Analyzing and Combating Attribute Bias for Face Restoration

Zelin Li[1,2], Dan Zeng[1,2*], Xiao Yan[2], Qiaomu Shen[1,2], Bo Tang[1,2]

[1]Research Institute of Trustworthy Autonomous Systems, Southern University of Science and Technology

[2]Department of Computer Science and Engineering, Southern University of Science and Technology

Paper code link: https://github.com/Seeyn/DebiasFR

Contact me

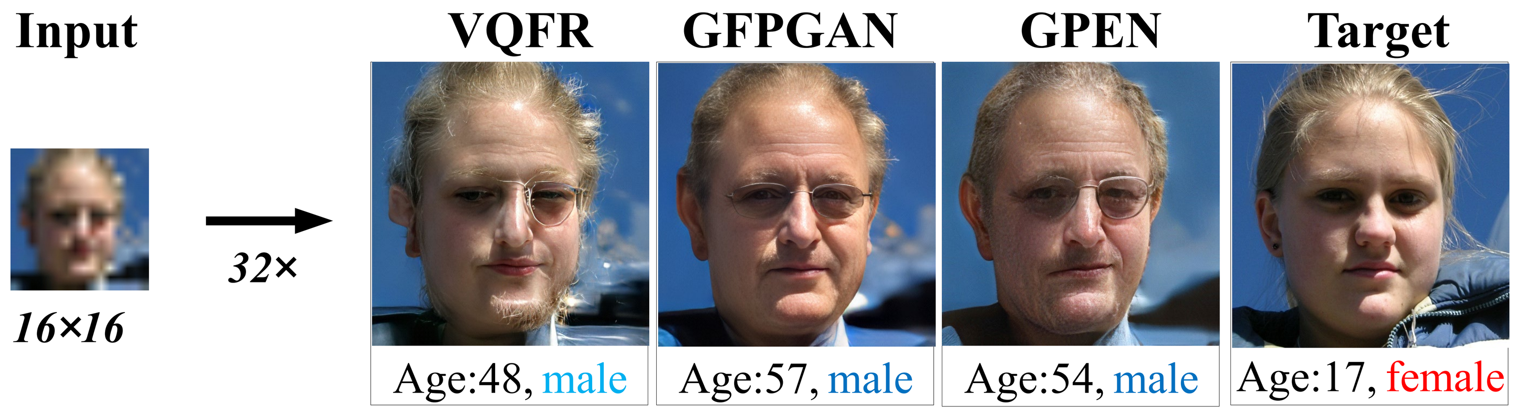
IJCAI/2023 MACAO

CONTRIBUTIONS

- Observe attribute bias in face restoration
- Analyze and trace it to two main causes
- Propose DebiasFR, which faithfully preserves input attribute information and produces quality HR faces

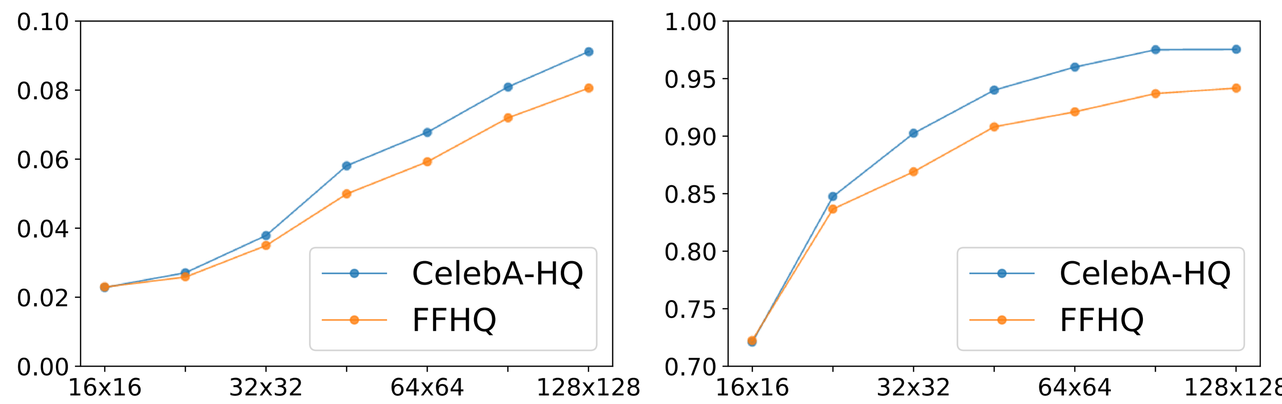
MOTIVATION

- **Attribute Bias: Key face attributes are dramatically different**



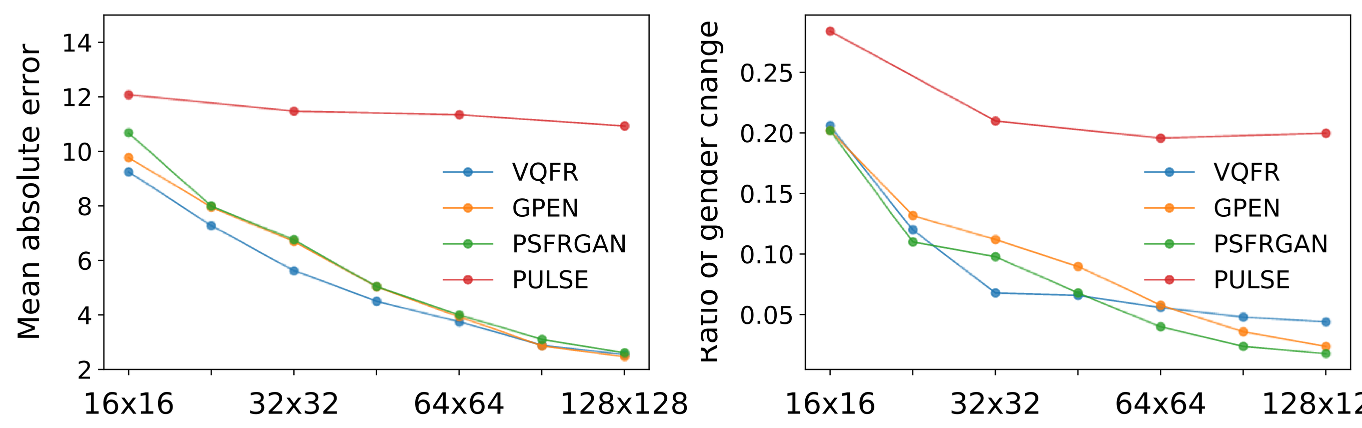
OBSERVATIONS

- **Two leading causes**
  - Lack of attribute information – Degradation is unavoidable
  - Training data prior – Hard to collect a large dataset with balanced attribute distribution
- **Attribute information loses as image resolution decreases**



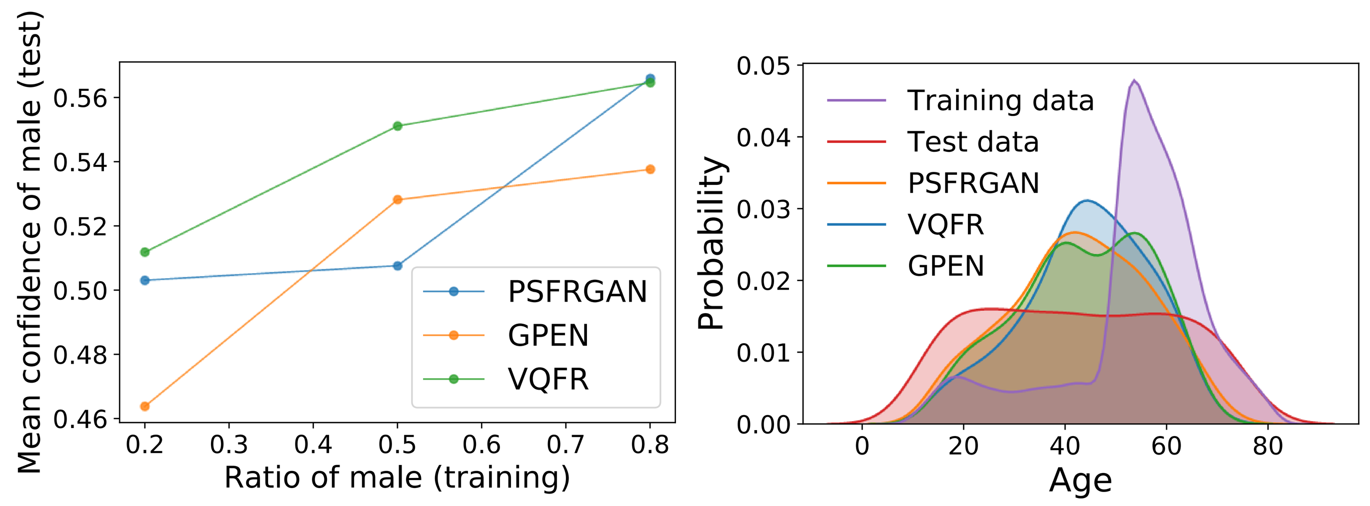
(a) Age Confidence. (b) Gender Confidence.

- **Attribute bias enlarges as image resolution decreases**



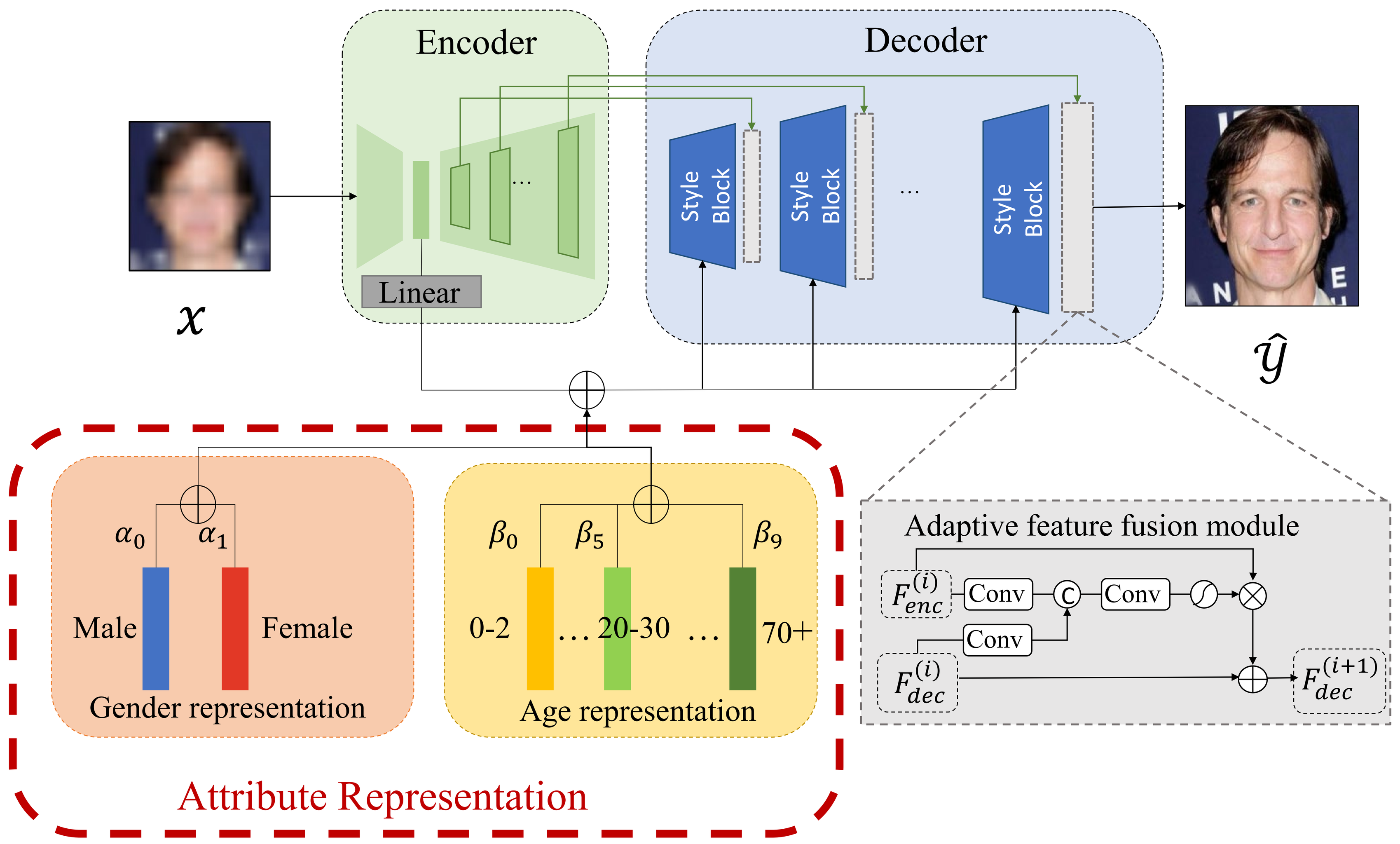
(a) Age. (b) Gender.

- **Affected by the training data distribution**



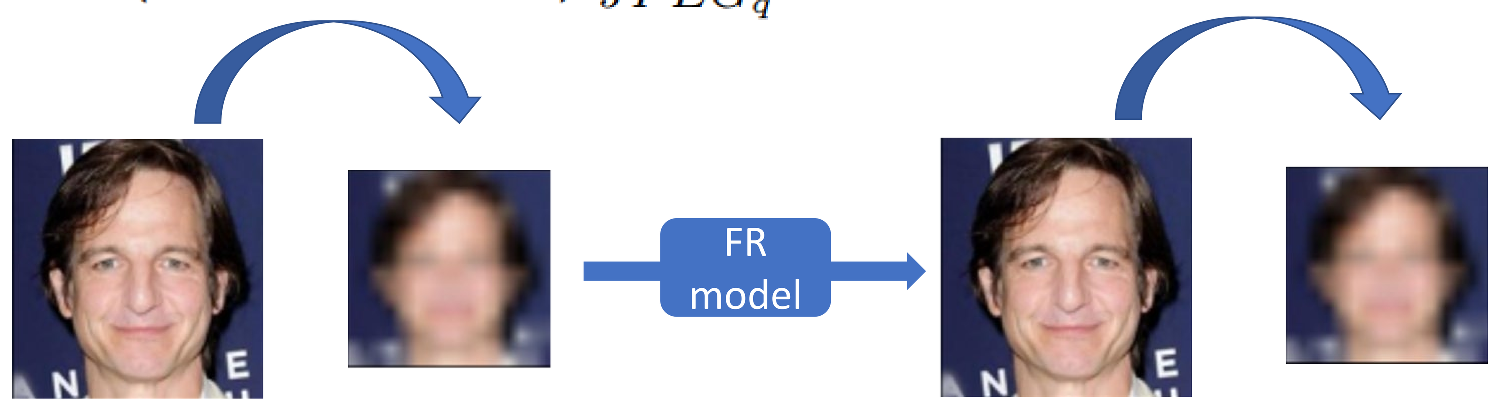
METHOD

- **Architecture Design**
  - Consists of encoder, decoder, attribute representation
  - Adjust the representation weights (α,β) to change the restored face attributes
  - Randomly initialize and update the representations during training



- **Training Strategies**
  - Given high-quality image $\mathcal{Y}$, degrade it by a series of operators to obtain low-quality image $\mathcal{X}$
  - Given $\mathcal{X}$ and its attribute label, face restoration model produce restored image $\hat{\mathcal{Y}}$
  - Degrade $\hat{\mathcal{Y}}$ in the exactly same manner as $\mathcal{Y}$ to obtain $\hat{\mathcal{X}}$

$$x = \left((y \otimes k)_{\downarrow_r} + n_\sigma\right)_{JPEG_q} \qquad \hat{x} = \left((\hat{y} \otimes k)_{\downarrow_r} + n_\sigma\right)_{JPEG_q}$$

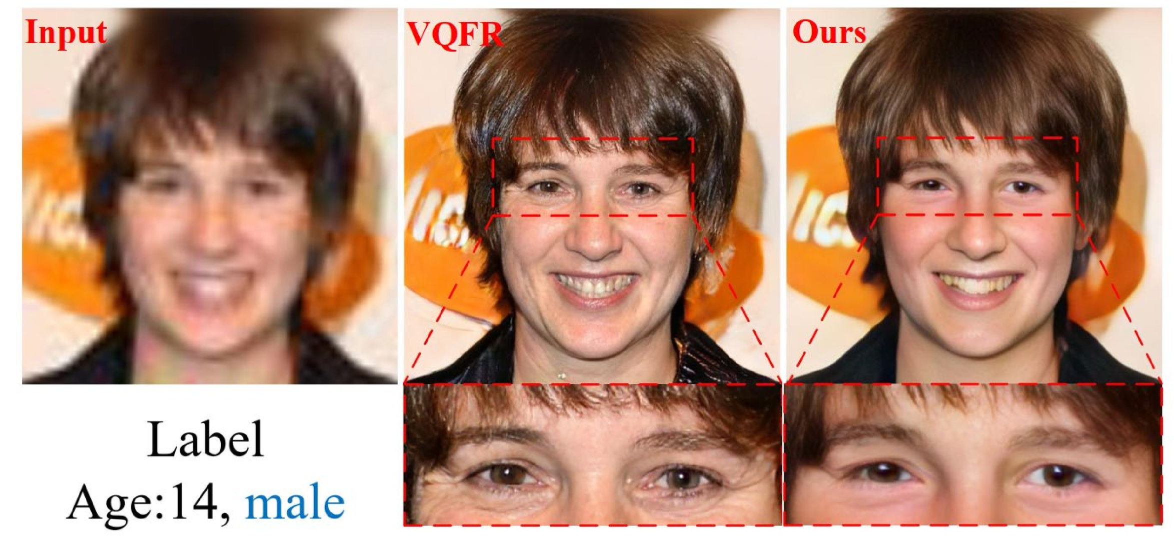


$\mathcal{Y}$ <$\mathcal{X}$, attribute> $\hat{\mathcal{Y}}$ $\hat{\mathcal{X}}$

- Calculate restoration loss based on $\mathcal{Y}$, $\hat{\mathcal{Y}}$ – Image quality
- Calculate attribute loss based on attribute, $\hat{\mathcal{Y}}$ – Attribute consistency
- Calculate degradation loss based on $\mathcal{X}$, $\hat{\mathcal{X}}$ – Data augmentation

EXPERIMENTS

- **Comparison**



Label
Age:14, male

- **Comparison on CelebA-HQ (Synthesized LQs)**

CelebA-HQ	NIQE↓	FID↓	PSNR↑	SSIM↑	Age error (%)↓	Gender error (%)↓
Bilinear	16.59	213.3	**23.25**	**0.6951**	60.90	14.37
AACNN	4.132	59.80	22.65	0.6052	43.60	2.930
PULSE	3.765	65.90	20.81	0.5695	68.90	14.16
PSFRGAN	3.982	55.88	21.71	0.6173	50.66	3.500
GFPGAN	3.800	53.87	20.72	0.6001	50.50	4.160
GPEN	3.877	**47.39**	22.12	0.6152	47.70	3.333
VQFR	**3.350**	52.09	20.48	0.5699	50.60	3.666
Ours	4.411	48.63	21.71	0.6247	**22.30**	**1.467**

Comparable image quality as baselines   Alleviate attribute bias

- **Comparison on IMDB&COX (Real LQs)**

IMDB	NIQE↓	FID↓	Age error (%)↓	Gender error (%)↓
PSFRGAN	4.316	39.04	50.72	1.790
GFPGAN	4.133	**32.30**	48.30	1.699
GPEN	4.719	54.13	49.25	1.155
VQFR	**3.540**	33.40	50.29	1.631
Ours	4.482	39.41	**26.57**	**0.929**
COX				
PSFRGAN	4.521	88.94	61.60	21.19
GFPGAN	5.036	82.52	59.50	20.59
GPEN	4.713	84.01	64.30	22.59
VQFR	**4.190**	**70.00**	59.40	22.90
Ours	5.238	80.45	**29.50**	**8.80**

- **Manipulate restoration result**