13th Chinese Conference on Biometric Recognition
(CCBR 2018 · Aug. 11-12th, 2018, Urumqi, China)

Face Synthesis for Eyeglass-Robust Face Recognition

Jianzhu Guo, Xiangyu Zhu, Zhen Lei, Stan Z. Li

CBSR&NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, China

University of Chinese Academy of Sciences, Beijing, China

1. Introduction

➢ Motivation

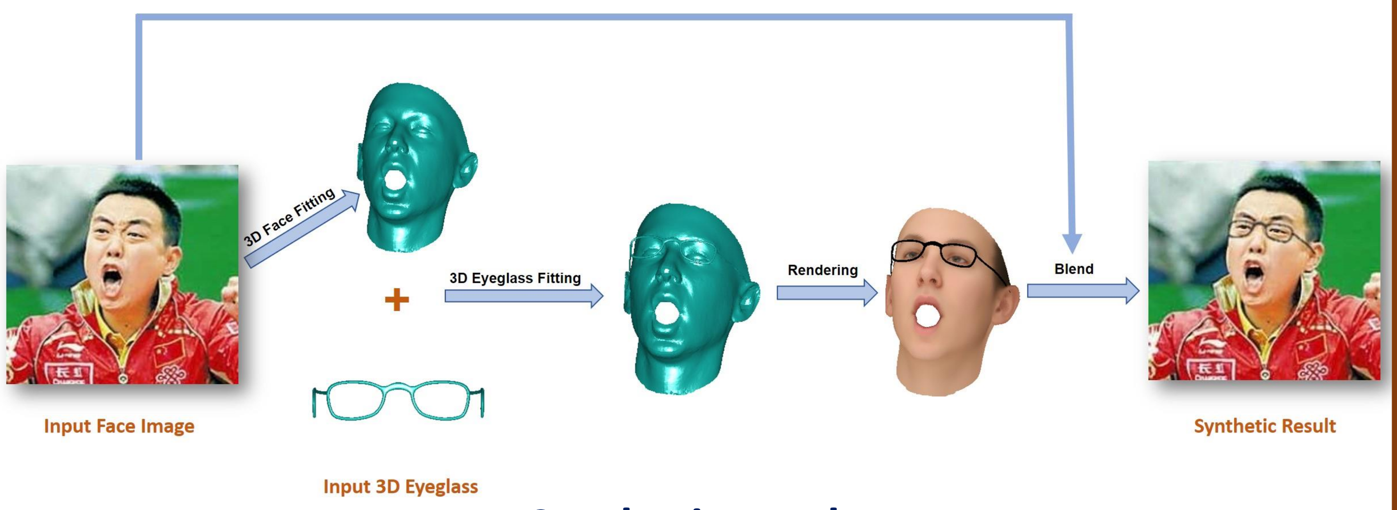
In the application of face recognition, eyeglasses could significantly degrade the recognition accuracy. However, it is difficult to collect the images with and without eyeglasses of the same identity. To address this problem, our core idea is to **synthesize high-fidelity face images with eyeglasses for training**.

➢ Contribution

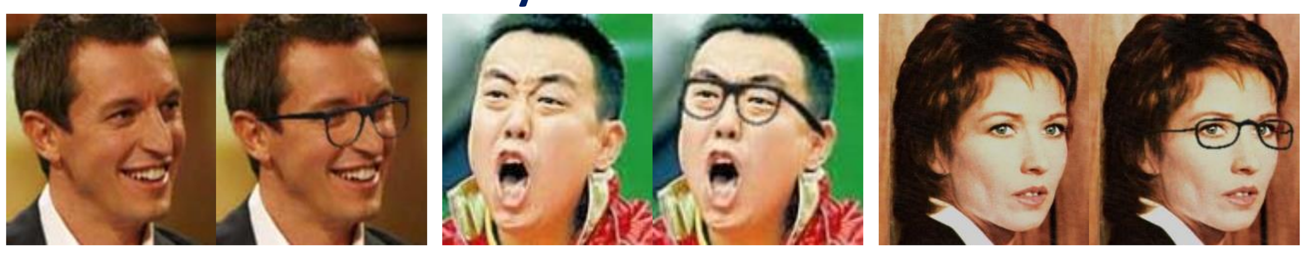
- A eyeglass face dataset named MeGlass, **including about 1.7K identities**, is collected and cleaned for eyeglass face recognition evaluation. It is released on https://github.com/cleardusk/MeGlass.
- A virtual high-fidelity eyeglass face image synthesis method is proposed to help improve the robustness to eyeglass.
- A novel metric loss is proposed to further improve the face recognition performance, which is designed to adequately utilize the synthetic training data.

2. Proposed Method

➢ Eyeglass image synthesis



Synthetic results



- **3D face fitting:** we first reconstruct the 3D face model based on pose adaptive 3DMM fitting method[1], which is robust to pose.
- **3D eyeglass fitting:** the 3D eyeglass is then fitted on the reconstructed 3D face model based on the corresponding anchor points on the 3D eyeglass and 3D fitted face model. The fitting progress is:

$$\arg\min_{f,R,t_{3d}} \left\| f * R * (p_g + t_{3d}) - p_f \right\|$$

- **Rendering and blending:** Z-Buffer and Phong illumination model are adopted for rendering. The rendered eyeglass image is blended on the original image to generate the final result.

➢ Network structure & Loss

- **Network:** We adapt a 22 layers residual network architecture to fit our task.

Layers	22-layer CNN
Conv1.x	[5×5, 32]×1, S2
Conv2.x	[3×3, 64]×1, S1
Conv3.x	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 3$, S2
Conv4.x	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 226 \end{bmatrix} \times 4$, S2
Conv5.x	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 3$, S2
Global Pooling	512

- **Mining-Contrasive loss:** We propose this loss base on [2] to further enlarge the inter-identity differences and reduce intra-identity variations.

$$L_{mc} = -\frac{1}{2|P|}\sum_{(i,j)\in P} d(f_i, f_j) + \frac{1}{2|N|}\sum_{(i,j)\in N} d(f_i, f_j)$$

- **Gradual sampling**: To make the model fit the synthetic training images in a gentle manner, we employ the gradual process into data sampling.

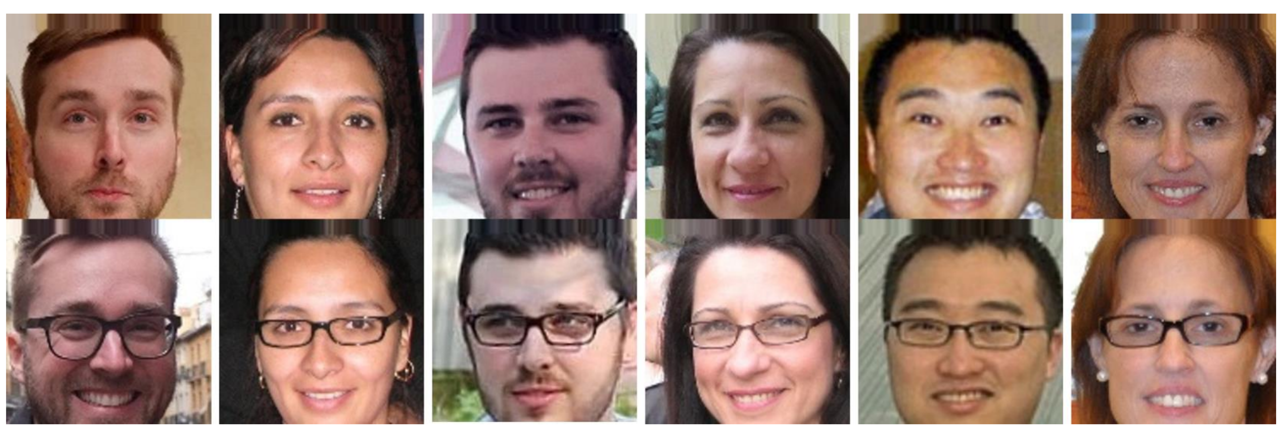
3. Dataset & Experiments

➢ Dataset description

Dataset	Identity	Images	G	NG
MeGlass	1,710	47,917	14,832	33,085
Testing set	1,710	6,840	3,420	3,420
Training set(MsCeleb)	78,765	5,001,877	-	-
Training set(Mixture)	78,765	10,003,754	-	-

- **Training set:** one is only the original MsCeleb and the other is the mixture of MsCeleb with synthetic MsCeleb-Eyeglass.
- **Testing set:** two faces with eyeglasses and two faces without eyeglasses are **selected from each identity of MeGlass** to build the testing set.

The released MeGlass dataset



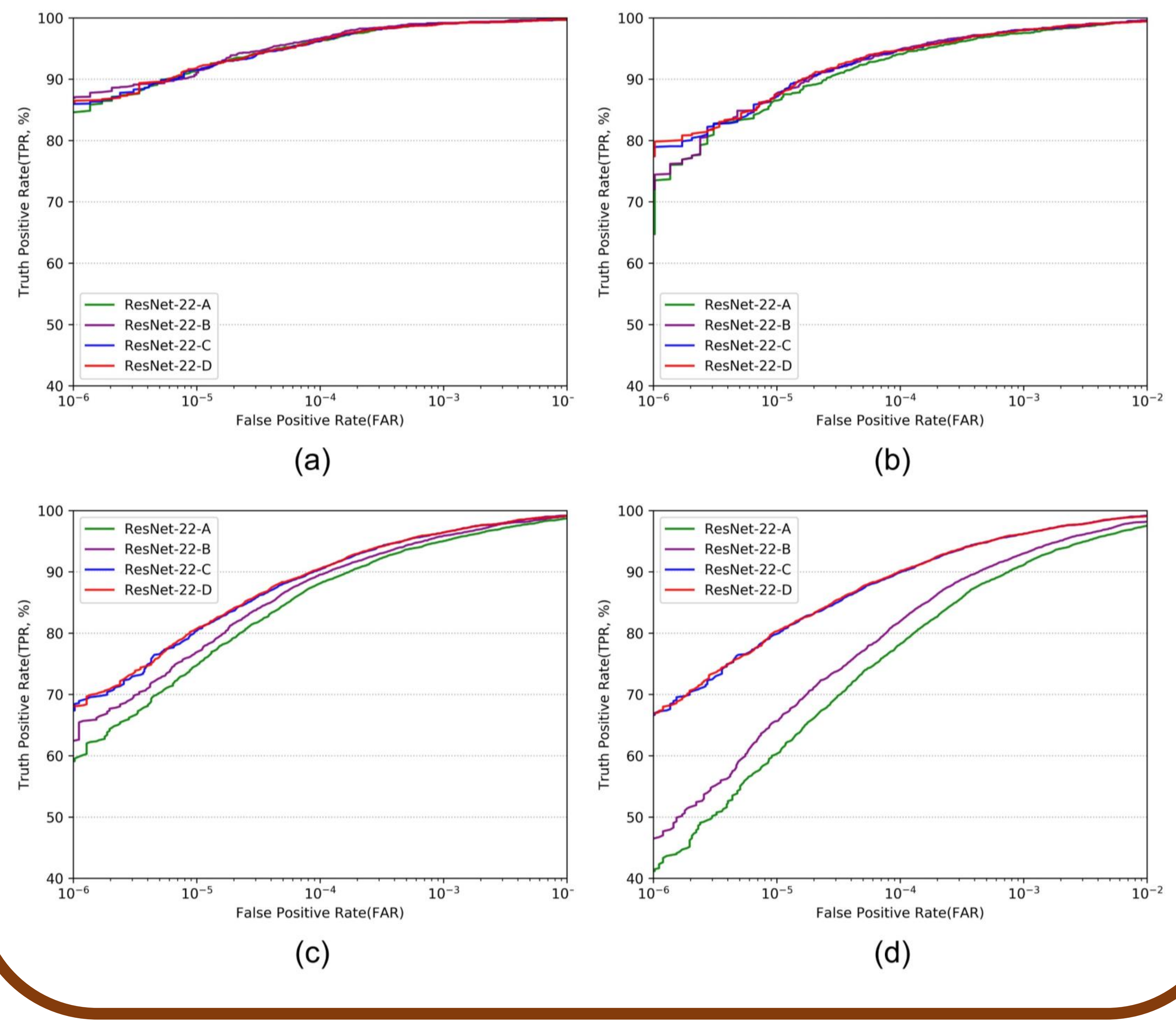
➢ Evaluation protocol & Benchmark

- **Protocol I:** all gallery and probe images are without eyeglasses.
- **Protocol II:** all gallery and probe images are with eyeglasses.
- **Protocol III:** all gallery images are without eyeglasses, and all probe images are with eyeglasses.
- **Protocol IV:** gallery images contain both eyeglass images and non-eyeglass images, so as probe images.

Training configurations

Model	Training Data	Loss	Strategy
ResNet-22-A	MsCeleb	A-Softmax	-
ResNet-22-B	MsCeleb	Mining-Contrasive	Finetune
ResNet-22-C	Mixture	Mining-Contrasive	Finetune
ResNet-22-D	Mixture	Mining-Contrasive	Finetune+GS

ROC curves of four protocols



4. Conclusion

- We propose to synthesize face images with eyeglasses as training data and propose a novel loss function to address the eyeglass robustness problem.
- Experiment results demonstrate that our proposed method is rather effective. the virtual-synthesis method may be extended to alleviate the impact of other factors on the robustness of face recognition
- We release the MeGlass dataset and build the evaluation protocols.

5. References

[1] Zhu X, Lei Z, Yan J, Yi D, Li SZ: High-fidelity pose and expression normalization for face recognition in the wild. CVPR (2015)

[2] Sun Y, Chen Y, Wang X, Tang X: Deep learning face representation by joint identification-verification. NIPS (2014)