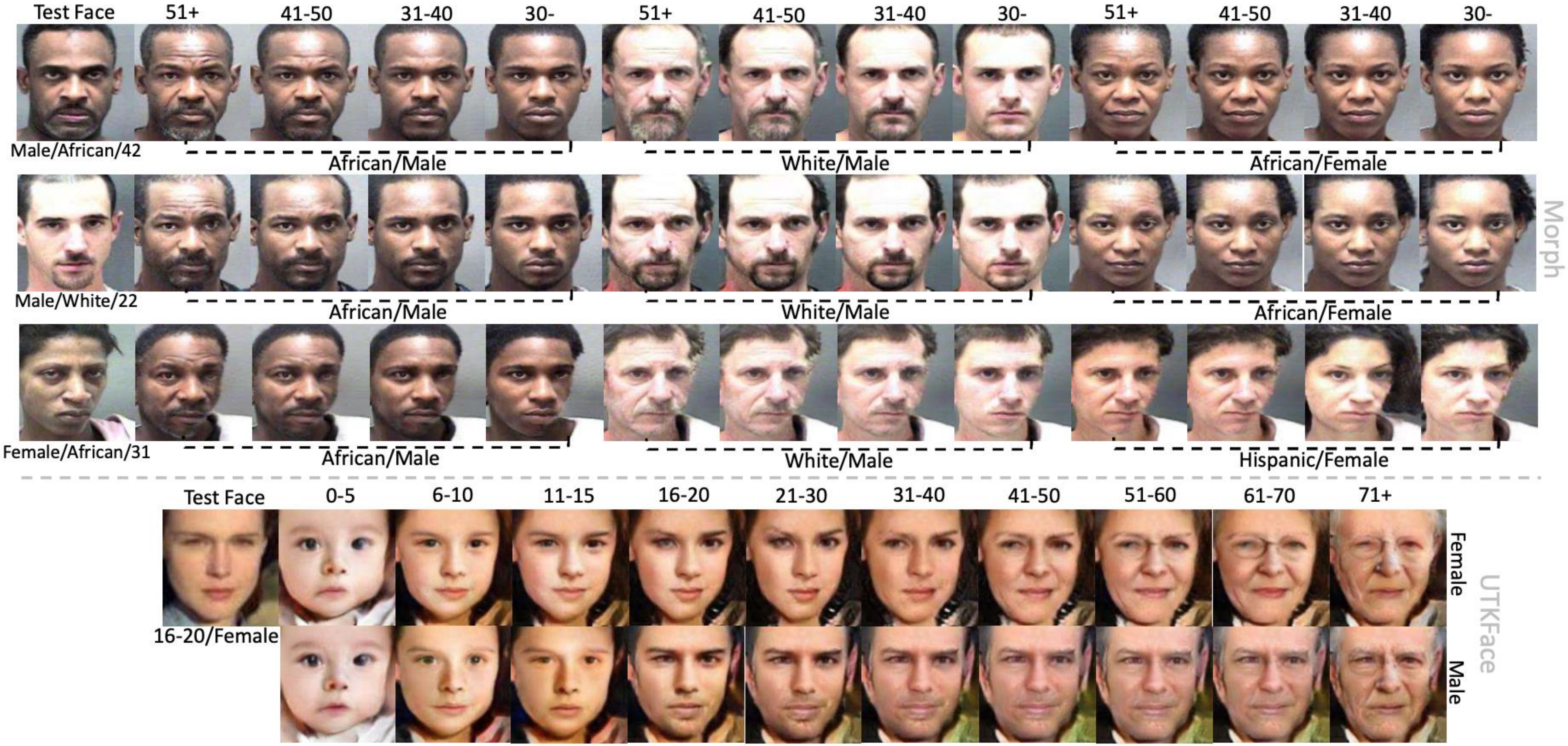
Comparison with prior work (the second row) of age progression. Four methods are considered and two sample results are presented for each. These four methods are: HFA, PAG-GAN, IPCGAN and Attribute-aware GAN.

Experiments



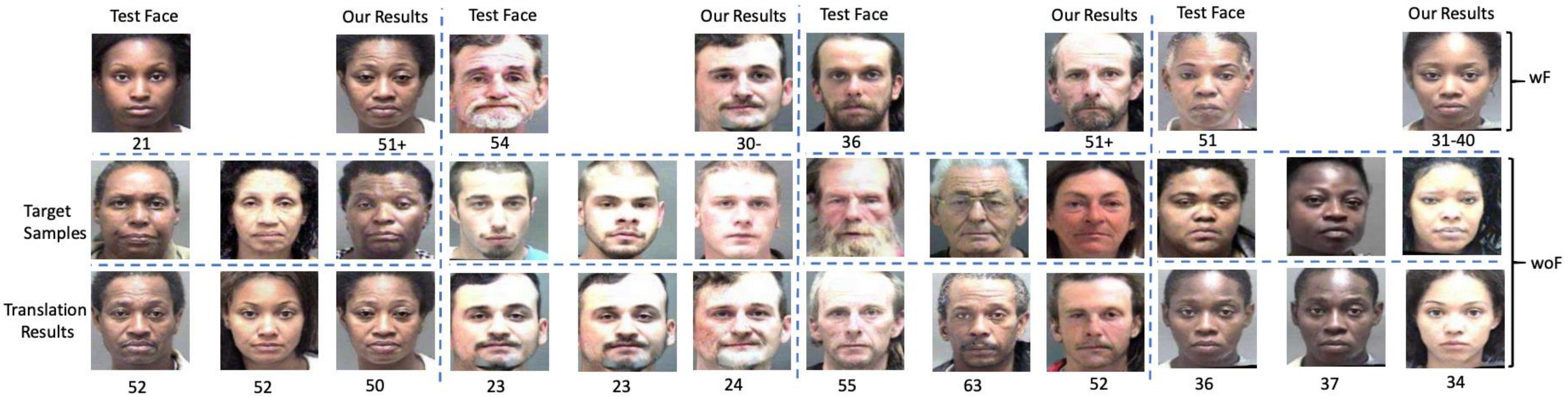
Age regression results on Morph and UTKFace. We compare our results with GLCA-GAN and CAAE, respectively.

Experiments



Results of attributes controllable synthesization experiment on Morph and UTKFace. The First column is the test faces while others are synthesis results on the conditions of different races or genders.

Experiments



Ablation study results of 4 different individuals. The first row is the test faces and output results of the model with style disentanglement F. The second row is the style images that required by the model without style disentanglement F and the last row is the output results of the model without F .

Thank you so much for watching!

Feel free to contact us for any question.