PF-cpGAN: Profile to Frontal Coupled GAN for Face Recognition in the Wild

Fariborz Taherkhani [*], Veeru Talreja [*], Jeremy Dawson, Matthew C. Valenti, and Nasser M. Nasrabadi
West Virginia University
{ft0009, vtalreja}@mix.wvu.edu, jeremy.dawson@mail.wvu.edu, valenti@ieee.org, nasser.nasrabadi@mail.wvu.edu

Abstract

In recent years, due to the emergence of deep learning, face recognition has achieved exceptional success. However, many of these deep face recognition models perform relatively poorly in handling profile faces compared to frontal faces. The major reason for this poor performance is that it is inherently difficult to learn large pose invariant deep representations that are useful for profile face recognition. In this paper, we hypothesize that the profile face domain possesses a gradual connection with the frontal face domain in the deep feature space. We look to exploit this connection by projecting the profile faces and frontal faces into a common latent space and perform verification or retrieval in the latent domain. We leverage a coupled generative adversarial network (cpGAN) structure to find the hidden relationship between the profile and frontal images in a latent common embedding subspace. Specifically, the cpGAN framework consists of two GAN-based sub-networks, one dedicated to the frontal domain and the other dedicated to the profile domain. Each sub-network tends to find a projection that maximizes the pair-wise correlation between two feature domains in a common embedding feature subspace. The efficacy of our approach compared with the state-of-the-art is demonstrated using the CFP, CMU Multi-PIE, IJB-A, and IJB-C datasets.

1. Introduction

Due to the emergence of deep learning, face recognition has achieved exceptional success in recent years [3]. However, many of these deep face recognition models perform relatively poorly in handling profile faces compared to frontal faces [36]. In other words, face recognition in the wild, or unconstrained face recognition, is a challenging problem. Pose, expression, and lighting variations are considered to be major obstacles in attaining high unconstrained face recognition performance. Some methods [3, 27] attempt to address pose-variation issue by learning pose-invariant features, while some other methods [7, 31, 47, 53, 55] try to normalize images (along with identity-preservation) to a single frontal pose before recognition. However, there are three major difficulties related to face frontalization or normalization in unconstrained environment:

Figure 1: Frontal and Profile Images from Celebrity in Frontal Profile (CFP) Dataset.

- Complicated face-variations besides pose: In comparison to a controlled environment, there are more complex face variations, e.g., lighting, head pose, expression, in real-world scenarios. It is difficult to directly warp the input face to a normalized view [31].
- Unpaired data: Undoubtedly, obtaining a strictly normalized face is expensive and time-consuming, but getting an effective pair of target normalized face (i.e., frontal-facing, neutral expression) and an input face is difficult due to highly imbalanced datasets [31].
- Presence of artifacts: Synthesized 'frontal' faces contain artifacts caused by occlusions and non-rigid expressions.

In this paper, we hypothesize that the profile face domain possesses a gradual connection with the frontal face domain in a latent deep feature subspace. We aim to exploit this connection by projecting the profile faces and frontal faces into a common latent subspace and perform verification or retrieval in this latent domain. We propose an embedding model for profile to frontal face verification based on a deep coupled learning framework which uses a gen-

[*]Authors Contributed Equally

978-1-7281-9186-7/20/$31.00 ©2020 IEEE

erative adversarial network (GAN) to find the hidden relationship between the profile face features and frontal face features in a latent common embedding subspace.

Our work is conceptually related to the embedding category of super-resolution [19,24,37,56] in that our approach also performs verification of profile and frontal face in the latent space but not in the image space. From our experiments, we observe that transforming profile and frontal face features to a latent embedding subspace could yield higher performance than image-level face frontalization, which is susceptible to the negative influence of artifacts as a result of image synthesis. To our best knowledge, this study is the first attempt to perform profile-to-frontal face verification in the latent embedding subspace using generative modeling.

This paper makes the following contributions:

- A novel profile to frontal face recognition model using coupled GAN framework with multiple loss functions is developed.
- Comprehensive experiments using different datasets and a comparison of the proposed method with the state-of-the-art methods have been performed, indicating the efficacy of the proposed GAN framework.
- The proposed framework can potentially be used to improve the performance of traditional face recognition methods by integrating it as a preprocessing procedure for face-frontalization schema.

2. Related Work

Face recognition using Deep Learning: Before the advent of deep learning, traditional methods for face recognition (FR) used one or more layer representations, such as the histogram of the feature codes, filtering responses, or distribution of the dictionary atoms [50]. FR research was concentrated more toward separately improving preprocessing, local descriptors, and feature transformation; however, overall improvement in FR accuracy was very slow. This all changed with the advent of deep learning, and now deep learning is the prominent technique used for FR.

Recently various deep learning models such as [8, 43] are used as baseline model for FR. Simultaneously, various loss functions have been explored and used in FR. These loss functions can be categorized as the Euclidean-distance-based loss, angular/cosine-margin-based loss, and softmax loss and its variations. The contrastive loss and the triplet loss are the commonly used Euclidean-distance-based loss functions [34,40–42]. For avoiding misclassification of difficult samples [45, 46], the learned face features need to be well separated. Angular/cosine-margin based loss [2,10,25] are commonly used to make the learned features more separable with a larger angular/cosine distance. Finally, in the category of softmax loss and its variants for FR [14,26,49], the softmax loss is modified to improve the FR performance as in [26], where the cosine distance among data features is optimized along with normalization of features and weights.

Profile-Frontal Face Recognition: Face recognition with pose variation in an unconstrained environment is a very challenging problem. Existing methods focus on the pose variation problem by training separate models for learning pose-invariant features [3, 27], elaborate dense 3D facial landmark detection and warping [44], and synthesizing a frontal, neutral expression face from a single image [7, 31, 47, 53, 55]. For instance, Cao *et al*. [3] exploit the inherent mapping between profile and frontal faces, and transform a deep profile face representation to a canonical pose by adaptively adding residuals. FF-GAN [55] solves the problem of large-pose face frontalization in the wild by incorporating a 3D face model into a GAN. Considering photorealistic and identity preserving frontal view synthesis, a domain adaptation strategy for pose invariant face recognition is discussed in [58]. Tran *et al*. [47] propose a GAN framework to rotate the face and disentangle the identity representation by using the pose code. In [31], a face normalization model (FNM) uses a generative adversarial network (GAN) network with 3 distinct losses for generating canonical-view and expression-free frontal images.

3. Generative Adversarial Network

GAN was first introduced by Goodfellow *et al*. [12] and has drawn great attention from the deep learning research community due to its remarkable performance on generative tasks. The GAN framework is based on two competing networks — a generator network G and a discriminator network D. The generator $G(z;\theta_g)$ is a differentiable function which maps the noise variable z from training noise distribution $p_z(z)$ to a data space with distribution p_{data} using the network parameters θ_g. On the other hand, the discriminator $D(.;\theta_d)$ is also a differentiable function, which discriminates between the real data y and the generated fake data $G(z)$ using a binary classification model. Specifically, the min-max two-player game between the generator and the discriminator provides a simple and powerful way to estimate target distribution and generate novel image samples [31]. The loss function $L(D,G)$ for GAN is given as:

$$\begin{aligned} L(D,G) = {} & E_{y\sim P_{data}(y)}[\log D(y)] \\ & + E_{z\sim P_z(z)}[\log(1-D(G(z)))]. \end{aligned} \tag{1}$$

The objective (two player minimax game) for GAN is as:

$$\begin{aligned} \min_G \max_D L(D,G) = \min_G \max_D [& E_{y\sim P_{data}(y)}[\log D(y)] \\ & + E_{z\sim P_z(z)}[\log(1-D(G(z)))]]. \end{aligned} \tag{2}$$

Another variant of GAN is the Conditional GAN, which was introduced by Mirza and Osindero [29]. In conditional

GAN, both the generator and discriminator are conditioned on an additional variable x. This additional variable could be any kind of auxiliary information such as discrete labels [29], and text [32]. The loss function for the conditional GAN is given as:

$$L_c(D,G) = E_{y\sim P_{data}(y)}[\log D(y|x)] + E_{z\sim P_z(z)}[\log(1 - D(G(z|x)))]. \quad (3)$$

Hereafter, we will denote the objective for the conditional GAN as $F_{cGAN}(D,G,y,x)$, which is given by:

$$F_{cGAN}(D,G,y,x) = \min_G \max_D [E_{y\sim P_{data}(y)}[\log D(y|x)] + E_{z\sim P_z(z)}[\log(1 - D(G(z|x)))]]. \quad (4)$$

4. Proposed Method

Here, we describe our method for profile to frontal face recognition. In contrast to the face normalization methods, we do not perform pose normalization (i.e., frontalization) on each profile image before matching. Instead, we seek to project the profile and frontal face images to a common latent low-dimensional embedding subspace using generative modeling. Inspired by the success of GANs [12], we explore adversarial networks to project profile and frontal images to a common subspace for recognition.

The framework of proposed profile to frontal coupled generative adversarial network (PF-cpGAN; shown in Fig. 2) consists of two modules, where each module contains a GAN architecture made of a generator and a discriminator. The generators that we have used in both modules are U-net auto-encoders that are coupled together using a contrastive loss function. In addition to adversarial, and contrastive loss, we propose to guide the generators using a perceptual loss [20] based on the VGG 16 architecture, as well as an L_2 reconstruction error. The perceptual loss helps in sharp and realistic reconstruction of the images.

4.1. Profile to Frontal Coupled GAN

The main objective of PF-cpGAN is the recognition of profile face images with respect to a gallery of frontal face images, which have not been seen during the training. The matching of the profile and the frontal face images is performed in a common embedding subspace. PF-cpGAN consists of two modules: a profile GAN module and a frontal GAN module, both consisting of a GAN (generator + discriminator), and a perceptual network based on VGG-16.

For the generators, we use a U-Net [33] auto-encoder architecture. The primary reason for using U-Net is that the encoder-decoder structure tends to extract global features and generates images by leveraging this overall information, which is very useful for global shape transformation

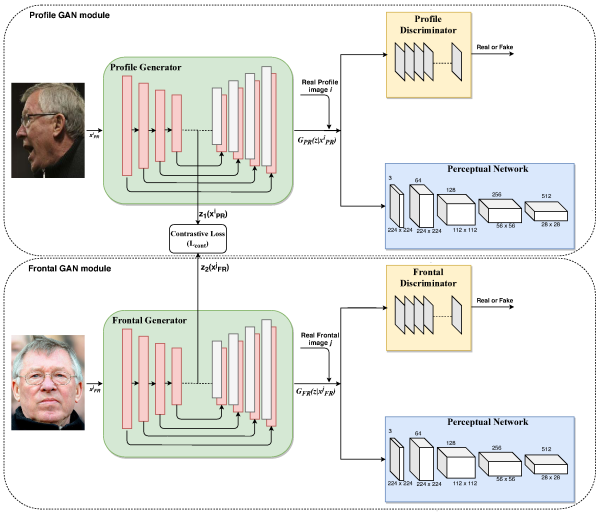


Figure 2: Block diagram of PF-cpGAN.

tasks such as profile to frontal image conversion. Moreover, for many image translation problems, there is a significant amount of low-level information that needs to be shared between the input and output, and it is desirable to pass this information directly across all the layers including the bottleneck. Therefore, using skip-connections, as in U-net, provides a means for the encoder-decoder structure to circumvent the bottleneck and pass the information over to other layers.

For discriminators, we have used patch-based discriminators [18], which are trained iteratively along with the respective generators. L_1 loss performs very well when trying to preserve the low frequency details but fails to preserve the high-frequency details. However, using a patch-based discriminator that penalizes structure at the scale of the patches, ensures the preservation of high-frequency details, which are usually lost when only L_1 loss is used.

The final objective of PF-cpGAN is to find the hidden relationship between the profile face features and frontal face features in a latent common embedding subspace. To find this common subspace between the two domains, we couple the two generators via a contrastive loss function, L_{cont}.

This loss function (L_{cont}) is a distance-based loss function, which tries to ensure that semantically similar examples (genuine pairs i.e., a profile image of a subject with its corresponding frontal image) are embedded closely in the common embedding subspace, and, simultaneously, semantic dissimilar examples (impostor pairs i.e., a profile image of a subject and a frontal image of a different subject) are pushed away from each other in the common embedding subspace. The contrastive loss function is defined as:

$$L_{cont}(z_1(x^i_{PR}), z_2(x^j_{FR}), Y) = (1-Y)\frac{1}{2}(D_z)^2 + (Y)\frac{1}{2}(\max(0, m - D_z))^2, \quad (5)$$

where x^i_{PR} and x^j_{FR} denote the input profile and frontal face image, respectively. The variable Y is a binary label, which is equal to 0 if x^i_{PR} and x^j_{FR} belong to the same class (i.e., genuine pair), and equal to 1 if x^i_{PR} and x^j_{FR} belong to the different class (i.e., impostor pair). $z_1(.)$ and $z_2(.)$ denote only the encoding functions of the U-Net auto-encoder to transform x^i_{PR} and x^j_{FR}, respectively into a common latent embedding subspace. The value m is the contrastive margin and is used to "tighten" the constraint. D_z denotes the Euclidean distance between the outputs of the functions $z_1(x^i_{PR})$ and $z_2(x^j_{FR})$.

$$D_z = \left\| z_1(x^i_{PR}) - z_2(x^j_{FR}) \right\|_2. \quad (6)$$

Therefore, if $Y = 0$ (i.e., genuine pair), then the contrastive loss function (L_{cont}) is given as:

$$L_{cont}(z_1(x^i_{PR}), z_2(x^j_{FR}), Y) = \frac{1}{2}\left\| z_1(x^i_{PR}) - z_2(x^j_{FR}) \right\|^2_2, \quad (7)$$

and if $Y = 1$ (i.e., impostor pair), then contrastive loss function (L_{cont}) is :

$$L_{cont}(z_1(x^i_{PR}), z_2(x^j_{FR}), Y) = \frac{1}{2}\max\left(0, m - \left\| z_1(x^i_{PR}) - z_2(x^j_{FR}) \right\|^2_2\right). \quad (8)$$

Thus, the total loss for coupling the profile generator and the frontal generator is denoted by L_{cpl} and is given as:

$$L_{cpl} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} L_{cont}(z_1(x^i_{PR}), z_2(x^j_{FR}), Y), \quad (9)$$

where N is the number of training samples. The contrastive loss in the above equation can also be replaced by some other distance-based metric, such as the Euclidean distance. However, the main aim of using the contrastive loss is to be able to use the class labels implicitly and find the discriminative embedding subspace, which may not be the case with some other metric such as the Euclidean distance. This discriminative embedding subspace would be useful for matching of the profile images with the frontal images.

4.2. Generative Adversarial Loss

Let the generators (profile generator and frontal generator) that reconstruct the corresponding profile and frontal image from the input profile and frontal image, be denoted as G_{PR} and G_{FR}, respectively. The patch-based discriminators used for the profile and frontal GANs are denoted as D_{PR} and D_{FR}. For the proposed method, we have used the conditional GAN, where the generator networks G_{PR} and G_{FR} are conditioned on input profile and frontal face images, respectively. We have used the conditional GAN loss function [29] to train the generators and the corresponding discriminators in order to ensure that the discriminators cannot distinguish the images reconstructed by the generators from the corresponding ground truth images. Let L_{PR} and L_{FR} denote the conditional GAN loss functions for the profile and the frontal GANs, respectively, where L_{PR} and L_{FR} are given as:

$$L_{PR} = F_{cGAN}(D_{PR}, G_{PR}, y^i_{PR}, x^i_{PR}), \quad (10)$$

$$L_{FR} = F_{cGAN}(D_{FR}, G_{FR}, y^j_{FR}, x^j_{FR}), \quad (11)$$

where function F_{cGAN} is the conditional GAN objective defined in (4). The term x^i_{PR} denotes the profile image used as a condition for the profile GAN, and y^i_{PR} denotes the real profile image. Note that the real profile image y^i_{PR} and the network condition given by x^i_{PR} are the same. Similarly, x^j_{FR} denotes the frontal image used as a condition for the frontal GAN and y^j_{FR} denotes the real frontal image. Again, the real frontal image y^j_{FR} and the network condition given by x^j_{FR} are the same. The total loss for the coupled conditional GAN is given by:

$$L_{GAN} = L_{PR} + L_{FR}. \quad (12)$$

4.3. L_2 Reconstruction Loss

We also consider the L_2 reconstruction loss for both the profile GAN and frontal GAN. The L_2 reconstruction loss measures the reconstruction error in terms of the Euclidean distance between the reconstructed image and the corresponding real image. Let $L_{2_{PR}}$ denote the reconstruction loss for the profile GAN and is defined as:

$$L_{2_{PR}} = \left\| G_{PR}(z|x^i_{PR}) - y^i_{PR} \right\|^2_2, \quad (13)$$

where y^i_{PR} is the ground truth profile image, $G_{PR}(z|x^i_{PR})$ is the output of the profile generator.

Similarly, Let $L_{2_{FR}}$ denote the reconstruction loss for the frontal GAN:

$$L_{2_{FR}} = \left\| G_{FR}(z|x^j_{FR}) - y^j_{FR} \right\|^2_2, \quad (14)$$

where y^j_{FR} is the ground truth frontal image, $G_{FR}(z|x^j_{FR})$ is the output of the frontal generator.

The total L_2 reconstruction loss function is given by:

$$L_2 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(L_{2_{PR}} + L_{2_{FR}}). \quad (15)$$

4.4. Perceptual Loss

In addition to the GAN loss and the reconstruction loss which are used to guide the generators, we have also used perceptual loss, which was introduced in [20] for style transfer and super-resolution. The perceptual loss function is used to compare high level differences, like content and style discrepancies, between images. The perceptual loss function involves comparing two images based on high-level representations from a pretrained CNN, such as VGG-16 [39]. The perceptual loss function is a good alternative to solely using L_1 or L_2 reconstruction error, as it gives better and sharper high quality reconstruction images [20].

In our proposed approach, perceptual loss is added to both the profile and the frontal module using a pre-trained VGG-16 [39] network. We extract the high-level features (ReLU3-3 layer) of VGG-16 for both the real input image and the reconstructed output of the U-Net generator. The L_1 distance between these features of real and reconstructed images is used to guide the generators G_{PR} and G_{FR} . The perceptual loss for profile network is defined as:

$$L_{P_{PR}} = \frac{1}{C_p W_p H_p} \sum_{c=1}^{C_p} \sum_{w=1}^{W_p} \sum_{h=1}^{H_p} \left\| V(G_{PR}(z|x^i_{PR}))^{c,w,h} - V(y^i_{PR})^{c,w,h} \right\|, \tag{16}$$

where $V(.)$ denotes a particular layer of the VGG-16, where the layer dimensions are given by C_p, W_p, and H_p.

Likewise the perceptual loss for frontal network is:

$$L_{P_{FR}} = \frac{1}{C_p W_p H_p} \sum_{c=1}^{C_p} \sum_{w=1}^{W_p} \sum_{h=1}^{H_p} \left\| V(G_{FR}(z|x^j_{FR}))^{c,w,h} - V(y^j_{FR})^{c,w,h} \right\|. \tag{17}$$

The total perceptual loss function is given by:

$$L_P = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (L_{P_{PR}} + L_{P_{FR}}). \tag{18}$$

4.5. Overall Objective Function

The overall objective function for learning the network parameters in the proposed method is given as the sum of all the loss functions defined above:

$$L_{tot} = L_{cpl} + \lambda_1 L_{GAN} + \lambda_2 L_P + \lambda_3 L_2, \tag{19}$$

where L_{cpl} is the coupling loss, L_{GAN} is the total generative adversarial loss, L_P is the total perceptual loss, and L_2 is the total reconstruction error. Variables λ_1, λ_2, and λ_3 are the hyper-parameters to weigh the different loss terms.

5. Experiments

We initially describe our training setup and the datasets that we have used in our experiments. We show the efficiency of our method for the task of frontal to profile face verification by comparing its performance with state-of the-art face verification methods across pose-variation. Moreover, we explore the effect of face yaw in our algorithm. Finally, we conduct an ablation study to investigate the effect of each term in our total training loss defined in (19).

5.1. Experimental Details

Datasets: The Celebrities in Frontal-Profile (CFP) dataset [36] is a mixture of constrained (i.e., carefully collected under different pose, illumination and expression conditions) and unconstrained (i.e., collected images from the Internet) settings. CFP includes 500 celebrities, averaging ten frontal and four profile face images per each celebrity. Following the standard 10-fold protocol [36], we divide the dataset into 10 folds, each of which consists of 350 same and 350 different pairs generated from 50 subjects (i.e., 7 same and 7 different pairs for each subject).

The CMU Multi-PIE database [13] contains 750,000 images of 337 subjects. Subjects were imaged from 15 viewing angles and 19 illumination conditions while exhibiting a range of facial expressions. It is the largest database for graded evaluation with respect to pose, illumination, and expression variations. For fair comparison, the database setting was made consistent with FNM [31], where 250 subjects from Multi-PIE have been used. The training set and testing split is consistent with FNM.

The IARPA Janus Benchmark A (IJB-A) [22] is a challenging dataset collected under complete unconstrained conditions covering full pose variation (yaw angles -90° to $+90^\circ$). IJB-A contains 500 subjects with 5,712 images and 20,414 frames extracted from videos. Following the standard protocol in [22], we evaluate our method on both verification and identification. The IARPA Janus Benchmark B (IJB-B) dataset [51] builds on the IJB-A by adding more 1345 subjects making it a total of 1845 subjects, and a total of 21,798 still images and 55,026 frames from 7,011 videos. The IARPA Janus Benchmark C (IJB-C) dataset [28] builds on IJB-A, and IJB-B datasets and has a total of 31,334 images for a total number of 3,531 subjects. We have also evaluated our method on IJB-A and IJB-C datasets.

Implementation Details: We have implemented the U-Net with ResNet-18 [15] encoder pre-trained on ImageNet. We have added an additional fully-connected layer after the average pooling layer for ResNet-18 for our U-Net encoder. The U-Net decoder has the same number of layers as the encoder. The entire framework has been implemented in Pytorch. For convergence, λ_1 is set to 1, and λ_2, and λ_3 are both set to 0.25. We used a batch size of 128 and an Adam optimizer [21] with first-order momentum of 0.5, and

Table 1: Performance comparison on CFP dataset. Mean Accuracy and equal error rate (EER) with standard deviation over 10 folds.

	Frontal-Profile		Frontal-Frontal	
Algorithm	Accuracy	EER	Accuracy	EER
HoG+Sub-SML [36]	77.31(1.61)	22.20(1.18)	88.34(1.31)	11.45(1.35)
LBP+Sub-SML [36]	70.02(2.14)	29.60(2.11)	83.54(2.40)	16.00(1.74)
FV+Sub-SML [36]	80.63(2.12)	19.28(1.60)	91.30(0.85)	8.85(0.74)
FV+DML [36]	58.47(3.51)	38.54(1.59)	91.18(1.34)	8.62(1.19)
Deep Features [5]	84.91(1.82)	14.97(1.98)	96.40(0.69)	3.48(0.67)
PR-REM [3]	93.25(2.23)	7.92(0.98)	98.10(2.19)	1.10(0.22)
PF-cpGAN	93.78(2.46)	7.21(0.65)	98.88(1.56)	0.93(0.14)

learning rate of 0.0004. We have used the ReLU activation function for the generator and Leaky ReLU with a slope of 0.3 for the discriminator.

For training, genuine and impostor pairs were required. The genuine/impostor pairs are created by frontal and profile images of the same/different subject. During the experiments, we ensure that the training set are balanced by using the same number of genuine and impostor pairs.

5.2. Evaluation on CFP with Frontal-Profile Setting

We first perform evaluation on the Celebrities in Frontal-Profile (CFP) dataset [36], a challenging dataset created to examine the problem of frontal to profile face verification in the wild. We follow the standard 10-fold protocol [36] in our evaluation. The same protocol is applied on both the Frontal-Profile and Frontal-Frontal settings. For fair comparison and as given in [36], we consider different types of feature extraction techniques like HoG [9], LBP [1], and Fisher Vector [38] along with metric learning techniques like Sub-SML [4], and Diagonal metric learning (DML) as reported in [38]. We also compare against deep learning techniques, including Deep Features [5], and PR-REM [3]. The results are summarized in Table 1.

We can observe from Table 1 that our proposed framework, PF-cpGAN, gives much better performance than the methods that use standard hand-crafted features of HoG, LBP, or FV, providing minimum of 13% improvement in accuracy with a 12% decrease in EER for the profile-frontal setting. PF-cpGAN also improves on the performance of the Deep Features by approximately 9% with a 7.5% decrease in EER for the profile-frontal setting. Finally, PF-cpGAN performs on-par with the best deep learning method of PR-REM, and, in-fact, does slightly better than PR-REM by $\approx 0.5\%$ improvement in accuracy with a 0.7% decrease in EER for the profile-frontal setting. This performance improvement clearly shows that usage of a GAN framework for projecting the profile and frontal images in the latent embedding subspace and maintaining the sematic similarity in the latent space is better than some other deep learning techniques such as Deep Features or PR-REM.

Table 2: Performance comparison on IJB-A benchmark. Results reported are the 'average±standard deviation' over the 10 folds specified in the IJB-A protocol. Symbol '-' indicates that the metric is not available for that protocol.

Method	Verification		Identification	
	GAR@ FAR= 0.01	GAR@ FAR= 0.001	@ Rank-1	@ Rank-5
OPENBR [23]	23.6 ± 0.9	10.4 ± 1.4	24.6 ± 1.1	37.5 ± 0.8
GOTS [23]	40.6 ± 1.4	19.8 ± 0.8	43.3 ± 2.1	59.5 ± 2.0
PAM [27]	73.3 ± 1.8	55.2 ± 3.2	77.1 ± 1.6	88.7 ± 0.9
DCNN [6]	78.7 ± 4.3	-	85.2 ± 1.8	93.7 ± 1.0
DR-GAN [48]	77.4 ± 2.7	53.9 ± 4.3	85.5 ± 1.5	94.7 ± 1.1
FF-GAN [54]	85.2 ± 1.0	66.3 ± 3.3	90.2 ± 0.6	95.4 ± 0.5
FNM [31]	93.4 ± 0.9	83.8 ± 2.6	96.0 ± 0.5	98.6 ± 0.3
PR-REM [3]	94.4 ± 0.9	86.8 ± 1.5	94.6 ± 1.1	96.8 ± 1.0
PF-cpGAN	95.8 ± 0.82	91.2 ± 1.3	97.6 ± 1.0	98.8 ± 0.4

Table 3: Performance comparison on IJB-C benchmark. Results reported are the 'average±standard deviation' over the 10 folds specified in the IJB-C protocol. Symbol '-' indicates that the metric is not available for that protocol.

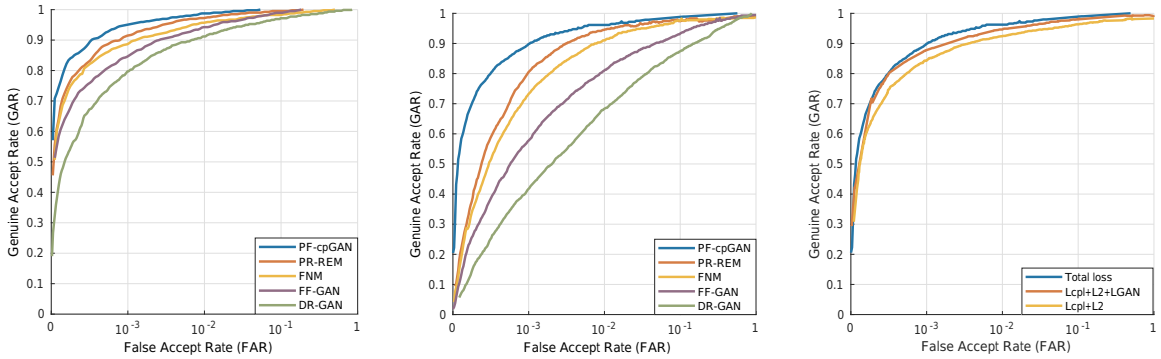
Method	Verification		Identification	
	GAR@ FAR= 0.01	GAR@ FAR= 0.001	@ Rank-1	@ Rank-5
GOTS [28]	62.1 ± 1.1	36.3 ± 1.2	38.5 ± 1.6	53.8 ± 1.8
FaceNet [35]	82.3 ± 1.18	66.3 ± 1.3	70.4 ± 1.2	78.8 ± 2.3
VGG-CNN [30]	87.2 ± 1.09	74.3 ± 0.9	79.6 ± 1.04	87.8 ± 1.3
FNM [31]	91.2 ± 0.8	80.4 ± 1.8	84.6 ± 0.6	93.7 ± 0.9
PR-REM [3]	92.1 ± 0.8	83.4 ± 1.5	83.1 ± 0.4	92.6 ± 1.1
PF-cpGAN	93.8 ± 0.67	86.1 ± 0.7	88.3 ± 1.2	94.8 ± 0.6

5.3. Evaluation on IJB-A and IJB-C

Here, we focus on unconstrained face recognition on IJB-A dataset to quantify the superiority of our PF-cpGAN for profile to frontal face recognition. Some of the baselines for comparison on IJB-A are DR-GAN [47], FNM [31], PR-REM [3],and FF-GAN [55]. We have also compared them with other methods as listed in [31] and shown in Table 2. As shown in Table 2, we perform better than the state-of-the-art methods for both verification and identification. Specifically, for verification, we improve the genuine accept rate (GAR) by at least 1.4% compared to other methods. For instance, at the false accept rate (FAR) of 0.01, the best previously-used method is PR-REM, with an average GAR of 94.4%. PF-cpGAN improves upon PR-REM and gives an average GAR of 95.8% at the same FAR. We also show improvement on identification. Specifically, the rank-1 recognition rate shows an improvement of around 1.6% in comparison to the best state-of-the-art method, FNM [31].

We have also plotted receiver operating characteristic (ROC) curve and compared with the baselines given above. The ROC curves for the IJB-A dataset are given in Fig. 3(a). As we can clearly see from the curves, the proposed PF-cpGAN method improves upon other methods and gives much better performance, even at a FAR of 0.001.

We have also performed the task of verification and identification using the IJB-C dataset according to the verification and the identification protocol given in the dataset. The



(a) IJBA (b) CMU Multi-PIE (c) Ablation Study

Figure 3: ROC curve comparison against the baselines for different datasets is shown in (a) and (b). In (c), we show the ROC curves showing the importance of different loss functions for ablation study.

Figure 4: Reconstruction of frontal images at the output of the frontal U-Net generator with profile images as input to the profile U-Net generator. Every odd number column represent the input profile image and every even number column represents the output frontal image. The input images belong to the CMU-MultiPIE dataset.

results are provided in Table 3, showing that PF-cpGAN improve on the existing state-of-the-art methods for both verification and identification. For instance, at the false accept rate (FAR) of 0.01, the best previously-used method is PR-REM, with an average GAR of 92.1%. PF-cpGAN improves upon PR-REM and gives an average GAR of 93.8% at the same FAR. We also observe that, for identification, specifically, rank-1 recognition, shows an improvement over the previous best state-of-the-art method FNM [31] by about 1.1%.

5.4. A Further Analysis on Influences of Face Yaw

In addition to complete profile to frontal face recognition, we also perform a more in-depth analysis on the influence of face yaw angle on the performance of face recognition to better understand the effectiveness of the PF-cpGAN for profile to frontal face recognition. We perform this experiment for the CMU Multi-PIE dataset [13] under setting-1 for fair comparison with other state-of-the-art methods.

Table 4: Rank-1 recognition rates ($\%$) across poses and illuminations under Multi-PIE Setting-1.

Method	$\pm 90°$	$\pm 75°$	$\pm 60°$	$\pm 45°$	$\pm 30°$	$\pm 15°$
HPN [11]	29.82	47.57	61.24	72.77	78.26	84.23
c-CNN [52]	47.26	60.7	74.4	89.0	94.1	97.0
TP-GAN [17]	64.0	84.1	92.9	98.6	99.99	99.8
PIM [57]	75.0	91.2	97.7	98.3	99.4	99.8
CAPG-GAN [16]	77.1	87.4	93.7	98.3	99.4	99.99
FNM+VGG-Face [31]	41.1	67.3	83.6	93.6	97.2	99.0
FNM+Light CNN [31]	55.8	81.3	93.7	98.2	99.5	99.9
PF-cpGAN	88.1	94.2	97.6	98.9	99.9	99.9

As shown in Table 4, we achieve comparable performance with other state-of-the-art methods for different yaw angles. Under extreme pose, PF-cpGAN achieves significant improvements (i.e., approx. 77% to 88% under $\pm 90°$).

For further testing on the Multi-PIE dataset under setting-1, we have also plotted ROC curves and compared with other state-of-the-art methods. The ROC curves for

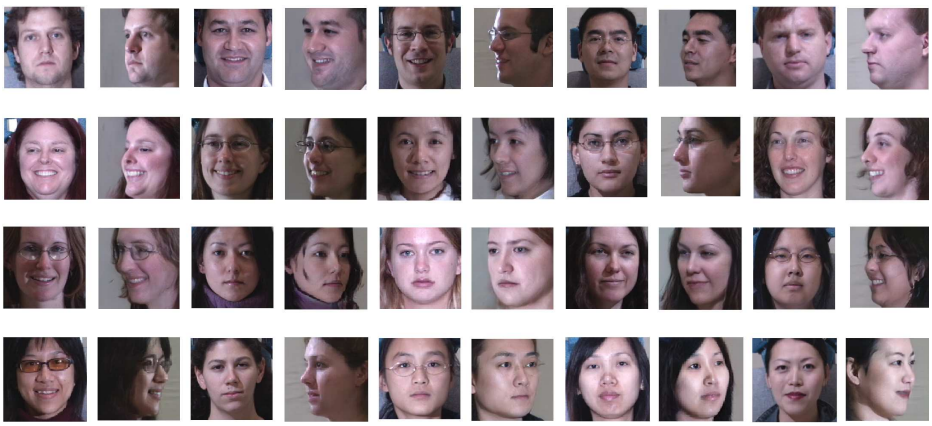


Figure 5: Reconstruction of profile images at the output of the profile U-Net generator with frontal images as input to the frontal U-Net generator. Every odd number column represents the input frontal image, and every even number column represents the output profile image. The input images belong to the CMU-MultiPIE dataset.

Multi-PIE dataset are given in Fig. 3(b). The curves clearly indicate that the proposed method of PF-cpGAN improves upon other methods and gives much better performance, even at FAR of 0.001.

5.5. Reconstruction of frontal and profile images

As noted in Sec. 1, the PF-cpGAN framework can also be used for reconstruction of frontal images by using profile images as input and vice versa. The results of reconstructing frontal images using the profile images as input are given in Fig. 4, and the results of reconstructing profile images using the frontal images as input is given in Fig. 5. The reconstruction procedure for frontal images is given as follows: The profile image is given as input to the profile U-Net generator and the feature vector generated at the bottleneck of the profile generator (i.e., at the output of the encoder of the profile U-Net generator) is passed through the decoder section of the frontal U-Net generator to reconstruct the frontal image. Similarly the reconstruction procedure for profile images is given as follows: The frontal image is given as input to the frontal U-Net generator and the feature vector generated at the bottleneck of the frontal generator (i.e., at the output of the encoder of the frontal U-Net generator) is passed through the decoder section of the profile U-Net generator to reconstruct the profile image. As we can see from Fig. 4 and Fig. 5, the PF-cpGAN can preserve the identity and generate high-fidelity faces from an unconstrained dataset such as CMU-MultiPIE. These results show the robustness and effectiveness of PF-cpGAN for multiple use of profile to frontal matching in the latent common embedding subspace, as well as in the reconstruction of facial images.

5.6. Ablation Study

The objective function defined in (19) contains multiple loss functions: coupling loss (L_{cpl}), perceptual loss (L_P), L_2 reconstruction loss (L_2), and GAN loss (L_{GAN}). It is important to understand the relative importance of different loss functions and the benefit of using them in our proposed method. For this experiment, we use different variations of PF-cpGAN and perform the evaluation using the IJB-A dataset. The variations are: 1) PF-cpGAN with only coupling loss and L_2 reconstruction loss ($L_{cpl} + L_2$); 2) PF-cpGAN with coupling loss, L_2 reconstruction loss, and GAN loss ($L_{cpl} + L_2 + L_{GAN}$); 3) PF-cpGAN with all the loss functions ($L_{cpl} + L_2 + L_{GAN} + L_P$).

We use these three variations of our framework and plot the ROC for profile to frontal face verification using the features from the common embedding subspace. We can see from Fig. 3(c) that the generative adversarial loss helps improve the profile to frontal verification performance, and adding the perceptual loss (blue curve) results in an additional performance improvement. The reason for this improvement is that using perceptual loss along with the contrastive loss leads to a more discriminative embedding subspace leading to a better face recognition performance.

6. Conclusion

We proposed a new framework which uses a coupled GAN for profile to frontal face recognition. The coupled GAN contains two sub-networks which project the profile and frontal images into a common embedding subspace, where the goal of each sub-network is to maximize the pairwise correlation between profile and frontal images during the process of projection. We thoroughly evaluated our model on several standard datasets and the results demonstrate that our model notably outperforms other state-of-the-art algorithms for profile to frontal face verification. Moreover, the improvement achieved by different losses in our proposed algorithm has been studied in an ablation study.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, Dec 2006.

[2] S. Boumi, A. Vela, and J. Chini. Quantifying the relationship between student enrollment patterns and student performance. *arXiv preprint arXiv:2003.10874*, 2020.

[3] K. Cao, Y. Rong, C. Li, X. Tang, and C. C. Loy. Pose-robust face recognition via deep residual equivariant mapping. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5187–5196, 2018.

[4] Q. Cao, Y. Ying, and P. Li. Similarity metric learning for face recognition. In *2013 IEEE International Conference on Computer Vision*, pages 2408–2415, Dec 2013.

[5] J. Chen, V. M. Patel, and R. Chellappa. Unconstrained face verification using deep cnn features. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9, March 2016.

[6] J.-C. Chen, V. M. Patel, and R. Chellappa. Unconstrained face verification using deep cnn features. In *2016 IEEE winter conference on applications of computer vision (WACV)*, pages 1–9. IEEE, 2016.

[7] F. Cole, D. Belanger, D. Krishnan, A. Sarna, I. Mosseri, and W. T. Freeman. Synthesizing normalized faces from facial identity features. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3386–3395, 2017.

[8] A. Dabouei, F. Taherkhani, S. Soleymani, J. Dawson, and N. Nasrabadi. Boosting deep face recognition via disentangling appearance and geometry. In *The IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020.

[9] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893, June 2005.

[10] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

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

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Proc. Neural Information Processing Systems*, pages 2672–2680. 2014.

[13] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. In *2008 8th IEEE International Conference on Automatic Face Gesture Recognition*, pages 1–8, Sep. 2008.

[14] M. A. Hasnat, J. Bohné, J. Milgram, S. Gentric, and L. Chen. von mises-fisher mixture model-based deep learning: Application to face verification. *ArXiv*, abs/1706.04264, 2017.

[15] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2015.

[16] Y. Hu, X. Wu, B. Yu, R. He, and Z. Sun. Pose-guided photorealistic face rotation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8398–8406, 2018.

[17] 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 *Proceedings of the IEEE International Conference on Computer Vision*, pages 2439–2448, 2017.

[18] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.

[19] J. Jiang, R. Hu, Z. Wang, and Z. Cai. CDMMA: Coupled discriminant multi-manifold analysis for matching low-resolution face images. *Signal Processing*, 124:162–172, 2016.

[20] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Proc. European Conference on Computer Vision (ECCV)*, 2016.

[21] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2015.

[22] B. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. E. Allen, P. Grother, A. Mah, M. Burge, and A. K. Jain. Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus benchmark a. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1931–1939, 2015.

[23] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, and A. K. Jain. Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus benchmark a. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1931–1939, 2015.

[24] P. Li, L. Prieto, D. Mery, and P. J. Flynn. On low-resolution face recognition in the wild: Comparisons and new techniques. *IEEE Transactions on Information Forensics and Security*, page 1–1, 2019.

[25] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song. Sphereface: Deep hypersphere embedding for face recognition. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6738–6746, June 2017.

[26] Y. Liu, H. Li, and X. Wang. Rethinking feature discrimination and polymerization for large-scale recognition. *ArXiv*, abs/1710.00870, 2017.

[27] I. Masi, S. Rawls, G. Medioni, and P. Natarajan. Pose-aware face recognition in the wild. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4838–4846, June 2016.

[28] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney, and P. Grother. IARPA Janus benchmark - c: Face dataset and protocol. In *2018 International Conference on Biometrics (ICB)*, pages 158–165, Feb 2018.

[29] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[30] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *BMVC*, 2015.

[31] Y. Qian, W. Deng, and J. Hu. Unsupervised face normalization with extreme pose and expression in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[32] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. Generative adversarial text to image synthesis. In *International Conference on Machine Learning (ICML)*, 2016.

[33] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[34] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proc. IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[35] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832, 2015.

[36] S. Sengupta, J. Chen, C. Castillo, V. M. Patel, R. Chellappa, and D. W. Jacobs. Frontal to profile face verification in the wild. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9, March 2016.

[37] S. Shekhar, V. M. Patel, and R. Chellappa. Synthesis-based robust low resolution face recognition. *arXiv preprint arXiv:1707.02733*, 2017.

[38] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman. Fisher vector faces in the wild. In *BMVC*, 2013.

[39] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[40] Y. Sun, Y. Chen, X. Wang, and X. Tang. Deep learning face representation by joint identification-verification. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, pages 1988–1996, Cambridge, MA, USA, 2014. MIT Press.

[41] F. Taherkhani, H. Kazemi, A. Dabouei, J. Dawson, and N. M. Nasrabadi. A weakly supervised fine label classifier enhanced by coarse supervision. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 6459–6468, 2019.

[42] F. Taherkhani, H. Kazemi, and N. M. Nasrabadi. Matrix completion for graph-based deep semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 5058–5065, 2019.

[43] F. Taherkhani, N. M. Nasrabadi, and J. Dawson. A deep face identification network enhanced by facial attributes prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 553–560, 2018.

[44] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, June 2014.

[45] V. Talreja, T. Ferrett, M. C. Valenti, and A. Ross. Biometrics-as-a-service: A framework to promote innovative biometric recognition in the cloud. In *2018 IEEE ICCE*, pages 1–6. IEEE, 2018.

[46] V. Talreja, M. C. Valenti, and N. M. Nasrabadi. Multibiometric secure system based on deep learning. In *2017 IEEE Global conference on signal and information processing (globalSIP)*, pages 298–302. IEEE, 2017.

[47] L. Tran, X. Yin, and X. Liu. Disentangled representation learning gan for pose-invariant face recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1283–1292, July 2017.

[48] L. Tran, X. Yin, and X. Liu. Disentangled representation learning gan for pose-invariant face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1415–1424, 2017.

[49] F. Wang, X. Xiang, J. Cheng, and A. L. Yuille. Normface: L2 hypersphere embedding for face verification. In *Proceedings of the 25th ACM International Conference on Multimedia*, pages 1041–1049, 2017.

[50] M. Wang and W. Deng. Deep face recognition: A survey. *ArXiv*, abs/1804.06655, 2018.

[51] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. Adams, T. Miller, N. Kalka, A. K. Jain, J. A. Duncan, K. Allen, J. Cheney, and P. Grother. IARPA Janus benchmark-b face dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 592–600, July 2017.

[52] C. Xiong, X. Zhao, D. Tang, K. Jayashree, S. Yan, and T.-K. Kim. Conditional convolutional neural network for modality-aware face recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3667–3675, 2015.

[53] J. Yim, H. Jung, B. Yoo, C. Choi, D.-S. Park, and J. Kim. Rotating your face using multi-task deep neural network. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 676–684, 2015.

[54] X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker. Towards large-pose face frontalization in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3990–3999, 2017.

[55] X. Yin, X. Yu, K. Sohn, X. Liu, and M. K. Chandraker. Towards large-pose face frontalization in the wild. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4010–4019, 2017.

[56] P. Zhang, X. Ben, W. Jiang, R. Yan, and Y. Zhang. Coupled marginal discriminant mappings for low-resolution face recognition. *Optik*, 126(23):4352–4357, 2015.

[57] J. Zhao, Y. Cheng, Y. Xu, L. Xiong, J. Li, F. Zhao, K. Jayashree, S. Pranata, S. Shen, J. Xing, et al. Towards pose invariant face recognition in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2207–2216, 2018.

[58] J. Zhao, Y. Cheng, Y. Xu, L. Xiong, J. Li, F. Zhao, K. Jayashree, S. Pranata, S. Shen, J. Xing, S. Yan, and J. Feng. Towards pose invariant face recognition in the wild. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2207–2216, June 2018.