Studies in Big Data 132

Gilberto Rivera
Laura Cruz-Reyes
Bernabé Dorronsoro
Alejandro Rosete *Editors*

Data Analytics and Computational Intelligence: Novel Models, Algorithms and Applications

Springer

Studies in Big Data

Volume 132

Series Editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland

Gilberto Rivera · Laura Cruz-Reyes ·
Bernabé Dorronsoro · Alejandro Rosete
Editors

Data Analytics and Computational Intelligence: Novel Models, Algorithms and Applications

Springer

Editors
Gilberto Rivera
División Multidisciplinaria de Ciudad Universitaria
Universidad Autónoma de Ciudad Juárez
Chihuahua, Mexico

Bernabé Dorronsoro
School of Engineering
Universidad de Cádiz
Cadiz, Spain

Laura Cruz-Reyes
Tecnológico Nacional de México/Instituto Tecnológico de Ciudad Madero
Ciudad Madero, Tamaulipas, Mexico

Alejandro Rosete
Universidad Tecnológica de La Habana "José Antonio Echeverría"
La Habana, Cuba

ISSN 2197-6503 ISSN 2197-6511 (electronic)
Studies in Big Data
ISBN 978-3-031-38324-3 ISBN 978-3-031-38325-0 (eBook)
https://doi.org/10.1007/978-3-031-38325-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Predictive Analytics

Prescriptive Analytics

Data Augmentation Techniques for Facial Image Generation: A Brief Literature Review

Blanca Elena Cazares, Rogelio Florencia, Vicente García, and J. Patricia Sánchez-Solís

Abstract Image processing has gained notoriety over the last few years in performing various tasks through deep learning (DL) algorithms, such as face recognition and identity verification. Unfortunately, most of them require a large set of images for training, usually manually labeled, which is a costly task both in time and effort, not to mention being prone to human error. Data Augmentation (DA) techniques have been used to mitigate this situation, as they generate images by applying variations to real image sets. This chapter presents a brief literature review on various DA methods dedicated to image generation. The technique that has presented outstanding results in the task of generating artificial images is Generative Adversarial Networks (GANs). Some recent research papers in which GANs have been used for the generation of artificial images are presented. General aspects of GANs, such as their definition, architecture, training, and challenges, are described. Additionally, the implementation of a GAN architecture for the generation of artificial face images from a public set of images is presented. The need for a great computational capacity to generate images with great sharpness and realism is highlighted.

Keywords Face image generation · Data augmentation · Generative Adversarial Networks

B. E. Cazares · R. Florencia (✉) · V. García · J. P. Sánchez-Solís
Departamento de Ingeniería Eléctrica y Computación, Universidad Autónoma de Ciudad Juárez, Cd. Juárez, Chihuahua, México
e-mail: rogelio.florencia@uacj.mx

B. E. Cazares
e-mail: blancacazaress@gmail.com

V. García
e-mail: vicente.jimenez@uacj.mx

J. P. Sánchez-Solís
e-mail: julia.sanchez@uacj.mx

G. Rivera et al. (eds.), *Data Analytics and Computational Intelligence: Novel Models, Algorithms and Applications*, Studies in Big Data 132,
https://doi.org/10.1007/978-3-031-38325-0_9

1 Introduction

Facial recognition systems have been discussed for several decades. The scientific community has shown great interest in this subject, given the human ability to recognize others through the face, being a unique part of the body. One of the goals is that these systems possess this same intelligence as part of the biometric identification measures.

Facial recognition has been developed in the field known as computer vision [1]. This area supports a variety of critical applications, e.g., identity verification. This task is done using what is known as a face analyzer, which is software that confirms the identity of people based on their faces [2]. It is achieved through the identification and measurement of facial features in images. It can also be used to associate human faces in the latest or even in videos to check how similar they are to one or several specific individuals. In addition, it can determine the level of similarity between two photographs to find out if they belong to the same person or search for them among a large set of images in a collection. For example, biometric security systems use facial recognition to uniquely identify individuals as users at login to strengthen user authentication. In addition, mobile devices often use this type of technology to protect the data they contain [3]. Unfortunately, large amounts of facial images are required for these processes to function properly, so much of the research has focused on artificial image generation. It aims to increase the image sets used to train different architectures because of their significant impact on the results [4].

Collecting and labeling data samples with good quality is costly, both in time and effort and is prone to human error. For this purpose, various techniques known as Data Augmentation (DA) have been implemented, and their performance is suitable in different domains.

DA techniques allow the size of existing image sets to be increased considerably through simulations [4], which helps significantly with the image requirements for learning systems. Also, DA techniques have faced several limitations of their own [5], for example:

The images generated lack realistic variations such as makeup, skin color, and background change, which means that they would have a different distribution than the real ones.

Creating high-quality facial images is very difficult due to the complexity of facial details.

After many years of research and application of different modeling methods, in 2014, a technique was proposed that allows the generation of realistic images using two artificial neural networks, Generative Adversarial Networks (GAN). Although these are not the first method used for artificial data generation, their results and versatility distinguished them from the rest since they have achieved outstanding results that were still considered impossible for artificial systems [6].

GANs are a machine learning (ML) technique that integrates two simultaneously trained models, the generator and the discriminator. The generator is trained to create the artificial images, and the discriminator is to discern between the fake images created by the generator and the real ones from the original training set. Its performance has far exceeded expectations in the field of artificial image generation, according to a variety of authors [5, 7–9].

Like ML algorithms, the data generated by a GAN completely depends on the training set provided to perform the learning. For example, if a GAN is required to learn how to create images of handwritten numbers, it is necessary to use a training set containing several images of handwritten numbers [6].

GANs are derived from a gaming perspective. Hence the word adversarial denotes a competitive dynamic between the two models that compose them. The generator aims to generate real images indistinguishable from the training set. On the other hand, the discriminator aims to distinguish these generated images from the real ones. Therefore, the better the generator generates realistic images, the better the discriminator must be to distinguish the real ones from the false ones [6].

The fast growth and progress of GANs have been due to research and development, generating new architectures to stabilize the outputs and generate images of higher quality and realism. In this way, the aim is to create impossible images for the human eye to differentiate.

This chapter presents a brief literature review on data augmentation methods for artificial image generation, focusing mainly on GANs, which have been used throughout recent years to create artificial facial images. Additionally, the general architecture of a GAN and related concepts are described. Lastly, the implementation of a GAN architecture for generating artificial face images, trained with a small part of the public image set "CelebA," is described.

The main contributions of this work are: (a) the compilation of recent state-of-the-art works that demonstrate the different areas in which GANs have been used to improve the training of various systems or algorithms, such as ML, (b) the description of some architectures and parameter variations that can be modified according to the objectives pursued and, (c) an example of a GAN architecture used to generate artificial faces.

The chapter is structured as follows. Section 2 presents the background to the generation of artificial facial images. Section 3 describes the general architecture of a GAN, the training process, and its challenges, among other information of interest. Section 4 presents a brief literature review on data augmentation methods for generating artificial images, focusing mainly on GANs. Section 5 describes the methodology used to conduct the literature review. Section 6 shows the implementation of a GAN architecture to generate artificial facial images. Lastly, Sect. 7 presents the conclusions and future work.

2 Generation of Artificial Facial Images

Generating new images from others is a widely researched task in computer vision. In recent years, the development of Artificial Intelligence (AI) techniques has motivated the idea of producing images of high quality and realism. It has enabled the creation of realistic human faces that are difficult to distinguish between fake and real, even for the human eye. This task has been evolving exponentially since implementing GANs [10]. However, before GANs, other computational methods were used to generate variations in facial images to train systems such as facial recognition systems.

Four main categories can be broadly discussed in the DA task for generating artificial images: generic, component, attribute, and age transformations. Each one is described in the following sections.

2.1 Generic Transformations

These transformations focus on modifying the entire image, ignoring high-level components such as composition, light, volume, symmetry, shape, and texture. They are usually divided into two main groups: geometric and photometric transformations.

Geometric transformations are commonly applied in multiple computer vision tasks [5], such as face recognition, healthcare, and manufacturing applications. In general terms, geometric transformations alter the pixels of an image by placing them in new positions. Some examples of these can be:

- Reflection: consists of flipping the image around its vertical or horizontal axis.
- Rotation: rotates the image θ degrees around its center, bringing each pixel (x, y) to its position $(x\prime, y\prime)$.
- Cropping: consists of cutting the images to a specific size.

For example, in the case of convolutional neural networks (CNNs), this type of transformation helps to minimize their sensitivity to changes in position and orientation [11]. Some examples can be seen in Fig. 1.

On the other hand, photometric transformations, shown in Fig. 2, generally work by altering the RGB (red, green, and blue) channels, shifting each pixel value (r, g, b) to new values (r', g', b') of an image according to predefined rules. These transformations adjust the illumination and color, leaving the geometry unaffected [11]. For example, it can be mentioned color manipulations, such as inverting them or adding some filters, such as blurring or grayscale [5].

These transformations are mainly used in computer vision tasks to enrich the training sets and prevent a common problem in this field, overfitting. [12] is a paper dedicated to evaluating the performance of CNNs, trained with images enriched using geometric and photometric transformations. In [13], different transformations were employed to increase the image set and prevent overfitting.

Fig. 1 Geometric transformation examples [5]

Fig. 2 Photometric transformation examples [5]

2.2 Component Transformation

There are transformations dedicated to enriching the sets of facial images by modifying the components of the person. These images are used to train the algorithm so that it is able to recognize the person, even if their appearance is altered.

The hairstyle can be considered one of the components to be generated since, although it is not considered a facial component, hairstyle affects face detection and recognition because it tends to hide certain features of a person's appearance. Therefore, DA techniques focus on generating facial images with different variations in hair, for example, color, shape, and bangs. In [14], a method using DiscoGAN was proposed. This variant learns to discover relationships between different domains and develops the ability to translate features between them, for example, by transforming hair color.

Makeup transfer and accessory addition techniques can also be identified due to the difficulty for recognition systems to effectively perform their tasks when some features of the face look different depending on the makeup or accessories a person wears. Most studies based on this type of transformation can be divided into two categories: traditional image processing [15] and those based on DL [16]. Some examples of these can be seen in Fig. 3.

Another component that impacts facial recognition is the use and removal of accessories, including glasses, earrings, and piercings, among others. Of all these, glasses are the most commonly used, as they are used for different reasons, for example, vision correction, prevention against sunlight, eye protection, and aesthetics, among others. They significantly affect the accuracy of facial recognition, as they usually cover a large area of the face.

In [17], a fusion of virtual lenses onto faces was performed using the Augmented Reality (AR) technique. In [18], a method of facial attribute manipulation based on image residuals was proposed, defining this as the difference between the input image and the desired output image.

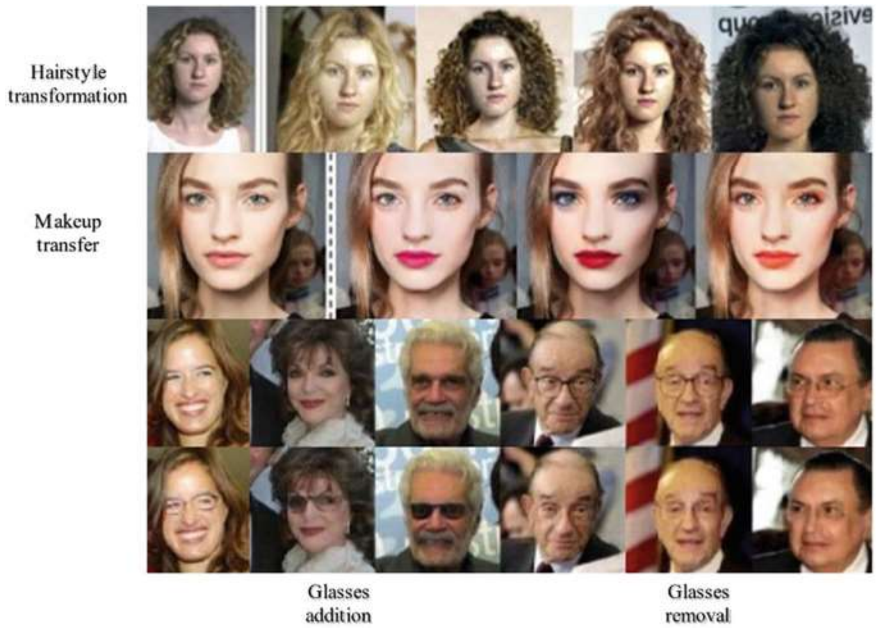


Fig. 3 Component transformation examples [5]

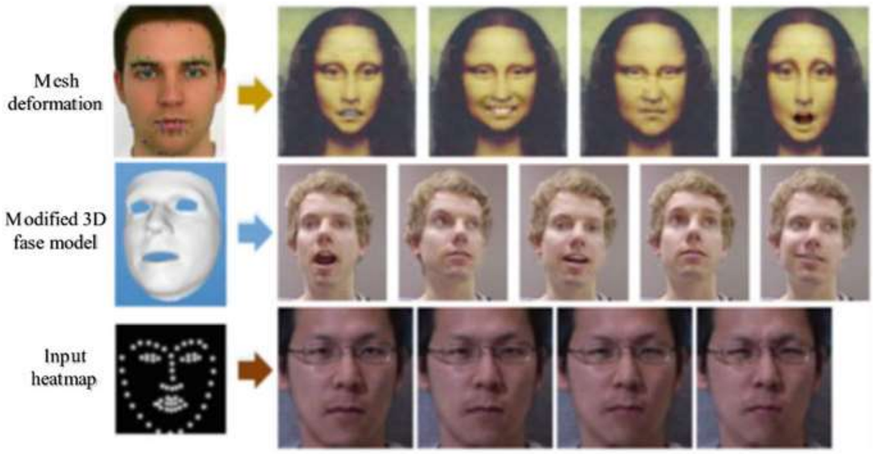


Fig. 4 Facial expression synthesis [22, 23]

2.3 Attribute Transformation

There are also some transformations dedicated to modifying aspects such as pose. In this case, the position of the head in a photograph is considered a significant challenge in facial recognition tasks since any variation in it tends to modify visual aspects of the face, i.e., it can hide or show different facial details. In addition, it has been considered an essential aspect since facial photographs in several legal processes are requested from the frontal side. This leads to the research of recreating how a face may look from other angles.

Facial expressions are also considered critical: happy, annoyed, scared, and surprised, among others. These techniques help to improve the performance of emotion classification and recognition systems. Primarily 2D, 3D, and learning-based modeling approaches are used to achieve this goal and usually focus on modifying the expression of a face using expression templates by concentrating on a series of points, for example, the corners of the mouth, the cheeks position, and the location of the eyebrows. Others focus on simulations to recreate face parts that are hidden by the pose in which the photograph was taken [4, 19–21]. Some examples of this can be found in Fig. 4.

2.4 Age Progression and Regression

Finally, age regression or progression seeks to predict a current face's appearance in the past or future, respectively, while preserving its facial features. It has become a widely explored research topic because it significantly affects various applications, including missing person tracking, facial recognition, and aesthetic studies, among others. The two main concerns in attempting age regression or progression are identity preservation and prediction accuracy [24].

These facial images enrich the image sets by adding different features at different time stages of the same face, which makes the models more robust to age variations.

Age modification methods are mainly divided into two: prototype-based and model-based. The first one creates an average face for different age ranges, learns the shapes and textures of these, and applies these features to transfer them to a new face. However, the individual's characteristics are often lost with these methods [5].

Model-based methods build, as the name implies, models of biological changes due to age in faces, e.g., musculature, wrinkles, and skin texture, among others. They are called generative because they are such a powerful tool for creating new data by learning to imitate the probabilistic distribution of a training set. Nowadays, generative models have gained importance and attention due to their good performance in data creation. Among the most popular are autoregressive models, Variational Autoencoders (VAEs), and GANs. The disadvantage of these models is their complexity and computational cost [5].

GANs, as mentioned above, are an alternative architecture for training generative models since they handle probabilistic computations very well. Recent work has begun to apply them to the age regression and progression task, and many variants of the model have been generated. Some examples of this generative task are shown in Fig. 5.

Before GAN, two approaches to age progression and regression in faces were prototype-based and modeling-based. In the latter, critical points in the image, such

Fig. 5 Facial age regression and progression examples [5]

as eyes, nose, and jaw, track, in turn, the temporal changes such as wrinkles, musculature, and color in these are identified. However, this method requires a large amount of age-labeled data over a long period for each individual, which is difficult to find and computationally expensive. The prototype-based method creates an average face based on a set of images of a particular age group, using it to transfer those features from one age range to another. The disadvantage of this is that personal features are often lost. Another possibility is to use neural networks to transform faces across ages. It generates smoother images but still requires images labeled with the person's age through the years [8].

On the other hand, the GAN consists of a discriminator and a generator competing with each other based on the *min–max* games. The generator starts by receiving as input a noise vector z and creates an image that it gives to the discriminator to receive feedback from it. Some variants of this architecture are [8]:

- DCGAN (Deep Convolutional GAN) has demonstrated that GAN can be successfully applied to generate indoor scenarios and human faces.
- StyleGAN significantly extended the basic GAN to progressively generate high-resolution images from those with very low resolutions.

cGAN (Conditional Generative Adversarial Networks) introduces an identity preservation vector with the optimization approach when generating faces so there is a better match between the original and the created face.

- Pyramid GAN simulates the effects of age more sharply and presents a suite of methods for assessing accuracy and fidelity to the original image.

3 Generative Adversarial Networks (GANs)

GANs are powerful AI-based unsupervised learning algorithms that aim to learn the estimated probability distribution in a specific training set. GANs were proposed by Ian Goodfellow in 2014 [25]. They are based on a competition system between two neural network models that try to maximize their performance while minimizing that of their adversary, developing the ability to analyze, capture and copy the variations presented by a particular set of images [8].

These networks can generate artificial data, being one of the generative models with the highest quality of results, especially when its potential to generate high-resolution images is analyzed [6]. Section 3.1 gives a definition of what a GAN is. Section 3.2 describes the general architecture with which GANs were originally proposed. Section 3.3 describes the GAN training process. Section 3.4 mentions the challenges researchers face when training a GAN. Finally, Sect. 3.5 presents the face image generation evolution with GANs.

3.1 Definition

It has been demonstrated that most neural networks can be easily tricked into misclassifying by adding only a small amount of noise to the original data. Surprisingly, after this addition, the model develops a higher confidence level in the wrong predictions than in the correct ones. It is because most ML algorithms learn from a limited amount of data, which is prone to model overfitting [26]. It motivated the creation of GANs. They can be described in three parts:

- Generative: They are considered generative models since they describe how new data are generated in terms of probabilistic models.
- Adversarial: The model is trained by competition among its neural networks, i.e., they are considered adversarial to each other.
- Networks: They use neural networks as the primary training algorithms.

They are based on game theory, which considers players to be both ML models, typically implemented using neural networks. A network is called a generator, which can learn the distribution obtained from an original data set to try to replicate it. It is achieved by inserting a noise vector z, i.e., random numbers with a Gaussian distribution. The main objective of the generator is to learn how to transform unstructured noise z into realistic samples [25].

The other player is called the discriminator. It examines each example x received as input and outputs an estimate of whether it is true or false.

Each player has a cost, so they try to minimize their own, i.e., the discriminator's cost encourages it to correctly classify the data as real or fake. In contrast, the cost of the generator encourages it to generate data that the discriminator incorrectly classifies as real.

A typical example when talking about GANs is to imagine that one network represents a money counterfeiter and the other a policeman. The former generates counterfeit bills while the policeman tries to arrest him for benefiting the production of legitimate ones. The competition leads to the production of more and more realistic counterfeit bills until, eventually, the counterfeiter produces them so realistically that the policeman cannot tell the difference between authentic and counterfeit [25].

3.2 Architecture

In the GAN architecture, the generator will start training simultaneously as the discriminator does. The latter will need a few epochs before starting the adversarial training since it must be able to classify the images correctly.

The architecture consists of two competing deep neural networks: the *generator* and the *discriminator*. The generator produces new data instances, while the discriminator tries to distinguish accurately between the real data, i.e., those coming from the original set or fake data produced by the generator.

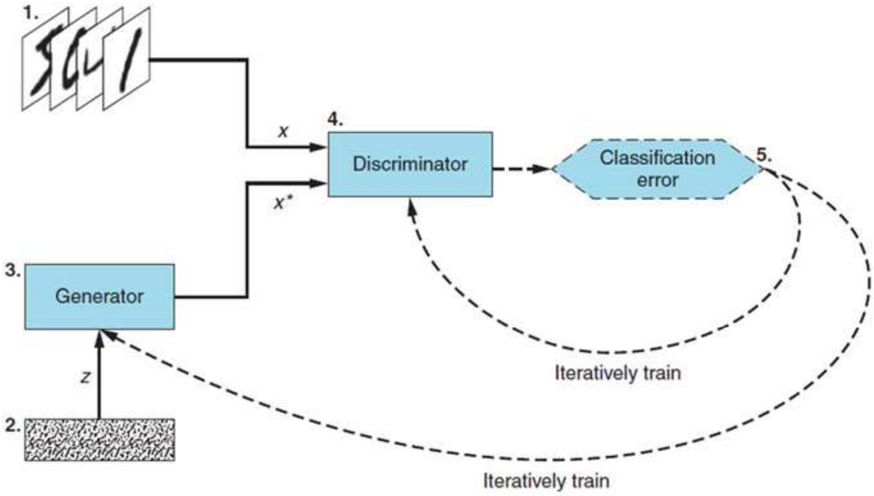


Fig. 6 The two GAN subnetworks with their inputs, outputs, and interactions [6]

This competition will continue until the generator can create realistic artificial data, which can then be used as input to other neural networks [6]. Figure 6 shows a GAN schematic with its basic components.

Since its development, many different architectures have been proposed to deal with a wide variety of domains. It could even be said that several scientific papers are published weekly [27]. In [28], an extensive literature review is done on multiple architectures developed for GANs.

3.3 Training Process

The training phase requires the two networks and a set of data from which the artificial data will be generated. First of all, it is worth mentioning that the training of the generator is much more complex than the discriminator, which can be seen more as a binary classifier.

The discriminator receives real inputs that must be labeled as such and inputs from the generator, which must be labeled false, 1 if the input data is real and 0 if it is false. On the other hand, the generator must be trained with the only condition that the data created mislead the discriminator, i.e., it must minimize its loss and maximize that of the opponent. To achieve this, the generator's output must be the input of the discriminator so that the output of the whole model gives as output the probability that the data are real, according to the discriminator. In this way, the generator obtains feedback from the discriminator, which it uses to create progressively more similar data in the training set [25].

3.4 Challenges

The GANs are still facing research challenges regarding their training. Among them, the following can be pointed out [29]:

- Non-convergence: this is when the generator and the discriminator fail to reach the desired equilibrium (50%). Their respective loss functions begin to fluctuate without being able to reach stability.
- Modal collapse: occurs when the generator produces similar data, even though the inputs vary in characteristics. It finds a small set of samples that successfully deceive the discriminator and are thus incapable of producing others. In these cases, the gradient of the loss function is stuck at a value close to 0.
- Non-informative loss: the general intuition is that the lower the loss of the generator, the higher the quality of the data it produces. However, the loss must be compared with the discriminator's, which constantly improves. Therefore, the issue of model evaluation is more complex. The generator may produce better-quality samples even as the loss function increases.

Due to these challenges, techniques have been developed with the aim of minimizing them as much as possible. However, they are still the subject of research at the moment [6].

3.5 Face Image Generation Evolution with GANs

Progress in facial image generation has gone hand in hand with GANs, so it is now possible to control the resolution or quality of artificial images. Table 1 shows the main advances of the GANs.

4 Related Work

The following research works have focused on implementing some GAN variants with different objectives.

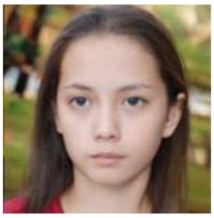
In [9], it is mentioned that the main objective is to generate a new face by editing facial attributes in the images, preserving its identity. Subsequently, it proceeds to perform a literature review, finding that GAN architecture with an encoder is usually incorporated for such tasks with promising results. Thus, it is proposed to apply an attribute classification constraint to the generated images, in addition to the previous architecture, forming together what is called AttGAN. Experiments were performed on the public set of images, "Celeba," to manipulate attributes such as hair color, beard, and age, achieving realism and preservation of facial details.

Table 1 GAN evolution in artificial facial image generation

Year	Image	Model description
2014		The proposed GAN architecture is composed of two models: generative *G* and discriminative *D*, represented by a multilayer perceptron [25]
2015		The proposed DCGAN architecture for unsupervised learning is composed of two models: *G* and *D*, represented by CNNs [30]
2016		Coupled Generative Adversarial Network (CoGAN) architecture proposed, which consists of two models for *G* and two for *D* [31]
2017		Proposed new training for GANs by progressively increasing both *G* and *D*, starting with low resolution and adding new layers. It is argued that it improves the training speed and stabilizes it [32]

(continued)

Table 1 (continued)

Year	Image	Model description
2018		Proposed alternative architecture for *G* in GANs, called StyleGAN. This leads to automatically learned and unsupervised high-level attribute separation [33]
2019		Improvements in the StyleGAN architecture and its training method are proposed. In addition, further work is being done on the quality of the generated images [34]

In [8], a CNN was developed to generate images considering facial age progression. To do this, the authors conducted an exhaustive review of the literature. They analyzed several sets of images to select the one with the greatest variety and fit the domain they wanted to develop the model. Then, they configured parameters and loss functions for network stabilization and the creation of realistic photographs. They used qualitative and quantitative metrics for the evaluation, resulting in images with appropriate characteristics for age progression.

The makeup transfer is the object of study to transfer a specific style to a clean face, preserving the identity of the same. This type of problem at the instance level is considered a great challenge since the styles vary greatly, for example, eye shadows, lipsticks, and foundations, among many others. In [16], BeautyGAN is proposed so that the networks can perform the transfer at the instance level through unsupervised adversarial learning. At the end of the study, a new set of high-resolution makeup images was constructed.

GANs even have an impact in the medical area, as can be seen in [35]. A study on coronavirus (COVID-19), a viral disease caused by the SARS-CoV-2 respiratory syndrome, begins, encouraged by its global effect on health and economics. Chest X-rays from infected patients were a crucial step in controlling this virus. Thus began the introduction of various DL systems and studies that demonstrated the efficiency of using chest X-rays for patient detection. In this context, since CNNs require a significant amount of training data to perform adequately and the virus was too recent to have enough Chest X-rays to generate systems that learn to detect it, the authors present a method to generate images from Chest X-rays in short times. It was achieved by introducing a model named CovidGAN, proving that the artificial images produced by it helped improve the performance of the trained CNNs for COVID-19 detection, increasing their accuracy to 95%.

In [36], it focuses on another important application of GAN: the generation of facial images from text. It has multiple applications in public safety and forensic analysis, such as finding criminals or suspects described by eyewitnesses. This chapter proposes synthesizing facial images from text using a fully-trained generative adversarial network (FTGAN), trained with a text encoder and an image decoder, to generate good-quality images from the input sentences. Multiple experiments were performed on the CUB public set, providing good results regarding the main objective. It was measured by comparison against methods found in the literature, using the Frechet Inception Distance (FID) and Face Semantic Distance (FSD) metrics. Additionally, human ratings were used to validate the generated images.

Another application in the medical field was discussed in [37], where a method was developed to synthesize photorealistic microscope images of red blood cells using cGAN. These images were combined with real ones to augment reduced datasets. Using cGANs as a synthetic data generator for the DA task benefits the development of a model. However, they also raise the fact that all that is involved in developing a GAN model: unstable and time-consuming training, heavy computational requirements, and other challenges may not be worth the small margin of improvement it may provide in some tasks.

A literature review is presented in [38], mainly focused on generating medical images using GAN frameworks since it is stated that these have proven to be useful in many cases of image augmentation, medical image generation, and image reconstruction, among others. These characteristics were the ones that encouraged the research of GANs in the medical field to improve image analysis. It is mentioned that several GAN frameworks have gained popularity in medical image interpretation, such as DCGAN, Laplacian GAN (LAPGAN), pix2pix, Cycle-GAN, and unsupervised image-to-image translation model (UNIT). Although despite all these approaches, it was concluded that they are still in very early stages, and in-depth research in the area is still needed to reach the most reliable level of progress for GAN applications in clinical imaging.

Among a wide range of applications where GANs have been implemented, one can be found in [39]. In this study, Pix2PixHD was applied to perform image-to-image translation of solar images to give scientists access to complete sets of images for analysis. The end of the training and testing proved the ability of such algorithms to generate high-resolution images. In addition, the images obtained through this model proved to be better than those obtained in previous works also intended to generate these images. It opens the possibility of using these models to create images when unavailable and thus have a better understanding of space weather. It also allows researchers to have the ability to predict solar events, for example, Solar Flares or Coronal Mass Ejections.

In [40], research is found to be motivated by the good performance of CNNs in traffic sign detection (TSD) and recognition (TSR), which are critical tasks in the development of autonomous driving systems. It is difficult because training neural networks requires a large amount of labeled data, and the visual design of traffic signs varies greatly, especially when dealing with different countries. This work focused on Taiwanese traffic signs, motivated by not finding a database, images, or study that focused on identifying them. It also pursued the generation of synthetic images when it was impossible to collect reference images through research and development. DCGANs and Wasserstein generative adversarial networks (Wasserstein GANs, WGANs) were used. Each demonstrated exceptional outputs when creating synthetic graphics. The evaluation of the efficiency of DCGAN and WGAN is also included by measuring the quality of the generated images using Structural Similarity Index (SSIM) and the Mean Square Error (MSE). It was concluded that synthetic traffic signal images could be generated with a minimal training image set. Although the resulting images' quality was different obtained with a large training set, they can still be used to solve difficulties when getting genuine photographs is complicated.

5 Methodology

The search and selection of the articles presented in this chapter were based on the Prisma (Preferred Reporting Items for Systematic Reviews and Meta-analyses) methodology, which includes a guide to identify, select, evaluate, and synthesize studies [41]. Prisma was designed for systematic reviews of health-related studies, but it can also be applied to social and educational interventions. In this way, researchers interested in reviewing the scientific literature can recreate the author search process.

The Google Scholar search engine was used to perform the literature review, which indexes full texts or metadata of the academic literature from a wide range of publication formats. The search was started considering the terms "Generative Adversarial Networks" to have a first vision of the research works that have been carried out, available in public databases. Subsequently, the search results were filtered based on the year of publication, "Generative Adversarial Networks," 2019, to identify recently published papers. In addition, a logical operator was added to further narrow the search for "Generative Adversarial Networks" AND "image generation." This allowed filtering of the papers found to identify only those focused on image generation.

Since GANs are used for text, images, videos, voice, and statistics, among others [6], a logical operator was added to further narrow the search to "Generative Adversarial Networks" AND "image generation." This allowed filtering of the articles found to identify those focused on image generation. The search returned approximately 16,700 results (February 3, 2023). Of these, we selected and analyzed those that we considered being the most like this work. The search returned approximately 16,700 results (February 3, 2023). Of these, we selected and analyzed those that we consider being the most similar to the purposes pursued by this work.

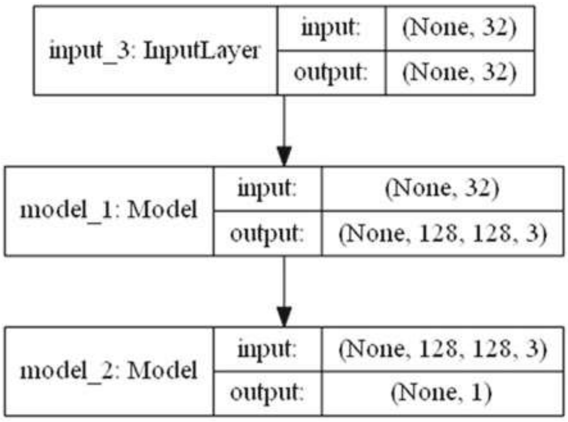


Fig. 7 Architecture of a DCGAN model for generating facial images where *model_1* represents the generator and *model_2* the discriminator

6 Face Image Generation with GANs

In this section, the architecture of a DCGAN for generating facial images is presented to exemplify the use of GANs. The model was trained using the public image set "Celeba" [42], comprising approximately 200,000 celebrity facial images. It was selected since it contains many variations such as hair color, smile, glasses, poses, backgrounds, and diversity of people that make it very suitable for training face detection models. In addition, its elements are labeled for efficient use in computer vision. A virtual machine generated and configured by the Laboratorio Nacional de Tecnologías de Información (LaNTI)[1] at Universidad Autónoma de Ciudad Juárez (UACJ) was used. The system specifications are Ubuntu 20.04, dual-core, 100 GB storage, and 8 GB RAM. The architecture is shown in Fig. 7.

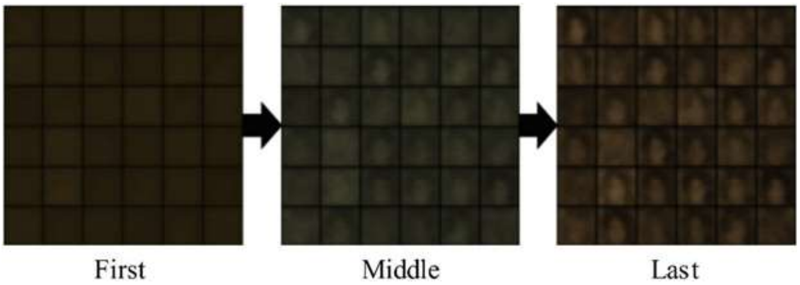
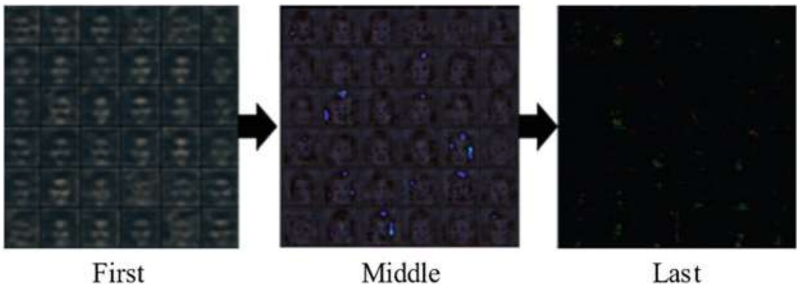
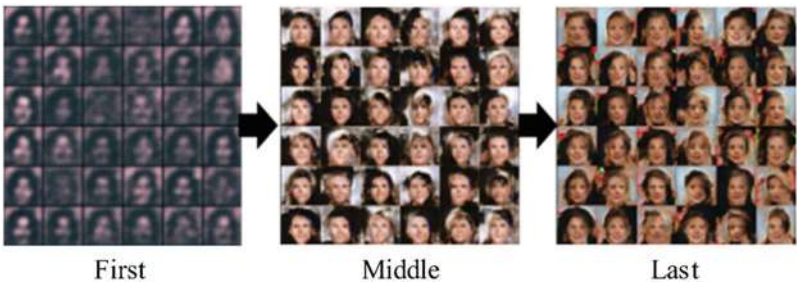
The input layer (input_3:InputLayer) represents the latent vector defined as 32, and the generator (model_1:Model) receives this as input and outputs an image of 128*128 size with three color channels (RGB). The discriminator (model_2:Model) receives this image as input and outputs a binary classification to identify the image as *true* or *false*.

Since training this model is computationally expensive, some tests were performed to evaluate the time required with different parameters, as shown in Table 2. The first column indicates the number of images used, the second the number of iterations, and the third the batch size, i.e., the number of examples introduced into the network for each training. Also, in the line below the parameters are the images resulting from different parts of the training: first, middle, and last.

The tests made obtaining better results possible but required considerable time investment. Additionally, tests allowed the selection of the model parameters, balancing the time and quality of the images generated to present the example.

[1] Laboratorio Nacional de Tecnologías de Información: http://www.lanti.org.mx/lanti/

Table 2 Tests for the selection of model parameters

<table>
<tr><th>Images</th><th>Iterations</th><th>Batch size</th></tr>
<tr><td>8</td><td>20</td><td>8</td></tr>
<tr><td colspan="3">
</td></tr>
<tr><td>50</td><td>20</td><td>8</td></tr>
<tr><td colspan="3">
</td></tr>
<tr><td>10,000</td><td>5,000</td><td>3</td></tr>
<tr><td colspan="3">
</td></tr>
</table>

Following the testing phase, the development of the final model began. The selected parameters, image preprocessing, model construction (generator, discriminator), and the final model (GAN) training can be seen in the pseudocode presented in Algorithm 1. Subsequently, the steps followed during the implementation process are detailed.

```
NO_ITER = 5,000
NO_IMAGES = 10,000
BATCH = 6
IMG_WIDTH = 178
IMG_HEIGHT = 218
TARG_WIDTH = 128
TARG_HEIGHT = 128
dif = (IMG_HEIGHT – IMG_WIDTH // 2)
crp = (0, diff, IMG_WIDTH, IMG_HEIGHT – dif)
IMGS = []
LATENT_DIM = 32
CHANNELS = 3
IMGS = Crop_Images()
generator = build_generator(gen_input=LAT_DIM,
                                      LAYERS=8,
                                      ACT_FUNC_LAY=‘Leaky ReLU’,
                                      ACT_FUNC_OUT=‘tanh’,
                                      )

discriminator = build_discriminator(disc_input=(TARG_WIDTH,
                  TARG_HEIGH, CHANNELS),
                                          LAYERS=8,

ACT_FUN_LAY=‘LeakyReLU’,
                                          ACT_FUNC_OUT =‘sigmoid’,
                                          OPTIMIZER=‘RMSprop’,
                                          LOSS_FUN=‘binary
                       crossentropy’
                  )

GAN_input = Input(shape=(LAT_DIM, ))
GAN_output = discriminator(generator(GAN_input))
GAN = Model(GAN_input, GAN_output)
GAN.compile(optimizer=‘RMSprop’,
      loss=‘binary_crossentropy’)

for t in range(NO_ITER)
  latent_vectors = samples(size=(batch_size, LATENT_DIM))
  generated = generator.predict(latent_vectors)
  real = IMGS[batch_size]
  combined_images = np.concatenate([generated, real])
  dis_loss = discriminator.train_on_batch(combined_images)
  gen_loss = GAN.train_on_batch(latent_vectors)
end for
```

Lines 1–12 contain the variables declaration. A total of 10,000 images, 5,000 iterations, and a batch of size six were used to train the model. The initial size of the images was 178*218 with three color channels (RGB). Subsequently, the images were cropped to center the face and standardize them, leaving a size of 128*128. Each cropped image was stored in an array. It is represented in Line 13, and a sample grid can be seen in Fig. 8, which contains some resized images.

The generator is built by giving as input the latent vector, which was defined in size 32. The model is also created with its respective layers; in this case, eight layers were constructed, and the Leaky ReLU activation function was applied to them. On the other hand, in the output layer, the hyperbolic tangent activation function was used to produce sharper images. It is also important to mention that the loss of this network is calculated from the discriminator's loss. The definition of the generator can be seen in Line 14.

The discriminator receives as input an image with three RGB color channels. It was built with a structure very similar to the generator's, adding eight layers, and the Leaky ReLU activation function was also applied to them. But the sigmoid activation function was used in its output layer, which gives the binary classification 0 and 1.

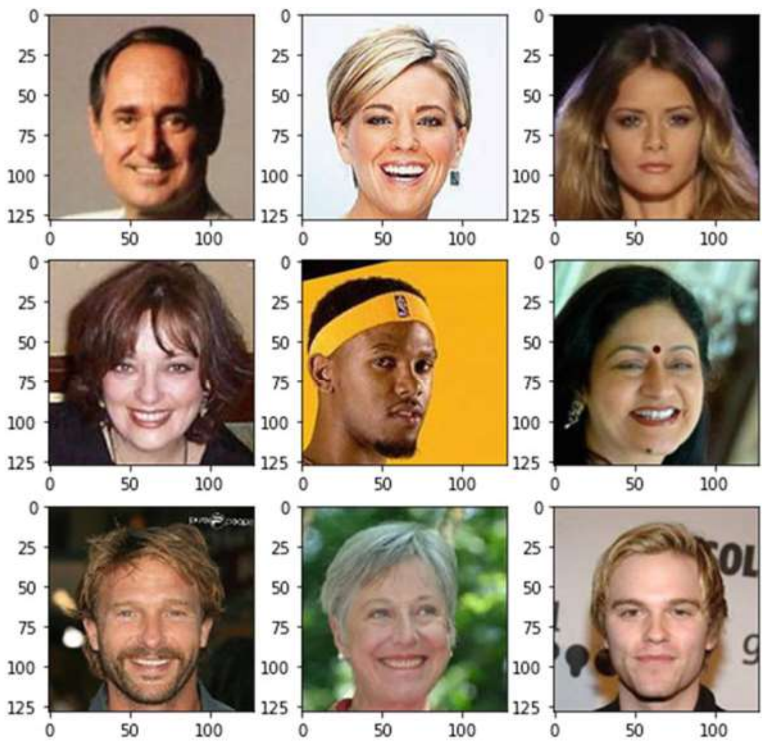


Fig. 8 Preview of the resulting images after preprocessing, belonging to the Celeba public image set

The RMSprop optimizer was employed, and the binary cross entropy was used as a loss function due to its utility in classification problems to measure the difference between the probabilities calculated between two possible classes. The definition of the model can be seen in Line 15.

The GAN model outputs the classification resulting from the discriminator receiving as a parameter the generator, which in turn gets the latent vector. As in the discriminator, the RMSprop optimizer was applied to the GAN model and used the binary cross-entropy as a loss function. This process is defined in Lines 16–19.

The training is performed during the previously defined iterations. The “latent_vectors” variable contains the noise with which the generator will be trained. Subsequently, the “real” variable has a subset of images from the original set, and then they are merged with those coming from the generator in the “combined_images” variable. This part can be seen in Lines 20–24.

Line 25 shows how the discriminator’s loss is calculated from its training by receiving the subset created in “combined_images” as input.

Furthermore, Line 26 shows the generator’s loss calculated from the GAN’s training that receives as input the noise vector, and this will be passed to the generator to start the training of the complete model. Finalizing with the loop in Line 27 when the last epoch is reached.

After training, the resulting images can be seen in Fig. 9. It can be seen how the characteristics of the images improve, learning to generate parts of the face such as eyes, mouth, nose, and textures for the background. The loss values of each network after training can be seen in Table 3, while Fig. 10 presents a plot to display the behavior of the respective loss functions.

Fig. 9 Images generated by the trained GAN

Table 3 Model losses at the end of the training

Epochs	Dis_loss	Gen_loss
4,900/5,000	0.7826	0.8347
4,950/5,000	0.6075	1.0423
5,000/5,000	0.8056	0.6091

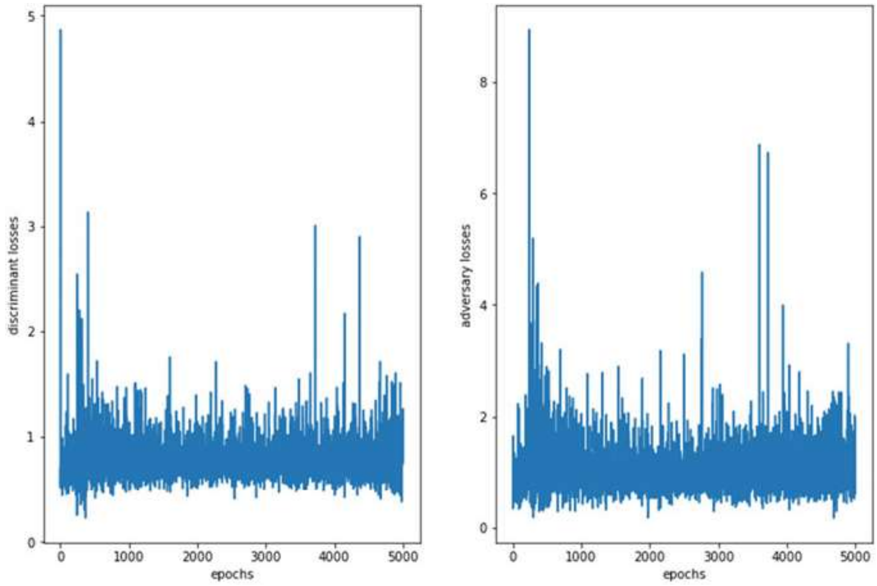


Fig. 10 Model losses during GAN training

Since the GAN model comprises two neural network models, understanding its behavior is highly complex. It is irregular, with highs and lows at seemingly random points (cf. [43]).

7 Conclusion and Future Work

The literature review, which describes the most widely used methods over time for artificial image generation, shows that GANs are effective models in this task to date, being successfully implemented in different domains. Despite the fact that they were recently proposed and the challenges their implementation represents, they show promising results. The more they are investigated, the more possibilities open up in the world of AI. One of the main challenges that were observed not only in research papers but also in experiments like the one presented above in Sect. 6, is the high computational power that they require. The higher the computational power, the better the quality and realism of the generated images.

In addition to the above, there is a large open area of research on the functioning of the GAN. The reason for the behavior they display during training is still unclear. For example, it is unknown why securities sometimes trigger when measuring their losses. Despite the above, it is only known that the images generated improve as the training progresses.

The metric for evaluating an artificial image is a significant limitation. Assessing whether an image looks realistic represents a high level of complexity for the human eye. Many papers resort to multiple attempts to measure it. Some use quantitative metrics, while others focus more on the human side with qualitative measures.

As future work, it is planned to analyze, evaluate, and implement a GAN architecture with the objective of simulating the age progression in the facial images of Mexican women reported missing. The idea is to visualize their appearance after a lapse of 10–20 years of adulthood. With this, it is intended to use GANs as a DA technique to create robust image sets that allow training artificial intelligence algorithms to help find missing women.

8 Code Repository

A repository that contains the code implemented for the generation of artificial faces is presented below, where is available a Jupyter Notebook that shows the results of each compiled code block.

https://github.com/BlancaECS/Face_generation/blob/5584e7fb9c7041353182151174407aa487048d53/FirstFace_GAN.ipynb.

References

1. Saragih, R.E., To, Q.H.: A survey of face recognition based on convolutional neural network. Indones. J. Inf. Syst. **4**(2), 122–139 (2022). https://doi.org/10.24002/ijis.v4i2.5439
2. Jain, A.K., Li, S.Z.: Handbook of Face Recognition, vol. 1. Springer, New York (2011). https://doi.org/10.1007/978-0-85729-932-1
3. aws, ¿Qué es el reconocimiento facial? (2022). https://aws.amazon.com/es/what-is/facial-recognition/
4. Masi, I., Trân, A.T., Hassner, T., Leksut, J.T., Medioni, G.: Do we really need to collect millions of faces for effective face recognition? Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9909. LNCS, pp. 579–596 (2016). https://doi.org/10.1007/978-3-319-46454-1_35
5. Wang, X., Wang, K., Lian, S.: A Survey on Face Data Augmentation. No. 1 (2019). https://doi.org/10.1007/s00521-020-04748-3
6. Langr, J., Bok, V.: GANs in action: deep learning with generative adversarial networks. Manning (2019)
7. Perarnau, G., van de Weijer, J., Raducanu, B., Álvarez, J.M.: Invertible conditional GANs for image editing. No. Figure 1, pp. 1–9 (2016). https://doi.org/10.48550/arXiv.1611.06355
8. Sun, S., Wang, A.: Face Aging With GAN
9. He, Z., Zuo, W., Kan, M., Shan, S., Chen, X.: AttGAN: facial attribute editing by only changing what you want. IEEE Trans. Image Process. **28**(11), 5464–5478 (2019). https://doi.org/10.1109/TIP.2019.2916751
10. Mohamed, E., Elharrouss, O., Al-Maadeed, S., Tairi, H.: Image generation: a review. Neural Process. Lett. **54**(5), 4609–4646 (2022). https://doi.org/10.1007/s11063-022-10777-x

11. Calderon, J.D., Fabrico, L., Ticliahuanca, M., Roncal, M.: Uso de algoritmos de machine learning para el diagnóstico de melanomas Trabajo de Investigación para optar el Grado de Bachiller en Ingeniería Mecánico-Eléctrica Jose Dario Calderon Ortiz María Esther Roncal Moscol, no. May, 2021. https://doi.org/10.13140/RG.2.2.20397.00483
12. Taylor, L., Nitschke, G.: Improving Deep Learning using Generic Data Augmentation (2017). https://doi.org/10.48550/arXiv.1708.06020
13. Hsu, W.H.: Investigating Data Augmentation Strategies for Advancing Deep Learning Training (2018)
14. Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J.: Learning to discover cross-domain relations with generative adversarial networks. In: 34th International Conference on Machine Learning, ICML 2017, vol. 4, pp. 2941–2949 (2017). https://doi.org/10.48550/arXiv.1703.05192
15. Guom, D., Sim, T.: Digital face makeup by example. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 73–79 (2009). https://doi.org/10.1109/CVPRW.2009.5206833
16. Li, T., Qian, R., Dong, C., Liu, S., Yan, Q., Zhu, W., Lin, L.: BeautyGAN, pp. 645–653 (2018). https://doi.org/10.1145/3240508.3240618
17. Guo, J., Zhu, X., Lei, Z., Li, S.Z.: Face synthesis for eyeglass-robust face recognition. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10996. LNCS, no. June, pp. 275–284 (2018). https://doi.org/10.1007/978-3-319-97909-0_30
18. Shen, W., Liu, R.: Learning residual images for face attribute manipulation. In: Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 1225–1233 (2017). https://doi.org/10.1109/CVPR.2017.135
19. Feng, Z.H., Kittler, J., Christmas, W., Huber, P., Wu, X.J.: Dynamic attention-controlled cascaded shape regression exploiting training data augmentation and fuzzy-set sample weighting. In: Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 3681–3686 (2017). https://doi.org/10.1109/CVPR.2017.392
20. Lv, J.J., Shao, X.H., Huang, J.S., Zhou, X.D., Zhou, X.: Data augmentation for face recognition. Neurocomputing **230**, 184–196 (2017). https://doi.org/10.1016/j.neucom.2016.12.025
21. Deng, J., Cheng, S., Xue, N., Zhou, Y., Zafeiriou, S.: UV-GAN: adversarial facial UV map completion for pose-invariant face recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 7093–7102 (2018). https://doi.org/10.1109/CVPR.2018.00741
22. Xie, W., Shen, L., Yang, M., Jiang, J.: Facial expression synthesis with direction field preservation based mesh deformation and lighting fitting based wrinkle mapping. Multimed. Tools Appl. **77**(6), 7565–7593 (2018). https://doi.org/10.1007/s11042-017-4661-6
23. Thies, J., Zollhöfer, M., Nießner, M., Valgaerts, L., Stamminger, M., Theobalt, C.: Real-time expression transfer for facial reenactment. ACM Trans. Graph **34**(6), 1–14 (2015). https://doi.org/10.1145/2816795.2818056
24. Chandaliya, P.K., Nain, N.: Child face age progression and regression using self-attention multi-scale patch GAN. In: 2021 International Joint Conference on Biometrics, IJCB 2021, pp. 1–8 (2021). https://doi.org/10.1109/IJCB52358.2021.9484329
25. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Commun. ACM **63**(11), 139–144 (2014). https://doi.org/10.1145/3422622
26. Nguyen, A., Yosinski, J., Clune, J.: Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 07-12-June, pp. 427–436 (2015). https://doi.org/10.1109/CVPR.2015.7298640
27. Calcagni, L.R.: Redes Generativas Antagónicas y sus aplicaciones, p. 72 (2020)
28. Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X., Wang, F.Y.: Generative adversarial networks: introduction and outlook. IEEE/CAA J. Autom. Sin. **4**(4), 588–598 (2017). https://doi.org/10.1109/JAS.2017.7510583

29. Raj, B.: Advances in Generative Adversarial Networks (GANs) (2019). https://medium.com/beyondminds/advances-in-generative-adversarial-networks-7bad57028032
30. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. In: 4th International Conference on Learning Representations, ICLR 2016-Conference Track Proceedings, pp. 1–16 (2016). https://doi.org/10.48550/arXiv.1511.06434
31. Liu, M.Y., Tuzel, O.: Coupled generative adversarial networks. Adv. Neural Inf. Process. Syst., no. Nips, 469–477 (2016). https://doi.org/10.48550/arXiv.1606.07536
32. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. In: 6th International Conference on Learning Representations, ICLR 2018-Conference Track Proceedings, pp. 1–26 (2018). https://doi.org/10.48550/arXiv.1710.10196
33. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. IEEE Trans. Pattern Anal. Mach. Intell. **43**(12), 4217–4228 (2019). https://doi.org/10.1109/TPAMI.2020.2970919
34. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of StyleGAN. In: Cvpr, pp. 8107–8116 (2019). https://doi.org/10.48550/arXiv.1912.04958
35. Waheed, A., Goyal, M., Gupta, D., Khanna, A., Al-Turjman, F., Pinheiro, P.R.: CovidGAN: data augmentation using auxiliary classifier GAN for improved Covid-19 detection. IEEE Access **8**, 91916–91923 (2020). https://doi.org/10.1109/ACCESS.2020.2994762
36. Jabeen, S., Khan, M.U.G., Saba, T., Rehmat, A., Rehman, A., Tariq, U.: A realistic image generation of face from text description using the fully trained generative adversarial networks. IEEE Access **9**, 1250–1260 (2021). https://doi.org/10.1109/ACCESS.2020.3015656
37. Bailo, O., Ham, D., Shin, Y.M.: Red blood cell image generation for data augmentation using conditional generative adversarial networks. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, vol. 2019-June, pp. 1039–1048 (2019). https://doi.org/10.1109/CVPRW.2019.00136
38. Metri, O., Mamatha, H.R.: Medical image generation using generative adversarial networks. In: Generative Adversarial Network Image-to-Image Translation, pp. 235–262 (2021). https://doi.org/10.1016/B978-0-12-823519-5.00007-5
39. Dash, A., Ye, J., Wang, G., Jin, H.: High resolution solar image generation using generative adversarial networks. Ann. Data Sci., 1–10 (2022). https://doi.org/10.1007/s40745-022-00436-2
40. Yasmina, D., Karima, R., Ouahiba, A.: Traffic signs recognition with deep learning. In: Proceedings of the 2018 International Conferences on Applied Smart Systems, ICASS 2018, no. June (2019). https://doi.org/10.1109/ICASS.2018.8652024
41. Yepes-Nuñez, J.J., Urrútia, G., Romero-García, M., Alonso-Fernández, S.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Rev. Esp. Cardiol. **74**(9), 790–799 (2021). https://doi.org/10.1016/j.recesp.2021.06.016
42. Kaggle, Celeba dataset. https://www.kaggle.com/datasets/jessicali9530/celeba-dataset
43. Mejía, J., Ochoa-Zezzatti, A., Contreras-Masse, R., Rivera, G.: Intelligent system for the visual support of caloric intake of food in inhabitants of a smart city using a deep learning model. Appl. Hybrid Metaheuristic Algorithms Image Process., 441–455 (2020). https://doi.org/10.1007/978-3-030-40977-7_19

May 31, 2023

Who may concern:

By this letter, the editors certify that the acceptance of the following chapter was the result of a double-blind peer-review process:

Chapter title:	Data augmentation techniques for facial image generation: a brief literature review
Authors:	Blanca Elena Cazares Salais, Rogelio Florencia Juárez, Vicente García Jiménez and Julia Patricia Sánchez Solís
Book title:	Data Analytics & Computational Intelligence: Novel Models, Algorithms and Applications
Editors:	Laura Cruz-Reyes, Bernabé Dorronsoro, Gilberto Rivera, and Alejandro Rosete
Book series:	Studies in Big Data (Springer)

Furthermore, the editorial review process complied with the publishing agreement stated in the Springer Contract # 145403. *Studies in Big Data* is currently indexed in SCOPUS, SCImago, and EI Compendex.

This contributed book was an editorial initiative of the Eurekas Community. Eurekas is an international and multidisciplinary scientific research network that joins professionals in mathematics, computer sciences, engineering, administration, economics, and social sciences. It was founded in 2008 and is currently integrating more than 60 research groups in more than 20 countries, mainly in America and Europe. The submitted chapters were accepted only after a stringent review process by our collaborators worldwide, coordinated by the editors.

As evidence, the following documents are enclosed: (i) Initial version of the manuscript, (ii) review report, (iii) revision notes, (iv) revised manuscript, (v) decision letter, and (vi) preprint.

Please, do not hesitate to contact me with any doubts or questions regarding this letter.

Sincerely yours,

Dr. Gilberto Rivera
VICE PRESIDENT OF THE EUREKAS COMMUNITY
COORDINATOR OF PUBLICATION PROJECTS

with the editors' approval,

Dr. Laura Cruz-Reyes	Dr. Bernabé Dorronsoro	Dr. Alejandro Rosete
TECNOLÓGICO NACIONAL DE MÉXICO (MÉXICO)	UNIVERSIDAD DE CÁDIZ (SPAIN)	UNIVERSIDAD TECNOLÓGICA DE LA HABANA "JOSÉ ANTONIO ECHEVERRÍA" (CUBA)

(i) Initial version of the manuscript

Face image generation: A review

Abstract. Image processing has gained notoriety over the last few years in performing various tasks through deep learning algorithms, such as face recognition and identity verification. Unfortunately, most of them require a large set of images for training, usually manually labelled, which is a costly task both in time and effort, not to mention being prone to human error. Data Augmentation (DA) techniques have been used to mitigate this situation, as they generate new images by applying variations to real image sets. For this purpose, Generative Adversarial Networks (GANs) have recently been successfully used to create synthetic images. This chapter presents a brief literature review on various DA methods dedicated to image generation. After this, the technique that has presented outstanding results in the face generation task are GANs. General background about them is given: their definition, architecture, training and challenges. Some works that have used GANs in artificial image generation are presented. Subsequently, a GAN architecture for face generation is implemented, obtaining one hundred 6 x 6 grids with faces artificially created from a public set of images. After this, the relevance of having a large computational capacity for training was identified when seeking to generate images with great sharpness and realism.

Keywords: Face image generation, Data augmentation, Generative Adversarial Networks.

1 Introduction

Facial recognition systems have been discussed for several decades. This is investigated with such attention given the tremendous human ability to recognize others through the face, being a unique part of the body. The search is on for such systems to possess this same intelligence, also known as biometric identification.

Facial recognition has been developed with great interest in the field known as computer vision [1]. This task supports a variety of critical applications, e.g., identity verification. This is done using what is known as a face analyzer, which is software that confirms the identity of people based on their faces [2]. This is achieved through the identification and measurement of facial features in images. It can also be used to associate human faces in the latest or even in videos to check how similar they are to one or several specific individuals. In addition, it can determine the level of similarity between two photographs to find out if they belong to the same person or search for them among a large set of images in a collection. For example, biometric security systems use facial recognition to uniquely identify individuals as users at login to strengthen user authentication. In addition, mobile devices often use this type of technology to protect the data they contain [3]. Unfortunately, large quantities of facial images are required for the correct operation of these processes, which is why much of the research has focused on the generation of artificial images. This aims to augment the sets of images used to train different architectures because this significantly impacts the results [4].

The task of collecting and labelling data samples with good quality is a costly task, both in time and effort, as well as is prone to human error. For this purpose, various techniques known as DA have been implemented, and their performance is suitable in different domains.

DA techniques allow the size of existing image sets to be increased considerably through simulations [4], which helps significantly with the image requirements for learning systems. Also in addition, AD techniques have faced several limitations of their own [5], for example:

- The images generated lack realistic variations such as makeup, skin color, and background change, which means that they would have a different distribution than the real ones.
- Creating high-quality facial images is very difficult due to the complexity of facial details.

After many years of research and application of different modelling methods, in 2014 a technique was proposed that allows the generation of realistic images through the use of two artificial neural networks, called GANs. Although these are not the first method used for artificial data generation, their results and versatility distinguished them from the rest since they have achieved outstanding results that were still considered impossible for artificial systems[6].

GANs are a machine learning technique that consists of integrating two simultaneously trained models: one called "generator", which is trained to create the artificial images and on the other hand, the "discriminator", whose objective is to discern between the fake images created by the generator and the real ones coming from the original training set. Its performance has far exceeded expectations in the field of artificial image generation, according to a variety of authors [7][8][9][5].

As with ML algorithms, the data generated by the GAN will depend entirely on the training set provided to perform the learning. For example, if a GAN is required to learn how to create images of handwritten numbers, it is necessary to use a training set containing multiple images of the handwritten numbers [10].

GANs are derived from a gaming perspective. Hence the word adversarial in their name denotes a competitive dynamic between the two models that compose them. The generator's goal is to create real images that are indistinguishable from the actual training set. On the other hand, the discriminator aims to distinguish these generated samples from the real ones. Therefore, the better the generator is at creating realistic images, the better the discriminator has to become at distinguishing the real ones from the fake ones [6].

The fast growth and progress of GANs have been due to research and development, generating new architectures to stabilize the outputs and generate images of higher quality and realism. In this way, the aim is to create impossible images for the human eye to differentiate.

This chapter describes in a general way the model based on deep learning: GAN, which has been used throughout recent years to create artificial facial images, highlighting the efficiency it has demonstrated. Implementing, in turn, an architecture, which was trained with a small part of the public image set "CelebA" and a 5000 epochs training. The paper is structured as follows. Section 2 describes different types of transformations commonly performed on images to augment the image sets. Section 3 describes a GAN architecture, training and challenges, among other data of interest. Section 4 presents a collection of related works that have employed GANs for artificial image generation. Section 5 shows the architecture of a GAN dedicated to facial imaging. Finally, Section 6 presents conclusions and future work.

2 Face image generation

Four main categories can be broadly discussed in the DA task for generating artificial images: generic, component, attribute and age transformations. Section 2.1 presents the generic transformations. Section 2.2 presents the modifications produced on image components. Section 2.3 presents the facial attribute transformations. Section 2.4 presents the regression and age progression.

2.1 Generic transformations

These transformations focus on modifying the entire image, ignoring high-level components such as composition, light, volume, symmetry, shape and texture. They are usually divided into two main groups: geometric and photometric transformations.

Geometric transformations are commonly applied in multiple computer vision tasks [5], such as face recognition, healthcare and manufacturing applications. In general terms, geometric transformations alter the pixels of an image by placing them in new positions. Some examples of these can be:

- Reflection: consists of flipping the image around its vertical or horizontal axis.
- Rotation: rotates the image θ degrees around its center, bringing each pixel (x, y) to its position (x', y').
- Cropping: consists of cutting the images to a specific size.
- For example, in the case of convolutional neural networks (CNNs), this type of transformations help to minimize their sensitivity to changes in position and orientation [11]. Some examples can be seen in Fig. 1.

Fig. 1 Geometric transformation examples [5]

On the other hand, photometric transformations, shown in Fig. 2, generally work by altering the RGB (red, green and blue) channels, shifting each pixel value (r, g, b) to new values (r', g', b') of an image according to predefined rules. These transformations adjust the illumination and color, leaving the geometry unaffected [11]. For example, we can mention color manipulations, such as inverting them or adding some filters, such as blurring or grayscale [5].

Fig. 2 Photometric transformation examples [5]

These transformations are mainly used in computer vision tasks to enrich the training sets and prevent a common problem in this field, overfitting. In [12] is an article dedicated to evaluating the performance of CNNs, trained with a set of images enriched using both geometric and photometric transformations. In [13] various transformations were employed to increase the image set and prevent overfitting.

2.2 Component transformation

In turn, there are transformations dedicated to enriching the sets of facial images by modifying the person's components. This is to train the algorithm to recognize the person even if their appearance is altered.

The hairstyle can be considered one of the components to be generated since, although it is not considered a facial component, hairstyle affects face detection and recognition because it tends to hide certain features of a person's appearance. Therefore, DA techniques focus on generating facial images with different variations in hair, for example, color, shape and bangs. In [14], a method using DiscoGAN was proposed. This variant learns to discover relationships between different domains and develops the ability to translate features between them, for example, by transforming hair color.

Makeup transfer and accessory addition techniques can also be identified due to the difficulty for recognition systems to effectively perform their tasks when some features of the face look different depending on the makeup or accessories a person wears. Most studies based on this type of transformation can be divided into two categories: traditional image processing [15] and those based on deep learning [16]. Some examples of these can be seen in Fig. 3.

Another component that impacts facial recognition is the use and removal of accessories, including glasses, earrings, and piercings, among others. Of all these, glasses are the most commonly used, as they are used for different reasons, for example, vision correction, prevention against sunlight, eye protection, and aesthetics, among others. They significantly affect the accuracy of facial recognition, as they usually cover a large area of the face.

In [17], a fusion of virtual lenses onto faces was performed using the Augmented Reality (AR) technique. Whereas in [18] a method of facial attribute manipulation based on image residuals was proposed, defining this as the difference between the input image and the desired output image.

Fig. 3 Component transformation examples [5]

2.3 Attribute transformation

There are also some transformations dedicated to modifying aspects such as pose. In this case, the position of the head in a photograph is considered a significant challenge in facial recognition tasks since any variation in it tends to modify visual aspects of the face, i.e., it can hide or show different facial details. In addition, it has been considered an essential aspect since facial photographs in several legal processes are requested from the frontal side. This leads to the research of recreating how a face may look from other angles.

Facial expressions are also considered critical: happy, annoyed, scared, and surprised, among others. These techniques help to improve the performance of emotion classification and recognition systems. Primarily 2D, 3D modelling and learning-based modelling approaches are used to achieve this goal and usually focus on modifying the expression of a face using expression templates by concentrating on a series of points, for example, the corners of the mouth, the cheeks position, and the location of the eyebrows. Others focus on simulations to recreate face parts that are hidden by the pose in which the photograph was taken [19][4][20][21]. Some examples of this can be found in Fig. 4.

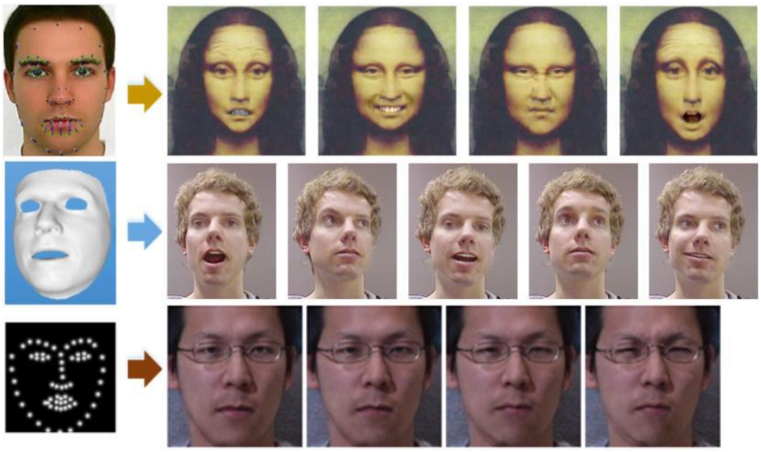


Fig. 4 Facial expresión synthesis [22][23]

2.4 Age progression and regression

Finally, age progression or regression is based on the predicting ability of what a current face will look like in past or future while preserving the characteristics of each individual. It has become a widely explored research topic because it significantly affects various applications, including tracing missing persons, facial recognition, and cosmetic studies, among others. The two main concerns in attempting age regression or progression are the preservation of identity and the accuracy of predictions [24].
These facial images enrich the image sets by adding different features at different time stages of the same face, which makes the models more robust to age variations.
Age modification methods are mainly divided into two: prototype-based and model-based. The first one creates an average face for different age ranges, learns the shapes and textures of these and applies these features to transfer them to a new face. However, the individual's characteristics are often lost with these methods [5].
Model-based methods build, as the name implies, models of biological changes due to age in faces, e.g. musculature, wrinkles, and skin texture, among others. They are called generative because they are such a powerful tool for creating new data by learning to imitate the probabilistic distribution of a training set. Nowadays, generative models have gained importance and attention due to their good performance in data creation. Among the most popular are autoregressive models, Variational Autoencoders (VAEs) and GANs. The disadvantage of these models is their complexity and computational cost [5].
GANs, as mentioned above, are an alternative architecture for training generative models since they handle probabilistic computations very well. Recent work has begun to

apply them to the age regression and progression task, and many variants of the model have been generated. Some examples of this generative task are shown in Fig. 5.

Fig. 5 Facial age regression and progression examples [5]

Before GAN, two approaches to age progression and regression in faces were prototype-based and modelling-based. In the latter, critical points in the image, such as eyes, nose, and jaw, track, in turn, the temporal changes such as wrinkles, musculature and color in these are identified. However, this method requires a large amount of age-labelled data over a long period for each individual, which is difficult to find and computationally expensive. The prototype-based method creates an average face based on a set of images of a particular age group, using it to transfer those features from one age range to another. The disadvantage of this is that personal features are often lost. Another possibility is to use neural networks to transform faces across ages. This generates smoother images but still requires images labelled with the person's age through the years [8].

On the other hand, the GAN consists of a discriminator and a generator competing with each other based on the min-max games. The generator starts by receiving as input a noise vector z and creates an image that it gives to the discriminator to receive feedback from it. Some variants of this architecture are [8]:

- DCGAN has demonstrated that GAN can be successfully applied to generate indoor scenarios and human faces.
- StyleGAN made a significant extension to the basic GAN to progressively generate high resolution images from those with very low resolution.

- cGAN introduces an identity preservation vector with the optimization approach when generating faces so there is a better match between the original and the created face.
- Pyramid GAN simulates the effects of age more sharply and presents a suite of methods for assessing accuracy and fidelity to the original image.

3 Generative Adversarial Networks (GANs)

GANs are a powerful type of unsupervised learning algorithm based on Artificial Intelligence (AI), which aims to learn the estimated probability distribution in a specific training set. Ian Goodfellow developed it in 2014 [10]. It is a competition system between two neural network models which try to maximize their performance while minimizing their adversary's, developing the ability to analyze, capture and copy the variations presented by a particular set of images [8].

These networks can generate artificial data and have been one of the generative models with the highest result quality, especially when analyzing their potential to generate high-resolution images [6]. Section 3.1 defines what a GAN consists. Section 3.2 describes the architecture with which GANs were originally proposed. Section 3.3 explains how GANs are trained. Section 3.4 mentions the challenges faced by researchers when using a GAN.

3.1 Definition

It has been demonstrated that most neural networks can be easily tricked into misclassifying by adding only a small amount of noise to the original data. Surprisingly, after this addition, the model develops a higher confidence level in the wrong predictions than in the correct ones. This is because most machine learning (ML) algorithms learn from a limited amount of data, which is prone to model overfitting [25]. This motivated the creation of GANs. They can be described in three parts:

- Generative: They are considered generative models since they describe how new data are generated in terms of probabilistic models.

- Adversarial: The model is trained by competition among its neural networks, i.e., they are considered adversarial to each other.
- Networks: They use neural networks as the primary training algorithms.

They are based on game theory, which considers players to be both ML models, typically implemented using neural networks. A network is called a generator, which can learn the distribution obtained from an original data set to try to replicate it. This is achieved by inserting a noise vector z, i.e. random numbers with a Gaussian distribution. The main objective of the generator is to learn how to transform unstructured noise z into realistic samples [10].

The other player is called the discriminator. It examines each example x received as input and outputs an estimate of whether it is true or false.

Each player has a cost, so they try to minimize their own, i.e., the discriminator's cost encourages it to classify the data as real or fake correctly. In contrast, the cost of the generator encourages it to generate data that the discriminator incorrectly classifies as real.

A typical example when talking about GANs is to imagine that one network represents a money counterfeiter and the other a policeman. The former generates counterfeit bills while the policeman tries to arrest him for benefiting the production of legitimate ones. The competition leads to the production of more and more realistic counterfeit bills until, eventually, the counterfeiter produces them so realistic that the policeman cannot tell the difference between authentic and counterfeit [6].

3.2 Arquitectura

In GAN architecture, the generator will start training simultaneously as the discriminator. The same will need a few epochs before starting the adversarial training since it must be able to classify the images correctly.

Its architecture consists of two deep neural networks, one generative and the other discriminative, which compete with each other. The generator produces new data instances, while the discriminator tries to distinguish accurately between the real data, i.e., those coming from the original set or fake data produced by the generator.

This competition will continue until the generator can create realistic artificial data, which can then be used as input to other neural networks [10]. Fig. 6 shows a GAN schematic with its basic components.

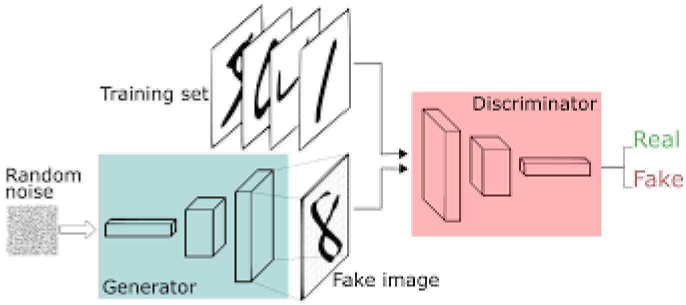


Fig. 6 GAN architecture

Furthermore, since its development, many different architectures have been proposed to deal with a wide variety of domains. It could even be said that weekly, several scientific papers are published [26]. In [27] an extensive literature review is done on multiple architectures developed for GANs.

3.3 Training

The training phase requires the two networks and a set of data from which the artificial data will be generated. First of all, it is worth mentioning that the training of the generator is much more complex than the discriminator, which can be seen more as a binary classifier.

Discriminator receives inputs, where some are real and need to be labelled as such, while others are from the generator and must be labelled as false. Its responses are then 1 if the input data is real or 0 if it is false.

On the other hand, the generator must be trained with the only condition that the data created mislead the discriminator, i.e., it must minimize its loss and maximize that of the opponent. To achieve this, the generator's output must be the input of the discriminator so that the output of the whole model gives as output the probability that the data are real, according to the discriminator. In this way, the generator obtains feedback from the discriminator, which it uses to create progressively more similar data in the training set [10].

3.4 Challenges

The GANs are still facing research challenges regarding their training. Among them, the following can be pointed out [28]:

- Non-convergence: this is when the generator and the discriminator fail to reach the desired equilibrium (50%). Their respective loss functions begin to fluctuate without being able to reach stability.

- Modal collapse: occurs when the generator produces similar data, even though the inputs vary in characteristics. It finds a small set of samples that successfully deceive the discriminator and are thus incapable of producing others. In these cases, the gradient of the loss function is stuck at a value close to 0.
- Non-informative loss: the general intuition is that the lower the loss of the generator, the higher the quality of the data it produces. However, its loss must be compared with the discriminator's, which is constantly improving. Therefore, the issue of model evaluation is more complex. The generator may produce better-quality samples even as the loss function increases.

Due to these challenges, techniques have been developed with the aim of reducing them as much as possible. However, they are still the subject of research at the moment [10].

4 Related work

The following works have focused on implementing some GAN variants with different objectives.

In [9], it is mentioned that the main objective is to generate a new face by editing facial attributes in the images, preserving its identity. Subsequently, it proceeds to perform a literature review, finding that GAN architecture with an encoder is usually incorporated for such tasks with promising results. Thus, it is proposed to apply an attribute classification constraint to the generated images, in addition to the previous architecture, forming together what is called AttGAN. With this, experiments were performed on the public set of images "Celeba" to manipulate attributes such as hair color, beard, and age, achieving realism and preservation of facial details.

In [8], a CNN was developed to generate age progression in faces. In order to do this, the authors conducted a thorough literature review. They analyzed several sets of images to select the one with the greatest variety and fit the domain for which they wanted to develop the model. Then, they configured parameters and loss functions for network stabilization and the creation of realistic photographs. The evaluation metrics were qualitative and quantitative, resulting in images with appropriate characteristics for age progression.

The makeup transfer is the object of study to transfer a specific style to a clean face, preserving the identity of the same. This type of problem at the instance level is considered a great challenge since the styles vary greatly, for example, eye shadows, lipsticks, and foundations, among many others. In [16], BeautyGAN is proposed so that the networks can perform the transfer at the instance level through unsupervised adversarial learning. At the end of the study, a new set of high-resolution makeup images was constructed.

GANs even have an impact in the medical area, as can be seen in [29]. A study on coronavirus (COVID-19), a viral disease caused by the SARS-CoV-2 respiratory syndrome, begins, encouraged by its global effect on health and economics. Chest X-rays from infected patients were a crucial step in controlling this virus. Thus began the introduction of various deep learning systems and studies that demonstrated the efficiency of using chest X-rays for patient detection. In this context, since CNNs require a significant amount of training data to perform adequately and the virus was too recent to have enough chest X-rays to generate systems that learn to detect it, the authors present a method to generate images from chest X-rays in short times. This was achieved by introducing a model named CovidGAN, proving that the artificial images produced by it helped improve the performance of the trained CNNs for COVID-19 detection, increasing their accuracy to 95%.

5 Face image generation with GANs

The architecture of a GAN for face image generation, trained from the public image set "Celeba" [30], is shown below. This consists of more than 200,000 celebrity facial images. It was selected because it has several suitable variations for training face detection models, such as hair color, smile, and glasses. Besides the mentioned above, it also varies in poses, backgrounds and people diversity, with many images and annotations for efficient use in computer vision.

For the GAN implementation, a virtual machine generated and configured by the Laboratorio Nacional de Tecnologías de Información (LaNTI)[1].

at the Universidad Autónoma de Ciudad Juárez (UACJ) was used. The system specifications are: Ubuntu 20.04, dual-core, 100 GB storage and 8 GB RAM.

Since a large computational capacity is required for the GAN training, a series of tests had to be performed by modifying its parameters until a training that met the requirements in time and form was achieved.

A total of 10,000 images were used for model training, 5,000 epochs and a batch of size 3. A variable was used to print the percentage of losses every 50 epochs to visualise training performance.

```
NUMERO_ITERACIONES_A_GUARDAR = 50
NUMERO_IMAGENES = 10000
NUMERO_ITERACIONES = 5000
BATCH = 3
```

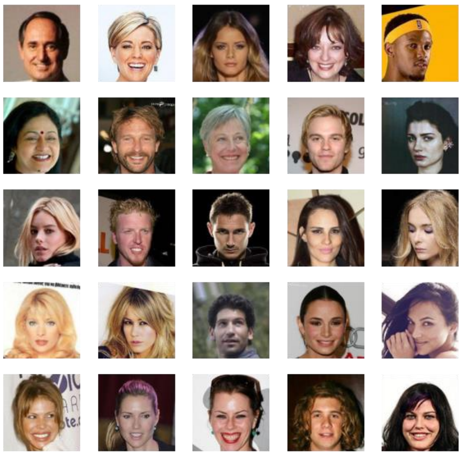
First, the directory that would contain the generated images was selected. Then each element was resized to standardize, using the "crop" method to cut each side of the image. At the end, a size of 128 x 128 pixels was defined. In addition, 3 color channels, i.e. RGB, were used.

```
ORIG_WIDTH = 178
ORIG_HEIGHT = 208
diff = (ORIG_HEIGHT - ORIG_WIDTH) // 2

WIDTH = 128
HEIGHT = 128

crop_rect = (0, diff, ORIG_WIDTH, ORIG_HEIGHT - diff)

images = []
for pic_file in tqdm(os.listdir(PIC_DIR)[:IMAGES_COUNT]):
    pic = Image.open(PIC_DIR + pic_file).crop(crop_rect)
    pic.thumbnail((WIDTH, HEIGHT), Image.ANTIALIAS)
    images.append(np.uint8(pic))

images = np.array(images) / 255
print(images.shape)
```

[1] Laboratorio Nacional de Tecnologías de Información: http://www.lanti.org.mx/lanti/

In a display, it was possible to check that the faces were not affected by the resizing, i.e. the faces were not cropped of any important attributes of the faces, as this would affect the generated images.

```
 plt.figure(1, figsize=(10, 10))
for i in range(25):
    plt.subplot(5, 5, i+1)
    plt.imshow(images[i])
    plt.axis('off')
plt.show()
```



The generator and discriminator were created with very similar structures. The former is composed of 8 layers and uses a hyperbolic tangent activation function, which will help to produce sharper images.

```
def create_generator():
    gen_input = Input(shape=(LATENT_DIM, ))

    x = Dense(128 * 16 * 16)(gen_input)
    x = LeakyReLU()(x)
    x = Reshape((16, 16, 128))(x)

    x = Conv2D(256, 5, padding='same')(x)
    x = LeakyReLU()(x)

    x = Conv2DTranspose(256, 4, strides=2, pad-
ding='same')(x)
    x = LeakyReLU()(x)

    x = Conv2DTranspose(256, 4, strides=2, pad-
ding='same')(x)
    x = LeakyReLU()(x)
```

```
      x = Conv2DTranspose(256, 4, strides=2, pad-
ding='same')(x)
      x = LeakyReLU()(x)

      x = Conv2D(512, 5, padding='same')(x)
      x = LeakyReLU()(x)
      x = Conv2D(512, 5, padding='same')(x)
      x = LeakyReLU()(x)
      x = Conv2D(CHANNELS, 7, activation='tanh', pad-
ding='same')(x)

      generator = Model(gen_input, x)
      return generator
```

The discriminator was also built with 8 layers, but in this case, the activation function was sigmoid, which helps to give as output the typical classification ranges from 0 to 1. A Binary Cross Entropy loss function is also applied, generally used in binary classification problems to measure the difference between the calculated probabilities of two possible classes.

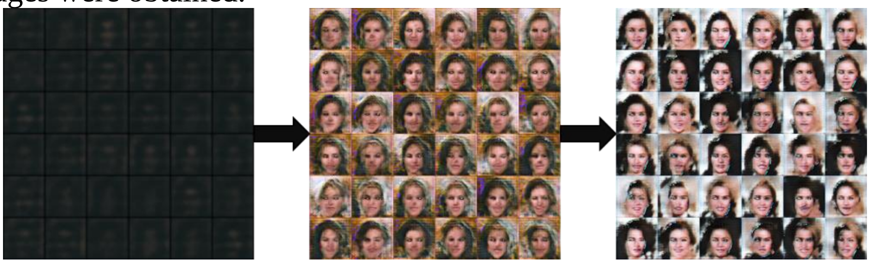
```
  def create_discriminator():
      disc_input = Input(shape=(HEIGHT, WIDTH, CHANNELS))

      x = Conv2D(256, 3)(disc_input)
      x = LeakyReLU()(x)

      x = Conv2D(256, 4, strides=2)(x)
      x = LeakyReLU()(x)

      x = Conv2D(256, 4, strides=2)(x)
      x = LeakyReLU()(x)

      x = Conv2D(256, 4, strides=2)(x)
      x = LeakyReLU()(x)

      x = Conv2D(256, 4, strides=2)(x)
      x = LeakyReLU()(x)

      x = Flatten()(x)
      x = Dropout(0.4)(x)

      x = Dense(1, activation='sigmoid')(x)
      discriminator = Model(disc_input, x)
```

```
    optimizer = RMSprop(
        lr=.0001,
        clipvalue=1.0,
        decay=1e-8
    )

    discriminator.compile(
        optimizer=optimizer,
        loss='binary_crossentropy'
    )

    return discriminator
```

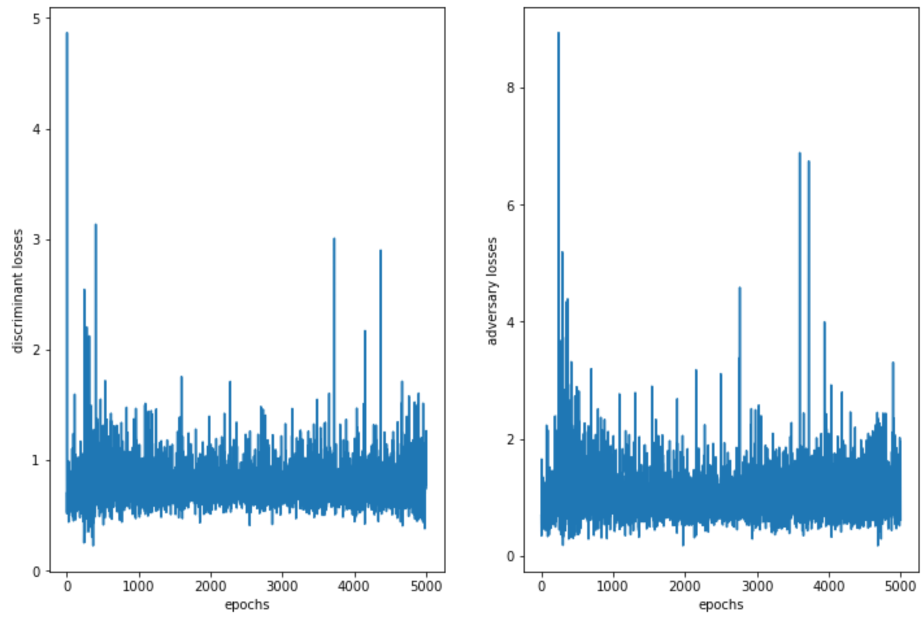
Both models form the GAN to proceed with the training.

```
generator = create_generator()
discriminator = create_discriminator()
discriminator.trainable = False
```

Finally, training is performed. With these parameters the following images were obtained.



During training, the GAN had the following behaviour with respect to its loss functions.



Since the GAN model comprises two neural network models, understanding its behaviour is highly complex. It is irregular, with highs and lows at seemingly random points.

6 Conclusions

The literature review, which describes the most widely used methods over time for artificial image generation, shows that GANs are the most effective model in this task to date, being successfully implemented in a variety of domains. Despite their difficulties, they show promising results due to their recent development. The more they are investigated, the more possibilities open up in the world of AI. One of the main obstacles, which was observed not only in research but also in experiments such as the one presented above, is the high computational power they require. They have high complexity, as mentioned in other works. The higher the computational power, the better the quality and realism of the images generated.

Besides the above, there is still a large open field of research on the functioning of GANs. It is not yet clear why they behave as they do during training. For example, it is unknown why the values are sometimes triggered when measuring their losses. Despite the above, it is only known that the images generated improve as training progresses.

The metrics of evaluating an artificial image can be seen as a major limitation. It is a high level of complexity to assess whether an image looks realistic to the human eye. In fact, many papers resort to many attempts to measure this. Some use quantitative metrics, while others focus more on the human side with qualitative measures.

As future work, it is intended to study GAN architectures that focus on modifying facial attributes, specifically age. This is to analyze the variations in terms of functions, parameters and other features that are added to the architecture to be applied in other areas. As future work, it is intended to study GAN architectures that focus on modifying facial attributes, specifically age. This is to analyze the variations in terms of functions, parameters and other features that are added to the architecture to be applied in other areas.

7 Code repository

A repository which contains the code implemented for the generation of artificial faces is presented below, where is available a Jupyter Notebook shows the results of each compiled code block.

https://github.com/BlancaECS/Face_generation/blob/4acee363c5cdade33b51d7af013c8defab92dd2b/FirstFace_GAN.ipynb

References

[1] R. E. Saragih and Q. H. To, “A Survey of Face Recognition based on Convolutional Neural Network,” *Indones. J. Inf. Syst.*, vol. 4, no. 2, pp. 122–139, 2022, doi: 10.24002/ijis.v4i2.5439.

[2] U. Arora, A. R. Khan, A. Majumdar, D. Ginhac, and R. Jain, *Handbook of Face Recognition*. 2005.

[3] aws, “¿Qué es el reconocimiento facial?,” 2022. https://aws.amazon.com/es/what-is/facial-recognition/.

[4] I. Masi, A. T. Trân, T. Hassner, J. T. Leksut, and G. Medioni, “Do we really need to collect millions of faces for effective face recognition?,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9909 LNCS, pp. 579–596, 2016, doi: 10.1007/978-3-319-46454-1_35.

[5] X. Wang, K. Wang, and S. Lian, “A Survey on Face Data Augmentation,” no. 1, 2019, doi: 10.1007/s00521-020-04748-3.

[6] J. Langr, *GANs in action*. 2019.

[7] G. Perarnau, J. van de Weijer, B. Raducanu, and J. M. Álvarez, “Invertible Conditional GANs for image editing,” no. Figure 1, pp. 1–9, 2016.

[8] S. Sun and A. Wang, “Face Aging With GAN.”

[9] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, “AttGAN: Facial Attribute Editing by only Changing What You Want,” *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5464–5478, 2019, doi: 10.1109/TIP.2019.2916751.

[10] I. Goodfellow *et al.*, “Generative adversarial networks,” *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020, doi: 10.1145/3422622.

[11] J. D. Calderon, L. Fabrico, M. Ticliahuanca, and M. Roncal, “Uso de

algoritmos de machine learning para el diagnóstico de melanomas Trabajo de Investigación para optar el Grado de Bachiller en Ingeniería Mecánico - Eléctrica Jose Dario Calderon Ortiz María Esther Roncal Moscol," no. May, 2021, doi: 10.13140/RG.2.2.20397.00483.

[12] L. Taylor and G. Nitschke, "Improving Deep Learning using Generic Data Augmentation," 2017.

[13] W. H. Hsu, "Investigating Data Augmentation Strategies for Advancing Deep Learning Training," 2018.

[14] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," *34th Int. Conf. Mach. Learn. ICML 2017*, vol. 4, pp. 2941–2949, 2017.

[15] D. Guo and T. Sim, "Digital face makeup by example," *2009 IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2009*, pp. 73–79, 2009, doi: 10.1109/CVPRW.2009.5206833.

[16] T. Li *et al.*, "BeautyGAN," pp. 645–653, 2018, doi: 10.1145/3240508.3240618.

[17] J. Guo, X. Zhu, Z. Lei, and S. Z. Li, "Face synthesis for eyeglass-robust face recognition," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10996 LNCS, no. June, pp. 275–284, 2018, doi: 10.1007/978-3-319-97909-0_30.

[18] W. Shen and R. Liu, "Learning residual images for face attribute manipulation," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 1225–1233, 2017, doi: 10.1109/CVPR.2017.135.

[19] Z. H. Feng, J. Kittler, W. Christmas, P. Huber, and X. J. Wu, "Dynamic attention-controlled cascaded shape regression exploiting training data augmentation and fuzzy-set sample weighting," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 3681–3686, 2017, doi: 10.1109/CVPR.2017.392.

[20] J. J. Lv, X. H. Shao, J. S. Huang, X. D. Zhou, and X. Zhou, "Data augmentation for face recognition," *Neurocomputing*, vol. 230, pp. 184–196, 2017, doi: 10.1016/j.neucom.2016.12.025.

[21] J. Deng, S. Cheng, N. Xue, Y. Zhou, and S. Zafeiriou, "UV-GAN: Adversarial Facial UV Map Completion for Pose-Invariant Face Recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 7093–7102, 2018, doi: 10.1109/CVPR.2018.00741.

[22] W. Xie, L. Shen, M. Yang, and J. Jiang, "Facial expression synthesis with direction field preservation based mesh deformation and lighting fitting based wrinkle mapping," *Multimed. Tools Appl.*, vol. 77, no. 6, pp. 7565–7593, 2018, doi: 10.1007/s11042-017-4661-6.

[23] J. Thies and M. Zollh, "a183-Thies," vol. 34, no. 6, pp. 1–14, 2015.

[24] P. K. Chandaliya and N. Nain, "Child Face Age Progression and Regression using Self-Attention Multi-Scale Patch GAN," *2021 IEEE Int. Jt. Conf. Biometrics, IJCB 2021*, pp. 1–8, 2021, doi: 10.1109/IJCB52358.2021.9484329.

[25] A. Nguyen, J. Yosinski, and J. Clune, "Deep neural networks are easily fooled:

High confidence predictions for unrecognizable images," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 427–436, 2015, doi: 10.1109/CVPR.2015.7298640.

[26] L. R. Calcagni, "Redes Generativas Antagónicas y sus aplicaciones," p. 72, 2020.

[27] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F. Y. Wang, "Generative adversarial networks: Introduction and outlook," *IEEE/CAA J. Autom. Sin.*, vol. 4, no. 4, pp. 588–598, 2017, doi: 10.1109/JAS.2017.7510583.

[28] B. Raj, "Advances in Generative Adversarial Networks (GANs)," 2019. https://medium.com/beyondminds/advances-in-generative-adversarial-networks-7bad57028032.

[29] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, "CovidGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved Covid-19 Detection," *IEEE Access*, vol. 8, pp. 91916–91923, 2020, doi: 10.1109/ACCESS.2020.2994762.

[30] Kaggle, "Celeba dataset." https://www.kaggle.com/datasets/jessicali9530/celeba-dataset.

[31] "Imágenes generadas por GAN." https://drive.google.com/drive/folders/1iaeLCv9xILmc33iqzQXX6hNewY3yG9_5?usp=sharing.

(ii) Review report

DA&CI 2022 - Springer Book (chair)

Help / Log out

Submissions | Reviews | Status | PC | Events | Email | Administration | Premium | Conference | News | EasyChair

Email Instance

To	Rogelio Florencia <rogelio.florencia@uacj.mx>
Time	Jan 26, 20:45 GMT
Subject	DA&CI 2022 - Springer Book notification for paper 5352
Body	Dear Rogelio Florencia, The review of your chapter, "Face image generation: A review," has just been completed. Although our reviewers find the topic pertinent, they believe you should strengthen the coverage before publishing the chapter. I have compiled the feedback from reviewer evaluations for your perusal to emphasize particular changes that I feel would be best for you to make to your chapter. Please study the evaluations carefully and let me know if you have any questions about any comments or suggestions. Once you have completed the revisions, you must upload a PDF file with the following parts: PART 1. A list of your responses to every single one of the reviewers' comments. Also, when applicable, you should indicate where the revised manuscript addresses the review comments by referencing line numbers. PART 2. A revised version of your chapter with line numbering. Here, the revisions should be explicitly marked. Please, provide this revision by no later than FEBRUARY 21 (2023), uploading the document as an update of your previous submission (https://easychair.org/conferences/?conf=daci2022springerbook). Please, be advised that a revision does not guarantee acceptance. The decision regarding the approval of your chapter depends on additional review. Before you upload the revision, you should: (a) Check all requirements and guidelines have been met as outlined in the Manuscript Preparation guide: https://www.springer.com/de/authors-editors/book-authors-editors/resources-guidelines/book-manuscript-guidelines/manuscript-preparation/5636 (see section "Chapters"). (b) Use the Word/LaTeX template provided for book chapters. For your convenience, I have shared a folder with the (LaTeX and Word) templates and a brief description of the reference style (https://drive.google.com/drive/folders/1HJSs5s2O3C1WGRO95aqZfcqoondC0yDW?usp=share_link). (c) Consider an extension of 10,000–16,000 words for the full manuscript. This direction is not mandatory but preferable. Please, contact the editors for documents with a different extension. (d) Ensure proper use of the English language, formal grammatical structure, and correct spelling and punctuation. If necessary, consult a professional (e.g., https://www.proof-reading-service.com/). (e) Reduce the similarity index of your manuscript. As a guidance, I share with you the

following similarity report:
https://drive.google.com/file/d/1sdcTVBWeDnPBHwFkA3_J29OdokahznRM/view?usp=share_link

Thank you for your interest and diligent work in your contribution to "Data Analytics and Computational Intelligence: Novel Models, Algorithms and Applications," I greatly value your manuscript and look forward to seeing your revision! If you have any questions, please do not hesitate to contact me, Gilberto Rivera, at gilberto.rivera@uacj.mx (with a copy to riveragil@gmail.com).

SUBMISSION: 5352
TITLE: Face image generation: A review

----------------------- REVIEW 1 ---------------------
SUBMISSION: 5352
TITLE: Face image generation: A review
AUTHORS: Blanca Elena Cazares, Rogelio Florencia and Vicente García

----------- Overall evaluation -----------
SCORE: 2 (Accept after minor revision)
----- TEXT:
• Give significant contribution or novelty
• Compare proposed methodology which is not properly presented in the manuscript
• Major contributions of the work must be highlighted in a separate paragraph of the introduction section.
• Give future work in this domain in conclusion section.
• Give Face image generation:recent years comparative analysis with some chart.

----------------------- REVIEW 2 ---------------------
SUBMISSION: 5352
TITLE: Face image generation: A review
AUTHORS: Blanca Elena Cazares, Rogelio Florencia and Vicente García

----------- Overall evaluation -----------
SCORE: 1 (Accept after major revision)
----- TEXT:
At introduction chapter, at the end, it needs to restructure the explanation of paper's organization. For instance, it says "this chapter talks about..." and then describes how the paper is structured. It doesn't show coherence.

Also, the abstract need rewrite. It describes a GAN architecture will be implemented, while at introduction just says a GAN architecture will be presented. Implementation is not part of a review, but a novel approach. Please correct this.

3.2 title is written in Spanish

Chapter 4 is very thin. It is hard to believe there are only four related works in this field. It is necessary to expand this section or re-arange the paper. (As an example, a quick search in google scholar finds 39000 papers since 2019)

The proposed architecture is missing, but the code implements it. It is recommended to include the model and its description of the proposed GAN architecture.

The title doesn't make sense. There are two styles in the paper. First, it focus on a review. That is done in acceptable way. However, there is experimentation to implement an architecture and show results. Therefore, The title of this paper needs to be adjusted to reflect the content of the paper.

----------------------- REVIEW 3 ---------------------
SUBMISSION: 5352
TITLE: Face image generation: A review
AUTHORS: Blanca Elena Cazares, Rogelio Florencia and Vicente García

----------- Overall evaluation -----------
SCORE: 1 (Accept after major revision)
----- TEXT:
The document presents a good introduction, which allows to understand the reason for the work carried out. The theme is relevant, and the methodology is appropriate. An appropriate and easy-to-follow literature review is made.

The implementation of a GAN architecture is presented for experimentation purposes; however, a more detailed analysis of the results is needed. If the purpose of the experiment is to demonstrate that the technique and architecture are the most suitable, it is necessary to contrast it with at least another technique or with different parameters. On the other hand, if the purpose of the paper is to highlight that the technique requires high levels of computing power, it would be necessary to present results with different computing capabilities and/or different GAN parameters. For these reasons, it is highly recommended to improve the work with a detailed section on the analysis of the results of the experiment performed.

The following are corrections that should be made in order to have a more refined work:

Abstract
What do you mean saying “after this” in this sentence? “After this, the technique that has presented outstanding results in the face generation task are GANs.” Do you mean that among the different techniques that have been reviewed, GANs are the best evaluated? If so, it’s necessary to improve the sentence.

Introduction
Biometric identification can be done with a variety of elements, not just facial recognition. The following sentence can be misinterpreted if it is stated that intelligent facial recognition systems are also known as biometric identification. “The search is on for such systems to possess this same intelligence, also known as biometric identification.”
Although the term “DA” is defined in the abstract, it is recommended that it be defined the first time it is used in the body of the text as well. The same for “GANs”.

The words “Also in addition” are redundant.
“AD techniques have faced several limitations of their own…” should be DA techniques…

Section 2.2
Figure 3 presents examples of transformation of components, however it is not clear what type of transformation each image or set of images in the figure corresponds to. Consider expanding the image description.

Section 3.2.

The tittle should be “Architecture” instead of “Arquitectura”.

Section 5.
The figures presented in this section should have a figure caption. In addition, it is recommended that the programming code shown be part of a figure or give it a more suitable format in order to make it more presentable.
The choice of parameters should be justified in more detail, rather than simply saying that they were chosen due to computational capabilities or to meet time and form criteria, or otherwise justify what "time and form" refers to.

General comments

The document has some form errors, for example, different font sizes are used.

Some figures do not contain a figure caption, while others are not described in detail.

Responses to reviews

Reviewer 1

Give significant contribution or novelty

This content was added to the end of “introduction” section. (lines 88-93)

> The main contributions of this work are: a) the compilation of recent state-of-the-art works that demonstrate the different areas in which GANs have been used to improve the training of various systems or algorithms, such as ML, b) the description of some architectures and parameter variations that can be modified according to the objectives pursued and, c) an example of a GAN architecture used to generate artificial faces.

Compare proposed methodology which is not properly presented in the manuscript

The manuscript was modified by integrating the “methodology” section, where it is described the research for papers presented in the chapter using a Prisma-based methodology. (lines 451-476)

Major contributions of the work must be highlighted in a separate paragraph of the introduction section.

This was attended to by adding this paragraph in the “introduction” section. (lines 88-93)

Give future work in this domain in conclusion section.

The "conclusion" section was modified by adding the future work to be done after this investigation of the GANs. (lines 611-616)

> As future work, it is planned to analyze, evaluate, and implement a GAN architecture with the objective of simulating the age progression in the facial images of Mexican women reported missing. The idea is to visualize their appearance after a lapse of 10 to 20 years of adulthood. With this, it is intended to use GANs as a DA technique to create robust image sets that allow training artificial intelligence algorithms to help finding missing women.

Give Face image generation:recent years comparative analysis with some chart.

The section "Face image generation evolution with GANs" was added to visualize the fast progress GANs have made in face image generation with a table. (lines 349-355)

3.5 Face image generation evolution with GANs

Progress in facial image generation has gone hand in hand with GANs, so it is now possible to control the resolution or quality of artificial images [26]. Table 1 shows the main advances of the GANs.

Table 1. GAN evolution in artificial facial images generation.

Year	Image	Model description
2014		Proposed GAN architecture composed of two models: generative *G* and discriminative *D*, represented by a multilayer perceptron [26].

Reviewer 2

At introduction chapter, at the end, it needs to restructure the explanation of paper's organization. For instance, it says "this chapter talks about..." and then describes how the paper is structured. It doesn't show coherence.

The paragraph indicates that the chapter is divided into sections. (lines 94-101)

The chapter is structured as follows. Section 2 presents the background to the generation of artificial facial images. Section 3 describes the general architecture of a GAN, the training process and its challenges, among other information of interest. Section 4 presents a brief literature review on data augmentation methods for generating artificial images, focusing mainly on GANs. Section 5 describes the methodology used to conduct the literature review. Section 6 shows the implementation of a GAN architecture to generate artificial facial images. Lastly, Section 7 presents the conclusions and future work.

Also, the abstract need rewrite. It describes a GAN architecture will be implemented, while at introduction just says a GAN architecture will be presented. Implementation is not part of a review, but a novel approach. Please correct this.

The paragraph was corrected to make it clear that the GAN architecture would be implemented. (lines 99-100)

methodology used to conduct the literature review. Section 6 shows the implementation of a GAN architecture to generate artificial facial images. Lastly, Section 7

3.2 title is written in Spanish

The title was corrected to “3.2 Architecture” (line 297)

3.2 Architecture

In the GAN architecture, the generator will start training simultaneously as the discriminator does. The latter will need a few epochs before starting the adversarial training since it must be able to classify the images correctly.

Chapter 4 is very thin. It is hard to believe there are only four related works in this field. It is necessary to expand this section or rearange the paper. (As an example, a quick search in google scholar finds 39000 papers since 2019)

The research was further expanded and some works in different areas was added to the "related work" section. (lines 394-450)

The proposed architecture is missing, but the code implements it. It is recommended to include the model and its description of the proposed GAN architecture.

The GAN model architecture was included with its respective description of the elements. (lines 487-506) dual-core, 100 GB storage and 8 GB RAM. The architecture is shown in Fig. 7.

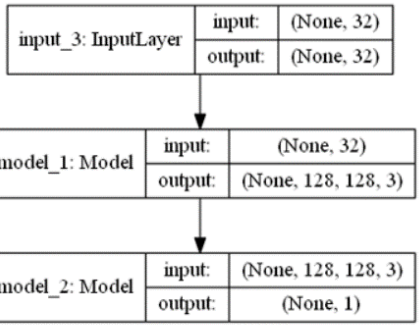


Fig 7. Architecture of a DCGAN model for generating facial images where *model_1* represents the generator and *model_2* the discriminator.

The input layer (input_3:InputLayer) represents the latent vector defined as 32, the generator (model_1:Model) receives this as input and outputs an image of 128*128 size with three color channels (RGB). The discriminator (model_2:Model) receives this image as input and outputs a binary classification to identify the image as *true* or *false*.

Since training this model is computationally expensive, some tests were performed to evaluate the time required with different parameters, as shown in Table 2. The first column indicates the number of images used, the second the number of iterations and the third the batch size, i.e., the number of examples introduced into the network for each training. Also, in the line below the parameters are the images resulting from different parts of the training: first, middle and last.

The tests made obtaining better results possible but required considerable time investment. Additionally, tests allowed the selection of the model parameters, bal-

The title doesn't make sense. There are two styles in the paper. First, it focus on a review. That is done in acceptable way. However, there is experimentation to implement an architecture and show results. Therefore, The title of this paper needs to be adjusted to reflect the content of the paper.

The title was modified to reflect the chapter's content, and the intention of experimentation was stated in the contributions. (lines 99-100)

The main contributions of this work are: a) the compilation of recent state-of-the-art works that demonstrate the different areas in which GANs have been used to improve the training of various systems or algorithms, such as ML, b) the description of some architectures and parameter variations that can be modified according to the objectives pursued and, c) an example of a GAN architecture used to generate artificial faces.

Reviewer 3

Abstract

What do you mean saying "after this" in this sentence? "After this, *the technique that has presented outstanding results in the face generation task are GANs." Do you mean that among the different techniques that have been reviewed, GANs are the best evaluated? If so, it's necessary to improve the sentence.*

> The sentence was restructured to better explain the intended idea. (lines 10-12)
>
> ations to real image sets. This chapter presents a brief literature review on various DA methods dedicated to image generation. The technique that has presented outstanding results in the task of generating artificial images is Generative Adversarial Networks (GANs). Some recent research papers in which GANs have been used for

Introduction

Biometric identification can be done with a variety of elements, not just facial recognition. The following sentence can be misinterpreted if it is stated that intelligent facial recognition systems are also known as biometric identification. "The search is on for such systems to possess this same intelligence, also known as biometric identification."

> The phrase was modified so as not to be misinterpreted. (lines 23-25)
>
> recognize others through the face, being a unique part of the body. One of the goals is that these systems possess this same intelligence as part of the biometric identification measures.

Although the term "DA" is defined in the abstract, it is recommended that it be defined the first time it is used in the body of the text as well. The same for "GANs".

> The acronyms have already been defined in the first mention in the document's body. (Lines 44 and 57)

The words "Also in addition" are redundant.

> The sentence was corrected. (Lines 48-49)

"AD techniques have faced several limitations of their own…" should be DA techniques…

The sentence was corrected. (Lines 48-49)

image requirements for learning systems. Also, DA techniques have faced several limitations of their own [5], for example:

Section 2.2
Figure 3 presents examples of transformation of components, however it is not clear what type of transformation each image or set of images in the figure corresponds to. Consider expanding the image description.

The figures were modified by adding their respective descriptions to each set of images. (Lines 176-177)

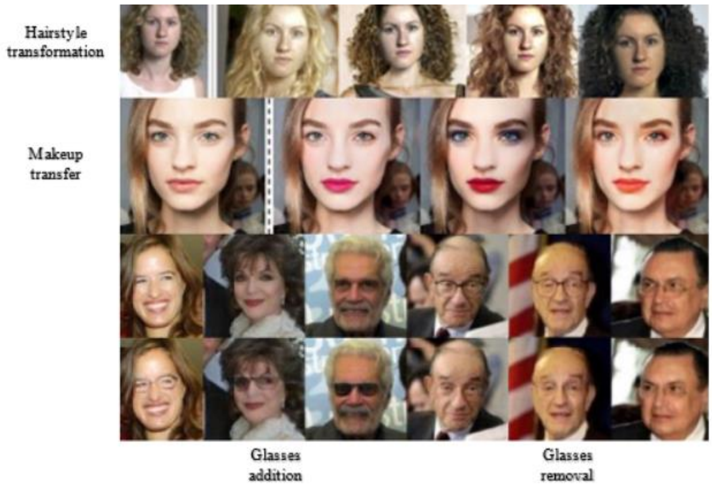


Fig 3. Component transformation examples [5]

Section 3.2.
The tittle should be "Architecture" instead of "Arquitectura"

The title was corrected to "3.2 Architecture" (line 297)

3.2 Architecture

In the GAN architecture, the generator will start training simultaneously as the discriminator does. The latter will need a few epochs before starting the adversarial training since it must be able to classify the images correctly.

Section 5.

The figures presented in this section should have a figure caption. In addition, it is recommended that the programming code shown be part of a figure or give it a more suitable format in order to make it more presentable.

Caption was added to the figures mentioned. Also, the code was rewritten to pseudocode using the style format provided in the template. (Figures caption lines 540-541, 581, 586) (Code line 531)

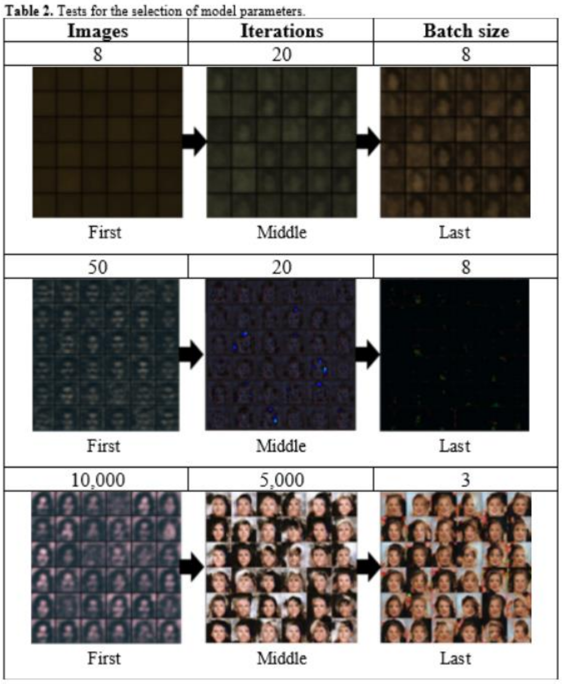
The choice of parameters should be justified in more detail, rather than simply saying that they were chosen due to computational capabilities or to meet time and form criteria, or otherwise justify what "time and form" refers to.

The parameter selection justification was rewritten, showing earlier tests to demonstrate why the article submission time was vital in parameter selection. (Lines 498-506, 519)

Since training this model is computationally expensive, some tests were performed to evaluate the time required with different parameters, as shown in Table 2. The first column indicates the number of images used, the second the number of iterations and the third the batch size, i.e., the number of examples introduced into the network for each training. Also, in the line below the parameters are the images resulting from different parts of the training: first, middle and last.

The tests made obtaining better results possible but required considerable time investment. Additionally, tests allowed the selection of the model parameters, balancing the time and quality of the images generated to present the example.

Table 2. Tests for the selection of model parameters.



Images	Iterations	Batch size
8	20	8
First	Middle	Last
50	20	8
First	Middle	Last
10,000	5,000	3
First	Middle	Last

General comments

The document has some form errors, for example, different font sizes are used.

> The provided template is now being used to follow the appropriate formats.

Some figures do not contain a figure caption, while others are not described in detail.

> Caption was added to the corresponding figures.

Data augmentation techniques for facial image generation: a brief literature review

Abstract Image processing has gained notoriety over the last few years in performing various tasks through deep learning (DL) algorithms, such as face recognition and identity verification. Unfortunately, most of them require a large set of images for training, usually manually labelled, which is a costly task both in time and effort, not to mention being prone to human error. Data Augmentation (DA) techniques have been used to mitigate this situation, as they generate images by applying variations to real image sets. This chapter presents a brief literature review on various DA methods dedicated to image generation. The technique that has presented outstanding results in the task of generating artificial images is Generative Adversarial Networks (GANs). Some recent research papers in which GANs have been used for the generation of artificial images are presented. General aspects of GANs, such as their definition, architecture, training, and challenges, are described. Additionally, the implementation of a GAN architecture for the generation of artificial face images from a public set of images is presented. The need for a great computational capacity to generate images with great sharpness and realism is highlighted.

Keywords Face image generation, Data augmentation, Generative Adversarial Networks.

1. Introduction

Facial recognition systems have been discussed for several decades. The scientific community has shown great interest in this subject given the human ability to recognize others through the face, being a unique part of the body. One of the goals is that these systems possess this same intelligence as part of the biometric identification measures.

Facial recognition has been developed in the field known as computer vision [1]. This area supports a variety of critical applications, e.g., identity verification. This task is done using what is known as a face analyzer, which is software that confirms the identity of people based on their faces [2]. It is achieved through the identification and measurement of facial features in images. It can also be used to associate human faces in the latest or even in videos to check how similar they are to one or several specific individuals. In addition, it can determine the level of similarity between two photographs to find out if they belong to the same person or search for them among a large set of images in a collection. For example, biometric security systems use facial recognition to uniquely identify individuals as users at login to strengthen user authentication. In addition, mobile devices often use this type of technology to protect the data they contain [3]. Unfortunately, large amounts of facial images are required for these processes to function properly, so much of the research has focused on artificial image generation. It aims to increase the image sets used to train different architectures because of their significant impact on the results [4].

Collecting and labelling data samples with good quality is costly, both in time and effort, and is prone to human error. For this purpose, various techniques known as Data Augmentation (DA) have been implemented, and their performance is suitable in different domains.

DA techniques allow the size of existing image sets to be increased considerably through simulations [4], which helps significantly with the image requirements for learning systems. Also, DA techniques have faced several limitations of their own [5], for example:

- The images generated lack realistic variations such as makeup, skin color, and background change, which means that they would have a different distribution than the real ones.
- Creating high-quality facial images is very difficult due to the complexity of facial details.

After many years of research and application of different modelling methods, in 2014, a technique was proposed that allows the generation of realistic images using two artificial neural networks, Generative Adversarial Network (GAN). Although these are not the first method used for artificial data generation, their results and versatility distinguished them from the rest since they have achieved outstanding results that were still considered impossible for artificial systems [6].

GANs are a machine learning (ML) technique that integrates two simultaneously trained models, the *generator* and the *discriminator*. The generator is trained to create the artificial images and the discriminator is to discern between the fake images created by the generator and the real ones from the original training set. Its performance has far exceeded expectations in the field of artificial image generation, according to a variety of authors [7][8][9][5].

Like ML algorithms, the data generated by a GAN completely depends on the training set provided to perform the learning. For example, if a GAN is required to learn how to create images of handwritten numbers, it is necessary to use a training set containing several images of handwritten numbers [10].

GANs are derived from a gaming perspective. Hence the word adversarial denotes a competitive dynamic between the two models that compose them. The generator aims to generate real images indistinguishable from the training set. On the other hand, the discriminator aims to distinguish these generated images from the real ones. Therefore, the better the generator generates realistic images, the better the discriminator must be to distinguish the real ones from the false ones [6].

The fast growth and progress of GANs have been due to research and development, generating new architectures to stabilize the outputs and generate images of higher quality and realism. In this way, the aim is to create impossible images for the human eye to differentiate.

This chapter presents a brief literature review on data augmentation methods for artificial image generation, focusing mainly on GANs, which have been used throughout recent years to create artificial facial images. Additionally, the general architecture of a GAN and related concepts are described. Lastly, the implementation of a GAN architecture for generating artificial face images, trained with a small part of the public image set “CelebA”, is described.

The main contributions of this work are: a) the compilation of recent state-of-the-art works that demonstrate the different areas in which GANs have been used to improve the training of various systems or algorithms, such as ML, b) the description of some architectures and parameter variations that can be modified according to the objectives pursued and, c) an example of a GAN architecture used to generate artificial faces.

The chapter is structured as follows. Section 2 presents the background to the generation of artificial facial images. Section 3 describes the general architecture of a GAN, the training process and its challenges, among other information of interest. Section 4 presents a brief literature review on data augmentation methods for generating artificial images, focusing mainly on GANs. Section 5 describes the methodology used to conduct the literature review. Section 6 shows the implementation of a GAN architecture to generate artificial facial images. Lastly, Section 7 presents the conclusions and future work.

2. Generation of artificial facial images

Generating new images from others is a widely researched task in computer vision. In recent years, the development of Artificial Intelligence (AI) techniques has motivated the idea of producing images of high quality and realism. It has enabled the creation of realistic human faces that are difficult to distinguish between fake and real, even for the human eye. This task has been evolving exponentially since implementing GANs [11]. However, before GANs, other computational methods were used to generate variations in facial images to train systems such as facial recognition systems.

Four main categories can be broadly discussed in the DA task for generating artificial images: generic, component, attribute, and age transformations. Each one is described in the following sections.

2.1 Generic transformations

These transformations focus on modifying the entire image, ignoring high-level components such as composition, light, volume, symmetry, shape and texture. They are usually divided into two main groups: geometric and photometric transformations.

Geometric transformations are commonly applied in multiple computer vision tasks [5], such as face recognition, healthcare and manufacturing applications. In general terms, geometric transformations alter the pixels of an image by placing them in new positions. Some examples of these can be:

- Reflection: consists of flipping the image around its vertical or horizontal axis.
- Rotation: rotates the image θ degrees around its center, bringing each pixel (x, y) to its position (x', y').
- Cropping: consists of cutting the images to a specific size.

For example, in the case of convolutional neural networks (CNNs), this type of transformations help to minimize their sensitivity to changes in position and orientation [12]. Some examples can be seen in Fig 1.

Fig 1. Geometric transformation examples [5]

On the other hand, photometric transformations, shown in Fig 2, generally work by altering the RGB (red, green, and blue) channels, shifting each pixel value (r, g, b) to new values (r', g', b') of an image according to predefined rules. These transformations adjust the illumination and color, leaving the geometry unaffected [12]. For example, it can be mentioned color manipulations, such as inverting them or adding some filters, such as blurring or grayscale [5].

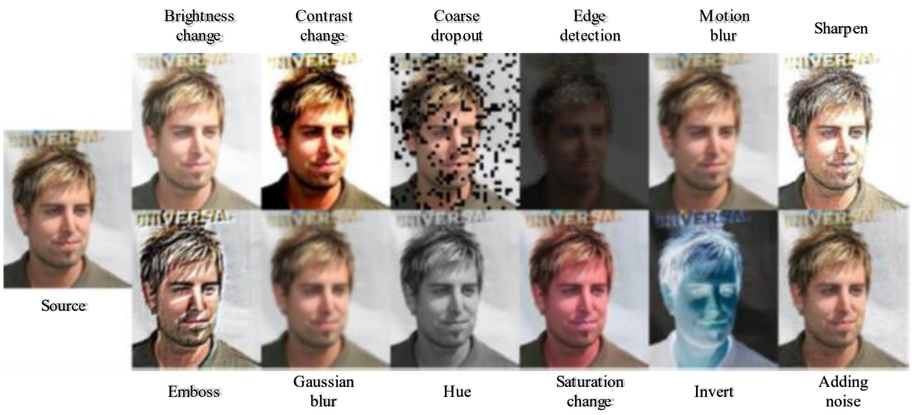


Fig 2. Photometric transformation examples [5]

These transformations are mainly used in computer vision tasks to enrich the training sets and prevent a common problem in this field, overfitting. In [13] is a paper dedicated to evaluating the performance of CNNs, trained with images enriched using geometric and photometric transformations. In [14] different transformations were employed to increase the image set and prevent overfitting.

2.2 Component transformation

There are transformations dedicated to enriching the sets of facial images by modifying the components of the person. These images are used to train the algorithm so that it is able to recognize the person, even if their appearance is altered.

The hairstyle can be considered one of the components to be generated since, although it is not considered a facial component, hairstyle affects face detection and recognition because it tends to hide certain features of a person's appearance. Therefore, DA techniques focus on generating facial images with different variations in hair, for example, color, shape, and bangs. In [15], a method using DiscoGAN was proposed. This variant learns to discover relationships between different domains and develops the ability to translate features between them, for example, by transforming hair color.

Makeup transfer and accessory addition techniques can also be identified due to the difficulty for recognition systems to effectively perform their tasks when some features of the face look different depending on the makeup or accessories a person wears. Most studies based on this type of transformation can be divided into two categories: traditional image processing [16] and those based on DL [17]. Some examples of these can be seen in Fig 3.

Another component that impacts facial recognition is the use and removal of accessories, including glasses, earrings, and piercings, among others. Of all these, glasses are the most commonly used, as they are used for different reasons, for example, vision correction, prevention against sunlight, eye protection, and aesthetics, among others. They significantly affect the accuracy of facial recognition, as they usually cover a large area of the face.

In [18], a fusion of virtual lenses onto faces was performed using the Augmented Reality (AR) technique. In [19], a method of facial attribute manipulation based on image residuals was proposed, defining this as the difference between the input image and the desired output image.

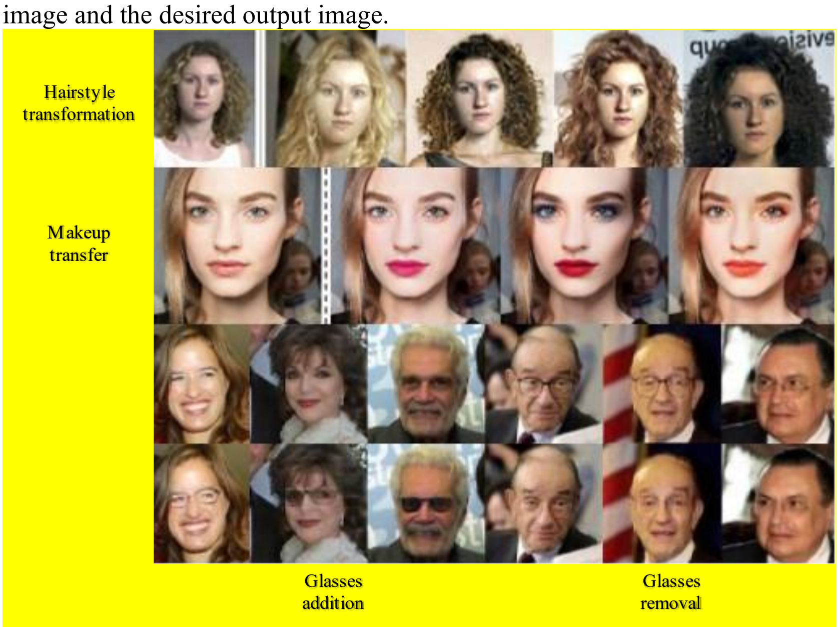


Fig 3. Component transformation examples [5]

2.3 Attribute transformation

There are also some transformations dedicated to modifying aspects such as pose. In this case, the position of the head in a photograph is considered a significant challenge in facial recognition tasks since any variation in it tends to modify visual aspects of the face, i.e., it can hide or show different facial details. In addition, it has been considered an essential aspect since facial photographs in several legal processes are requested from the frontal side. This leads to the research of recreating how a face may look from other angles.

Facial expressions are also considered critical: happy, annoyed, scared, and surprised, among others. These techniques help to improve the performance of emotion classification and recognition systems. Primarily 2D, 3D and learning-based modelling approaches are used to achieve this goal and usually focus on modifying the expression of a face using expression templates by concentrating on a series of points, for example, the corners of the mouth, the cheeks position, and the location of the eyebrows. Others focus on simulations to recreate face parts that are hidden by the pose in which the photograph was taken [20][4][21][22]. Some examples of this can be found in Fig 4.

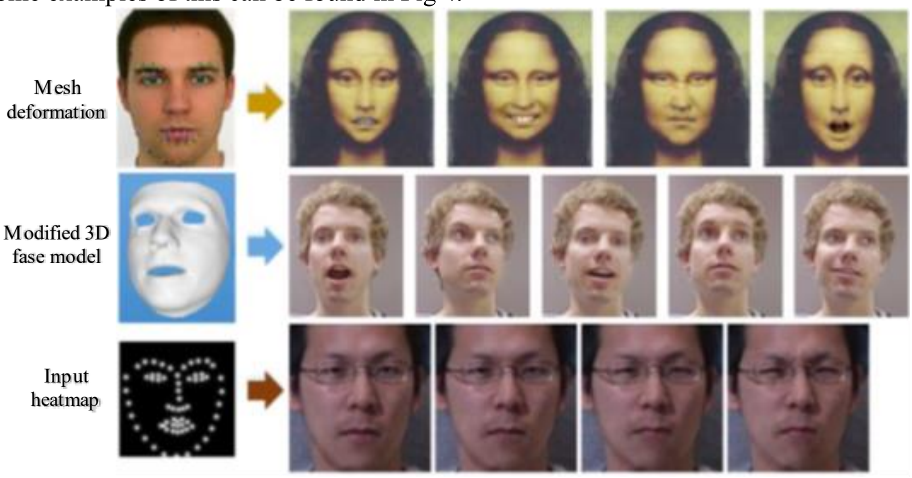


Fig 4. Facial expression synthesis [23][24]

2.4 Age progression and regression

Finally, age regression or progression seeks to predict a current face's appearance in the past or future, respectively, while preserving its facial features. It has become a widely explored research topic because it significantly affects various applications, including missing person tracking, facial recognition, and aesthetic studies, among others. The two main concerns in attempting age regression or progression are identity preservation and prediction accuracy [25].

These facial images enrich the image sets by adding different features at different time stages of the same face, which makes the models more robust to age variations.

Age modification methods are mainly divided in two: prototype-based and model-based. The first one creates an average face for different age ranges, learns the shapes and textures of these and applies these features to transfer them to a new face. However, the individual's characteristics are often lost with these methods [5].

Model-based methods build, as the name implies, models of biological changes due to age in faces, e.g., musculature, wrinkles, and skin texture, among others. They are called generative because they are such a powerful tool for creating new data by learning to imitate the probabilistic distribution of a training set. Nowadays, generative models have gained importance and attention due to their good performance in data creation. Among the most popular are autoregressive models, Variational Autoencoders (VAEs) and GANs. The disadvantage of these models is their complexity and computational cost [5].

GANs, as mentioned above, are an alternative architecture for training generative models since they handle probabilistic computations very well. Recent work has begun to apply them to the age regression and progression task, and many variants of the model have been generated. Some examples of this generative task are shown in Fig 5.

Fig 5. Facial age regression and progression examples [5]

Before GAN, two approaches to age progression and regression in faces were prototype-based and modelling-based. In the latter, critical points in the image, such as eyes, nose, and jaw, track, in turn, the temporal changes such as wrinkles, musculature and color in these are identified. However, this method requires a large amount of age-labelled data over a long period for each individual, which is difficult to find and computationally expensive. The prototype-based method creates an average face based on a set of images of a particular age group, using it to transfer those features from one age range to another. The disadvantage of this is that personal features are often lost. Another possibility is to use neural networks to transform faces across ages. It generates smoother images but still requires images labelled with the person's age through the years [8].

On the other hand, the GAN consists of a discriminator and a generator competing with each other based on the *min-max* games. The generator starts by receiving as input a noise vector *z* and creates an image that it gives to the discriminator to receive feedback from it. Some variants of this architecture are [8]:

- DCGAN (Deep Convolutional GAN) has demonstrated that GAN can be successfully applied to generate indoor scenarios and human faces.
- StyleGAN significantly extended the basic GAN to progressively generate high-resolution images from those with very low resolutions.

- cGAN (Conditional Generative Adversarial Networks) introduces an identity preservation vector with the optimization approach when generating faces, so there is a better match between the original and the created face.
- Pyramid GAN simulates the effects of age more sharply and presents a suite of methods for assessing accuracy and fidelity to the original image.

3. Generative Adversarial Networks (GANs)

GANs are powerful AI-based unsupervised learning algorithms that aim to learn the estimated probability distribution in a specific training set. GANs were proposed by Ian Goodfellow in 2014 [10]. They are based on a competition system between two neural network models that try to maximize their performance while minimizing that of their adversary, developing the ability to analyze, capture and copy the variations presented by a particular set of images [8].

These networks can generate artificial data, being one of the generative models with the highest quality of results, especially when its potential to generate high-resolution images is analyzed [6]. Section 3.1 gives a definition of what a GAN is. Section 3.2 describes the general architecture with which GANs were originally proposed. Section 3.3 describes the GAN training process. Section 3.4 mentions the challenges researchers face when training a GAN. Finally, Section 3.5 presents the face image generation evolution with GANs.

3.1 Definition

It has been demonstrated that most neural networks can be easily tricked into misclassifying by adding only a small amount of noise to the original data. Surprisingly, after this addition, the model develops a higher confidence level in the wrong predictions than in the correct ones. It is because most ML algorithms learn from a limited amount of data, which is prone to model overfitting [32]. It motivated the creation of GANs. They can be described in three parts:

- Generative: They are considered generative models since they describe how new data are generated in terms of probabilistic models.
- Adversarial: The model is trained by competition among its neural networks, i.e., they are considered adversarial to each other.
- Networks: They use neural networks as the primary training algorithms.

They are based on game theory, which considers players to be both ML models, typically implemented using neural networks. A network is called a generator, which can learn the distribution obtained from an original data set to try to replicate it. It is achieved by inserting a noise vector z, i.e., random numbers with a

Gaussian distribution. The main objective of the generator is to learn how to transform unstructured noise z into realistic samples [10].

The other player is called the discriminator. It examines each example x received as input and outputs an estimate of whether it is true or false.

Each player has a cost, so they try to minimize their own, i.e., the discriminator's cost encourages it to correctly classify the data as real or fake. In contrast, the cost of the generator encourages it to generate data that the discriminator incorrectly classifies as real.

A typical example when talking about GANs is to imagine that one network represents a money counterfeiter and the other a policeman. The former generates counterfeit bills while the policeman tries to arrest him for benefiting the production of legitimate ones. The competition leads to the production of more and more realistic counterfeit bills until, eventually, the counterfeiter produces them so realistic that the policeman cannot tell the difference between authentic and counterfeit [6].

3.2 Architecture

In the GAN architecture, the generator will start training simultaneously as the discriminator does. The latter will need a few epochs before starting the adversarial training since it must be able to classify the images correctly.

The architecture consists of two competing deep neural networks: the *generator* and the *discriminator*. The generator produces new data instances, while the discriminator tries to distinguish accurately between the real data, i.e., those coming from the original set, or fake data produced by the generator.

This competition will continue until the generator can create realistic artificial data, which can then be used as input to other neural networks [10]. Fig 6 shows a GAN schematic with its basic components.

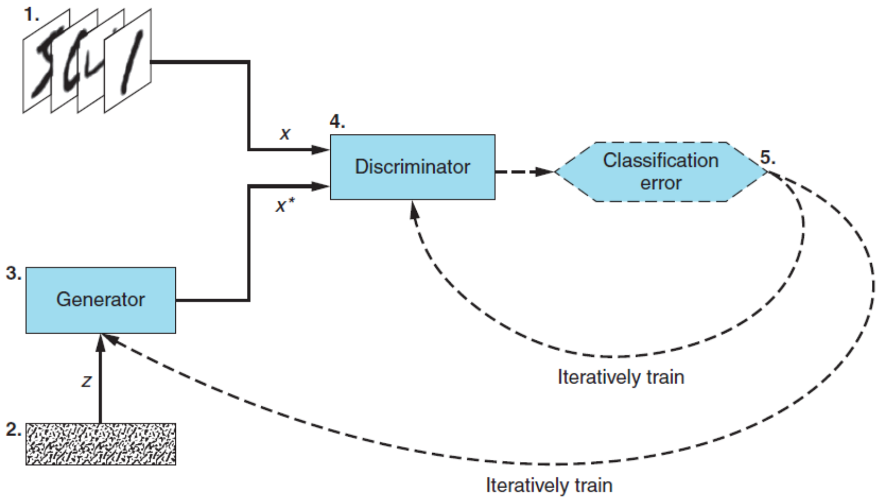


Fig 6. The two GAN subnetworks with their inputs, outputs and interactions [6].

Since its development, many different architectures have been proposed to deal with a wide variety of domains. It could even be said that several scientific papers are published weekly [33]. In [34] an extensive literature review is done on multiple architectures developed for GANs.

3.3 Training process

The training phase requires the two networks and a set of data from which the artificial data will be generated. First of all, it is worth mentioning that the training of the generator is much more complex than the discriminator, which can be seen more as a binary classifier.

The discriminator receives real inputs that must be labelled as such and inputs from the generator which must be labelled false, 1 if the input data is real and 0 if it is false. On the other hand, the generator must be trained with the only condition that the data created mislead the discriminator, i.e., it must minimize its loss and maximize that of the opponent. To achieve this, the generator's output must be the input of the discriminator so that the output of the whole model gives as output the probability that the data are real, according to the discriminator. In this way, the generator obtains feedback from the discriminator, which it uses to create progressively more similar data in the training set [10].

3.4 Challenges

The GANs are still facing research challenges regarding their training. Among them, the following can be pointed out [35]:

- Non-convergence: this is when the generator and the discriminator fail to reach the desired equilibrium (50%). Their respective loss functions begin to fluctuate without being able to reach stability.
- Modal collapse: occurs when the generator produces similar data, even though the inputs vary in characteristics. It finds a small set of samples that successfully deceive the discriminator and are thus incapable of producing others. In these cases, the gradient of the loss function is stuck at a value close to 0.
- Non-informative loss: the general intuition is that the lower the loss of the generator, the higher the quality of the data it produces. However, the loss must be compared with the discriminator's, which constantly improves. Therefore, the issue of model evaluation is more complex. The generator may produce better-quality samples even as the loss function increases.

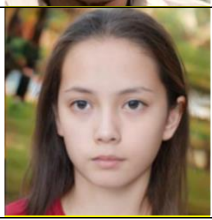
Due to these challenges, techniques have been developed with the aim of minimizing them as much as possible. However, they are still the subject of research at the moment [10].

3.5 Face image generation evolution with GANs

Progress in facial image generation has gone hand in hand with GANs, so it is now possible to control the resolution or quality of artificial images [26]. Table 1 shows the main advances of the GANs.

Table 1. GAN evolution in artificial facial images generation.

Year	Image	Model description
2014		Proposed GAN architecture composed of two models: generative *G* and discriminative *D*, represented by a multilayer perceptron [26].

2015		Proposed DCGAN architecture for unsupervised learning composed of two models: *G* and *D*, represented by CNNs [27].
2016		Coupled Generative Adversarial Network (CoGAN) architecture proposed, which consists of two models for *G* and two for *D* [28].
2017		Proposed new training for GANs by progressively increasing both *G* and *D*, starting with low resolution and adding new layers. It is argued that it improves the training speed and stabilizes it [29].
2018		Proposed alternative architecture for *G* in GANs, called StyleGAN This leads to automatically learned and unsupervised high-level attribute separation [30].
2019		Improvements in the StyleGAN architecture and its training method are proposed. In addition, further work is being done on the quality of the generated images [31].

4. Related work

The following research works have focused on implementing some GAN variants with different objectives.

In [9], it is mentioned that the main objective is to generate a new face by editing facial attributes in the images, preserving its identity. Subsequently, it proceeds to perform a literature review, finding that GAN architecture with an encoder is usually incorporated for such tasks with promising results. Thus, it is proposed to apply an attribute classification constraint to the generated images, in addition to the previous architecture, forming together what is called AttGAN. Experiments were performed on the public set of images "Celeba" to manipulate attributes such as hair color, beard, and age, achieving realism and preservation of facial details.

In [8], a CNN was developed to generate images considering facial age progression. To do this, the authors conducted an exhaustive review of the literature. They analyzed several sets of images to select the one with the greatest variety and fit the domain they wanted to develop the model. Then, they configured parameters and loss functions for network stabilization and the creation of realistic photographs. They used qualitative and quantitative metrics for the evaluation, resulting in images with appropriate characteristics for age progression.

The makeup transfer is the object of study to transfer a specific style to a clean face, preserving the identity of the same. This type of problem at the instance level is considered a great challenge since the styles vary greatly, for example, eye shadows, lipsticks, and foundations, among many others. In [17], BeautyGAN is proposed so that the networks can perform the transfer at the instance level through unsupervised adversarial learning. At the end of the study, a new set of high-resolution makeup images was constructed.

GANs even have an impact in the medical area, as can be seen in [36]. A study on coronavirus (COVID-19), a viral disease caused by the SARS-CoV-2 respiratory syndrome, begins, encouraged by its global effect on health and economics. Chest X-rays from infected patients were a crucial step in controlling this virus. Thus began the introduction of various DL systems and studies that demonstrated the efficiency of using chest X-rays for patient detection. In this context, since CNNs require a significant amount of training data to perform adequately and the virus was too recent to have enough Chest X-rays to generate systems that learn to detect it, the authors present a method to generate images from Chest X-rays in short times. It was achieved by introducing a model named CovidGAN, proving that the artificial images produced by it helped improve the performance of the trained CNNs for COVID-19 detection, increasing their accuracy to 95%.

In [37], it focuses on another important application of GAN: the generation of facial images from text. It has multiple applications in public safety and forensic analysis, such as finding criminals or suspects described by eyewitnesses. This chapter proposes synthesizing facial images from text using a fully-trained generative adversarial network (FTGAN), trained with a text encoder and an image decoder, to generate good-quality images from the input sentences. Multiple experiments were performed on the CUB public set, providing good results regarding the main objective. It was measured by comparison against methods found in the literature, using the Frechet Inception Distance (FID) and Face Semantic Distance (FSD) metrics. Additionally, human ratings were used to validate the generated images.

Another application in the medical field was discussed in [38], where a method was developed to synthesize photorealistic microscope images of red blood cells using cGAN. These images were combined with real ones to augment reduced datasets. Using cGANs as a synthetic data generator for the DA task benefits the development of a model. However, they also raise the fact that all that is involved in developing a GAN model: unstable and time-consuming training, heavy computational requirements and other challenges may not be worth the small margin of improvement it may provide in some tasks.

A literature review is presented in [39], mainly focused on generating medical images using GAN frameworks since it is stated that these have proven to be useful in many cases of image augmentation, medical image generation, and image reconstruction, among others. These characteristics were the ones that encouraged the research of GANs in the medical field to improve image analysis. It is mentioned that several GAN frameworks have gained popularity in medical image interpretation, such as DCGAN, Laplacian GAN (LAPGAN), pix2pix, Cycle-GAN and unsupervised image-to-image translation model (UNIT). Although despite all these approaches, it was concluded that they are still in very early stages, and in-depth research in the area is still needed to reach the most reliable level of progress for GAN applications in clinical imaging.

Among a wide range of applications where GANs have been implemented, one can be found in [40]. In this study, Pix2PixHD was applied to perform image-to-image translation of solar images to give scientists access to complete sets of images for analysis. The end of the training and testing proved the ability of such algorithms to generate high-resolution images. In addition, the images obtained through this model proved to be better than those obtained in previous works also intended to generate these images. It opens the possibility of using these models to create images when unavailable and thus have a better understanding of space weather. It also allows researchers to have the ability to predict solar events, for example, Solar Flares or Coronal Mass Ejections.

In [41], research is found to be motivated by the good performance of CNNs in traffic sign detection (TSD) and recognition (TSR), which are critical tasks in the development of autonomous driving systems. It is difficult because training neural networks requires a large amount of labelled data, and the visual design of traffic signs varies greatly, especially when dealing with different countries. This work focused on Taiwanese traffic signs, motivated by not finding a database, images or study that focused on identifying them. It also pursued the generation of synthetic images when it was impossible to collect reference images through research and development. DCGANs and Wasserstein generative adversarial networks (Wasserstein GANs, WGANs) were used. Each demonstrated exceptional outputs when creating synthetic graphics. The evaluation of the efficiency of DCGAN and WGAN is also included by measuring the quality of the generated images using Structural Similarity Index (SSIM) and the Mean Square Error (MSE). It was concluded that synthetic traffic signal images could be generated with a minimal training image set. Although the resulting images' quality was different obtained with a large training set, they can still be used to solve difficulties when getting genuine photographs is complicated.

5. Methodology

The search and selection of the articles presented in this chapter was based on the Prisma (Preferred Reporting Items for Systematic Reviews and Meta-analyses) methodology, which includes a guide to identify, select, evaluate, and synthesize studies [32]. Prisma was designed for systematic reviews of health-related studies, but it can also be applied to social and educational interventions. In this way, researchers interested in reviewing the scientific literature can recreate the author search process.

The Google Scholar search engine was used to perform the literature review, which indexes full texts or metadata of the academic literature from a wide range of publication formats. The search was started considering the terms "Generative Adversarial Networks" to have a first vision of the research works that have been carried out, available in public databases. Subsequently, the search results were filtered based on the year of publication, "Generative Adversarial Networks" 2019, to identify recently published papers. In addition, a logical operator was added to further narrow the search “Generative Adversarial Networks” AND “image generation”. This allowed filtering the papers found to identify only those focused on image generation.

Since GANs are used for text, images, videos, voice, and statistics, among others [6], a logical operator was added to further narrow the search to “Generative Adversarial Networks” AND “image generation”. This allowed filtering the articles found to identify those focused on image generation. The search returned approximately 16,700 results (February 3, 2023).Of these, we selected and analyzed those that we considered to be the most like this work. The search returned approximately 16,700 results (February 3, 2023). Of these, we selected and analyzed those that we consider to be the most similar to the purposes pursued by this work.

6. Face image generation with GANs

In this section, the architecture of a DCGAN for generating facial images is presented to exemplify the use of GANs. The model was trained using the public image set "Celeba" [42], comprising approximately 200,000 celebrity facial images. It was selected since it contains many variations such as hair color, smile, glasses, poses, backgrounds, and diversity of people that make it very suitable for training face detection models. In addition, its elements are labelled for efficient use in computer vision. A virtual machine generated and configured by the Laboratorio Nacional de Tecnologías de Información (LaNTI)[1] at Universidad Autónoma de

[1] Laboratorio Nacional de Tecnologías de Información: http://www.lanti.org.mx/lanti/

Ciudad Juárez (UACJ) was used. The system specifications are Ubuntu 20.04, dual-core, 100 GB storage and 8 GB RAM. The architecture is shown in Fig. 7.

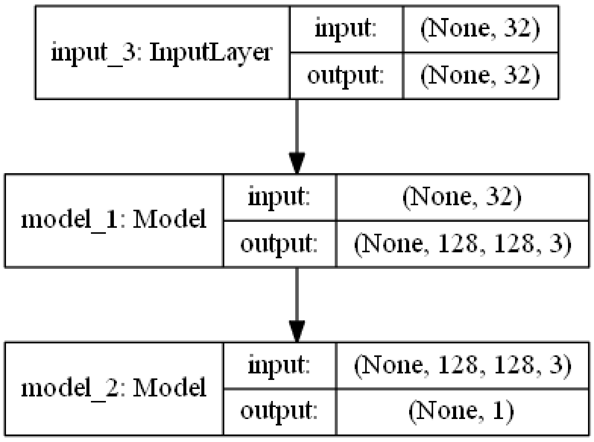


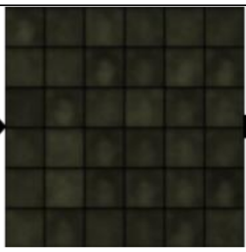
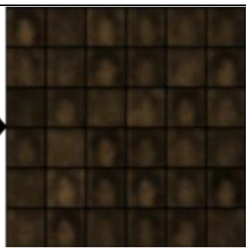
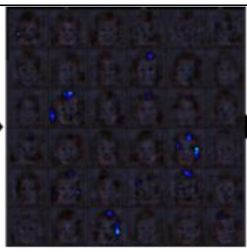
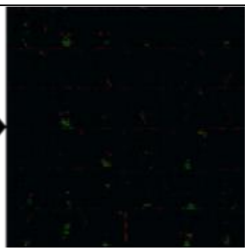
Fig 7. Architecture of a DCGAN model for generating facial images where *model_1* represents the generator and *model_2* the discriminator.

The input layer (input_3:InputLayer) represents the latent vector defined as 32, the generator (model_1:Model) receives this as input and outputs an image of 128*128 size with three color channels (RGB). The discriminator (model_2:Model) receives this image as input and outputs a binary classification to identify the image as *true* or *false*.

Since training this model is computationally expensive, some tests were performed to evaluate the time required with different parameters, as shown in Table 2. The first column indicates the number of images used, the second the number of iterations and the third the batch size, i.e., the number of examples introduced into the network for each training. Also, in the line below the parameters are the images resulting from different parts of the training: first, middle and last.

The tests made obtaining better results possible but required considerable time investment. Additionally, tests allowed the selection of the model parameters, balancing the time and quality of the images generated to present the example.

Table 2. Tests for the selection of model parameters.

Images	Iterations	Batch size
8	20	8
		
First	Middle	Last
50	20	8
		
First	Middle	Last
10,000	5,000	3
		
First	Middle	Last

Following the testing phase, the development of the final model began. The selected parameters, image preprocessing, model construction (generator, discriminator) and the final model (GAN) training can be seen in the pseudocode presented in Algorithm 1. Subsequently, the steps followed during the implementation process are detailed.

Algorithm 1. Face Generation GAN

```
NO_ITER ← 5,000
NO_IMAGES ← 10,000
BATCH ← 6
IMG_WIDTH ← 178
IMG_HEIGHT ← 218
TARG_WIDTH ← 128
TARG_HEIGHT ← 128
dif ← (IMG_HEIGHT - IMG_WIDTH // 2)
crp ← (0, diff, IMG_WIDTH, IMG_HEIGHT - dif)
IMGS ← []
LATENT_DIM ← 32
CHANNELS ← 3
IMGS ← Crop_Images()
generator ← build_generator(gen_input=LAT_DIM,
                            LAYERS=8,
                            ACT_FUNC_LAY='Leaky ReLU',
                            ACT_FUNC_OUT='tanh',
                            )

discriminator←build_discriminator(disc_input=(TARG_WIDTH,
                                  TARG_HEIGH, CHANNELS),
                                  LAYERS=8,
                                  ACT_FUN_LAY='LeakyReLU',
                                  ACT_FUNC_OUT ='sigmoid',
                                  OPTIMIZER='RMSprop',
                                  LOSS_FUN='binary
                                            crossentropy'
                                  )

GAN_input = Input(shape=(LAT_DIM, ))
GAN_output = discriminator(generator(GAN_input))
GAN = Model(GAN_input, GAN_output)
GAN.compile(optimizer='RMSprop',
            loss='binary_crossentropy')

for t in range(NO_ITER)
  latent_vectors = samples(size=(batch_size, LATENT_DIM))
  generated = generator.predict(latent_vectors)
  real = IMGS[batch_size]
  combined_images = np.concatenate([generated, real])
  dis_loss = discriminator.train_on_batch(combined_images)
  gen_loss = GAN.train_on_batch(latent_vectors)
end for
```

Lines 1-12 contain the variables declaration. A total of 10,000 images, 5,000 iterations and a batch of size 6 were used to train the model. The initial size of the images was 178*218 with three color channels (RGB). Subsequently, the images were cropped to center the face and standardize them, leaving a size of 128*128. Each cropped image was stored in an array. It is represented in line 13, and a sample grid can be seen in Fig 8, which contains some resized images.

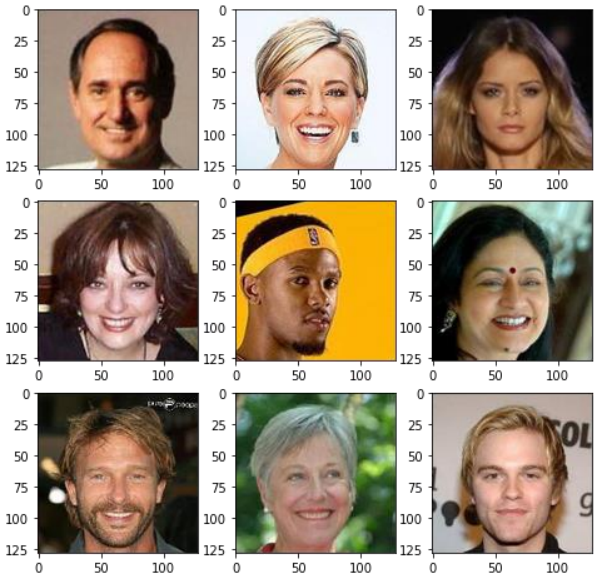


Fig 8. Preview of the resulting images after preprocessing, belonging to the Celeba public image set.

The generator is built by giving as input the latent vector, which was defined in size 32. The model is also created with its respective layers; in this case, 8 layers were constructed, and the Leaky ReLU activation function was applied to them. On the other hand, in the output layer, the hyperbolic tangent activation function was used to produce sharper images. It is also important to mention that the loss of this network is calculated from the discriminator's loss. The definition of the generator can be seen in line 14.

The discriminator receives as input an image with three RGB color channels. It was built with a structure very similar to the generator's, adding 8 layers, and the Leaky ReLU activation function was also applied to them. But the sigmoid activation function was used in its output layer, which gives the binary classification 0 and 1. The RMSprop optimizer was employed, and the binary cross entropy was used as loss function due to its utility in classification problems to measure the difference between the probabilities calculated between two possible classes. The definition of the model can be seen in line 15.

The GAN model outputs the classification resulting from the discriminator receiving as a parameter the generator, which in turn gets the latent vector. As in the discriminator, the RMSprop optimizer was applied to the GAN model and used the binary crossentropy as a loss function. This process is defined in lines 16 to 19.

The training is performed during the previously defined iterations. The "latent_vectors" variable contains the noise with which the generator will be trained. Subsequently, the "real" variable has a subset of images from the original set, and then they are merged with those coming from the generator in the "combined_images" variable. This part can be seen in lines 20 to 24.

Line 25 shows how the discriminator's loss is calculated from its training by receiving the subset created in "combined_images" as input.

Furthermore, line 26 shows the generator's loss calculated from the GAN's training that receives as input the noise vector, and this will be passed to the generator to start the training of the complete model. Finalizing with the loop in line 27 when the last epoch is reached.

After training, the resulting images can be seen in Fig 9. It can be seen how the characteristics of the images improve, learning to generate parts of the face such as eyes, mouth, nose, and textures for the background. The loss values of each network after training can be seen in Table 3, while Fig 10 presents a plot to display the behavior of the respective loss functions.

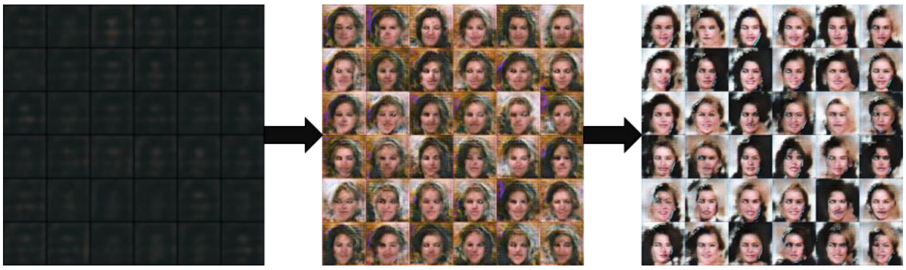


Fig 9. Images generated by the trained GAN.

Table 3. Model losses at the end of training

Epochs	Dis_loss	Gen_loss
4,900/5,000	0.7826	0.8347
4,950/5,000	0.6075	1.0423
5,000/5,000	0.8056	0.6091

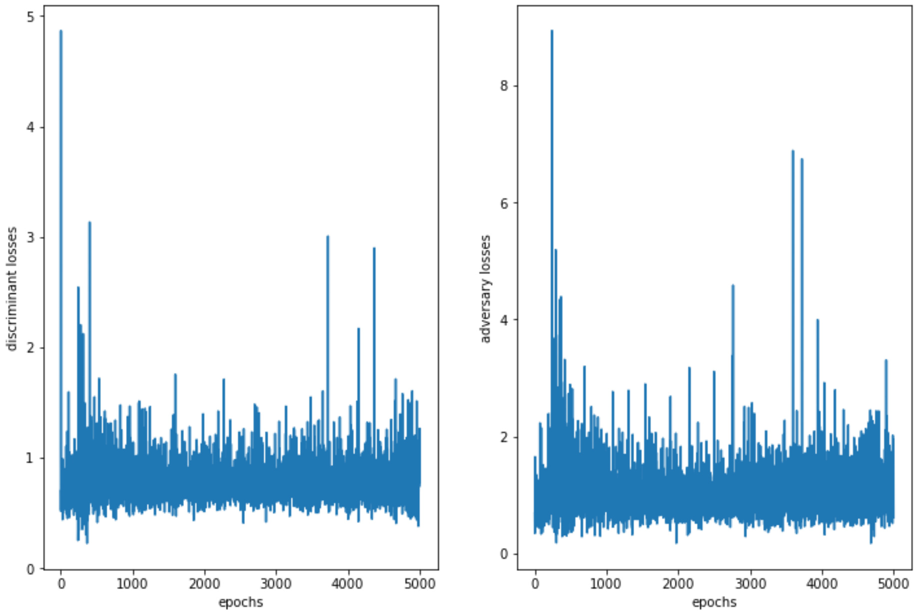


Fig 10. Model losses during GAN training

Since the GAN model comprises two neural network models, understanding its behavior is highly complex. It is irregular, with highs and lows at seemingly random points.

Conclusion and future work

The literature review, which describes the most widely used methods over time for artificial image generation, shows that GANs are effective models in this task to date, being successfully implemented in different domains. Despite the fact that they were recently proposed and the challenges their implementation represents, they show promising results. The more they are investigated, the more possibilities open up in the world of AI. One of the main challenges that was observed not only in research papers but also in experiments like the one presented above in Section 6, is the high computational power that they require. The higher the computational power, the better the quality and realism of the generated images.

In addition to the above, there is a large open area of research on the functioning of the GAN. The reason for the behavior they display during training is still unclear. For example, it is unknown why securities sometimes trigger when measuring their losses. Despite the above, it is only known that the images generated improve as the training progresses.

The metric for evaluating an artificial image is a significant limitation. Assessing whether an image looks realistic represents a high level of complexity for the human eye. Many papers resort to multiple attempts to measure it. Some use quantitative metrics, while others focus more on the human side with qualitative measures.

As future work, it is planned to analyze, evaluate, and implement a GAN architecture with the objective of simulating the age progression in the facial images of Mexican women reported missing. The idea is to visualize their appearance after a lapse of 10 to 20 years of adulthood. With this, it is intended to use GANs as a DA technique to create robust image sets that allow training artificial intelligence algorithms to help finding missing women.

Code repository

A repository which contains the code implemented for the generation of artificial faces is presented below, where is available a Jupyter Notebook shows the results of each compiled code block.

https://github.com/BlancaECS/Face_generation/blob/5584e7fb9c704135318215117 4407aa487048d53/FirstFace_GAN.ipynb

References

[1] R. E. Saragih and Q. H. To, "A Survey of Face Recognition based on Convolutional Neural Network," *Indones. J. Inf. Syst.*, vol. 4, no. 2, pp. 122–139, 2022, doi: 10.24002/ijis.v4i2.5439.

[2] U. Arora, A. R. Khan, A. Majumdar, D. Ginhac, and R. Jain, *Handbook of Face Recognition*. 2005.

[3] aws, "¿Qué es el reconocimiento facial?," 2022. https://aws.amazon.com/es/what-is/facial-recognition/.

[4] I. Masi, A. T. Trân, T. Hassner, J. T. Leksut, and G. Medioni, "Do we really need to collect millions of faces for effective face recognition?," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9909 LNCS, pp. 579–596, 2016, doi: 10.1007/978-3-319-46454-1_35.

[5] X. Wang, K. Wang, and S. Lian, "A Survey on Face Data Augmentation," no. 1, 2019, doi: 10.1007/s00521-020-04748-3.

[6] J. Langr, *GANs in action*. 2019.

[7] G. Perarnau, J. van de Weijer, B. Raducanu, and J. M. Álvarez, "Invertible Conditional GANs for image editing," no. Figure 1, pp. 1–9, 2016.

[8] S. Sun and A. Wang, “Face Aging With GAN.”

[9] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, “AttGAN: Facial Attribute Editing by only Changing What You Want,” *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5464–5478, 2019, doi: 10.1109/TIP.2019.2916751.

[10] I. Goodfellow *et al.*, “Generative adversarial networks,” *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020, doi: 10.1145/3422622.

[11] E. Mohamed, O. Elharrouss, S. Al-Maadeed, and H. Tairi, “Image Generation: A Review,” *Neural Process. Lett.*, vol. 54, no. 5, pp. 4609–4646, 2022, doi: https://doi.org/10.1007/s11063-022-10777-x.

[12] J. D. Calderon, L. Fabrico, M. Ticliahuanca, and M. Roncal, “Uso de algoritmos de machine learning para el diagnóstico de melanomas Trabajo de Investigación para optar el Grado de Bachiller en Ingeniería Mecánico - Eléctrica Jose Dario Calderon Ortiz María Esther Roncal Moscol,” no. May, 2021, doi: 10.13140/RG.2.2.20397.00483.

[13] L. Taylor and G. Nitschke, “Improving Deep Learning using Generic Data Augmentation,” 2017.

[14] W. H. Hsu, “Investigating Data Augmentation Strategies for Advancing Deep Learning Training,” 2018.

[15] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, “Learning to discover cross-domain relations with generative adversarial networks,” *34th Int. Conf. Mach. Learn. ICML 2017*, vol. 4, pp. 2941–2949, 2017.

[16] D. Guo and T. Sim, “Digital face makeup by example,” *2009 IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2009*, pp. 73–79, 2009, doi: 10.1109/CVPRW.2009.5206833.

[17] T. Li *et al.*, “BeautyGAN,” pp. 645–653, 2018, doi: 10.1145/3240508.3240618.

[18] J. Guo, X. Zhu, Z. Lei, and S. Z. Li, “Face synthesis for eyeglass-robust face recognition,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10996 LNCS, no. June, pp. 275–284, 2018, doi: 10.1007/978-3-319-97909-0_30.

[19] W. Shen and R. Liu, “Learning residual images for face attribute manipulation,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 1225–1233, 2017, doi: 10.1109/CVPR.2017.135.

[20] Z. H. Feng, J. Kittler, W. Christmas, P. Huber, and X. J. Wu, “Dynamic attention-controlled cascaded shape regression exploiting training data augmentation and fuzzy-set sample weighting,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 3681–3686, 2017, doi: 10.1109/CVPR.2017.392.

[21] J. J. Lv, X. H. Shao, J. S. Huang, X. D. Zhou, and X. Zhou, “Data augmentation for face recognition,” *Neurocomputing*, vol. 230, pp. 184–196, 2017, doi: 10.1016/j.neucom.2016.12.025.

[22] J. Deng, S. Cheng, N. Xue, Y. Zhou, and S. Zafeiriou, “UV-GAN: Adversarial Facial UV Map Completion for Pose-Invariant Face Recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 7093–

7102, 2018, doi: 10.1109/CVPR.2018.00741.

[23] W. Xie, L. Shen, M. Yang, and J. Jiang, “Facial expression synthesis with direction field preservation based mesh deformation and lighting fitting based wrinkle mapping,” *Multimed. Tools Appl.*, vol. 77, no. 6, pp. 7565–7593, 2018, doi: 10.1007/s11042-017-4661-6.

[24] J. Thies and M. Zollh, “a183-Thies,” vol. 34, no. 6, pp. 1–14, 2015.

[25] P. K. Chandaliya and N. Nain, “Child Face Age Progression and Regression using Self-Attention Multi-Scale Patch GAN,” *2021 IEEE Int. Jt. Conf. Biometrics, IJCB 2021*, pp. 1–8, 2021, doi: 10.1109/IJCB52358.2021.9484329.

[26] I. Goodfellow *et al.*, “Generative adversarial networks,” *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2014, doi: 10.1145/3422622.

[27] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” *4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.*, pp. 1–16, 2016.

[28] M. Y. Liu and O. Tuzel, “Coupled generative adversarial networks,” *Adv. Neural Inf. Process. Syst.*, no. Nips, pp. 469–477, 2016.

[29] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive growing of GANs for improved quality, stability, and variation,” *6th Int. Conf. Learn. Represent. ICLR 2018 - Conf. Track Proc.*, pp. 1–26, 2018.

[30] T. Karras, S. Laine, and T. Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4217–4228, 2019, doi: 10.1109/TPAMI.2020.2970919.

[31] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, “Analyzing and Improving the Image Quality of StyleGAN,” *Cvpr*, pp. 8107–8116, 2019.

[32] A. Nguyen, J. Yosinski, and J. Clune, “Deep neural networks are easily fooled: High confidence predictions for unrecognizable images,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 427–436, 2015, doi: 10.1109/CVPR.2015.7298640.

[33] L. R. Calcagni, “Redes Generativas Antagónicas y sus aplicaciones,” p. 72, 2020.

[34] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F. Y. Wang, “Generative adversarial networks: Introduction and outlook,” *IEEE/CAA J. Autom. Sin.*, vol. 4, no. 4, pp. 588–598, 2017, doi: 10.1109/JAS.2017.7510583.

[35] B. Raj, “Advances in Generative Adversarial Networks (GANs),” 2019. https://medium.com/beyondminds/advances-in-generative-adversarial-networks-7bad57028032.

[36] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, “CovidGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved Covid-19 Detection,” *IEEE Access*, vol. 8, pp. 91916–91923, 2020, doi: 10.1109/ACCESS.2020.2994762.

[37] M. Z. Khan *et al.*, “A Realistic Image Generation of Face from Text Description Using the Fully Trained Generative Adversarial Networks,” *IEEE*

Access, vol. 9, pp. 1250–1260, 2021, doi: 10.1109/ACCESS.2020.3015656.

[38] O. Bailo, D. Ham, and Y. M. Shin, “Red blood cell image generation for data augmentation using conditional generative adversarial networks,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2019-June, pp. 1039–1048, 2019, doi: 10.1109/CVPRW.2019.00136.

[39] O. Metri and H. R. Mamatha, “Medical image generation using generative adversarial networks,” *Gener. Advers. Networks Image-to-Image Transl.*, pp. 235–262, 2021, doi: 10.1016/B978-0-12-823519-5.00007-5.

[40] A. Dash, J. Ye, G. Wang, and H. Jin, “High Resolution Solar Image Generation Using Generative Adversarial Networks,” *Ann. Data Sci.*, pp. 1–10, 2022, doi: 10.1007/s40745-022-00436-2.

[41] D. Yasmina, R. Karima, and A. Ouahiba, “Traffic signs recognition with deep learning,” *Proc. 2018 Int. Conf. Appl. Smart Syst. ICASS 2018*, no. June, 2019, doi: 10.1109/ICASS.2018.8652024.

(v) Decision letter

DA&CI 2022 - Springer Book (chair)

Help / Log out

Submissions | Reviews | Status | PC | Events | Email | Administration | Premium | Conference | News | EasyChair

Email Instance

To	Rogelio Florencia Juárez <rogelio.florencia@uacj.mx>
Time	Mar 09, 18:05 GMT
Subject	DA&CI 2022 - Springer Book notification for paper 5352

Body

Dear Rogelio Florencia Juárez,

I am pleased to inform you that your chapter "Face image generation: A review," submitted to "Data Analytics and Computational Intelligence: Novel Models, Algorithms and Applications," has satisfactorily passed the review phase. Next, you must upload your editable files to the shared drive
https://drive.google.com/drive/folders/14MoAMMEIgdr9vmQhoXAAekeLMfyyTHHl?usp=share_link

You will find the following folders:
1. MANUSCRIPT: Here, you must upload a clean version of your approved manuscript (without line numbering and revision marks). Provide an editable document (i.e., LaTeX or Word). If you used the Word template, ".docx" files are welcome (please, don't upload ".docm" files). Lastly, provide all pertinent information about the authors: full name, affiliation, email, and ORCID (if available). Note that this is the last chance to add (or remove) names to (from) the list of authors.
2. FIGURES: Here, you must upload the figure files with a high resolution using the same names in the document (that is, "Figure1," "Figure2," and so on). Consider the following points:
2.1. Do not submit tabular material as figures.
2.2. Graphics and diagrams should be saved as EPS or TIFF files with embedded fonts.
2.3. MS Office figures can be presented in the original format (.xlsx, .pptx).
2.4. Scanned graphics in TIFF format should have a minimum resolution of 1200 dpi.
2.5. Photos or drawings with fine shading should be saved as TIFF with a minimum resolution of 300 dpi.
2.6. A combination of halftone and line art (e.g., photos containing line drawings or extensive lettering, color diagrams, etc.) should be saved as TIFF with a minimum resolution of 600 dpi.

To ensure the timely and efficient completion of this publication, please check all requirements and guidelines have been met as outlined in the Manuscript Preparation Guide: https://www.springer.com/de/authors-editors/book-authors-editors/resources-guidelines/book-manuscript-guidelines/manuscript-preparation/5636 (see section "Chapters"). No chapter will be finally published unless it strictly follows the manuscript guidelines. That is:
(a) It must be professionally copyedited, with proper use of the English language, formal grammatical structure, and correct spelling and punctuation.
(b) The references and citations are formatted according to guidelines. We encourage the authors to provide the DOI of the references.
(c) It is free of any plagiarism practices (in both figures and text). In this regard, the figures you used in the chapter must be original artwork, not taken from previous publications.

I kindly request you to upload the editable files of your approved chapter by **MARCH 23, 2023**. If you have any questions, feel free to contact me, Gilberto Rivera, at gilberto.rivera@uacj.mx (with a copy to gilberto.rivera@eurekascommunity.org).

Thank you for your diligent work in your contribution to "Data Analytics and Computational Intelligence: Novel Models, Algorithms and Applications," I greatly value your manuscript.

Sincerely yours, Gilberto RIVERA.
On behalf of the editors: Laura CRUZ-REYES, Bernabé DORRONSORO, Gilberto RIVERA, Alejandro ROSETE

Data augmentation techniques for facial image generation: a brief literature review

Blanca Elena Cazares, Rogelio Florencia, Vicente García, J. Patricia Sánchez-Solís

Abstract: Image processing has gained notoriety over the last few years in performing various tasks through deep learning (DL) algorithms, such as face recognition and identity verification. Unfortunately, most of them require a large set of images for training, usually manually labeled, which is a costly task both in time and effort, not to mention being prone to human error. Data Augmentation (DA) techniques have been used to mitigate this situation, as they generate images by applying variations to real image sets. This chapter presents a brief literature review on various DA methods dedicated to image generation. The technique that has presented outstanding results in the task of generating artificial images is Generative Adversarial Networks (GANs). Some recent research papers in which GANs have been used for the generation of artificial images are presented. General aspects of GANs, such as their definition, architecture, training, and challenges, are described. Additionally, the implementation of a GAN architecture for the generation of artificial face images from a public set of images is presented. The need for a great computational capacity to generate images with great sharpness and realism is highlighted.

Keywords: Face image generation, Data augmentation, Generative Adversarial Networks.

Blanca Elena Cazares, Rogelio Florencia (correspondence), Vicente García, J. Patricia Sánchez-Solís
Departamento de Ingeniería Eléctrica y Computación, Universidad Autónoma de Ciudad Juárez, Cd. Juárez, Chihuahua, México.
Email: blancacazaress@gmail.com (B.E.C.); rogelio.florencia@uacj.mx (R.F.); vicente.jimenez@uacj.mx (V.G.); julia.sanchez@uacj.mx (J.P.S.S.)

1. Introduction

Facial recognition systems have been discussed for several decades. The scientific community has shown great interest in this subject, given the human ability to recognize others through the face, being a unique part of the body. One of the goals is that these systems possess this same intelligence as part of the biometric identification measures.

Facial recognition has been developed in the field known as computer vision [1]. This area supports a variety of critical applications, e.g., identity verification. This task is done using what is known as a face analyzer, which is software that confirms the identity of people based on their faces [2]. It is achieved through the identification and measurement of facial features in images. It can also be used to associate human faces in the latest or even in videos to check how similar they are to one or several specific individuals. In addition, it can determine the level of similarity between two photographs to find out if they belong to the same person or search for them among a large set of images in a collection. For example, biometric security systems use facial recognition to uniquely identify individuals as users at login to strengthen user authentication. In addition, mobile devices often use this type of technology to protect the data they contain [3]. Unfortunately, large amounts of facial images are required for these processes to function properly, so much of the research has focused on artificial image generation. It aims to increase the image sets used to train different architectures because of their significant impact on the results [4].

Collecting and labeling data samples with good quality is costly, both in time and effort and is prone to human error. For this purpose, various techniques known as Data Augmentation (DA) have been implemented, and their performance is suitable in different domains.

DA techniques allow the size of existing image sets to be increased considerably through simulations [4], which helps significantly with the image requirements for learning systems. Also, DA techniques have faced several limitations of their own [5], for example:

The images generated lack realistic variations such as makeup, skin color, and background change, which means that they would have a different distribution than the real ones.

Creating high-quality facial images is very difficult due to the complexity of facial details.

After many years of research and application of different modeling methods, in 2014, a technique was proposed that allows the generation of realistic images using two artificial neural networks, Generative Adversarial Networks (GAN). Although these are not the first method used for artificial data generation, their results and versatility distinguished them from the rest since they have achieved outstanding results that were still considered impossible for artificial systems [6].

GANs are a machine learning (ML) technique that integrates two simultaneously trained models, the generator and the discriminator. The generator is trained to create the artificial images, and the discriminator is to discern between the fake images created by the generator and the real ones from the original training set. Its performance has far exceeded expectations in the field of artificial image generation, according to a variety of authors [5], [7]–[9].

Like ML algorithms, the data generated by a GAN completely depends on the training set provided to perform the learning. For example, if a GAN is required to learn how to create images of handwritten numbers, it is necessary to use a training set containing several images of handwritten numbers [6].

GANs are derived from a gaming perspective. Hence the word adversarial denotes a competitive dynamic between the two models that compose them. The generator aims to generate real images indistinguishable from the training set. On the other hand, the discriminator aims to distinguish these generated images from the real ones. Therefore, the better the generator generates realistic images, the better the discriminator must be to distinguish the real ones from the false ones [6].

The fast growth and progress of GANs have been due to research and development, generating new architectures to stabilize the outputs and generate images of higher quality and realism. In this way, the aim is to create impossible images for the human eye to differentiate.

This chapter presents a brief literature review on data augmentation methods for artificial image generation, focusing mainly on GANs, which have been used throughout recent years to create artificial facial images. Additionally, the general architecture of a GAN and related concepts are described. Lastly, the implementation of a GAN architecture for generating artificial face images, trained with a small part of the public image set "CelebA," is described.

The main contributions of this work are: a) the compilation of recent state-of-the-art works that demonstrate the different areas in which GANs have been used to improve the training of various systems or algorithms, such as ML, b) the description of some architectures and parameter variations that can be modified according to the objectives pursued and, c) an example of a GAN architecture used to generate artificial faces.

The chapter is structured as follows. Section 2 presents the background to the generation of artificial facial images. Section 3 describes the general architecture of a GAN, the training process, and its challenges, among other information of interest. Section 4 presents a brief literature review on data augmentation methods for generating artificial images, focusing mainly on GANs. Section 5 describes the methodology used to conduct the literature review. Section 6 shows the implementation of a GAN architecture to generate artificial facial images. Lastly, Section 7 presents the conclusions and future work.

2. Generation of artificial facial images

Generating new images from others is a widely researched task in computer vision. In recent years, the development of Artificial Intelligence (AI) techniques has motivated the idea of producing images of high quality and realism. It has enabled the creation of realistic human faces that are difficult to distinguish between fake and real, even for the human eye. This task has been evolving exponentially since implementing GANs [10]. However, before GANs, other computational methods were used to generate variations in facial images to train systems such as facial recognition systems.

Four main categories can be broadly discussed in the DA task for generating artificial images: generic, component, attribute, and age transformations. Each one is described in the following sections.

2.1 Generic transformations

These transformations focus on modifying the entire image, ignoring high-level components such as composition, light, volume, symmetry, shape, and texture. They are usually divided into two main groups: geometric and photometric transformations.

Geometric transformations are commonly applied in multiple computer vision tasks [5], such as face recognition, healthcare, and manufacturing applications. In general terms, geometric transformations alter the pixels of an image by placing them in new positions. Some examples of these can be:

- Reflection: consists of flipping the image around its vertical or horizontal axis.
- Rotation: rotates the image θ degrees around its center, bringing each pixel (x, y) to its position (x', y').
- Cropping: consists of cutting the images to a specific size.

For example, in the case of convolutional neural networks (CNNs), this type of transformation helps to minimize their sensitivity to changes in position and orientation [11]. Some examples can be seen in Fig 1.

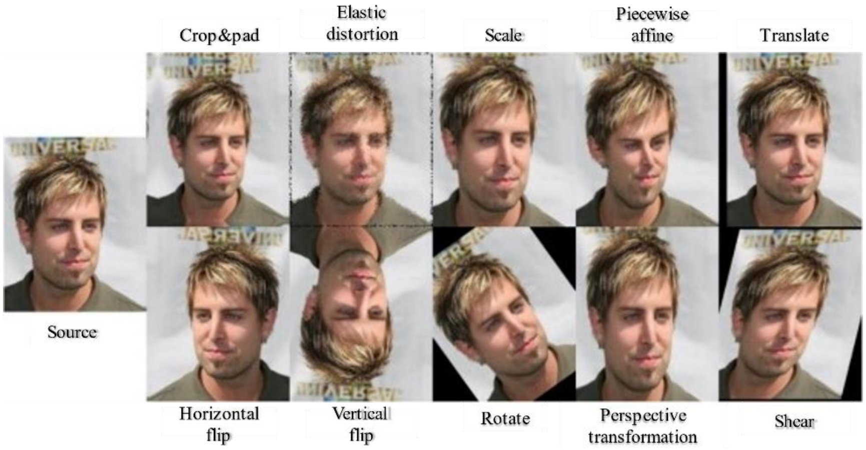


Fig 1. Geometric transformation examples [5]

On the other hand, photometric transformations, shown in Fig 2, generally work by altering the RGB (red, green, and blue) channels, shifting each pixel value (r, g, b) to new values (r', g', b') of an image according to predefined rules. These transformations adjust the illumination and color, leaving the geometry unaffected [11]. For example, it can be mentioned color manipulations, such as inverting them or adding some filters, such as blurring or grayscale [5].

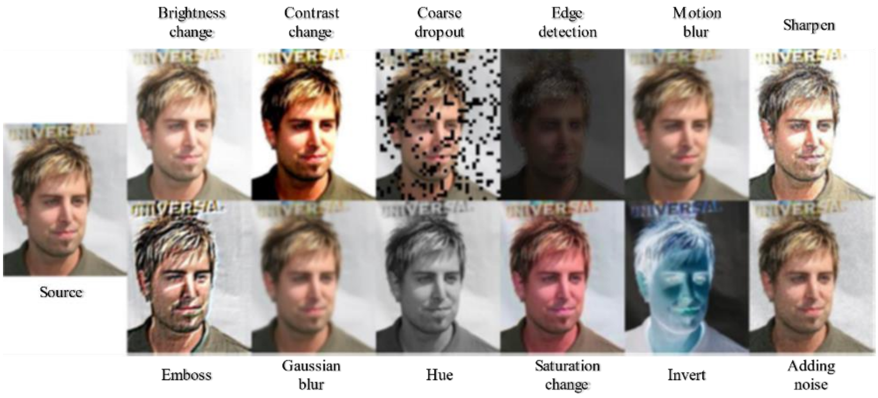


Fig 2. Photometric transformation examples [5]

These transformations are mainly used in computer vision tasks to enrich the training sets and prevent a common problem in this field, overfitting. [12] is a paper dedicated to evaluating the performance of CNNs, trained with images enriched using geometric and photometric transformations. In [13], different transformations were employed to increase the image set and prevent overfitting.

2.2 Component transformation

There are transformations dedicated to enriching the sets of facial images by modifying the components of the person. These images are used to train the algorithm so that it is able to recognize the person, even if their appearance is altered.

The hairstyle can be considered one of the components to be generated since, although it is not considered a facial component, hairstyle affects face detection and recognition because it tends to hide certain features of a person's appearance. Therefore, DA techniques focus on generating facial images with different variations in hair, for example, color, shape, and bangs. In [14], a method using DiscoGAN was proposed. This variant learns to discover relationships between different domains and develops the ability to translate features between them, for example, by transforming hair color.

Makeup transfer and accessory addition techniques can also be identified due to the difficulty for recognition systems to effectively perform their tasks when some features of the face look different depending on the makeup or accessories a person wears. Most studies based on this type of transformation can be divided into two categories: traditional image processing [15] and those based on DL [16]. Some examples of these can be seen in Fig 3.

Another component that impacts facial recognition is the use and removal of accessories, including glasses, earrings, and piercings, among others. Of all these, glasses are the most commonly used, as they are used for different reasons, for example, vision correction, prevention against sunlight, eye protection, and aesthetics, among others. They significantly affect the accuracy of facial recognition, as they usually cover a large area of the face.

In [17], a fusion of virtual lenses onto faces was performed using the Augmented Reality (AR) technique. In [18], a method of facial attribute manipulation based on image residuals was proposed, defining this as the difference between the input image and the desired output image.

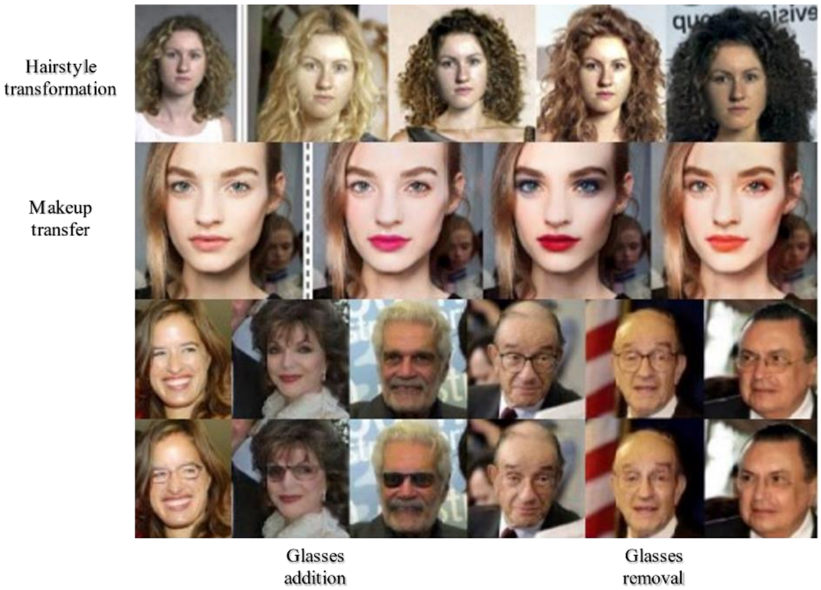


Fig 3. Component transformation examples [5]

2.3 Attribute transformation

There are also some transformations dedicated to modifying aspects such as pose. In this case, the position of the head in a photograph is considered a significant challenge in facial recognition tasks since any variation in it tends to modify visual aspects of the face, i.e., it can hide or show different facial details. In addition, it has been considered an essential aspect since facial photographs in several legal processes are requested from the frontal side. This leads to the research of recreating how a face may look from other angles.

Facial expressions are also considered critical: happy, annoyed, scared, and surprised, among others. These techniques help to improve the performance of emotion classification and recognition systems. Primarily 2D, 3D, and learning-based modeling approaches are used to achieve this goal and usually focus on modifying the expression of a face using expression templates by concentrating on a series of points, for example, the corners of the mouth, the cheeks position, and the location of the eyebrows. Others focus on simulations to recreate face parts that are hidden by the pose in which the photograph was taken [4], [19]–[21]. Some examples of this can be found in Fig 4.

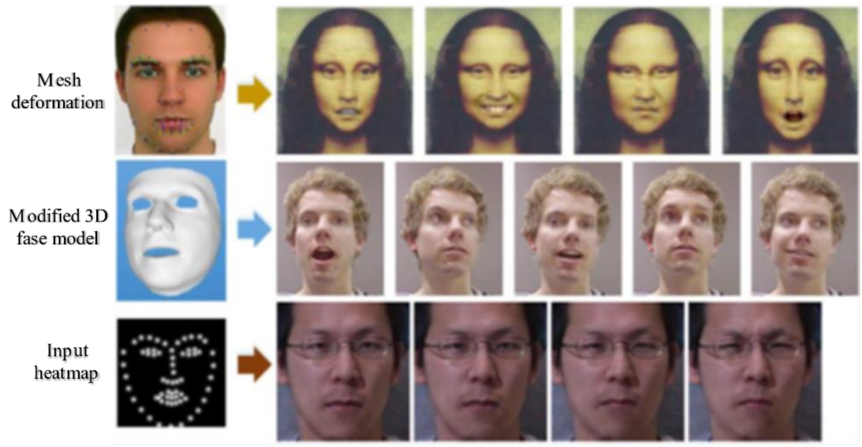


Fig 4. Facial expression synthesis [22], [23]

2.4 Age Progression and Regression

Finally, age regression or progression seeks to predict a current face's appearance in the past or future, respectively, while preserving its facial features. It has become a widely explored research topic because it significantly affects various applications, including missing person tracking, facial recognition, and aesthetic studies, among others. The two main concerns in attempting age regression or progression are identity preservation and prediction accuracy [24].

These facial images enrich the image sets by adding different features at different time stages of the same face, which makes the models more robust to age variations.

Age modification methods are mainly divided into two: prototype-based and model-based. The first one creates an average face for different age ranges, learns the shapes and textures of these, and applies these features to transfer them to a new face. However, the individual's characteristics are often lost with these methods [5].

Model-based methods build, as the name implies, models of biological changes due to age in faces, e.g., musculature, wrinkles, and skin texture, among others. They are called generative because they are such a powerful tool for creating new data by learning to imitate the probabilistic distribution of a training set. Nowadays, generative models have gained importance and attention due to their good performance in data creation. Among the most popular are autoregressive models, Variational Autoencoders (VAEs), and GANs. The disadvantage of these models is their complexity and computational cost [5].

GANs, as mentioned above, are an alternative architecture for training generative models since they handle probabilistic computations very well. Recent work has begun to apply them to the age regression and progression task, and many variants of the model have been generated. Some examples of this generative task are shown in Fig 5.

Fig 5. Facial age regression and progression examples [5]

Before GAN, two approaches to age progression and regression in faces were prototype-based and modeling-based. In the latter, critical points in the image, such as eyes, nose, and jaw, track, in turn, the temporal changes such as wrinkles, musculature, and color in these are identified. However, this method requires a large amount of age-labeled data over a long period for each individual, which is difficult to find and computationally expensive. The prototype-based method creates an average face based on a set of images of a particular age group, using it to transfer those features from one age range to another. The disadvantage of this is that personal features are often lost. Another possibility is to use neural networks to transform faces across ages. It generates smoother images but still requires images labeled with the person's age through the years [8].

On the other hand, the GAN consists of a discriminator and a generator competing with each other based on the *min-max* games. The generator starts by receiving as input a noise vector *z* and creates an image that it gives to the discriminator to receive feedback from it. Some variants of this architecture are [8]:

- DCGAN (Deep Convolutional GAN) has demonstrated that GAN can be successfully applied to generate indoor scenarios and human faces.
- StyleGAN significantly extended the basic GAN to progressively generate high-resolution images from those with very low resolutions.

cGAN (Conditional Generative Adversarial Networks) introduces an identity preservation vector with the optimization approach when generating faces so there is a better match between the original and the created face.

- Pyramid GAN simulates the effects of age more sharply and presents a suite of methods for assessing accuracy and fidelity to the original image.

3. Generative Adversarial Networks (GANs)

GANs are powerful AI-based unsupervised learning algorithms that aim to learn the estimated probability distribution in a specific training set. GANs were proposed by Ian Goodfellow in 2014 [25]. They are based on a competition system between two neural network models that try to maximize their performance while minimizing that of their adversary, developing the ability to analyze, capture and copy the variations presented by a particular set of images [8].

These networks can generate artificial data, being one of the generative models with the highest quality of results, especially when its potential to generate high-resolution images is analyzed [6]. Section 3.1 gives a definition of what a GAN is. Section 3.2 describes the general architecture with which GANs were originally proposed. Section 3.3 describes the GAN training process. Section 3.4 mentions the challenges researchers face when training a GAN. Finally, Section 3.5 presents the face image generation evolution with GANs.

3.1 Definition

It has been demonstrated that most neural networks can be easily tricked into misclassifying by adding only a small amount of noise to the original data. Surprisingly, after this addition, the model develops a higher confidence level in the wrong predictions than in the correct ones. It is because most ML algorithms learn from a limited amount of data, which is prone to model overfitting [26]. It motivated the creation of GANs. They can be described in three parts:

- Generative: They are considered generative models since they describe how new data are generated in terms of probabilistic models.
- Adversarial: The model is trained by competition among its neural networks, i.e., they are considered adversarial to each other.
- Networks: They use neural networks as the primary training algorithms.

They are based on game theory, which considers players to be both ML models, typically implemented using neural networks. A network is called a generator, which can learn the distribution obtained from an original data set to try to replicate it. It is achieved by inserting a noise vector z, i.e., random numbers with a Gaussian distribution. The main objective of the generator is to learn how to transform unstructured noise z into realistic samples [25].

The other player is called the discriminator. It examines each example x received as input and outputs an estimate of whether it is true or false.

Each player has a cost, so they try to minimize their own, i.e., the discriminator's cost encourages it to correctly classify the data as real or fake. In contrast, the cost of the generator encourages it to generate data that the discriminator incorrectly classifies as real.

A typical example when talking about GANs is to imagine that one network represents a money counterfeiter and the other a policeman. The former generates counterfeit bills while the policeman tries to arrest him for benefiting the production of legitimate ones. The competition leads to the production of more and more realistic counterfeit bills until, eventually, the counterfeiter produces them so realistically that the policeman cannot tell the difference between authentic and counterfeit [25].

3.2 Architecture

In the GAN architecture, the generator will start training simultaneously as the discriminator does. The latter will need a few epochs before starting the adversarial training since it must be able to classify the images correctly.

The architecture consists of two competing deep neural networks: the *generator* and the *discriminator*. The generator produces new data instances, while the discriminator tries to distinguish accurately between the real data, i.e., those coming from the original set or fake data produced by the generator.

This competition will continue until the generator can create realistic artificial data, which can then be used as input to other neural networks [6]. Fig 6 shows a GAN schematic with its basic components.

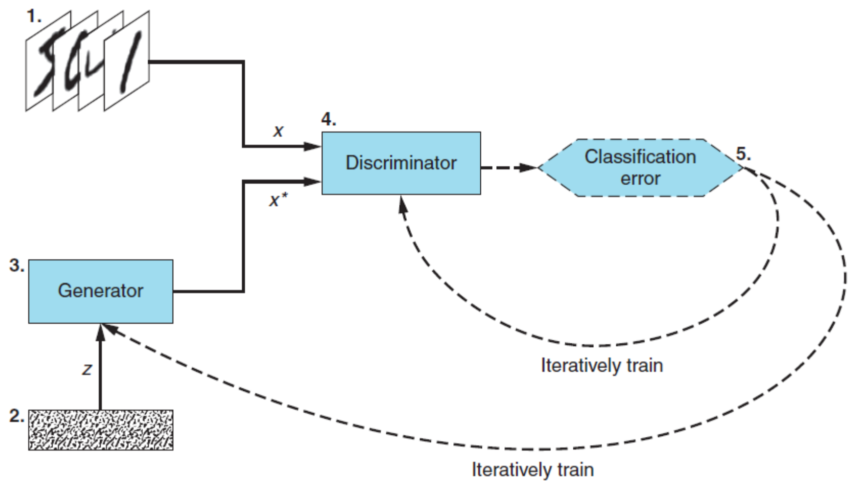


Fig 6. The two GAN subnetworks with their inputs, outputs, and interactions [6]

Since its development, many different architectures have been proposed to deal with a wide variety of domains. It could even be said that several scientific papers are published weekly [27]. In [28], an extensive literature review is done on multiple architectures developed for GANs.

3.3 Training process

The training phase requires the two networks and a set of data from which the artificial data will be generated. First of all, it is worth mentioning that the training of the generator is much more complex than the discriminator, which can be seen more as a binary classifier.

The discriminator receives real inputs that must be labeled as such and inputs from the generator, which must be labeled false, 1 if the input data is real and 0 if it is false. On the other hand, the generator must be trained with the only condition that the data created mislead the discriminator, i.e., it must minimize its loss and maximize that of the opponent. To achieve this, the generator's output must be the input of the discriminator so that the output of the whole model gives as output the probability that the data are real, according to the discriminator. In this way, the generator obtains feedback from the discriminator, which it uses to create progressively more similar data in the training set [25].

3.4 Challenges

The GANs are still facing research challenges regarding their training. Among them, the following can be pointed out [29]:

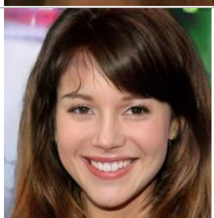
- Non-convergence: this is when the generator and the discriminator fail to reach the desired equilibrium (50%). Their respective loss functions begin to fluctuate without being able to reach stability.
- Modal collapse: occurs when the generator produces similar data, even though the inputs vary in characteristics. It finds a small set of samples that successfully deceive the discriminator and are thus incapable of producing others. In these cases, the gradient of the loss function is stuck at a value close to 0.
- Non-informative loss: the general intuition is that the lower the loss of the generator, the higher the quality of the data it produces. However, the loss must be compared with the discriminator's, which constantly improves. Therefore, the issue of model evaluation is more complex. The generator may produce better-quality samples even as the loss function increases.

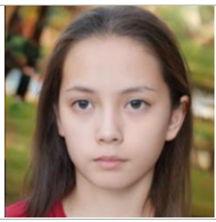
Due to these challenges, techniques have been developed with the aim of minimizing them as much as possible. However, they are still the subject of research at the moment [6].

3.5 Face image generation evolution with GANs

Progress in facial image generation has gone hand in hand with GANs, so it is now possible to control the resolution or quality of artificial images. Table 1 shows the main advances of the GANs.

Table 1. GAN evolution in artificial facial image generation.

Year	Image	Model description
2014		The proposed GAN architecture is composed of two models: generative *G* and discriminative *D*, represented by a multilayer perceptron [25].
2015		The proposed DCGAN architecture for unsupervised learning is composed of two models: *G* and *D*, represented by CNNs [30].
2016		Coupled Generative Adversarial Network (CoGAN) architecture proposed, which consists of two models for *G* and two for *D* [31].
2017		Proposed new training for GANs by progressively increasing both *G* and *D*, starting with low resolution and adding new layers. It is argued that it improves the training speed and stabilizes it [32].
2018		Proposed alternative architecture for *G* in GANs, called StyleGAN. This leads to automatically learned and unsupervised high-level attribute separation [33].

2019		Improvements in the StyleGAN architecture and its training method are proposed. In addition, further work is being done on the quality of the generated images [34].

4. Related work

The following research works have focused on implementing some GAN variants with different objectives.

In [9], it is mentioned that the main objective is to generate a new face by editing facial attributes in the images, preserving its identity. Subsequently, it proceeds to perform a literature review, finding that GAN architecture with an encoder is usually incorporated for such tasks with promising results. Thus, it is proposed to apply an attribute classification constraint to the generated images, in addition to the previous architecture, forming together what is called AttGAN. Experiments were performed on the public set of images, “Celeba,” to manipulate attributes such as hair color, beard, and age, achieving realism and preservation of facial details.

In [8], a CNN was developed to generate images considering facial age progression. To do this, the authors conducted an exhaustive review of the literature. They analyzed several sets of images to select the one with the greatest variety and fit the domain they wanted to develop the model. Then, they configured parameters and loss functions for network stabilization and the creation of realistic photographs. They used qualitative and quantitative metrics for the evaluation, resulting in images with appropriate characteristics for age progression.

The makeup transfer is the object of study to transfer a specific style to a clean face, preserving the identity of the same. This type of problem at the instance level is considered a great challenge since the styles vary greatly, for example, eye shadows, lipsticks, and foundations, among many others. In [16], BeautyGAN is proposed so that the networks can perform the transfer at the instance level through unsupervised adversarial learning. At the end of the study, a new set of high-resolution makeup images was constructed.

GANs even have an impact in the medical area, as can be seen in [35]. A study on coronavirus (COVID-19), a viral disease caused by the SARS-CoV-2 respiratory syndrome, begins, encouraged by its global effect on health and economics. Chest X-rays from infected patients were a crucial step in controlling this virus. Thus began the introduction of various DL systems and studies that demonstrated the efficiency of using chest X-rays for patient detection. In this context, since CNNs require a significant amount of training data to perform adequately and the virus was too recent to have enough Chest X-rays to generate systems that learn to detect it, the authors present a method to generate images from Chest X-rays in short times. It was achieved by introducing a model named CovidGAN, proving that the artificial images produced by it helped improve the performance of the trained CNNs for COVID-19 detection, increasing their accuracy to 95%.

In [36], it focuses on another important application of GAN: the generation of facial images from text. It has multiple applications in public safety and forensic analysis, such as finding criminals or suspects described by eyewitnesses. This chapter proposes synthesizing facial images from text using a fully-trained generative adversarial network (FTGAN), trained with a text encoder and an image decoder, to generate good-quality images from the input sentences. Multiple experiments were performed on the CUB public set, providing good results regarding the main objective. It was measured by comparison against methods found in the literature, using the Frechet Inception Distance (FID) and Face Semantic Distance (FSD) metrics. Additionally, human ratings were used to validate the generated images.

Another application in the medical field was discussed in [37], where a method was developed to synthesize photorealistic microscope images of red blood cells using cGAN. These images were combined with real ones to augment reduced datasets. Using cGANs as a synthetic data generator for the DA task benefits the development of a model. However, they also raise the fact that all that is involved in developing a GAN model: unstable and time-consuming training, heavy computational requirements, and other challenges may not be worth the small margin of improvement it may provide in some tasks.

A literature review is presented in [38], mainly focused on generating medical images using GAN frameworks since it is stated that these have proven to be useful in many cases of image augmentation, medical image generation, and image reconstruction, among others. These characteristics were the ones that encouraged the research of GANs in the medical field to improve image analysis. It is mentioned that several GAN frameworks have gained popularity in medical image interpretation, such as DCGAN, Laplacian GAN (LAPGAN), pix2pix, Cycle-GAN, and unsupervised image-to-image translation model (UNIT). Although despite all these approaches, it was concluded that they are still in very early stages, and in-depth research in the area is still needed to reach the most reliable level of progress for GAN applications in clinical imaging.

Among a wide range of applications where GANs have been implemented, one can be found in [39]. In this study, Pix2PixHD was applied to perform image-to-image translation of solar images to give scientists access to complete sets of images for analysis. The end of the training and testing proved the ability of such algorithms to generate high-resolution images. In addition, the images obtained through this model proved to be better than those obtained in previous works also intended to generate these images. It opens the possibility of using these models to create images when unavailable and thus have a better understanding of space weather. It also allows researchers to have the ability to predict solar events, for example, Solar Flares or Coronal Mass Ejections.

In [40], research is found to be motivated by the good performance of CNNs in traffic sign detection (TSD) and recognition (TSR), which are critical tasks in the development of autonomous driving systems. It is difficult because training neural networks requires a large amount of labeled data, and the visual design of traffic signs varies greatly, especially when dealing with different countries. This work focused on Taiwanese traffic signs, motivated by not finding a database, images, or study that focused on identifying them. It also pursued the generation of synthetic images when it was impossible to collect reference images through research and development. DCGANs and Wasserstein generative adversarial networks (Wasserstein GANs, WGANs) were used. Each demonstrated exceptional outputs when creating synthetic graphics. The evaluation of the efficiency of DCGAN and WGAN is also included by measuring the quality of the generated images using Structural Similarity Index (SSIM) and the Mean Square Error (MSE). It was concluded that synthetic traffic signal images could be generated with a minimal training image set. Although the resulting images' quality was different obtained with a large training set, they can still be used to solve difficulties when getting genuine photographs is complicated.

5. Methodology

The search and selection of the articles presented in this chapter were based on the Prisma (Preferred Reporting Items for Systematic Reviews and Meta-analyses) methodology, which includes a guide to identify, select, evaluate, and synthesize studies [41]. Prisma was designed for systematic reviews of health-related studies, but it can also be applied to social and educational interventions. In this way, researchers interested in reviewing the scientific literature can recreate the author search process.

The Google Scholar search engine was used to perform the literature review, which indexes full texts or metadata of the academic literature from a wide range of publication formats. The search was started considering the terms “Generative Adversarial Networks” to have a first vision of the research works that have been carried out, available in public databases. Subsequently, the search results were filtered based on the year of publication, “Generative Adversarial Networks,” 2019, to identify recently published papers. In addition, a logical operator was added to further narrow the search for “Generative Adversarial Networks” AND “image generation.” This allowed filtering of the papers found to identify only those focused on image generation.

Since GANs are used for text, images, videos, voice, and statistics, among others [6], a logical operator was added to further narrow the search to “Generative Adversarial Networks” AND “image generation.” This allowed filtering of the articles found to identify those focused on image generation. The search returned approximately 16,700 results (February 3, 2023). Of these, we selected and analyzed those that we considered being the most like this work. The search returned approximately 16,700 results (February 3, 2023). Of these, we selected and analyzed those that we consider being the most similar to the purposes pursued by this work.

6. Face image generation with GANs

In this section, the architecture of a DCGAN for generating facial images is presented to exemplify the use of GANs. The model was trained using the public image set “Celeba” [42], comprising approximately 200,000 celebrity facial images. It was selected since it contains many variations such as hair color, smile, glasses, poses, backgrounds, and diversity of people that make it very suitable for training face detection models. In addition, its elements are labeled for efficient use in computer vision. A virtual machine generated and configured by the Laboratorio Nacional de Tecnologías de Información (LaNTI)[2] at Universidad Autónoma de Ciudad Juárez (UACJ) was used. The system specifications are Ubuntu 20.04, dual-core, 100 GB storage, and 8 GB RAM. The architecture is shown in Fig. 7.

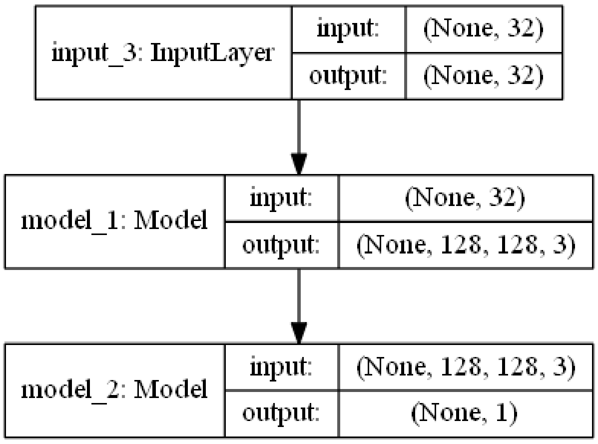


Fig 7. Architecture of a DCGAN model for generating facial images where *model_1* represents the generator and *model_2* the discriminator.

The input layer (input_3:InputLayer) represents the latent vector defined as 32, and the generator (model_1:Model) receives this as input and outputs an image of 128*128 size with three color channels (RGB). The discriminator (model_2:Model) receives this image as input and outputs a binary classification to identify the image as *true* or *false*.

Since training this model is computationally expensive, some tests were performed to evaluate the time required with different parameters, as shown in Table 2. The first column indicates the number of images used, the second the number of

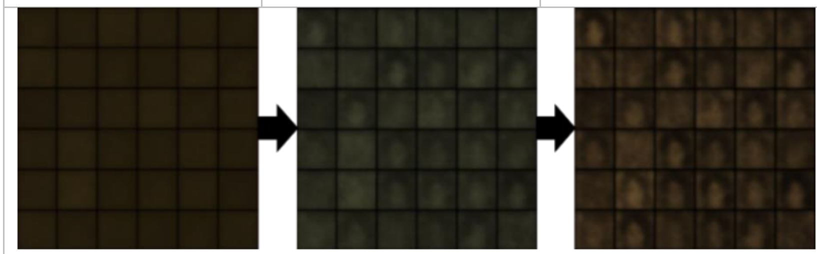
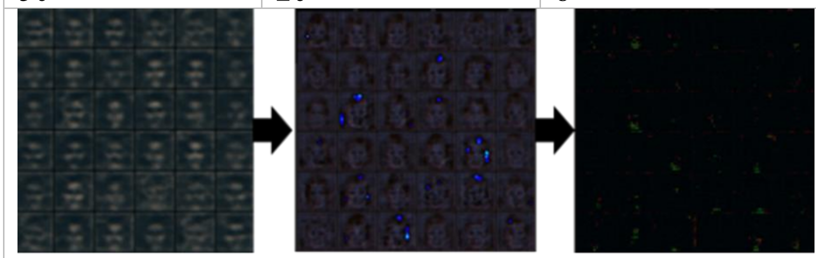
---

[2] Laboratorio Nacional de Tecnologías de Información: http://www.lanti.org.mx/lanti/

iterations, and the third the batch size, i.e., the number of examples introduced into the network for each training. Also, in the line below the parameters are the images resulting from different parts of the training: first, middle, and last.

The tests made obtaining better results possible but required considerable time investment. Additionally, tests allowed the selection of the model parameters, balancing the time and quality of the images generated to present the example.

Table 2. Tests for the selection of model parameters.

Images	**Iterations**	**Batch size**
8	20	8
		
First	Middle	Last
50	20	8
		
First	Middle	Last
10,000	5,000	3
		
First	Middle	Last

Following the testing phase, the development of the final model began. The selected parameters, image preprocessing, model construction (generator, discriminator), and the final model (GAN) training can be seen in the pseudocode presented in Algorithm 1. Subsequently, the steps followed during the implementation process are detailed.

Algorithm 1. Face Generation GAN

```
NO_ITER ← 5,000
NO_IMAGES ← 10,000
BATCH ← 6
IMG_WIDTH ← 178
IMG_HEIGHT ← 218
TARG_WIDTH ← 128
TARG_HEIGHT ← 128
dif ← (IMG_HEIGHT – IMG_WIDTH // 2)
crp ← (0, diff, IMG_WIDTH, IMG_HEIGHT – dif)
IMGS ← []
LATENT_DIM ← 32
CHANNELS ← 3
IMGS ← Crop_Images()
generator ← build_generator(gen_input=LAT_DIM,
                            LAYERS=8,
                            ACT_FUNC_LAY=‘Leaky ReLU’,
                            ACT_FUNC_OUT=‘tanh’,
                            )

discriminator←build_discriminator(disc_input=(TARG_WIDTH,
              TARG_HEIGH, CHANNELS),
                            LAYERS=8,

ACT_FUN_LAY=‘LeakyReLU’,
                            ACT_FUNC_OUT =‘sigmoid’,
                            OPTIMIZER=‘RMSprop’,
                            LOSS_FUN=‘binary
                  crossentropy’
              )

GAN_input = Input(shape=(LAT_DIM, ))
GAN_output = discriminator(generator(GAN_input))
GAN = Model(GAN_input, GAN_output)
GAN.compile(optimizer=‘RMSprop’,
      loss=‘binary_crossentropy’)

for t in range(NO_ITER)
  latent_vectors = samples(size=(batch_size, LATENT_DIM))
  generated = generator.predict(latent_vectors)
  real = IMGS[batch_size]
  combined_images = np.concatenate([generated, real])
  dis_loss = discriminator.train_on_batch(combined_images)
  gen_loss = GAN.train_on_batch(latent_vectors)
```

27: **end for**

Lines 1–12 contain the variables declaration. A total of 10,000 images, 5,000 iterations, and a batch of size six were used to train the model. The initial size of the images was 178*218 with three color channels (RGB). Subsequently, the images were cropped to center the face and standardize them, leaving a size of 128*128. Each cropped image was stored in an array. It is represented in line 13, and a sample grid can be seen in Fig 8, which contains some resized images.

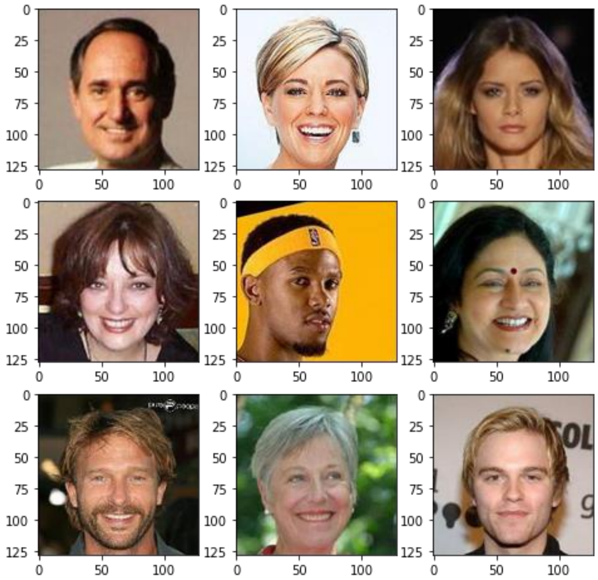


Fig 8. Preview of the resulting images after preprocessing, belonging to the Celeba public image set.

The generator is built by giving as input the latent vector, which was defined in size 32. The model is also created with its respective layers; in this case, eight layers were constructed, and the Leaky ReLU activation function was applied to them. On the other hand, in the output layer, the hyperbolic tangent activation function was used to produce sharper images. It is also important to mention that the loss of this network is calculated from the discriminator's loss. The definition of the generator can be seen in line 14.

The discriminator receives as input an image with three RGB color channels. It was built with a structure very similar to the generator's, adding eight layers, and the Leaky ReLU activation function was also applied to them. But the sigmoid activation function was used in its output layer, which gives the binary classification 0 and 1. The RMSprop optimizer was employed, and the binary cross entropy was used as a loss function due to its utility in classification problems to measure the difference between the probabilities calculated between two possible classes. The definition of the model can be seen in line 15.

The GAN model outputs the classification resulting from the discriminator receiving as a parameter the generator, which in turn gets the latent vector. As in the discriminator, the RMSprop optimizer was applied to the GAN model and used the binary cross-entropy as a loss function. This process is defined in lines 16 to 19.

The training is performed during the previously defined iterations. The “latent_vectors” variable contains the noise with which the generator will be trained. Subsequently, the “real” variable has a subset of images from the original set, and then they are merged with those coming from the generator in the “combined_images” variable. This part can be seen in lines 20 to 24.

Line 25 shows how the discriminator’s loss is calculated from its training by receiving the subset created in “combined_images” as input.

Furthermore, line 26 shows the generator’s loss calculated from the GAN’s training that receives as input the noise vector, and this will be passed to the generator to start the training of the complete model. Finalizing with the loop in line 27 when the last epoch is reached.

After training, the resulting images can be seen in Fig 9. It can be seen how the characteristics of the images improve, learning to generate parts of the face such as eyes, mouth, nose, and textures for the background. The loss values of each network after training can be seen in Table 3, while Fig 10 presents a plot to display the behavior of the respective loss functions.

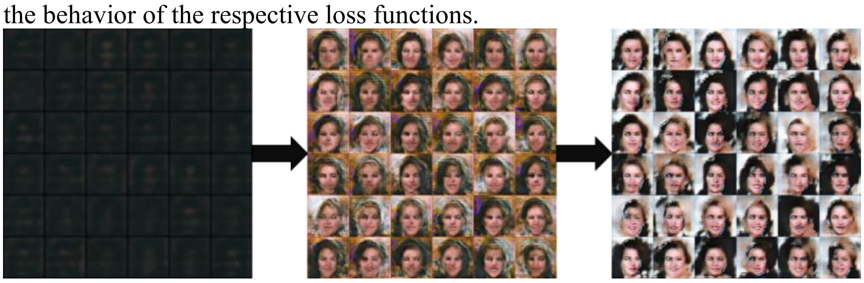


Fig 9. Images generated by the trained GAN.

Table 3. Model losses at the end of the training

Epochs	Dis_loss	Gen_loss
4,900/5,000	0.7826	0.8347
4,950/5,000	0.6075	1.0423
5,000/5,000	0.8056	0.6091

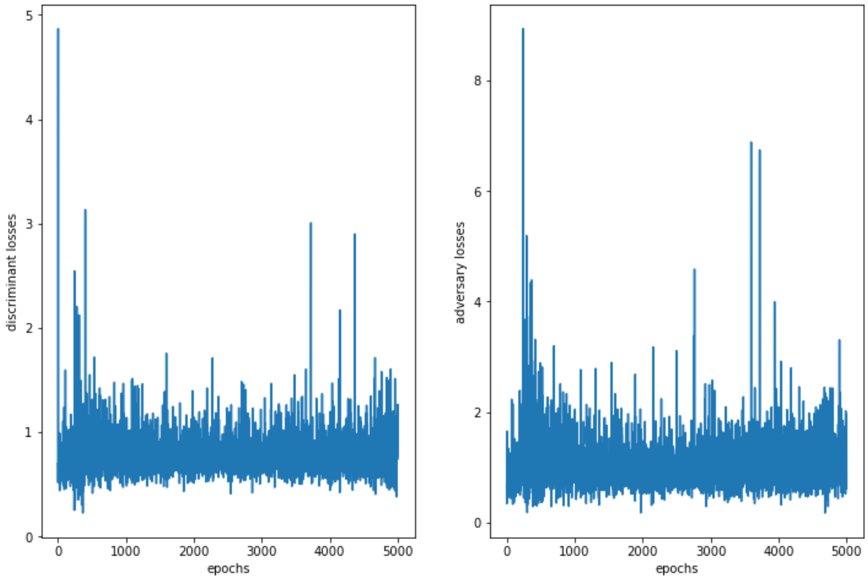


Fig 10. Model losses during GAN training

Since the GAN model comprises two neural network models, understanding its behavior is highly complex. It is irregular, with highs and lows at seemingly random points (cf. [43]).

Conclusion and future work

The literature review, which describes the most widely used methods over time for artificial image generation, shows that GANs are effective models in this task to date, being successfully implemented in different domains. Despite the fact that they were recently proposed and the challenges their implementation represents, they show promising results. The more they are investigated, the more possibilities open up in the world of AI. One of the main challenges that were observed not only in research papers but also in experiments like the one presented above in Section 6, is the high computational power that they require. The higher the computational power, the better the quality and realism of the generated images.

In addition to the above, there is a large open area of research on the functioning of the GAN. The reason for the behavior they display during training is still unclear. For example, it is unknown why securities sometimes trigger when measuring their losses. Despite the above, it is only known that the images generated improve as the training progresses.

The metric for evaluating an artificial image is a significant limitation. Assessing whether an image looks realistic represents a high level of complexity for the human eye. Many papers resort to multiple attempts to measure it. Some use quantitative metrics, while others focus more on the human side with qualitative measures.

As future work, it is planned to analyze, evaluate, and implement a GAN architecture with the objective of simulating the age progression in the facial images of Mexican women reported missing. The idea is to visualize their appearance after a lapse of 10 to 20 years of adulthood. With this, it is intended to use GANs as a DA technique to create robust image sets that allow training artificial intelligence algorithms to help find missing women.

Code repository

A repository that contains the code implemented for the generation of artificial faces is presented below, where is available a Jupyter Notebook that shows the results of each compiled code block.

https://github.com/BlancaECS/Face_generation/blob/5584e7fb9c7041353182151174407aa487048d53/FirstFace_GAN.ipynb

References

1 R. E. Saragih and Q. H. To, "A Survey of Face Recognition based on Convolutional Neural Network," Indones. J. Inf. Syst., vol. 4, no. 2, pp. 122–139, 2022. https://doi.org/10.24002/ijis.v4i2.5439

2 Jain, A. K., & Li, S. Z., "Handbook of face recognition" (Vol. 1). New York: Springer (2011). https://doi.org/10.1007/978-0-85729-932-1

3 aws, "¿Qué es el reconocimiento facial?," 2022. https://aws.amazon.com/es/what-is/facial-recognition/

4 I. Masi, A. T. Trân, T. Hassner, J. T. Leksut, and G. Medioni, "Do we really need to collect millions of faces for effective face recognition?," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9909 LNCS, pp. 579–596, 2016. https://doi.org/10.1007/978-3-319-46454-1_35

5 X. Wang, K. Wang, and S. Lian, "A Survey on Face Data Augmentation," no. 1, 2019. https://doi.org/10.1007/s00521-020-04748-3

6 J. Langr and V. Bok, GANs in action: Deep learning with Generative Adversarial Networks. Manning (2019).

7 G. Perarnau, J. van de Weijer, B. Raducanu, and J. M. Álvarez, "Invertible Conditional GANs for image editing," no. Figure 1, pp. 1–9, 2016. https://doi.org/10.48550/arXiv.1611.06355

8 S. Sun and A. Wang, "Face Aging With GAN."

9 Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, "AttGAN: Facial Attribute Editing by only Changing What You Want," IEEE Trans. Image Process., vol. 28, no. 11, pp. 5464–5478, 2019. https://doi.org/10.1109/TIP.2019.2916751

10 E. Mohamed, O. Elharrouss, S. Al-Maadeed, and H. Tairi, “Image Generation: A Review,” Neural Process. Lett., vol. 54, no. 5, pp. 4609–4646, 2022. https://doi.org/https://doi.org/10.1007/s11063-022-10777-x

11 J. D. Calderon, L. Fabrico, M. Ticliahuanca, and M. Roncal, “Uso de algoritmos de machine learning para el diagnóstico de melanomas Trabajo de Investigación para optar el Grado de Bachiller en Ingeniería Mecánico - Eléctrica Jose Dario Calderon Ortiz María Esther Roncal Moscol,” no. May, 2021. https://doi.org/10.13140/RG.2.2.20397.00483

12 L. Taylor and G. Nitschke, “Improving Deep Learning using Generic Data Augmentation,” 2017. https://doi.org/10.48550/arXiv.1708.06020

13 W. H. Hsu, “Investigating Data Augmentation Strategies for Advancing Deep Learning Training,” 2018.

14 T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, “Learning to discover cross-domain relations with generative adversarial networks,” 34th Int. Conf. Mach. Learn. ICML 2017, vol. 4, pp. 2941–2949, 2017. https://doi.org/10.48550/arXiv.1703.05192

15 D. Guo and T. Sim, “Digital face makeup by example,” 2009 IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2009, pp. 73–79, 2009. https://doi.org/10.1109/CVPRW.2009.5206833

16 Li, T., Qian, R., Dong, C., Liu, S., Yan, Q., Zhu, W. and Lin, L., “BeautyGAN,” pp. 645–653, 2018. https://doi.org/10.1145/3240508.3240618

17 J. Guo, X. Zhu, Z. Lei, and S. Z. Li, “Face synthesis for eyeglass-robust face recognition,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10996 LNCS, no. June, pp. 275–284, 2018. https://doi.org/10.1007/978-3-319-97909-0_30

18 W. Shen and R. Liu, “Learning residual images for face attribute manipulation,” Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 1225–1233, 2017. https://doi.org/10.1109/CVPR.2017.135

19 Z. H. Feng, J. Kittler, W. Christmas, P. Huber, and X. J. Wu, “Dynamic attention-controlled cascaded shape regression exploiting training data augmentation and fuzzy-set sample weighting,” Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 3681–3686, 2017. https://doi.org/10.1109/CVPR.2017.392

20 J. J. Lv, X. H. Shao, J. S. Huang, X. D. Zhou, and X. Zhou, “Data augmentation for face recognition,” Neurocomputing, vol. 230, pp. 184–196, 2017. https://doi.org/10.1016/j.neucom.2016.12.025

21 J. Deng, S. Cheng, N. Xue, Y. Zhou, and S. Zafeiriou, “UV-GAN: Adversarial Facial UV Map Completion for Pose-Invariant Face Recognition,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 7093–7102, 2018. https://doi.org/10.1109/CVPR.2018.00741

22 W. Xie, L. Shen, M. Yang, and J. Jiang, “Facial expression synthesis with direction field preservation based mesh deformation and lighting fitting based wrinkle mapping,” Multimed. Tools Appl., vol. 77, no. 6, pp. 7565–7593, 2018. https://doi.org/10.1007/s11042-017-4661-6

23 J. Thies and M. Zollhöfer, M. Nießner, L. Valgaerts, M. Stamminger and C. Theobalt, “Real-time expression transfer for facial reenactment,” ACM Trans.Graph, vol. 34, no. 6, pp. 1–14, Nov. 2015. https://doi.org/10.1145/2816795.2818056

24 P. K. Chandaliya and N. Nain, “Child Face Age Progression and Regression using Self-Attention Multi-Scale Patch GAN,” 2021 IEEE Int. Jt. Conf. Biometrics, IJCB 2021, pp. 1–8, 2021. https://doi.org/10.1109/IJCB52358.2021.9484329

25 Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., “Generative adversarial networks,” Commun. ACM, vol. 63, no. 11, pp. 139–144, 2014. https://doi.org/10.1145/3422622

26 A. Nguyen, J. Yosinski, and J. Clune, “Deep neural networks are easily fooled: High confidence predictions for unrecognizable images,” Proc. IEEE Comput. Soc. Conf. Comput.

Vis. Pattern Recognit., vol. 07-12-June, pp. 427–436, 2015. https://doi.org/10.1109/CVPR.2015.7298640
27 L. R. Calcagni, "Redes Generativas Antagónicas y sus aplicaciones," p. 72, 2020.
28 K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F. Y. Wang, "Generative adversarial networks: Introduction and outlook," IEEE/CAA J. Autom. Sin., vol. 4, no. 4, pp. 588–598, 2017. https://doi.org/10.1109/JAS.2017.7510583
29 B. Raj, "Advances in Generative Adversarial Networks (GANs)," 2019. https://medium.com/beyondminds/advances-in-generative-adversarial-networks-7bad57028032
30 A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc., pp. 1–16, 2016.
https://doi.org/10.48550/arXiv.1511.06434
31 M. Y. Liu and O. Tuzel, "Coupled generative adversarial networks," Adv. Neural Inf. Process. Syst., no. Nips, pp. 469–477, 2016. https://doi.org/10.48550/arXiv.1606.07536
32 T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 6th Int. Conf. Learn. Represent. ICLR 2018 - Conf. Track Proc., pp. 1–26, 2018. https://doi.org/10.48550/arXiv.1710.10196
33 T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 12, pp. 4217–4228, 2019. https://doi.org/10.1109/TPAMI.2020.2970919
34 T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," Cvpr, pp. 8107–8116, 2019. https://doi.org/10.48550/arXiv.1912.04958
35 A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, "CovidGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved Covid-19 Detection," IEEE Access, vol. 8, pp. 91916–91923, 2020. https://doi.org/10.1109/ACCESS.2020.2994762
36 Jabeen, S., Khan, M. U. G., Saba, T., Rehmat, A., Rehman, A., & Tariq, U. "A Realistic Image Generation of Face from Text Description Using the Fully Trained Generative Adversarial Networks," IEEE Access, vol. 9, pp. 1250–1260, 2021. https://doi.org/10.1109/ACCESS.2020.3015656
37 O. Bailo, D. Ham, and Y. M. Shin, "Red blood cell image generation for data augmentation using conditional generative adversarial networks," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., vol. 2019-June, pp. 1039–1048, 2019. https://doi.org/10.1109/CVPRW.2019.00136
38 O. Metri and H. R. Mamatha, "Medical image generation using generative adversarial networks," Gener. Advers. Networks Image-to-Image Transl., pp. 235–262, 2021. https://doi.org/10.1016/B978-0-12-823519-5.00007-5
39 A. Dash, J. Ye, G. Wang, and H. Jin, "High Resolution Solar Image Generation Using Generative Adversarial Networks," Ann. Data Sci., pp. 1–10, 2022. https://doi.org/10.1007/s40745-022-00436-2
40 D. Yasmina, R. Karima, and A. Ouahiba, "Traffic signs recognition with deep learning," Proc. 2018 Int. Conf. Appl. Smart Syst. ICASS 2018, no. June, 2019. https://doi.org/10.1109/ICASS.2018.8652024
41 J. J. Yepes-Nuñez, G. Urrútia, M. Romero-García, and S. Alonso-Fernández, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," Rev. Esp. Cardiol., vol. 74, no. 9, pp. 790–799, 2021. https://doi.org/10.1016/j.recesp.2021.06.016
42 Kaggle, "Celeba dataset." https://www.kaggle.com/datasets/jessicali9530/celeba-dataset.
43 Mejía, J., Ochoa-Zezzatti, A., Contreras-Masse, R., & Rivera, G., "Intelligent system for the visual support of caloric intake of food in inhabitants of a smart city using a deep

learning model." Applications of Hybrid Metaheuristic Algorithms for Image Processing, 441–455, 2020. https://doi.org/10.1007/978-3-030-40977-7_19