Taylor & Francis
Taylor & Francis Group

Applied Artificial Intelligence
An International Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uaai20

Improving Face Recognition by Integrating Decision Forest into GAN

Yea-Shuan Huang & Mahmood HB Alhlffee

To cite this article: Yea-Shuan Huang & Mahmood HB Alhlffee (2023) Improving Face Recognition by Integrating Decision Forest into GAN, Applied Artificial Intelligence, 37:1, 2175108, DOI: 10.1080/08839514.2023.2175108

To link to this article: https://doi.org/10.1080/08839514.2023.2175108

APPLIED ARTIFICIAL INTELLIGENCE
2023, VOL. 37, NO. 1, e2175108 (565 pages)
https://doi.org/10.1080/08839514.2023.2175108

Taylor & Francis
Taylor & Francis Group

Improving Face Recognition by Integrating Decision Forest into GAN

Yea-Shuan Huang[a] and Mahmood HB Alhlffee[b]

[a]Department of Computer Science and information Engineering, Chung Hua University, Hsinchu, Taiwan; [b]College of Computer Science and Electrical Engineering, Chung Hua University, Hsinchu, Taiwan

ABSTRACT

Posture variation and self-occlusion are well-known factors that can compromise the accuracy and robustness of face recognition systems. There are a variety of ways to combat the challenges listed above, include using Generative Adversarial Networks (GANs). Nevertheless, many GAN methods cannot guarantee high-quality frontal-face images, which can improve recognition accuracy and verification when applied to multiple datasets. Recent results have proven that the two-pathway GAN (TP-GAN) method is superior to many traditional GAN deep learning methods that provide better face-texture details due to a unique architecture that enables the method to perceive global structure and local details in a supervised fashion. Although the TP-GAN overcomes some of the difficulties associated with generating photorealistic frontal views through the use of texture information provided by landmark detection and synthesis functions, it is difficult to replicate across different datasets. Particularly, under extreme pose scenarios, TP-GAN fails to further boost photo-realistic face frontalization image samples, minimizes the training time, and reduces computational resources, all of which result in substantially lower performance. This paper proposes simple adaptive strategies for overcoming TP-GAN's inherent limitations. First, we incorporate the powerful discrimination capabilities of a decision forest into the discriminator of a TP-GAN. This method will result in a more stable discriminator model over time. Secondly, we acclimate a data augmentation technique along with a method which reduces training errors and accelerates the convergence of existing learning algorithms. Our proposed approaches are evaluated on three datasets, Multi-PIE, FEI and CAS-PEAL. We demonstrate both quantitatively and qualitatively that our proposed approaches can enhance TP-GAN performance by restoring identity information contaminated by variations in posture and self-occlusion, resulting in high quality visualizations and rank-1 individual face identification.

ARTICLE HISTORY
Received 24 August 2022
Revised 17 January 2023
Accepted 27 January 2023

CONTACT Mahmood HB Alhlffee Mahmood.bidir1985@gmail.com College of Computer Science and Electrical Engineering, Chung Hua University, Hsinchu, Taiwan

Supplemental data for this article can be accessed online at https://doi.org/10.1080/08839514.2023.2175108

Introduction

Nowadays, face-recognition systems are one of the most widely used biometric methods for identifying people and objects on digital media platforms, mobile devices, and other electronic devices through their facial characteristics. Nonetheless, the current methods of face recognition that are used in these biometric devices are very sensitive to changes in facial features caused by changes in posture, illumination, or occlusion. In some cases, posture changes can be difficult to detect if some parts of the face are not captured or if the entire face cannot be viewed in the image. This generally happens when a person is not facing the camera during video recording or photo surveillance. Unconstrained environments usually display a wide range of pose and appearance variations that can severely hamper existing frontal approach methods, which detect only faces in their frontal or near-frontal orientations. Due to the pose and appearance variations, it is extremely difficult to extract facial features or use facial landmarks to match individual faces. Many promising algorithms are typically applied to overcome these difficulties, such as GANs. In these methods, a set of pre-processing, post-processing, and feature representation techniques are used to identify individuals by utilize a few facial features or the entire face to ensure high accuracy while maintaining the face's identity across a wide range of benchmark datasets (Junho et al. 2015; Mehdi and Simon 2014; Tal et al. 2015). Unfortunately, these methods were unable to produce high-quality synthesis facial images from a single face image, minimizes the training time, and reduces computational resources that could greatly enhance their recognition accuracy (Mahmood, Yea-Shuan, and Yi-An 2022; Rui et al. 2017; Yi et al. 2021). This could be due to several factors, such as the difficulty of collecting annotated large-scale datasets, or manually collecting and annotating images, which are regarded as error-prone, or the existing face datasets may not contain adequate face samples data for a given individual, or other issues related to the reliability of face recognition algorithms. There are two approaches available to address these limitations. The first approach is to synthesize a frontal face using face frontalization (Omar et al. 2022; Yanfei and Junhua 2020), whereby traditional methods are applicable for face recognition. In certain facial datasets, these methods produced excellent results, but when dealing with benchmark datasets that contained low resolution and poor illumination conditions, the results declined dramatically. A further problem with these methods is that the probe images have poor quality due to noise variations. Various techniques can be used to enhance the robustness of the first approach, such as super-resolution (Xiaoguang et al. 2017), illumination normalization (Ziyi et al.

2018), and so forth. Nevertheless, the above methods are insufficient for handling cases involving multiple challenging factors simultaneously and are not ideal for images with extreme poses due to the sensitivity of the face texture images representation levels and variations in each area's smoothness, coarseness, and regularity. In the second approach, discriminative representations are learned directly from non-frontal faces. This can be achieved with single joint models or multiple pose-specific models. These methods are best suited for near-frontal images, but fail to produce satisfactory results for profiles taken in extreme poses due to severe texture loss and artifact. Due to the poor performance of the first approach, researchers are exploring data-driven methods to improve facial image reconstruction. For instance (Eric et al. 2022), introduced an unsupervised model for learning 3D representations from a collection of single-view 2D photographs that improves rendering efficiency while maintaining its true 3D grounded neural style. (Tero et al. 2021) proposed a small change in the GAN architecture that will prevent unwanted information from getting into hierarchical synthesis. Both translation and rotation produce identical results, but the internal representations of the networks differ significantly. (Bassel, Ilya, and Yuri 2021) presented an improved variant of GAN that utilizes label conditioning to yield high-resolution images with global coherence. This method is prone to false results in classification, which can make it difficult to distinguish between extreme poses of the same faces or objects. (Duc My et al. 2021) proposed solutions to address the serious denoising issues. First, a generator is used to restore high-frequency features such as edges and textures. A combination of the generator and discriminator was trained together to improve the model ability to preserve essential details. The second generator eliminates instabilities caused by the discriminator and restores low-frequency features in noisy images. (Xiaoguang et al. 2021) developed a method called Multi-Degradation Face Restoration or MDFR which restores high-quality frontalized facial images from low-quality images. However, some extreme poses lack fine detail and appear incomplete. Thus, we cannot guarantee the induction of successful results or achieve the same level of accuracy across different datasets. (Puja and Vinit Kumar 2020) proposed two methods for finding the descriptors or signatures of an image using face images: The Heuristic and the Local Binary Pattern. The method begins with converting the image to grayscale and ends with generating a 2D array for classification with the value [0,1], and etc. In high face poses, most GAN algorithms produces some characteristic artifact variations, such as blur, sharpness, smoothness, and clarity, which are caused by the progressive growth of neural network layers. Furthermore, some existing GAN methods are quite complex because

the generative models at each level are trained independently, i.e., they do not receive updates from each other, which affects their performance. Despite the outstanding performance of few GAN methods in generating large and high-fidelity images, random sampling produces images with lower diversity than images that are the same size in real life. Additionally, this method has limited capabilities for augmenting large-scale datasets. Others, such as (Martin, Soumith, and Léon 2017) reduce sample diversity by penalizing outliers excessively, whereas (Mahmood, Yea-Shuan, and Yi-An 2022) and (Duc My et al. 2021) is time-consuming and requires additional resources. Further, learning about neural network training loops in (Mahmood, Yea-Shuan, and Yi-An 2022) and (Duc My et al. 2021) is difficult due to their deep neural network architecture complexity and other factors related to classification performance. Few others (Rui et al. 2017) (Yi et al. 2021), and (Yanfei and Junhua 2020) are not suitable for large face poses due to neural structure weaknesses and other factors related to classification performance. Those limitations are just a few of several that are associated with face recognition in general and the GAN method in particular. Face frontalization, such as TP-GAN (Rui et al. 2017), has led to significant progress in face synthesis, providing a powerful feature extraction method that overcomes some of the limitations in generating photorealistic frontals. We argue that TP-GAN (Rui et al. 2017) has some major limitations when replicated across different datasets. Due to its reliance on landmark detection and synthesis functions, its generalization capabilities for synthesis of faces are limited. In particular, the final synthesis involves inferring global structural information and transforming local texture details into corresponding feature maps that may be sensitive to factors such as large poses, inaccurate in automatic localization landmark, or changes in facial characteristics – all of which can lead to undesirable results such as a color bias between synthetic frontal faces and input profiles, which further reduces recognition and verification accuracy. A growing concern regarding discriminators includes instability and non-convergence, which can result in weak classifiers and further difficulties in interpreting and decomposing non-linear data, affecting the discriminator's ability to model non-linear joint distributions associated with image data. Another determining factor is the complex design and deep entanglement of generator network architectures, which are susceptible to multi-factor failures, such as difficulty in training, slow learning capacity, and high resource consumption, which adversely affect face recognition accuracy and reliability. We propose a multitask learning approach to overcome the limitations of TP-GAN that integrates decision forest into TP-GAN discriminators, utilize a data augmentation technique along with

a method that reduces training-related errors and aids existing learning algorithms in accelerating convergence for face frontalization that can be used for facial recognition in conditions where identity information needs to be preserved, as well as generating high-quality images under extreme face circumstances.

Our Paper Makes the Following Contributions

We incorporate a decision forest approach to tackle the task of improving TP-GAN from a different perspective, improving its architecture, while also realigning faces across multiple poses. In particular, we consider the final layer of the discriminator network, typically a fully-connected layer (FC). This layer interprets the incapability of non-linear data to be interpreted and broken out easily, which impacts the capability of the discriminator to model non-linear joint distributions that are associated with image data. In fact, these characteristics are ingrained in decision forest (Rich and Alexandru 2006). By including a decision forest in the discriminator structure, one can improve the discriminator classification performance in a GAN setup by increasing the empirical validity's complexity.

We introduce a simple, but surprisingly effective data augmentation technique for image classification tasks in order to improve classifier performance for faces with inadequate or insufficient samples in extreme pose situations. The purpose of data augmentation is to maximize training data by increase the amount size of an existing dataset by generating completely new synthetic data from it. Methods such as this provide information that affects the accuracy of the trained network, which can be used to address issues with the training data or a lack of class balance within the datasets for each individual faces. Each generated face pose is treated with a set of strategies using appropriate parameter initializations.

Finally, we upgrade and modify the existing TP-GAN deep learning framework, such as TensorFlow and Keras, thus allowing the method to increase performance and reduce the information acquisition process. This will allow us to a build high-performance model with overall faster running time, lower GPU loads, and better memory utilization, which will require less advanced hardware and accelerate the learning process while maintaining the highest level of accuracy.

Related Work

This section focuses on the most recent approaches that have used deep learning to address the issues associated with posture variation.

Face Frontalization

Frontalization is a simple, yet effective method for maintaining a person's identity while simultaneously removing variance among different faces of the same individual. Face frontalization can be classified into three categories: 2/3D-based methods (Junho et al. 2015; Tal et al. 2015), Statistical method (Changxing and Dacheng 2017), and Deep learning method (Zhenyao et al. 2014). The 2/3D-based methods uses face geometric transformations in order to render a frontal face with either a typical 3D model or a model specific to an identity (Junho et al. 2015; Tal et al. 2015). Although they perform well under near-face front poses, they suffer greatly under large poses due to severe texture loss as well as matching face image content. Statistical method such as (Changxing and Dacheng 2017) use constrained low-rank minimization to solve a statistical model for joint frontal view reconstruction and landmark localization. Nevertheless, such a method does not have good generalizability when applied to faces in extreme poses, resulting in unreal textures, lacking identity information, and being computationally expensive. Deep learning method such as CNNs (Zhenyao et al. 2014), are another well-known method. Despite the high recognition rate, the synthesized images lack fine details and might appear blurry in large poses. As a result, these methods are insufficient to counteract face frontalization.

Generative Adversarial Networks (GANs)

In deep learning, GANs (Ian et al. 2014) have gained considerable attention because of their superior ability to generate data. The GAN consists of a generator (G) and discriminator (D) that are learned from a minmax game. In this process, the G attempts to produce realistic images that look like real, while the D learns to identify the real from the fake, guiding the G to more realistic outcomes. One characteristic that distinguishes GAN models from traditional generative models is that they produce entire images instead of pixels by pixels. The ability of GAN to produce photo-realistic images has made it one of the most popular models in a variety of fields, including image super-resolution (Yu et al. 2020), style transfer (Christian et al. 2017), and so forth. Many GAN models have been developed recently that can handle the most complex unconstrained face image situations, such as self-occlusion, illumination, or issues related to facial expression. (Junbo, Michael, and Yann 2016) developed an Energy-based GAN or EBGAN in which regions near a data manifold have low energy and other regions have higher energy. Using a regularizer loss term, this method prevents the G

from producing samples that fall into a few modes, demonstrating better convergence and higher resolution. Due to the fixed margin *m* phenomena used in EBGAN, the *D* network can't adapt to changing dynamics of the *D* and *G*, which makes reconstructed real samples difficult since energy values vary near margin *m*. (Luan, Xi, and Xiaoming 2017) developed a model called Disentangled Representation Learning GAN or DRGAN which takes a face image of any pose as input and generates a synthetic face, even for extreme profiles beyond 60º. In DRGAN, the *D* always wins easily because the convergence of *D* and *G* is unbalanced, and the training of such methods becomes unstable due to the uncertainty of prediction boundaries and the massive parameters of the traditional binary *D*. (Martin, Soumith, and Léon 2017) adapted an alternative method to traditional GAN training called Wasserstein GAN or WGAN to improve the stability of learning, avoid mode collapse, and provide meaningful learning curves that can be useful for debugging and hyper-parameter searches. Such a method demonstrates sound optimization and reveals connections with other distribution distances. Although weight clipping is used to enforce lipschitz constraints on the critic network representing the *D* network, the WGAN model still suffer/produce poor quality images and fail to converge. Further, WGAN model is restricted to a limited number of functions. (Tian et al. 2018) introduced a Load Balanced GAN or LBGAN, the authors contribution is to rotate the input face image to a target angle based on a set of known poses. Due to the high success rate of the *D*, the gradient of the *G* does not appear. An imbalance between the *G* and *D* leads to overfitting. (Yanfei and Junhua 2020) proposed a Pose-Conditional (PC) method that extends Cycle-GAN method to create pose-invariant frontal images that preserve subject identity. Synthetic frontals can reduce recognition and verification accuracy because deep neural networks are required to explore high-value features. (Rui et al. 2017) Combined adversarial loss, symmetry loss, and identity preservation loss to constrain ill-posed problem which called TP-GAN. Using pre-trained discriminative deep face models to infer frontal views from profiles. Compared to the original GAN model, the TP-GAN performs better under extended face poses, the image generation process can be more controlled, and the outcome can be interpreted more easily. However, such method still produces poor quality faces due to the unbalanced training between *G* and *D*, and its generalization abilities are limited due to its dependence on landmark detection and synthesis loss functions. (Duc My et al. 2021) presented a useful solution to address these serious denoising issues where the use of a generator restores high-frequency features such as edges and textures. The proposed method improves *D* stability by developing an adversarial loss function as well as developing

a multi-generation that prevents mode collapse more effectively than a single-generation. However, the training process takes longer and requires more resources than usual, and the loops in neural networks are difficult to understand because of their complexity. (Mahmood, Yea-Shuan, and Yi-An 2022) extend the work of (Rui et al. 2017) by introducing additional landmark detection and denoising methods called LFM. The method shows high quality face samples that are superior to many traditional methods. However, the consumption of hardware and training time is quite high. These methods are highly regarded because they are capable of sustaining some complex facial poses and are considered a base for achieving good accuracy. The solution to face pose issues requires a great deal of analysis, dedication, and the right amount of effort. The uniqueness of our method lies in the fact that we consider every aspect of face recognition and deal with it simultaneously.

Decision Tree and Forest

Decision tree and forest are well-known for their strong discriminating capabilities for calculating a target value by applying a set of binary rules. "The term forest refers to models made up of multiple decision trees. Forest predictions are the summation of the predictions of its decision trees." Deep learning incorporates decision forest in two different ways: implicitly, by using the decision trees to influence the training process (Alvaro-H and Robert 2020; Yani et al. 2016), or explicitly, by incorporating decision trees into the core architecture (Peter et al. 2015; Yan, Gil, and Tom 2021). To that end, it would be necessary to review some of these approaches in more detail. For instance (Yan, Gil, and Tom 2021), proposed a decision forest method to modify GAN architectures, resulting in strong discrimination. Such kind of framework exploits facial landmarks to disentangle pose-invariant features and pose-adaptive losses to handle the imbalance issue adaptively. (Peter et al. 2015) presented deep neural decision forest, which combine the advantages of representation learning and decision trees. In particular datasets, this method produces good results while maintaining low error rates. By contrast, these methods learn representations that are semantically correlated across layers, whereas our method learns semantically independent representations across layers. This type of representation is used in order to learn category-specific features and control the generation of face images accordingly.

Data Augmentation

Face recognition datasets commonly contain near-frontal faces, which are used to train deep learning models to achieve desired accuracy levels. Nowadays, the amount of face pose data required to train most GAN

models is limited or insufficient, which can inhibit generalization. The use of data augmentation techniques improves the generalizability of neural networks by exploiting existing training data more efficiently. However, standard data augmentation methods produce limited plausible alternative data. It has been found that GANs can optimize the amount size of training data by creating completely new synthetic data from existing data. (Subhajit et al. 2022) proposed a GAN-based model for generating synthetic images. This modification attempt to increase the image synthesis quality and reduced mode collapses by using a lightweight GAN model that consists of *G*, *D* and an auto-encoding structure to capture essential parts of the input image. (Chao, Ziqi, and Shiwen 2022) proposed a GAN-based data augmentation method, called RFPose-GAN, to create synthetic datasets for multi-model neural networks. Experimental results demonstrate that the proposed data augmentation approach improves 3D human pose tracking performance with limited training data. (Ngoc-Trung et al. 2021) proposed a principled framework called data augmented for GAN or DAG that allows augmented data to be incorporated into GAN training. Then DAG is compared to the original GAN to demonstrate that it minimizes the Jensen – Shannon (JS) divergence between the original and model distributions for better *G* and *D* learning. (Tongyu et al. 2021) adapted an approach to handling missing data. In this approach, noise patterns are extracted from target data, and the source data is adapted with the extracted target noise patterns while preserving supervision signals. After that, it retrains the model using the adapted data to better serve the target. In (Tongyu et al. 2021), data augmentation is based on an unsupervised framework called depth-aware or DAGAN based on GAN. The DAGAN method generalizes any data item within the source domain to produce other data items within the same class. The generative process does not depend on the classes themselves, so it can be applied to novel classes of data. Our data augmentation approach differs in a few ways (e.g., large training samples for each face pose, light-CNN for identification, label verification, among other technical aspects). For each face pose generated, a set of strategies is applied based on appropriate parameter initializations that are primarily focused on low balanced-pose representations. Deep neural networks can become more accurate as training face samples increase in size and minimize overfitting among layers by learning and adapting more details each time. Moreover, we provide an easy way to enhance model accuracy without building a complex application system or utilizing an additional model. The goal is to enhance the TP-GAN performance by synthesizing face images for each class pose and reducing the model complexity.

Several relevant points can be summarized based on our related work. Although the existing methods produced reasonable results in some face image datasets for which they were designed and had robust performance across poses, there is no guarantee that the same result would be obtained if replicated to other datasets. For tasks such as facial normalization or other face synthesis tasks, deep learning methods still fail to generate high-quality image samples under excessive pose scenarios and illumination variation, leading to significantly lower final performance results that are unavoidable. The purpose of this study is to enables TP-GAN model to deal with the face pose problem more efficiently in order to produce adequate results.

Proposed Method

This section provides a brief overview of the TP-GAN architecture and then discusses our proposed methods in detail.

TP-GAN Architecture

As shown in Figure 1, the TP-GAN (Rui et al. 2017) layout consists of two neural network structures. The first layout is a two-pathway convolutional neural network G_{θ_G} parameterized by θ_G. Every pathway contains an encoder G_{θ_E} and decoder G_{θ_D}networks, along with a set of loss functions. A local pathway $\left(G_{\theta_E^l}, G_{\theta_D^l}\right)$ consists of four landmark patch networks $G_{\theta_i^l}, i \in \{0, 1, 2, 3\}$ representing local textures surrounding a facial landmark. On the other hand, the global pathway network $\left(G_{\theta_E^g}, G_{\theta_D^g}\right)$ processes global facial structures. The G_{θ_G}, which is the output of $G_{\theta_E^g}$, is normally used for classification tasks using $L_{cross-entropy}$. The second layout consists of a discriminator D_{θ_D} that distinguishes ground truth frontal views $(GT)$from synthetic frontal views (SF). A detailed description is given in (Rui et al. 2017).

Discriminator with Decision Trees and Forest

In our approach, we incorporate a decision forest by replacing the final fully connected (FC) layer of the TP-GAN discriminator network with a decision forest architecture. The decision forest is designed in such a way that it can be easily integrated into the discriminator network, and the whole model can be trained supervised. As in (Yan, Gil, and Tom 2021), we replaced the hard decision routing function with a soft, different sigmoid function at each decision node. In contrast to hard internal nodes, soft decision nodes redirect instances to all of their children with a certain probability. Essentially, we follow every path to every leaf, and each leaf makes a contribution to the final

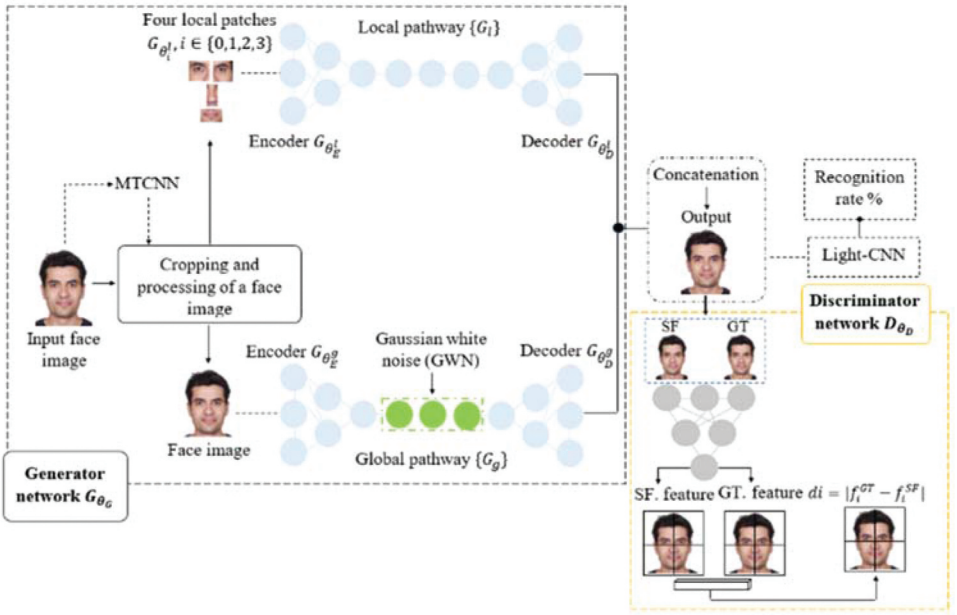

Figure 1. A brief overview of the TP-GAN model architecture. modifications have been made to the existing TP-GAN model to produce this architecture. The generator network that combines a two-pathway layout (local and global pathways) and a discriminator with a single deep neural structure, followed by a light-CNN model, determines the accuracy of identity-preserving properties.

decision, but with a different probability. In this way, we can minimize the need for a stochastic hard routing approximation on a forward pass through the trees by using the soft functionality of the decision nodes in our ensemble. Moreover, in our method each face pose (e.g. $\pm 60^{o}$, $\pm 75^{o}$ and $\pm 90^{o}$) is trained and optimized according to a $K - class$ misclassification error over the parameters of the node, thereby reducing misclassification errors. Both leaf and decision nodes are updated simultaneously instead of alternately. Figure 2 illustrates our modification network architecture.

Decision trees and forest are well-known for their discriminative abilities that enable better decision making and faster learning. Specifically, we aim to produce a representation of a face that preserves feature such as poses, structure, texture details and other regarded variables, as well as the ability to achieve high accuracy in a variety of face situations, allowing us to perform advanced face-related classification analyses that may use a portion of the facial images, thereby improving the model performance. The facts above were not mentioned or addressed in (Ma et al. 2021; Peter et al. 2015; Shuhui et al. 2020; Yan, Gil, and Tom 2018, 2021). These methods work well when the dataset is perfect or does not present any challenges, and can achieve high accuracy in face classification. From a point of view, these methods

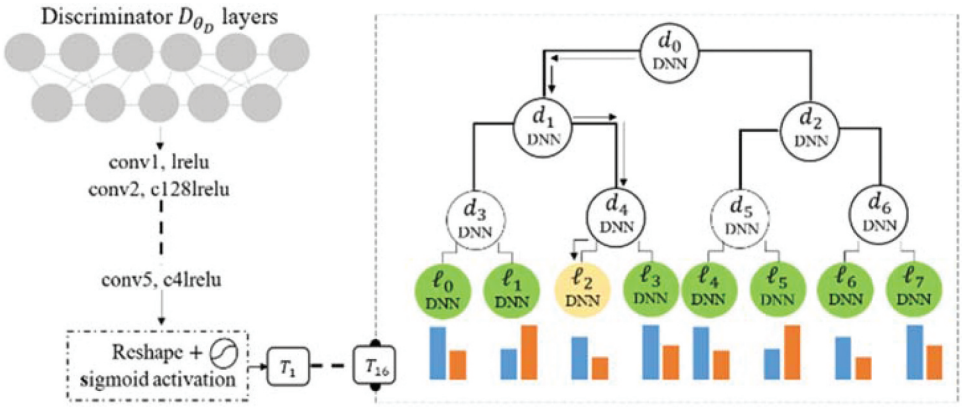

Figure 2. Overview of our approach to modifying the TP-GAN *D* network. where *d*and ℓ represent the decision node and leaf node, respectively. each decision node/leaf is implemented by a deep neural network (DNN) structure. our forest consist of 16 trees with 9 depth levels (i.e. $T_1, \ldots, T_{16}$). arrows represent paths used to route information of the sample *x*along a tree to reach leaf ℓ_2, which has probability $\mu_{\ell_2} = d_0(x)\bar{d}_1(x)\bar{d}_4(x)$.

often deal with frontal and near-frontal problems by using large amount of datasets for training. Our framework can be viewed as special cases, and these works may not have the similar characteristics to our case studies. In this paper, we consider multiple factors such as the network structure, the process transferring of the face poses, diversity of intensity levels and other related face recognition issues while constructing the decision tree and forest, which allows for sustained accuracy with fewer parameters and enabled optimization models to learn functions from one or two-class data. One of the key features that distinguish our work from existing ones, is the way we train decision trees and forest model, which consist of a multi-step process to ensure high accuracy for each face poses.

Preliminaries: Let $x \in X$ represents a set of features variable $(x_1, \ldots, x_n)$, where X is the total feature set of x, and $y \in Y$ represents a set of target variable $(y_1, \ldots, y_n)$, where Y is the total variable set of y. The decision tree is made up of N_d, representing the decision nodes, and N_ℓ, representing the leaf nodes. Here d is a decision node and ℓ is a leaf node. Data is routed down the decision tree using decision nodes by splitting it up and sending it to the left or right child node, whilst leaf nodes hold the prediction distribution. N_ℓ, are those nodes that have no additional branches coming off of them. They do not further split the data; they simply classify the examples located within each node. All the other N_d, node in the tree can be referred to as split nodes, decision nodes, or internal nodes. Each leaf node $\ell \in N_\ell$ holds a probability distribution π_ℓ over Y. Each decision node $n \in N_d$ is paired with

a decision function $d_n : (x; \theta) \rightarrow \{0, 1\}$ parametrized by θ, which determines where the samples will be routed to left child node when d_n is 0 and to right child node when d_n is 1. The N_d map input sample, X, from the root node to the final leaf node: $\ell = N_d(x; \theta)$. Whenever a sample ends in ℓ, a class-label distribution π_ℓ provides a prediction of the related tree.

Soft Decision Tree (T)

In soft decision tree, we combine the values of leaves based on their proportions in order to arrive at a single prediction. Each decision node returns a value indicating the proportion of its left and right subtrees:

$$d_n(x; \theta) = \sigma(f_n(x; \theta)) \tag{1}$$

where $\sigma(x) = (1 + e^{-x})^{-1}$ represents a sigmoid function, while the $f_n(x; \theta) : X \rightarrow \mathbb{R}$ is represented as real-valued function depending on the x, θ and x is the total set of x(whichmeans$x \in X$). f_n responds as a linear output unit with DNN functionality, which then further applies a σ function to reach a response within $[0, 1]$. All the nodes are designed with deep neural network (DNN) advantages to produce an accurate feature representation and a powerful classifier that can learn the capabilities of different classes with minimal complexity. For each ℓ, we define $\mu_\ell(x; \theta)$ as the blending function that determines its proportional contribution to the tree's final output:

$$\mu_\ell(x|\theta) = \prod_{n \in N_d} d_n(x; \theta)^{1_{\ell \swarrow n}} \bar{d}_n(x; \theta)^{1_{n \searrow \ell}} \tag{2}$$

where $\bar{d}_n(x; \theta) = 1 - d_n(x; \theta)$, 1_C is an indicator function that is set to 1 when its condition C is met, and 0 otherwise, $\ell \swarrow n$ means ℓ belongs to *left* n node and $n \searrow \ell$ means ℓ belongs to *right* n node. Among all $\mu_\ell(x|\theta)$, only one has a single value of 1 while the others are 0. Despite the fact that the product in Eq. (2) runs over all nodes, only decisions that lead to the root ℓ node contribute to μ_ℓ. Please refer to Figure 2. Therefore, our final prediction value is determined by:

$$Q(y|x, \theta, \pi) = \sum_{\ell \in N_\ell} \pi_{\ell_y} \mu_\ell(x|\theta) \tag{3}$$

where θ represents the gathered parameters of all the N_d and N_ℓ values, π_{ℓ_y}denotes the probability of a sample reaching ℓ to take on class y, $\mu_\ell(x|\theta)$ presents as the routing function providing the probability that sample x reachs ℓ. In addition, because we use decision trees that make soft decisions rather than hard ones, we can easily generate gradients for updating our model by error backpropagation.

Soft Decision Forest (F)

We follow the discerption in (Yongxin, Irene, and Timothy 2018), in which we do not strictly split to left or right nodes, because we use differential binning that can split nodes into multiple leaves. Additionally, the most relevant input features are taken into account in order to ensure interpretability of complex data distributions. As part of a soft decision forest, all prediction contributions are based on the number of trees in the ensemble. A decision forest is an ensemble of decision trees $F = (T_1, \ldots, T_N)$ that predicts the outcome by a majority voting mechanism, which can be calculated using the following formula:

$$F[y|x] = \sum_{n=1}^{N} Q(y|x, \theta_n, \pi_n) \tag{4}$$

here N represents the number of the trees, and Q represents the prediction value of a single tree. By using the soft approach, Fcan easily be trained in forward and backward passes. A modification of this kind allows us to train F supervised, where N_ℓ and N_d are updated simultaneously.

Learning Tree Nodes

Learning a tree requires estimating both the leaf predictions π and decision node θ, for back-propagation purpose. With respect to the given data set $\mathcal{T} \subset X \times Y$ under log-loss, we adhere to the minimum empirical risk (R) principle, i.e. determine the minimizers of the following risk term:

$$R(\theta, \pi; \mathcal{T}) = \frac{1}{|\mathcal{T}|} \sum_{(x,y)\in\mathcal{T}} L(\theta, \pi; x, y) \tag{5}$$

where L is a weighted cross-entropy loss (weighted by its path probability $\mu_\ell(x|\theta)$ and the target distribution π) with respect to a given data set sample $(x, y) \in \mathcal{T}$, which can be calculate as follow:

$$L(\theta, \pi; x, y) = -log(Q(y|x, \theta, \pi)) \tag{6}$$

where Q can be referred to Eq. (2). The next paragraph describes a two-step optimization approach in order to minimize Eq. (5).

Each decision function in Eq. (1) is parametrized by the same θ, which in turn parametrizes the function f_n. Until now, we have not made assumptions about the types of functions in f_n, so there is no reason that optimizing risk with respect to θ for a given π may ultimately become a difficult and large-scale optimization problem. Hence, to

minimize the risks associated with θ, we will use Stochastic Gradient Descent (SGD), which is common to DNN:

$$\theta^{(t+1)} = \theta^{(t)} - \frac{\eta}{|\beta|} \sum_{(x,y)\in\beta} \frac{\partial L}{\partial \theta}\left(\theta^{(t)}, \pi; x, y\right) \tag{7}$$

where, $0 < \eta$ represents the learning rate, t represents the time step, and $\beta \subseteq \mathcal{T}$ represents the subset (β represented as mini-batch) of samples from the training set. According to the chain rule, the gradient of loss L with respect to θ can be decomposed as follows:

$$\frac{\partial L}{\partial \theta}(\theta, \pi; x, y) = \sum_{n\in N_d} \frac{\partial L(\theta, \pi; x, y)}{\partial f_n(x;\theta)} \frac{\partial f_n(x;\theta)}{\partial \theta} \tag{8}$$

based on the decision tree, the gradient term will be as follows:

$$\frac{\partial L(\theta, \pi; x, y)}{\partial f_n(x;\theta)} = d_n(x;\theta) A_{n_r} - \bar{d}_n(x;\theta) A_{n_l} \tag{9}$$

where n_r and n_l indicate the *right* and *left* children of node n, respectively. Here A_m is defined for generic node $m \in N$ which can be computed as follows:

$$A_m = \frac{\sum_{\ell\in N_m} \pi_{\ell_y} \mu_\ell(x|\theta)}{Q(y|x,\theta,\pi)} \tag{10}$$

here N_m represents the set of leaves held by the subtrees rooted in node m. The detailed derivation of Eq. (10) can be found in subsection 2.5, which describes an efficient way to compute A_m for all nodes m in a single traversal of T.

Taking into account the rules for updating the decision function θ from the previous subsection, we now consider the problem of minimizing Eq. (5) with respect to π when θ is fixed:

$$\min_{\pi} R(\theta, \pi; \mathcal{T}) \tag{11}$$

due to the fact that this is a convex optimization problem, it is easy to recover the global solution. This problem was encountered in (Samuel Rota and Peter 2014) but only at a single node level. Nevertheless, we consider the whole tree here, and all predictions at each ℓ are jointly estimated. A global minimizer of Eq. (11) is computed by the following iterative scheme that updates the distribution of N_ℓ in iteration $t+1$:

$$\pi_{\ell_y}^{(t+1)} = \frac{1}{P_\ell^{(t)}} \sum_{(x,y)\in\beta} \frac{\pi_{\ell_y}^{(t)} \mu_\ell(x|\theta)}{Q(y|x,\theta,\pi^{(t)})} \tag{12}$$

for all $\ell \in N_\ell$ and $y \in Y$, where $P_\ell^{(t)}$ is a normalizing factor ensuring that $\sum_y \pi_{\ell_y}^{(t+1)} = 1$. As long as every element is positive, the starting point π^0 can be

arbitrary. A typical choice would be to use the uniform distribution in all leaves, i.e. $\pi^0_{\ell_y} = |Y|^{-1}$.

Training Decision Forest

Until now, we have outlined the process for a single decision tree. The next step is to discuss an ensemble of F, where each T have a same structure (as mentioned in subsection 2.2 and subsection 2.3). Due to the fact that each T in F has its own set of ℓ parameters π, we can update the prediction nodes of each tree independently based on the contemporary estimate of θ.

As for, instead, we randomly select a T in F for each β, and then use an update mechanism the same way as described in subsection 2.3. The strategy is somewhat similar to Dropout (Nitish et al. 2014), in which SGD updates are applied to different network topologies based on a specific distribution. In addition, updating individual Trather than the entire forest reduces computational workload.

Additional Implementation Details

Decision Nodes: The d_n is defined using real-valued functions $f_n(x;\theta)$, which are not necessarily independent. The aim is to endow the trees with feature learning capabilities by embedding functions f_n within a CNN with θ. Each function f_n can be viewed as a linear output unit of a DNN that will generate a probabilistic routing decision after d_n applies a σ activation in order to obtain a $[0, 1]$ response. As proposed, DNN output units do not directly deliver predictions, e.g. via softmax layers, but rather drive the d_n in a forest through their own decisions. When a data sample x passes through the DNN forward, the soft activations of the routing decisions of T induce a mixture of ℓ predictions based on Eq. (3), which is the final outcome. Last but not least, we obtain a similar model to oblique forest by assuming linear and independent (via separate parametrizations θ) functions $f_n(x;\theta_n) = \theta_n^\top x$, the model recover is similar to that in (Sreerama, Simon, and Steven 1994).

Routing Function: The routing function μ_ℓ can be conducted by traversing the T once. Let $\top \in N$ be the root node and for each node $n \in N$ let n_r denote as *right* child and n_l denote as *left* child, respectively. We start from the root by setting $\mu_\top = 1$ and for each $n \in N$ that we visit in breadth-first order we set $\mu_{n_l} = d_n(x;\theta)\mu_n$ and $\mu_{n_r} = \bar{d}_n(x;\theta)\mu_n$. The desired routing function values can be read from the leaves at the end.

Learning Decision Nodes: Back-propagation algorithms precompute the routing function $\mu_\ell(x;\theta)$ and T prediction $Q(y|x,\theta,\pi)$ for each sample (x,y) in the β during the forward pass. Every sample (x,y) in the β must have its gradient term computed in Eq. (9) through backward pass. In this case, a single traversal of the tree can be performed from the bottom way to up. To begin, let's set:

$$A_\ell = \frac{\pi_{\ell_y}\mu_\ell(x|\theta)}{Q(y|x,\theta,\pi)} \tag{13}$$

for each $\ell \in N_\ell$, afterword, we visit the T in reversed breadth-first order (bottom-up). As long as we can read A_{n_r} and A_{n_l} from the children, we can compute the partial derivative in Eq. (9) since $A_n = A_{n_r} + A_{n_l}$ will be needed by the parent (Sreerama, Simon, and Steven 1994).

A summary of the learning procedure can be found in Algorithm. A training set $\mathcal{T}$ is used to start with random initializations of decision nodes parameters θ and then iterate the learning procedure for a predetermined number of iterations.

Summary of Algorithm

Require: $\mathcal{T}$: training set, *nEpochs* $(x,y) \in \mathcal{T}$

1. random initialization of θ
2. **for all** $i \in \{1,\ldots,nEpochs\}$ **do**
   Compute π by iterating $\pi_{\ell_y}^{(t+1)} = \frac{1}{P_\ell^{(t)}} \sum_{(x,y)\in\beta} \frac{\pi_{\ell_y}^{(t)}\mu_\ell(x|\theta)}{Q(y|x,\theta,\pi^{(t)})}$
3. break $\mathcal{T}$ into a set of random mini-batches
4. **for all** β: mini-batch from $\mathcal{T}$ **do**
5. Update θ by SGD step in $\theta^{(t+1)} = \theta^{(t)} - \frac{\eta}{|\beta|}\sum_{(x,y)\in\beta}\frac{\partial L}{\partial\theta}\left(\theta^{(t)},\pi;x,y\right)$
6. **end for**
7. **end for**

Deep Learning Framework

In this section, we upgrade and modify the existing TP-GAN deep learning framework, such as TensorFlow and Keras, and then compare the TP-GAN and our method frameworks by training them on Multi-PIE, FEI and CAS-PEAL datasets. This modification includes updating packages and sub-packages, reshaping some of the code, and utilizing multi-proceeding technique for fast execution, among other things. Each framework is evaluated according to its overall running time "number of epoch," GPU load, and memory utilization. The supplementary material contains additional details. The frameworks were compared on the same machine, which had the following specifications: Intel (R) Core i7-9700K, CPU@ 3.60-GHz, 3600-MHz, 8-Cores, 8-Logical Port, 64-bit OS, x64-based processor, NVIDIA GeForce RTX-2080-Ti, and 32.0-GB RAM. Our upgrade and modification will enable us to develop model that are significantly more accurate and time-efficient than existing ones,

by eliminating the resources consumed by each framework to reach a certain level of accuracy. A model of this kind can perform deep transfer learning and result in a facial recognition model that can be fine-tuned to make accurate predictions on several facial datasets. This task improved the overall speed training time by *8-to-11%*.

Experiments

Extensive experiments have been conducted to demonstrate the superiority of our method over TP-GAN. Thus, we evaluated three factors: the visual quality of the face, the accuracy of face recognition, and the efficiency of deep learning.

Experimental Settings

The TP-GAN model and our model are trained and evaluated on the Multi-PIE, FEI, and CAS-PEAL datasets. Multi-PIE (Ralph et al. 2010) consists of a large dataset of 75000 images of 337individual faces, taken under constrained conditions in a variety of poses, illuminations, and expressions. For each individual face, there are 15 poses ranging around $\pm 90^0$, as well as20 illumination levels. FEI (FEI 2005–2006) is a Brazilian face database consisting of images taken at the artificial intelligence laboratory of FEI in Sao Bernardo do Campo, Brazil. For a total of 2800 images, each face is composed of 14 images. Face images are taken in an upright frontal position with a profile rotated 90^0 to the left or right against a uniformly white background. Faces range in age from 19 to 40 years old, with distinct features, hairstyles, and adornments. CAS-PEAL (Wen et al. 2009) is a large-scale gray Chinese face dataset with 99594 images of 1040 individuals (595males and 445 females). Each face is represented by 27 images in three shots with different poses ranging $\pm 90^0$ "straight ahead, up, down and sideways." Furthermore, the database includes 5 facial expressions, 6 accessories, and 15 illuminations variations. 80% of the subjects in the above datasets were used for training, while 20% were used for testing.

Quality Results of Face Frontalization

We compare the quality of the face frontalization of our method to TP-GAN on three different datasets: Multi-PIE, FEI, and CAS-PEAL. In Figure 3, Figures 4 and 5, the first column represents the face pose, the second column represents the profile images taken under different posing conditions, the third column represents the TP-GAN generated face images, the fourth column represents our method generated face images, and the last column represents a random ground-

truth image from each category. Our method is capable of inferring and reconstructing better frontal views with more accurate facial texture and appearance than TP-GAN. TP-GAN results can be drastically affected if the occlusion factor is even slightly changed, as in the case of the CAS-PEAL dataset. As a result, the TP-

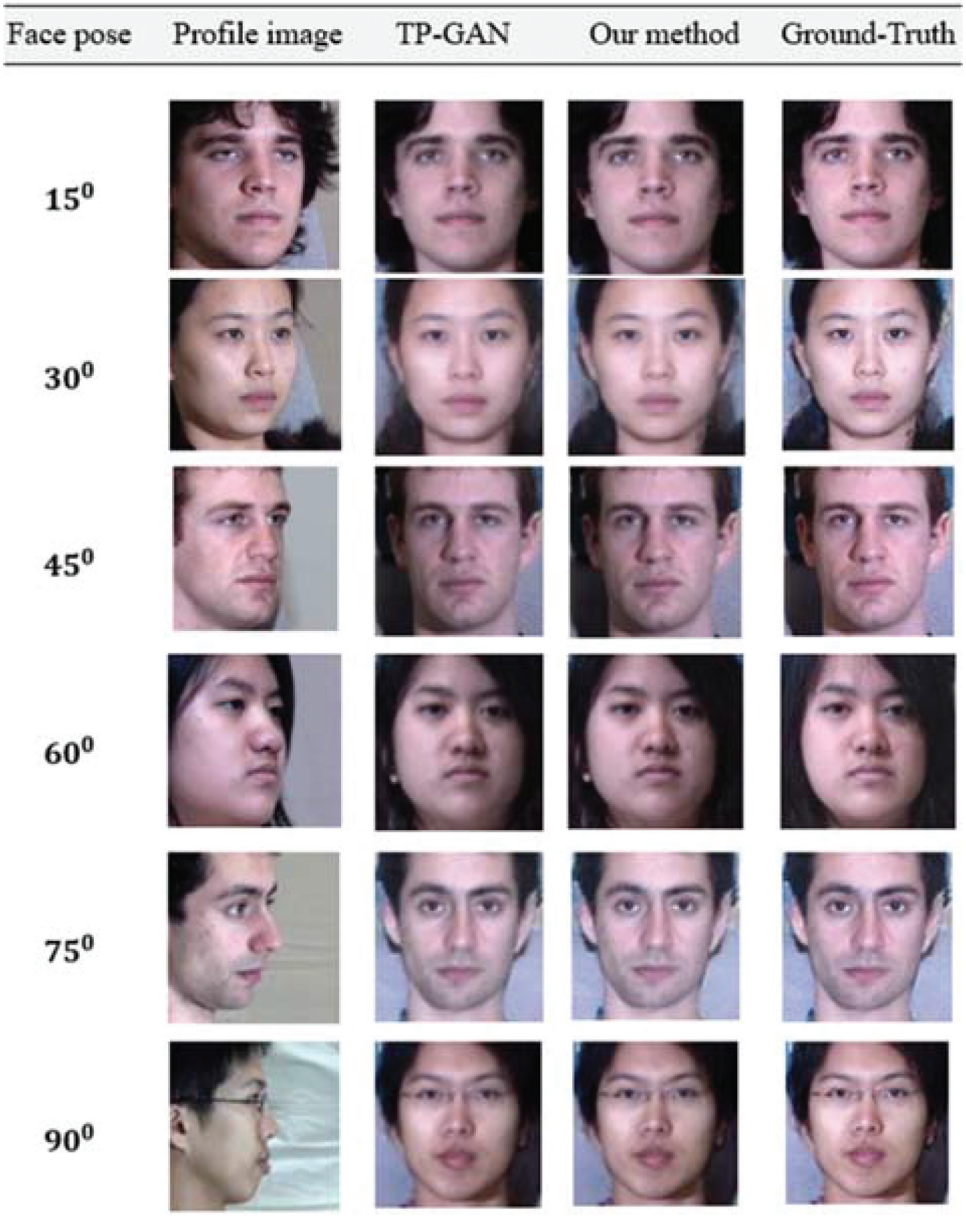

Figure 3. Comparison of our method's generated facial images with those generated by TP-GAN on the multi-PIE database. despite significant abnormalities in the faces image, our synthetic faces seem convincing. The dataset was downloaded from the TP-GAN GitHub repository at: https://github.com/HRLTY/TP-GAN.

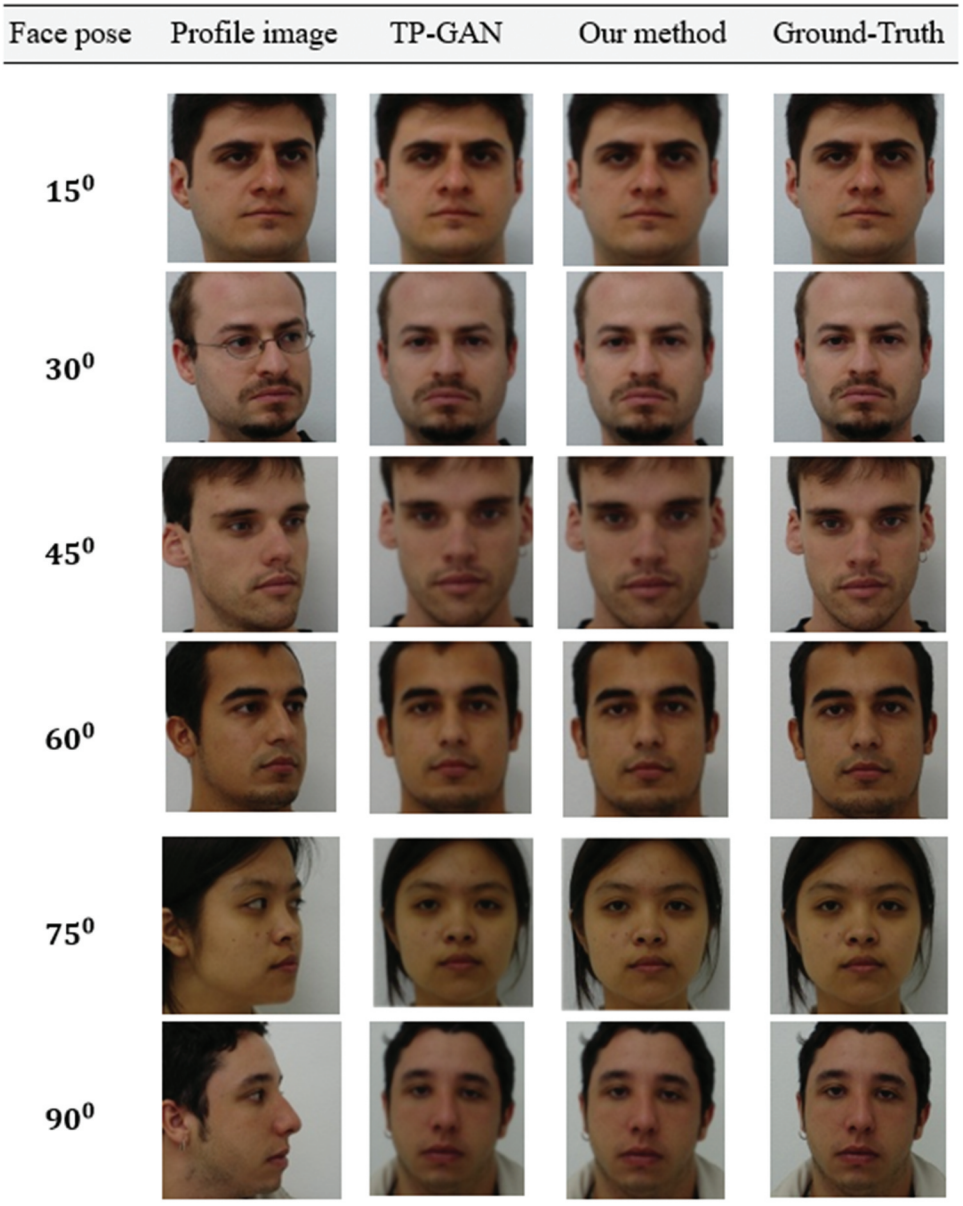

Figure 4. Comparison of our method's generated facial images with those generated by TP-GAN on the FEI database. Our method consistently produced better texture detail for all face poses. We downloaded the dataset from the FEI official repository at: https://fei.edu.br/~cet/facedatabase.html.

GAN is unable to handle large occluded regions, resulting in a flawed facial structure that lacks details. In other words, TP-GAN isn't very resistant to occlusion. Our method has the ability to comprehend the facial structural information, allowing us to provide more stable and visually satisfying results over the TP-GAN.

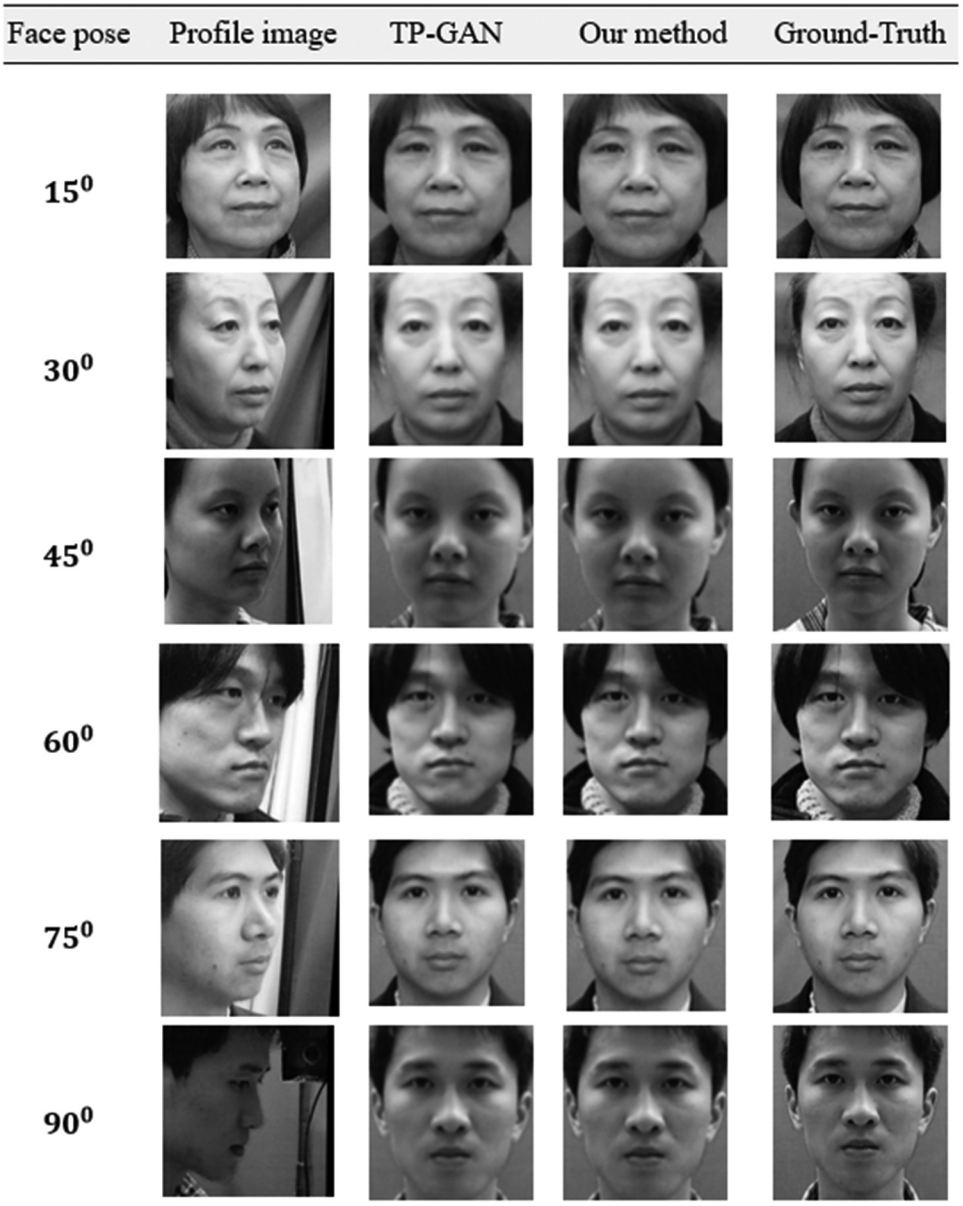

Figure 5. Comparison of our method's generated facial images with those generated by TP-GAN on the CAS-PEAL database. Despite illumination variations such as grey faces, our method consistently produced better texture detail. We downloaded the dataset from CAS-PEAL official repository at: https://github.com/YuYin1/DA-GAN.

Generally, most facial recognition methods assume that if the posture is greater than 60°, it is difficult to reconstruct the image of the frontal view. It is particularly difficult for most methods to handle the large-pose problem due to the significant semantic information that is lost

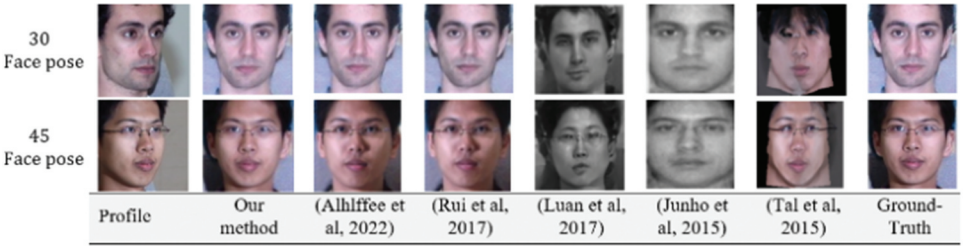

Figure 6. A comparison of our frontal-profile synthesis results with those from various methods in the Multi-PIE dataset, using 30° and 45° face poses. We downloaded the dataset from the TP-GAN GitHub repository at: https://github.com/HRLTY/TP-GAN.

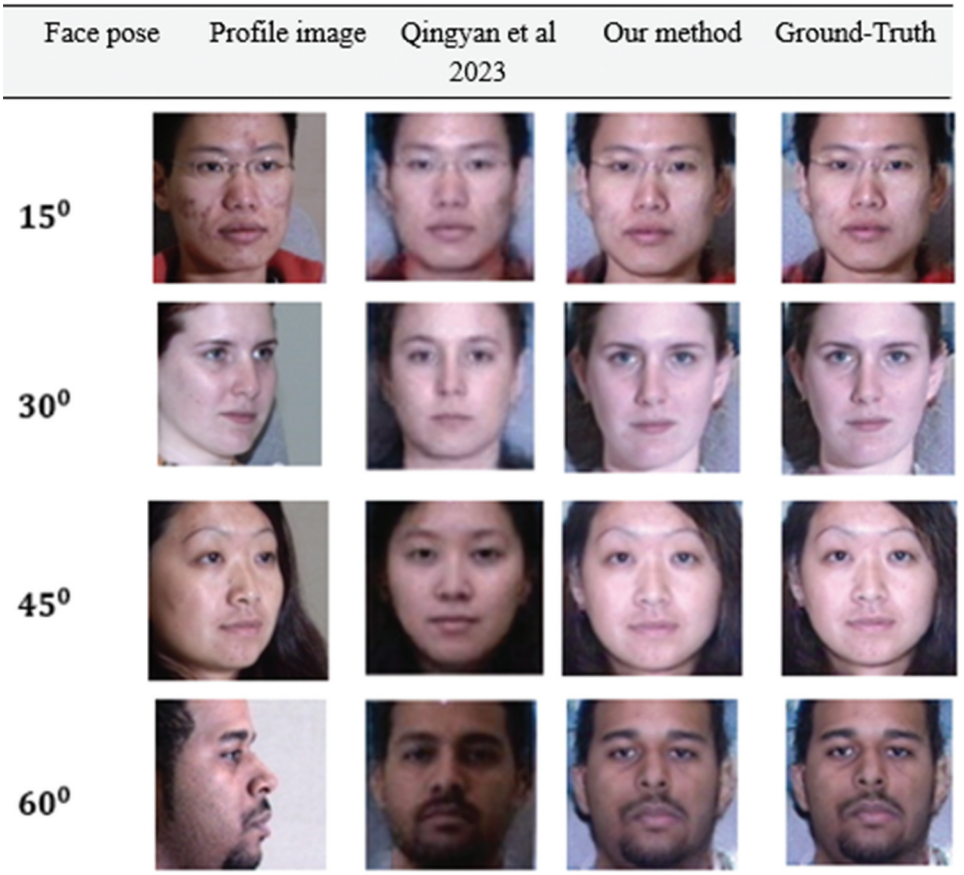

Figure 7. A comparison of our frontal-profile synthesis results with those generated on Multi-PIE dataset, using various face poses. We downloaded the dataset from the TP-GAN GitHub repository at: https://github.com/HRLTY/TP-GAN.

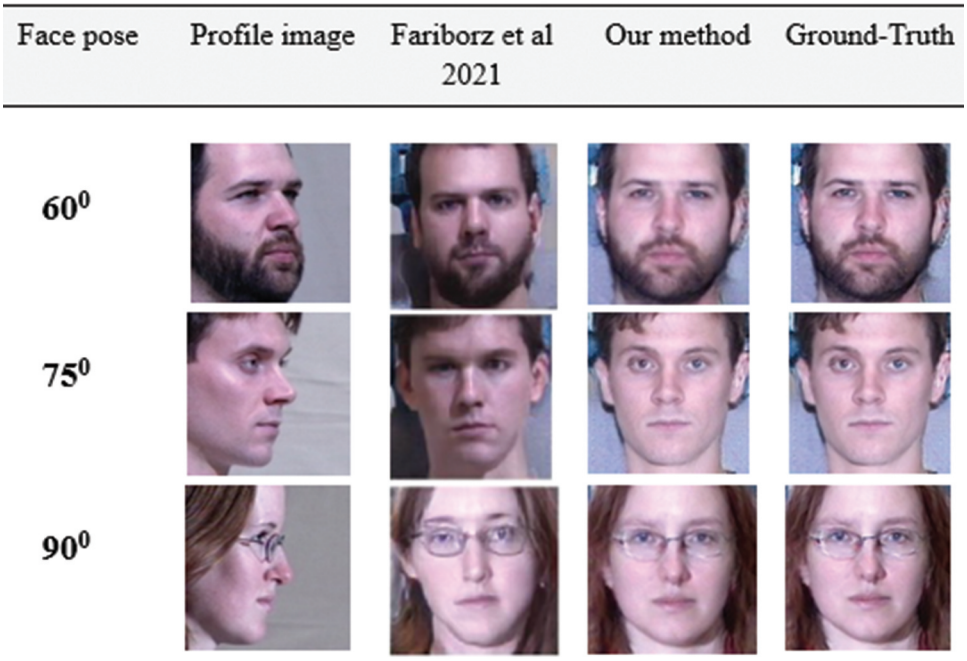

Figure 8. A comparison of our frontal-profile synthesis results with those generated on Multi-PIE dataset, using various face poses. We downloaded the dataset from the TP-GAN GitHub repository at: https://github.com/HRLTY/TP-GAN.

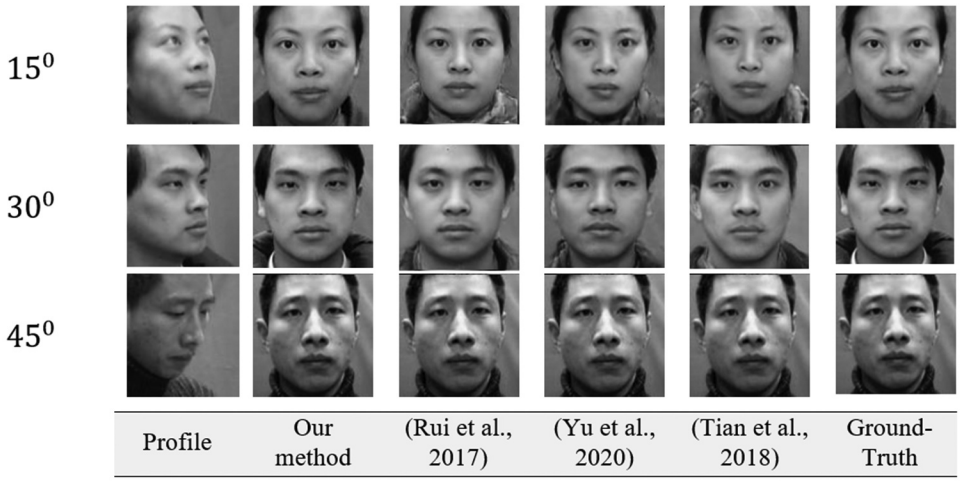

Figure 9. Our frontal-profile synthesis results have been compared with those obtained from various methods using 15°, 30° and 45° face poses from the CAS-PEAL dataset. We downloaded the dataset from CAS-PEAL official repository at: https://github.com/YuYin1/DA-GAN.

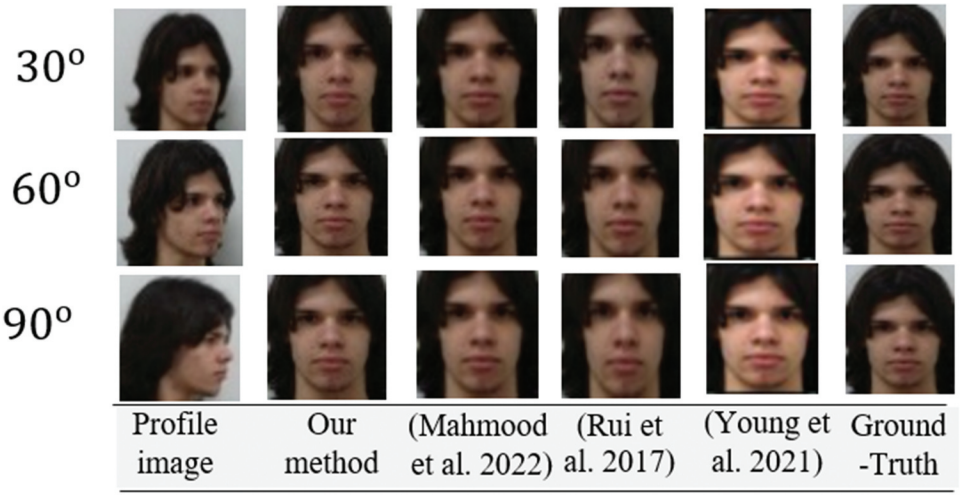

Figure 10. Comparing our frontal-profile synthesis results with those from other methods in the FEI dataset, using 30, 75 and 90 degree face poses. The dataset was downloaded from FEI official repository at https://fei.edu.br/~cet/facedatabase.html.

when high-pose occurs. Nevertheless, we will demonstrate that outstanding results can be achieved by providing enough training data and a carefully designed algorithm, as in Figure 6, Figure 7, Figure 8, Figure 9 and Figure 10 which illustrate our method's visual effects and variations on synthetic face images as compared to the state-of-the-art methods. Due to the constraint of diverging from the identity feature of the input image, many high performance (Junho et al. 2015; Luan, Xi, and Xiaoming 2017; Mahmood, Yea-Shuan, and Yi-An 2022; Rui et al. 2017; Tal et al. 2015; Yanfei and Junhua 2020) methods produce inferior visual effects as compared to ours in terms of image quality and outline details. Further, the synthesis images in some large poses have some distortion in texture information (brightness, sharpens, clarity and blur) (Fariborz et al. 2021; Qingyan et al. 2023; Young, Byung-Gyu, and Partha-Pratim 2021) in contrast to our results, such as the regain covering the four areas of the mouth, nose, and eyes, a constraint that preserves the context of details. The synthesized face images generated by our model are clearer than those generated by state-of-the-art methods which reveal their lack of fine texture details and missing face content information. Thus, our model performs better than its variants, indicating that the structure of the model has been strengthened to withstand large facial poses.

Face Identity Preservation

In order to demonstrate the effectiveness of our proposed method, we compared the classification accuracy % of synthetic frontal-face images using two datasets: Multi-PIE and CAS-PEAL, Table 1 and Table 2 compares the recognition rate %. We first extract deep features using Light-CNN, and compute the similarity of these features using a cosine-distance metric. This provides a baseline against which our results can be compared. We demonstrate that our method preserves texture information related to identity effectively when compared with TP-GAN. In particular, our method consistently yields superior results as compared to many state-of-the-art methods for both verification and identification of facial pose. Despite the fact that deep learning methods have been proposed to synthesize frontal images, none of these methods has proven to be highly effective for tasks such as recognition and verification. (Yu et al. 2020) The authors report that using a high resolution face image directly from a CNN will lower performance rather than boost it. Typically, larger faces provide less information, making preserving the synthesizer's identity more difficult. The performance of existing methods, such as those shown in Tables 1 and 2, drops significantly when face poses increase, while our method still exhibits compelling results. In addition, Table 3 illustrates the accelerated process at which our model performs faster without sacrificing recognition accuracy.

Table 1. Comparing our approach recognition rate (%) against various methods on multi-PIE dataset. Despite the large poses, rank-1 recognition was achieved in almost all of the face poses.

Methods Face pose	Our method	(Mahmood, Yea-Shuan, and Yi-An 2022)	(Xiaoguang et al. 2021)	(Yanfei and Junhua 2020)	(Rui et al. 2017)	(Xi et al. 2017)
$\pm 90^{o}$	74.30%	65.23%	70.20%	67.04%	64.03%	61.02%
$\pm 75^{o}$	90.78%	85.30%	85.31%	87.01%	84.10%	77.02%
$\pm 60^{o}$	95.96%	94.13%	91.81%	93.09%	92.93%	85.02%
$\pm 45^{o}$	99.54%	98.80%	98.05%	98.04%	98.58%	89.07%
$\pm 30^{o}$	99.88%	99.88%	99.82%	99.07%	99.85%	92.05%
$\pm 15^{o}$	99.87%	99.80%	99.83%	99.09%	99.78%	94.06%
Average ACC	93.38%	90.52%	90.83%	90.55%	89.87%	83.04%

Table 2. Comparing our approach recognition rate (%) against various methods on CAS-PEAL dataset. Despite the large poses, rank-1 recognition was achieved in almost all of the face poses.

Methods	Our method	(Rui et al. 2017)	(Yu et al. 2020)	(Tian et al. 2018)
Face pose Pitch at (-30^{0})				
15^{o}	99.86%	98.94%	83.91%	99.72%
30^{o}	99.55%	98.89%	83.17%	99.56%
45^{o}	98.89%	97.63%	80.83%	98.99%
Average ACC	99.43%	98.48%	82.63%	99.42%
Pitch at ($+45^{0}$)				
15^{o}	98.92%	97.73%	89.44%	98.98%
30^{o}	99.00%	97.45%	87.95%	98.86%
45^{o}	98.56%	95.83%	83.90%	98.13%
Average ACC	98.82%	97.00%	87.09%	98.65%

Table 3. A comparison of deep learning frameworks that focus on systemic details. Each deep learning model was tested on the same computer, but with different environment settings.

Evaluated item	TP-GAN (Rui et al. 2017) (TensorFlow 0.12v, Keras 2.0.1v)	Our method (TensorFlow 1.12v, Keras 2.2.5v)
Overall running time "number of epoch"	30 epochs = 2.20 Hrs. 1 = 279.96 (sec per-epoch)	30 epochs = 2.00 Hrs. 1 = 240.00 (sec per-epoch)
GPU load	$12-14GB$	$8-12GB$
Memory utilization	$74-86\%$ (step-per-epoch)	$67-77\%$ (step-per-epoch)

Improving Face Recognition with Data Augmentation

Despite the outstanding performance of decision forest, they still have some drawbacks when used alone to deal with a specific face pose problem. In TP-GAN, decision forest is capable of handling generated faces that have adequate samples, but struggle to handle low-quality faces with insufficient samples under extreme pose situations. As a result, the decision forest will lose some of its classification capabilities if there are only a few facial samples in the subsets of data. To eliminate these limitations, we employ a data augmentation strategy. We concentrated our attention on the facial images larger than a 60^{o} poses because they lacked fine details and inadequate facial samples, as illustrate in Table 4. As shown in Figure 11, the training consists of multiple steps to

Table 4. The amount of data augmentation that TP-GAN model synthesizes for each dataset. During the entire process, three datasets are used: training and testing.

Dataset types	Face poses	A total of all available face poses iamges before data augmentation	A total of all available face poses image after data augmentation
Multi-PIE	$\pm 60^0$ to $\pm 90^0$	2010	6707
CAS-PEAL	$\pm 60^0$ to $\pm 90^0$	10000	32430
FEI	$\pm 60^0$ to $\pm 90^0$	1800	5280

ensure high accuracy for each face poses as shown below:

1- Step-1: This step involves training our decision trees and forests (DF) with the Multi-PIE, FEI, and CASE-PEAL datasets, respectively. Our ideal is to train the model using different face pose samples, 60° to 90°, which provides a well-balanced pose features that can adapt and learn in response to changes in facial features and provide high classification accuracy for each pose.

2- Step-2: Utilize the TP-GAN model for data augmentation to increase the size of a face pose dataset by generating new synthetic data (Set 1) from it. In this step, TP-GAN is trained for each specific face pose such as 60°, 75° and 90°, respectively, so that it can be train with more balanced-pose synthesized face images to adapt and learn more facial features, which means large-pose faces will generate more face images than small-pose ones.

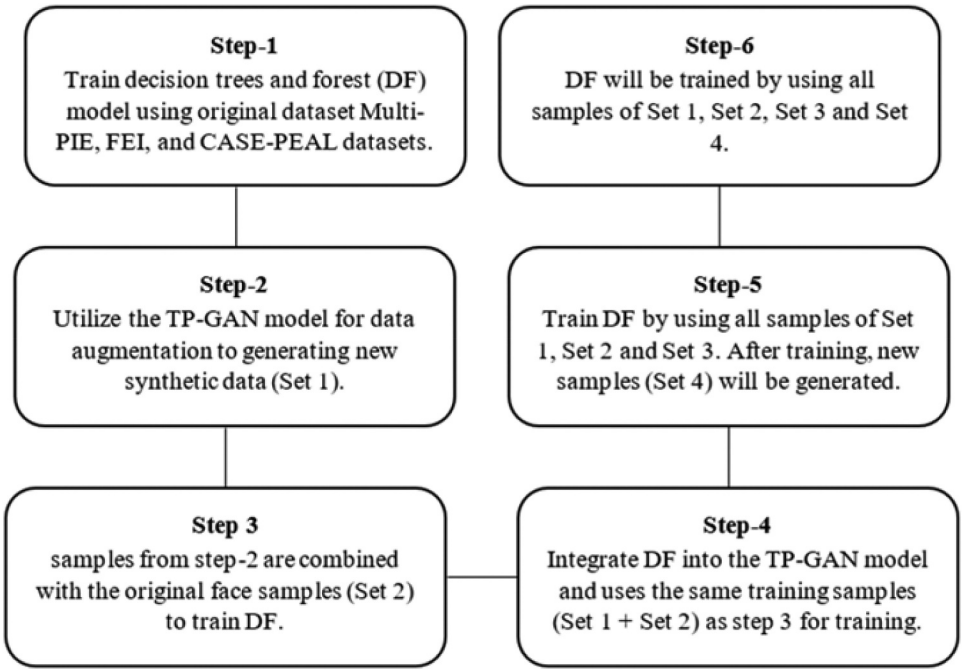

Figure 11. An overview of the training procedure steps for decision trees and forests.

3- Step 3: The generated face pose samples from step-2 are combined with the original face samples (Set 2) to train DF so that the model can adapt and learn to the change of facial features.

4- Step-4: Integrate DF into the TP-GAN model by replacing the final fully connected layer of the light-CNN discriminator DF, and train the model in an end-to-end manner. This step uses the same training samples (Set 1 + Set 2) as step 3 for training. After training, TP-GAN will generate new samples (Set 3) for every face poses.

5- Step-5: DF will be trained again by using all samples of Set 1, Set 2 and Set 3. After training, new samples (Set 4) will be generated by the TP-GAN generator. Here we would like to mentioned that the generated samples of both step 4 and step 5 are classified by a light CNN, only the correctly classified samples are collected, otherwise, they are discarded. By this design, the augmented data are clean and identity maintained.

6- Step-6: DF will be trained by using all samples of Set 1, Set 2, Set 3 and Set 4.

Our data augmentation model was built using the TP-GAN model in conjunction with a face detection and alignment method based on Multi-Task Cascaded Convolutional Networks or MTCNN (Kaipeng et al. 2016) that provides better classification features using face point

information. MTCNN has gained popularity in a broad range of situations, and is widely used in the detection and alignment of facial features. Regardless of the potential inconsistencies or over-smoothing that may occur due to factors beyond our control, our method nevertheless remains an effective method for incorporating pose information during the learning process.

Model Performance Based on the Loss Curve

Further analysis of our method is presented in this section in comparison with the TP-GAN method. In our analysis, we evaluate the impact of our model on two tradeoff parameters: generator loss and pixel loss, as shown in Figure 12. The loss performance of our method drops sharply when the number of epochs exceeds 130, while the TP-GAN method drops slightly. We calculated the optimization learning curve based on the metric we used to optimize the

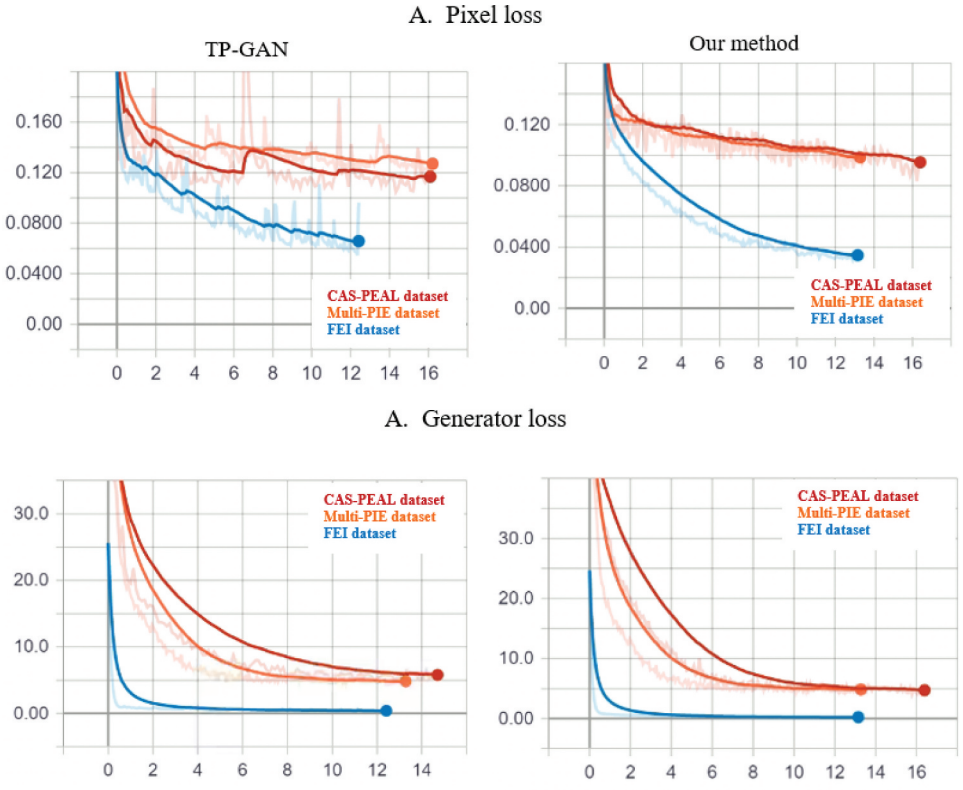

Figure 12. Plots of the TP-GAN and our method loss curves on the multi-PIE, FEI and CAS-PEAL datasets. Pixel loss is shown on A, while generator loss is shown on B. An axis in the horizontal direction indicates the number of epochs, or the number of times that all of the training data has been trained. The vertical axis indicates the accuracy of the model after each epoch; the smaller the loss, the better it performs.

model parameters, i.e., loss. As a result, our approach proved more effective and produced better outcomes than TP-GAN. Our method is easily replicated and can be applied to a variety of face recognition systems that are affected by unique face angles that directly affect detection accuracy. In addition, our approach is applicable across a wide range of datasets without difficulty or compromise the final results, even when faces are posed in extreme ways (as shown in Figures 3, Figure 4, Figure 5, Figure 6, Figures 7, Figures 8, Figure 9 and Figure 10), and its recognition performance is about 3.51% higher than that in (Rui et al. 2017), 2.83% higher than the result obtained from (Yanfei and Junhua 2020), and around 2.89% higher than the result obtained by (Xiaoguang et al. 2021), and finally 2.86% higher that the result achieved by (Mahmood, Yea-Shuan, and Yi-An 2022) on Multi-PIE dataset, as shown in Table 1.

Results and Discussions

The goal is to encourage the discriminator network to push the generator to produce more accurate faces and texture details each time it fails to fool the discriminator, thus resulting in reject samples with high confidence. Such claims are the result of extreme face poses. The TP-GAN relies heavily on the generator alone for its face image synthetization, which is one of the main reasons why it was able to better preserve facial features in updated, synthesized views. The statement holds true when a dataset is considered to be under constrained conditions, such as the Multi-PIE. It is important to keep in mind that not all datasets have the same unique constrained environment, face shape characteristics, illumination conditions, and etc., such as FEI and CAS-PEAL. Experiments indicate that using such datasets can be difficult and produce counterproductive and harmful results, rather than boosting the accuracy of the final face recognition. The limitations outlined above can be overcome by using careful design methods. In the first step, we integrate the decision forest with careful analysis to have our discriminator work more closely with our generator for better feature representations and identity-preserving inferences. In this way, we are able to eliminate images that produce less desirable results and train the network simultaneously to achieve high-level face identification without measuring or calculating errors between D and G. Decision forest require no extensive modification of D, nor do they require the inclusion of a large external neural network, which would complicate the training process. As a second step, we enlarge the training dataset using a data augmentation technique in order to diminish the problem of low-face samples with extreme poses present in Multi-PIE, FEI and CAS-PEAL datasets.

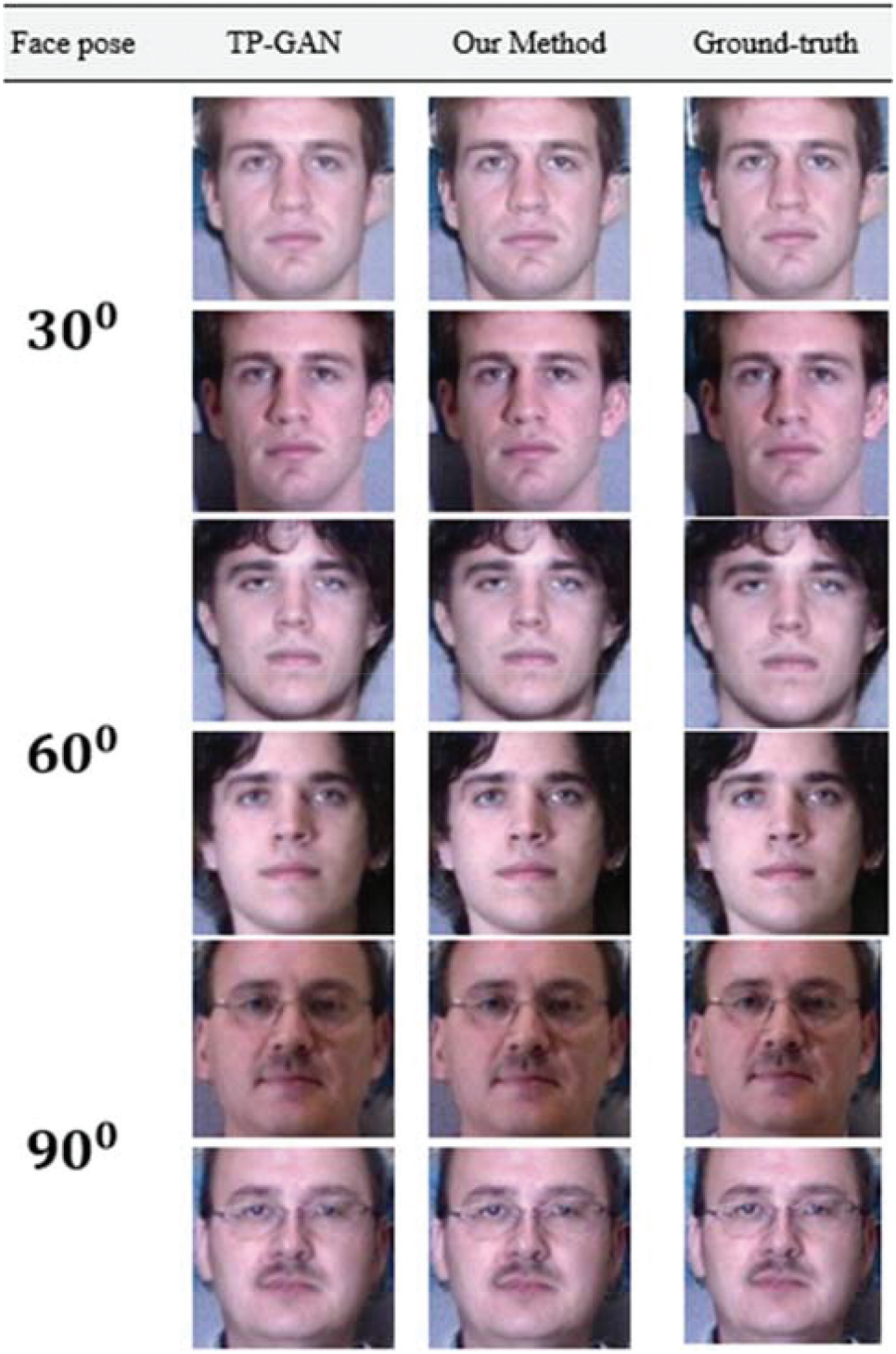

Figure 13. Examples of illumination levels on the multi-PIE dataset. For instance, illumination is a combination of brightness, exposure, contrast, and shadows. Various effects of quality can also be observed, including sharpness, smoothness, and blurriness. Overall, those qualities can contribute to a low level of face recognition. We downloaded the dataset from the TP-GAN GitHub repository at https://github.com/HRLTY/TP-GAN.

Lastly, upgrade and modify existing deep learning frameworks such as TensorFlow and Keras to improve performance and reduce information acquisition, and build a high-performance model with faster running times, lower GPU loads, and better memory utilization, requiring less advanced hardware and accelerating learning process. The loss function has been modified and some parts of the kernel layers have been removed. These layers are used to generate 32×32 and 64×64 image pixels, and we realized that their usage had two major impacts, including slowing down the training process and increasing design complexity. Therefore, we devoted our efforts to 128×128 pixels, which we found to be the most advantageous. Our achieved result is shown in Figure 6, Figures 7, Figures 8, Figure 9 and Figure 10 compared to state-of-the-art methods. Moreover, we have included additional results based on illumination conditions. The benefits provided by our method will be further supported by this evidence, as shown in Figure 13. This is followed by a different type of facial recognition method Table 1 and Table 2. Overall, we attain a $rank-1$ performance rate (%).

Conclusion

In this paper, we propose a sample adaptive strategies to improve the TP-GAN method for unconstrained pose-invariant face recognition. Our goal was accomplished by designing tasks that can be integrated into the TP-GAN without requiring the existing algorithm to be expanded. In fact, we reduced the existing complexity of the model while ensuring high face recognition accuracy. Comparing our method to TP-GAN, we are capable of generating frontal images with superior texture details while conserving identity information. Combining decision forest allows us to strengthen our discriminator and make it more cooperative with the generator, producing better synthetic frontal images with rich texture details and higher classification accuracy. The use of data augmentation enabled us to improve the accuracy of TP-GAN model in categorizing images and further improve their classification strength in comparison with traditional augmentation, which required considerable computational resources. Additionally, the enhanced deep learning framework allows us to create highly accurate model that require less powerful hardware and accelerate the learning process. Multi-PIE, FEI, and CAS-PEAL results indicate that our method produces superior perceptual facial images over TP-GAN results in extreme poses. We would like to summarize our work here. Using decision trees forests to model complex real-world problems has earned them a reputation for being able to handle non-linear and high-dimensional data. A method such as this

enables features to be jointly learned through linear nodes, in an end-to-end trainable manner, which can boost the classifiers capabilities. Data augmentation aims to increase the quantity and quality of face samples to allow the model to learn more features each time it is trained. Upgrades and modifications to the existing framework speed up the training process without sacrificing the model's performance. Despite the good results achieved by our method, we believe that other optimization algorithms or different facial analysis and recognition techniques can still be used to improve it further. Future research will incorporate multiple face-analysis techniques into two pathways structures, culminating in a highly precise and super-resolution generative model.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Mahmood HB Alhlffee http://orcid.org/0000-0001-6640-0582

References

Alvaro-H, C., P. Robert 2020. Joints in random forests. Conference on Neural Information Processing Systems. Vancouver, Canada: 1–12. doi: 10.48550/arXiv.2006.14937.

Bassel, Z., K. Ilya, and M. Yuri. 2021. *PFA-GAN: Pose face augmentation based on generative adversarial network* 425–40. doi: 10.15388/21-INFOR443.

Changxing, D., and T. Dacheng. 2017. Pose-invariant face recognition with homography-based normalization. *Pattern Recognition Journal* 66:144–52. doi:10.1016/j.patcog.2016.11.024.

Chao, Y., W. Ziqi, M. Shiwen 2022. Rfpose-GAN: Data augmentation for RFID based 3D human pose tracking. IEEE 12th International Conference on RFID Technology and Applications (RFID-TA). Cagliari, Italy: 1–4. DOI: 10.1109/RFID-TA54958.2022.9924133.

Christian, L., T. Lucas, H. Ferenc, C. Jose, C. Andrew, A. Alejandro, A. Andrew, T. Alykhan, T. Johannes, W. Zehan, S. Wenzhe 2017. Photo-realistic single image super-resolution using a generative adversarial network. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). HI, USA: 105–14. doi: 10.1109/CVPR.2017.19.

Duc My, V., N. Duc Manh, L. Thao Phuong, and L. Sang-Woong. 2021. *HI-GAN: A hierarchical generative adversarial network for blind denoising of real photographs* 225–50. doi: 10.1016/j.ins.2021.04.045.

Eric, R. -C., Z. -L. Connor, A. -C. Matthew, N. Koki, P. Boxiao, M. Shalini De, G. Orazio, G. Leonidas, T. Jonathan, K. Sameh, K. Tero, W. Gordon 2022. Efficient geometry-aware 3D generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition: 16123–33. 10.48550/arXiv.2112.07945.

Fariborz, T., T. Veeru, D. Jeremy, C. -V. Matthew, and M. -N. Nasser. 2021. *Profile to frontal face recognition in the wild using coupled conditional generative adversarial network* 149–61. doi: 10.48550/arXiv.2107.13742.

FEI2005-2006*Artificial intelligence laboratory of FEI*São Bernardo do Campo, Brazildoihttps://fei.edu.br/~cet/facedatabase.html

Ian, J. G., P. -A. Jean, M. Mehdi, X. Bing, W. -F. David, O. Sherjil, C. Aaron, B. Yoshua. 2014. Generative adversarial networks. International Conference on Neural Information Processing Systems: 1–9. doi: 10.48550/arXiv.1406.2661.

Junbo, Z., M. Michael, L. Yann 2016. Energy-based generative adversarial network. The 5th International Conference on Learning Representations: 1–17 doi: 10.48550/arXiv.1609.03126.

Junho, Y., J. Heechul, Y. ByungIn, C. Changkyu, P. Dusik, K. Junmo 2015. Rotating your face using multi-task deep neural network. IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: 676–84. doi: 10.1109/CVPR.2015.7298667.

Kaipeng, Z., Z. Zhanpeng, L. Zhifeng, Q. Yu 2016. Joint face detection and alignment using multi-task cascaded convolutional networks. Journal of IEEE Signal Processing Letters: 1499–503. doi: 10.1109/LSP.2016.2603342.

Luan, T., Y. Xi, L. Xiaoming 2017. Disentangled representation learning gan for pose-invariant face recognition. IEEE Conference on Computer Vision and Pattern Recognition, (CVPR). Honolulu, USA: 1415–24. doi: 10.1109/CVPR.2017.141.

Mahmood, A., H. Yea-Shuan, and C. Yi-An. 2022. 2D facial landmark localization method for multi-view face synthesis image using a two-pathway generative adversarial network approach. *Journal of PeerJ Computer Science* 1–28. doi:10.7717/peerj-cs.897.

Martin, A., C. Soumith, and B. Léon. 2017. Wasserstein GAN. 1–31. doi:10.48550/arXiv.1701.07875.

Ma, R., Y. Xiao, T. Uno, L. Ma, K. Khorasani. 2021. A new facial expression recognition system using decision tree and deep neural networks. IEEE 6th International Conference on Signal and Image Processing (ICSIP). Nanjing, China: 49–54. doi: 10.1109/ICSIP52628.2021.9688871.

Mehdi, M., and O. Simon. 2014. *Conditional generative adversarial nets* 676–84. doi: 10.48550/arXiv.1411.1784.

Ngoc-Trung, T., T. Viet-Hung, N. Ngoc-Bao, N. Trung-Kien, C. Ngai-Man 2021. On data augmentation for GAN training. IEEE Transactions on Image Processing: 1882–97. doi: 10.1109/TIP.2021.3049346.

Nitish, S., H. Geoffrey, K. Alex, S. Ilya, S. Ruslan 2014. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research: 1929–85. doi: 10.5555/2627435.2670313.

Omar, E., A. Noor, A. Somaya, and K. Fouad. 2022. Pose-invariant face recognition with multitask cascade networks. *Journal of Neural Computing and Applications* 34:6039–52. doi:10.1007/s00521-021-06690-4.

Peter, K., F. Madalina, C. Antonio, R. Samuel 2015. Deep neural decision forests. IEEE International Conference on Computer Vision (ICCV). Santiago, Chile:1467–75. doi: 10.1109/ICCV.2015.172.

Puja, P., G. Vinit Kumar 2020. Feature descriptors for face recognition. Conference. The IEEE 17th India Council International Conference (INDICON): 1–4. doi: 10.1109/INDICON49873.2020.9342424.

Qingyan, D., Z. Lei, Z. Yan, and G. Xinbo. 2023. PSGAN: Revisit the binary discriminator and an alternative for face frontalization. *Journal of Neurocomputing* 518:360–72. doi:10.1016/j.neucom.2022.11.033.

Ralph, M., C. Jeffrey, K. Takeo, B. Simon, and S. Baker. 2010. Multi-PIE. *Image and Vision Computing* 28:ls. 5. 1–23. doi:10.1016/j.imavis.2009.08.002.

Rich, C., N. -M. Alexandru 2006. An empirical comparison of supervised learning algorithms. In Proceedings of the 23rd international conference on Machine learning, Pittsburgh, PA, USA: 161–68. doi: 10.1145/1143844.1143865.

Rui, H., Z. Shu, L. Tianyu, H. Ran 2017. Beyond face rotation: Global and local perception GAN for photorealistic and identity preserving frontal view synthesis. In: IEEE International Conference on Computer Vision (ICCV). Venice, Italy: 2439–48. doi: 10.1109/ICCV.2017.267.

Samuel Rota, B., K. Peter 2014. Neural decision forests for semantic image labelling. IEEE Conference on Computer Vision and Pattern Recognition: Columbus, OH, USA. 1–7. doi: 10.1109/CVPR.2014.18.

Shuhui, J., M. Haiyi, D. Zhengming, and F. Yun. 2020. Deep decision tree transfer boosting. *IEEE Transactions on Neural Networks and Learning Systems* 31 (2):383–95. doi:10.1109/TNNLS.2019.2901273.

Sreerama, K. -M., K. Simon, and S. Steven. 1994. *A system for induction of oblique decision trees* 1–32. doi: 10.1613/jair.63.

Subhajit, C., H. Debapriya, B. Yung-Cheol, and K. Yong-Woon. 2022. *Enhancement of image classification using transfer learning and GAN-based synthetic data augmentation* 1–16. doi: 10.3390/math10091541.

Tal, H., H. Shai, P. Eran, E. Roee 2015. Effective face frontalization in unconstrained images. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: 1–10. doi: 10.1109/CVPR.2015.7299058.

Tero, K., A. Miika, L. Samuli, H. Erik, H. Janne, L. Jaakko, L. Jaakko 2021. Alias-free generative adversarial networks. Advances in Neural Information Processing Systems. The 35th Conference on Neural Information Processing Systems: 1–12 doi: 10.48550/arXiv.2106.12423.

Tian, Y., X. Peng, L. Zhao, S. Zhang, D. -N. Metaxas 2018. CR-GAN: Learning complete representations for multi-view generation. IJCAI International Joint Conference on Artificial Intelligence: 1–7. doi: 10.48550/arXiv.1806.11191.

Tongyu, L., F. Ju, L. Yinqing, T. Nan, L. Guoliang, and D. Xiaoyong. 2021. Adaptive data augmentation for supervised learning over missing data. *Proceedings of the VLDB Endowment* 14 (7):1–13. doi:10.14778/3450980.3450989.

Wen, G., B. -C. Senior, S. Shiguang, C. Xilin, Z. Delong, Z. Xiaohua, and Z. Debin. 2009. The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man and Cybernetics Part A, Systems and Humans* 38 (1):149–61. doi:10.1109/TSMCA.2007.909557.

Xiaoguang, T., Z. Jian, L. Qiankun, A. Wenjie, G. Guodong, L. Zhifeng, L. Wei, and F. Jiashi. 2021. Joint face image restoration and frontalization for recognition. *IEEE Transactions on Circuits and Systems for Video Technology* 32 (3):1–14. doi:10.1109/TCSVT.2021.3078517.

Xiaoguang, T., G. Jingjing, X. Mei, Q. Jin, and M. Zheng. 2017. Illumination normalization based on correction of large-scale components for face recognition. *Journal of Neurocomputing* 266:465–76. doi:10.1016/j.neucom.2017.05.055.

Xi, Y., Y. Xiang, S. Kihyuk, L. Xiaoming, C. Manmohan 2017. Towards large-pose face frontalization in the wild. IEEE International Conference on Computer Vision (ICCV). Venice, Italy: 1–10. doi: 10.1109/ICCV.2017.430.

Yanfei, L., and C. Junhua. 2020. Unsupervised face frontalization for pose-invariant face recognition. *Image and Vision Computing* 106:1–9. doi:10.1016/j.imavis.2020.104093.

Yan, Z., A. Gil, and D. Tom. 2018. *Generative adversarial forests for better conditioned adversarial learning* 1–17. doi: 10.1080/1805.05185.

Yan, Z., A. Gil, D. Tom 2021. Improved training of generative adversarial networks using decision forests. IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA: 3492–501. doi: 10.1109/WACV48630.2021.00353.

Yani, I., R. Duncan, Z. Darko, K. Peter, S. Jamie, B. Matthew, and C. Antonio. 2016. Decision forests, convolutional networks and the models in-between. *Computer Vision and Pattern Recognition* 1–9. doi:10.48550/arXiv.1603.01250.

Yi, Z., F. Keren, H. Cong, and C. Peng. 2021. Identity-and-pose-guided generative adversarial network for face rotation. *Journal of Neurocomputing* 33–47:33–47. doi:10.1016/j.neucom.2021.04.007.

Yongxin, Y., G. Irene, and H. Timothy. 2018. *Deep Neural Decision Tree: 1467–1475*. doi:10.1109/1806.06988.

Young, H., K. Byung-Gyu, and R. Partha-Pratim. 2021. Face generation algorithm from multi-view images based on generative frontal adversarial network. *Journal of Multimedia Information System* 8 (2):85–92. doi:10.33851/JMIS.2021.8.2.85.

Yu, Y., J. Songyao, R. Joseph, F. Yun 2020. Dual-attention GAN for large-pose face frontalization. IEEE International Conference on Automatic Face and Gesture Recognition. Buenos Aires, Argentina: 1–8. doi: 10.1109/FG47880.2020.00004.

Zhenyao, Z., L. Ping, W. Xiaogang, and T. Xiaoou. 2014. Multi-view perceptron: A deep model for learning face identity and view representations. *Advances in Neural Information Processing Systems* 217–25.

Ziyi, S., L. Wei-Sheng, X. Tingfa, K. Jan, Y. Ming-Hsuan 2018. Deep semantic face deblurring. IEEE/CVF Conference of Compute Vision and Pattern Recognition. Salt Lake City, UT, USA: 8260–69. doi: 10.1109/CVPR.2018.00862.