Received: 14 June 2021 | Revised: 17 January 2022 | Accepted: 17 February 2022

DOI: 10.1049/bme2.12069

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Profile to frontal face recognition in the wild using coupled conditional generative adversarial network

Fariborz Taherkhani | **Veeru Talreja** | **Jeremy Dawson** | **Matthew C. Valenti** | **Nasser M. Nasrabadi**

Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, West Virginia, USA

Correspondence
Fariborz Taherkhani, Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, USA.
Email: ft0009@mix.wvu.edu

Funding information
Center for Identification Technology Research, Grant/Award Number: 1650474

Abstract

In recent years, with the advent of deep-learning, face recognition (FR) has achieved exceptional success. However, many of these deep FR models perform much better in handling frontal faces compared to profile faces. The major reason for poor performance in handling of profile faces is that it is inherently difficult to learn pose-invariant deep representations that are useful for profile FR. In this paper, the authors hypothesise that the profile face domain possesses a latent connection with the frontal face domain in a latent feature subspace. The authors look to exploit this latent connection by projecting the profile faces and frontal faces into a common latent subspace and perform verification or retrieval in the latent domain. A coupled conditional generative adversarial network (cpGAN) structure is leveraged to find the hidden relationship between the profile and frontal images in a latent common embedding subspace. Specifically, the cpGAN framework consists of two conditional 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 maximises the pair-wise correlation between the two feature domains in a common embedding feature subspace. The efficacy of the authors' approach compared with the state of the art is demonstrated using the CFP, CMU Multi-PIE, IARPA Janus Benchmark A, and IARPA Janus Benchmark C datasets. Additionally, the authors have also implemented a coupled convolutional neural network (cpCNN) and an adversarial discriminative domain adaptation network (ADDA) for profile to frontal FR. The authors have evaluated the performance of cpCNN and ADDA and compared it with the proposed cpGAN. Finally, the authors have also evaluated the authors' cpGAN for reconstruction of frontal faces from input profile faces contained in the VGGFace2 dataset.

KEYWORDS
biometric applications, face biometrics, face recognition, feature extraction, image retrieval

1 | INTRODUCTION

Due to the emergence of deep-learning, face recognition (FR) has achieved exceptional success in recent years [1]. However, many of these deep FR models perform relatively poorly in handling profile faces compared to frontal faces [2]. When faces are captured in an unconstrained environment, in the wild, they are often in a profile orientation. Thus, there is an equivalency between the challenging problems of unconstrained FR and profile FR. Pose, expression, and lighting variations are considered to be major obstacles in attaining high unconstrained FR performance. Some methods [1, 3] attempt to address the pose-variation issue by learning pose-invariant features, while some other methods [4–8] try to

Fariborz Taherkhani and Veeru Talreja contributed equally.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *IET Biometrics* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

normalise images while preserving the identity to a single frontal pose before recognition. However, there are three major difficulties related to face frontalisation or normalisation in unconstrained environments:

- Complicated face variations besides pose: In comparison to a controlled environment, there are more complex face variations, for example, lighting, head pose, expression, in real-world scenarios. It is a difficult task to directly warp the input face to a normalised view [7].
- Unpaired data: Undoubtedly, obtaining a strictly normalised face is expensive and time consuming, but getting an effective pair of images consisting of a target normalised face (i.e. frontal-facing and neutral expression) and an input face is difficult due to highly imbalanced datasets [7].
- Presence of artefacts: Synthesised 'frontal' faces contain artefacts caused by occlusions and non-rigid expressions.

In this paper, we hypothesise that the profile face domain shares a latent 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 generative adversarial network (GAN) to find the hidden relationship between the profile face features and frontal face features in a latent common embedding subspace. This is motivated by the fact that given an input image of arbitrary pose, we can actually map its feature to the frontal space through a mapping function that adds residual. This observation is closely connected to the notion of feature equivariance [9], which finds the representation of many deep layers depends upon transformations of the input image. Interestingly, such transformations can be learnt by a mapping function from data, and the function can be subsequently applied to manipulate the representation of an input image to achieve the desired transformation. Figure 1 shows the illustration of embedding features of a subject in different poses.

Our work is conceptually related to the embedding category of super-resolution [10–13] in which our approach also performs verification of profile and frontal faces in the latent space but not in the original image space. From our experiments, we observe that transforming profile and frontal face features into a latent embedding subspace could yield higher performance than image-level face frontalisation, which is susceptible to the negative influence of artefacts as a result of image synthesis. To our best knowledge, this study is the first attempt to perform profile-to-frontal face verification in a latent embedding subspace using generative modelling. The proposed framework can potentially be used to improve the performance of traditional FR methods by integrating it as a preprocessing procedure for a face-frontalisation schema. This work is an extension of our previous work [14]. This paper makes the following contributions, where in contributions three to six are the new contributions different from Ref. [14]:

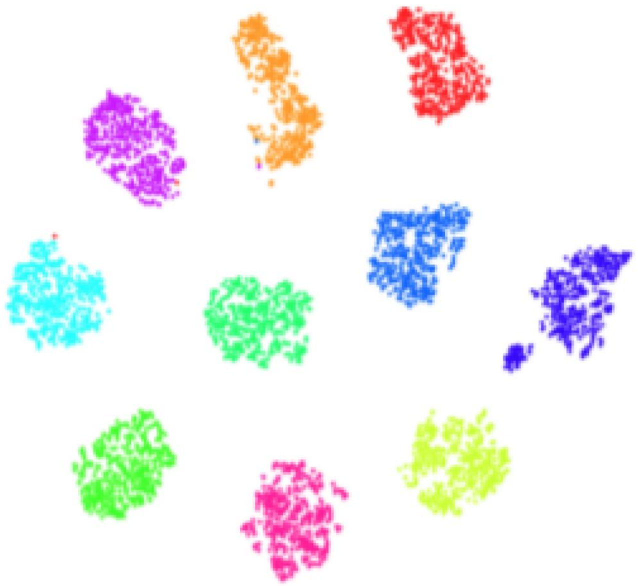


FIGURE 1 Illustration of embedding feature of a subject in different poses.

1. The paper develops a profile to frontal FR model using a coupled conditional generative adversarial network (cpGAN) framework with multiple loss functions.
2. The paper includes comprehensive experiments using different datasets and a comparison of the proposed method with the state-of-the-art methods, indicating the efficacy of the proposed GAN framework.
3. The paper includes experiments to evaluate the frontalisation performance of the cpGAN by using a face matcher (verifier) to compare off-pose faces with a gallery of frontal faces and also compare the frontalised images with the gallery to see if frontalising the face would increase the face matcher performance.
4. The paper implements a coupled CNN (cpCNN) and includes experiments to evaluate the benefits of using the GAN by comparing the performance of a cpCNN with our proposed approach (cpGAN).
5. The paper implements an adversarial discriminative domain adaptation (ADDA) framework for profile to frontal FR and includes experiments to compare the performance of our proposed cpGAN with an ADDA network.
6. The paper includes generated qualitative results for the VGGFace2 dataset to test the robustness and reconstruction ability of our proposed coupled GAN framework.

2 | RELATED WORK

2.1 | Face recognition using deep-learning

Before the advent of deep-learning, traditional methods for FR used one or more layer representations, such as the histogram of the feature codes, filtering responses, or distribution of the dictionary atoms [15]. Face recognition research was concentrated more towards separately improving preprocessing, local descriptors, and feature transformation; however, the 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 the models in Ref. [16, 17] have been used as baseline models for FR. Simultaneously, various loss functions have been explored and used in FR. These loss functions can be categorised 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 [18–23]. For avoiding miss-classification of difficult samples [24, 25], the learnt face features need to be well separated. Angular/cosine-margin-based loss [26–28] is commonly used to make the learnt features more separable with a larger angular/cosine distance. Finally, in the category of softmax loss and its variants for FR [29–31], the softmax loss is modified to improve the FR performance as in Ref. [31], where the cosine distance among data features is optimised along with normalisation of features and weights.

In the SphereFace method [26], angular discriminative features are learnt using CNNs by using an angular softmax (A-Softmax) loss. The notion behind using A-Softmax loss is that, geometrically, it can be viewed as imposing discriminative constraints on a hypersphere manifold. Recently, in order to maximise face class separability, a prominent line of research is to integrate margins in well-established loss functions. For example, in ArcFace approach [27], an Additive Angular Margin Loss is proposed to obtain highly discriminative features for FR. The ArcFace has a clear geometric interpretation due to its exact correspondence to geodesic distance on a hypersphere [27]. In the UniformFace method [32], a new supervised objective function named Uniform loss has been proposed to learn deep equidistributed representations for FR, where the complete feature space on the hypersphere manifold has been exploited by uniformly spreading the class centres on the manifold. A survey of deep-learning methods for FR can be found in Ref. [33].

2.2 | Generative adversarial networks

Introduced by Goodfellow et al. [34], the GAN learns a generator network, G, and a discriminator network, D, with a minimax optimisation procedure. Using this minimax optimisation over a generator and a discriminator provides a simple yet powerful way to map from a source data distribution to a target distribution. GANs have been used for a wide range of applications such as image generation [35–37], 3D object generation [38] etc. In addition to the original GAN, there have been other flavours of GAN [39–41] that have been developed to resolve some of the issues with the original GAN. The Wasserstein GAN [39] proposed the use of Wassertein distance in order to provide a more stable training of GANs. Deep Convolutional GAN [40] was an extension of the original GAN, where the multi-layer perceptron structure is replaced by convolutional structures. Another popular extension of GAN is the Conditional GAN, which was introduced by Mirza and Osindero in Ref. [41]. 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 [41] or text [42]. The most recent GAN models achieve better synthesis by utilising these conditional settings and introducing latent factors to disentangle the objective space. For instance, Info-GAN [43] employs the latent code for information loss to regularise the generative network. There have also been many instance of GAN usage for face frontalisation or generating pose-invariant features. Yin et al. [44] integrated 3D Morphable Model (3DMM) into the GAN structure to propose 3DMM conditioned Face Frontalisation GAN, termed as FF-GAN. Tran et al. [8] combined face frontalisation and learning a pose-invariant representation from a non-frontal face image and integrated it with a GAN structure to propose a Disentangled Representation Learning-Generative Adversarial Network (DR-GAN).

2.3 | 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 [1, 3], elaborate dense 3D facial landmark detection and warping [45], and synthesising a frontal, neutral expression face from a single image [4–8].

2.3.1 | Pose-invariant feature representation

Face frontalisation may be considered as an image-level pose-invariant representation. However, feature-level pose invariant representations have also been a mainstay for FR. Canonical Correlation Analysis was used in earlier studies to analyse the commonality among pose-variant samples. Recently, with the advent of deep-learning, deep-learning-based methods have become popular for pose-invariant feature representation. Cao et al. [1] exploit the inherent mapping between profile and frontal faces and transform a deep profile face representation to a canonical pose by adaptively adding residuals. Additionally, deep-learning methods consider several aspects, such as multi-view perception layers [46], to learn a model separating identity from viewpoints. In Ref. [46], given a single 2D face image, a deep neural net, named Multi-View Perceptron (MVP) can untangle the identity and view features, and infer a full spectrum of multi-view images. Multi-View Perceptron can also predict images under viewpoints that are unobserved in the training data. To allow a single network structure for multiple pose inputs, feature pooling across different poses is proposed in Ref. [47]. There have also been methods related to pose-invariant feature disentanglement [48] or identity preservation [49, 50] that aim to factorise out the non-identity part with a meticulously designed network. In Ref. [50], a new learning-based face representation, the Face Identity-Preserving (FIP) features, has been proposed. The FIP features are learnt by using a deep neural network that combines the feature extraction layers and the reconstruction layer. The former layer generates FIP features from a face image, while the latter layer transforms the FIP features into an image in the canonical view.

2.3.2 | Face frontalisation

Using a single image with large pose variation, it is very challenging to synthesise face with a frontal view with a neutral expression face due to two major reasons: (a) recovering the 3D information from 2D projections is obscure and uncertain and (b) presence of self-occlusion. Seminal works date back to the 3D Morphable Model (3DMM) [51], which models both the shape and appearance as PCA spaces. Hassner et al. [52] adopt a 3D shape model combined with input images to register and produce the frontalised face. Based on 3DMM, Zhu et al. [53] provide a high-fidelity pose and expression normalisation method. However, 3D-based methods often do not provide reasonable results and suffer from a significant performance drop with large pose variations due to artefacts and severe texture losses. Some deep-learning-based methods have shown promising performance in terms of face frontalisation [5–8, 54–56]. In Ref. [55], a recurrent transform unit is proposed to incrementally rotate faces in fixed yaw angles and synthesise discrete 3D views. FF-GAN [5] solves the problem of large-pose face frontalisation in the wild by incorporating a 3D face model into a GAN. Considering photo realistic and identity-preserving frontal view synthesis, a domain adaptation strategy for pose invariant FR is discussed in Ref. [56–60]. Tran et al. [8] propose a GAN framework to rotate a face and disentangle the identity representation by using a given pose code. In Ref. [7], a face normalisation model (FAN) uses a GAN network with three distinct losses for generating canonical-view and expression-free frontal images.

3 | GENERATIVE ADVERSARIAL NETWORK

Generative adversarial network was first introduced by Goodfellow, et al. [34] 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 a 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 [7]. The loss function, $L(D, G)$, for GAN is given as

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

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

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

In conditional GAN [41], both the generator and discriminator are conditioned on an additional variable, x. The loss function for the conditional GAN is given as follows:

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

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

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

4 | PROPOSED METHOD

Here, we describe our method for profile to frontal FR. In contrast to the face normalisation methods, we do not perform pose normalisation (i.e. frontalisation) 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 modelling. Inspired by the success of GANs [34], we explore adversarial networks to project profile and frontal images to a common subspace for recognition.

The framework of proposed profile to frontal cpGAN (PF-cpGAN, shown in Figure 2) consists of two modules, where each module contains a GAN architecture comprised 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 the perceptual loss [61] based on the VGG 16 architecture, as well as an L_2 reconstruction error. The perceptual loss helps to generate a 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,

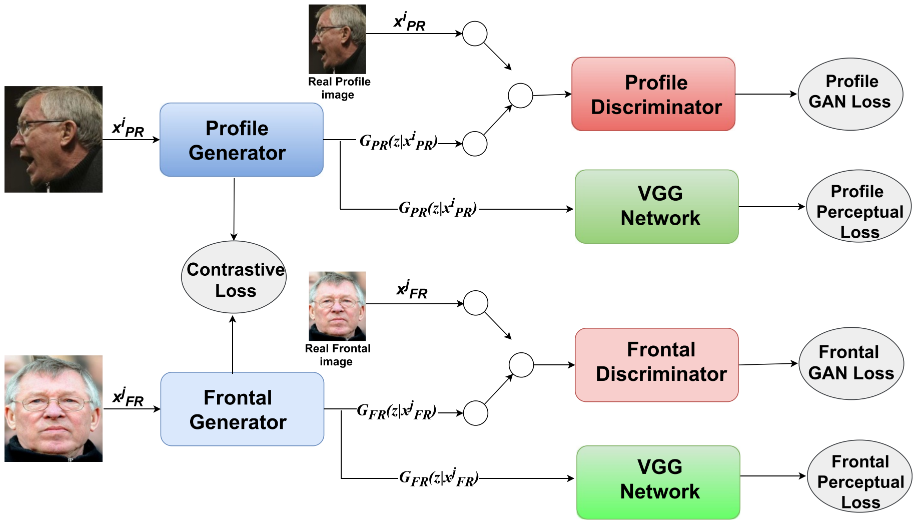


FIGURE 2 Block diagram of profile to frontal coupled conditional generative adversarial network (PF-cpGAN).

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 [62] auto-encoder architecture (shown in Figure 3a). The primary reason for using U-Net is that the encoder–decoder structure tends to extract global features and generate images by leveraging this overall information, which is very useful for global shape transformation 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, the use of 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 [63](shown in Figure 3b), 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 penalises structure at the scale of the patches ensures the preservation of high-frequency details, which are usually eliminated 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:

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

where x^i_{PR} and x^j_{FR} denote the *i-th* profile and *j-th* 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 a 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 $z_1\left(x^i_{PR}\right)$ and $z_2\left(x^j_{FR}\right)$, which is given by,

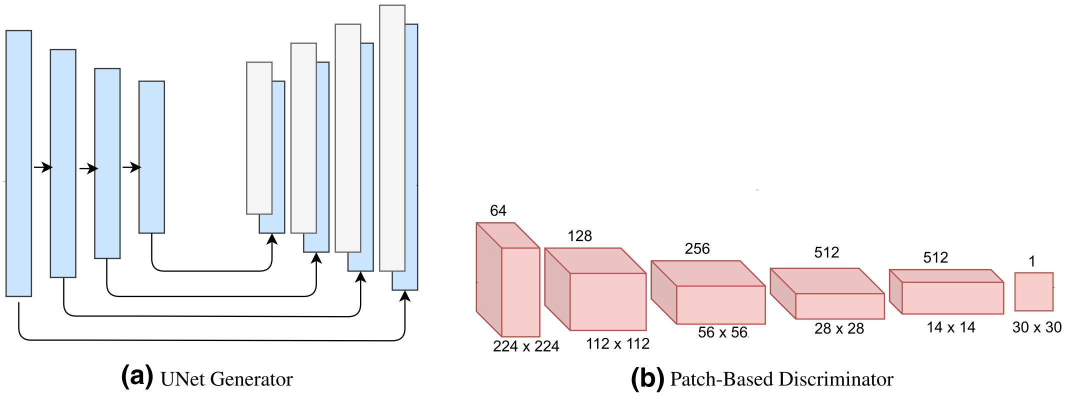
$$D_z = \left\| z_1\left(x^i_{PR}\right) - z_2\left(x^j_{FR}\right) \right\|_2. \tag{6}$$

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

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

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

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



(a) UNet Generator

(b) Patch-Based Discriminator

FIGURE 3 Generative adversarial network (GAN) architectures

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

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

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 a profile image against a frontal image.

4.2 | Generative adversarial loss

Let the profile and frontal generators 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}, respectively. 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 [41] 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

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

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

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,

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

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 be defined as

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

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

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

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

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

The total L_2 reconstruction loss function is given by,

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

4.4 | Perceptual loss

In addition to the GAN loss and the reconstruction loss that are used to guide the generators, we have also used the perceptual loss, which was introduced in Ref. [61] for style transfer and super-resolution. The perceptual loss function is used to compare high-level differences, such as content and style discrepancies, between images. The perceptual loss function involves comparing two images based on high-level representations from a pre-trained CNN, such as VGG-16 [64]. 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 [61].

In our proposed approach, perceptual loss is added to both the profile and the frontal module using a pre-trained VGG-16 network [64]. We extract the high-level features (ReLU3-3 layer) of the 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 the profile network is defined as

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

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

Likewise, the perceptual loss for the frontal network is

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

The total perceptual loss function is given by

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

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 given by (9), L_{GAN} is the total generative adversarial loss given by (12), L_P is the total perceptual loss given by (18), and L_2 is the total reconstruction error given by (15). 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 the state-of-the-art face verification methods across pose-variation. We also explore the effect of face yaw in our algorithm. Additionally, we have implemented a cpCNN and an ADDA for profile to frontal FR. We have evaluated the performance of cpCNN and ADDA and compared it with that of the proposed PF-cpGAN. We have also evaluated our PF-cpGAN for reconstruction of frontal images from input profile images. Finally, we conduct an ablation study to investigate the effect of each term in our total training loss function in (20).

5.1 | Experimental details

The CMU Multi-PIE database [65] 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. There are four sessions in this database. For fair comparison, the database setting was made consistent with CAPG-GAN [66], where 250 subjects from Multi-PIE have been used. Consistent with CAPG-GAN, face images with neutral expression under 20 illuminations and 13 poses within ±90° are used. We follow the setting-1 testing protocol provided in CAPG-GAN.

In setting-1, only images from session 1, which contains faces of 250 subjects, were used. First, 150 identities were used in the training set and remaining 100 identities were used for testing. The training set consists of all the images (13 poses and 20 illumination levels) of 150 identities, that is, 150 × 13 × 20 = 39,000 images in total. For testing, one gallery image with frontal view and normal illumination is used for each of the remaining 100 subjects. The numbers of the probe and gallery sets are 24,000 and 100, respectively.

The IARPA Janus Benchmark A (IJB-A) [67] is a challenging dataset collected under complete unconstrained conditions covering full pose variation (yaw angles −90° to +90°). IARPA Janus Benchmark A contains 500 subjects with 5712 images and 20,414 frames extracted from videos. Following the standard protocol in Ref. [67], we evaluate our method on both verification and identification. The IARPA Janus Benchmark C (IJB-C) dataset [68] builds on IJB-A and IJB-B [69] datasets and has a total of 31,334 images for a total number of 3531 subjects. We have also evaluated our method on IJB-A and IJB-C datasets.

VGGFace2 is a large-scale FR dataset, where the images are downloaded from Google Image Search and have large variations in pose, age, illumination, and ethnicity. The dataset contains about 3.3 million images corresponding to more than 9000 identities with an average of 364 images per subject.

5.1.1 | Implementation details

We have implemented a U-Net autoencoder with a ResNet-18 [70] architecture pre-trained on ImageNet. We have added an additional fully connected layer after the average pooling layer for the 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 optimiser [71] with first-order momentum of 0.5, and a 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 CFP dataset [2], a challenging dataset created to examine the problem of frontal to profile face verification in the wild. The same 10-fold protocol is applied on both the Frontal-Profile and Frontal-Frontal settings. For fair comparison and as given in Ref. [2], we consider different types of feature extraction techniques such as HoG [72], LBP [73], and Fisher Vector [74] along with metric learning techniques such as Sub-SML [75], and the diagonal metric learning as reported in Ref. [74]. We also compare against deep-learning techniques, including Deep Features [76], and PR-REM [1]. The results are summarised 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 equal error rate (EER) for the profile-frontal setting. The 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, the PF-cpGAN performs on-par with the best deep-learning method of PR-REM, and in-fact, does slightly better than PR-REM by ≈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 semantic similarity in the latent space is better than some other deep-learning techniques such as Deep Features or PR-REM.

5.3 | Evaluation on IJB-A and IJB-C

Here, we focus on unconstrained FR on the IJB-A dataset to quantify the superiority of our PF-cpGAN for profile to frontal FR. Some of the baselines for comparison on IJB-A are DR-GAN [8], FNM [7], PR-REM [1], and FF-GAN [5]. We have also compared them with other methods as listed in Ref. [7] 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 that of 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%. The PF-cpGAN improves upon PR-REM and gives an average GAR of 95.8% at the same FAR. We also show improvement in 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 [7].

We have also plotted the receiver operating characteristic (ROC) curve and compared with the baselines given above. The ROC curves for the IJB-A dataset are given in Figure 4a. 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 that dataset. The results are provided in Table 3, showing that the proposed PF-

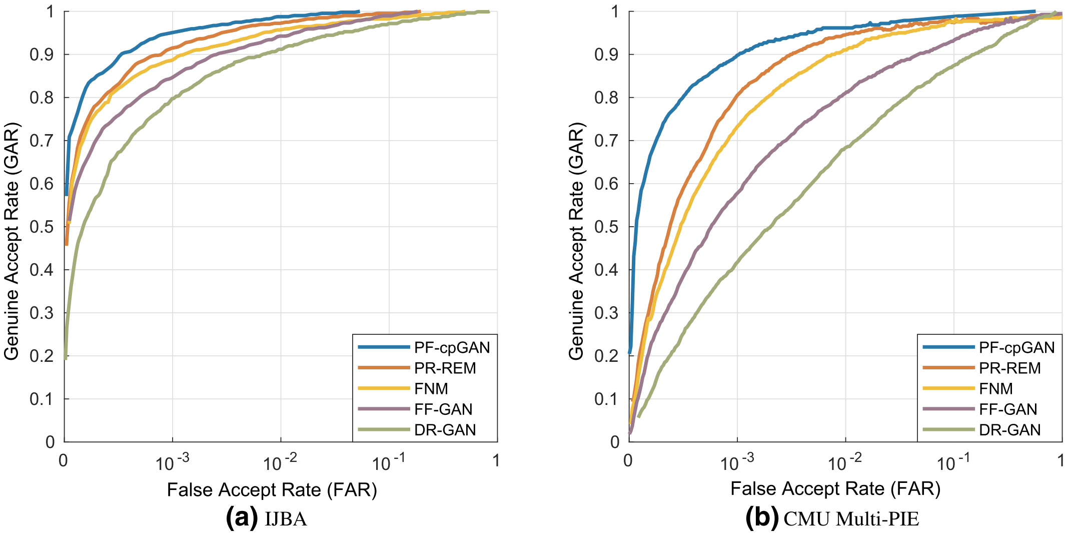
TABLE 1 Performance comparison on the CFP dataset. Mean Accuracy and equal error rate (EER) with standard deviation over 10 folds.

	Frontal-profile		Frontal-frontal	
Algorithm	Accuracy	EER	Accuracy	EER
HoG + Sub-SML [2]	77.31 ± 1.61	22.20 ± 1.18	88.34 ± 1.31	11.45 ± 1.35
LBP + Sub-SML [2]	70.02 ± 2.14	29.60 ± 2.11	83.54 ± 2.40	16.00 ± 1.74
FV + Sub-SML [2]	80.63 ± 2.12	19.28 ± 1.60	91.30 ± 0.85	8.85 ± 0.74
FV + DML [2]	58.47 ± 3.51	38.54 ± 1.59	91.18 ± 1.34	8.62 ± 1.19
Deep features [76]	84.91 ± 1.82	14.97 ± 1.98	96.40 ± 0.69	3.48 ± 0.67
PR-REM [1]	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

TABLE 2 Performance comparison on IARPA Janus Benchmark A (IJB-A) benchmark

	Verification		Identification	
Method	**GAR@ FAR = 0.01**	**GAR@ FAR = 0.001**	**@ Rank-1**	**@ Rank-5**
OPENBR [67]	23.6 ± 0.9	10.4 ± 1.4	24.6 ± 1.1	37.5 ± 0.8
GOTS [67]	40.6 ± 1.4	19.8 ± 0.8	43.3 ± 2.1	59.5 ± 2.0
PAM [3]	73.3 ± 1.8	55.2 ± 3.2	77.1 ± 1.6	88.7 ± 0.9
DCNN [76]	78.7 ± 4.3	–	85.2 ± 1.8	93.7 ± 1.0
DR-GAN [77]	77.4 ± 2.7	53.9 ± 4.3	85.5 ± 1.5	94.7 ± 1.1
FF-GAN [44]	85.2 ± 1.0	66.3 ± 3.3	90.2 ± 0.6	95.4 ± 0.5
FNM [7]	93.4 ± 0.9	83.8 ± 2.6	96.0 ± 0.5	98.6 ± 0.3
PR-REM [1]	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

Note: 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.



(a) IJBA **(b)** CMU Multi-PIE

FIGURE 4 Receiver operating characteristic (ROC) curve comparison against the baselines for different datasets is shown in (a) and (b).

cpGAN improves on the existing state-of-the-art methods for both verification and identification. For instance, at the FAR of 0.01, the best previously used method is PR-REM, with an average GAR of 92.1%. The 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, the PF-cpGAN shows an improvement over the previous best state-of-the-art method FNM [7] by about 1.1%.

5.4 | A further analysis on influences of face yaw

In addition to complete profile to frontal FR, we also perform a more in-depth analysis on the influence of face yaw angle on the performance of FR to better understand the effectiveness of the PF-cpGAN for profile to frontal FR. We perform this experiment for the CMU Multi-PIE dataset [65] under setting-1 for fair comparison with other state-of-the-art methods. As shown in Table 4, we achieve comparable performance with other state-of-the-art methods for different yaw angles. Under extreme pose, the PF-cpGAN achieves significant improvements (i.e., approx. 77%–88% under ±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 Multi-PIE dataset are given in Figure 4b. 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 Section 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

TABLE 3 Performance comparison on IARPA Janus Benchmark C (IJB-C) benchmark

	Verification		Identification	
Method	GAR@ FAR = 0.01	GAR@ FAR = 0.001	@ Rank-1	@ Rank-5
GOTS [68]	62.1 ± 1.1	36.3 ± 1.2	38.5 ± 1.6	53.8 ± 1.8
FaceNet [19]	82.3 ± 1.18	66.3 ± 1.3	70.4 ± 1.2	78.8 ± 2.3
VGG-CNN [78]	87.2 ± 1.09	74.3 ± 0.9	79.6 ± 1.04	87.8 ± 1.3
FNM [7]	91.2 ± 0.8	80.4 ± 1.8	84.6 ± 0.6	93.7 ± 0.9
PR-REM [1]	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

Note: 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.

frontal images using the profile images as input are given in Figure 5, and the results of reconstructing profile images using the frontal images as input is given in Figure 6. 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 Figures 5 and 6, 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.

TABLE 4 Rank-1 recognition rates (%) across poses and illuminations under Multi-PIE Setting-1.

Method	±90°	±75°	±60°	±45°	±30°	±15°
HPN [79]	29.82	47.57	61.24	72.77	78.26	84.23
c-CNN [80]	47.26	60.7	74.4	89.0	94.1	97.0
TP-GAN [81]	64.0	84.1	92.9	98.6	99.99	99.8
PIM [56]	75.0	91.2	97.7	98.3	99.4	99.8
CAPG-GAN [66]	77.1	87.4	93.7	98.3	99.4	99.99
FNM + VGG-Face [7]	41.1	67.3	83.6	93.6	97.2	99.0
FNM + Light CNN [7]	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

5.6 | Evaluation of the frontalisation by cpGAN as a preprocessing for face matching

As mentioned earlier, our coupled GAN framework can also be used for frontalisation, which can be an important preprocessing step for other face-recognition tasks. Here, we conducted experiments to indicate the effectiveness of the frontalisation performed using our cpGAN for the face verification task. In this set of experiments, we have used an Inception [82]-based FaceNet [19] model for the face verification task, which is specifically the NN2 model from Ref. [19]. We have performed this set of experiments on the VGGFace2 dataset [83].

The VGGFace2 dataset provides annotation to enable evaluation of face matching across different poses [83]. In the dataset, six pose templates corresponding to three poses (i.e. two templates for a single pose) have been provided for about 300 identities. A template corresponds to five faces from the same subject with a consistent pose. This pose can be frontal, three-quarter or profile view. Consequently, for the 300 identities, there are a total of 1.8K templates with 9K images in total [83]. For this set of experiments, we have used only the profile and the frontal templates, which correspond to about 6K images corresponding to 300 identities.

Here, we perform face verification using FaceNet in three different settings. In the first setting, we choose about 2.5K frontal images corresponding to 250 identities. Using these images, we fine-tune the Inception model NN2 from FaceNet for frontal to frontal face verification. Next, using this FaceNet model, we evaluate the frontal to frontal face verification on the remaining 50 identities. This setting will be called *Original Frontal to Frontal*. In the second setting, we choose about 5K images corresponding to 250 identities, which have both profile and frontal images. Using these images, we fine-tune the Inception model NN2 from FaceNet for profile to frontal face verification. Next, using this FaceNet model, we evaluate the profile to frontal face verification on the remaining 50 identities. This setting will be called *Profile to Frontal*. In the third setting, we used our cpGAN to frontalise the profile images from the datasets used in the second setting (300 subjects with about 3K profile images) using the method outlined in Section 5.5. We call this frontalised dataset synthesised frontal dataset. Next, using the fine-tuned FaceNet model from the first setting, we evaluate the frontal to frontal face verification

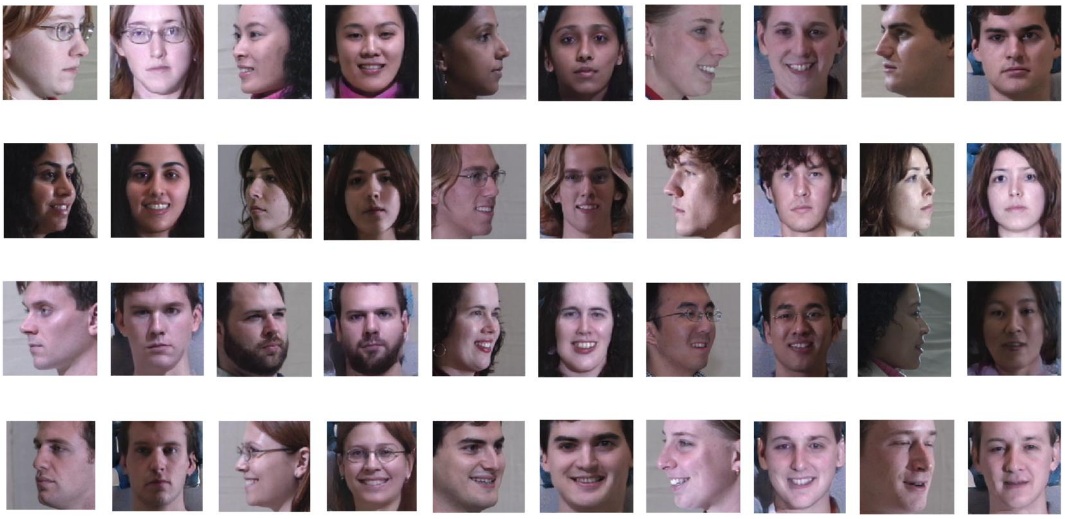


FIGURE 5 Reconstruction of frontal images at the output of the frontal U-Net generator with profile images as input to the profile U-Net generator. Every odd number column represent the input profile image and every even number column represents the output frontal image. The input images belong to the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

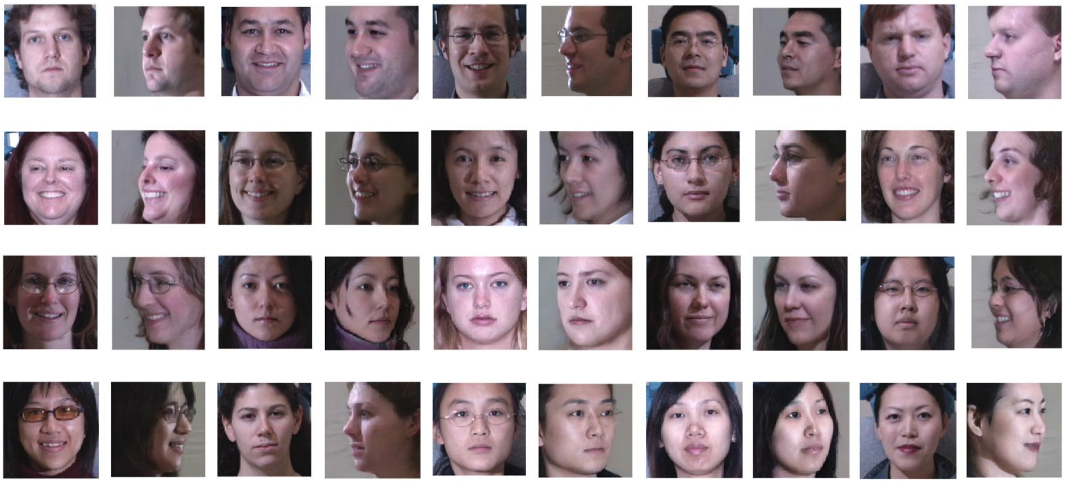


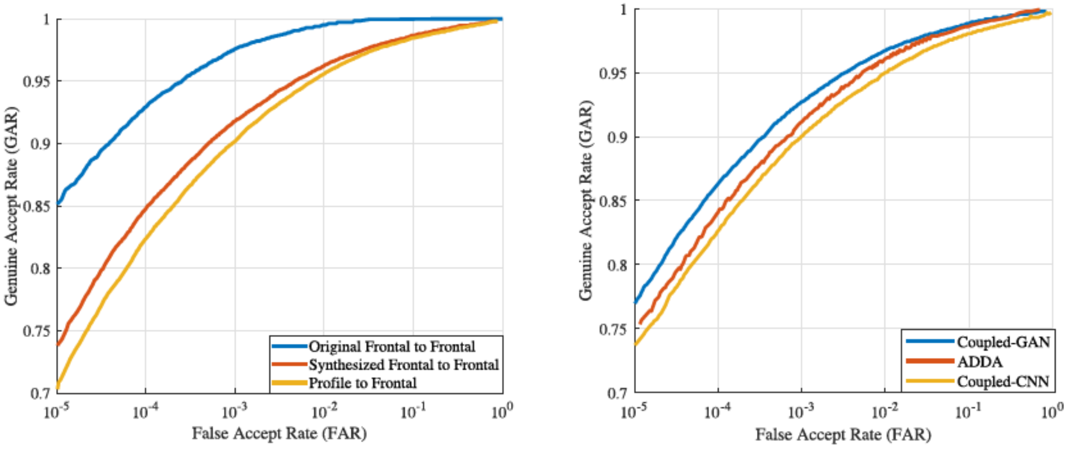
FIGURE 6 Reconstruction of profile images at the output of the profile U-Net generator with frontal images as input to the frontal U-Net generator. Every odd number column represents the input frontal image, and every even number column represents the output profile image. The input images belong to the CMU-MultiPIE dataset.

on 50 identities from the synthesised frontal dataset. Specifically, in the third setting, we are trying to check how well the proposed cpGAN is able to frontalise the images by running the frontal to frontal face verification model on the synthesised frontal dataset. This setting will be called *Synthesised Frontal to Frontal*. Note that we try to keep the 50 identities used for evaluation consistent across all the three settings.

Using the ROC curve as our performance metric, we have compared the performance of these three settings to evaluate the effectiveness of frontalisation performed using our proposed cpGAN. The performance curves are provided in Figure 7a. As expected the first setting (Original Frontal to Frontal) gives us the best performance, and it is the upper bound as we are using the original frontal dataset for training and evaluation in this setting. On comparing the curves for the second (Profile to Frontal) and third settings (Synthesised Frontal to Frontal), it can be observed that the Synthesised Frontal to Frontal outperforms the Profile to Frontal FR model. This shows that the preprocessing in the form of frontalisation performed using the proposed cpGAN framework improves the performance of a FaceNet model for profile to frontal face verification.

5.7 | Implementation of couplesd CNN and domain adaptation network for profile to frontal face matching

Before the advent of GAN, many deep-learning applications used CNNs for classification, regression, or reconstruction. To



(a) Frontalization Performance

(b) Different network comparisons

FIGURE 7 Performance comparison for (a) frontalisation using coupled generative adversarial network (cpGAN) as a preprocessing and (b) PF-cpGAN (Coupled-GAN) versus coupled convolutional neural network (Coupled-CNN) versus profile to frontal adversarial domain adaptation (PF-ADDA).

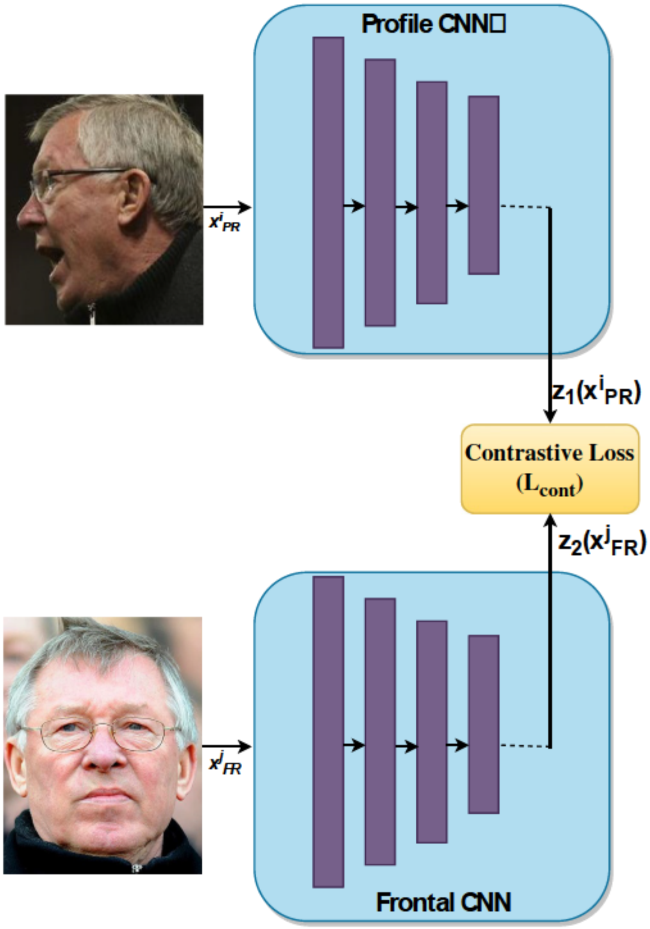


FIGURE 8 Block diagram of Coupled convolutional neural network (CNN)

showcase the advantage of using a GAN model in our proposed approach for profile to frontal FR, we have also implemented two other frameworks that will be explained in this section. The performance comparison of the proposed PF-cpGAN with these new frameworks will be discussed in the following section.

5.7.1 | Coupled CNN

In the literature, it has been shown that GAN is better than CNN for some deep-learning applications. To confirm this hypothesis for our proposed application, we have implemented a coupled CNN and compared its performance with our proposed coupled GAN architecture. The coupled CNN (cpCNN) architecture is shown in Figure 8. For fair comparison, we have used ResNet18 [70] pre-trained on ImageNet network as our CNN architecture for both Frontal CNN and Profile CNN. Additionally, we have added an extra fully connected layer after the average pooling layer of ResNet18 for our coupled CNNs.

The frontal and profile CNNs are coupled together at their output layer using a contrastive loss function ($\mathcal{L}_{cont}$). This loss function ($\mathcal{L}_{cont}$) is a distance-based loss function, which is similar to the contrastive loss function (5) that we have used for PF-cpGAN. For ease of understanding, we have used the same naming convention for cpCNN as in PF-cpGAN.

We have used the VGGFace2 dataset for training and testing of the cpGAN. As in Section 5.6, we choose about 5K images corresponding to 250 identities, which have both profile and frontal images for fine-tuning the cpCNN for profile to frontal face verification. We have tested the cpCNN on the 50 disjoint identities from VGGFace2. The performance comparison is discussed in the following section.

5.7.2 | Domain adaptation network

A profile to frontal recognition network could very well be implemented using deep-learning-based domain adaptation techniques. These domain adaptation techniques attempt to alleviate the negative effects of domain shift (frontal domain to profile domain in our case) by learning deep neural transformations that map both domains into a common feature space. Recently, adversarial adaptation methods, which are based on reconstructing the target domain from the source representation have become increasingly popular. These adversarial methods seek to reduce an approximate domain discrepancy distance through an adversarial objective function with respect to a domain discriminator [84].

Taking a cue from Ref. [84], we have implemented an unsupervised discriminative domain adaption network for profile to frontal FR. Hereafter, this network will be known as profile to frontal adversarial domain adaptation (PF-ADDA). For this adversarial domain adaptation network, we consider the source domain as the frontal images and the target domain as the profile images. The architecture of PF-ADDA is shown in Figure 9. Profile to Frontal Adversarial Domain Adaptation has been implemented and optimised in two steps:

- In the first step of pre-training a frontal convolutional neural network (CNN), a discriminative representation is learnt using the labels in the frontal image domain (source domain). This implies we first pre-train a frontal image encoder CNN using labelled frontal image examples. The optimisation for this step is given as

$$\min_{z_2,C} L_{cls}(X_{FR}, Y_{FR}) = -E_{(x_{FR},y_{FR})\sim(X_{FR},Y_{FR})}\sum_{k=1}^{K}\mathbb{1}_{[k=y_{FR}]}\log C\left(z_2\left(x^j_{FR}\right)\right), \tag{20}$$

where the classification loss L_{cls} is optimised over z_2 and frontal image classifier, C, by training using the labelled source data, X_{FR}, and Y_{FR}.

- In the second step of adversarial adaptation, a separate encoding that maps the profile image data to the same space as the frontal image domain using an asymmetric mapping is learnt through a combination of domain-adversarial loss and the contrastive loss. In other words, this implies that we perform adversarial adaptation by learning a profile image encoder CNN such that a discriminator that sees encoded frontal and profile images cannot accurately predict their domain label. In addition to the discriminator loss, the frontal and profile domain CNNs are also coupled through a contrastive loss. The optimisations for this step are given as

$$\min_{D} L_{adv_D}(X_{FR}, X_{PR}, z_2, z_1) = -E_{x_{FR}\sim X_{FR}}\left[\log D\left(z_2\left(x^j_{FR}\right)\right)\right] - E_{x_{PR}\sim X_{PR}}\left[\log\left(1 - D\left(z_1\left(x^i_{PR}\right)\right)\right)\right], \tag{21}$$

$$\min_{z_1} L_{adv_G}(X_{FR}, X_{PR}, D) = -E_{x_{PR}\sim X_{PR}}\left[\log D\left(z_1\left(x^i_{PR}\right)\right)\right], \tag{22}$$

and

$$L_{cont}\left(z_1\left(x^i_{PR}\right), z_2\left(x^j_{FR}\right), Y\right) = (1-Y)\frac{1}{2}(D_z)^2 + (Y)\frac{1}{2}(\max(0, m - D_z))^2. \tag{23}$$

As shown in Equations (21) and (22), the frontal image encoder CNN (source CNN) is fixed during the second stage, we just need to optimise the discriminator loss L_{adv_D} and profile encoder loss L_{adv_G} over the profile encoder CNN to generate z_1 without revisiting the source domain encoder. Finally, along with the adversarial losses, we also optimise the contrastive loss, L_{cont}, between the output of the Frontal CNN and Profile CNN as shown in (23). This contrastive loss is similar to the loss used for cpCNN and PF-cpGAN.

During testing, profile images (target domain) are mapped with the profile image encoder to the shared feature space, and frontal images (source images) are mapped with the frontal image encoder to the shared feature space. Finally, the profile to frontal matching is performed in the shared feature space. Dashed lines in Figure 9 indicate fixed network parameters.

We have used the VGGFace2 dataset for training and testing of the PF-ADDA. For fair comparison, the train and test split of the dataset for the PF-ADDA is consistent with the split for cpCNN.

5.8 | Performance comparison of PF-cpGAN versus cpCNN versus PF-ADDA

We have performed several experiments to compare the performance of our proposed PF-cpGAN approach with cpCNN and PF-ADDA. These experiments are performed on the VGGFace2 dataset [83]. As already mentioned in the previous Section 5.7, for implementing cpCNN and PF-ADDA, we have used a common network architecture as PF-cpGAN.

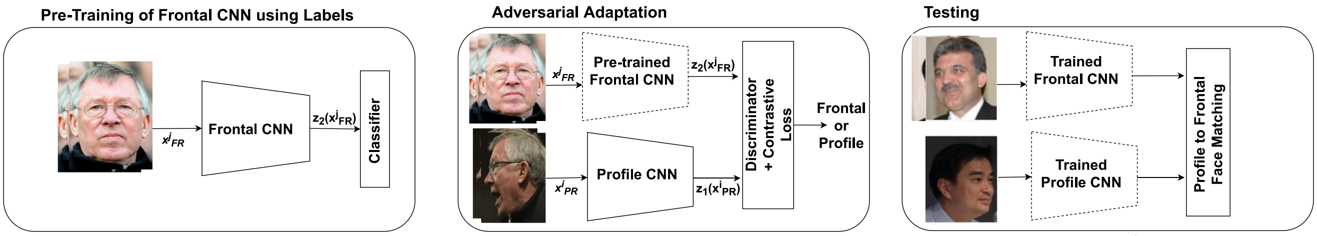


FIGURE 9 Block diagram of profile to frontal adversarial domain adaptation (PF-ADDA)

Furthermore, we have been consistent in the training procedure (optimiser, batch-size, learning rate decay schedule etc.). The performance comparison is plotted in terms of ROC and shown in Figure 7b. The ROC result curves show that the proposed PF-cpGAN method outperforms other methods and gives much better performance for face verification under different pose variations. This demonstrates the effectiveness of coupled-GAN compared to other implementations. The improvement in performance using PF-cpGAN could be attributed to individual discriminators in the PF-cpGAN, which generate more domain specific features. The improvement can also be attributed to the sharpening of the features due to the perceptual loss terms.

5.9 | Coupled-GAN qualitative results on VGGFace2

In this section, we test the robustness of our proposed approach Pf-cpGAN on the VGGFace2 dataset by reconstructing frontal images from input profile images. In VGGFace2 [83], two networks are trained to estimate the pose of images in the dataset. Specifically, a 5-way classification ResNet-50 is trained on the large-scale CASIA-WebFace dataset [85] to estimate head pose (roll, pitch, and yaw). This model is then leveraged to predict pose of all the images in the dataset. As a result, VGGFace2 published different pose templates for 368 identities. Specifically, there are six templates for each subject: two templates each for frontal view, three-quarter view and profile view. There are five images per template. Here, we used about 250 identities to construct our pairs for training our coupled-GAN framework. Next, we test our network for frontalisation of profile images. We follow the same procedure as discussed in Section 5.5 to frontalise our images. The results for frontalised images are shown in Figure 10. From these images, it can be observed that the PF-cpGAN can preserve the identity and generate high-fidelity faces from the VGGFace2 dataset. These results demonstrate the robustness and effectiveness of our coupled-GAN framework for frontalising pose-variant images in the latent common embedding subspace.

5.10 | Ablation study

The objective function defined in (20) contains multiple loss functions: coupling loss (L_{cpl}), perceptual loss (L_P) and L_2

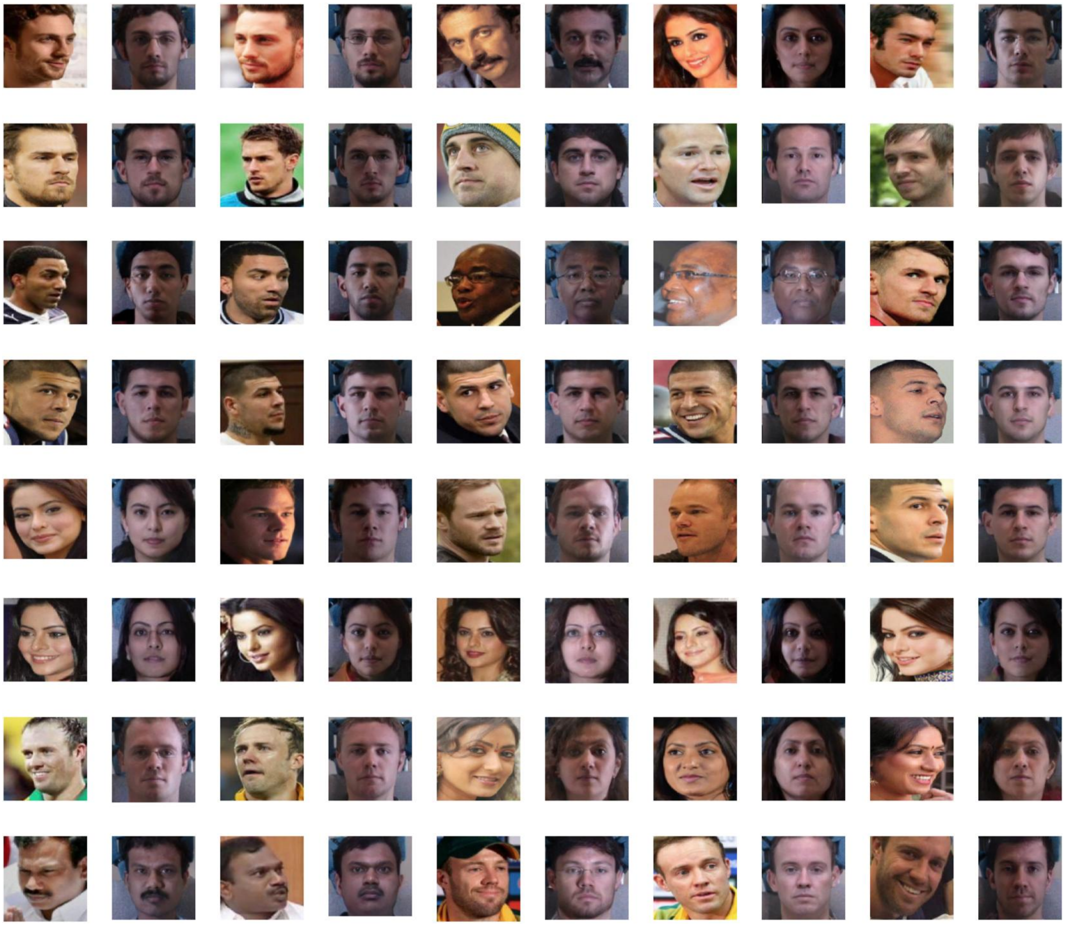


FIGURE 10 Reconstruction of frontal images at the output of the frontal U-Net generator with profile images as input to the profile U-Net generator. Every odd number column represents the input profile image, and every even number column represents the output frontal image. The input images belong to the VGGFace2 dataset.

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 as follows: (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 Figure 11 that the generative adversarial loss helps to 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 resulting in better FR performance.

6 | CONCLUSION

We proposed a new framework, which uses a coupled GAN for profile to frontal FR. The coupled GAN contains two subnetworks, which project the profile and frontal images into a common embedding subspace, where the goal of each subnetwork is to maximise the pair-wise 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. For instance, under the extreme pose of ±90°, the PF-cpGAN achieves improvements of approx. 11% (i.e. 77%–88%), when compared to the state-of-the-art methods for CMU-MultiPIE dataset. We have also explored two other similar implementations in the form of coupled CNN (cpCNN) and domain adaptation network (ADDA) for profile to frontal FR. We have compared the performance of the proposed approach with cpCNN and ADDA and shown that the proposed approach performs much better than these two implementations. Moreover, we have also evaluated the frontal image reconstruction performance of the proposed approach. Finally, the improvement achieved by different losses including perceptual and GAN losses in our proposed algorithm has been investigated in an ablation study.

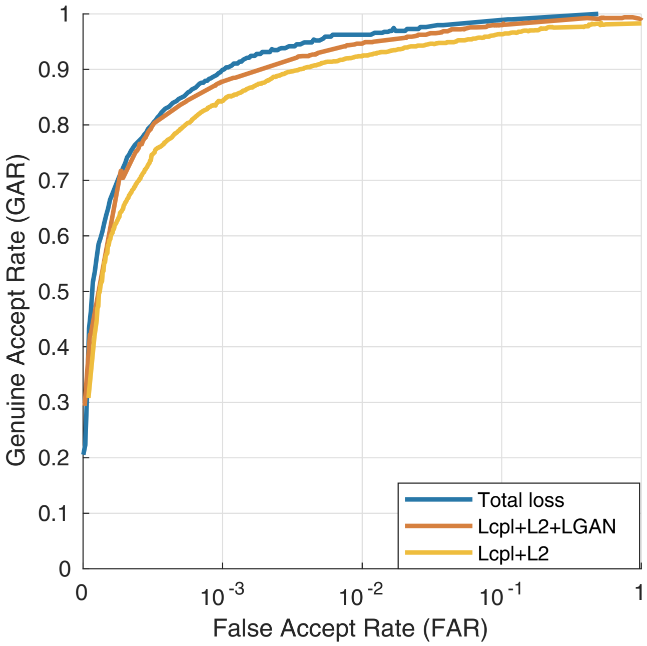


FIGURE 11 Receiver operating characteristic (ROC) curves showing the importance of different loss functions for ablation study.

ACKNOWLEDGEMENT

Center for Identification Technology Research, grant number: 1650474.

CONFLICT OF INTEREST

No conflict of interest.

PERMISSION TO REPRODUCE MATERIALS FROM OTHER SOURCES

None.

DATA AVAILABILITY STATEMENT

Data openly available in a public repository that does not issue DOIs.

ORCID

Fariborz Taherkhani https://orcid.org/0000-0001-7966-734X

REFERENCES

1. Cao, K., et al.: Pose-robust face recognition via deep residual equivariant mapping. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5187–5196 (2018)
2. Sengupta, S., et al.: Frontal to profile face verification in the wild. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1–9 (2016)
3. Masi, I., et al.: Pose-aware face recognition in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4838–4846 (2016)
4. Yim, J., et al.: Rotating your face using multi-task deep neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 676–684 (2015)
5. Yin, X., et al.: Towards large-pose face frontalization in the wild. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 4010–4019 (2017)
6. Cole, F., et al.: Synthesizing normalized faces from facial identity features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3386–3395 (2017)
7. Qian, Y., Deng, W., Hu, J.: Unsupervised face normalization with extreme pose and expression in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
8. Tran, L., Yin, X., Liu, X.: Disentangled representation learning GAN for pose-invariant face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1283–1292 (2017)

9. Lenc, K., Vedaldi, A.: Understanding image representations by measuring their equivariance and equivalence. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2015)
10. Shekhar, S., Patel, V.M., Chellappa, R.: Synthesis-based robust low resolution face recognition. arXiv preprint arXiv:170702733 (2017)
11. Zhang, P., et al.: Coupled marginal discriminant mappings for low-resolution face recognition. Optik. 126(23), 4352–4357 (2015)
12. Jiang, J., et al.: CDMMA: coupled discriminant multi-manifold analysis for matching low-resolution face images. Signal Process. 124, 162–172 (2016)
13. Li, P., et al.: On low-resolution face recognition in the wild: comparisons and new techniques. IEEE Trans. Inf. Forensics Secur. 14(8), 2000–2012 (2019)
14. Taherkhani, F., et al.: PF-cpGAN: profile to frontal coupled GAN for face recognition in the wild. In: Proceedings of the IEEE/IAPR International Joint Conference on Biometrics (IJCB) arXiv preprint arXiv:200502166 (2020)
15. Wang, M., Deng, W.: Deep face recognition: a survey. ArXiv. abs/1804.06655 (2018)
16. Taherkhani, F., Nasrabadi, N.M., Dawson, J.: A deep face identification network enhanced by facial attributes prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 553–560 (2018)
17. Dabouei, A., et al.: Boosting deep face recognition via disentangling appearance and geometry. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV) (2020)
18. Sun, Y., et al.: Deep learning face representation by joint identification-verification. In: Advances in Neural Information Processing Systems, pp. 1988–1996 (2014)
19. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815–823 (2015)
20. Taherkhani, F., et al.: A weakly supervised fine label classifier enhanced by coarse supervision. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 6459–6468 (2019)
21. Taherkhani, F., Kazemi, H., Nasrabadi, N.M.: Matrix completion for graph-based deep semi-supervised learning. Proc. AAAI Conf. Artif. Intell. 33, 5058–5065 (2019)
22. Talreja, V., Valenti, M.C., Nasrabadi, N.M.: Deep hashing for secure multimodal biometrics. IEEE Trans. Inf. Forensics Secur. 16, 1306–1321 (2021)
23. Taherkhani, F., et al.: Error-corrected margin-based deep cross-modal hashing for facial image retrieval. IEEE Trans. Biom. Behav. Identity Sci. 2(3), 279–293 (2020)
24. Talreja, V., Valenti, M.C., Nasrabadi, N.M.: Multibiometric secure system based on deep learning. In: Proceedings of the IEEE Global conference on signal and information processing (GlobalSIP), pp. 298–302 (2017)
25. Talreja, V., et al.: Biometrics-as-a-service: a framework to promote innovative biometric recognition in the cloud. In: 2018 IEEE ICCE, pp. 1–6 (2018)
26. Liu, W., et al.: Sphereface: deep hypersphere embedding for face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6738–6746 (2017)
27. Deng, J., et al.: Arcface: Additive angular margin loss for deep face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
28. Mohsenvand, M.N., Izadi, M.R., Maes, P.: Contrastive representation learning for electroencephalogram classification. In: Machine Learning for Health, pp. 238–253. PMLR (2020)
29. Hasnat, M.A., et al.: Von mises-Fisher mixture model-based deep learning: application to face verification. ArXiv. abs/1706.04264 (2017)
30. Wang, F., et al.: Normface: L2 hypersphere embedding for face verification. In: Proceedings of the ACM International Conference on Multimedia, pp. 1041–1049 (2017)
31. Liu, Y., Li, H., Wang, X.: Rethinking feature discrimination and polymerization for large-scale recognition. ArXiv. abs/1710.00870 (2017)
32. Duan, Y., Lu, J., Zhou, J.: Uniformface: learning deep equidistributed representation for face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3415–3424 (2019)
33. Guo, G., Zhang, N.: A survey on deep learning based face recognition. Comput. Vis. Image Understand. 189, 102805 (2019)
34. Goodfellow, I., et al.: Generative adversarial nets. In: Proceedings of the Neural Information Processing Systems (NIPS), pp. 2672–2680 (2014)
35. Denton, E.L., et al.: Deep generative image models using a Laplacian pyramid of adversarial networks. In: Advances in Neural formation Processing Systems, pp. 1486–1494 (2015)
36. Dosovitskiy, A., Tobias Springenberg, J., Brox, T.: Learning to generate chairs with convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1538–1546 (2015)
37. Taigman, Y., Polyak, A., Wolf, L.: Unsupervised cross-domain image generation. arXiv preprint arXiv:161102200 (2016)
38. Wu, J., et al.: Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In: Advances in Neural Information Processing Systems, pp. 82–90 (2016)
39. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein GAN. arXiv preprint arXiv:170107875 (2017)
40. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:151106434 (2015)
41. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:14111784 (2014)
42. Reed, S., et al.: Generative adversarial text to image synthesis. In: Proceedings of the International Conference on Machine Learning (ICML) (2016)
43. Chen, X., et al.: InfoGAN: interpretable representation learning by information maximizing generative adversarial nets. In: Advances in neural information processing systems, pp. 2172–2180 (2016)
44. Yin, X., et al.: Towards large-pose face frontalization in the wild. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 3990–3999 (2017)
45. Taigman, Y., et al.: Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1701–1708 (2014)
46. Zhu, Z., et al.: Multi-view perceptron: a deep model for learning face identity and view representations. In: Proceedings of the International Conference on Neural Information Processing Systems—Volume 1. NIPS'14, pp. 217–225. MIT Press, Cambridge (2014)
47. Kan, M., Shan, S., Chen, X.: Multi-view deep network for cross-view classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4847–4855 (2016)
48. Peng, X., et al.: Reconstruction-based disentanglement for pose-invariant face recognition. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1632–1641 (2017)
49. Yin, X., Liu, X.: Multi-task convolutional neural network for pose-invariant face recognition. IEEE Trans. Image Process. 27, 964–975 (2017)
50. Zhu, Z., et al.: Deep learning identity-preserving face space. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 113–120 (2013)
51. Blanz, V., Vetter, T.: Face recognition based on fitting a 3d morphable model. IEEE Trans. Pattern Anal. Mach. Intell. 25(9), 1063–1074 (2003)
52. Mallikarjun, B.R., Chari, V., Jawahar, C.V.: Efficient face frontalization in unconstrained images. In: Proceedings of the National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), pp. 1–4 (2015)
53. Zhu, X. et al.: High-fidelity pose and expression normalization for face recognition in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 787–796 (2015)
54. Bousmalis, K., et al.: Unsupervised pixel-level domain adaptation with generative adversarial networks. In: Proceedings of the IEEE Conference

on Computer Vision and Pattern Recognition (CVPR), pp. 3722–3731 (2017)
55. Yang, J., et al.: Weakly-supervised disentangling with recurrent transformations for 3d view synthesis. In: Advances in Neural Information Processing Systems, pp. 1099–1107 (2015)
56. Zhao, J., et al.: Towards pose invariant face recognition in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2207–2216 (2018)
57. Sanyal, S., Mandal, D., Biswas, S.: Aligned discriminative pose robust descriptors for face and object recognition. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 820–824. IEEE (2017)
58. Sanyal, S., Mudunuri, S.P., Biswas, S.: Discriminative pose-free descriptors for face and object matching. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3837–3845 (2015)
59. Ren, C.X., et al.: Transfer learning of structured representation for face recognition. IEEE Trans. Image Process. 23(12), 5440–5454 (2014)
60. Kang, G., et al.: Contrastive adaptation network for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4893–4902 (2019)
61. Johnson, J., Alahi, A., Fei Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Proceedings of the European Conference on Computer Vision (ECCV) (2016)
62. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. Springer (2015)
63. Isola, P., et al.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967–5976 (2017)
64. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556 (2014)
65. Gross, R., et al.: Multi-pie. In: Proceedings of the IEEE International Conference on Automatic Face Gesture Recognition, pp. 1–8 (2008)
66. Hu, Y., et al.: Pose-guided photorealistic face rotation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8398–8406 (2018)
67. Klare, B., et al.: 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 (CVPR), pp. 1931–1939 (2015)
68. Maze, B., et al.: IARPA Janus Benchmark - C: face dataset and protocol. In: Proceedings of the International Conference on Biometrics (ICB), pp. 158–165 (2018)
69. Whitelam, C., et al.: IARPA Janus Benchmark-B face dataset. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 592–600 (2017)
70. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2015)
71. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. CoRR. abs/1412.6980 (2015)
72. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 886–893 (2005)
73. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 28(12), 2037–2041 (2006)
74. Simonyan, K., et al.: Fisher vector faces in the wild. In: Proceedings of the British Machine Vision Conference (BMVC) (2013)
75. Cao, Q., Ying, Y., Li, P.: Similarity metric learning for face recognition. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2408–2415 (2013)
76. Chen, J., Patel, V.M., Chellappa, R.: Unconstrained face verification using deep CNN features. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1–9 (2016)
77. Tran, L., Yin, X., Liu, X.: Disentangled representation learning GAN for pose-invariant face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1415–1424 (2017)
78. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: Proceedings of the British Machine Vision Conference (BMVC) (2015)
79. Ding, C., Tao, D.: Pose-invariant face recognition with homography-based normalization. Pattern Recogn. 66, 144–152 (2017)
80. Xiong, C., et al.: Conditional convolutional neural network for modality-aware face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3667–3675 (2015)
81. Huang, R., et al.: Beyond face rotation: global and local perception GAN for photorealistic and identity preserving frontal view synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2439–2448 (2017)
82. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9 (2015)
83. Cao, Q., et al.: VGGFace2: a dataset for recognising faces across pose and age. In: International Conference on Automatic Face and Gesture Recognition (2018)
84. Tzeng, E., et al.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7167–7176 (2017)
85. Yi, D., et al.: Learning face representation from scratch. arXiv preprint arXiv:14117923 (2014)

How to cite this article: Taherkhani, F., et al.: Profile to frontal face recognition in the wild using coupled conditional generative adversarial network. IET Biome. 11(3), 260–276 (2022). https://doi.org/10.1049/bme2.12069