IMAGE GENERATION WITH GANS-BASED TECHNIQUES: A SURVEY

Shirin Nasr Esfahani[1]and Shahram Latifi[2]

[1]Department of Computer Science, UNLV, Las Vegas, USA
[2]Department of Electrical & Computer Eng., UNLV, Las Vegas, USA

ABSTRACT

In recent years, frameworks that employ Generative Adversarial Networks (GANs) have achieved immense results for various applications in many fields especially those related to image generation both due to their ability to create highly realistic and sharp images as well as train on huge data sets. However, successfully training GANs are notoriously difficult task in case ifhigh resolution images are required. In this article, we discuss five applicable and fascinating areas for image synthesis based on the state-of-the-art GANs techniques including Text-to-Image-Synthesis, Image-to-Image-Translation, Face Manipulation, 3D Image Synthesis and DeepMasterPrints. We provide a detailed review of current GANs-based image generation models with their advantages and disadvantages.The results of the publications in each section show the GANs based algorithmsAREgrowing fast and their constant improvement, whether in the same field or in others, will solve complicated image generation tasks in the future.

KEYWORDS

Conditional Generative Adversarial Networks (cGANs), DeepMasterPrints, Face Manipulation, Text-to-Image Synthesis, 3D GAN

1. INTRODUCTION

Image synthesis has applications in many fields like arts, graphics, and machine learning. This is done by computing the correct color value for each pixel in an image with desired resolution. Although various approaches have been proposed, image synthesis remains a challenging problem. Generative Adversarial Networks (GANs), a generative model based on game theory, have made a breakthrough in Machine Learning applications. Due to the power of the competitive training manner as well as deep networks, GANs are capable of producing realistic images, and have shown great advances in many image generations and editing models.

GANs were proposed by Goodfellowetal. (2014) [1] as a novel way to train a generative model. GANs are typically employed in a semi-supervised setting. They consist of two adversarial models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The only way G learns is through interaction with D (G has no direct access to real images). In contrast, D has access to both the synthetic samples and real samples. Unlike FVBNs (Fully Visible Belief Networks) [2] and VAE (Variational Autoencoder) [3], they do not explicitly model the probability distribution that generates the training data.In fact, G maps anoise vector z in the latent space to an image and$D$ is defined as classifying an input as a real image (close to 1) or as a fake image (close to 0). The loss function is defined as:

DOI: 10.5121/ijcsit.2019.11503

$$\min_{G} \max_{D} E_{x\in X}\,[\log D(x)] + E_{x\in X}\left[\log\left(1 - D\left(G(z)\right)\right)\right] \qquad (1)$$

Images generated by GANs are usually less blurred and more realistic than ones produced with other previousgenerative models. In an unconditioned generative model, there is no control on modes of the data being generated. Conditioning the model on additional information will direct the data generation process. This makes it possible to engage the learned generative model in different "modes" by providing it with different contextual information. Conditional Generative Adversarial Networks (cGANs) was introduced by M. Mirza and S. Osindero [4]. In cGANs, both *G* and *D* are conditioning on some extra information (*c*) that can be class labels, text or sketches.

Providing additional controls on the type of data being generated, makes cGANs popular for almost all image generating applications. The structure of GANs and cGANsare illustrated as Figure 1.

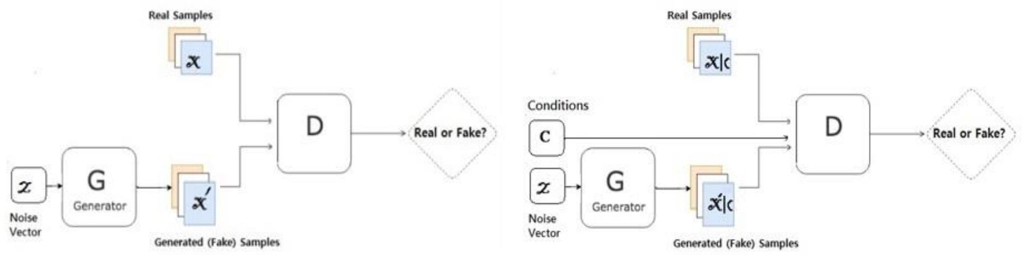

Figure 1. Structure of GANs (left) and cGANs (right)

In this survey, we discuss the ideas, contributions and drawbacks of state-of-the artmodelsin four fields of image synthesis by using GANs. So, it is not intended to be a comprehensive review of all image generation fields of GANs; many excellent papers are notdescribed here, simply because they were not relevant to our chosen subjects.This survey is structuredas follows: Sections2 and 3 provide state-of-the-art GAN-based techniques in text-to-image and image-to-image translation fields, respectively, thensection 4 and 5are related to Face Manipulation and 3D generative adversarial networks (3GANs). Finally, Section 6 isrelevant materials to DeepMasterPrints.

An earlier version of this work was presented at [5]. This paper expands on that paper by including DeepMasterPrints assection 6 (last section) and changing section 4 from Face Aging to Face Manipulation by removing some materials and adding new ones.

2. TEXT-TO-IMAGE SYNTHESIS

Synthesizing high-quality images from text descriptions, is one of the exciting and challenging problems in Computer Vision which has many applications, including photo editing and computer-aided content creation. The task of text to image generation usually means translating text in the form of single-sentence descriptions directly into prediction of image pixels. This can be done by different approaches.One of difficult problems is the distribution of images conditioned on a text description is highly multimodal. In other words, there are many plausible configurations of pixels that correctly illustrate the description. For example, more than one suitable image would be found with "this small bird has a short, pointy orange beak and white belly" in a bird dataset. S. Reed et al. [6] were the first to propose a CGAN-based model (GAN-CLS), which successfully generated realistic images (64 × 64) for birds and flowers that are described by natural language descriptions. By conditioning both generator and discriminator on side information (also used before by Mirza et al. [4]), they were able to naturally model multimodal issue since the discriminator plays as a "smart" adaptive loss function. Their approach was to train a deep convolutional generative adversarial network (DCGAN) conditioned on text features encoded by a hybrid character-level convolutional recurrent neural network. The network architecture follows the guidelines of DCGAN [7]. Both the generator *G* and the discriminator *D* performed feed-forward inference conditioned on the text feature. The architecture can be seen in Figure 2.

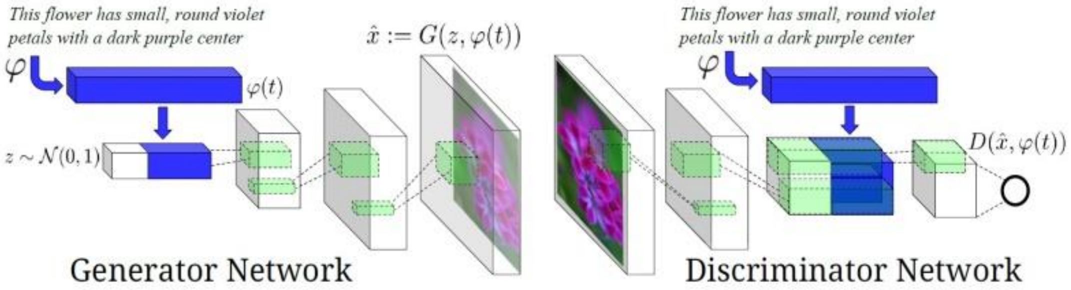
This flower has small, round violet petals with a dark purple center
φ
$\varphi(t)$
$z \sim \mathcal{N}(0,1)$
$\hat{x} := G(z, \varphi(t))$
Generator Network
This flower has small, round violet petals with a dark purple center
φ
$D(\hat{x}, \varphi(t))$
Discriminator Network

Figure 2. DCGANs architecture: Text encoding φ(t) is used by both *G* and *D*. It is projected to a lower-dimension and depth concatenated with image feature maps for further stages of convolutional processing [6]

They improved their model to generate 128 × 128 images by utilizing the locations of the content to draw (GAWWN) [8]. Their methods are not directly suitable for cross-media retrieval, but their ideas and models are valuable because they use *ten*single-sentence descriptions for each bird image. In addition, each image marked the bird location with a bounding box, or key point's coordinates for each bird's parts as well as an extra bit used in each part to show whether or not the part can be visible in the each. Both *G* and *D* are conditioned on the bounding box and the text vector (represents text description). The model has two branches for *G*: a global stage that apply on full image and local stage which only operates on the inside of bounding box. Several new approaches have been developed based on GAN-CLS. In a similar way, S. Zhu et al. [9] presented a novel approach for generating new clothing on a wearer based on textual descriptions. S. Sharma et al. [10] improved the inception scores of synthesis images with several objects by adding a dialogue describing the scene (Chat Painter). However, a large text input is not desirable for users. Z. Zhang et al.'s model [11](HDGAN) was a multi-purpose adversarial loss for generating more effective images. Furthermore, they defined a new visual-semantic similarity measure to evaluate the semantic consistency of output images. M. Cha et al. [12]extended the model by improving perceptual quality of generated images. H. Dong at al. [13] defined a new condition (the given images) in the image generation process to reduce the searching space of synthesized images. H. Zhang et al. [14] followed Reed's [6] approach to decompose the challenging problem of generating realistic high-resolution images into more manageable sub-problems by proposing StackGAN-v1 and StackGAN-v2. S. Hong [15] designed a model to generate complicated images which preserve semantic details and highly relevant to the text expression by generating a semantic layout of the objects in the image and then conditioning on the map and the caption. Y. Li et al. [16]did similar work to generate video from text. J. Chen et al. [17] designeda Language-Based Image Editing (LBIE) system to create an output image automatically by editing the input image based on the language instructions that users provide. Another text-to-image generation model (TAC-GAN) was proposed by A. Dash et al. [18]. It is designed based on Auxiliary Classifier GAN[19] but uses a text description condition instead of a class label condition. Comparisons between different text-to-image GAN-based models are given in Table 1.

Although, the application of Conditional GAN is very promising in generating realistic nature images, training GAN to synthesize high-resolution images using descriptors is a very difficult task. S. Reed et al. [6] succeeded to generate reasonable 64 × 64 images which didn't have much details. Later, [8] they were able to synthesize higher resolution (128 × 128) only with additional annotations of objects. Additionally, the training of their CGANs was unstable and highly related to the choices of hyper-parameters [20]. T. Xu et al. [21] proposed an attention-driven model (AttnGAN) to improve fine-grained detail. It uses a word-level visual-semantic that fundamentally relies on a sentence vector to generate images.

Table 1.Different text-to image models.

Model	**Input**	**Output**	**Characteristics**	**Resolution**
GAN-INT-CLS [6]	text	image	---------	64 × 64
GAWWM [8]	text + location	image	location-controllable	128 × 128
StackGAN [14]	text	image	high quality	256 × 256
TAC-GAN [18]	text	image	diversity	128 × 128
ChatPainter [10]	text + dialogue	image	high inception score	256 × 256
HDGAN [11]	text	image	high quality and resolution	512 × 512
AttnGAN [21]	text	image	high quality and the highest inception score	256 × 256
Hong et al. [15]	text	image	Second highest inception score and complicated images	128 × 128

T. Salimans et al. [22] defined Inception Scores as a metric for automatically evaluating the quality of image generative models. This metric was shown to correlate well with human judgment of image quality. In fact, inception score tries to formalize the concept of realism for a generated set of images. The inception scores of generated images on the MS COCO data set for some different models is provided in Table 2. [10]

Table 2.Inception scores of different models.

Model	**Inception Score**
GAN-INT-CLS [6]	7.88 ± 0.07
StackGAN [14]	8.45 ± 0.03
Hong et al. [15]	11.46 + 0.09
ChatPainter (non-current) [10]	9.43 ± 0.04
ChatPainter (recurrent) [10]	9.74 ± 0.02
AttnGAN [21]	25.89 ± 0.47

3. IMAGE-TO-IMAGE-TRANSLATION

Many visual techniques including in painting missing image regions (predicting missing parts in a damaged image in such a way that the improved region cannot be detected by observer), adding color to grayscale images and generate photorealistic images from sketches, involve translating one visual representation of an image into another. Application-specific algorithms are usually used to solve these problems with the same setting (map pixels to pixels). However, applying generative modeling to train the model is essential because some translating processes may have more than one correct output for each input image. Many researchers of image processing and computer graphic area have tried to design powerful translation models with supervised learning when they can have training image pairs (input, output), but producing paired images can be difficult and expensive. Moreover, these approaches are suffering from the fact that they usually formulated as per-pixel classification or regression which means that each output pixel is conditionally independent from all others in the input image.

P. Isola et al. [23] designed a general-purpose image-to-image-translation model using conditional adversarial networks. The new model (Pix2Pix), not only learned a mapping function, but also constructed a loss function to train this mapping. In particular, a high-resolution source grid is mapped to a high-resolution target grid. (The input and output differ in surface appearance, but both are renderings of the same underlying structure). In Pix2Pix model, *D* learns to classify between fake (synthesized by the generator) and real {input map, photo} tuples. *G* learns to fool *D*. *G* and *D* can access to the input map. (Figure 3)

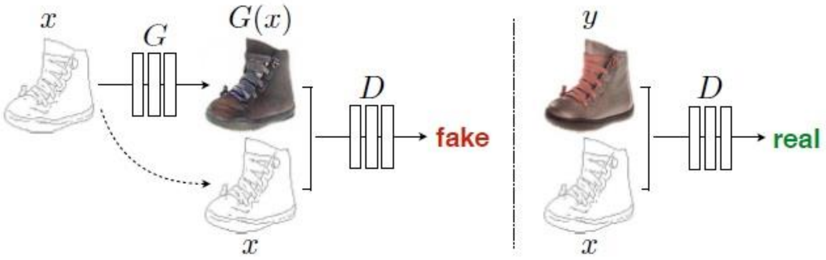

Figure 3. Training a cGANs to map edges to the photo. (Here, input map is map edges) [23]

The Pix2Pix model has some important advantages: (1) it is a general-purpose model which means it is a common framework for all automatic problems defining as the approach of translating one possible instance of an image into another(predicting pixels from pixels) by giving sufficient training data; and (2) instead of hand designing the loss function, the networks learn a loss function sensitive to data and task, to train the mapping. Finally (3), by using the fact that there is a lot of information sharing between input and output, Pix2Pix model takes advantages of them more directly by skipping connections between corresponding layers in the encoder following the general shape of a “U-Net” to create much higher quality results. The main drawback of Pix2Pix model is that it requires significant number of labeled image pairs, which is generally not available in domain adaptation problems. Later, they improved their method and designed a new model (CycleGAN) to overcome to this issue by translating an image from a source domain to a target domain in the absence of paired examples using combination of adversarial and cycle-consistent losses. [23].A comparison against other baselines (CoGAN) [25], BiGAN [26]/ALI [27], SimGAN [10] and CycleGAN for mapping aerial photos can be seen in Figure 4.

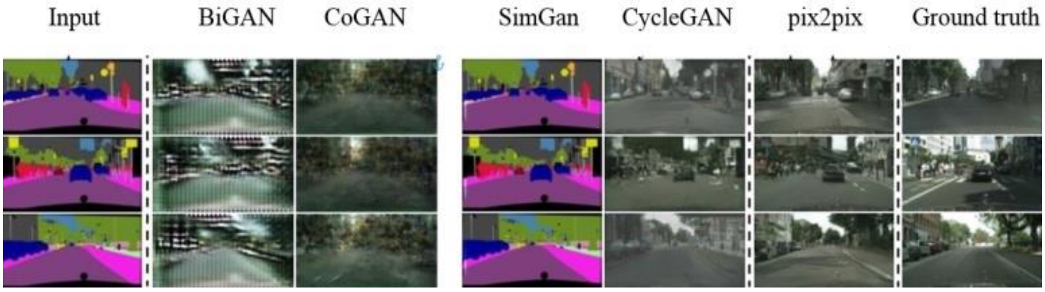

Figure 4. Different methods for mapping labels ↔ photo on Cityscapes images. From left to right: input, BiGAN/ALI, CoGAN, SimGAN, CycleGAN, Pix2Pix trained on paired data, and ground truth [24]

To measure the performance of photo↔ *l*abels, the standard metrics of the Cityscapes benchmark is used that includes per-pixel accuracy, per-class accuracy, and mean class Intersection-Over-Union (Class IOU) [28]. Comparison results are provided in Table 3 [11].

Table 3.Classification performance for different models on images of the Cityscapes dataset.

Model	**Per-pixel Accuracy**	**Per-class Accuracy**	**Class IOU**
CoGAN [25]	0.45	image	0.08
BiGAN/ALI [26, 27]	0.41	image	0.07
SimGAN [10]	0.47	image	0.07
CycleGAN [24]	0.58	image	0.16
Pix2Pix [23]	0.85	image	0.32

Later, Q. Chen and V. Koltun [29] suggest that because of the training instability and optimization issues of CGANs, it is hard and prone to failure to generate images with high resolution. Instead, they used a direct regression objective based on a perceptual loss and produced the first model that can generate 2048 × 1024 images. However, their results often don't have fine details and realistic textures [30].Following the Pixt2Pix model's architecture, Lample et al. [31]designed *Fader Networks*, with *G* and *D*competing in the latent space to generates realistic images of high resolution without needing to apply a GAN to the decoder output. Their model provided a new direction towards robust adversarial feature learning. D. Michelsanti and Z.-H Tan [32] used Pix2Pix to create a new framework for speech enhancement. Their model learned a mapping between noisy and clean speech spectrograms as well as to learn a loss function for training the mapping.

4. FACE MANIPULATION

Face manipulation has been an attractive field in the media and entertainment industry for well over two decades. Generally, face manipulation includes modifying facial attributes such as age, hair and facial hair, eyes color, skin texture or adding glasses, smile, frown or swapping/morphing two faces. It is divided in two distinct groups: Face sample manipulation using original sample, and synthetic face image generation. The first group needs to have original face images manipulated without losing important attributes like identity, while algorithms in the second group synthesize face images using semantic domains.[33] Manipulating face's attributes is more challenging than other image generation problems due to the fact that some image's features have to be modified while others need to remain unchanged.[34]

Since the invention of GANs, many GAN-based methods have been designed for manipulating face images. Compared to traditional algorithms, GANs are able to produce more realistic faces, while most of them cannot prevent losing person’s identity during the transformation process. Age-cGAN by G. Antipov et al. was the first automatic face aging approach to generating realistic results with high quality.[35]The architecture of Age-cGAN consisted of cGAN networks combined with an encoder which mapped an input image to a latent vector. An optimal latent vector was computed based on the input face as well as the age number as additional information.The generator produced a new image mapping to the latent vector conditioned on age number. The output face was reconstructed in the final step(Figure 5).Age-cGAN had some important drawbacks. There was not any mechanism to preserve original face’s identity during modifying face’s attributes. Moreover, by using L-BFGS-B optimization algorithm [36] for each image, the process was time consuming.[37] Age-cGAN+LMA, a modified model of Age-cGAN, was introduced later. They used a Local Manifold Adaptation approach [38] to improve the accuracy of cross age verification by using age normalization. A comparison between two models is provided in Figure 6.

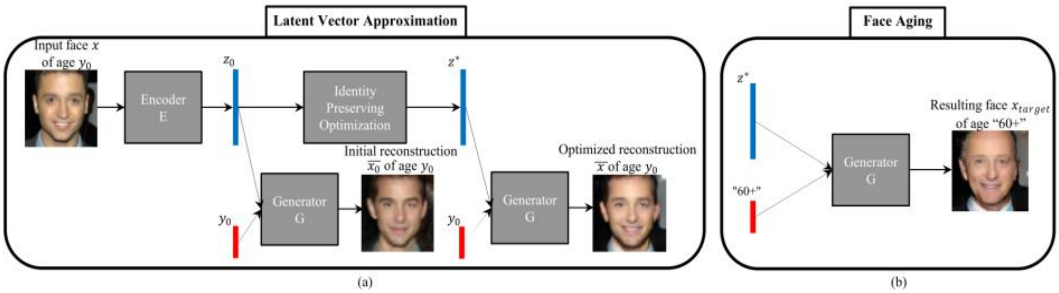
Latent Vector Approximation
Input face x of age y_0
Encoder E
z_0
Identity Preserving Optimization
z^*
Initial reconstruction $\overline{x_0}$ of age y_0
Generator G
y_0
y_0
Generator G
Optimized reconstruction $\overline{x}$ of age y_0
(a)
Face Aging
z^*
"60+"
Generator G
Resulting face x_{target} of age “60+”
(b)

Figure 5. (a): Approximation of the latent vector to reconstruct the input image, (b): Switching the age condition at the input of the generator to perform face aging[35]

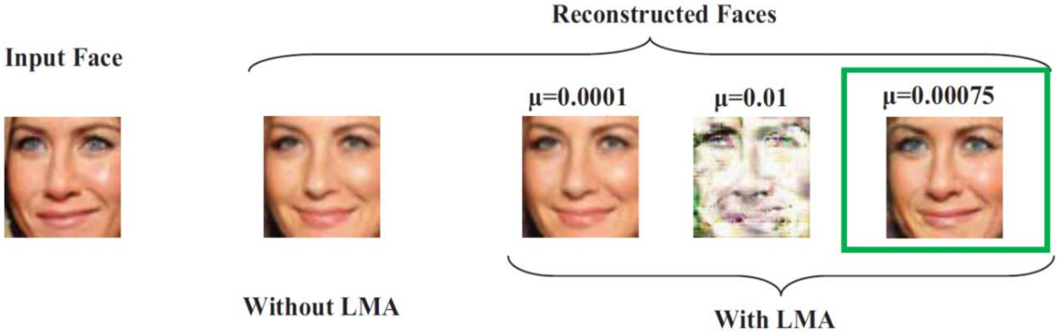

Figure 6. Face reconstruction with and without Local Manifold Adaptation (LMA) For LMA-enhanced reconstructions, the impact of the learning rate μ is illustrated [38]

The first solution for keeping facial identity was provided by Z. Zhang et al. [39].The face images were fed into a face recognition network before and after manipulation and their feature distances were penalized. The model’s problem was, it changed some parts of image (beyond identity) like background. In addition to this drawback, the model’s performance was unreasonable for a small amount of training data. Some other models like Fader network [31], an encoder-decoder framework, had ability to change various attributes of face such as age, agender, eye and mouth opening. (Figure 7) IcGAN, by Perarnau et al.[40], was a combination of a cGAN with an encoder that learned the inverse mapping of cGAN which led to regeneration of real images with deterministic complex modifications. For reconstruction or modification purposes, images were conditioned on arbitrary attributes, while the conditional vector was changed, and the latent vector was preserved.DIAT (Deep Identity- Aware Transfer of Facial Attributes) was another model following GAN networks which was designed by Li et al. [41].It consisted of an attribute transform network as well as a mask network to synthesize realistic faces with reference attributes.In addition to the mask network which avoided modifying irrelevant regions, there was a denoising network to suppress the artifacts in the transferred result.It also used an adversarial loss to learn the mapping from face images in different domains. Choi et al introduced StarGAN [42], a scalable system using GANs, which did multiple-domain image to image translations, such as skin, age and emotions (angry, happy, fearful) using only a single model.StarGAN was unable to preserve details. To overcome this problem, Chen et al. created a model, TDB-GAN, to generate an image with fine details using texture. [43]

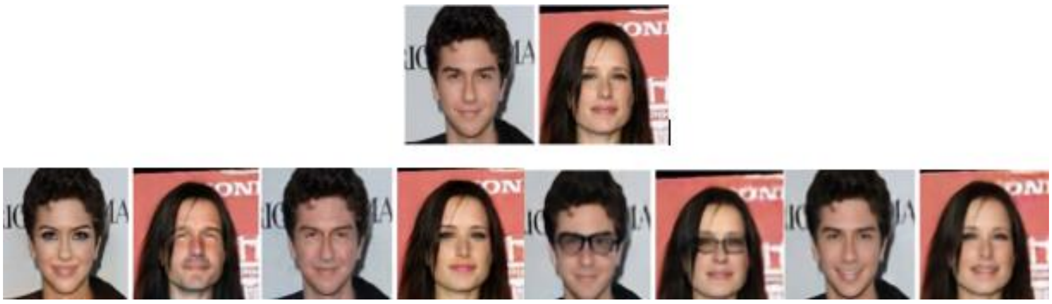

Figure 7. Swapping the attributes of faces. Top: original images, bottom from left to right: gender, age, glasses and mouth opening [31]

Figures 8. and 9. show the generated images by different models on CelebA (Celebrity Faces Attributes) and RaFD (Radcoud Faces) databases. To prevent most details (in regions of irrelevant attributes) from changing, a different framework was introduced by Shen & Liu [44] that used residual image learning (learn from the difference original image and target one) as well as dual learning (learn from each other). There were two image transformation networks (two generators) to manipulate attributes and do its dual function as well as a discriminator to distinguish fake images from real ones.

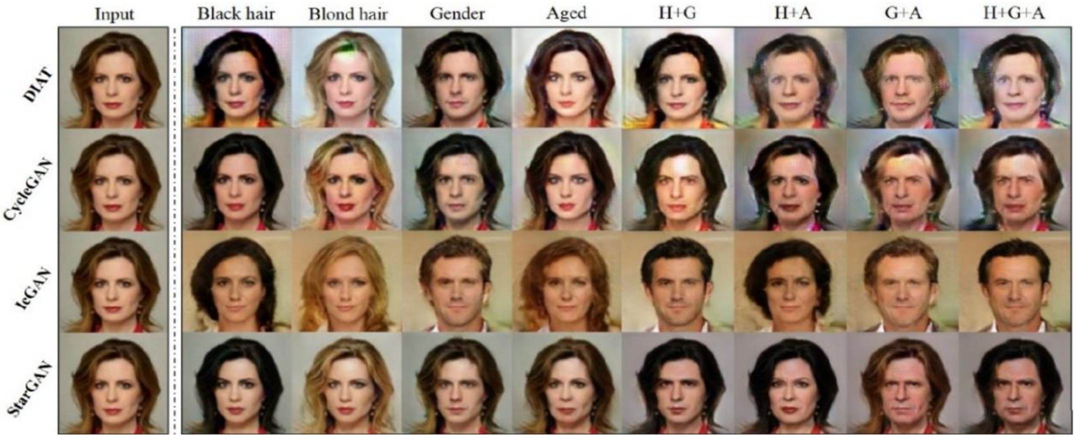

Figure 8. Facial attribute transfer on CelebA dataset. The first four leftmost columns (excluding the input column) are the results of single attribute transfer and others show the multi-attribute one. H: Hair color, G: Gender, A: Aged [42]

In order to have more lifelike facial details in synthesized faces, Yang et [45] al. designed pyramidal GANs at multiple scales, to create simulation of the aging effects with fine precision. Their approach was able to generate diverse faces with different pose and makeup, etc.There are also various open-source, commercial software packages for face image and video editing. Among them, is a popular and new emerging product, FaceApp, which is a mobile application developed by a Russian company, Wireless Lab, which uses GANs networks to generate highly realistic transformations of faces in photographs. it has a database with a huge number of faces which extracts features from faces and applies some changes to render the new face with different look, while the distinctive features which make the identify unique, are remained unchanged. [46]The app can transform a face to make it smile, look younger, look older, or change gender. (Figure 10)

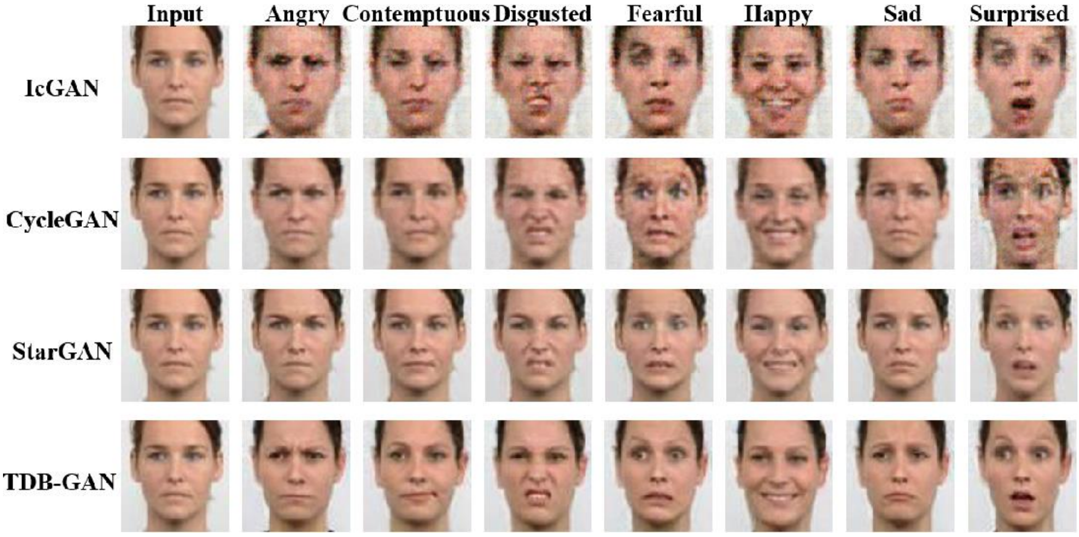

Figure 9. Facial expression synthesis results on RaFD dataset [43]

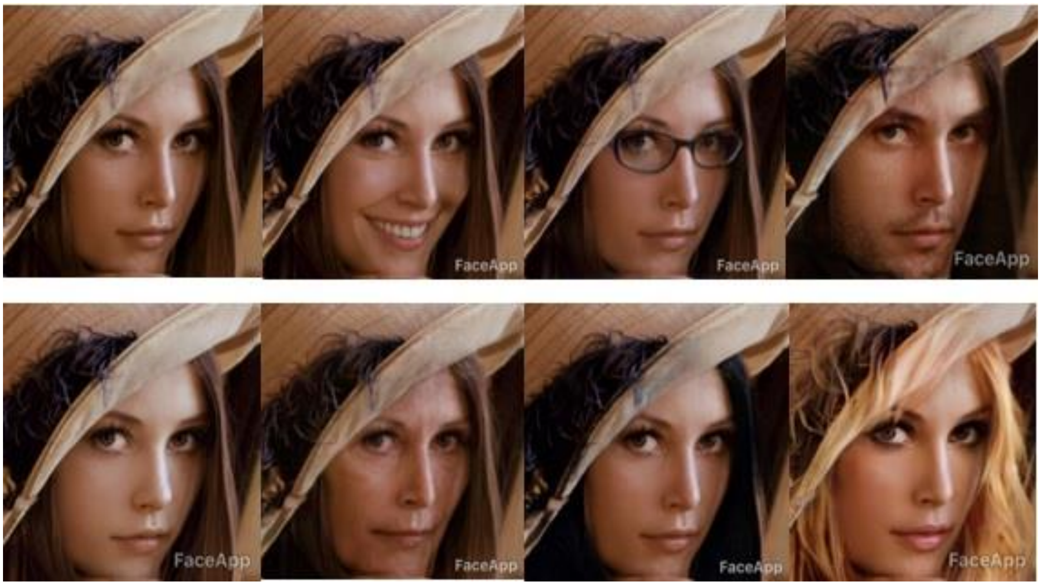

Figure 10. FaceApp results on Lena image. Top, from left to right: original image, smile, glasses and gender, bottom from left to right: young, old, dark hair and blond hair + make up

5. 3D IMAGE SYNTHESIS

3D object reconstruction of 2D images has always been a challenging task that try to define any object's 3D profile, as well as the 3D coordinate of every pixel. It is generally a scientific problem which has a wide variety of applications such as Computer Aided Geometric Design (CAGD), Computer Graphics, Computer Animation, Computer Vision, medical imaging etc. Researchers have done impressive works on 3D object synthesis, mostly based on meshes or skeletons. Using parts from objects in existing CAD model libraries, they have succeeded to generate new objects. Although the output objects look realistic, but they are not conceptually novel. J. Wu et al. [47] were the first that introduced 3D generative adversarial networks (3D GANs). Their state-of-the-art framework was proposed to model volumetric objects from a probabilistic domain (usually Gaussian or uniform distribution) by using recent progresses in volumetric convolutional networks and generative adversarial networks. They generated novel objects such as chairs, table and cars. Besides, they proposed a model which mapped 2D images to images having 3D versions of objects. 3DGAN is an all-convolutional neural network, showing in Figure11.

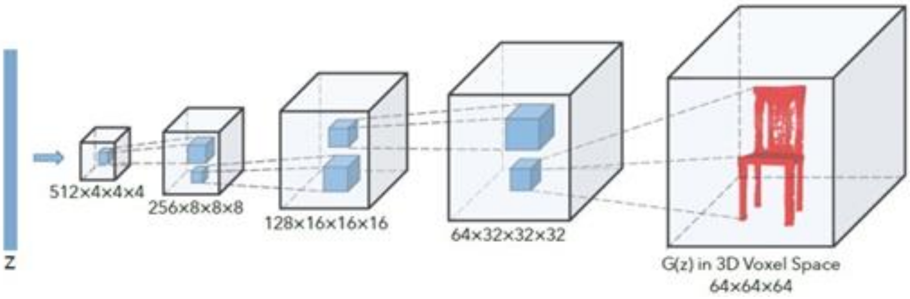

Figure 11. 3DGAN generator. The Discriminator mostly mirrors the generator[47]

The *G*has five volumetric fully convolutional layers with kernel sizes of 4 × 4 × 4 and strides 2. between the layers, batch normalization and ReLU layers have been added with a Sigmoid layer at the end. Instead of ReLU layers, The *D* uses Leaky ReLU while it is basically like the *G*. Neitherpooling nor linear layers are used in the network. The 3DGAN model has some important achieving results comparing with previous 3D models: (1) It samples objects without using a reference image or CAD model; (2) It has provided a powerful 3D shape descriptor that can be learned without supervision that makes it widely applicable in many 3D object recognition algorithms; (3) Having comparable performance against recent surprised methods, and outperforms other unsupervised methods by a large margin; (4) They have the capability to apply for different purposes including 3D object classification and 3D object recognition. However, there are significant limitations in using 3DGANs: (1) Their using memory and the computational costs grow cubically as the voxel resolution increases which make them unusablein generating high resolution 3D image as well as in interactive 3D modelling (2) They are largely restricted to partial (single) view reconstruction and rendered images. There is a noticeable drop in performance when applied to natural (non-rendered) images. Later, they proposed a new 3D model called Marr Net by improving the previous model(3DGANs) [48]. They enhanced the model's performance by using 2.5D sketches for single image 3D shape reconstruction. Besides, in order to have consistency between 3D shape and 2.5D sketches, they defined differentiable loss functions, so Marr Net is an end-to-end fine-tuned on real images without annotations. At first, it returns objects from an RGB image to their normal, depth, and silhouette image, then from the 2.5D sketches, regresses the 3D shape. It also applies an encoding-decoding nets as well as reprojection consistency loss function to ensure the estimated 3D shape aligns with the 2.5D sketches precisely. The whole architecture can be trained end-to-end. (Figure12)

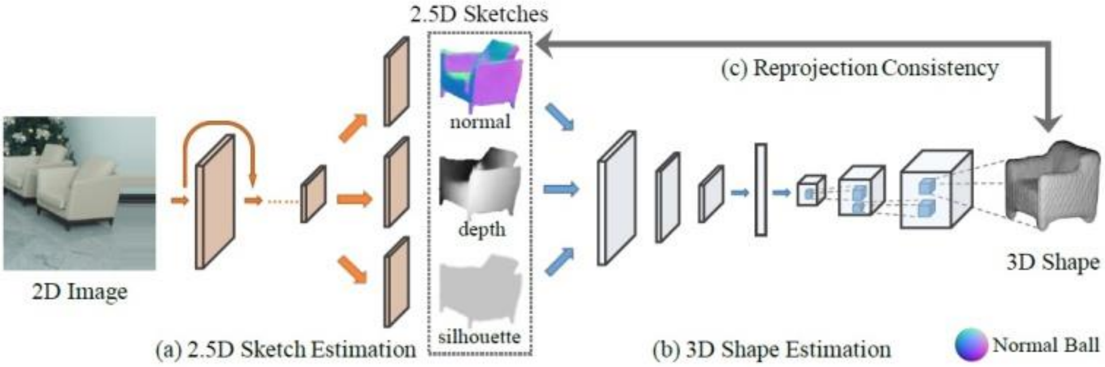

Figure 12. Components of MarrNet: (a) 2.5D sketch estimation, (b) 3D shape estimation, and (c) Loss function for reprojection consistency [48]

There are other 3D models that have been designed based on the 3DGAN architecture. Combining a 3D Encoder-Decoder GAN(3D-ED-GAN) with a Long term Recurrent Convolutional Network (LRCN), W. Wang et al. [49] proposed a hybrid framework. The model's purpose is in painting corrupted 3D objects and completing high-resolution 3D volumetric data. It gets significant advantage of completing complex 3D scene with higher resolution such as indoor area, since it is easily fit into GPU memory. E. J. Smith and D. Meger [50]improved 3DGAN and introduced a new model called 3D-IWGAN (Improved Wasserstein Generative Adversarial Network) to reconstruct 3D shape from 2D images and perform shape completion from occluded 2.5D range scans. Leaving the object of interest still and rotating the camera around it, they were able to extract partial 2.5D views, instead of enforcing it to be similar to a known distribution. P. Achlioptas et al. [51]explored AAE variant by using a specially designed encoder network for learning a compressed representation of point clouds before training GAN on the latent space. However, their decoder is restricted to be MLP that generate s*m* pre-defined and fixed number of points. On the other hand, the output of decoder is 3*m*(fixed)for 3D point clouds, while the output of the proposed G_x is only3 dimensional and it can generate arbitrarily many points by sampling different random noise *z* as input. The new model had the ability to jointly estimates intrinsic images and full 3D shape from a colour image and generates reasonable results on standard datasets [52]. It was able to recover more details compared to 3D GAN (Figure 13). A comparison between different 3D models can be shown in Table 4.

Table 4. Classification results on ModelNet dataset. [49]

Model	**ModelNet40**	**ModelNet10**
3DGAN [47]	83.3%	91.0%
3D-ED-GAN [49]	87.3%	92.6%
VoxNet [53]	92.0%	83.0%
DeepPano [54]	88.66%	82.54%
VRN [55]	91.0%	93.6%

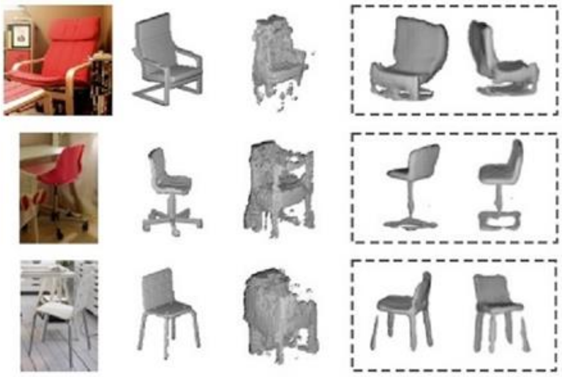

Figure 13. 3D construction of chairs on IKEA dataset. From left to right: input, ground truth, 3D estimation by 3DGAN and two view of MarrNet [48]

6. DEEP MASTER PRINTS

Fingerprinting is the oldest biometric trait that has been most widely used for human identification for over a century and has two important properties of persistence and uniqueness. In recent years, there have been various applications introduced by Apple and Samsung that use fingerprint for user's biometric identification in smart phones or other small electronic devices.

The main disadvantage of these applications is the tiny sensors they employ to capture fingerprints which are unable to obtain the whole image of user's fingerprint. Partial fingerprint-based applications may cause a threat for user's security. The probability that a fake partial fingerprint matches with the fingerprint data of user, are higher.(Figure 14) Furthermore, most of the features in fingerprints are very common and nearly identical for most people.

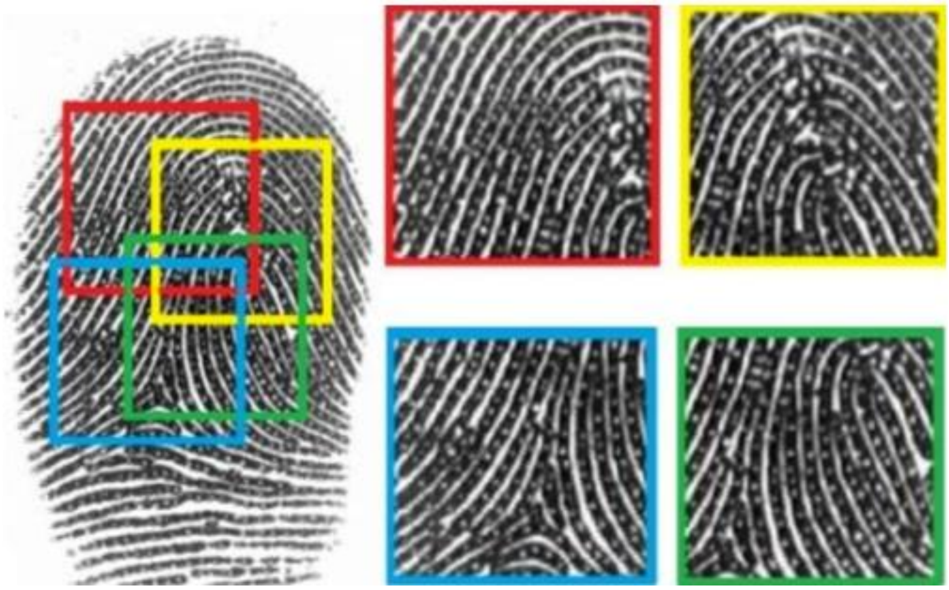

Figure 14. Left: a set of partial fingerprints, right: extracted from the full fingerprint [56]

Roy et al. observed this fact and introduced the concept of "MasterPrint" which are real or synthetic fingerprints (full or partial print) that effectively matches one or more of the stored templates for a significant number of users thereby undermining the security afforded by fingerprint systems. [56]. In their work, specific approaches were presented to generate Master Print at the feature level, then the vulnerability of fingerprint systems that use partial prints for authentication was carefully analysed. Later, Bontrager et al [57] generated new artificial fingerprints, Deep Master Prints, as a complete image-level Master Prints for hacking smart phones fingerprint applications. According to their research, the attack accuracy of Deep Master Prints is much superior to that of previous approaches. In order to fool a fingerprint matcher, they proposed a new approach, Latent Variable Evolution, to generate fake fingerprint images (Deep Master Prints). At first, a Wasserstein GAN (WGAN)[58] network was trained by using real images a fingerprint dataset. Then, Covariance Matrix Adaptation Evolution Strategy (CMA-ES) was used as the Stochastic search method to explore the latent input variables of the generator for (fake) image that maximize the number of successful matches with it (Figure 15). (A fingerprint recognizer was used to make an assessment). Their method has the potential of using broadly in applications relating to fingerprint security or fingerprint synthesis. Figure 16. shows WGAN generator's results after training process. In their research, both types of fingerprints images were used (inked-and-rolled impression samples scanned images and ones created by a capacitive sensor).Moreover, they generated different Deep Master Print based on the security level which they were optimized for. Images with FMR = 0.01% were at the highest level of security while those with FMR=1% belonged to the lowest level. Table 5 provides the results of false subject matches (They are samples in dataset that successfully matched against fake fingerprint images).

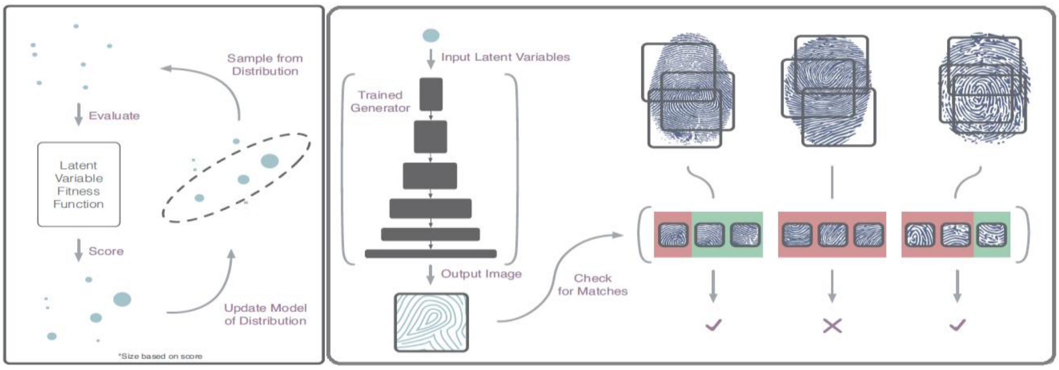

Figure 15. Latent Variable Evolution with a trained network, left: a high-level overview of CMA-ES, Right: how the latent variables are evaluated [57]

Table 5:Successful matches on the on the rolled and capacitive dataset. [57]

	Rolled DeepMasterPrint Matches			Capacitive DeepMasterPrint Matches		
	0.01% FMR	0. 1% FMR	1% FMR	0.01% FMR	0. 1% FMR	1% FMR
Training	5.00%	13.89%	67.50%	6.94%	29.44%	89.44%
Testing	0.28%	8.61%	78.06%	1.11%	22.50%	76.67%

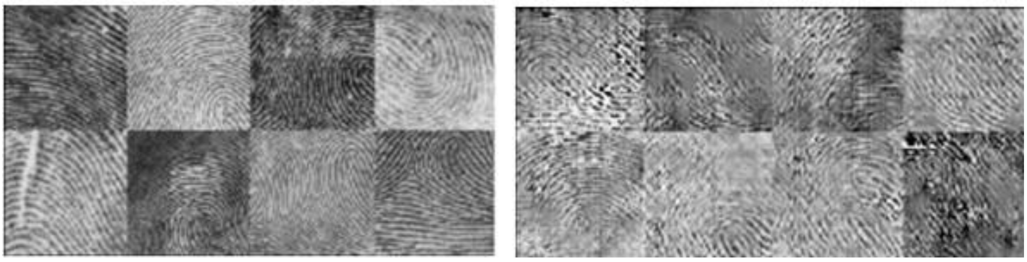

Figure 16. Left: Real Samples, right: generated samples for the NIST dataset [57]

7. CONCLUSION

Although digital image synthesis is as old as image processing and machine vision, automatic image generation techniques are still discussed, and new methods are introduced every day.In this paper, we presented an overview of state-of-art approaches in five common fields of GANs-based image generation including text-to-image synthesis, image-to-image translation, face aging,3D image generation and DeepMasterPrints.We have demonstrated pioneering frameworks in each part following with their advantages and disadvantages. In mentioned fields,text-to-image synthesis and image-to-image translation, older than others, have been the fields with most different proposed models and still have potential for improvement and expansion improved.3D image synthesis approaches face several limitations even despite the advancements.Face Manipulation has been the most attractive one due to its promising results in entertainment and media industry. While the idea of DeepMasterPrints is a novel and under-explored that will be essential, even crucialin many security domains in the future.

REFERENCES

[1] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014) “Generative adversarial nets”,*Advances in Neural Information Processing Systems 27 (NIPS 2014)*,Montreal, Canada.

[2] Frey, B. J. (1998) “Graphical models for machine learning and digital communication”, MIT press.

[3] Doersch, C. (2016) “Tutorial on variational autoencoders”, arXiv preprint arXiv:1606.05908.

[4] M. Mirza & S. Osindero (2014) “Conditional generative adversarial nets”, arXiv:1411.1784v1.

[5] Sh.Nasr Esfahani, &Sh. Latifi (2019) “A Survey of State-of-the-Art GAN-based Approaches to ImageSynthesis”, *9th International Conference on Computer Science, Engineering and Applications (CCSEA 2019),* Toronto, Canada, pp. 63-76.

[6] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele & H. Lee (2016) “Generative adversarial text to image synthesis”, *International Conference on Machine Learning,* New York, USA, pp. 1060-1069.

[7] A. Radford, L. Metz & S. Chintala (2016) “Unsupervised representation learning with deep convolutional generative adversarial networks”, 4th*International Conference of Learning Representations (ICLR 2016),* San Juan, Puerto Rico.

[8] S. Reed, Z. Akata, S. Mohan, S. Tenka, B. Schiele & H. Lee (2016) “Learning what and where to draw”, *Advances in Neural Information Processing Systems*, pp. 217–225.

[9] S. Zhu, S. Fidler, R. Urtasun, D. Lin & C. L. Chen (2017) “Be your own prada: Fashion synthesis with structural coherence”, *International Conference on Computer Vision (ICCV 2017),* Venice, Italy,pp. 1680-1688.

[10] S. Sharma, D. Suhubdy, V. Michalski, S. E. Kahou & Y. Bengio (2018) “ChatPainter: Improving text to image generation using dialogue*”, 6th International Conference on Learning Representations (ICLR 2018 Workshop),* Vancouver, Canada.

[11] Z. Zhang, Y. Xie & L. Yang (2018) “Photographic text-to-image synthesis with a hierarchically-nested adversarial network”, *Conference on Computer Vision and PatternRecognition (CVPR 2018),* Salt Lake City, USA,pp. 6199-6208.

[12] M. Cha, Y. Gwon & H. T. Kung (2017) "Adversarial nets with perceptual losses for text-to-image synthesis", *International Workshop on Machine Learning for Signal Processing (MLSP 2017),* Tokyo, Japan,pp. 1- 6.

[13] H. Dong, S. Yu, C. Wu & Y. Guo (2017) "Semantic image synthesis via adversarial learning", *International Conference on Computer Vision (ICCV 2017*), Venice, Italy,pp. 5706-5714.

[14] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. Metaxas (2017) "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks", *International Conference on Computer Vision (ICCV 2017),* Venice, Italy,pp. 5907-5915.

[15] S. Hong, D. Yang, J. Choi & H. Lee (2018) "Inferring semantic layout for hierarchical text-to-image synthesis", *Conference on Computer Vision and PatternRecognition (CVPR 2018),* Salt Lake City, USA,pp. 7986-7994.

[16] Y. Li, M. R. Min, Di. Shen, D. Carlson, and L. Carin (2018) "Video generation from text", *14th Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE 2018),* Edmonton, Canada.

[17] J. Chen, Y. Shen, J. Gao, J. Liu & X. Liu (2017) "Language-based image editing with recurrent attentive models", *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018),* Salt Lake City, USA, pp. 8721-8729.

[18] A. Dash, J. C. B. Gamboa, S. Ahmed, M. Liwicki & M. Z. Afzal (2017) "TAC-GAN-Text conditioned auxiliary classifier", arXiv preprint arXiv: 1703.06412, 2017.

[19] A. Odena, C. Olah & J. Shlens (2017) "Conditional image synthesis with auxiliary classifier GANs," *Proceeding of 34th International Conference on Machine Learning (ICML 2017),* Sydney, Australia.

[20] H. Zhang, I. Goodfellow, D. Metaxas & A. Odena (2018) "Self-attention, generative adversarial networks", arXiv preprint arXiv:1805.08318, 2018.

[21] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z.Gan, X. Huang & X. He (2018) "AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks", *The IEEE Conference on Computer Vision and PatternRecognition (CVPR 2018),* Salt Lake City, USA,pp. 1316-1324.

[22] T. Salimans, I. J. Goodfellow, W. Zaremba, V. Cheung, A. Radford & X. Chen (2016) "Improved techniques for training GANs", *Advances in Neural Information Processing Systems 29 (NIPS 2016),* Barcelona, Spain.

[23] P. Isola, J.-Y. Zhu, T. Park & A. A. Efros (2017) "Image-to-image translation with conditional adversarial networks",*The IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017),* Honolulu, Hawai, USA, pp. 1125-1134.

[24] J.-Y. Zhu, T. Park, P. Isola & A. A. Efros (2017) "Unpaired Image-to-Image Translation using Cycle-Consistent", *The IEEE International Conference on Computer Vision (ICCV2017*), Venice, Italy, pp. 2223-2232.

[25] M.-Y. Liu & O. Tuzel (2016) "Coupled generative adversarial networks", *2016 Conference on Neural Information Processing Systems (NIPS 2016)*, Barcelona, Spain, pp. 469–477.

[26] J. Donahue, P. Kr¨ahenb¨uhl & T. Darrell (2016) "Adversarial feature learning" ,4th*International Conference on Learning Representations (ICLR 2016),*San Juan, Puerto Rico.

[27] V. Dumoulin, I. Belghazi, B. Poole, A. Lamb, M. Arjovsky, O. Mastropietro & A. Courville (2017) "Adversarially learned inference", *5th International Conference on Learning Representations(ICLR 2017)*, Toulon, France.

[28] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, & B. Schiele (2016) “The cityscapes dataset for semantic urban scene understanding”, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016),* Las Vegas, USA, pp. 3213-3223.

[29] Q. Chen & V. Koltun (2017) “Photographic image synthesis with cascaded refinement networks”, *IEEE International Conference on Computer Vision (ICCV 2107),* Venice, Italy, pp. 1520–1529.

[30] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz & B. Catanzaro (2018) “High-resolution image synthesis and semantic manipulation with conditional GANs”, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018),* Salt Lake City, USA, pp. 8798-8807.

[31] G. Lample, N. Zeghidour, N. Usunier, A. Bordes, L. Denoyer & M. Ranzato (2017) “Fader networks: Manipulating images by sliding attributes”, *Advances in Neural Information Processing Systems 30 (NIPS 2017),* Long Beach, USA.

[32] D. Michelsanti & Z.-H. Tan (2017) “Conditional generative adversarial networks for speech enhancement and noise-robust speaker verification”, *Proceeding of Interspeech,* pp. 2008–2012.

[33] Z. Akhtar, D. Dasgupta, B. Baerjee (2019) “Face Authenticity: An Overview of Face Manipulation Generation, Detection and Recognition”, *International Conference on Communication and Information Processing (ICCIP-2019*)”, Pune, India.

[34] R. Sun, C. Huang, J. Shi, L. Ma (2018) “Mask-aware photorealistic face attribute manipulation”, arXiv preprint arXiv:1804.08882, 2018.

[35] .Antipov, M. Baccouche & J.-L. Dugelay (2017)“Face aging with conditional generative adversarial networks*”, IEEE International Conference on Image Processing (ICIP 2017),* pp.2089 – 2093.

[36] R. H. Byrd, P. Lu, J. Nocedal & C. Zhu (1995) “A limited memory algorithm for bound constrained optimization”, *SIAM Journal on Scientific Computing,* vol. 16, no. 5, pp. 1190–1208, 1995.

[37] Z. Wang, X. Tang, W. Luo & S. Gao (2018) “Face aging with identity preserved conditional generative adversarial networks”, *Proceeding IEEE Conference Computer Vision and Pattern Recognition, CVPR 2018),* Salt Lake City, USA, pp. 7939–7947.

[38] G. Antipov, M. Baccouche & J.-L. Dugelay (2017)” Boosting cross-age face verification via generative age normalization”, *International Joint Conference on Biometrics* (*IJCB 2017*), Denver, USA, pp. 17.

[39] Z. Zhang, Y. Song & H. Qi (2017) “Age progression/regression by conditional adversarial auto encoder”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017),* Honolulu, USA, pp. 4352 – 4360.

[40] G. Perarnau, J. van de Weijer, B. Raducanu and J.M. Alvarez (2016) “Invertible conditional gans for image editing”, arXiv preprint arXiv:1611.06355, 2016.

[41] M. Li, W. Zuo and D. Zhang (2016) “Deep identity-aware transfer of facial attributes”, arXiv preprint arXiv:1610.05586, 2016.

[42] Y. Choi, M. Choi, and M. Kim (2018) “Stargan: Unified generative adversarial networks for multi-domain image-to-image translation”,*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2018),* Salt Lake City, USA, pp. 8789–8797.

[43] W. Chen, X. Xie, X. Jia and L. Shen (2018) “Texture deformation based generative adversarial networks for face editing”, arXiv preprint arXiv:1812.09832, 2018.

[44] W. Shen and R. Liu (2017) “Learning residual images for face attribute manipulation”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017),* Honolulu, Hawaii, USA, pp. 4030–4038.

[45] H. Yang, D. Huang, Y. Wang and A. K. Jain (2017) “Learning face age progression: A pyramid architecture of gans”, arXiv preprint arXiv:1711.10352, 2017.

[46] H. Arian (2019) “FaceApp: How Neural Networks can do Wonders”, *Noteworthy - The Journal Blog*, https://blog.usejournal.com/tagged/faceapp.

[47] J. Wu, C. Zhang, T. Xue, W. T. Freeman & J. B. Tenenbaum (2016) “Learning a probabilistic of object shapes via 3d generative-adversarial modeling”, In *Advances in Neural Information Processing Systems 29 (NIPS 2016),* Barcelona, Spain.

[48] J. Wu, Y. Wang, T. Xue, X. Sun, B. Freeman & J. Tenenbaum (2017) “Marrnet: 3d shape reconstruction via 2.5 d sketches”, *Advances in Neural Information Processing Systems,*Long Beach, USA, pp. 540–550.

[49] W. Wang, Q. Huang, S. You, C. Yang & U. Neumann (2017) “Shape inpainting using 3d generative adversarial network and recurrent convolutional networks*”, The IEEE international Conference on Computer Vision (ICCV 2017),*Venice, Italy, pp. 2298-2306.

[50] E. J. Smith & D. Meger (2017) “Improved adversarial systems for 3d object generation and reconstruction”, *first Annual Conference on Robot Learning,*Mountain View, USA, pp. 87–96.

[51] P. Achlioptas, O. Diamanti, I. Mitliagkas & L. Guibas (2018) “Learning representations and generative models for 3d point clouds”, *6th International Conference on Learning Representations,*Vancouver, Canada.

[52] X. Sun, J. Wu, X. Zhang, Z. Zhang, C. Zhang, T. Xue, J. B. Tenenbaum & W. T. Freeman (2018) “Pix3d: Dataset and methods for single-image 3d shape modeling”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018),* Salt Lake City, USA, pp. 2974-2983.

[53] D. Maturana &S. Scherer (2015) “VoxNet: A 3D Convolutional Neural Network for real-time object recognition”, *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* Hamburg, Germany, pp. 922 – 928.

[54] B. Shi, S. Bai, Z. Zhou & X. Bai (2015) “DeepPano: Deep Panoramic Representation for 3-D Shape Recognition”, *IEEE Signal Processing Letters* ,vol. 22(12) , pp. 2339 – 2343.

[55] A. Brock, T. Lim, J. Ritchie & N. Weston (2016) “Generative and discriminative voxel modeling withconvolutional neural networks”, arXiv:1608.04236.

[56] A. Roy, N. Memon, and A. Ross (2017) “MasterPrint: Exploring the vulnerability of partial fingerprint-based authentication systems”,*IEEE Trans. Inf. Forensics Security,* vol. 12, no. 9, pp. 2013–2025.

[57] P. Bontrager, A. Roy, J. Togelius, N. Memon and A. Ross (2018)“DeepMasterPrints: Generating MasterPrintsforDictionaryAttacks via Latent Variable Evolution”, https://arxiv.org/pdf/1705.07386.pdf

[58] M. Arjovsky, S.Chintala, and L. Bottou(2017), “ Wasserstein generative adversarial networks”, *InProceedingsof the 34th International Conference on Machine Learning(ICML2017),* Sydney, Australia,Vol. 70, pp. 214–223.

AUTHORS

Shirin Nasr Esfahani received her M.S. degree in computer science – scientific computation from Sharif University of technology, Tehran- Iran. She is currently a Ph.D. candidate in computer science, University of Nevada, Las Vegas (UNLV). Her fields of interest include, hyper spectral image processing, neural networks, deep learning and data mining.

Shahram Latifi received the Master of Science and the PhD degrees both in Electrical and Computer Engineering from Louisiana State University, Baton Rouge, in 1986 and 1989, respectively. He is currently a Professor of Electrical Engineering at the University of Nevada, Las Vegas.

.