Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Frontal Face GAN: Generating Pose-Corrected Human Faces

Ayush R. Barai[1], Santosh G. Giram[2], Pratik N. Salve[3], Harshad H. Pawar[4], Omprakash K.Parihar[5]

[1,2,3,4,5]Dept. of Computer Science and Design K. K. Wagh Institute of Engineering Education and Research,Nashik.

Abstract

Generating identity-preserving frontal face images from unconstrained real-world profile views represent a fascinating problem in computer vision, the solution of which is critical for enhancing downstream tasks such as face recognition and attribute analysis. Current approaches often struggle with balancing the realistic requirements of 3D geometrical fidelity against real-world-nuisance variations such as uncontrolled light, occlusions, and heterogeneous backgrounds. This paper presents a novel two-stage transfer learning approach for training a Generative Adversarial Network (GAN) for the task. Our model, which follows the Pix2pix architecture with a U-Net generator and a PatchGAN discriminator, is first pre-trained on the CMU Multi-PIE dataset, which is controlled and makes the model learn the necessary 3D-aware geometric transformations from facial rotation. Thereafter, the pre-trained model is fine-tuned on realistic difficulties presented by the CFPW dataset. In the interest of practical application, our system integrates an MTCNN face detector for the automated pre-processing of ’wild’ images. According to our study, this two-phase training setting turns out to be very powerful, yielding a robust model capable of generating high-fidelity geometrically correct frontal images from unconstrained challengingly posed profile images.

Keywords: {(Face Frontalization, Encoder, Decoder, Generator, Discriminator,Generative Adversarial Networks (GANs))}

1. Introduction

The training data for facial recognition technology cover a lot of aspects that make it so prevalent in such modern applications as biometric identification, surveillance, human-computer interaction, and digital forensics-greatly increasing the possibility of accidents. One of the principal problems, though, in face recognition systems is that of pose variation-affixing a subject’s face to non-frontal angles. These variations result in a loss of geometric and texture information, which diminishes recognition accuracy and reduces the reliability of the systems themselves.

Recent advancements in deep learning, especially the initiation of Generative Adversarial Networks (GANs), have thrown new light on the option of realizing image creation and transformation at levels hitherto inconceivable. The framework for a GAN comprises two neural networks: a Generator that synthesizes new images and a Discriminator that evaluates their authenticity. Through a process of competitive adversarial learning, the generator progressively improves its outcome until the realization of high-fidelity images impossible to distinguish from real data.

This proposed project is titled “Frontal Face GAN: Generating Pose-Corrected Human Faces”. It features a deep learning-based algorithm that generates realistic frontal facial views from side-profile inputs. A U-Net-based Generator is used here to construct the frontal facial structures with identity preservation, while a PatchGAN Discriminator makes sure that photorealism at texture level and geometric consistency are maintained.

A significant enhancement of this work, however, is found under the two-stage transfer learning strategy. The model is first pre-trained on the CMU Multi-PIE dataset, which contains organized 3D pose variations shot under controlled lighting; thus the GAN can learn the appropriate way to correctly represent face geometry. Then fine-tuning would occur on the CFPW (Celebrities in Frontal-Profile in the Wild) data set, which introduces real variations in illumination, occlusions, and diverse backgrounds. This hybrid training concept enables model generalization over both constrained and unconstrained environments.

In this way, by utilizing pose-corrected photorealistic frontal images for input, the Frontal Face GAN system shall significantly improve existing face recognition pipelines in terms of robustness and fidelity. In fact, by such preprocessing, recognition rates may see a substantial increase in practice. Applications include security systems, access control, and digital identity verification, along with photo restoration.

Thus, the proposed work will eventually contribute to combining computer vision, deep learning, and generative modeling into an intelligent framework that could understand and reconstruct human facial appearances. The scope of transformation that GANs hold in solving pose variation problems and setting a benchmark in high-fidelity identity-preserving facial synthesis is enormous.

2. LITERATURE REVIEW

Wasseem N. Ibrahem Al-Obaydy, Shahrel Azmin Suandi: proposed “Automatic pose normalization for open-set single-sample face recognition in video surveillance” [1], a model in which the large number of poses that are captured by video surveillance cameras, where the face recognition systems are often degraded by these images. Currently, there are methods that can normalize the poses prior to recognition. However, these methods require manual facial landmarks annotation. In this work, the authors present a method that can automatically perform facial landmarks detection and pose normalization without requiring manual intervention. The method is based on a thin-plate splines warping algorithm and a facial landmark detection algorithm. The resultant 2D surface points are used to transform the facial images into a desired frontal pose. Results of the experiments conducted on the FERET database revealed that the proposed method has leveraged a better performance.

ChaorongLi, WeiHuang, YuanyuanHuang authors proposed "Gabor Log-Euclidean Gaussian and its fusion with deep network based on self-attention for face recognition", [2] a method to extract facial features from a multivariate Gaussian with the help of a whitening principal component analysis (WPCA). The method is mainly used for extracting the raw features from the various components of the Gaussian. Due to the nature of the space of the multivariate Gaussian, it is difficult to incorporate its learning mechanism into the model. In this paper, we present a method that combines the log-linear approach and the multi-component analysis of the multivariate Gaussian. The result of this method is that it is able to extract discriminative facial features.

Tianyang Cao, Chang Liu,and Jiamin Chen proposed "Nonfrontal Expression Recognition in the Wild Based on PRNet Frontalization and Muscle Feature Strengthening" [3] a model in which the ability to recognize non-frontal facial expressions in the wild is very important for developing artificial intelligence and human machine interaction. In this paper, we introduce a face reconstruction method called PRNet that can build 3D frontal faces for 2D head photos. Unfortunately, it is still hard to recognize non frontal facial expressions due to the effects of frontalization. In addition, the muscle parameters of the frontalization face are weaker than those of the real one. This paper proposes a method that combines the 3D facial contours and the Frechet distance to build a stronger and more complete facial expression. Through this process, the muscles are reinforced by moving directions from each muscle area.

DanielSaezTriguerosaLiMengaMargaretHartnettb proposed "Ǵenerating photo-realistic training data to improve face recognition accuracy" [4]. Due to the high accuracy of face recognition systems, developed by CNNs has become a widely used biometric in law enforcement and forensics. However, to achieve a best performance, these systems need to be trained with large datasets. A novel approach to enhance face datasets by training a generative adversarial network will identify non identifiable attributes from the identity-related ones. This method is implemented by mapping discrete identity labels to a space known as identity latent space. They presented a novel approach to train face recognition systems by generating both synthetic images of the subjects and new images of the unknown subjects. Through the use of GAN training, we show that the models can produce realistic images to improve the system accuracy.

Guoli Wang, Jiaqi Ma, Qian Zhang, Jiwen Lu, Jie Zhou proposed "Pseudo Facial Generation with Extreme Poses for Face Recognition, CVPR 2021 open access" [5]. Despite the success of face recognition in recent years, it is still hard to identify the facial images with extreme poses. Traditional methods try to generate fake frontal face from the input image, which is very hard to maintain and it also consumes a lot of computational energy. An experimental study shows that the precision of identifying the faces with extreme poses drops dramatically after making small changes in the image. This helps to believe that there is a way to improve accuracy without changing the discriminators. A novel method has been proposed to generate lightweight pseudo facial images, which can be used to identify the faces with extreme poses. It can also depict the facial contours and modify the information to preserve the identity. The proposed method achieves this by minimizing the difference between the original and pseudo profile.

Zeno Bassel, Kalinovskiy Ilya, Matveev Yuri proposed "PFA-GAN: Pose Face Augmentation Based on Generative Adversarial Network" [6]. A novel framework known as PCA-GAN is proposed to enable a pose-based augmentation of a face image by considering the driving face and its identity. It can be trained in a fully self-supervised manner. In [6], the proposed framework can be used to expand the training set by introducing various augmentation strategies. The results of the experiments show that the framework can perform well in terms of face verification.

Cong Hu, Zhenhua Feng, Xiaojun Wu, Josef Kittler proposed "Dual Encoder-Decoder Based Generative Adversarial Networks for Disentangled Facial Representation Learning" [7], both the generator and discriminator aredesigned with deep encoder-decoder architectures. The purpose of the encoder-decoder structure generator is to learn a face representation that is disentangled from its original state. The discriminator is then tasked with performing various face reconstruction and classification tasks. We propose a network architecture that avoids the additional loss of the Wasserstein distance due to the discriminator's output. In addition, we consider face pose variation as continuous, which can be used to inject richer information into our model. The proposed network is designed to perform a pose estimation task that disentangles identity information from variations in the face. It is evaluated on the various face recognition and face synthesis tasks. The results of the study revealed that the proposed network outperforms the existing state-of-the-art methods.

Douglas M. Souza, Duncan D. Ruiz proposed "Towards High-Resolution Face Pose Synthesis" [8], the authors present a novel approach to synthesize different face views by using GANs. This method can be used in various computer graphics and vision applications. The method involves controlling the rotation of the facial poses along the three axes of space. We start by estimating their pose and storing a vector that contains the rotation angles. We then train a state of-the-art GAN by combining the images and angles. The results show that the synthesis method can achieve a high-quality output and control the pose of the images.

Hong Li; Nanfeng Xiao proposed "Face Rotation and Recognition Based on Attention Mechanism and Generative Adversarial Networks" [9]. Due to the increasing complexity of facial data, face rotation can improve the performance of face recognition. However, most of the previous approaches mainly focus on the information related to the identity. This paper proposes a novel approach that combines face rotation with a framework for face recognition. The proposed model [9] is composed of a heatmap-based generator and dual discriminators. The latter focuses on preserving the structure and critical facial organs during face rotation. The proposed model combines the features of face rotation with an expression learning framework to improve the performance of facial recognition. It also features attribute purification and dynamic loss weights.

Anil Celik, Nafiz Arica, enhance the face pose normalization with deep learning [10]. The paper proposes a hybrid method by combining the 3-D model based face pose normalization with the SDAE deep network. It is performed through three consecutive stages. After the facial landmark points have been aligned, the next step is to feed 2 D images into the 3D pose normalization stage. This step involves estimating the pose generation and fine-tuning the resultant image.

3. METHODOLOGY USED

The proposed methodology for face frontalization is structured around a Generative Adversarial Network (GAN) and is systematically organized into five key modules. The major elements of the system are described below.

1) Data Pre-Processing: The primary and most critical module is Data Pre-Processing, which readies the raw images for training. The model utilizes the CFPW (Celebrities in Frontal-Profile in the Wild) dataset sourced from Kaggle. In order for the program to access this dataset, it must be organized in a specific format: a separate folder is established for each individual, and within that folder, there should be 10 frontal-view images and 4 side-view (profile) images of the same person. The technical procedure for loading this data begins by employing Python's `os` module to inspect the parent directory of the dataset and compile a list of all individual subject folders. Once the images of a subject are loaded, they undergo a crucial pre-processing stage: each image is resized to a uniform dimension of 256 × 256 pixels using the `cv2.resize` function. This standardization is essential for the neural network, which requires that all inputs are of identical size. Ultimately, each resized image is appended to a source list, which is then converted into a single, large NumPy array, preparing it for input into the model for training.

2) The Discriminator Network: The Discriminator functions as the primary of the two critical components of the GAN. Its primary role is to act as a binary classifier, or a "critic," responsible for distinguishing between two categories of images: authentic frontal faces from the dataset and artificially created "fake" frontal faces generated by the Generator. In this specific architecture, the Discriminator also performs dimensionality reduction. The network is designed to accept a 256 × 256 3-channel (color) image as input. This image is then processed through a deep sequence of Conv2D (Convolutional) layers. As the image advances deeper into the network, its spatial dimensions are progressively reduced, while the number of feature channels is increased. This process effectively "funnels" the image from its original 256 × 256 size down to a compact 16 × 16 feature map. This reduction forces the network to develop a comprehensive and in-depth set of features for high-accuracy image recognition. The model utilizes the LeakyReLU activation function to ensure that the gradient flow remains active and does not stagnate or vanish when encountering small or negative input values.

3) The Generator Network: The Generator serves as the second essential component, designed to operate in direct opposition to the Discriminator. Its primary function is to produce the final output image. It takes the low-dimensional 16 × 16 feature map and "reforms" all the layers, upsampling the data back to a complete 256 × 256 image. This generated image is subsequently submitted to the Discriminator for evaluation. The architecture of the Generator is structured as an Encoder-Decoder model. The Encoder section comprises down-sampling layers that diminish the "signal rate" (akin to the Discriminator), while the Decoder section includes up-sampling layers that enhance the signal rate to reconstruct the full-sized image. Within this framework, the Generator's ultimate objective is to become proficient at generating realistic frontal faces to the extent that the Discriminator is "fooled" into misclassifying its synthetic images as authentic.

4) Model Training: With both the Generator and Discriminator established, they are integrated to form the complete GAN for the training process. This training is characterized as an adversarial "game" consisting of three primary steps. Initially, the `define_discriminator()` function is invoked to construct the critic network (D model). Subsequently, the `define_generator()` function is executed to develop the synthesis network (G model). Thirdly, and most crucially, a composite GAN model is formed, which links the Generator and Discriminator together. In this integrated model, the Discriminator's weights are fixed (rendered non-trainable). This aspect is fundamental to GAN training: during the training of the GAN model, the error signal is backpropagated through the frozen Discriminator, ensuring that only the weights associated with the Generator are updated. This compels the Generator to evolve and improve its ability to deceive the static Discriminator. The entire procedure is initiated by invoking the `train()` function, which oversees this alternating training process across numerous epochs.

5) Model Testing and Inference: The Discriminator serves as the primary component among the two essential elements of the GAN. Its main function is to operate as a binary classifier, or a "critic," tasked with differentiating between two types of images: genuine frontal faces from the dataset and artificially generated "fake" frontal faces produced by the Generator. Within this particular architecture, the Discriminator also undertakes dimensionality reduction. The network is structured to accept a 256 × 256 3-channel (color) image as input. This image is subsequently processed through a deep series of Conv2D (Convolutional) layers. As the image progresses deeper into the network, its spatial dimensions are gradually diminished, while the number of feature channels is augmented. This procedure effectively "funnels" the image from its initial 256 × 256 size down to a compact 16 × 16 feature map. This reduction compels the network to cultivate a thorough and detailed set of features for high-accuracy image recognition. The model employs the LeakyReLU activation function to guarantee that the gradient flow remains active and does not stagnate or disappear when faced with small or negative input values.

4. **PROPOSED SYSTEM A.**

System Architecture

The proposed framework for face frontalization is founded on a Generative Adversarial Network (GAN) that comprises two main components: a Generator and a Discriminator. The Generator is tasked with transforming a given profile image into its corresponding frontal face image, whereas the Discriminator's role is to distinguish between authentic and generated frontal faces. The overall architecture of both networks is illustrated in The Generator adopts an encoder–decoder architecture that incorporates skip connections to preserve spatial and textural information throughout the transformation process. The encoder systematically compresses the input profile image through multiple convolutional layers (Conv) to extract hierarchical features. Each convolutional layer is succeeded by Batch Normalization (BN) and LeakyReLU activation functions to ensure stability during training and to introduce non-linearity. As the encoding progresses, the spatial resolution diminishes while the depth of feature maps increases, allowing the model to learn intricate facial attributes. Skip connections are integrated between corresponding encoder and decoder layers to maintain low-level spatial information that could otherwise be lost during downsampling. The decoder reconstructs the frontal face image from the latent representation generated by the encoder. It

Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

employs transposed convolutional (UpConv) layers to upsample the encoded features back to the original resolution. Each up-convolution layer is also followed by BN and LeakyReLU activations, akin to the encoder. The skip connections combine encoder features with decoder layers, assisting the model in recovering high-frequency details and producing a structurally coherent output. The final layer of the decoder utilizes a Tanh activation function to create the frontal face image within the normalized range $[-1, 1]$. The resulting output maintains a spatial dimension of 128imes128 pixels, which is identical to the size of the input image. The Discriminator functions as a binary classifier that assesses whether a given frontal image is authentic or artificially created. It takes in concatenated pairs of real and generated images and processes them through a sequence of convolutional blocks. Each block comprises a Conv–BN–LeakyReLU sequence designed to extract distinguishing features and progressively downsample the input. The architecture yields an 8imes8imes1 output patch rather than a single scalar value, thereby establishing a PatchGAN discriminator. This patch-based assessment enables the network to concentrate on local features and textures, ensuring that both the overall facial structure and intricatedetails appear convincingly realistic.

The training procedure incorporates two categories of loss functions: Adversarial Loss and L1 Loss. The Adversarial Loss motivates the Generator to create realistic frontal faces capable of misleading the Discriminator, while the Discriminator is trained to accurately differentiate between real and fake samples. Concurrently, the L1 Loss calculates the pixel-wise discrepancy between the generated image and the corresponding ground truth frontal image, assisting the model in maintaining facial identity and structural precision. The integration of these two loss functions empowers the network to produce photorealistic, identity-preserving frontal faces with enhanced visual fidelity.

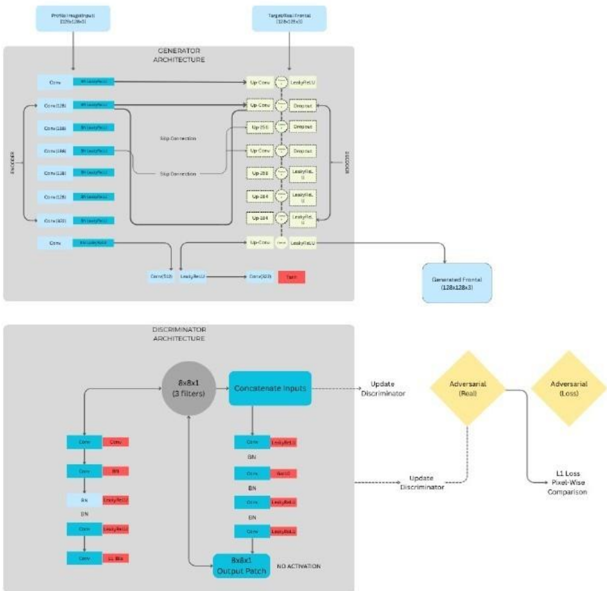

Fig. 1: System Flow Diagram of the Proposed GAN-based Face Frontalization Model

B. Performance Parameters

1) Accuracy of Generated Images: The main performance metric for this model is the visual and quantitative accuracy of the produced frontal images. The evaluation of how closely the generated outputs (G(y)) match the ground truth frontal images (x) is conducted using the Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR).
PSNR is derived from the Mean Squared Error (MSE) and is defined as:

$$PSNR = 10 \cdot \log_{10} \frac{MAX_I^2}{MSE}$$

where MAX_I is the maximum possible pixel value (e.g., 1.0 for normalized images) and MSE is

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [x(i,j) - G(y)(i,j)]^2$$

SSIM is a perceptual metric that quantifies image qualitydegradation as a change in structural information:

$$SSIM(x, G(y)) = \frac{(2\mu_x \mu_{G(y)} + c_1)(2\sigma_{xG(y)} + c_2)}{(\mu_x^2 + \mu_{G(y)}^2 + c_1)(\sigma_x^2 + \sigma_{G(y)}^2 + c_2)}$$

where µ is the mean, σ2 is the variance, and $\sigma_{xG(y)}$ isthe covariance of the two images.

2) Loss Function Evaluation: Both the Generator (G)and Discriminator (D) performance is monitored through adversarial loss functions. The standard GAN "minimax"objective is:

$$\min_G \max_D V(D,G) = E_{x \sim p_{real}}[\log D(x)] + E_{y \sim p_{side}}[\log(1 - D(G(y)))] \quad (1)$$

where x is the real frontal image and y is the side-profile input.
The Generator's objective is to minimize both this adversarial loss and a reconstruction loss (such as L1 loss), which enforces that the generated image is a plausible translation of the input:

$$L_{L1}(G) = E_{x,y}[\|x - G(y)\|_1]$$

The Generator's total loss is a weighted sum: $L_G = L_{GAN}(G) + \lambda L_{L1}(G)$. A lower generator loss combined with a stable discriminator loss (often converging to ≈ 0.5 in a balanced system) signifies a well-balanced learning process.

3) Training Time and Convergence Rate: Additionally, the overall training duration and the convergence rate are vital parameters. These metrics influence the computational viability of the model. Due to hardware limitations, each epoch requires several minutes to finalize, with convergence typically reached after around 50–70 epochs, contingent on the batch size.

4) Frontalization Quality Assessment: Beyond numerical assessments, the quality of the generated images is evaluated visually. The model's proficiency in maintaining facial identity and geometry across various poses is scrutinized through side-by-side comparisons of real and generated faces. This visual examination confirms the perceptual realism accomplished by the GAN.

5) Robustness to Variations: The model’s efficacy is further evaluated on images exhibiting a range of poses, occlusions, and lighting conditions. Consistent accuracy in these scenarios reflects robustness and stability. The proposed GAN demonstrates strong performance even with non-frontal angles, showcasing its adaptability to unconstrained environments.

C. Dataset Used

The proposed model was trained and assessed utilizing a publicly accessible benchmark dataset to guarantee the robustness and comparability of the results.

- Source: Kaggle
- Dataset Name: CFPW (Celebrities in Frontal-Profile in the Wild)
- Contents: Contains facial images of 100 unique individuals.
- Structure: Each individual features 10 frontal-view images and 4 profile (side-view) images, accompanied by annotations that support both ”Frontal–Frontal” (FF) and ”Frontal–Profile” (FP) verification protocols.

D. Software Requirements

The execution of the proposed GAN-based face frontalization system was performed in the following software environment:

- Programming Language: Python
- Development Platform: Google Colab • Operating System: Windows 11
- Libraries and Frameworks: NumPy, Pandas, Matplotlib, Keras, TensorFlow, and Dlib.

E. Hardware Requirements

The model was trained and evaluated under the following hardware specifications:

- RAM: 4 GB
- Storage: 6 GB of available disk space

5. CONCLUSION

The proposed system addresses a critical limitation in many facial recognition pipelines: the significant drop in performance when presented with non-frontal, side-view facial images. To overcome this, we introduce a Generative Adversarial Network (GAN) specifically designed for face frontalization. This model consists of two core components: a generator and a discriminator. The generator is built on an encoder-decoder architecture; it first encodes the side-view input into a low-dimensional latent representation, which is then passed through a series of deconvolutional layers (the decoder) to reconstruct a photorealistic frontal view. Concurrently, the discriminator learns to extract deep feature representations, enabling it to distinguish between authentic frontal images and the generator’s synthetic outputs.

REFERENCES

1. W. N. I. Al-Obaydy and S. A. Suandi, ”Automatic pose normalization for open-set single-sample face recognition in video surveillance,” Springer.
2. C. Li, W. Huang, and Y. Huang, ”Gabor Log Euclidean Gaussian and its fusion with deep network

based on self-attention for face recognition," ScienceDirect.
3. T. Cao, C. Liu, and J. Chen, "Nonfrontal Expression Recognition in the Wild Based on PRNet Frontalization and Muscle Feature Strengthening," Hindawi.
4. D. Saez Trigueros, L. Meng, and M. Hartnett, "Generating photo-realistic training data to improve face recognition accuracy," ScienceDirect.
5. "Pseudo Facial Generation with Extreme Poses for Face Recognition,"in Proc. CVPR 2021 Open Access.
6. Z. Bassel, K. Ilya, and M. Yuri, "PFA-GAN: Pose Face Augmentation Based on Generative Adversarial Network," Informatica, vol. 32, 2021.
7. C. Hu, Z. Feng, X. Wu, and J. Kittler, "Dual Encoder-Decoder Based Generative Adversarial Networks for Disentangled Facial Representation Learning," IEEE Access, 2020.
8. D. M. Souza and D. D. Ruiz, "Towards High Resolution Face Pose Synthesis," in Proc. International Joint Conference on Neural Networks (IJCNN).
9. H. Li and N. Xiao, "Face Rotation and Recognition Based on Attention Mechanism and Generative Adversarial Networks," IEEE Xplore.
10. A. Celik and N. Arica, "Enhancing face pose normalization with deep learning," Scientific and Technological Research Council of Turkey, 2019.
11. V. Balasubramainam, "Facemask Detection Algorithm on COVID Community Spread Control using EfficientNet Algorithm," Journal of Soft Computing Paradigm, vol. 3, no. 2, 2021.
12. Y. Yin, S. Jiang, J. P. Robinson, and Y. Fu, "Dual-Attention GAN for Large-Pose Face Frontalization," in Proc. IEEE International Conference on Automatic Face and Gesture Recognition, 2020.
13. X. Tu, J. Zhao, Q. Liu, and W. Ai, "Joint Face Image Restoration and Frontalization for Recognition," IEEE Transactions on Circuits and Systems for Video Technology, 2021.
14. Q. Duan and L. Zhang, "Look More Into Occlusion: Realistic Face Frontalization and Recognition with BoostGAN," IEEE Transactions on Neural Networks and Learning Systems, 2021.
15. Y.-J. Heo and B.-G. Kim, "Frontal Face Generation Algorithm from Multi-view Images Based on Generative Adversarial Network," Journal of Multimedia Information System, 2021.