Generative Adversarial Networks for face generation: A survey

AMINA KAMMOUN*, CES Laboratory, University of Sfax, Tunisia
RIM SLAMA, LINEACT Laboratory, CESI Lyon, France
HEDI TABIA, IBISC, Univ. Evry, Université Paris-Saclay, France
TAREK OUNI, CES Laboratory, University of Sfax, Tunisia
MOHMED ABID, CES Laboratory, University of Sfax, Tunisia

Recently, Generative Adversarial Networks (GANs) have received enormous progress, which makes them able to learn complex data distributions in particular faces. More and more efficient GAN architectures have been designed and proposed to learn the different variations of faces, such as cross pose, age, expression and style. These GAN based approaches need to be reviewed, discussed, and categorized in terms of architectures, applications, and metrics. Several reviews that focus on the use and advances of GAN in general have been proposed. However, the GAN models applied to the face, that we call facial GANs, have never been addressed. In this article, we review facial GANs and their different applications. We mainly focus on architectures, problems and performance evaluation with respect to each application and used datasets. More precisely, we reviewed the progress of architectures and we discussed the contributions and limits of each. Then, we exposed the encountered problems of facial GANs and proposed solutions to handle them. Additionally, as GANs evaluation has become a notable current defiance, we investigate the state of the art quantitative and qualitative evaluation metrics and their applications. We concluded the article with a discussion on the face generation challenges and proposed open research issues.

CCS Concepts: • **Computing methodologies → Computer vision tasks**.

Additional Key Words and Phrases: Survey, Generative Adversarial Networks (GANs), face generation, applications, training, loss functions, benchmarks, and algorithms

1 INTRODUCTION

Generative Adversarial Networks (GANs) have recently proved remarkable performance on several tasks, in particular face generation and face manipulation. The important effect of GANs, which gives them an invaluable role to play, is their power to generate realistic faces from noise [1, 33, 67] taking into consideration the little details in texture and structure without artifacts which make them undistinguished by human eyes. GANs belong to the family of generative models [114, 143], which generate samples by learning a given data distribution from latent representations. Among the most popular generative models, we mainly find (1) VAEs [71], (2) Flow-based [70], and (3) GANs [46] models. Variational autoencoders models (VAE) optimize the log-likelihood of the data by maximizing the evidence lower bound. They learn a given distribution by comparing its inputs to its outputs, which are efficient for learning hidden representations of data but usually generate blurry images for new data. Flow-based generative models [72] are constructed by a sequence of invertible transformations. Contrarily to other types of generative models, they explicitly learn the data distribution. GANs use a discriminator model

Authors' addresses: Amina Kammoun, amina.kammoun@enis.tn, CES Laboratory, University of Sfax, Sokra Street km4, Sfax, Sfax, Tunisia, 3038; Rim Slama, LINEACT Laboratory, CESI Lyon, , Lyon, France, rsalmi@cesi.fr; Hedi Tabia, IBISC, Univ. Evry, Université Paris-Saclay, Paris, France; Tarek Ouni, CES Laboratory, University of Sfax, Sfax, Tunisia; Mohmed Abid, CES Laboratory, University of Sfax, Sfax, Tunisia.

0360-0300/2022/3-ART $15.00
https://doi.org/10.1145/1122445.1122456

Table 1. Related reviews

Year	Survey title	Reference	Focus
2017	Generative adversarial networks: introduction and outlook	Wang et al. [133]	GANs theory and future challenges
2017	Stable and improved generative adversarial nets (GANs): A constructive survey	Zhang et al. [155]	GANs pipeline including shallow neural networks and reinforcement learning
2018	Convergence problems with generative adversarial networks	S.A.Barnett [12]	Dissertation on Theoretic problems with GANs convergence
2018	Generative Adversarial Networks: An overview	Creswell et al. [32]	GANs theory and applications
2018	Comparative study on generative adversarial networks	Hitawala [55]	Evolution in GANs architecture
2019	Recent progress on generative adversarial networks (GANs): A survey	Pan et al. [109]	Recent progress in GANs
2019	A survey on face data augmentation	Wang et al. [137]	Face data augmentation methods including GANs
2019	A large-scale study on regularization and normalization in GANs	Kurach et al. [74]	Regularization and normalization methods in GANs
2019	How generative adversarial networks and their variants work: An overview	Hong et al. [56]	GANs theory
2020	Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy	Wang et al. [138]	GANs architectures and applications
2020	Stabilizing Generative Adversarial Networks: A Survey	Wiatrak et al. [140]	GANs stability training methods
2020	A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications	Gui et al. [49]	GANs algorithms and applications
2020	Generative Adversarial Networks (GANs): Challenges, Solutions, and Future Directions	Saxena et al. [118]	Advancements in GANs design and optimization solutions to handle GANs challenges.
2020	A Survey on Generative Adversarial Networks: Variants, Applications, and Training	A.Jabbar et al. [62]	GANs theory and proposed stabilizing training solutions in history
2020	Loss Functions of Generative Adversarial Networks (GANs): Opportunities and Challenges	Pan et al. [108]	Loss functions in GANs
2020	Image Synthesis with Adversarial Networks: a Comprehensive Survey and Case Studies	Shamsolmoali et al. [119]	GANs for image synthesis generation
2021	Generative adversarial networks for image and video synthesis: Algorithms and applications	Liu et al. [86]	Algorithms and applications for visual synthesis
2021	Generative adversarial network: An overview of theory and applications	Aggarwal et al. [2]	Theory and applications
2021	Generative adversarial networks and their application to 3D face generation: A survey	Toshpulatov et al. [126]	GANs used for 3D face generation
2021	**Generative Adversarial Networks for face generation: A survey**	**Ours**	**GANs used for face generation and their architectures, applications, semantic, and evaluation**

to measure the distance between the generated and the real data, which usually allows better quality generation. Additionally, GANs are able to control sharp estimated density functions and eliminate deterministic bias [138] which allowed them to achieve great progress in the field of computer vision.

They attract the interest of deep learning committees, from which they surveyed their improvements. Table 1 summarizes most of the surveys on GANs. GANs have undergone a great evolution in generating images from an unconstrained context to specific domains and from grayscale face to high resolution photorealistic face images.

This success is done thanks to the modifications over the original model. The milestones of face generation using GANs over the past years are presented in Fig. 1, in which we categorize facial GANs into three basic categories: Conditional, Controllable and Progressive GANs. In conditional GANs, a condition or code is injected to both the generator and the discriminator for disentangled factor generation. The controllable GANs where the facial attribute is edited in the latent space. The progressive GANs in which the generation is learned continuously. The generator improves progressively, which allows better output quality and more stable training. Other GANs are also presented in the history, which we reviewed also in the milestone. The progress of GANs has reshaped the research landscape of face generation and improved state-of-the-art face recognition models in laboratory uncontrollable cases with limited data. Among popular works, [82, 137] showed that adding generated data to a small database increases performance by 15%. Recently GANs progress is no more restricted to data augmentation,

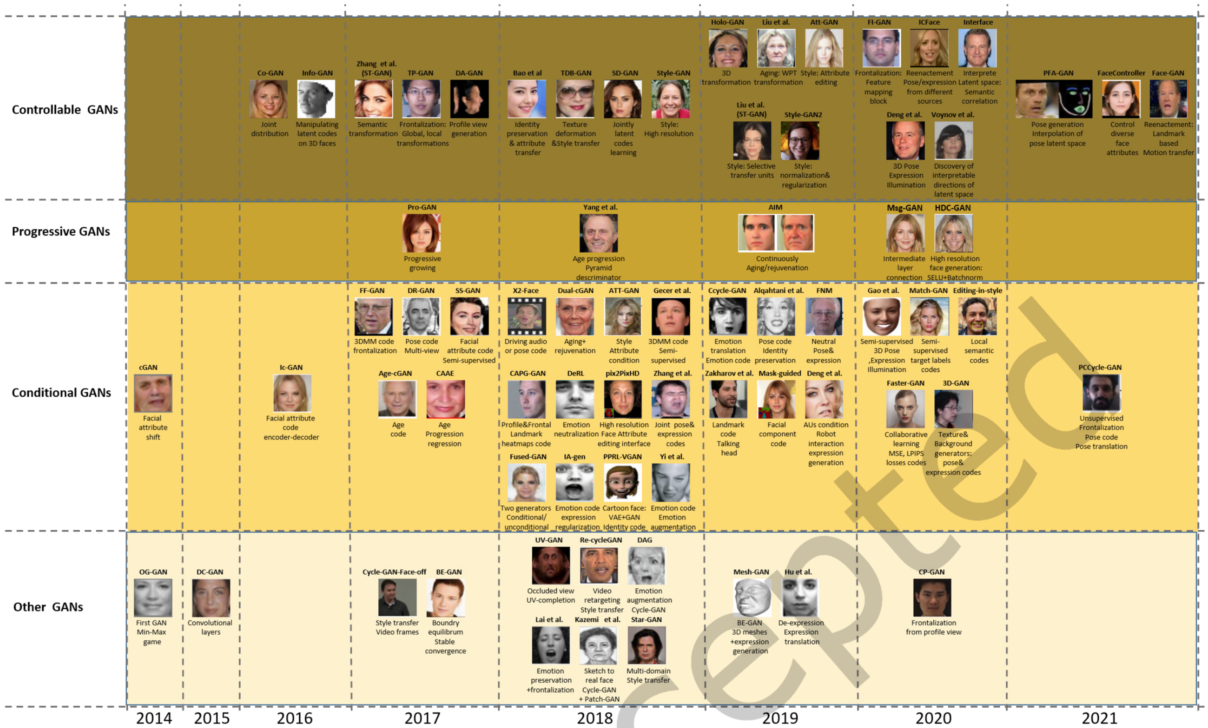


Fig. 1. Milestone of GANs for face generation: Facial GANs are reviewed over time and category. Three main categories of GANs are presented: conditional, controllable and progressive GANs, in addition to other GANs which are presented in the first row.

but covers facial image preprocessing, in particular deblurring and inpainting. In addition, GANs architectures have received important improvements to be able to generate not only uncontrollable faces, but also 2D and 3D faces, static and dynamic, as well as cross-factors face generation, especially cross-pose, cross-facial expression, cross-age, and cross-style. Several surveys on face recognition and processing techniques have been published in recent years. They focus on network evolution from traditional to deep learning based CNNs ones [51, 99]. On the other hand, they are also many surveys as shown in Table 1 that reviewed the GANs architecture and progress in various applications. However, there is no exhaustive survey which reviews GAN in the facial context, despite the multitude of works that have emerged. A recent work by Toshpulatov et al. [126] focuses on GANs applied to 3D face generation. In this article, we review generative adversarial networks for face generation of different modalities including 2D and 3D, static and dynamic face generation. We propose a taxonomy for the existent models and discuss their different subtitle architectures and used techniques. We cover about two hundred works, most of them within the recent five years. We aim to analyze, summarize and compare state-of-the art GANs architectures, submitted over the last years, in terms of image quality and accuracy. We present a complete overview of facial GANs models considering architectures, databases, applications, and evaluation metrics. Various proposed taxonomies are proposed in order to have a global view that brings together the existing works. Specifically, the major contributions of this survey are as follows:

- A Review of GANs applications on the face, including face generation and preprocessing, is presented. In this survey, applications using 2D/3D, and static/dynamic (video) data are covered. Real world challenges treated by GANs especially cross-pose, cross-expression, cross-age, and style variations are also reviewed.
- An overview of state-of-the-art metrics used for facial GANs evaluation is conferred.
- A taxonomy of facial GANs models in terms of architectures, in particular loss functions, applications, and training mode is proposed.
- A qualitative and quantitative comparison between the different solutions is exposed.
- An overview of the available benchmarks used to train and validate facial GANs applications is reported, with a proposed taxonomy and detailed user guide to be easily used in a specific application.
- An analytic study of facial GANs problems related to training and output quality with the proposed solutions

The remainder of this survey is organized as follows. Section 2 introduces the basic theory and types of GANs. Section 3 reviews the facial GANs applications. Section 4 proposes a taxonomy for GANs models for face generation in terms of architecture, training mode, and semantic. Section 5 presents an overview of the available datasets used for face generation and summarizes the existing evaluation metrics of GANs models. Section 6 discusses the facial GANs problems related to training stability and output fidelity, with the proposed solutions to handle. Section 7 presents challenges related to face generation with open issues. Section 8 concludes the article.

2 BACKGROUND CONCEPT AND TERMINOLOGY

GANs were first introduced to generate images from random noise. First works has been employed on different contexts and evaluated on data including digits, animals, flowers and faces. As shown in table 1, many surveys reviewed general GAN architectures. However, GANs applied to face generation have never been reviewed before. The face is a complex object which can be generated in 2D, 3D, static and dynamic samples. It is more difficult to synthesis compared to other objects as handwritten digits. The face generation comes with many difficulties like identity preservation, symmetry, and texture. Moreover, many semantics such as age, pose, expression and style have to be considered for realistic generation. Recent works with specific GANs architectures addressed these issues and progress has been made to improve the quality of generated faces. In this article, we reviewed the majority of these works. From the baseline GANs that were applied to face as a case study to specific GANs models that have been proposed to resolve the various face generation challenges. In this section, we first present the basic theory behind GANs and then give their different types, especially conditional GANs, controllable GANs, and progressive GANs.

2.1 Basic theory of GANs

GANs were first introduced in 2014 by Goodfellow et al. [46] inspired by the game theory, where two networks, a generator and a discriminator, compete to achieve the Nash equilibrium. The generator seeks for generating face samples as realistic as possible to fraud discriminator, which tries to differentiate between real and fake samples. The generator gets the discriminator's feedback and keeps improving to trick the discriminator into classifying the generated fake data as true (Fig. 2(a)). The generator and the discriminator play min-max game with the corresponding Value function (Eq1).

$$\underset{G}{min}\,\underset{D}{max}\mathbb{E}_{x\sim P_r}[log(D(x))] + \mathbb{E}_{z\sim Pz(z)}[log(1-D(G(z)))] \tag{1}$$

where G, D represent, respectively, the generator and the discriminator models. The generator estimates the probability distribution of samples G(z) given an input distribution $z \sim Pz(z)$ and approximates the real data distribution P_r.

2.1.1 *The generator.*
The Generative model learns to synthesize a realistic representation of some classes. By taking a random set of inputs, it captures the probability distribution, $P(X|Y)$, of different features X in the desired class Y to generate a set of features that look as realistic as possible. The generated samples are fed to the discriminator for inspecting how far they are from being real. Basing on the obtained score from the discriminator, the generator looks for the direction to move and update network parameters to generate more realistic samples that fraud the discriminator.

2.1.2 *The descriminator.*
The discriminative model is a classification model that learns to distinguish between two classes of real and fake data. Specifically, it models the probability of class Y given a set of features X; $P(Y|X)$. The estimated probability of generated samples being fake guides the generator to improve overtime for generating more realistic outputs.

2.2 Types of GANs

Divers GANs are proposed in the history. They can be classified into three basic types; Conditional, controllable, and progressive GANs.

2.2.1 *Conditional GANs.*
Traditionally, GANs generate random face samples unconditionally and have no control over the class of objects to be generated. In addition to random inputs, conditional GANs [43, 100, 110, 115] requires an additional code to guide the generator outputting samples from a specific class. They require labelled training datasets to learn to generate samples in desired classes. The generator takes one-hot class representation and a noise vector as inputs and outputs samples from the corresponding class. While the discriminator, takes one-hot labels in depth in addition to the fake generated image to estimate if it is real/ fake and belongs to the desired class. Fig. 2(a) shows an example of cGAN for face generation, where the condition c can be style, pose, expression, age code, etc. The cGAN aims to minimize the corresponding objective function. Equation (2) is inspired from (1). Actually, the conditional GAN architecture differs from the original GAN with the code y conditioned on the generator and discriminator inputs.

$$\underset{G}{min}\underset{D}{max}\left(\mathbb{E}_{x,y\sim p_r(x,y)}\left[logD(x|y)\right]\right)+\mathbb{E}_{y\sim p_y,z\sim p_z(z)}\left[log\left(1-D(G(z|y),y)\right)\right] \quad (2)$$

2.2.2 *Controllable GANs.*
Controllable GANs aim to generate objects with specific features by tweaking the input noise. Actually, any change in the noise vector affects the generated output. For this purpose, a pretrained classifier is used to look for the presence or not of desired features in the generated images, which permit to define noise vectors for different features. Fig. 2(c) presents an example of controllable GANs [66].

2.2.3 *Progressive GANs.*
Progressive GAN is a training methodology of GANs, which consists in increasingly learning to generate the desired faces from low to high resolution or from structure to fine details. This process allows speed and stable training with high quality generated outputs. This concept was defined in Progressive Growing GANs (Pro-GAN) [65] which consists in growing both generator and discriminator progressively. The model starts which small input in low resolution, then incrementally new blocks of layers are added which increase the output size of the generator and the input of the discriminator until achieving the desired image size (Fig. 2(b)).

3 FACIAL GANS APPLICATIONS

There are diverse applications that use GANs with faces. In this section, we give an overview of facial GANs, and we group these applications as face inpainting, face deblurring, and face generation. Fig. 3 exposes the different categories of facial GANs. We notice that face generation is the facial GAN category which is most explored in researches. Therefore, we focus in this study on face generation in depth. We group two main classes of 2D and 3D face generation. 2D faces are generated in static and dynamic form and different applications. We categorize

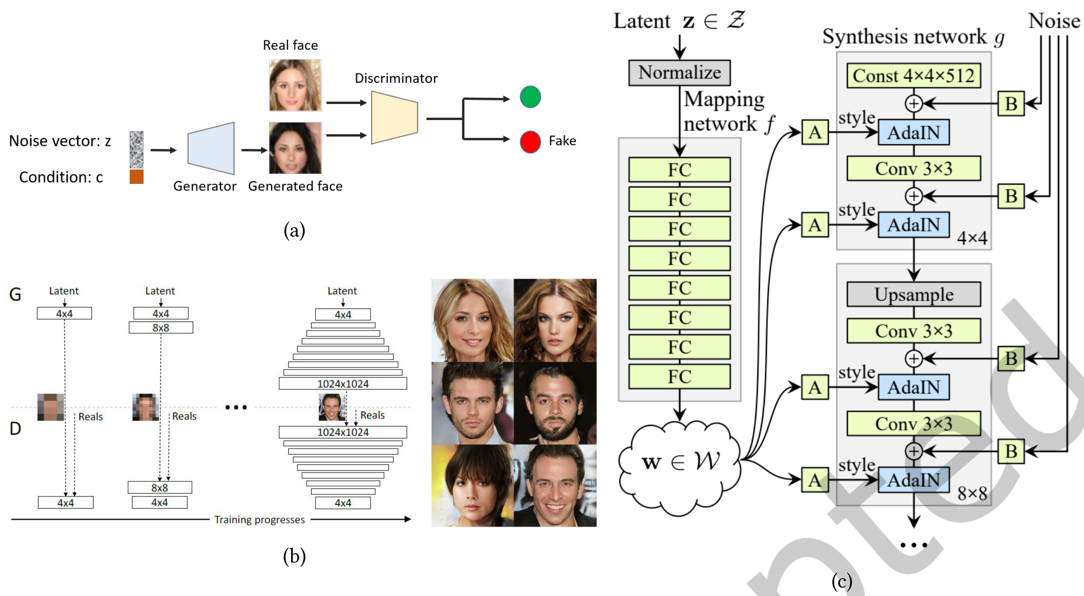


Fig. 2. Types of GANs: (a) Baseline GANs architecture for face generation/ Conditional generative networks cGAN [110] inject a condition code c in the latent noise input. (b) Example of Progressive GAN architecture taken from [65]: The training starts with both the generator 'G' and discriminator 'D' having a low spatial resolution of 4×4 pixels. Progressively, the number of layers is incremented, which increases the spatial resolution of the generated images. The N × N convolutional layers operate on N × N spatial resolution. The images on the right corresponds to generated samples using progressive growing at 1024 × 1024. (c) Example of controllable GANs taken from Style-GAN [66]:Contrarily to traditional generator which feeds the latent code though the input layer only, Style-GAN map the input to an intermediate latent space W, which then controls the generator through adaptive instance normalization (AdaIN) at each convolution layer.

the applications into 4 basic families; namely, image super-resolution, image-to-image translation, image matting, and image completion.

3.1 Face inpainting

Face inpainting GANs [84] learn the reconstruction of arbitrary corrupted resource and generate missing blocks (Fig. 4(a)). Inpainting based GANs [9, 26, 27, 44, 78, 96, 134, 135, 139, 151] aim to restore the corrupted regions of faces caused by extreme lighting variations, occlusion, or disguise. Many GANs architectures are proposed to perform well in image inpainting due to the generator and the discriminator for filling, missing or replacing damaged parts. Wang et al. [134, 135] proposed different types of deep neural networks based on semantic segmentation for face detail recovery. In [134], a Laplacian-pyramid-based convolutional network framework has been designed to predict missing regions under different resolutions. A residual learning architecture has also been used to progressively generate high-frequency details and eliminate color discrepancies which ensure visual consistency in the completed image. In [135], Wang et al. proposed an end-to-end generative adversarial network with recurrent architecture for face completion. The Recurrent Generative Adversarial Network takes advantage of multi-scale features, and facilitates both spatial information and high-frequency details recovery.

Some evaluation metrics such as SIM, PSNR, FID [9, 26, 139] are generally used for image improvement and photorealism optimization.

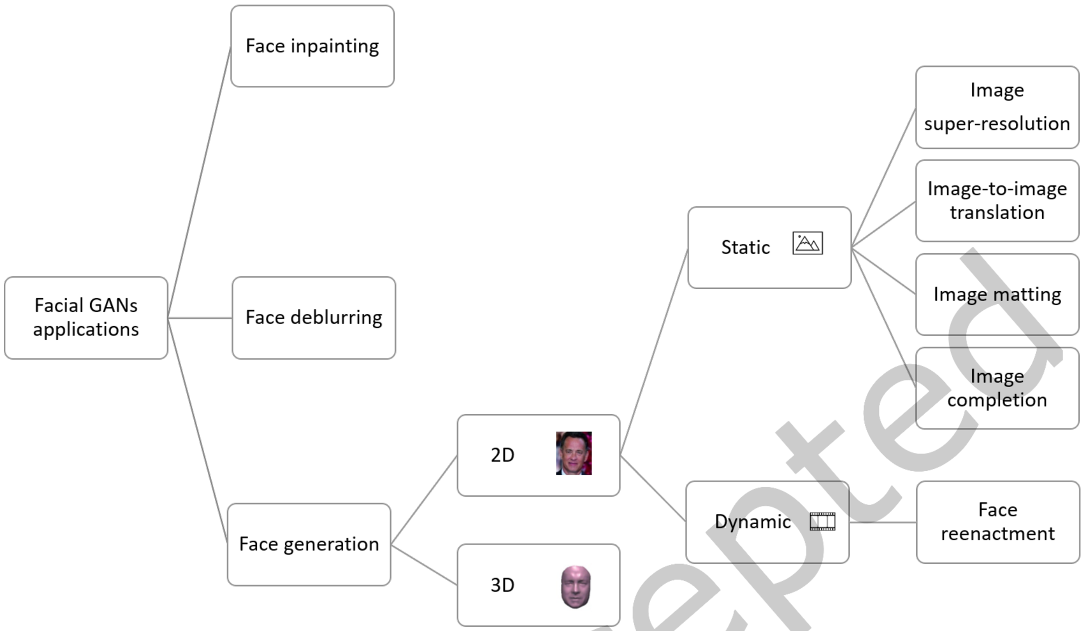
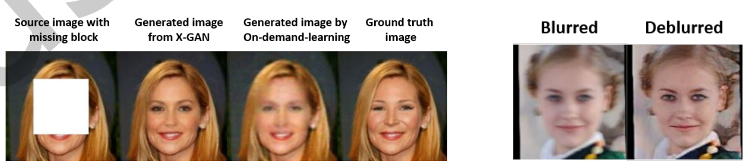


Fig. 3. Facial GANs applications

3.2 Face deblurring

Face deblurring consists in eliminating the motion blur, especially from the video feed of low quality surveillance cameras. An example of face deblurring is presented in Fig. 4(b). Many approaches [30, 39, 90, 94, 104, 121] proposed to use GANs to resolve this problem by restoring the latent sharp faces from the corresponding blurred ones. The generator learns to reproduce a similar visual appearance from the blurred version, while the discriminator helps to evaluate the realism concerning finer details. Multi-scale progressive GANs [21] are used for high-quality face restoration.



(a) (b)

Fig. 4. GANs for face inpainting and deblurring: (a) Example of GANs for face inpainting taken from X-GAN [84] which compared GANs with on-demand-learning framework [41] to show the power of GANs to generate photorealistic missing blocks. (b) Example of GANs for face deblurring taken from [90]

3.3 Face generation

Face generation is the major focus of facial GANs. First GANs focus on data augmentation to reduce the overfitting of deep neural networks [2, 138]. However, with the improvement of GANs architectures, they become effectively able to generate objects in specific applications [119]. In this section, we propose a taxonomy for face generation GANs applications, based on the one hand on 2D and 3D generation, and on the other hand on static and dynamic generation.

3.3.1 2D generation. 2D face generation approaches include image super-resolution, image-to-image translation, image matting, and image completion.

Image super-resolution approaches consist in generating face images in high resolution. [33, 64, 65, 77] use GANs to learn progressively to generate face images from low to high resolution. Fig. 5(a) explains how super-resolution based GANs [64] improves continuously image resolution over epochs.

Image-to-image translation consists of taking face images from one domain and transforming them to the style of images from another domain. Fig. 5(b) [125] exposed an example of GANs application for transferring facial images from the original expression to other target expressions. Several researches applied GANs [10, 24, 88, 91, 125, 168, 169] to learn the mapping between source and target domains of facial semantic, or style.

Image matting consists of estimating the foreground object in images and videos. In [60, 93], the generator is trained to predict alpha mattes from input images while the discriminator is trained to distinguish good images composited from the ground-truth alpha from images composited with the predicted alpha (Fig. 5(c)).

Face completion consists in generating or repairing the missing occluded parts of a face. [34] applied GANs for UV face completion by synthesizing the content of occluded regions in faces during pose generation (Fig. 5(d)).

Face reenactment is an emerging conditional face synthesis task that aims to transfer a source face shape to a target face while preserving the appearance and identity of the target face. As presented in Fig. 5(e), reenactment GANs [129, 141], take face identity from a source frame or image, and head pose and expressions from one or many driving sources. The models output a dynamic face of a source identity which imitates facial motions from the driving sources.

3.3.2 3D generation. 3D generation approaches consist of 3D face synthesis for data augmentation and face deformation. GANs are powerful in generating 3D faces [20, 28, 76, 98, 102, 103, 126, 166]. Fig. 5(f) shows an example of 3D face generation in various expressions.

4 GANS FOR FACE GENERATION

Face generation has become the focus of recent researches in GANs. It has received more interest in research committees compared to face inpainting and deblurring. On account of this, we loop in the face generation applications in this research. Thanks to the power of GANs architecture to learn distributions with its finest details, various GANs models have been proposed to generate faces. Facial GANs have received important progress from uncontrollable face generation in low resolution to photorealistic faces in specific semantics. This progress has been made in different axes. Some researches focused on improving GANs models based on the output target semantics, specially pose, facial expression, style, age, and videos. Specific architectures were proposed for this purpose using the progress of conditional, controllable and progressive GANs, and introduced further, additional losses to stabilize the training of GANs in different variation, in addition to tuning the facial details and eliminating imperfections. Other researches had a go at improving GANs models with the limited available datasets. Different problems of image-to-image translation from impaired datasets, unlabeled and as well as in the wild data were the concern of various researches. In this section, we first present the evolution of GANs over time by presenting the milestone contributions. Second, we reviewed the improvements in architectures, in particular loss functions. Then, we studied the advancement in facial models for generating faces in different

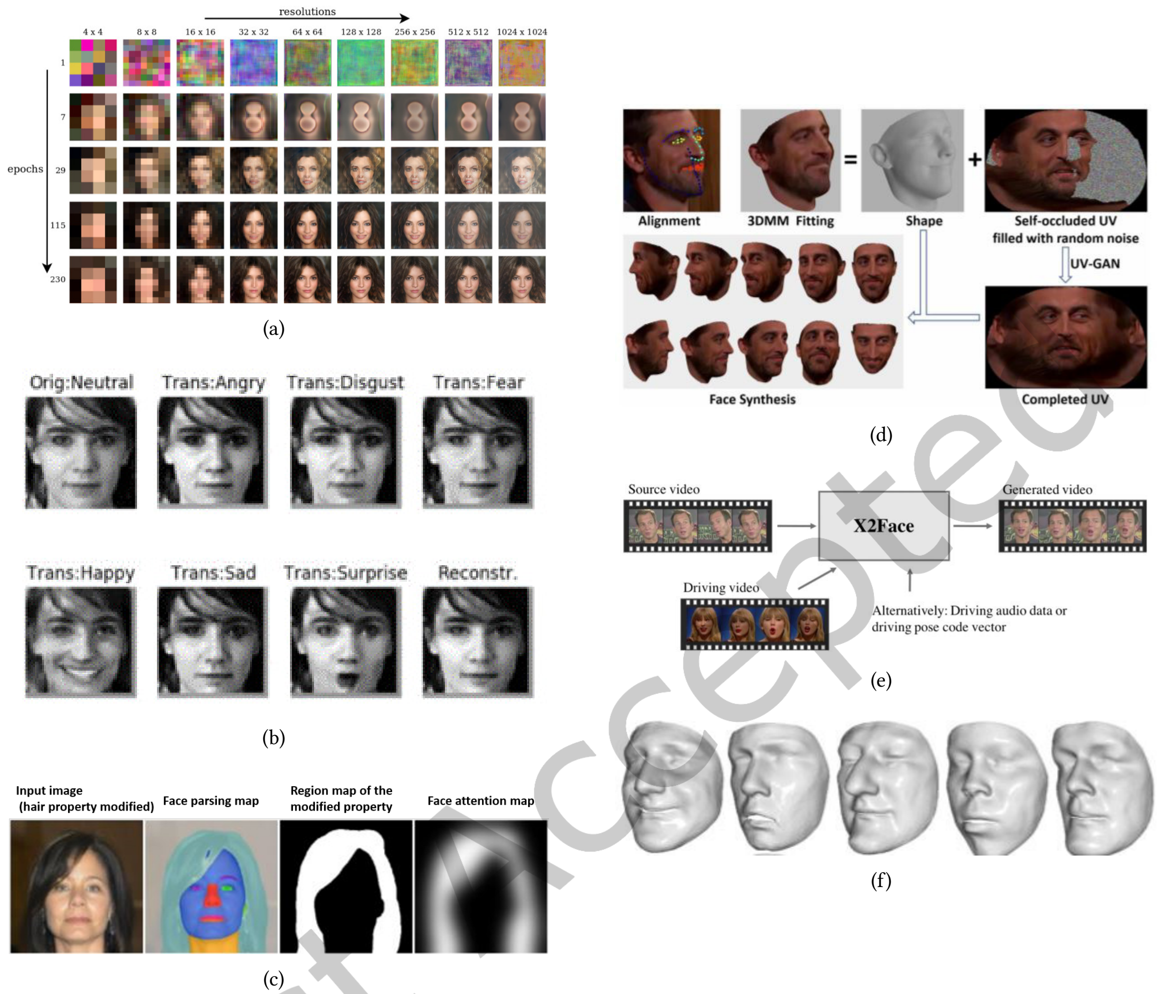


Fig. 5. GANs for face generation: (a) Example of super-resolution GANs taken from [64]: the resolution of generated images progresses over epochs. (b) Example of image-to-image translation GANs for facial expression transformation taken from [125]: the original expression image is translated into 6 other expressions. (c) Example of GANs for image matting taken from [60]: the image matting is used for predicting the fakeness region while comparing between two face images. (d) Example of GANs for face completion taken from UV-GAN [34]. (e) Example of GANs video face generation [141]. (f) Example of GANs for 3D face generation [28] in various expressions.

training modes. Finally, we confer a detailed study on the GANs progress for every semantic generation, by presenting the specific developments proposed to solve the problem of every semantic.

4.1 Evolution of facial GANs

Facial GANs models have progressed from generating grayscale uncontrollable faces to photorealistic faces with high resolution and style control [19, 33, 52, 65–67, 87, 113, 120, 152]. From the milestone of face generation using GANs over the past years presented in Fig. 1, We notice huge number of researches in conditional GANs especially between 2018 and 2019. From 2019, the number of researches in controllable GANs increased. From these researches, some integrated progressive GANs in the framework to improve image quality. Progressive GANs can cause artifacts in the progressive learning, which is improved by adding skip connections and regularization. In 2021, most researches used controllable GANs and focus on interpreting the latent space and control many attributes. This is because the controllable GANs proved better quality output and better convergence. However, they need larger datasets and computation requirements. In each category, many improvements are made over time, especially in loss functions to stabilize the training of different face variations and optimize the quality of outputs.

4.2 Facial GANs architectures

GAN architectures have received great improvements to address challenges related to resolution and fidelity. The first architecture of OG-GAN [46] composed of fully connected layers that had limited capacity to generate grayscale uncontrollable faces. The application of convolutional layers in DC-GAN [113] improved the learning of more efficient facial features. Moreover, with the appearance of conditional architectures, GANs become able to learn the face generation in different variations such as pose via the injection of the corresponding codes in the inputs of the generator and the discriminator. In addition, controllable architectures [66, 67] were introduces for more fluid face control via the modification of facial attributes. Joint architectures [87] facilitated face generation with two or more parallel variations. Progressive models [65] adapt pyramid architecture for progressive learning of face distribution, which allows better resolution of generated faces. Additionally, some recent researches [52, 120] manipulate two or more face attributes and control the variation intensity via interpreting the separation boundaries between different facial styles. In the following we are going to present the progress of loss functions.

Loss functions One of the most challenging in generative adversarial networks is the choice of loss functions. Actually, GANs look for an equilibrium between the adversarial players, the generator, and the discriminator. The convergence of the model corresponds to the optimization of the adversarial loss function. Most of the researches focused on the improvement of this function. Among them, some works [7, 40, 50, 92] proposed improved versions of the adversarial loss to stabilize GANs training and convergence, while others introduced additional losses [10, 58, 75, 83] for specific task learning. Equation (3) presents the general form of GANs objective function.

$$L_{adversarial} + \sum_i \lambda_i L_i \tag{3}$$

Where $L_{adversarial}$ is the adversarial loss and $\lambda_i L_i$ is the additional optional losses, λ_i are the trade-off parameters. Table 2 presents the main loss functions [7, 10, 50, 58, 75, 83, 92] for face generation. A detailed specification for each function is presented to better understand the difference and specificity of each loss equation.

4.3 Facial GANs training modes

The recent researches in facial GANs are no more limited to generating uncontrollable faces and go further for generating faces in various applications with limited data. The generation of face cross different semantic variations consists in training networks to disentangle between the invariant content and the specific information related to the facial variation. The disentanglement representation learning (DRL) was previously applied in multitask CNNs [149, 158, 170] using joint models to learn feature transformations in the target subdomains. Similarly, several GANs works [3, 11, 25, 36, 38, 103, 127, 127, 142, 150] proposed complex GANs architectures for facial style disentanglement. The disentanglement between facial content and the specific semantic can be done in either supervised, semi-supervised or unsupervised way depending on the availability of data and the adapted learning mode.

Table 2. Most applied loss functions for face generation

Name	Equation	Specification
Min-Max [46]	$L_{adv} = \underset{G}{min}\underset{D}{max}\mathbb{E}_{x_r\sim P_r}[log(D(x_r))]+$ $\mathbb{E}_{x_g\sim Pg(x_g)}[log(1-D(G(x_g)))]$	Derives from binary cross-entropy between real and generated distributions, Suffers from vanishing gradients and mode collapse
W-GAN [7]	$L_{W-GAN} = \underset{\gamma\in\prod_{(P_r,P_g)}}{inf} \mathbb{E}_{(x_r,x_g)\sim\gamma}[x_r - x_g]$	Computes Earth mover Distance (ED) between real and generated distributions,don't have flat regions, Eliminates vanishing gradients and reduces mode collapse, Uses weight clipping to enforce Lipshitz contraint on the critic, May cause poor samples generation or convergence failure
WGAN-GP [50]	$L_{WGAN-GP} = \underset{x_g\sim\mathbb{P}_g}{\mathbb{E}}\left[D(x_g)\right] - \underset{x_r\sim\mathbb{P}_r}{\mathbb{E}}\left[D(x_r)\right]$ $+\lambda\underset{x_g\sim\mathbb{P}_{x_g}}{\mathbb{E}}\left[\left(\left\Vert\nabla_{x_g}D(x_g)\right\Vert_2 - 1\right)^2\right]$	Improved version of W-GAN, Proposes to penalize the norm of gradient critic with respect to its inputs, More stable training without hyper-parameters tuning
cGAN [100]	$L_{cGAN} = \underset{G}{min}\underset{D}{max}\left(\mathbb{E}_{x_r,y\sim p_r(x_r,y)}\left[logD(x_r\vert y)\right]\right)$ $+\mathbb{E}_{y\sim p_y,x_g\sim p_g(x_g)}\left[log\left(1-D(G(x_g\vert y),y)\right)\right]$	Incorporates domain knowledge from the condition domain distribution
Pixel-wise [58]	$L_{pixel} = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{W_sH_sC}\sum_{w,h,c=1}^{W_s,H_s,C}\left\vert x_{g_{s,w,h,c}} - x_{r_{s,w,h,c}}\right\vert$	Smoothes the synthesized images, Speeds up the optimization and reconstruction the global information
Feature [75]	$L_{feature} = \frac{1}{N}\sum_{i=1}^{N}\left\vert F(x_{r_i}) - F(x_{g_i})\right\vert$	Facilitates the generation of images that match the probability distribution of real face images to prevent from over training on the discriminator
Total variation [58]	$L_{tv} = \sum_{c=1}^{C}\sum_{w,h=1}^{W,H}\left\vert x_{g_{w+1,h,c}} - x_{g_{w,h,c}}\right\vert$ $+\left\vert x_{g_{w,h+1,c}} - x_{g_{w,h,c}}\right\vert$	Reduces artifacts in the generated images
Patch-wise [83]	$L_{patch} = \frac{1}{3}\sum_{l=1}^{3}\frac{1}{\Vert p^l\Vert}\sum_i\sum_j\left\vert x_{g_{p^l_{i.j}}} - x_{r_{p^l_{i.j}}}\right\vert$	Ensures the consistency of the critical regions based on the fact: the pixel-wise variation inside a local landmark-based patch is consistent
Symmetry [83]	$L_{sym} = \frac{1}{W/2\times H}\sum_{i=1}^{W/2}\sum_{j=1}^{H}\left\vert x_{g_{i.j}} - x_{g_{W-(i-1).j}}\right\vert$	alleviates the self-occlusion in large pose variations especially for frontal faces generation
Identity preservation [58]	$L_{ip} = \left\Vert D^p_{ip}(x_g) - D^p_{ip}(x_r)\right\Vert_F^2 + \left\Vert D^{fc}_{ip}(x_g) - D^{fc}_{ip}(x_r)\right\Vert_2^2$	Is commonly used in face generation GANs to integrate domain knowledge of identities
Cycle-consistency [10]	$L_{cycle}(G_{D1}, G_{D2}) = \sum_t \Vert x_i - G_{D1}(G_{D2}(x_i))\Vert^2$	Enforces forward-backward consistency in impaired image-to-image translation

G: generator, D: discriminator, P_r: real data distribution, P_g: generated data distribution, x_r: real features, x_g: generated features, y: condition code, $\prod_{(P_r,P_g)}$: set of all joint distributions, $\gamma(x,y)$ the mass that must be transported from x_r to x_g in order to transform the distributions P_r into the distribution P_g, S:number of scales, H_s: height in each image scale, W_s: width in each image scale, N: total number of features, F: features for matching, W: width of the final generated images, H: height of the final generated images, C: the number of image channel, l=1; 2; 3: the number of the eyes, nose, and mouse patch respectively, $\left\Vert p^l\right\Vert$: the size of the patch, (i; j): the pixel inside the patch, $D^p_{ip}(.)$: the output of the last pooling layer, $D^{fc}_{ip}(.)$: the output of the fully connected layer, $\Vert\Vert_2$: the vector 2-norm, $\Vert\Vert_F$: the matrix F-norm, G : D1 → D2 : the mapping from domain D1 to domain D2, x_i: individual sample

4.3.1 Supervised GANs. The supervised GANs aim to generate facial semantics using either labelled data of the desired semantic or paired data of faces. In this article, we categorize supervised GANs into conditional GANs and paired image-to-image translation GANs.

Conditional Generative Adversarial Networks [43, 100, 115] are hugely used for supervised face generation. Actually, the inputs to the generator and discriminator are conditioned on the codes of the facial semantic. They mostly use labelled datasets of faces in the desired semantic to generate.

Paired Image-to-image translation GANs aim to translate an input image from one domain to another domain, given input-output pairs of facial images as training data. In fact, in the paired image-to-image translation GANs, each identity has a pair of images; an input image in the source style and an output image in the target style. So that, the generative model learns image-to-image translation to transfer faces from the input style to the output style using the facial images pairs. Pix2Pix [61] model is the first image-to-image translation model which uses paired images from source and target domains based on conditional generative networks where an image input is inserted in the generator instead of a class label. Image-to-image translation and image-to-video translation [47] are very common for face image mapping from one style domain to another or from a stable face image to talking face. Paired translation uses a pair of images for every person, mapping from the original style to target style. Pix2PixHD [136] inspired from Pix2pix model, proposed a new objective function as well as a novel multiscale generator and discriminator, in order to stabilize the training of conditional GANs on high-resolution images. Chen et al. [23] show that direct supervised training of generative models still suffers from unstable training and optimization which make them unable to generate photorealistic images.

4.3.2 Semi-supervised GANs. Various semi-supervised GANs architectures are presented for face generation with limited labelled data. Actually, the generation of faces with specific semantics, for example, pose needs a large labelled dataset of head poses. The generation of faces in this variation, with limited available datasets, results in unrealistic output of faces. Semi-supervised GANs try to tackle this problem by learning features related to face structure from large amounts of unlabeled data, while the limited labelled data fine-tunes the details for better quality face generation [42] introduced a semi-supervised GAN which reconstructs a 3D face using joint constraints from unconstrained photo collections of the person captured under different times, backgrounds, and expressions. First, an encoder network extracts four disentangled representations: identity, expression, pose, and lighting codes from the input face image. Then an albedo decoder network reconstructs the face albedo from the identity code, while a shape decoder network reconstructs the face shape from the combination of the identity and expression codes. Finally, the rendering layer renders back the face image from the face albedo, shape, pose, and lighting. Semi-supervised Regression GAN (SR-GAN) [107] introduced feature contrasting loss function, which allows semi-supervised GANs to be applied to regression problems and a discriminator that distinguishes between fake and real images through feature statistics. It is demonstrated that adding the SR-GAN generator and objectives to a CNN with unlabeled data almost increases the predictive accuracy of CNN models like age estimation. Bodla et al. [17] presented Fused-GAN conditional image synthesis with controllable sampling of diverse images by fusing two generators, one generates an unconditional image and the other generates a conditional image where the two partly share a common latent space disentangling the generation. The unconditional generator learns a good structure prior to synthesizing a better conditional face, by taking the structure as part of its input in addition to the condition. The dual generators allow disentangled representation for style and structure for face generation with high fidelity, diversity, and more controllability in sampling. Liang et al. [81] proposed a multi-modality network structure (CT-GAN) for better semi-supervised emotion recognition GAN with acoustic and visual modalities. Match-GAN [124] applied self-supervised learning for training conditional GANs under a semi-supervised setting. Facial data is augmented by sampling sensible labels from available labelled examples and assigning them as target labels to the unlabeled examples from the same distribution of labelled data. Triplets faces are generated containing both matched and mismatched pairs of target labels in the synthetic domain and source labels in the real domain. The triplet-like constraint forces the generator to maintain consistency on translated attributes and synthesize better images. Gecer et al. [45] presented an end-to-end semi-supervised adversarial framework to generate photorealistic face images of new identities in different expressions, poses, and illuminations conditioned by synthetic images sampled from a 3D morphable model. SS-GAN [123] shows that learning conditional models in a semi-supervised setting marked improvement in performance compared to semi-supervised approaches and performs almost as well as fully supervised conditional GANs.

4.3.3 Unsupervised GANs. With the limited available labelled datasets of faces in some specific semantics, recent researches proposed new approaches based on unsupervised learning. The unsupervised GANs instead of using the semantic label for face generation with the desired variation, they try to learn the difference between the common content of the face and the proposed variation unsupervised. Therefore, they generate semantic faces either by transforming the noise input vector or by computing the difference between two face styles for style transformation. In this section, we categorize unsupervised GANs into latent space control GANs and impaired image-to-image translation GANs.

Latent space control GANs Latent space or the embedding space [4], is the space in which a compressed representation of data lies. Modifying facial attributes such as style in image space directly highly difficult because of the high dimensions and complexity of the image distributions manifolds. However, the latent representation of the face expresses the specific features of the corresponding image in a compressed way. So that, the manipulation of the latent space for change of facial style is softer than the manipulation of image space. Zhang et al. [157] proposed a semantic transformation generative network (STGAN), in which facial semantic features are mapped into the latent vector and then the transformation is performed by controlling the latent vector. Voynov et al. [131] presented the first unsupervised approach for the discovery of semantically meaningful directions in the GAN latent space which can be universally applied to pretrained generators. The directions of different semantics are defined with model-agnostic approach. [162] introduced Unsupervised Adversarial Attacks on Generative Adversarial Networks (UAA-GAN) to attack deep feature-based image retrieval systems. UAA-GAN requires a small amount of unlabeled data for training and once trained, it can generate perturbations according to each query image. It was proved that this approach is effective in content-based image retrieval, person re-identification and face search. The Info-GAN architecture [25] learns interpretable representations and disentangles visual concepts like emotions from faces unsupervised by maximizing the mutual information between conditional variables and the generative data. SD-GAN [38] incorporates a decomposed latent space and Siamese discriminators to disentangle factors of variation from identity.

Style-based GANs [31, 48, 66, 67] learn the unsupervised separation of high-level attributes and stochastic variation in the generated images. As illustrated in Fig. 2(c), Style-GAN [66] introduced mapping network which converts the latent code from latent space into an intermediate code in the intermediate space. The goal of this mapping network is to create untangled features for the generator and avoid feature combinations in the training dataset. Then, Adaptive Instance Normalization (AdaIN) [66] separately transfers style via intermediate code in each layer. Style-GAN computes a pair of style values (y(s, i) and y(b, i)) as scale and bias from w in each layer to apply the style to the spatial feature map i. The deformation in face can be coarse or fineness, and it depends on where style mixing or noise is added in the network. Coarser variations(like head pose) are produced when noise is added in earlier layers while finer variations(like smile), in later layers.

Unpaired image-to-image translation The problem of unpaired translation is common in face datasets, in which not all persons have face images in different styles(emotions, poses, etc.). Unpaired image-to-image translation uses two piles of style images, where each pile is composed of images from different persons. The unpaired translation learns mapping between style piles by keeping in both piles the common content and transferring the style as the different content. Fig. 5(b) presents an example of impaired image-to-image translation in the case of an impaired expression dataset (not all persons have 7 expressions). The translation between the original expression(neutral) and the transferred expressions (for example, angry) is done by learning the differences between both expressions domains. Moniz et al. [101] presented DepthNet as unsupervised approach for learning 3D facial structure estimation from a single image. It predicted 3D viewpoint transformations that match pose and facial geometry by inferring the depth of facial keypoints of an input image in an unsupervised manner. DepthNets are used for face rotation and replacement, and when combined with adversarial repair, it can blend warped faces to synthesize the background. The use of adversarial image translation techniques as a post-processing step to DepthNet, allows a complete headshot resynthesizing for faces retargeted to different poses or identities. Minisha et al. [104] introduced the first unsupervised GAN architecture for deblurring class-specific data. Two additional modules are introduced: A self-supervised reblurring module that guides the generator to generate a deblurred output corresponding to the input blurred image. And a gradient module which constrains the solution space of generated clean images. Zhou et al. [166] proposed an unsupervised GAN architecture that synthesize rotated faces using only single-view image collections in the wild. Faces are rotated in the 3D space back and forth, and re-rendered to the 2D plane, which preserves the details and illumination condition and can apply to arbitrary angles. Faces are translated from rendered image to real-image domain. Recycle-GAN [10] combines both spatial and temporal information along with adversarial losses for content translation and style preservation. It learns unsupervised video retargeting by face-to-face translation by keeping content from one face and capturing stylistic expression from the other face. Face normalization model (FNM) [112] generates frontal faces with neutral expression unsupervised. Face expert network is introduced to produce face identity. The reconstruction of the normal face is stabilized with pixel-wise loss, while the texture is refined with the face attention discriminators. Cycle-GAN [63] transforms an image from one pile to another pile style, then transforms the image generated back to the original style forming a cycle consistency. Two different GANs are usually used to transfer styles when keeping the content of the image. Cycle-GAN introduced a Cycle Consistency loss in additional to the adversarial loss for both generators as the difference in both directions between reconstructed images. The cycle of mapping from original style to target style, then from target to original one allows to learn from the difference between the original image style and the target one to improve generation. The cycle consistency transfers the uncommon style while preserving the common content. Cycle consistency loss works as a regularizer to prevent mode collapse and add realism to the generated images. Kazemi et al. [68] presented a generative approach to synthesize a photorealistic image from a face sketch unsupervised. In addition to a texture-wise discriminator, a novel perceptual discriminator is introduced to learn face geometry. Learning facial information helps the network to remove geometrical artifacts in the face sketch and improve face quality and recognition rate. The application of cycle consistency, in addition to the discriminator improvement, achieved promising results in image translation in terms of facial texture and geometry.

Conditional-Cycle-GANs In addition to impaired data from two domains, offers the style attribute as a condition to the generator. Lu et al. [91] proposed attribute-guided face generation based on Conditional Cycle-GAN. An additional attribute vector is embedded with the input to train the generator. Conditioning Cycle-GAN can handle the problem of impaired training data, and the additional attribute extracted from high resolution images allows easy control of the appearance of the generated face. Liu et al. [88] proposed Pose Conditional Cycle-GAN (PCCycle-GAN) for pose-invariant face recognition in an unconstrained environment. PCCycle-GAN uses Cycle-GAN architecture with a conditional pose label guiding the inverse mapping of face frontalization, which makes it advantageous with impaired dataset training. Pixel-level loss, feature space perception loss, and identity preserving loss are introduced to generate photorealistic and identity-preserving frontal face images.

4.4 Facial GANs semantics

In the recent years, we notice the emergence of specific facial GANs, where the architectures are modelled to focus on one or joint semantics. Fig. 6. presents the different facial semantics treated by GANs. In Fig. 7, we present the models related to each semantic; namely, pose, expression, age, style, as well as the models joining two or more semantics. Each model is presented in the color of the corresponding training mode: red for supervised; blue for semi-supervised, and green for unsupervised.

4.4.1 Cross-pose GANs. Cross-pose GANs group all the GAN models which are interested in the problems related to the facial pose. We categorize cross-pose GANs into 5 classes: Frontalization, profile generation, multi-view generation, pose completion, and 3D generation.

Frontalization. Frontalization-based GANs focus on generating a frontal face view from face in the other pose angle. CP-GAN [83] is introduced to frontalize profile faces with unaltered identity based on u-net auto-encoder generator and Siamese discriminator in addition to a combination of losses which learn identity preservation and

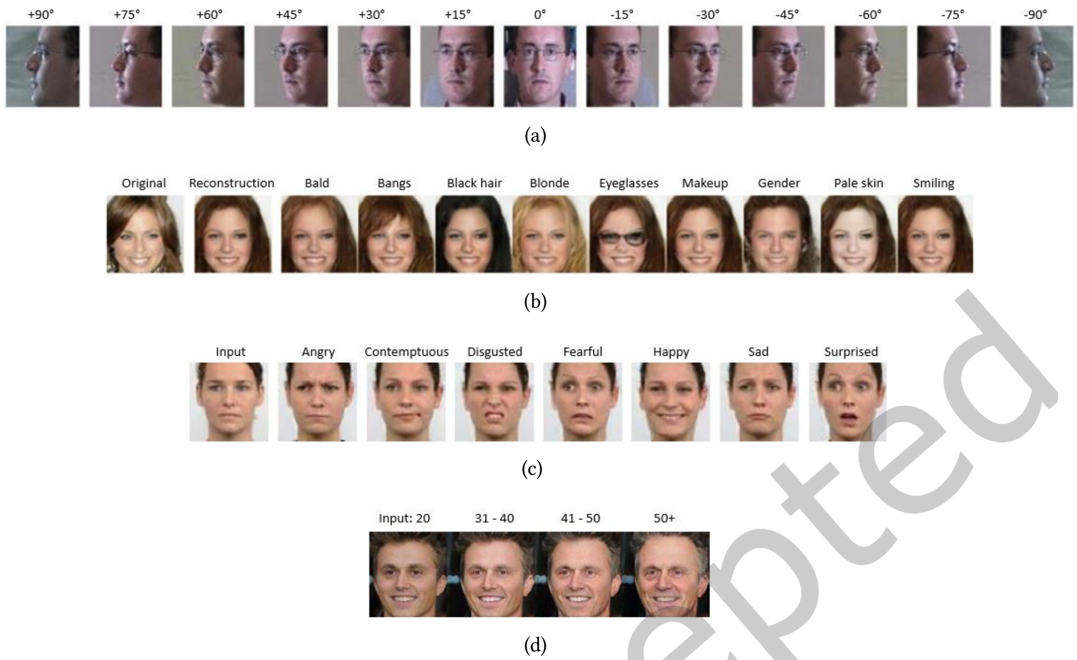


Fig. 6. Examples of GANs Facial Semantics. (a) Head pose generation [58], (b) Style editing [110], (c) Facial expression generation [24], (d) Age generation [146]

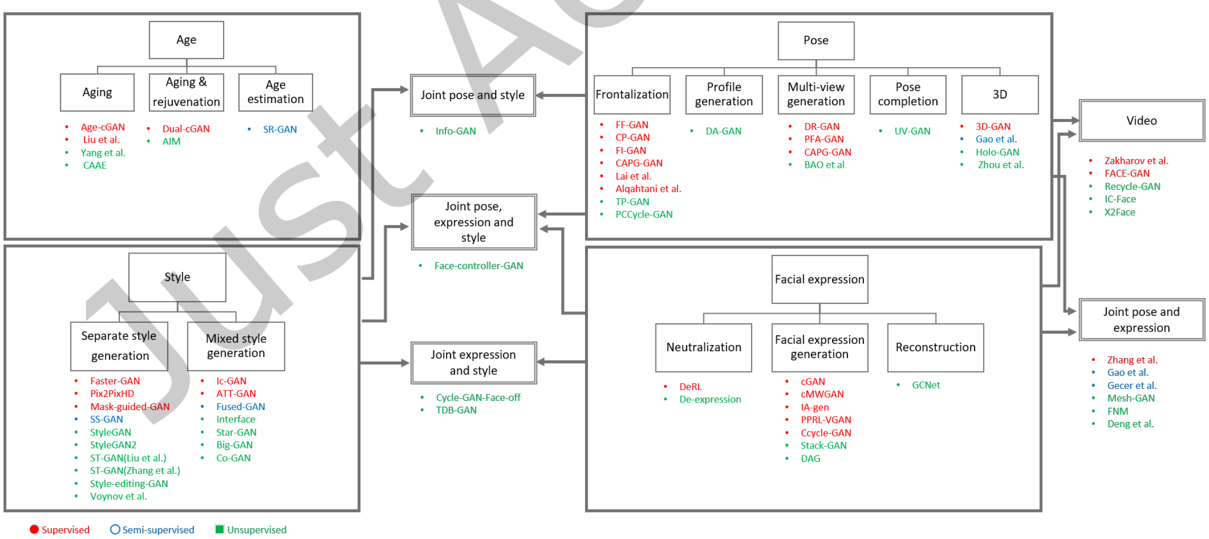


Fig. 7. Taxonomy of GAN models for face generation in terms of semantics and training modes

loss variation. Lai et al. [75] proposed a multi-task learning approach based on GANs that frontalizes input face images while preserving the identity and expression characteristics. FF-GAN [150] proposes a 3DMM-conditioned GAN framework to frontalize faces under all pose ranges. In addition to the generator and discriminator modules, FF-GAN has a 3DMM reconstruction module providing an important shape and appearance prior to guide the generator to rotate faces and a recognition engine to preserve identity to the generated image. Feature-Improving GAN (FI-GAN) [115] has a compact generation module named Feature-Mapping Block. The generation module that helps to map the features of profile face images to the frontal space. The proposed discriminator distinguishes the features of profile face images from those of ground true frontal images in order to guide the generation module to provide high-quality features of profile faces. Huang et al. [59] proposed a Two-Pathway Generative Adversarial Network (TP-GAN) for photorealistic frontal view synthesis by simultaneously perceiving global structures and local details. Local textures are learned from four landmark located patch networks in addition to the global encoder-decoder network. FF-GAN applied a combined loss function of adversarial loss, symmetry loss, and identity preserving loss for an identity preserving inference of frontal views from profiles. Alqahtani et al. [3] disentangled pose attributes from identity based on encoder-decoder structure. Liu et al. [88] proposed Pose Conditional Cycle-GAN (PCCycle-GAN) for pose-invariant face recognition in an unconstrained environment. PCCycle-GAN uses Cycle-GAN architecture with a conditional pose label guiding the inverse mapping of face frontalization.

Profile face generation. Zhao et al. [164] introduced Dual Agent GAN (DA-GAN) for identity preserving profile face generation by projecting a 3D face into 2D image space.

Multi-view pose generation. Disentangled Representation learning Generative Adversarial Network (DR-GAN) [127, 128] proposes an encoder-decoder structure of the generator to learn a generative and discriminative representation, in addition to image synthesis. DR-GAN takes one or multiple images as the input. The pose code provided to the decoder and pose estimation in the discriminator explicitly disentangle pose features from other face variations. The decoder generates one unified representation with an arbitrary number of synthetic images. Zeno et al. [153] introduced a pose face augmentation framework (PFA-GAN) which consists of two subpaths; a reconstruction subpath for generating face images with the target pose and a transformation subpath to synthesize the target face while preserving the subject identity. Hu et al. [58] proposes a novel Couple-Agent Pose-Guided Generative Adversarial Network (CAPG-GAN) to generate both neutral and profile head pose face images. Head pose information is encoded by facial landmark heatmaps. In addition to the encoder-decoder network, a couple-agent discriminator is introduced to reinforce the realism of synthetic arbitrary view faces. CAPG-GAN employs a weighted sum of generator, conditional adversarial loss, and identity preserving loss with total variation regularization to synthesize photorealistic faces while preserving identity and pose information.

Pose completion. The generation of facial pose from another pose with a wide variation, for example, from frontal to profile and vise versa, may cause occluded parts in the generated images. Pose completion aims to complete the missing part of the face caused by this transition. UV-GAN [34] presents a facial UV map completion solution by fitting 3D face models in images. Global and local adversarial networks are employed to learn identity-preserved UV completion. The attachment of completed UV map to the fitted 3D mesh, provides faces with arbitrary poses (Fig. 5(f)).

3D face generation. Mariott et al. [98] incorporated a 3D morphable model into the generator in order to learn a nonlinear texture model from in-the-wild images, which allows faces generation in different poses and expressions without compromising the identity. Holo-GAN [103] disentangles 3D pose and identity objects through rigid-body transformation of the learned 3D features. Moniz et al. [101] presented a generative approach for learning 3D facial structure estimation from a single image. Zhou et al. [166] proposed a GAN architecture that synthesize rotated faces using only single-view image collections in the wild. Faces are rotated in the 3D space back and forth, and re-rendered to the 2D plane, which preserves the details and illumination condition. Then, the faces are translated from rendered image to real-image domain. SofGAN [20] decouples the latent space of portraits into a geometry space and a texture space. The latent codes sampled from the two subspaces are fed to two networks to generate separately, 3D geometry of portraits with canonical pose, and textures. 3DFaceGAN [102] models the distribution of 3D facial surfaces, while retaining the high frequency details of 3D face shapes. The generator retrieves a facial UV map as input and generates a realistic one close to the real target facial UV map. Lattas et al. proposed AvatarMe [76] as the first approach which produces high-quality rendering-ready face reconstructions from arbitrary "in-the-wild" images.

Interpretations on cross-pose generation. Cross-pose GANs [75, 83, 88, 115, 150, 164] are first, designed to generate either frontal or profile faces. Most of these researches are based on the pose code and supervised training mode. Nevertheless, in the wild pose frontalization [112] is still a challenging problem which need unsupervised settings. However, the large pose variation makes it challenging to preserve identity, as well as face texture. Some researches [34] go for face completion by stabilizing the architecture for better quality generation. Recent researches on cross-pose generation are based on 3D GANs [20, 76, 98, 101–103, 166], which have additional challenging optimization problems [126, 165] compared to 2D GANs such as symmetry preservation and texture reconstruction caused by occlusion especially for extreme profile or frontal face generation.

4.4.2 Cross-facial expression GANs. GANs are used in history for several cross-facial expression problems [79] especially facial expression generation, de-expression and reconstruction.

facial expressions generation. Facial expression classification suffers from imbalanced label distribution, because some classes of emotions like disgusted are relatively rare comparing to other labels like happy or sad. GANs based data augmentation can increase emotions classification accuracy. Yi et al. [148] uses conditional generative adversarial networks to generate facial expressions. Convolution networks are used as the body architectures of generator and discriminator and Wasserstein distance as the loss function to improve the stability of our network and augment facial training sets. Results show that with facial expression augmentation, pipeline accuracy increases (4% absolute and 7% relative). The mean accuracy of all facial expressions improves from 61% to 65%. Deng et al. [35] presented a conditional generative adversarial network (cGAN) for human robot interaction. The framework consists of a generator G and three discriminators (Di, Da, and Dexp). Three loss functions are designed to learn Action Units(AUs), expressions and realism while preserving identity. Zhang et al. [160] propose a general framework of data augmentation using GANs in feature space for imbalanced classification. The framework consists of a feature extractor, feature generator, and classifier. The conditional mixture WGANs (cMWGANs) is used to generate preserved labels and diverse features by learning the distribution of features extracted from text or face datasets. EM distance is used to guarantee the stability of the GANs training process. The results on FER2013 show that training on subclasses is better than on full classes. Li et al. [80] proposed a framework for facial expression generation and photo-to-cartoon transfer. First, separately, Star-GAN was trained for facial expression generation and Cartoon-GAN for cartoon style transfer. Then, two stacked models were presented. Stack-GAN-A, in which Star-GAN is considered as the first GAN and Cartoon-GAN as the second, and Stack-GAN-B where the order of GANs is flipped. Zhu et al. [169] used GANs for data augmentation in emotion classification(DAG). The proposed framework consists of CNN based classifier and a cycle-consistent adversarial network (Cycle-GAN) as the generator. To avoid the gradient vanishing problem during training, the least-squared loss is combined with the original adversarial loss. Results show that the classifier can be better learned to find margins or hyper-planes of neighboring classes when possessing a more complete data manifold. After employing the GAN based data augmentation techniques, the classification accuracy increased by 5%–10%. Identity-adaptive generation [147] (IA-gen) consists of two parts. The first part generates images of the same subject with different expressions using cGANs. The second part is fine-tuned for facial expression recognition using a regular CNN. The classification is based on the minimum distance between the input image and the generated expression images in the feature space. CCYLE-GAN [125] synthesizes realistic face images conditioning on the emotion of facial expression in absence of paired examples. A combination of adversarial loss and cycle consistency loss are used to learn the mapping between different domains. Chen et al. [22] proposed a privacy-preserving representation-learning variational GAN (PPRL-VGAN) that combines VAE and GAN to learn an identity-invariant representation of a face image that permits synthesis of an expression-preserving. The network learns to disentangle the identity information from the latent representation via the identity code given to the generator.The model is tested on various applications like expression morphing and image completion.

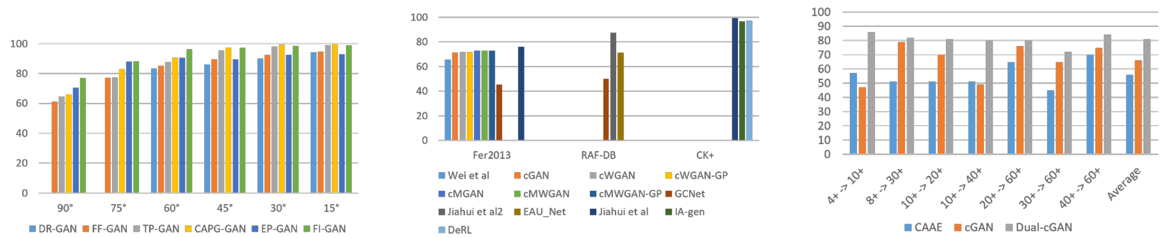
De-expression GANs. Facial expression can be decomposed into an expressive component and a neutral component. De-expression GANs aim to neutralize the expression component and keep only the neutral one, which allows generating faces with neutral expression. Hu et al. [57] proposed a facial de-expression and expression recognition model, able to accomplish the facial de-expression for both laboratory-controlled and in-the-wild datasets. The different expressions are treated as different image domains, so that domain transfer techniques are adopted to generate neutral face images for in-the-wild datasets.The main advantage of this approach is the possibility to deal with unpaired in the wild datasets. Yang et al. [145] proposed a De-expression Residue Learning procedure (DeRL) to recognize facial expressions by extracting information of the expressive component.A generative model is trained by CGAN to filter out the expressive information and generate the corresponding neutral face image for any input face image (de-expression procedure). Then, a learning procedure is performed on the internal layers to capture the expressive components of facial expressions that have been recorded in the generative model.

Facial expression reconstruction. Facial expression reconstruction GANs learn, given an input expression image, to reproduce the same expression. This process is useful for improving facial expression recognition when the expressive depth of an emotional face is varied among individuals, expressions, or situations [69]. GCNet [69] presented a facial expression recognition method based on contrastive representation learning. The contrastive representation is calculated in the embedding layers of deep networks by comparing a given image with a reference image. The reference image is generated by deep generative. This representation is provided as the input to the final classification layer for facial expression recognition.

Interpretations on cross-facial expression generation. Most of cross-facial expression GANs [22, 35, 57, 79, 80, 125, 145, 148, 169] are designed for expression neutralization or augmentation. The neutralization procedure is less challenging because the suppression of expression can be done with in the wild or impaired datasets. However, expression generation is more difficult when it uses impaired datasets. Consequently, the results look unrealistic for some expressions like disgust because with the lack of samples, expression-to-expression translation is an unstable process.

4.4.3 Cross-age GANs. Cross-age GANs group the GAN models specialized in age generation problems. We categorize cross age GANs into aging face generation, face rejuvenation, and age estimation.

Aging face generation. consists in generating an aging face of a person from a younger face image. Age Conditional Generative Adversarial Network (Age-cGAN) [5] synthesis aging faces while preserving the original person's identity. First, the input face reconstruction preserves the identity by looking for an optimal latent approximation, then face aging is performed by a change of condition at the input of the generator. Yang et al. [146] proposed an age transformation GAN based on a pyramidal adversarial discriminator. The progressive learning simulates aging effects in a finer manner while preserving personalized properties stable. The age transformation loss achieved age affect while pixel level loss and identity preserving loss kept the information related to image quality and identity. Conditional Adversarial Auto-encoder (CAAE) [161] achieved face age progression and



(a) (b) (c)

Fig. 8. Performance for semantic generation. (a) Head pose generation(MultiPIE database under setting2), (b) Facial expression generation, (c) Age generation

regression in a holistic framework. The face is first mapped to a latent vector through a convolutional encoder, and then projected to the face manifold conditional on age through a deconvolutional generator. Personalized face features were preserved through the latent vector while age features are controlled via the age condition. Liu et al. [89] introduced an attribute-aware face aging model with wavelet-based Generative Adversarial Networks for aged face image synthesis. Facial attribute vectors are embedded to both the generator and discriminator to learn to preserve facial attributes of the corresponding input image in the generated images. Wavelet packet transform is introduced at multiple scales to extract textual features which improves the visual fidelity of the generated face images.

Face Rejuvenation. consists in generating a younger version of the face from an elder image version. Dual conditional GANs (Dual cGANs) [122] trains face aging and rejuvenation from multiple sets of unlabeled face images with different ages. The primal conditional GAN transforms a face image to other ages based on the age condition, while the dual conditional GAN learns to invert the task. The face identity is preserved via the defined loss function. Zhao et al. [163] proposed Age-Invariant Model (AIM) extended from auto-encoder based Generative Adversarial Network which includes a disentangled Representation Learning subNet (RLN) and a Face Synthesis subNet (FSN) for age-invariant face recognition. The model learns to generate age-invariant facial representations disentangled from the age variation, which allows face rejuvenation/aging with identity-preserving.

Age estimation. is a challenging problem because most of the available labelled datasets are composed of age groups and not decimal numbers. This does not give a precise age estimation because the discontinuity of this function. SR-GAN [107] introduced semi-supervised approach for improving the predictive accuracy in regression problems like age estimation. Fig. 8 presents the performance of pose, expression, and age models.

Interpretations on cross-age generation. Most of cross-age GANs [5, 89, 146, 161] generate aging faces. However, only few models [122, 163] are capable of both aging and rejuvenation face generation. Actually, age generation is a challenging problem, because it needs a continuous function [107] to learn age features from different age groups.

4.4.4 Style generation. Style generation consists in synthesizing faces in specific styles such as hair color and glasses. Many models [17, 47, 66, 91, 110, 123, 124, 136, 157] have been proposed for facial style generation. Style-GAN [66] learns unsupervised separation of high-level attributes like pose and identity and stochastic variation in the generated images like hair. The input latent code was introduced into an intermediate latent space in addition to two automated metrics: perceptual path length and linear separability, with more linear and less entangled representation of different factor variation. StyleGAN2 [67] improved unconditional image modeling in terms of distribution quality metrics and image quality. Bhattarai et al. [15] proposed unsupervised facial image semantic transformation generative network (STGAN). Facial semantic features are mapped into the latent vector, and then the transformation is performed by controlling the latent vector. Collins et al. [31] proposed a model for local semantic part editing in Style-GAN images. In Editing-in-StyleGAN, the latent object representation guides the style interpolation to generate realistic part transfers without introducing artifacts. Guan et al. [48] proposed a collaborative learning framework for unsupervised real-time Style-GAN based semantic image editing applications. Faster Style GAN consists of an efficient embedding network and an optimization-based iterator. STGAN [85] incorporates the difference attribute vector and selective transfer units(STUs) in an encoder-decoder network. The bottleneck layer in encoder-decoder gives rise to blurry and low quality editing results. STGAN adds skip connections to improve image quality at the cost of weakened attribute manipulation ability. In addition, by taking the difference attribute vector rather than the target attribute vector as model input, the model can focus on attribute changes, which enhances the translation of attributes with better image reconstruction quality.

Mixed styles generation

Various studies used GANs for generating faces in mixed styles. Expressly, this corresponds to modifying, simultaneously, two or more facial attributes corresponding to the desired style. Shen et al. [120] proposed a framework called Interface-GAN for semantic face editing by interpreting the latent semantics learned by GANs. The study shows that the latent code of well-trained GANs learns a disentangled representation after linear transformation. Interface-GAN learns to separate entangled semantics with subspace projection to improve control of facial attributes. [29] presented Star-GAN as scalable image-to-image translation multitask model. Star-GAN acquires generalization ability by learning from multiple datasets with different sets of domain labels. Voynov et al. [131] presented the first unsupervised approach for the discovery of semantically meaningful directions in the GAN latent space which can be universally applied to pretrained generators. Invertible Conditional GAN (Ic-GAN) [110] introduced an encoder in a conditional setting of GAN model. Ic-GAN is an extension of CGAN which inverses the mapping of a real image into a latent space and conditional representation. The experiences done on the optimal position of conditional information show that it should be inserted in the input level of the generator and first layer of the discriminator. Moreover, training two separate encoders for conditional information and latent representation has given the best results. He et al. [53] considers the relation between attributes and the latent representation and proposes a facial attribute editing method called Att-GAN, which incorporates attribute classification constraint, reconstruction learning, and adversarial learning for high-quality facial attribute editing.

Interpretations on cross-style generation. Recently, cross-style GANs [29, 53, 110, 120, 131] go for generating mixed styles. The study of the entangled features is challenging. For example, some styles are related to age or gender which make it important to study the projection of some features in the feature space. In addition, some details like artifacts in hair generation interested recent researches [67] to optimize the realism of cross-style face generation.

4.4.5 Video face generation/Face reenactment. Face reenactment [130] is a facial animation technique where the person's identity is taken from the source image and the facial motion from the driving image. Recent researches have shown high quality results by incorporating facial motion representations with generative adversarial networks. GANs based reenactment models perform well when the source and driving images are from the same person or have similar structures. The current researches seek to avoid face reenactment distortion and improve outputs when identities of source and driving images are different. Face reenactment can be useful for video or image editing in addition to controlling facial poses and expressions. Zakharov et al. proposed a model [152] in which face landmarks are passed to the GAN architecture to drive animation of the model. FACE-GAN [130] used Action Unit (AU) representation to transfer the facial motion from the driving face since AUs are unlike facial landmarks independent of the facial structure. This can prevent identity leak and facilitate the reenactment control. Wiles et al. [141] presented a self-supervised framework X2Face that controls the pose and facial expression of a given face. The objective of this framework is driving face generation in an unconstrained setting like unseen identity and can be used as post training for driving face from audio or head pose information. Tripathy et al. [129] presents Interpretable and controllable face reenactment framework to control pose and expressions from different sources and driving identities. The source face is transformed into a canonical presentation, acting as template for subsequent animation steps. The animation was controlled via human interpretable attributes consisting of action unit activations and head pose angles. The control signal can be mixed from different sources, such as pose from one image and expression from another. Recycle-GAN [10] learns unsupervised video retargeting by face-to-face translation by keeping content from one face and capturing stylistic expression from the other face.

Interpretations on face reenactment. Dynamic face generation become the focus of facial GANs for its importance in real world applications [10, 130, 141]. However, face reenactment has more challenges compared to static face generation. Actually, when the identities of source and driving faces are different, the quality of generated faces can be affected. Texture distortion, identity loss, as well as shape bias are the mane problems of face reenactment. Current researches [129, 130] enhanced models either with inputs improvement (such actions units), or with better architectures (additional blocks).

4.4.6 Multi-domain generation. Multi-domain generation consists in learning to jointly generate faces in two or more semantics. Contrarily to mixed style generation approaches, which use the same database of different styles, these models can use more than one database of semantic faces, such as generating multi-view expressing faces from pose and facial expression data. [24, 25, 36, 42, 45, 63, 112, 144] are some works that proposed solutions on this task. Cheng et al. [28] show that Mesh-GAN is able to generate 3D facial meshes of different identities and different expressions and model the distribution of faces better than Auto-encoders. Zhang et al. [154] presented a joint model for simultaneous facial image synthesis and pose-invariant facial expression recognition. The encoder-decoder structure of the generator can learn the identity representation for face images and disentangle the identity from both expression and pose variations, through the expression and pose codes. Then the decoder generates facial images with arbitrary expressions and poses.

Interpretations on multi-domain generation. Some researches [25, 36, 42, 45, 144, 154] in facial GANs studied not only one facial semantic but also two or more semantics. These models learn to jointly generate faces in mixed domains. Despite the researches in some joint domains like pose and expressions, there is no model which generates age with other facial domains.

4.4.7 Discussion on facial semantics. As shown in Fig. 7, facial GAN semantics can be categorized in 4 main groups; namely, pose, expression, age, and style. We present also the groups which combine one or more semantics. About 77% of the models generate separate semantics, while others generate jointly two or more semantics. In particular, about 27% of facial GANs are specialized in cross-pose generation, 15% cross-expression, 10% cross-age, 25% cross-style, 7% face reenactment or video generation and others in multi-domain face generation. More precisely, 44% of cross-pose GANs synthesizes frontal views, 22% multi-view and 22% 3D faces. 70% of cross-expression GANs generate 7 expressions and 20% generate neutral faces. 57% of cross-age GANs generate aging faces, while only 29% can both aging and rejuvenation faces. 59% of cross-style GANs generate separately styles and 41% generate mixed styles. Besides, we remark that only 7% of Facial GAN models generate dynamic, and few works generate 3D faces. However, most of GANs generate static 2D faces. Moreover, about 73% of multi-domain facial GANs joint pose and expression for static or dynamic face generation. Furthermore, about 51% of facial GANs are unsupervised, 40% supervised, and only 9% semi-supervised models. Particularly, most of the style models are unsupervised, while most pose models are based on conditional GANs and expression models on Cycle-GAN. This can be explained by the type of datasets available in each semantic.

5 GANS EVALUATION

In this section, we present a study on facial GANs evaluation. First, the state-of-the-art datasets used for facial GANs training and validation are exposed. Then an overview of the metrics used for GANs evaluation is conferred. Actually, the GANs evaluation is a challenging problem from the original model until now because of the nonexistence of common evaluation in the existence researches. According to Boadla et al. [17], GANs evaluation consists in evaluating fidelity and diversity of the generated images. where, the fidelity measures the quality of generated images while the diversity reflects the capacity of the generative model to generate variety of samples. Various metrics have been proposed in history to evaluate GANs. [18] presented the pros and cons of some GANs evaluation metrics. In this research, we are interested in facial GANs evaluation, and we present a taxonomy for the existent metrics. Actually, we categorize them into quantitative and qualitative metrics. In addition, we present a comparison between the performance of the models that used these metrics.

5.1 Datasets

We present specific datasets used for face generation using GANs. We propose to categorize these benchmarks to Head pose, Facial expression, Age, Style, and video datasets. The Table (**in the supplementary material**) summarizes all used datasets for training and validating facial GANs. We present a detailed study of each dataset; by giving the number of samples, subjects, and types of data. Facial GANs were trained on large datasets collected from the web or from the combination of many benchmarks. In addition, we present in Fig. 9 the datasets used for facial GANs validation for different semantics with the number of use of each one. We notice that most pose, facial expression, age, and style GANs work on static and 2D data. However, face reenactment and video generating GANs, are trained and validated using dynamic datasets. In the following, we present in Tables 3 and 4, and Fig. 8 the results of facial GANs performance in the corresponding datasets. The large variety of facial GAN generation applications makes it complex to compare the different state-of-the models similarly. For example, the cross-pose models are designed to generate frontal or profile faces with respect to the target pose and pose constraints such as occlusion, symmetry and identity preservation. However, the cross-expression architectures are designed to generate expressions with impaired datasets or limited paired data. The cross-style models are often designed to disentangle the style features from identity features and generate faces from the feature vectors obtained from the projection in the latent space. Because of this heterogeneity, we compare the facial GAN models depending on the application and the variation factors of evaluation and training mode. In Fig. 8, the performance of each pose, age group and mean of expressions is presented in the following histograms. However, in table 3, we present a comparison of the facial GANs basing on the most used qualitative evaluation metrics and datasets. While table 4 presents the accuracy results of GANs for each application and datasets.

5.2 Qualitative evaluation of facial GANs

The qualitative evaluation of facial GANs consists in evaluating the quality of the generated face images. Actually, the quality includes the resolution of generated faces as well as the lack of blemish like artifacts and unrealistic patches in the faces. In this paragraph, we present first the description of the existent metrics. Then, we give a comparative study based on the commonly used ones [132] in Table 3. Up arrows present the metrics which higher value corresponds to better performance, while down arrows present whose lower value corresponds to better performance. The best state-of-the-art performance for each metric is presented in bold.

The Frechet Inception Distance **(FID)** [54] uses inception-v3 network to extract feature embedding of real and fake images. The distributions are fitted to multivariate normal distributions. The FID score looks at the mean and variance to compare between the real and fake distributions [15, 60]. Lower FID score means the real and fake distributions are closer and so the features are more similar to each other.

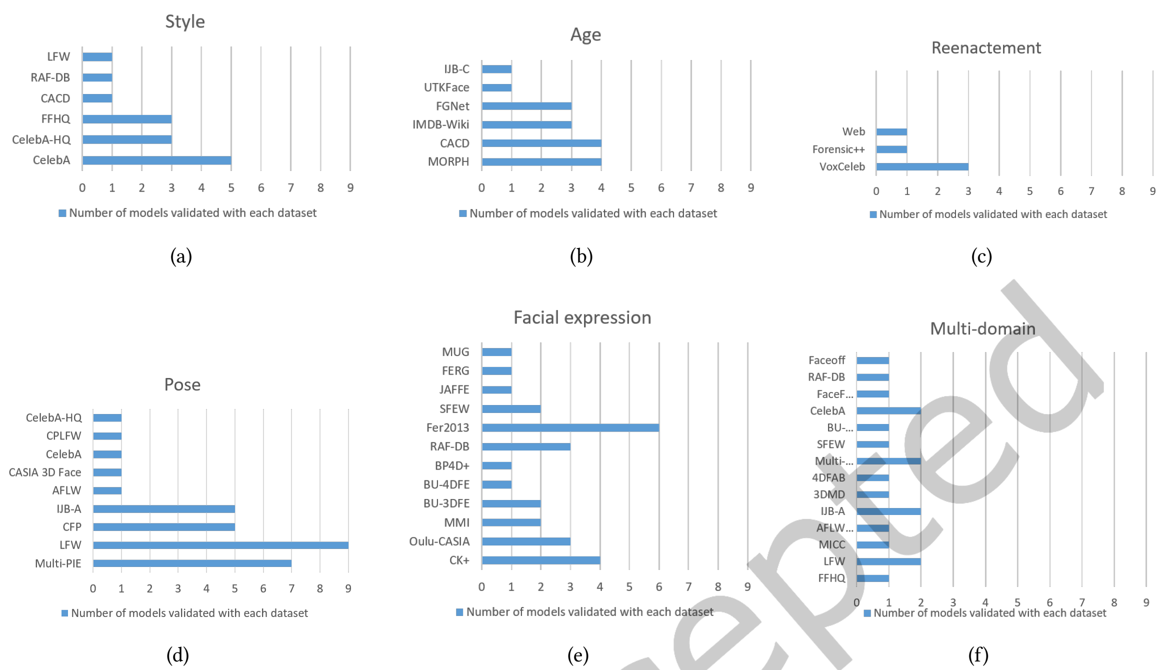


Fig. 9. Validation datasets of facial GANs: each dataset is presented with the number of use. (a) Style, (b) Age, (c) Reenactment/video, (d) Pose, (e) Facial expression, (f) Multi-domain

Kernel inception distance **(KID)** [16] evaluates the discrepancy between distributions by computing the squared Maximum Mean Discrepancy (MMD) between Inception representations. Makhmudkhujaev et al. [95] computed KID between real and generated distributions for their respective age-group. Lower KID scores indicate that input images are better regenerated, and thus the model has the capability of preserving the identity.

Perceptual path length **(PPL)** [66], is a metric which estimates the quality of the latent space. It estimates the smoothness of the mapping from a latent space to the output image by measuring the average LPIPS distances [159] between generated images under small perturbations in latent space. [66], explained the correlation between the generated image quality and perceptual path length. Actually, during training, the discriminator penalizes bad images; so that stretching the region of the latent space that produces good images can help the generator to improve. Smaller PPL correlates with higher image quality.

Structural Similarity Index **(SSIM)** [152], measures the low-level similarity to the ground truth images. Higher SSIM corresponds to better similarity between the generated images and the ground truth ones.

Multi-scale structural similarity **(MS-SSIM)** [33, 105] is an improved version of SSIM conducted over multiple scales through a process of multiple stages of subsampling. MS-SSIM values range between 0.0 and 1.0 and higher MS-SSIM values correspond to perceptually more similar images.

The Inception Score **(IS)** was proposed by Tim Salimans et al. [117] to remove the subjective human evaluation of images. IS uses the pretrained Inception v3 for image classification to classify the generated images. The score captures two properties of the generated images; image quality and diversity. The inception score ranges between 1.0 as the lowest value and as the highest value, the number of classes supported by the classification model.The Inception scores specifically uses KL divergence, which captures fidelity from difference in the conditional label distribution and diversity from the marginal label distribution.The Inception Score sums over all images and averages over all the classes. A high Inception Score means the entropy of the conditional probability distribution is low, finding relevant objects and features, while the marginal probability distribution is high, when capturing a diverse set of features. However, a low score means both distributions either have low entropy, having high peaks (no diversity) or both distributions have high entropy (no clear features found). The main drawback of Inception score [13] is the mode collapse, when GAN model generates one single image for each class which give perfect IS when the model can do better. In addition, IS look only at the generated samples and doesn't compare them with real images. Moreover, the Inception score is trained on ImageNet dataset and may do not give relevant features with some semantics of faces unknown like head pose. The inception score doesn't focus on the spatial relationships of features like nose or eyes position. Inception Score is usually used for evaluating conditional GANs but is actually replaced by FID which is more relevant.

Cosine Similarity **(CSIM)** [37] is used to calculate the similarity between two embedding vectors. Pre-trained face recognition networks are applied to obtain embeddings from the source and driving images. Higher score indicates that the source identity is better preserved in the reenactment process.

Peak Signal to Noise Ratio **(PSNR)** [48, 157] measures the ratio between the maximum possible value called the power of a signal and the power of the distorted noise that affects the quality of its representation. This ratio measures the quality between the original and a reconstructed image. The higher PSNR, corresponds to better quality of the reconstructed image.

Some models proposed other evaluation metrics for their models. [167] defined Human Eye Perceptual Evaluation **(HYPE)**, as human benchmark for generative realism. Two variants are introduced: $HYPE_{time}$ measures visual perception under adaptive time constraints to determine the threshold at which a model's outputs appear real, and $HYPE_{\infty}$ measures the human error rate on fake and real images without time constraints. Huang et al. [60] introduced FakeLokator is introduced to solve the GAN based fake localization problem on facial images.

We can notice from Table 3. That there are few Facial GANs models that evaluate the quality of the generated faces. In addition, there are no common metrics used by all models, which can give clearer comparison between all existent approaches.

5.3 Quantitative evaluation of facial GANs

Despite the lack of quality evaluation of the generated images, most of the facial GANs presented a quantitative evaluation of their models. Actually, they train classification networks with images generated by GAN models, and then evaluate their performance on a test set composed of real-world images. In this survey, we gathered some works that used common benchmarks for evaluating the performance of their models. Table 4 presents the quantitative evaluation score namely, accuracy(ACC), area under ROC curve (AUC) and probability sum of facial GANs cross-pose, facial expressions, age, style, and video. Best state-of-the-art performance are presented in bold.

6 GANS PROBLEMS AND PROPOSED SOLUTIONS

Despite the power of facial GANs, they suffer from common problems related to training stability and output quality. In this section, we present these problems, namely, mode collapse, vanishing gradients, convergence failure, generated face quality, and problems related to the training modes and the semantics. In addition, we review the proposed state-of-the-art regularization and normalization approaches [67, 74, 106, 108] for tackling them.

Table 3. Qualitative evaluation of generated faces

Model	Database	FID↓	PPL↓	SSIM↑	MS-SSIM↑	IS↑	CSIM↑	PSNR↑
Zakharov et al. [152]	VoxCeleb1	29.5		0.74			0.19	
X2Face [152]	VoxCeleb1	56.5		0.75			0.18	
Pix2pixHD [152]	VoxCeleb1	24		0.70			0.16	
Style-GAN [67]	FFHQ	4.4	212.1					
Style-GAN2 [67]	FFHQ	**2.84**	145					
HDC-GAN [33]	CelebA	8.44			0.1978			
Style-editing-GAN [31]	FFHQ	5.4						
Faster-GAN [48]	CelebA-HQ			0.83				**31.47**
Match-GAN [124]	CelebA	5.58				**3.07**		
	RaF-DB	5.06				1.58		
Mask-guided-GAN [47]	Helen	8.92						
Deng et al. [36]	FFHQ	12.9	**123**					
Mesh-GAN-ID [28]	3dMD+4DFAB	10.82						
Face-controller [144]	FaceForensics++	3.51						
Pro-GANs [65]	FFHQ	8.04			0.28			
MSG-StyleGAN [64]	CelebA-HQ	6.37						
	FFHQ	5.8						
MSG-ProGAN [64]	CelebA-HQ	8.02						
	FFHQ	8.36						
SS-GAN [123]	CelebA			0.217				
SA-GAN [123]	CelebA			0.167				
SC-GAN [123]	CelebA			0.143				
AC-GAN [123]	CelebA			0.076				
c-GAN [123]	CelebA			0.497				
Conditional -CycleGAN [91]	CelebA			**0.92**				
Ic-GAN [91]	CelebA			0.74				
Disco-GAN [91]	CelebA			0.87				
Face-GAN [130]	IJB-C+Forensic++						**0.747**	
ST-GAN [157]	Celeb-A				0.77 ± 0.01			28.11 ± 0.01

6.1 Mode collapse

The main purpose of GAN is to produce a diversity of outputs which allows data augmentation and reduce the overfitting of deep neural networks. However, the generator usually fails to learn the full distribution of datasets, which is famous for mode collapse or partial mode convergence. Actually, when the generator learns to output samples from single or partial distributions, it gets used to produce similar outputs that seem most plausible to the discriminator. However, the discriminator learns to reject all outputs except the best sample produced. This phenomenon is caused when the generator and the discriminator forget some distribution modes and fail to achieve an equilibrium between them during training. The generator learns only a part of the inputs and no longer generate samples from one part of the space, while the discriminator is used to classify the inputs as fakes and forget them later when they are no longer generated. Therefore, one of the players can get stack in local optimum in gradient descent. Generative Adversarial Network Games (GANGs) [106] inspired from game theoretic techniques introduced Resource Bounded Nash equilibrium (RB-NE) to escape the local optima of gradient descent and provide convergence grantees for methods with random restarts by approximating the best response computations. Some researches [7, 40, 50, 92, 108] focused on improving the loss functions for more

Table 4. Quantitative evaluation of generated faces

Semantic	Database	Model	Performance								
Head Pose			ACC%		AUC%						
	LFW	FF-GAN [115]	96,42		99.45						
		DR-GAN [115]	96.9		99.6						
		CAPG-GAN [58]	99.37		99.9						
		FI-GAN [115]	98.3		99.6						
		3D-GAN [98]	**99.85**								
		PCCycle-GAN [88]	99.52		**99.93**						
	CFP	UV-GAN [34]	94.05								
			FF-ACC%	FF-AUC	FP-ACC	FP-AUC					
		DR-GAN [115]	97.1	91							
		Alqahtani et al. [3]	98.23 pm 0.83	91.92 ± 0.59							
		FI-GAN [115]	98.9	94.2	99.7	98.9					
		3D-GAN [52]	**99.9**	**99.85**							
	Multi-PIE		ACC%								
			90°	75°	60°	45°	30°	15°	0°		
		FI-GAN [115]	**77.03**	**88.24**	**96.22**	97.37	98.52	98.8			
		DR-GAN [115]			83.2	86.2	90.1	94			
		FF-GAN [115]	61.2	77.2	85.2	89.7	92.5	96.6			
		TP-GAN [115]	64.64	77.43	87.72	95.38	98.06	98.68			
		CAPG-GAN [115]	66.05	83.05	90.63	97.33	**99.56**	**99.82**			
		Lai et al. [75]	70.05	88	90.5	89.5	92.5	93	92		
		PCCycle-GAN [88]	64.5	82.7	92.3	**97.4**	99.3	99.4			
Facial Expression			ACC%								
			Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral	Total	
	Fer2013	CGAN (Yi et al) [148]	63	52	41	85	50	75	64	65.5	
		cMWGAN [160]								72.9	
		De-expression [57]								**75.84**	
	RAF-DB	EAU-Net [35]	65.5	45.6	58.7	96.3	74.5	80.2	77.5	71.18	
		StackGAN-A [80]	72			76	89			79	
		StackGAN-B [80]	86			84	86			85	
		De-expression [57]								**87.32**	
	BU-3DFE	IA-gen [147]	70	75	71	87.5	65	97.5		76.83	
		DeRL [145]								**84.17**	
	BU-4DFE	IA-gen [147]	82.5	95	84.5	95	89	92.5		89.55	
	BP4D+	DeRL [145]									
	Oulu-CASIA	IA-gen [147]	83	95.5	83.5	94.5	89	93		**88.92**	
		DeRL [145]								88	
	CK+	IA-gen [147]	96	100	99	100	97.5	100		96.57	
		De-expression [57]								**99.38**	
		DeRL [145]								97.3	
	MMI(6 classes)	DeRL [145]								73.23	
Age			ACC%								
			4+ ->10+	8+ ->30+	10+ ->20+	10+ ->40+	20+ ->60+	30+ ->60+	40+ ->60+	Average	
	UTKFace+CACD	CAAE [122]	57	51	51	51	65	45	70	56	
	+MORPH+FGNet	c-GAN [122]	47	79	70	49	76	65	75	66	
	+IMDB-	Dual-cGAN [122]	86	82	81	80	80	72	84	**81**	
Style	wiki		ACC%								
			Smile	Pale skin	Glasses	Gender/Age	Bangs	Bald	Moustache	Hair	Entire
	CelebA/CelebA-HQ	Ic-GAN [48]						**99.95**			
		Star-GAN [80]	90.56	96.56	97.08	90.21/-	96		89.87	95.48	
		ST-GAN [60]	99.85		99.7		**99.85**	99.85	**99.35**	99.85	
		Star-GAN [60]				**99.9/99.9**				**99.9**	
		Att-GAN [60]	99.85		**99.85**			99.85			
		PG-GAN [60]									**97.75**
		Style-GAN [60]	**99.9**			99.9/99.85					92.1
		Style-GAN2 [60]									85.5
			Probability Sum								
		Ic-GAN [24]	2.33	2	0	1.33/0.33					
		Cycle-GAN [24]	21.33	37	28	35/20					
		Star-GAN [24]	19	36.67	30.33	9.67/17.67					
		TDB-GAN [24]	57.33	24.33	41.67	54/62					
Video			ACC%								
	VoxCeleb	X2Face [129]	**64.76**								
		ICFace [129]	62.86								

stable training, which make the real/fake classification softer and help to avoid generation stack. In fact, the earlier version of GANs using min max evaluation loss, suffers from the hard evaluation of the discriminator which rejects almost samples without giving a relative score to the generator to know in which direction it can go to improve its performance.

6.2 Vanishing gradients

Perfect GANs aim to generate samples as real as possible, which means making the generated distribution close to the real one. This is done by minimizing the corresponding cost function, which measures the difference between both distributions. The first GANs suffer from vanishing gradients caused by the use of binary cross-entropy loss. Actually, when the discriminator starts learning faster than the generator, it can easily differentiate between real and fake samples, so that reject almost generated samples. When the discriminator becomes too perfectionist, it becomes unhelpful to the generator that needs informative feedback to know in which direction may improve. The vanishing gradients make learning slow and sometimes stops. Wiatrak et al. [140] studied the cause of vanishing gradients and convergence failure and mentioned that the loss function of the original GAN leads to the oscillating and cyclical behavior, especially when the distributions of generated and target data are disjoint.

Some researches proposed to replace BCE loss with least squares loss, while others proposed to inject continuously noise to the discriminator input to smooth the distribution of probability mass.

6.3 Convergence failure

Similar to mode collapse, convergence failure is a common problem in generative adversarial networks, caused by the imbalance between the generator and the discriminator. When the discriminator progresses faster, and its loss is close to zero, while the generator loss may not settle down and continue to rise. This happens when the generated samples are rejected easily by the discriminator [12, 140]. Many researches [6, 73, 108, 116] were proposed to stabilize training to avoid the convergence failure. Actually, [6, 73] proposed adding noise to the discriminator outputs to smooth the probability distribution. Roth et al. [116] demonstrated that this can be a costly solution, because it demands a high-dimensional noise which cause significant variance in the parameter estimation process and needs a lot of samples. Instead, the research proposed a noise-induced regularization solution, which consists in a weighted penalty function on the norm of the gradients of the discriminator rather than being a part of the generative process.

Lucic et al. [92] had compared FID and training failures of different models on CELEBA dataset. Some models indicates the presence of significant outlier runs, usually severe mode collapses or training failures (up to 20% failures). In addition, FID score can be improved from 87.5 to 30 depending on the architecture and the additional optimizations.

6.4 Image quality

Most of GANs focus is to produce high quality outputs. The first GAN architecture [46] based on fully connected layers had limited capacity to generate images with high resolution. Many improvements were proposed for this purpose. Some works [58, 66, 75, 107] proposed improved loss functions like WGAN-GP, pixel-level, and feature level losses to penalize the quality shift. In addition, the use of an Auto-encoder architecture for the discriminator [14] that compares between real and generated images in the pixel level, which allows better generation quality. Some other works [33, 65, 113] focused on the architecture blocks. DC-GAN [113] proposed deconvolutional and upsampling layers for learning better resolution, Pro-GAN [65] proposed a progressive training process to pass from low to high resolution. However, Checkerboard artifacts can appear in generated images when passing from low to high resolution via transposed convolution (Fig. 10). In addition to 2D facial GANs, 3D GANs [165] also suffers from big performance drop for profile and near-profile faces due to severe

Fig. 10. Examples of artifacts appeared in generated faces taken from [67]: (a) Droplet -like artifacts caused by Instance normalization in Style-GAN images. (b) Artifacts caused by Progressive growing. In this example, the teeth do not follow the pose but stay aligned to the camera, as indicated by the blue line.

texture loss and artifacts. The improvement of loss function can avoid the blurry effect, alleviate artifacts and produce visually pleasing results. Arora et al. [8] proposed MIX+DCGAN model which consists of a mixture of such DCGANs so that it has 3 generators and 3 discriminators. The improvement of architecture generates more faithful and more diverse samples from a face database than the baseline. Karras et al. [67] showed that the improvement of Style-GAN to Style-GAN2 by improving the architecture and loss function improved the model recall from 0.399 to 0.492, FID from 4.4 to 2.84 and PPL from 212.1 to 145. This normalization improved the quality of the generated images and reduced face artifacts. StyleGAN2 [67] proposes an alternative design for solving issues that occurred in using progressive growing which its purpose is to stabilize high-resolution training with great depth using skip connection design and residual concepts similar to ResNet.

6.5 Problems related to the training mode

The main problems related to the training facial GANs modes are the availability of data and resources. The supervised architectures are based on labelled and/or paired datasets. However, these datasets are expensive and difficult to collect. Recently, the limited labelled face datasets pushed researchers [31, 48, 66, 67, 112] to propose unsupervised GANs trained on big face datasets which are gourmand to computation requirements. Additionally, others researches [17, 42, 107] proposed semi-supervised architectures which need small labelled datasets in additional to from the web datasets. These models reduce the problem of dataset availability. But, they are generally complex and need sophisticated computational resources.

6.6 Problems related to face semantic

In addition to the common GANs and facial GANs problems, There are specific problems related to each face semantic. The main concerns and challenges include:

- Pose generation with large feature variations. It consists of generating extreme profile face from a frontal face and vice versa. 3D face and dynamic face generation are also challenging problems. The large pose variation can disturb the training process which affects the face texture and realism, as well as the identity preservation
- Facial expression generation with impaired datasets. It consists of generating faces in different expressions. Contrarily to the pose generation, the facial expression variation is smaller. Nevertheless, most available datasets are impaired. The recent models are based on image-to-image translation which transfer the common feature of target domain to the source input. This affects the realism of the generated face in different expression which preserve common expression features and not those related to the identity.

- Age generation with continuous functions. With limited age datasets of persons in specific age groups, it is important to learn the distributions of all age groups to generate face in the desired age.
- Style generation by learning the disentanglement of style features. Most of style based GANs are based on transformations in the latent space. This is based on learning the disentanglement of style features from identity features. And the entanglements between different styles, and mixing styles. For example, some styles are common to gender or age. In addition, many aspects in human portraits can be adopted as stochastic [66], such as the exact placement of hairs, stubble, freckles, or skin pores.

7 DISCUSSION AND FUTURE DIRECTIONS

We present in this article, a detailed study of facial GANs, in which we present a full taxonomy of existent models. Actually, we present a taxonomy based on architecture, training mode and facial semantic. We notice that conditional GANs are easy to implement using the semantic code. However, they suffer from unstable training and artifacts in the generated images. Most of GANs for pose generation used conditional GANs thanks to the availability of paired datasets. While most of the facial expression generation, GANs used impaired image-to-image translation models because of the limited paired data. Therefore, the accuracy scores are still poor compared to paired datasets. Fig. 8 illustrates the difference in facial expression accuracy using an impaired database (41%-65.5% in fer2013), compared to paired database (95.8%-99.38% in CK+). Furthermore, most of the recent GANs used controllable architectures thanks to their power to generate better images. However, they suffer from important computational requirements. Moreover, we present the existent models in each semantic and with the corresponding training mode(7). We notice that few works used architectures which jointly generate faces in two or more face variations. However, there are, as far as our knowledge, no researches joining aging faces with pose facial expression and style generation. In addition, few works [17, 24, 53, 120] applied control on face to generate faces in mixed styles. In fact, almost all works are tested on uni-style face generation. Furthermore, most of existent researches focused on facial image generation. However, few works generated 3D faces or faces in reenactment. These studies did not optimize the generated faces in terms of photo realism quality. Additionally, we present a quantitative and qualitative evaluation of facial GANs using common benchmarks and metrics. We notice that not all the existent models do not use common qualitative evaluation metrics which give a clear qualitative comparison between them. Yet, we try to gather in table 3, the common evaluation metrics to have better based on state-of-the-art scores. This can be an issue for the future of facial GANs; the evaluation of the generation quality using a common metric. On top of the problem of computation requirement still presents an important problem of facial GANs, in terms of datasets and GPUs, which makes it challenging to propose new unsupervised models with less requirements.

Despite the important progress in facial GANs, face generation in different semantics and using different training models still suffer from some weaknesses. We propose the following future directions for better face generation:

- Computation requirements: Recent facial GANs focus on improving the realism of the generated faces. Face generation acquires tens of thousands of training samples, and with datasets including a lot of intrinsic variation [67]. Therefore, it is crucial to find new techniques to reduce the training data requirements.
- Limited large paired datasets: Domain-to-domain translation is one of the same focus of face generation for different semantics. However, the use of supervised setting needs paired data. And, the availability of the paired data is either difficult or expensive. Only few datasets are available for some semantics, like pose, but they are either small or expensive. For other semantics, with the lack data, unsupervised or semi-supervised approaches are proposed, which is more challenging because faces are not available for the corresponding face domain. These models need more regularization for identity preservation and reducing artifacts.

- Quality of generated faces: Artifacts in generated face images and perceptual quality distortion can be caused by many factors [111] such as losses, unstable training process, and network architectures. Some solutions are proposed recently to alleviate this problem, such improving [111] or regularizing [156] the network structure or incorporating additional objective functions, or attention mechanism to capture better global structure. However, in some face variation like extreme profile face normalization, the existent GANs based solutions still suffer from imperfections in face texture. This explains the need to new GANs architectures more efficient to learn and generate extreme profile faces
- Multi-semantic generation: A good multi-domain image generation model [97] should be able to generate pairs of corresponding images which share common semantics but are of different domain-specific semantics. Most of multi-domain facial GANs are able to generate faces in only two domains. However, there are no models which generate faces in all possible domains or with the option of choosing one or many domains. Actually, the generation of faces in a specific domain is done either by learning a conditional distribution encoded in the conditioned domain variables or by unsupervised domain adaptation. This is challenging because of the absence of paired and large labelled datasets in some domains. Therefore, future facial GANs should be able to learn all-over face multi-domain variations in unsupervised or semi-supervised mode to be functional with the state-of-art benchmarks.
- Common evaluation metrics: Despite the variety of evaluation metrics used for evaluating the generated faces by GANs, there is no common convention on facial GANs evaluation. In this survey, we classify the existent metrics on qualitative and qualitative evaluation metrics. Nevertheless, in order to compare precisely between existent architecture, there is an important need to an experimental comparison using common metrics and datasets.

8 CONCLUSION

This article has presented a survey of the existing GAN models for face generation and facial semantics. A comparative analysis is accorded of different facial GANs. The proposed taxonomy of facial GANs and GANs for face generation can be a reference for better understanding of the existent solutions as well as the evolution of architecture cross every face variation. The present article focuses on face generation as the focal attention of GANs. Despite the important number of researches in this domain, there are never reviewed before. In this research, we presented the main improvement of GANs architecture and loss functions applied in the specific context of face generation and synthesis. The GAN based approaches are presented regarding their applications and are grouped into categories to better identify and differentiate them. The used datasets to generate faces using GANs are presented and listed. Besides, we discussed several open issues and pointed out different challenges that require research efforts to improve, like the GANs evaluation problem and quality comparison between existent solutions.

REFERENCES

[1] Mahla Abdolahnejad and Peter Xiaoping Liu. 2020. Deep learning for face image synthesis and semantic manipulations: a review and future perspectives. *Artificial Intelligence Review* 53, 8 (2020), 5847–5880.

[2] Alankrita Aggarwal, Mamta Mittal, and Gopi Battineni. 2021. Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights* (2021), 100004.

[3] Hamed Alqahtani and Manolya Kavakli-Thorne. 2019. Adversarial disentanglement using latent classifier for pose-independent representation. In *International Conference on Image Analysis and Processing (ICIAP).*

[4] Hamed Alqahtani, Manolya Kavakli-Thorne, and Gulshan Kumar. 2019. Applications of generative adversarial networks (gans): An updated review. *Archives of Computational Methods in Engineering* (2019), 1–28.

[5] Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. 2017. Face aging with conditional generative adversarial networks. In *2017 IEEE international conference on image processing (ICIP).* IEEE, 2089–2093.

[6] Martin Arjovsky and Léon Bottou. 2017. Towards principled methods for training generative adversarial networks. *arXiv preprint arXiv:1701.04862* (2017).

[7] Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein generative adversarial networks. In *International conference on machine learning*. PMLR, 214–223.

[8] Sanjeev Arora, Rong Ge, Yingyu Liang, Tengyu Ma, and Yi Zhang. 2017. Generalization and equilibrium in generative adversarial nets (gans). In *International Conference on Machine Learning*. PMLR, 224–232.

[9] Sandipan Banerjee, Walter Scheirer, Kevin Bowyer, and Patrick Flynn. 2020. On hallucinating context and background pixels from a face mask using multi-scale gans. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 300–309.

[10] Aayush Bansal, Shugao Ma, Deva Ramanan, and Yaser Sheikh. 2018. Recycle-gan: Unsupervised video retargeting. In *Proceedings of the European conference on computer vision (ECCV)*. 119–135.

[11] Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, and Gang Hua. 2018. Towards open-set identity preserving face synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 6713–6722.

[12] Samuel A Barnett. 2018. Convergence problems with generative adversarial networks (gans). *arXiv preprint arXiv:1806.11382* (2018).

[13] Shane Barratt and Rishi Sharma. 2018. A note on the inception score. *arXiv preprint arXiv:1801.01973* (2018).

[14] David Berthelot, Thomas Schumm, and Luke Metz. 2017. Began: Boundary equilibrium generative adversarial networks. *arXiv preprint arXiv:1703.10717* (2017).

[15] Binod Bhattarai and Tae-Kyun Kim. 2020. Inducing optimal attribute representations for conditional GANs. In *European Conference on Computer Vision*. Springer, 69–85.

[16] Mikołaj Bińkowski, Danica J Sutherland, Michael Arbel, and Arthur Gretton. 2018. Demystifying mmd gans. *arXiv preprint arXiv:1801.01401* (2018).

[17] Navaneeth Bodla, Gang Hua, and Rama Chellappa. 2018. Semi-supervised FusedGAN for conditional image generation. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 669–683.

[18] Ali Borji. 2019. Pros and cons of gan evaluation measures. *Computer Vision and Image Understanding* 179 (2019), 41–65.

[19] Andrew Brock, Jeff Donahue, and Karen Simonyan. 2018. Large scale GAN training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096* (2018).

[20] Anpei Chen, Ruiyang Liu, Ling Xie, Zhang Chen, Hao Su, and Jingyi Yu. 2020. Sofgan: A portrait image generator with dynamic styling. *arXiv preprint arXiv:2007.03780* (2020).

[21] Chaofeng Chen, Xiaoming Li, Lingbo Yang, Xianhui Lin, Lei Zhang, and Kwan-Yee K Wong. 2020. Progressive semantic-aware style transformation for blind face restoration. *arXiv preprint arXiv:2009.08709* (2020).

[22] Jiawei Chen, Janusz Konrad, and Prakash Ishwar. 2018. Vgan-based image representation learning for privacy-preserving facial expression recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1570–1579.

[23] Qifeng Chen and Vladlen Koltun. 2017. Photographic image synthesis with cascaded refinement networks. In *Proceedings of the IEEE international conference on computer vision*. 1511–1520.

[24] WenTing Chen, Xinpeng Xie, Xi Jia, and Linlin Shen. 2018. Texture deformation based generative adversarial networks for face editing. *arXiv preprint arXiv:1812.09832* (2018).

[25] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. 2016. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. *arXiv preprint arXiv:1606.03657* (2016).

[26] Yizhen Chen and Haifeng Hu. 2019. An improved method for semantic image inpainting with gans: Progressive inpainting. *Neural Processing Letters* 49, 3 (2019), 1355–1367.

[27] Yu-An Chen, Wei-Che Chen, Chia-Po Wei, and Yu-Chiang Frank Wang. 2017. Occlusion-aware face inpainting via generative adversarial networks. In *2017 IEEE International conference on image processing (ICIP)*. IEEE, 1202–1206.

[28] Shiyang Cheng, Michael Bronstein, Yuxiang Zhou, Irene Kotsia, Maja Pantic, and Stefanos Zafeiriou. 2019. Meshgan: Non-linear 3d morphable models of faces. *arXiv preprint arXiv:1903.10384* (2019).

[29] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. 2018. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 8789–8797.

[30] Grigorios G Chrysos, Paolo Favaro, and Stefanos Zafeiriou. 2019. Motion deblurring of faces. *International journal of computer vision* 127, 6 (2019), 801–823.

[31] Edo Collins, Raja Bala, Bob Price, and Sabine Susstrunk. 2020. Editing in style: Uncovering the local semantics of gans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5771–5780.

[32] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A Bharath. 2018. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine* 35, 1 (2018), 53–65.

[33] Joachim D Curtó, Irene C Zarza, Fernando De La Torre, Irwin King, and Michael R Lyu. 2017. High-resolution deep convolutional generative adversarial networks. *arXiv preprint arXiv:1711.06491* (2017).

[34] Jiankang Deng, Shiyang Cheng, Niannan Xue, Yuxiang Zhou, and Stefanos Zafeiriou. 2018. Uv-gan: Adversarial facial uv map completion for pose-invariant face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7093–7102.

[35] Jia Deng, Gaoyang Pang, Zhiyu Zhang, Zhibo Pang, Huayong Yang, and Geng Yang. 2019. cGAN based facial expression recognition for human-robot interaction. *IEEE Access* 7 (2019), 9848–9859.

[36] Yu Deng, Jiaolong Yang, Dong Chen, Fang Wen, and Xin Tong. 2020. Disentangled and controllable face image generation via 3d imitative-contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5154–5163.

[37] Sourabh Dhere, Suresh B Rathod, Sanket Aarankalle, Yash Lad, and Megh Gandhi. 2020. A Review on Face Reenactment Techniques. In *2020 International Conference on Industry 4.0 Technology (I4Tech)*. IEEE, 191–194.

[38] Chris Donahue, Zachary C Lipton, Akshay Balsubramani, and Julian McAuley. 2017. Semantically decomposing the latent spaces of generative adversarial networks. *arXiv preprint arXiv:1705.07904* (2017).

[39] Shiqing Fan and Ye Luo. 2021. Deblurring Processor for Motion-Blurred Faces Based on Generative Adversarial Networks. *arXiv preprint arXiv:2103.02121* (2021).

[40] Charlie Frogner, Chiyuan Zhang, Hossein Mobahi, Mauricio Araya-Polo, and Tomaso Poggio. 2015. Learning with a Wasserstein loss. *arXiv preprint arXiv:1506.05439* (2015).

[41] Ruohan Gao and Kristen Grauman. 2017. On-demand learning for deep image restoration. In *Proceedings of the IEEE International Conference on Computer Vision*. 1086–1095.

[42] Zhongpai Gao, Juyong Zhang, Yudong Guo, Chao Ma, Guangtao Zhai, and Xiaokang Yang. 2020. Semi-Supervised 3D Face Representation Learning From Unconstrained Photo Collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 348–349.

[43] Jon Gauthier. 2014. Conditional generative adversarial nets for convolutional face generation. *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester* 2014, 5 (2014), 2.

[44] Shiming Ge, Chenyu Li, Shengwei Zhao, and Dan Zeng. 2020. Occluded face recognition in the wild by identity-diversity inpainting. *IEEE Transactions on Circuits and Systems for Video Technology* 30, 10 (2020), 3387–3397.

[45] Baris Gecer, Binod Bhattarai, Josef Kittler, and Tae-Kyun Kim. 2018. Semi-supervised adversarial learning to generate photorealistic face images of new identities from 3d morphable model. In *Proceedings of the European conference on computer vision (ECCV)*. 217–234.

[46] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial networks. *arXiv preprint arXiv:1406.2661* (2014).

[47] Shuyang Gu, Jianmin Bao, Hao Yang, Dong Chen, Fang Wen, and Lu Yuan. 2019. Mask-guided portrait editing with conditional gans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3436–3445.

[48] Shanyan Guan, Ying Tai, Bingbing Ni, Feida Zhu, Feiyue Huang, and Xiaokang Yang. 2020. Collaborative learning for faster stylegan embedding. *arXiv preprint arXiv:2007.01758* (2020).

[49] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. 2020. A review on generative adversarial networks: Algorithms, theory, and applications. *arXiv preprint arXiv:2001.06937* (2020).

[50] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. 2017. Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028* (2017).

[51] Guodong Guo and Na Zhang. 2019. A survey on deep learning based face recognition. *Computer vision and image understanding* 189 (2019), 102805.

[52] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. 2020. Ganspace: Discovering interpretable gan controls. *arXiv preprint arXiv:2004.02546* (2020).

[53] Zhenliang He, Wangmeng Zuo, Meina Kan, Shiguang Shan, and Xilin Chen. 2019. Attgan: Facial attribute editing by only changing what you want. *IEEE Transactions on Image Processing* 28, 11 (2019), 5464–5478.

[54] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *arXiv preprint arXiv:1706.08500* (2017).

[55] Saifuddin Hitawala. 2018. Comparative study on generative adversarial networks. *arXiv preprint arXiv:1801.04271* (2018).

[56] Yongjun Hong, Uiwon Hwang, Jaeyoon Yoo, and Sungroh Yoon. 2019. How generative adversarial networks and their variants work: An overview. *ACM Computing Surveys (CSUR)* 52, 1 (2019), 1–43.

[57] Jiahui Hu, Bing Yu, Yun Yang, and Bailan Feng. 2019. Towards facial de-expression and expression recognition in the wild. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 157–163.

[58] Yibo Hu, Xiang Wu, Bing Yu, Ran He, and Zhenan Sun. 2018. Pose-guided photorealistic face rotation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 8398–8406.

[59] Rui Huang, Shu Zhang, Tianyu Li, and Ran He. 2017. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In *Proceedings of the IEEE international conference on computer vision*. 2439–2448.

[60] Yihao Huang, Felix Juefei-Xu, Run Wang, Xiaofei Xie, Lei Ma, Jianwen Li, Weikai Miao, Yang Liu, and Geguang Pu. 2020. FakeLocator: Robust Localization of GAN-Based Face Manipulations via Semantic Segmentation Networks with Bells and Whistles. *arXiv preprint arXiv:2001.09598* (2020).

[61] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1125–1134.
[62] Abdul Jabbar, Xi Li, and Bourahla Omar. 2020. A Survey on Generative Adversarial Networks: Variants, Applications, and Training. *arXiv preprint arXiv:2006.05132* (2020).
[63] Xiaohan Jin, Ye Qi, and Shangxuan Wu. 2017. Cyclegan face-off. *arXiv preprint arXiv:1712.03451* (2017).
[64] Animesh Karnewar and Oliver Wang. 2020. Msg-gan: Multi-scale gradients for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7799–7808.
[65] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2017. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* (2017).
[66] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4401–4410.
[67] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8110–8119.
[68] Hadi Kazemi, Fariborz Taherkhani, and Nasser M Nasrabadi. 2018. Unsupervised facial geometry learning for sketch to photo synthesis. In *2018 international conference of the biometrics special interest group (BIOSIG)*. IEEE, 1–5.
[69] Youngsung Kim, ByungIn Yoo, Youngjun Kwak, Changkyu Choi, and Junmo Kim. 2017. Deep generative-contrastive networks for facial expression recognition. *arXiv preprint arXiv:1703.07140* (2017).
[70] Diederik P Kingma and Prafulla Dhariwal. 2018. Glow: Generative flow with invertible 1x1 convolutions. *arXiv preprint arXiv:1807.03039* (2018).
[71] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).
[72] Ivan Kobyzev, Simon Prince, and Marcus Brubaker. 2020. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).
[73] Naveen Kodali, Jacob Abernethy, James Hays, and Zsolt Kira. 2017. On convergence and stability of gans. *arXiv preprint arXiv:1705.07215* (2017).
[74] Karol Kurach, Mario Lučić, Xiaohua Zhai, Marcin Michalski, and Sylvain Gelly. 2019. A large-scale study on regularization and normalization in GANs. In *International Conference on Machine Learning*. PMLR, 3581–3590.
[75] Ying-Hsiu Lai and Shang-Hong Lai. 2018. Emotion-preserving representation learning via generative adversarial network for multi-view facial expression recognition. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 263–270.
[76] Alexandros Lattas, Stylianos Moschoglou, Baris Gecer, Stylianos Ploumpis, Vasileios Triantafyllou, Abhijeet Ghosh, and Stefanos Zafeiriou. 2020. AvatarMe: Realistically Renderable 3D Facial Reconstruction" In-the-Wild". In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 760–769.
[77] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4681–4690.
[78] Soochan Lee, Junsoo Ha, and Gunhee Kim. 2019. Harmonizing maximum likelihood with GANs for multimodal conditional generation. *arXiv preprint arXiv:1902.09225* (2019).
[79] Shan Li and Weihong Deng. 2020. Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing* (2020).
[80] Xiaoyi Li and Xiaowen Yu. [n.d.]. Generating Cartoon Style Facial Expressions with StackGAN. ([n. d.]).
[81] Jingjun Liang, Shizhe Chen, and Qin Jin. 2019. Semi-supervised Multimodal Emotion Recognition with Improved Wasserstein GANs. In *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 695–703.
[82] David Liu and Nathan Hu. 2020. GAN-Based Image Data Augmentation. (2020).
[83] Jinjin Liu, Qingbao Li, Ming Liu, and Tongxin Wei. 2020. CP-GAN: A Cross-Pose Profile Face Frontalization Boosting Pose-Invariant Face Recognition. *IEEE Access* 8 (2020), 198659–198667.
[84] Longfei Liu, Sheng Li, Yisong Chen, and Guoping Wang. 2018. X-gans: Image reconstruction made easy for extreme cases. *arXiv preprint arXiv:1808.04432* (2018).
[85] Ming Liu, Yukang Ding, Min Xia, Xiao Liu, Errui Ding, Wangmeng Zuo, and Shilei Wen. 2019. STGAN: A unified selective transfer network for arbitrary image attribute editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3673–3682.
[86] Ming-Yu Liu, Xun Huang, Jiahui Yu, Ting-Chun Wang, and Arun Mallya. 2021. Generative adversarial networks for image and video synthesis: Algorithms and applications. *Proc. IEEE* 109, 5 (2021), 839–862.
[87] Ming-Yu Liu and Oncel Tuzel. 2016. Coupled generative adversarial networks. *arXiv preprint arXiv:1606.07536* (2016).
[88] Yanfei Liu and Junhua Chen. 2021. Unsupervised face Frontalization for pose-invariant face recognition. *Image and Vision Computing* 106 (2021), 104093.

- [89] Yunfan Liu, Qi Li, and Zhenan Sun. 2019. Attribute-aware face aging with wavelet-based generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11877–11886.
- [90] Boyu Lu, Jun-Cheng Chen, and Rama Chellappa. 2019. Unsupervised domain-specific deblurring via disentangled representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10225–10234.
- [91] Yongyi Lu, Yu-Wing Tai, and Chi-Keung Tang. 2018. Attribute-guided face generation using conditional cyclegan. In *Proceedings of the European conference on computer vision (ECCV)*. 282–297.
- [92] Mario Lucic, Karol Kurach, Marcin Michalski, Sylvain Gelly, and Olivier Bousquet. 2017. Are gans created equal? a large-scale study. *arXiv preprint arXiv:1711.10337* (2017).
- [93] Sebastian Lutz, Konstantinos Amplianitis, and Aljosa Smolic. 2018. Alphagan: Generative adversarial networks for natural image matting. *arXiv preprint arXiv:1807.10088* (2018).
- [94] Kushal Mahalingaiah and Bruce Matichuk. 2019. ProDeblurGAN: Progressive Growing of GANs for Blind Motion Deblurring in Face Recognition. In *International Conference on Smart Multimedia*. Springer, 439–450.
- [95] Farkhod Makhmudkhujaev, Sungeun Hong, and In Kyu Park. 2021. Re-Aging GAN: Toward Personalized Face Age Transformation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3908–3917.
- [96] Dejan Malesevic, Christoph Mayer, Shuhang Gu, and Radu Timofte. 2019. Photo-realistic and robust inpainting of faces using refinement gans. In *Inpainting and Denoising Challenges*. Springer, 129–144.
- [97] Xudong Mao and Qing Li. 2018. Unpaired multi-domain image generation via regularized conditional GANs. *arXiv preprint arXiv:1805.02456* (2018).
- [98] Richard T Marriott, Sami Romdhani, and Liming Chen. 2020. A 3D GAN for Improved Large-pose Facial Recognition. *arXiv preprint arXiv:2012.10545* (2020).
- [99] Iacopo Masi, Yue Wu, Tal Hassner, and Prem Natarajan. 2018. Deep face recognition: A survey. In *2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)*. IEEE, 471–478.
- [100] Mehdi Mirza and Simon Osindero. 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* (2014).
- [101] Joel Ruben Antony Moniz, Christopher Beckham, Simon Rajotte, Sina Honari, and Christopher Pal. 2018. Unsupervised depth estimation, 3d face rotation and replacement. *arXiv preprint arXiv:1803.09202* (2018).
- [102] Stylianos Moschoglou, Stylianos Ploumpis, Mihalis A Nicolaou, Athanasios Papaioannou, and Stefanos Zafeiriou. 2020. 3dfacegan: Adversarial nets for 3d face representation, generation, and translation. *International Journal of Computer Vision* 128, 10 (2020), 2534–2551.
- [103] Thu Nguyen-Phuoc, Chuan Li, Lucas Theis, Christian Richardt, and Yong-Liang Yang. 2019. Hologan: Unsupervised learning of 3d representations from natural images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7588–7597.
- [104] Thekke Madam Nimisha, Kumar Sunil, and AN Rajagopalan. 2018. Unsupervised class-specific deblurring. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 353–369.
- [105] Augustus Odena, Christopher Olah, and Jonathon Shlens. 2017. Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning*. PMLR, 2642–2651.
- [106] Frans A Oliehoek, Rahul Savani, Jose Gallego-Posada, Elise Van der Pol, Edwin D De Jong, and Roderich Groß. 2017. GANGs: Generative adversarial network games. *arXiv preprint arXiv:1712.00679* (2017).
- [107] Greg Olmschenk, Zhigang Zhu, and Hao Tang. 2019. Generalizing semi-supervised generative adversarial networks to regression using feature contrasting. *Computer Vision and Image Understanding* 186 (2019), 1–12.
- [108] Zhaoqing Pan, Weijie Yu, Bosi Wang, Haoran Xie, Victor S Sheng, Jianjun Lei, and Sam Kwong. 2020. Loss Functions of Generative Adversarial Networks (GANs): Opportunities and Challenges. *IEEE Transactions on Emerging Topics in Computational Intelligence* 4, 4 (2020), 500–522.
- [109] Zhaoqing Pan, Weijie Yu, Xiaokai Yi, Asifullah Khan, Feng Yuan, and Yuhui Zheng. 2019. Recent progress on generative adversarial networks (GANs): A survey. *IEEE Access* 7 (2019), 36322–36333.
- [110] Guim Perarnau, Joost Van De Weijer, Bogdan Raducanu, and Jose M Álvarez. 2016. Invertible conditional gans for image editing. *arXiv preprint arXiv:1611.06355* (2016).
- [111] Shengju Qian, Kwan-Yee Lin, Wayne Wu, Yangxiaokang Liu, Quan Wang, Fumin Shen, Chen Qian, and Ran He. 2019. Make a face: Towards arbitrary high fidelity face manipulation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10033–10042.
- [112] Yichen Qian, Weihong Deng, and Jiani Hu. 2019. Unsupervised face normalization with extreme pose and expression in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9851–9858.
- [113] Alec Radford, Luke Metz, and Soumith Chintala. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* (2015).
- [114] Xuanchi Ren, Tao Yang, Yuwang Wang, and Wenjun Zeng. 2021. Do Generative Models Know Disentanglement? Contrastive Learning is All You Need. *arXiv preprint arXiv:2102.10543* (2021).

[115] Changle Rong, Xingming Zhang, and Yubei Lin. 2020. Feature-improving generative adversarial network for face frontalization. *IEEE Access* 8 (2020), 68842–68851.
[116] Kevin Roth, Aurelien Lucchi, Sebastian Nowozin, and Thomas Hofmann. 2017. Stabilizing training of generative adversarial networks through regularization. *arXiv preprint arXiv:1705.09367* (2017).
[117] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. 2016. Improved techniques for training gans. *arXiv preprint arXiv:1606.03498* (2016).
[118] Divya Saxena and Jiannong Cao. 2021. Generative Adversarial Networks (GANs) Challenges, Solutions, and Future Directions. *ACM Computing Surveys (CSUR)* 54, 3 (2021), 1–42.
[119] Pourya Shamsolmoali, Masoumeh Zareapoor, Eric Granger, Huiyu Zhou, Ruili Wang, M Emre Celebi, and Jie Yang. 2021. Image synthesis with adversarial networks: A comprehensive survey and case studies. *Information Fusion* (2021).
[120] Yujun Shen, Ceyuan Yang, Xiaoou Tang, and Bolei Zhou. 2020. Interfacegan: Interpreting the disentangled face representation learned by gans. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).
[121] Ziyi Shen, Wei-Sheng Lai, Tingfa Xu, Jan Kautz, and Ming-Hsuan Yang. 2020. Exploiting semantics for face image deblurring. *International Journal of Computer Vision* 128, 7 (2020), 1829–1846.
[122] Jingkuan Song, Jingqiu Zhang, Lianli Gao, Xianglong Liu, and Heng Tao Shen. 2018. Dual Conditional GANs for Face Aging and Rejuvenation.. In *IJCAI*. 899–905.
[123] Kumar Sricharan, Raja Bala, Matthew Shreve, Hui Ding, Kumar Saketh, and Jin Sun. 2017. Semi-supervised conditional gans. *arXiv preprint arXiv:1708.05789* (2017).
[124] Jiaze Sun, Binod Bhattarai, and Tae-Kyun Kim. 2020. MatchGAN: a self-supervised semi-supervised conditional generative adversarial network. In *Proceedings of the Asian Conference on Computer Vision*.
[125] Gino Tesei. 2019. Generating Realistic Facial Expressions through Conditional Cycle-Consistent Generative Adversarial Networks (CCycleGAN). (2019).
[126] Mukhiddin Toshpulatov, Wookey Lee, and Suan Lee. 2021. Generative adversarial networks and their application to 3D face generation: A survey. *Image and Vision Computing* (2021), 104119.
[127] Luan Tran, Xi Yin, and Xiaoming Liu. 2017. Disentangled representation learning gan for pose-invariant face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1415–1424.
[128] Luan Tran, Xi Yin, and Xiaoming Liu. 2018. Representation learning by rotating your faces. *IEEE transactions on pattern analysis and machine intelligence* 41, 12 (2018), 3007–3021.
[129] Soumya Tripathy, Juho Kannala, and Esa Rahtu. 2020. Icface: Interpretable and controllable face reenactment using gans. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3385–3394.
[130] Soumya Tripathy, Juho Kannala, and Esa Rahtu. 2021. FACEGAN: Facial Attribute Controllable rEenactment GAN. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1329–1338.
[131] Andrey Voynov and Artem Babenko. 2020. Unsupervised discovery of interpretable directions in the gan latent space. In *International Conference on Machine Learning*. PMLR, 9786–9796.
[132] De Wang and Liang-Jie Zhang. 2019. *Artificial Intelligence and Mobile Services–AIMS 2019*. Springer.
[133] Kunfeng Wang, Chao Gou, Yanjie Duan, Yilun Lin, Xinhu Zheng, and Fei-Yue Wang. 2017. Generative adversarial networks: introduction and outlook. *IEEE/CAA Journal of Automatica Sinica* 4, 4 (2017), 588–598.
[134] Qiang Wang, Huijie Fan, Gan Sun, Yang Cong, and Yandong Tang. 2019. Laplacian pyramid adversarial network for face completion. *Pattern Recognition* 88 (2019), 493–505.
[135] Qiang Wang, Huijie Fan, Gan Sun, Weihong Ren, and Yandong Tang. 2020. Recurrent generative adversarial network for face completion. *IEEE Transactions on Multimedia* 23 (2020), 429–442.
[136] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. 2018. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 8798–8807.
[137] Xiang Wang, Kai Wang, and Shiguo Lian. 2019. A survey on face data augmentation. *arXiv preprint arXiv:1904.11685* (2019).
[138] Zhengwei Wang, Qi She, and Tomas E Ward. 2019. Generative adversarial networks in computer vision: A survey and taxonomy. *arXiv preprint arXiv:1906.01529* (2019).
[139] Tongxin Wei, Qingbao Li, Jinjin Liu, Ping Zhang, and Zhifeng Chen. 2020. 3D Face Image Inpainting with Generative Adversarial Nets. *Mathematical Problems in Engineering* 2020 (2020).
[140] Maciej Wiatrak, Stefano V Albrecht, and Andrew Nystrom. 2019. Stabilizing Generative Adversarial Networks: A Survey. *arXiv preprint arXiv:1910.00927* (2019).
[141] Olivia Wiles, A Koepke, and Andrew Zisserman. 2018. X2face: A network for controlling face generation using images, audio, and pose codes. In *Proceedings of the European conference on computer vision (ECCV)*. 670–686.
[142] Hongyi Xu, Eduard Gabriel Bazavan, Andrei Zanfir, William T Freeman, Rahul Sukthankar, and Cristian Sminchisescu. 2020. Ghum & ghuml: Generative 3d human shape and articulated pose models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and*

Pattern Recognition. 6184–6193.
[143] Jiarui Xu and Hao-Wen Dong. [n.d.]. Flow-based Deep Generative Models. ([n. d.]).
[144] Zhiliang Xu, Xiyu Yu, Zhibin Hong, Zhen Zhu, Junyu Han, Jingtuo Liu, Errui Ding, and Xiang Bai. 2021. FaceController: Controllable Attribute Editing for Face in the Wild. *arXiv preprint arXiv:2102.11464* (2021).
[145] Huiyuan Yang, Umur Ciftci, and Lijun Yin. 2018. Facial expression recognition by de-expression residue learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2168–2177.
[146] Hongyu Yang, Di Huang, Yunhong Wang, and Anil K Jain. 2018. Learning face age progression: A pyramid architecture of gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 31–39.
[147] Huiyuan Yang, Zheng Zhang, and Lijun Yin. 2018. Identity-adaptive facial expression recognition through expression regeneration using conditional generative adversarial networks. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 294–301.
[148] Wei Yi, Yaoran Sun, and Sailing He. 2018. Data augmentation using conditional GANs for facial emotion recognition. In *2018 Progress in Electromagnetics Research Symposium (PIERS-Toyama)*. IEEE, 710–714.
[149] Junho Yim, Heechul Jung, ByungIn Yoo, Changkyu Choi, Dusik Park, and Junmo Kim. 2015. Rotating your face using multi-task deep neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 676–684.
[150] Xi Yin, Xiang Yu, Kihyuk Sohn, Xiaoming Liu, and Manmohan Chandraker. 2017. Towards large-pose face frontalization in the wild. In *Proceedings of the IEEE international conference on computer vision*. 3990–3999.
[151] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. 2019. Free-form image inpainting with gated convolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4471–4480.
[152] Egor Zakharov, Aliaksandra Shysheya, Egor Burkov, and Victor Lempitsky. 2019. Few-shot adversarial learning of realistic neural talking head models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9459–9468.
[153] Bassel Zeno, Ilya Kalinovskiy, and Yuri Matveev. 2021. PFA-GAN: Pose Face Augmentation Based on Generative Adversarial Network. *Informatica* (2021), 1–16.
[154] Feifei Zhang, Tianzhu Zhang, Qirong Mao, and Changsheng Xu. 2018. Joint pose and expression modeling for facial expression recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3359–3368.
[155] Guanghao Zhang, Enmei Tu, and Dongshun Cui. 2017. Stable and improved generative adversarial nets (GANS): A constructive survey. In *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 1871–1875.
[156] Jiayuan Zhang, Ao Li, Yu Liu, and Minghui Wang. 2019. Adversarially regularized U-Net-based GANs for facial attribute modification and generation. *IEEE Access* 7 (2019), 86453–86462.
[157] Jichao Zhang, Fan Zhong, Gongze Cao, and Xueying Qin. 2017. St-gan: Unsupervised facial image semantic transformation using generative adversarial networks. In *Asian Conference on Machine Learning*. PMLR, 248–263.
[158] Ke-Yue Zhang, Taiping Yao, Jian Zhang, Ying Tai, Shouhong Ding, Jilin Li, Feiyue Huang, Haichuan Song, and Lizhuang Ma. 2020. Face anti-spoofing via disentangled representation learning. In *European Conference on Computer Vision*. Springer, 641–657.
[159] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 586–595.
[160] Yinghui Zhang, Bo Sun, Yongkang Xiao, Rong Xiao, and YunGang Wei. 2019. Feature augmentation for imbalanced classification with conditional mixture WGANs. *Signal Processing: Image Communication* 75 (2019), 89–99.
[161] Zhifei Zhang, Yang Song, and Hairong Qi. 2017. Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5810–5818.
[162] Guoping Zhao, Mingyu Zhang, Jiajun Liu, and Ji-Rong Wen. 2019. Unsupervised adversarial attacks on deep feature-based retrieval with GAN. *arXiv preprint arXiv:1907.05793* (2019).
[163] Jian Zhao, Yu Cheng, Yi Cheng, Yang Yang, Fang Zhao, Jianshu Li, Hengzhu Liu, Shuicheng Yan, and Jiashi Feng. 2019. Look across elapse: Disentangled representation learning and photorealistic cross-age face synthesis for age-invariant face recognition. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 9251–9258.
[164] Jian Zhao, Lin Xiong, Jayashree Karlekar, Jianshu Li, Fang Zhao, Zhecan Wang, Sugiri Pranata, Shengmei Shen, Shuicheng Yan, and Jiashi Feng. 2017. Dual-Agent GANs for Photorealistic and Identity Preserving Profile Face Synthesis.. In *NIPS*, Vol. 2. 3.
[165] Jian Zhao, Lin Xiong, Jianshu Li, Junliang Xing, Shuicheng Yan, and Jiashi Feng. 2018. 3d-aided dual-agent gans for unconstrained face recognition. *IEEE transactions on pattern analysis and machine intelligence* 41, 10 (2018), 2380–2394.
[166] Hang Zhou, Jihao Liu, Ziwei Liu, Yu Liu, and Xiaogang Wang. 2020. Rotate-and-render: Unsupervised photorealistic face rotation from single-view images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5911–5920.
[167] Sharon Zhou, Mitchell L Gordon, Ranjay Krishna, Austin Narcomey, Li Fei-Fei, and Michael S Bernstein. 2019. Hype: A benchmark for human eye perceptual evaluation of generative models. *arXiv preprint arXiv:1904.01121* (2019).
[168] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*. 2223–2232.

[169] Xinyue Zhu, Yifan Liu, Jiahong Li, Tao Wan, and Zengchang Qin. 2018. Emotion classification with data augmentation using generative adversarial networks. In *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 349–360.
[170] Zhenyao Zhu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. 2014. Multi-view perceptron: a deep model for learning face identity and view representations. *Advances in Neural Information Processing Systems 27 (NIPS 2014)* (2014).