POSE-WEIGHTED GAN FOR PHOTOREALISTIC FACE FRONTALIZATION

Sufang Zhang[1,2]*, Qinghai Miao*[1]*, Min huang*[1,⋆]*, Xiangyu Zhu*[2]*,Yingying Chen*[2]*, Zhen Lei*[1,2]*,Jinqiao Wang*[1,2]

[1] University of Chinese Academy of Sciences [2] National Laboratory of Pattern Recognition,
Institute of Automation Chinese Academy of Sciences, Beijing, China, 100190
$\{zhangsufang16\}$@mails.ucas.ac.cn,$\{miaoqh, huangm\}$@ucas.ac.cn,
$\{xiangyu.zhu, yingying.chen, zlei, jqwang\}$@nlpr.ia.ac.cn

ABSTRACT

Face recognition methods have achieved high accuracy when faces are captured in frontal pose and constrained scenes. However, severe drop in accuracy is observed when large pose variations exist. The main reason is that the large yaw angle leads to ID information loss. In this paper, we intend to solve the large pose variations in a generation manner. Specifically, we propose a Pose-Weighted Generative Adversarial Network (PW-GAN) for photorealistic frontal view synthesis. We find frontalizing the faces in large poses $(yaw\ angle\ larger\ than\ 60°)$ is so difficult that the results are not photorealistic and the ID information is lost. To simplify the problem, we first frontalize the face image through 3D face model, which is then used to guide the network predicting. Second, we refine the pose code in the loss function to make the network pay more attention to large poses. Quantitative and qualitative experimental results on the Multi-PIE and LFW demonstrate our method achieves state of the art.

Index Terms— face frontalization, pose-invariant, face recognition, PW-GAN

1. INTRODUCTION

Benefiting from the deep learning based methods and the easy access to a large amount of annotated face databases [1, 2], face recognition techniques [3] have achieved significant advances in recent years. However, identifying face images in large poses is still a big challenge due to the large appearance variations. Existing methods that address pose variations can be divided into two categories. The first category tries to adopt hand-crafted or learned pose-invariant features [4, 5]. The second category aims to normalize face to the frontal view [6–8] from a given pose and then uses the recovered frontal images for face recognition [9, 10]. Early efforts on face frontalization in computer vision relied on 3D face geometry. The 3D Morphable Model (3DMM) [11] explicitly models facial shape and texture to match an input image as closely as possible. In this paper, we concentrate on the second category which is called face frontalization. Specifically, we aim to synthesize the frontal view image from a randomly given faces through 3D face geometry.

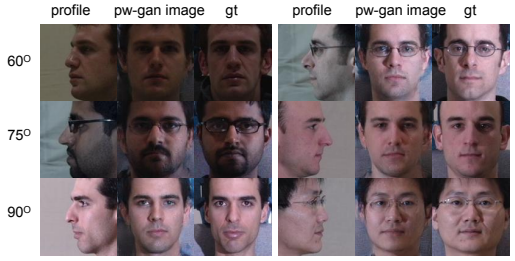

Fig. 1. Frontal face generated by PW-GAN,from first to third line correspond to $60°, 75°, 90°$, the first column is the profile image, the second column is generated image, the third column is the ground truth frontal image

Zhu $et\ al$.proposed High-fidelity Pose and Expression Normalization (HPEN) method [7] which can generate a natural face image in frontal pose based on 3D Morphable Model (3DMM). The HPEN method combines face landmark alignment and 3D face model fitting to achieve pose and expression normalization. The results are encouraging, but the synthesized images have artifacts and tend to be blurry in large poses. Other face frontalization methods are based on generative adversarial networks (GAN) [12], where visual realism has been improved significantly. Among them CAPG-GAN [13] and DR-GAN [14] encode pose into network in different ways. Inspired by this, we add the pose information to the loss function as weight to tell the network which image is in large pose and should pay more attention to. Another challenge in face recognition is that the frontalization method should predict the self-occluded regions, while the image frontalized by HPEN method can provide information of the self-occluded regions. PW-GAN combines pose code to loss function and image frontalized by HPEN to get a photorealistic result.

Based on the DR-GAN [14] network structure, we adopt the encoder-decoder structured generator to capture charac-

* corresponding author

teristics at different resolutions. We show some images generated by PW-GAN from pictures beyond 60° in Figure 1. The main contributions of our work in three folds:

- We design a novel loss function that can guide the network to pay more attention to the large angle images. The loss can better predict the lost pixels caused by self occlusion.
- Directly predicting the invisible facial regions through data-driven methods is too difficult due to the complicate face appearance variations and the lack of training data. In this paper, we find the face 3D geometry can well constrain the frontalization process and propose to utilize the HPEN results to help the GAN network learn the self-occluded regions.
- Through quantitative and qualitative experiment, PW-GAN method is proved effective for face recognition.

2. THE PROPOSED METHOD

Frontalization is to get a frontal face Y from a given profile image X in an arbitrary pose. Our training dataset contains M pairs $\{\mathrm{X,Y}\}$ from N identities. The network should not only learn to transform the profile X to frontal Y but also recover the self-occluded regions.

As shown in Figure 2, the network is based on DR-GAN. We refine the network by removing the step of adding pose code to the output of encoder and adding the pose code to loss functions to guide the predicting of network. Besides, we concatenate the source image and the frontalized image by HPEN as the input of the network. We show some pictures frontalized by HPEN in Figure 3. Although the synthesized image is not good enough, especially when the angle is larger than 60°, it still provides useful information for the self-occluded regions which will greatly assist the GAN to make a more accurate prediction. Since HPEN is based on the 3D face model, by adding the HPEN result we add the 3D structure constrains for the GAN network.

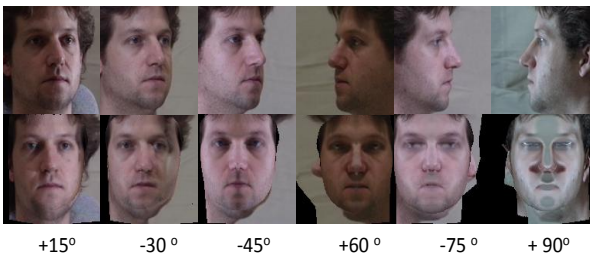

Fig. 3. Face images frontalized by HPEN. The angle gradually increases from left to right.

2.1. Network Architecture

The network is based on Generative Adversarial Network (GAN). In this paper, the formula of generator part is shown as:

$$\hat{Y} = G_{\theta_G}\left(X, I^F\right) \tag{1}$$

Where θ_G is the parameter of generator, X is the source profile image and I^F is the frontalized image by HPEN, which provides the self-occluded region's details. $\hat{Y}$ is the synthetic face.

2.2. Training Loss

PW-GAN is supervised by a weighted sum of several losses in an end-to-end manner, including pose-weighted pixel loss, symmetric loss, identity loss, classification loss and adversarial loss.

2.2.1. Pose-Weighted Pixel Loss

We propose a pose-weighted pixel loss on the synthesized face image$\hat{Y}$ to constrain the content consistency.

$$L_{pw-pix} = P \times \sum_{w,h,c}^{W,H,C} |\hat{Y} - Y|_1 \tag{2}$$

P is the pose code. which is 1 to 7 according to the angles $(0^{\circ} to\ 90^{\circ})$ in the Multi-PIE. W and H represent the width and height of the images. C is the number of channels.

2.2.2. Symmetric Loss

According to attributes of facial symmetry, the synthetic picture is symmetrical on both sides of the face.

$$L_{sym} = \sum_{c=1}^{C=3} |\hat{Y}_{x,y} - \hat{Y}_{W-x,y}|_1 \tag{3}$$

W is the width of images. The synthesis loss is indispensable. In experiments we found it must be appear with pose-weighted pixel loss.

2.2.3. Identity Preserving Loss

Preserving identity of the frontal view image is the most critical part in developing the 'recognition via generation' framework. In this work we adopt perceptual loss [15] that is originally for face recognition.

$$L_{cp} = |F(Y) - F\left(\hat{Y}\right)|_1 \tag{4}$$

F denotes the feature extraction network, where the feature is 256 dimensions, Y is the ground truth frontal face and $\hat{Y}$ is the synthesized face. We keep the synthetic identities is coincident with the ground truth image.

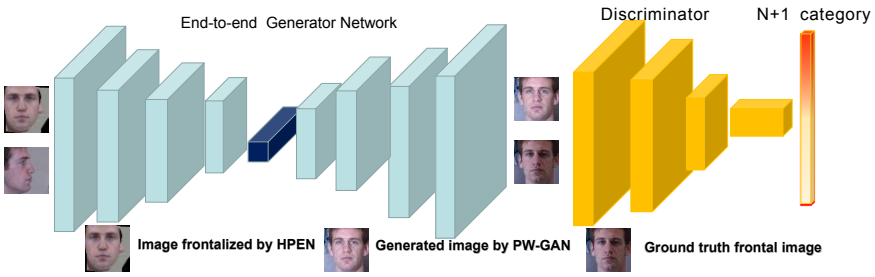

Fig. 2. Overview of the whole PW-GAN network. We concatenate the source image and the HPEN result as a 6-channels map as the CNN input which is fed into the encoder to get a pose-invariable feature. Then the pose invariable feature is further passed through the decoder to generate the frontal image. During the training stage, the predicted frontal image and the ground truth are grouped together to train the subsequent discriminator, which outputs N+1categories. N is the identity class and the last category represents the picture is true or false.

2.2.4. *Classification Loss*

In order to strengthen discriminative ability of discriminator, we add the classification loss to the discriminator, which can make faces with different IDs have the largest class differences:

$$L_{class} = |D_{\theta_G}\left(G_{\theta_G}\left(X, I^F\right)\right) - D_{\theta_D}\left(X\right)|_1 \tag{5}$$

L_{class} can strengthen the discriminator to distinguish the characteristics of different identity and push the generator to produce pictures with specific identity characteristics.

2.2.5. *Adversarial Loss*

The loss for distinguishing real frontal face images Y from synthesized frontal face $\hat{Y}$ is calculated as follows:

$$L_{adv} = D_{\theta_D}\left(G_{\theta_G}\left(X, I^F\right)\right) \tag{6}$$

L_{adv} serves as a supervision to push the synthesized image to match the original data distribution as much as possible. It can make the synthesized frontal face produce visually pleasing results. The discriminator is a simple shallow network with parameterD_θ.

2.2.6. *Overall Loss*

The total supervision loss is a weighted sum of the above losses. The PW-GAN network is trained alternately to optimize the following min-max problem

$$\begin{aligned} \min_{\theta_G} \max_{\theta_D} = \lambda_1 L_{pw-pix} + \lambda_2 L_{sym} + \lambda_3 L_{cp} \\ +\lambda_4 L_{class} + \lambda_5 L_{adv} \end{aligned} \tag{7}$$

where the parameters are $\lambda_1 = 20, \lambda_2 = 3 * 10^{-5}, \lambda_3 = 10, \lambda_4 = 0.1 \; and \; \lambda_5 = 0.3$. Reference to parameters published by TP-GAN and CAPG-GAN, we get the current optimal parameters.

3. EXPERIMENTS

3.1. Experimental settings and data set

The training set is Multi-PIE [16] which contains 337 identities, 20 illumination levels and 13 poses ranging from $-90°$ to $90°$. The training data is the first 200 subjects and the reminding 137 subjects are the test data. We follow the second setting protocol in [8, 17] for test: Each testing identity has one gallery image from his first appearance, probe and gallery sets respectively.

LFW [18] is widely used to evaluate synthesis and verification performance of various methods under unconstrained environments. For the verification protocol, face images are divided into 10 folds that contain different identities and 600 face pairs. We evaluate face verification performance on the frontalized images and compare PW-GAN with other face frontalization methods.

To train PW-GAN, images are $128\times128\times3$, adopting [19] for face alignment. Image's intensities are linearly scaled to the range of [-1,1]. We concatenate source profile image and HPEN image to six channels as the training data.

3.2. Face Synthesis

We conduct $\pm90°$ frontal face synthesis. Figure 4 shows the images generated by PW-GAN. We find that the images are identity-preserving and photographic.

Table 1 shows the quantitative face recognition performance compared with methods [8, 17] in the Multi-PIE. Follow the setting 2, we directly use the frontalized images following a 'recognition via generation'procedure.

Table 1 shows that PW-GAN is effective for large pose frontalization. In small pose $(less\ than\ 15°)$ we have very close performance with other methods. The reason is that in the Pose-Weighted Pixel Loss put more effort to large angle learning than small angle. Table 2 is the quantitative result of face synthesis in LFW. We compare PW-GAN with other

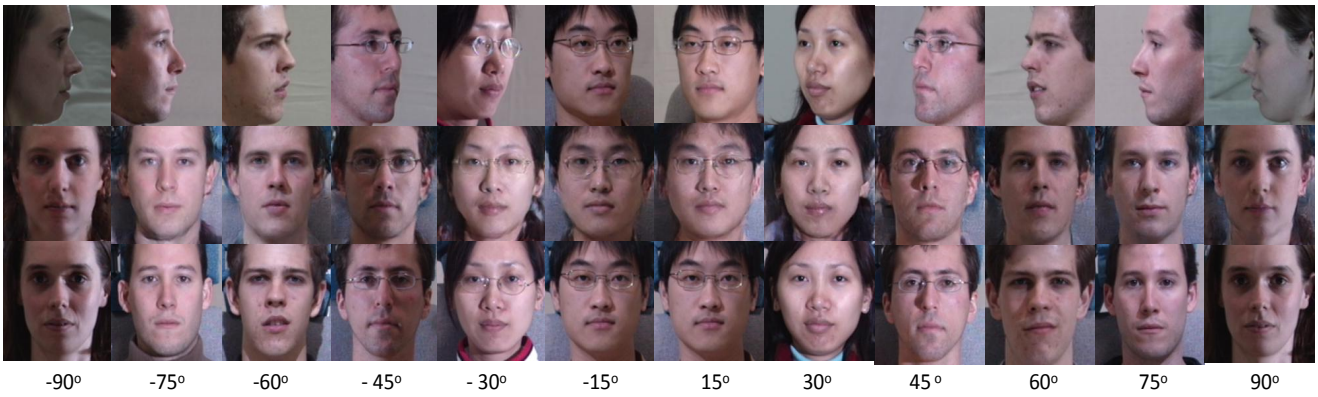

Fig. 4. Synthesis results from $0° - 90°$, the test images of different person generated by PW-GAN are shown from left to right. From the first to the third row are the source image, image generated by PW-GAN and the ground truth image respectively. The synthesized image preserves the facial attributes well, such as beard and eyeglasses.

Table 1. Rank-1 recognition rates $(\%)$ under setting 2.

Method	$\pm 90°$	$\pm 75°$	$\pm 60°$	$\pm 45°$	$\pm 30°$	$\pm 15°$
FIP+LDA [9]	-	-	45.9	64.1	80.7	90.7
MVP+LDA [10]	-	-	60.1	72.9	83.7	92.8
CPF [20]	-	-	61.9	79.9	88.5	95.0
DR-GAN [14]	-	-	83.2	86.2	90.1	94.0
Light-cnn [19]	5.51	24.18	62.09	92.13	97.38	98.59
HPEN [7]	38.17	63.21	80.33	95.31	99.03	99.15
FF-GAN [17]	61.2	77.2	85.2	89.7	92.5	94.6
TP-GAN [8]	64.64	77.43	87.72	95.38	98.06	98.68
PW-GAN	**72.48**	**79.71**	**93.27**	**96.89**	**98.32**	98.48

methods [21, 22]. Figure 5 is the synthesized images from LFW.

Table 2. Face verification accuracy (ACC)

Method	$ACC\ (\%)$
Ferrari et al [21]	-
LFW-3D [22]	93.62
LFW-HPEN [7]	96.25
FF-GAN [17]	96.42
PW-GAN	98.38

Table 2 shows the performances of frontalization methods in unconstrained scenes. By making full use of the 3D constrains from HPEN and the powerful generation ability of GAN, we get obvious improvements and achieves the state of the art.

3.3. Ablation Study

In this session, we investigate the improvements by adding the HPEN results.

Table 3. Ablation study about the proposed methods

Method	$\pm 90°$	$\pm 75°$	$\pm 60°$	$\pm 45°$	$\pm 30°$	$\pm 15°$
3c+pw-loss	70.56	78.91	90.47	96.39	98.06	98.06
6c+pw-loss	72.48	79.71	93.27	96.89	98.32	98.48

Fig. 5. The frontalized images for LFW, from the first line to the third line are the images from LFW, frontalized by PW-GAN and the HPEN respectively.

In table 3, we try to frontalize the faces from only the source image (3c+pw-loss) and concatenation of the source image and the HPEN result (6c+pw-loss). The 3c+pw-loss shows that PW-GAN is feasible for face frontalization but the 6c+pw-loss can provide more facial details to assist the network frontalizing face.

4. CONCLUSION

We propose two methods to improve the performance of GAN based face frontalization. Qualitative and quantitative experiments show that our method is effective.

5. ACKNOWLEDGEMENT

This work was supported by the National Science and Technology Major Project under grant 2016ZX05057007, Chinese National 921 Natural Science Foundation Projects #61876178, #61806196 and Chinese National Natural Science Foundation Projects #61772527,#61806200.

6. REFERENCES

[1] Qiong Cao, Shen Li, Weidi Xie, Omkar M. Parkhi, and Andrew Zisserman, "Vggface2: A dataset for recognising faces across pose and age," 2017.

[2] Yandong Guo, Zhang Lei, Yuxiao Hu, Xiaodong He, and Jianfeng Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," 2016.

[3] Yi Sun, Xiaogang Wang, and Xiaoou Tang, "Deep learning face representation from predicting 10,000 classes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1891–1898.

[4] Dong Chen, Xudong Cao, Fang Wen, and Jian Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3025–3032.

[5] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[6] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, , no. 12, pp. 2037–2041, 2006.

[7] Xiangyu Zhu, Zhen Lei, Junjie Yan, Dong Yi, and Stan Z Li, "High-fidelity pose and expression normalization for face recognition in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 787–796.

[8] Rui Huang, Shu Zhang, Tianyu Li, Ran He, et al., "Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis," *arXiv preprint arXiv:1704.04086*, 2017.

[9] Zhenyao Zhu, Luo Ping, Xiaogang Wang, and Xiaoou Tang, "Deep learning identity-preserving face space," in *IEEE International Conference on Computer Vision*, 2013.

[10] Z. Zhu, P. Luo, X. Wang, and X. Tang, "Multi-view perceptron: A deep model for learning face identity and view representations," in *International Conference on Neural Information Processing Systems*, 2014.

[11] Volker Blanz and Thomas Vetter, "Face recognition based on fitting a 3d morphable model," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.

[12] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672–2680, 2014.

[13] Yibo Hu, Xiang Wu, Bing Yu, Ran He, and Zhenan Sun, "Pose-guided photorealistic face rotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018.

[14] Luan Tran, Xi Yin, and Xiaoming Liu, "Disentangled representation learning gan for pose-invariant face recognition," in *CVPR*, 2017, vol. 3, p. 7.

[15] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*. Springer, 2016, pp. 694–711.

[16] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[17] Xi Yin, Xiang Yu, Kihyuk Sohn, Xiaoming Liu, and Manmohan Chandraker, "Towards large-pose face frontalization in the wild," in *Proc. ICCV*, 2017, pp. 1–10.

[18] Gary B. Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller, "Labeled Faces in the Wld: A Database forStudying Face Recognition in Unconstrained Environments," in *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, Oct. 2008, Erik Learned-Miller and Andras Ferencz and Frédéric Jurie.

[19] Xiang Wu, Ran He, Zhenan Sun, and Tieniu 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.

[20] Junho Yim, Heechul Jung, ByungIn Yoo, Changkyu Choi, Dusik Park, and Junmo Kim, "Rotating your face using multi-task deep neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 676–684.

[21] Claudio Ferrari, Giuseppe Lisanti, Stefano Berretti, and Alberto Del Bimbo, "Effective 3d based frontalization for unconstrained face recognition," in *Pattern Recognition (ICPR), 2016 23rd International Conference on*. IEEE, 2016, pp. 1047–1052.

[22] Tal Hassner, Shai Harel, Eran Paz, and Roee Enbar, "Effective face frontalization in unconstrained images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4295–4304.