# EFFICIENT PRACTICES FOR PROFILE-TO-FRONTAL FACE SYNTHESIS AND RECOGNITION

*Huijiao Wang*[1] *Xulei Yang*[2]

[1] School of Electronic Information, Wuhan University, China
[2] Institute for Infocomm Research (I2R), A*STAR, Singapore

## ABSTRACT

Despite the great progress of deep learning and generative adversarial networks, face frontalization (i.e., profile-to-frontal synthesis) and profile (i.e., non-frontal) face recognition still remain challenging tasks under uncontrolled environments. In this study, we propose three efficient practices to improve the performance of profile-to-frontal face synthesis and recognition. Firstly, the identity preserving module is embedded to constrain synthesized frontal images similar to true frontal faces in feature space. Secondly, facial consistency loss is employed to reduce the artifact of the generated frontal face in pixel space. Lastly, the multi-model embedded scheme enhances the representation learning through diverse facial features extracted by multiple facial feature extractors. The proposed practices are general, though specifically deployed to CR-GAN in this study for performance verification. Experimental results on Multi-PIE and VGGFace2 demonstrate that the proposed practices qualitatively generate more realistic photography frontal faces and quantitatively obtain better face recognition accuracy.

***Index Terms***— Face Frontalization, Face Recognition, Deep Learning, Generative Adversarial Networks

## 1. INTRODUCTION

In recent years, the face recognition technology based on deep learning has made great progress. In the ideal constrained environments, the performance of state-of-the-art algorithms has gone beyond the competence of human recognition. However, the recognition accuracy under unconstrained environments, such as low light, declines seriously. The recognition accuracy of profile-to-frontal face decreased by nearly $10\%$ compared with accuracy of frontal-to-frontal face [1]. This problem is becoming more and more important with the large-scale application of face recognition in the wild.

Face frontalization is the process of synthesizing frontal facing views of faces appearing in single unconstrained photos [2]. This, by transforming the challenging problem of recognizing faces viewed from unconstrained viewpoints to the easier problem of recognizing faces in constrained frontal facing poses, may provide a possible solution for profile face recognition. Recent methods [3, 4, 5, 6] have made great breakthroughs on the synthesis of frontal faces.

In this study, we propose three efficient practices to improve the performance of profile-to-frontal face synthesis and recognition. Firstly, the identity (ID) preserving module is embedded to the generator to restore personal ID feature. Secondly, facial consistency loss is used to force the symmetry and reduce the artifact of the generated frontal face. Lastly, multi-model embedded scheme is designed to enhance the representation learning through extracting diverse facial features by different facial feature extractors. The proposed practices are general and can be applied to any face frontalization approach. In this study, to verify the effectiveness of the proposed practices for profile face frontalizatoin and subsequent recognition, the complete representations generative adversarial network (CR-GAN [7]) is chosen as the base framework to generate frontal face from profile.

The main contributions of this paper are three folds: Firstly, three efficient practices are proposed to improve the performance of face fractalization and face recognition. The proposed practices can be generally deployed to any face fractalization architecture. Secondly, the effectiveness of the presented practices is verified by the extensive experiments on Multi-PIE [8] and VGGFace2 [9] datasets. To the best of our knowledge, we are the first to conduct face frontalization on the VGGFace2 wild face dataset. Lastly, the source codes of the implementation of the proposed practices are publicly shared[1], interested readers may re-use the source codes for their face frontalization related tasks.

## 2. RELATED WORKS

Face frontalization and recognition has obtained extensive attentions in past years. We briefly review the most used methods in this section.

**3D based Methods** Through local texture wrapping, 3D based methods, especially 3D morphable model (3DMM [10]) based methods, have achieved success in recovering 3D face shapes from single-view images. Hassner et al. [2] propose the earliest works to explore the approach of using a single 3D surface as an approximation to the shape of all input faces and generate new frontal views. Recent 3D based

[1] https://github.com/HJ0Wang/Face-Frontalization-Algorithm

methods are generally combined with deep learning techniques to further push the performance of face frontalization. In HF-PIM [11, 12], a high fidelity pose invariant model is proposed to bind 3D and 2D surface features to produce photographic and ID preserving results.

**GAN based Methods** In past few years, deep learning, especially GAN [13], based methods have been dominating the development of face frontalization. Occlusion-aware GAN [14] learns to reconstruct original images from synthesized images occluded with a limited set of objects. In TP-GAN [15], a two-pathway (global and local) GAN is proposed for photorealistic frontal view synthesis by simultaneously perceiving global structures and local details. While CR-GAN [7] proposes two learning pathways (labelled and unlabelled) to improve the performance of face synthesis to "unseen" data. And recently DA-GAN [3] uses a self-attention generator to yield better feature representations and a face-attention discriminator to reinforce the realism of synthetic frontal faces.

Compared to previous works, the proposed practices aim at retrieving ID information as much as possible during generating frontal images. Three practices constrain ID information from different perspectives: 1) In feature space, ID preserving module is embedded to constrain the generator to preserve personal ID feature. 2) In pixel space, facial consistency loss is used to force the symmetry and reduce the artifact of the generated frontal face. 3) In addition, multi-model embedded scheme is designed to recover diverse facial features from the input profile.

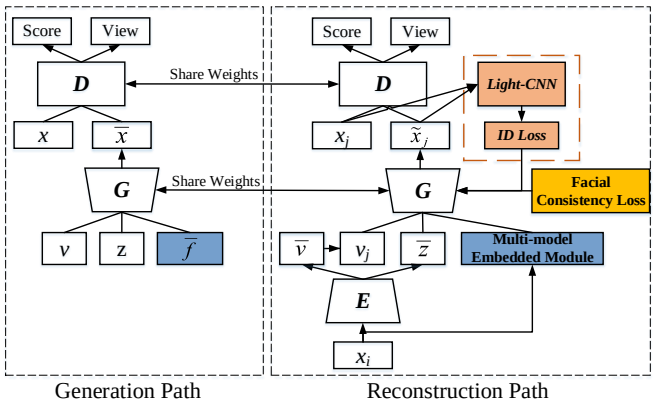


**Fig. 1**. Diagram of Optimized CR-GAN Structure. The three practices are highlighted by different colors. Zoom in for better visualization.

## 3. METHODOLOGY

We choose CR-GAN [7] as the baseline to generate frontal face from profile. In this section, we first brief the basic idea of CR-GAN and highlight the potential problems of CR-GAN, then present three efficient practices to improve the performance of the baseline.

### 3.1. Brief of CR-GAN

CR-GAN introduces a generative sideway with reconstruction path to maintain the completeness of the learned embedding space, which can improve the robustness of the model to face image generation task. This two-pathway framework solves the problem that the generator lacks the ability to deal with new data to some extent. However, CR-GAN lacks constrains of face identity feature and prior knowledge (e.g., symmetry) of the face, which declines the recognition performance of generated frontal images. We propose three measures to improve the weakness of CR-GAN and the quality of the generated frontal face.

### 3.2. ID Preserving Module

To ensure that enough identity information is retained during the generation of frontal face, an ID preserving module (brown-colored in Fig. 1) is embedded in the reconstruction path. The module consists of feature extraction network Light CNN-29 [16] and ID loss. As shown in Fig. 1, firstly, facial features of $x_j$ and $\tilde{x_j}$ are extracted by Light CNN-29. Then, the feature distances in feature space are calculated by ID loss. The loss function value is feedback to the generator $G$ in the reconstruction path for subsequent network's parameter updating. We choose cosine metric as the ID loss, which is shown as:

$$L_{\cos}(x_j, \tilde{x_j}) = \begin{cases} 1 - \cos(x_j, \tilde{x_j}), & \text{if } y = 1 \\ \max(0, \cos(x_j, \tilde{x_j}) - \text{m}), & \text{if } y = -1 \end{cases} \quad (1)$$

If $y$ is 1, it represents that $x_j$ and $\tilde{x_j}$ belong to the same person and $-1$ represents they are from two different persons. The $m$ is a constant which usually is set to $0.5$.

### 3.3. Facial Consistency Loss

The design of facial consistency loss (yellow-colored in Fig. 1) is based on the symmetry loss [15]. Facial symmetry is a prior information in human faces. Each pixel of the image relative to the central axis of the nose and mouth should be consistent in principle in Laplace space [15]. Therefore, the loss of facial consistency is the subtraction of the visible half of the face from the blocking half in pixel space. This loss is added to the generator of reconstruction path:

$$L_{sym} = \frac{1}{W/2 \times H} \sum_{i=1}^{W/2} \sum_{j=1}^{H} \left| I_{i,j}^{pred} - I_{W-(i-1),j}^{pred} \right|, \quad (2)$$

where $W$ and $H$ are the width and height of the input image respectively. $i$ and $j$ are pixel coordinates, and $I^{pred}$ represents the image generated by the generator $G$. During practice, adding a facial consistency loss to the generator helps to produce a more realistic image and meanwhile can accelerate the convergence speed of the network.

### 3.4. Multi-model Embedded Scheme

To retain as much facial information as possible from the input profiles, a multi-model embedded structure (blue-colored

in Fig. 1) is designed, where multi-model refers to multiple feature extractors to extract diverse face features. By using the combination of multiple facial feature extractors trained by different large datasets, facial features with wide diversity can be extracted from the same face. In the process of encoder $E$ estimation view, facial features can be input to the generator $G$ to a great extent, providing more identity information for the subsequent generation of frontal faces.

Considering inference time, lightweight models with high recognition accuracy and high speed are desirable. MobileFaceNet [17] and Light CNN-29 [16] are therefore selected as two feature extractors for the multi-model scheme. Since the generator $G$ shares weights between the generation path and reconstruction path, an additional input with the same dimension as the extracted features should be input to $G$ in the generation path. In order to speed up the convergence, an average feature vector $\bar{f}$ is calculated from 5 randomly selected images and added to the generation path as shown in Fig. 1.

### 3.5. Losses

**Generation path** The generation path trains generator $G$ and discriminator $D$. The encoder $E$ is not involved here since $G$ tries to generate from random noise [7]. $G$ aims to generate a realistic image $G(v, \mathbf{z}, \bar{f})$ from a view label $v$ , random noise $\mathbf{z}$ and the proposed average feature vector $\bar{f}$. $D$ tries to distinguish real data from $G$'s output, which minimizes:

$$\underset{\mathbf{z}\sim P_{\mathbf{z}}}{E}\Big[D_s\big(G(v,\mathbf{z},\bar{f})\big)\Big]-\underset{\mathbf{x}\sim P_{\mathbf{x}}}{E}\Big[D_s(\mathbf{x})\Big]+$$

$$\lambda_1\underset{\hat{\mathbf{x}}\sim P_{\hat{\mathbf{x}}}}{E}\Big[\big(\|\nabla_{\hat{\mathbf{x}}}D(\hat{\mathbf{x}})\|_2-1\big)^2\Big]-\lambda_2\underset{\mathbf{x}\sim P_{\mathbf{x}}}{E}\Big[P\big(D_v(\mathbf{x})=v\big)\Big], \tag{3}$$

where $P_{\mathbf{x}}$ is the data distribution and $P_{\mathbf{z}}$ is the noise uniform distribution, $P_{\hat{\mathbf{x}}}$ is an interpolation between pairs of points sampled from data distribution and the generator distribution, $G$ tries to fool $D$ and it maximizes:

$$\underset{\mathbf{z}\sim P_{\mathbf{z}}}{E}\Big[D_s\big(G(v,\mathbf{z},\bar{f})\big)\Big]+\lambda_3\underset{\mathbf{z}\sim P_{\mathbf{z}}}{E}\Big[P\big(D_v\big(G(v,\mathbf{z},\bar{f})\big)=v\big)\Big], \tag{4}$$

$D_v(\cdot)$ estimates the probability of being a specific view. $D_s(\cdot)$ describes the image quality, i.e., how real the image is.

**Reconstruction Path** The goal of reconstruction path is to synthesize the frontal face $\tilde{\mathbf{x}_j}$ from $\mathbf{x}_i$, where the $\tilde{\mathbf{x}_j}$ is the reconstruction of $\mathbf{x}_j$. D tries to minimizes:

$$\underset{\mathbf{x}_i,\mathbf{x}_j\sim P_{\mathbf{x}}}{E}\Big[D_s(\tilde{\mathbf{x}}_j)-D_s(\mathbf{x}_i)\Big]+$$
$$\lambda_1\underset{\hat{\mathbf{x}}\sim P_{\mathbf{x}}}{E}\Big[\big(\|\nabla_{\hat{\mathbf{x}}}D(\hat{\mathbf{x}})\|_2-1\big)^2\Big]-\lambda_2\underset{\mathbf{x}_i\sim P_{\boldsymbol{x}}}{E}\Big[P\big(D_v\left(\mathbf{x}_i\right)=v_i\big)\Big]. \tag{5}$$

E estimates views from $\mathbf{x}_i$ and it aims to maximize:

$$\begin{aligned}\underset{\mathbf{x}_i,\mathbf{x}_j\sim P_{\mathbf{x}}}{E}\big[&D_s\left(\tilde{\mathbf{x}}_j\right)+\lambda_3 P\left(D_v\left(\tilde{\mathbf{x}}_j\right)=v_j\right)-\\&\lambda_4 L_1\left(\tilde{\mathbf{x}}_j,\mathbf{x}_j\right)-\lambda_5 L_v\left(E_v\left(\mathbf{x}_i\right),v_i\right)\big].\end{aligned} \tag{6}$$

Combined with the proposed $L_{\cos}$ and $L_{sym}$, the loss function of the generator $G$ in the reconstruction path finally maximizes:

$$\begin{aligned}\underset{\mathbf{z}\sim P_{\mathbf{z}}}{E}\left[D_s(G(v,\mathbf{z}))\right]+\lambda_3\underset{\mathbf{z}\sim P_{\mathbf{z}}}{E}\left[P\left(D_v(G(v,\mathbf{z}))=v\right)\right]\\+\lambda_6 L_{\cos}-\lambda_7 L_{sym}.\end{aligned} \tag{7}$$

The $\lambda_1 \sim \lambda_7$ are weights of losses respectively, which are experimentally set to be $\lambda_1 = 10, \lambda_1 \sim \lambda_4 = 1, \lambda_5 = 0.01, \lambda_6 = 20, \lambda_7 = 1$.

## 4. EXPRIMENTS

We evaluate our network under both controlled and unconstrained settings for profiles recognition. For controlled settings, we perform ablation studies and comparisons with other methods on the CMU Multi-PIE database [8]. For unconstrained conditions, we show both qualitative and quantitative results on VGGFace2 database [9]. For quantitative evaluation, we evaluate face recognition performance using the learned facial representations by Light CNN-29 with a cosine distance metric.

### 4.1. Datasets and Experiment Setup

We choose the CMU Multi-PIE database [8] and VGGFace2 [9] to train and test the proposed network. Multi-PIE is the most commonly used database in profile-frontal transforming task, which is made in well-controlled conditions with precise labels of face angle. Similar to the CR-GAN [7], we conduct experiments on Setting-1 which concentrates on pose, illumination and minor expression variations. The setting-1 uses 150 identities for training and 100 identities for testing.

To further verify the robustness of our proposed network in unconstrained environments, we train our model on VGGFace2. The whole database is split to a training set (including 8631 identities) and a test set (including 500 identities). Due to the fact that VGGFace2 is collected from the Internet, without view labels which is a must if training on the proposed network. Therefore, we firstly estimate the view labels of images in VGGFace2 database using FSA-Net [18]. FSA-Net estimates roll, pitch and yaw angles for each head in VGGFace2. Then, we make view labels according to yaw angles of faces and divide the images with view interval of 15 degree, keeping a same format with Multi-PIE database. In order to promote convergence of the network, we filter the initial VGGFace2 by filtering out the hard examples, such as blurry images, sketches, images whose yaw angle is bigger than $60^\circ$ and so on. Finally, the training dataset contains 125149 images of 8631 identities and the testing dataset has 8690 images of 500 identities.

### 4.2. Ablation Studies

To gain an insight to different components of our proposed method, we compare variants of our model by combining various practices with the base CR-GAN network. Fig. 2 and Table 1 respectively show the qualitative and quantitative results of baseline (Light CNN-29), CR-GAN, CR-GAN

**Table 1**. Component analysis: rank-1 recognition rates (%) under Multi-PIE Setting-1

| Method | ±60° | ±45° | ±30° | ±15° |
|---|---|---|---|---|
| Light CNN-29 [16] | 73.30 | 97.45 | 99.80 | 99.78 |
| CR-GAN [7] | 74.12 | 77.15 | 78.42 | 78.58 |
| ID | 91.76 | 96.85 | 98.70 | 98.90 |
| MM | 85.59 | 94.78 | 94.50 | 96.00 |
| $\mathcal{L}_{sym}$ | 87.34 | 94.71 | 96.61 | 97.15 |
| *Ours | **93.52** | **98.64** | **99.80** | **99.80** |

with ID module (ID), CR-GAN with Multi-model Embedded (MM), CR-GAN with facial consistency loss ($\mathcal{L}_{sym}$), and our proposed network on the CMU Multi-PIE.

Fig. 2 shows synthesised frontal faces from different variants of our network. As expected, inference results with ID improves the perceptual performance of synthesised frontal faces most, no matter what the details of facial texture and lights. And variants with MM and $\mathcal{L}_{sym}$ both learn more realistic facial features compared with CR-GAN. Rank-1 recognition rates are compared under Setting-1 of the Multi-PIE in Table 1. We can observe an improvement of 20.22% under ±60° can be achieved with our proposed network. The ID preserving module contributes the most for improving the performance under large pose cases. For small view cases, our proposed method is slightly better than Light CNN-29.

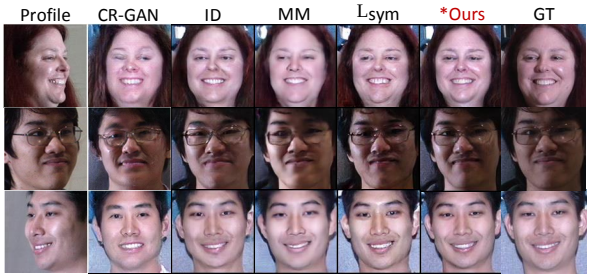


**Fig. 2**. Component analysis. Synthesized results of the proposed network and its variants

### 4.3. Comparison with Other Methods

*4.3.1. Evaluations on Multi-PIE Database*

Table 2 shows the Rank-1 recognition results of our proposed network with the baseline and other methods under Setting-1. The proposed network achieves the best performance across all poses (except compared with TP-GAN under ±30°), especially for large yaw angles. Compared to TP-GAN [15], our network has comparable results under ±30° and slightly higher performance under other angles. The proposed network outperforms the c-CNN [19] by 19.14% under ±60°.

*4.3.2. Evaluations on VGGFace2 Database*

To explore the performance of the proposed network under unconstrained conditions, we show the face recognition rates on VGGFace2. At a given threshold, the evaluation system measures the True Accept Rate (TAR), which is the fraction of genuine comparisons that correctly exceed the threshold, and

**Table 2**. Rank-1 recognition rates (%) under Multi-PIE Setting-1. "-" means the result is not reported

| Methods | ±60° | ±45° | ±30° | ±15° |
|---|---|---|---|---|
| Light CNN-29 [16] | 73.30 | 97.45 | 99.80 | 99.78 |
| CPF [20] | - | 71.65 | 81.05 | 89.45 |
| Hassner [2] | 44.81 | 74.68 | 89.59 | 96.78 |
| FV [21] | 68.71 | 80.33 | 87.21 | 93.30 |
| HPN [22] | 61.24 | 72.77 | 78.26 | 84.23 |
| FIP_40 [23] | 69.75 | 85.54 | 92.98 | 96.30 |
| c-CNN [19] | 74.38 | 89.02 | 94.05 | 96.97 |
| TP-GAN [15] | 92.93 | 98.58 | **99.85** | 99.78 |
| *Ours | **93.52** | **98.64** | 99.80 | **99.80** |

the False Accept Rate (FAR), which is the fraction of impostor comparisons that incorrectly exceed the threshold. Since the VGGFace2 contains so many images under unconstrained conditions, it is very difficult to do synthesised frontal face recognition task. Therefore, there is not so much results to compare on VGGFace2. When FAR is 0.01, the TAR of our proposed network is 75.92%. And under the same circumstance, the CR-GAN method with the weights file that the author provides is only 10.33% on the same testing dataset of VGGFace2.

Fig. 3 shows synthesised frontal images from profiles. The VGGFace2 does not contain profile-frontal pairs. So, we put the corresponding frontal image of input profile from the same person here. The synthesised frontal images in Fig. 3 recover identity features and global face shapes, which well verify that ID preserving module and Multi-model Embedded scheme can effectively advance generalizability. Especially considering a large varying in lights, background and poses, improving the synthesis quality is much harder on VGGFace2 than on the CMU Multi-PIE benchmark.

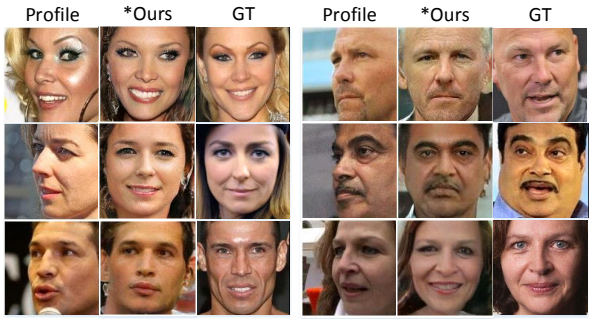


**Fig. 3**. Synthesized frontal faces on VGGFace2

## 5. CONCLUSION

The presented approach in this study combines three efficient practices to improve the accuracy of profile recognition, which transforms a profile face into a frontal face first then performs recognition on the generated frontal face. Experimental results on Multi-PIE [8] and VGGFace2 [9] database, especially a leap of accuracy on VGGface2, show the effectiveness of the presented practices.

## 6. REFERENCES


[1] S. Sengupta, J. Chen, C. Castillo, V. M. Patel, R. Chellappa, and D. W. Jacobs, "Frontal to profile face verification in the wild," in *WACV*, pp. 1–9, 2016.

[2] T. Hassner, S. Harel, E. Paz, and R. Enbar, "Effective face frontalization in unconstrained images," in *CVPR*, 2015.

[3] Y. Yin, S. Jiang, J. P. Robinson, and Y. Fu, "Dual-attention gan for large-pose face frontalization," in *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 249–256, 2020.

[4] Y. Ju, G. Lee, J. Hong, and S.-W. Lee, "Complete face recovery GAN: Unsupervised joint face rotation and de-occlusion from a single-view image," in *WACV*, pp. 1173–1183, 2022.

[5] H. Zhou, Y. Sun, W. Wu, C. C. Loy, X. Wang, and Z. Liu, "Pose-controllable talking face generation by implicitly modularized audio-visual representation," in *CVPR*, pp. 4176–4186, 2021.

[6] F. Taherkhani, V. Talreja, J. Dawson, M. C. Valenti, and N. M. Nasrabadi, "PF-cpGAN: Profile to frontal coupled gan for face recognition in the wild," in *IJCB*, pp. 1–10, 2020.

[7] Y. Tian, X. Peng, L. Zhao, S. Zhang, and D. N. Metaxas, "CR-GAN: Learning complete representations for multi-view generation," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pp. 942–948, 7 2018.

[8] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[9] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 67–74, 2018.

[10] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 187–194, 1999.

[11] J. Cao, Y. Hu, H. Zhang, R. He, and Z. Sun, "Towards high fidelity face frontalization in the wild," *IJCV*, vol. 128, pp. 1485–1504, 2020.

[12] J. Cao, Y. Hu, H. Zhang, R. He, and Z. Sun, "Learning a high fidelity pose invariant model for high-resolution face frontalization," in *NeurIPS*, pp. 1–11, 2018.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *International Conference on Neural Information Processing Systems*, pp. 2672–2680, 2014.

[14] J. Dong, L. Zhang, H. Zhang, and W. Liu, "Occlusion-aware gan for face de-occlusion in the wild," in *IEEE International Conference on Multimedia and Expo*, pp. 1–6, 2020.

[15] R. Huang, S. Zhang, T. Li, and R. He, "Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis," in *ICCV*, pp. 2439–2448, 2017.

[16] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2884–2896, 2018.

[17] S. Chen, Y. Liu, X. Gao, and Z. Han, "Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices," in *Biometric Recognition*, pp. 428–438, 2018.

[18] T. Yang, Y. Chen, Y. Lin, and Y. Chuang, "FSA-Net: Learning fine-grained structure aggregation for head pose estimation from a single image," in *CVPR*, 2019.

[19] C. Xiong, X. Zhao, D. Tang, K. Jayashree, S. Yan, and T.-K. Kim, "Conditional convolutional neural network for modality-aware face recognition," in *ICCV*, pp. 3667–3675, 2015.

[20] J. Yim, H. Jung, B. Yoo, C. Choi, D. Park, and J. Kim, "Rotating your face using multi-task deep neural network," in *CVPR*, pp. 676–684, 2015.

[21] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Fisher vector faces in the wild," in *BMVC*, pp. 4295–4304, 2015.

[22] C. Ding and D. Tao, "Pose-invariant face recognition with homography-based normalization," *Pattern Recognition*, vol. 66, pp. 144–152, 2017.

[23] Z. Zhu, P. Luo, X. Wang, and X. Tang, "Deep learning identity-preserving face space," in *ICCV*, pp. 113–120, 2013.