ITM Web of Conferences **70**, 04009 (2025) https://doi.org/10.1051/itmconf/20257004009
DAI 2024

Advances in Image Generation Technology: Exploring GANs and MirrorGANs

Lewuqiong Shi

The Faculty of Arts and Science, Queen's University, ON K7L 3N6 Kingston, Canada

Abstract. This paper is an in-depth study by delving into the latest in image generation technology, where thesis is focusing on the Generative Adversarial Networks (GANs) and MirrorGANs possibilities. Image Generation is the backbone of visual computing, mostly utilized in intelligent designs. It is for this reason that this research aims at unravelling the theoretical basis and consolidated practices of GANs when it comes to generating both high-quality and semantically consistent imagery. The study will investigate the whole of the image generation process, starting from data preprocessing to the use of GANs to generate images from textual descriptions. The work discussed the relevance as well as the limitations of these technologies from the artistic point of view, medical imaging, and virtual reality. The article concludes that the paper sketches the data and experiments that show that the realism and richness in picture quality are accentuated when GANs and MirrorGANs are incorporated. This suggests the scope of image-generation technology to enhance human-machine collaboration and allow for innovating in smart tech. Further studies will be geared to enhancing these methods and consequently drawing humanity and machines closer, which in turn will fuel the ongoing progress in this fast-paced sphere.

1 Introduction

Image generating as a significant part of the visualizing computing technology is mainly applied in intelligent design in recent years. In the beginning, generating images from machine were depended on algorithm iteration with some characteristics from the data, but the accuracy of image generating needed professional information from the specialists. Therefore, some improvements are necessary for image generating. This paper reviews the progress in generating image and highlight Generative Adversarial Networks (GANs) and how generating image is applied in text-image generation and Image to image translation.

Image generation has been improved from traditional methods to advanced techniques with leverage deep learning and generative models, which enhance the quality and the applicability of generating image. Few years before, the early approaches utilized the basic algorithms that can't produce the high-fidelity images from machines. Compare with the original method, the recent advancements focus on GANs, which Chen and others described as a transformative force in image synthesis [1], enabling the creation of realistic images

Corresponding author: 21ls57@queensu.ca

across variety of domains, including the fashion clothes described by Shi and Luo in2023 [2]. For some techniques such as the layout-based image generation proposed by Zhao and others, the layout-based image generation allows the structured synthesis that maintains semantic integrity [3]. At the same year, Qiao and others introduced the Mirror GANs, which is a frame that bridges the text and image generation, this framework can improve the alignment between textual and visual inputs. Furthermore, Li and his groups discussed the controllable text to image generation, which can help the users to specify attributes to influence the generated outputs [4]. Similarly, in 2018, Johnson and his team shows the use of scene graphs for generating images that accurately depicts the relationships among objects in complex scenes. In recent years, the discussion of the creative dimensions of text to image generation underscoring the artistic potential of these models [5-7]. Collectively, these developments highlights the interactive, contextually aware and creatively capable of image generation systems.

The primary objective of this study is to explore image generation techniques and provide a comprehensive overview of the underlying concepts and methodologies. The paper introduces fundamental theories and methods related to image generation, including GANs and Mirror GANs. It details how these approaches work to produce images, discussing their respective advantages and limitations based on experimental results. In addition to theoretical insights, the paper examines practical applications of these image generation methods and explores their real-world implementation. With advancements in intelligent technology, image generation is becoming a crucial tool for enhancing communication between machines and humans. Effective image generation techniques not only improve user interaction with machines but also facilitate more intuitive and meaningful exchanges.

The paper is structured as follows: Section 2 provides background information and an overview of various image generation theories. Section 3 presents experimental results and analyses of different generation methods. Finally, Section 4 summarizes the findings and discusses the implications of these advancements for the future of information technology and intelligent machine development.

2 Methodology

2.1 Dataset description and preprocessing

For this research, the datasets have been sourced measuring the extent of existing research. In that context, it was studied and found that GANs, as a reinforcement learning tool, could generate almost any imaginable image. Hence, it could also be used for learning the high-dimensional features of data and generating real data, such as real images [1]. In this regard, the authors Shi, Q., and Luo, R., in 2023, observed that the GANs could be now used to develop clothing image generation further. Emphasized was the role of high-quality datasets in modelling the GANs [2]. In fact, the data include numerous situations for the generation of the image of high resolution, which are mainly for training and testing the GANs model that generates images of high quality [2]. Models' input and general performance can be enhanced further by standardizing the datasets and improving the data inside them. Computer should ensure that all the operations it does like cropping, rotation, and flipping result in images with unified sizes. Additionally, it is essential that the images' labels are encoded to provide the model with the ability to work with them. These preprocessing procedures are the building blocks making GANs training successful and the modelling prompt to generate images with high-quality and the model's stability.

2.2 Proposed approach

The basic objective of this paper is to undertake a thorough examination of the principles and technical aspects of image iteration, which are grounded on computer-generated imagery. The work is laid out in the path illustrated in Fig. 1.

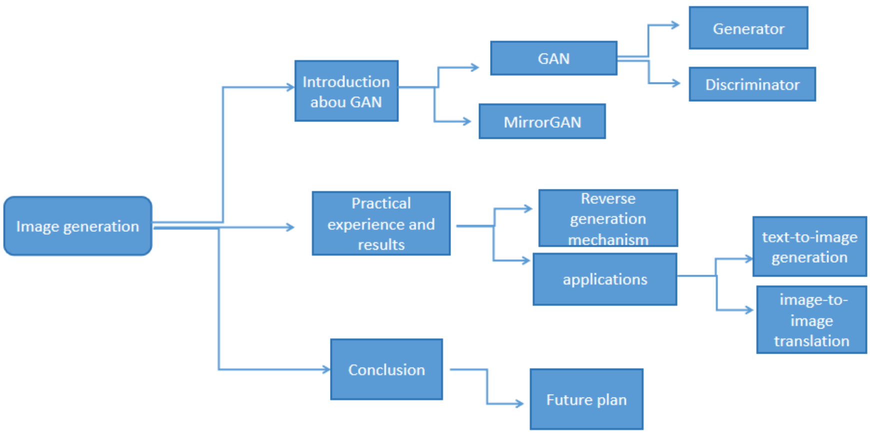

Fig. 1. Pipeline of the study (Picture credit: Original).

The first part discusses the principles and modules of image generation, which is GANs. GANs, developed by deep learning, employ two members, the discriminator and the generator. While the generator imitates low-noise images, the discriminator will label the imitated ones as either low-noise or high-noise images. GANs of the discrimination training method for barrier functionality are aroused to refine the result even if the boundaries between real pictures and GAN-generated outputs are made less accessible. Most of these techniques are inspired by GANs and MirrorGANs modes. MirrorGANs, on the other hand, applies a generation mechanism that is reverse in direction to deliver text descriptions with the images that have been produced. The approach guarantees that the generated images are not only high quality but also contain semantic meanings.

In the third part, different methods of image generation will be interpreted by practical usage and results on the subjects. By positively training, GANs are able to generate photo realism images, and vibrating text to image quality has also increased. Practical applications of these technologies involve text-to-image production and image-to-image translation, which can result in quality images being developed and increase the authentication and heterogeneity of the website [4,6]. The fourth chapter outlines and discusses the applicational issues of image generation technology, bringing forward text-to-image and image-to-image modification. Text-to-image generation targets the goal of the perfect replication of language sentences into images by text encoding into semantic feature vectors. That way, object GANs, and MirrorGANs can work on the image optimization and refining the objects drawn in the images [5]. In the fields of intelligent design, image search, and virtual reality, these technologies are among the most important. The findings section thus presents the overall research outputs, including the key findings and the role of MirrorGANs and GANs in solving image quality challenges. The conclusion will present the future approach and the main breakthrough in this technique realization; the development of artificial intelligence opens the road to a new size of communication between human and a machine [7]. This research is highlighted by the significant contribution of these technologies and how they can be instrumental in driving further investments into smart machines and the acceleration of the information technology evolution [8].

2.2.1 Introduction to basic GAN technology

The GANs is said to be composed of two main parts: the generator and the discriminator. Through game training between generator and discriminator, the generator continuously grows the quality of the training images, and the discriminator predominantly differentiates images generated from the truly real one [1,2]. The initial image generation technology contributed a lot to the generation of images and derivatives, but the quality of the images and the whole concept of artificial intelligence largely depended on the information coming from human experts [3]. In contrast, GANs makes use of deep learning and generative models, as a result enhancing the quality and versatility of the image production. The layout-based image generation technology introduced by Zhao, et al., uses a structured synthesis to preserve semantics in the generated images [4]. In addition, Qiao and other researchers have covered the aspect of improving the matching degree between text and visual inputs of their framework, MirrorGANs, by integrating text input and image generation via bridging text and visual context [5]. Li and other authors developed the research line of controllable text-to-image technology, which can deepen the control and precision of the generated images [6]. The application of such futuristic technologies in image generation has led to an evolution from conventional approaches to advanced ones that employ deep learning in order to enhance the quality and range of applications in creating pictures. This research will demonstrate the utility and communicability of machines; thus, it can further improve development of information technology [7].

2.2.2 MirrorGAN and application of text-to-image

A Real-World Use of Text-to-Image Generally, developing a reverse generation method shows good perspectives in many areas, especially in the context of GANs and their modes. GANs realize the generation of good quality images through adversarial training of the generator that comes up with the samples and the discriminator that is going to judge the samples. This mechanism does not only enhance the quality of generated images, but also changes the way the images are created, because the use of the algorithm helps to select the best image for the needed words in the new image. A technology that combines text descriptions into different pictures is one of the applications of style creation, advertising, design, and vision for the blind. The stable textual representation of images enhances images' descriptive value in a comprehensive and intuitive way. GANs successfully generate human face images with good visual quality and good likeness of the text descriptions with the help of adversarial training of generators and discriminators [5]. The generation step creates images that fit well with the Extracted Image, but the discriminator simply acts as an image checker, judging whether each generated image is a real image or not. Consequently, GANs are able to produce images that not only are semantically inferred from the input text, but also are noticeably consistent with this text data as well. Just as the basic model GANs, the modified GANs, the MirrorGANs, have two-way models for making up images and texts on fair terms. The generator has functions as emboldening images from the brief description of the script, the text in the other side, is written down as well. The discriminator, as its name implies, will compare the images with the textual passages to make sure the generated image is the correct output. GANs and its variants like MirrorGANs have delivered marvellous results in the field of text-to-image generation strategies. GANs use it for image recovery from hazy photographs, and images in black and white for colour conception, which becomes more realistic as a result [1]. MirrorGANs keeps the generated image alignment of the text description through the bidirectional generating mechanism, which is very efficient in the quality and semantic consistence of the generation [5].

The image-to-image conversion technology is one kind of media that can convert an image of type to another type, which is widely used in the fields like the image style transformation, image restoration, and the super high-resolution imaging [4]. In Pix2Pix, it is the kind of image-to-image conversion techniques that is derived from conditional generative adversarial Network (cGAN). The training is embarked on the canvas of the pairs of images where one generates the target image from the input image, and the discriminators are responsible for the meaningful separation of the generated imagery and the real one. This kind of technique performs well in the area of visual things restoration and visual pseudo-colour implementation [9]. Not to speak of, but even at the same time, the technique like GANs and its extension such as CycleGAN have empowered unsupervised image-to-image conversion with the cycled losses. It is designed specially, that it is able to convert one image to another style with the style that is not trained on the paired data, such as day to night and horse to zebra pictures. On a practical level, the spouses, Pix2Pix and CycleGAN, are able at image-switching, taking into account the above-mentioned ideas. Pix2Pix works right in fixing the images and transforming the way they look, and by using the pairs of training images and their targets, the generator can accurately generate high-quality target images [5]. With CycleGAN in the unattended image conversion, the cyclic loss takes effect, which makes the transition between different styles of images accurate without the need of paired training data. The mirror of the reverse generation mode, particularly in the area of art creation, advertising, image-improvement, style transformation, and high-resolution rebuilding, shows the prospect of use. People will get some works of art converted to simulate photographic scenes for artists to empower their creativity to the extreme and air what they have in mind confidently. Apart from this, quality images will get structured automatically according to the text of the advertisement, which will make the process of marketing cheaper and faster. Words are transformed into images to create a picturesque atmosphere for the blind. Image transformation technology is used to remedy defects or blur in specific images so that they can be in excellent quality. Finally, photos transform to any form that people can enhance the clarity by X-times prolifically [10]. Indeed, it should be mentioned that this is true only within the reverse generation technology, especially the GANs alkies, which brought some good outcomes in the field of the generated images with respect to the application value and popularity among users. These technologies are not devoid of the fact that they provide favourable conditions for scientific and practical development in the field of knowledge and image processing technologies that become an essential element of different industrial segments of the economy [11].

2.2.3 A future plan of the concluding section

Intelligent technology has developed rapidly these years, and during the near future, digital imaging will be one of the main directions in many areas. The integration of deep learning and the generative modelling will be the primary aspect for the technology of image creation, and they would not only be more similar to original real-world images, but also have more varieties and high quality. Text-to-image technology will become the central thing in the fields of content creation, design of advertising, and additional vision display, where users can select pictures through text descriptions fast. Image-to-image transformation technologies, such as well-known Pix2Pix and CycleGAN, offer important techniques in image restoration, style transformation, and super-resolution improvement with the new, more capable image editors of today [9,11]. The next level of artificial design and human-computer interaction will focus more on the intuitive and reactive machines, by which they can be aware of user’s requests and wisdom and grasp all of the human components to meet the user's demands and contribute to the desirable interactive with intelligent resolution [7]. Nowadays, the multi-modal technology, generated by the models, takes into consideration different forms of input like text, image, or sound [10]. This sooner or later, opens up more opportunities for creativity and innovation in the work of designers. The intelligence level will be raised quite remarkably in the industries in the framework of image generation technology, and promote the smart digital transition and development of society as big.

3 Result and Discussion

According to Fig. 2, the diagram represents the correlation or association of quality, diversity, and complexity of different GANs. The diagram shows the evolutionary trend of GANs architectures, which were updated over time, and important modifications were made for their improvement in performance [12].

The plot symbolizes complexity with the green and the blue streams separating and passing next to each other, which suggests hierarchy with the higher river capturing more aspects of complexity, while stream captures other aspects. These contraptions are referred to in the diagram where types of operations used are included in the architecture, like convolution, deconvolution, and self-attention. The diagram attempts to illustrate this concept of GANs accuracy and diversity improvement upon increasing complexity. One of these examples is that GANs, at the beginning, used just fully connected layers through fully conditional GAN (FCGAN), which was a process flag value for 1 that was advocating lack of variety in the operations. This structuring simplification aids in computational savings but at the cost of complexity and diversity where these are lacking in the outcomes. Other advanced GANs have adopted new function operations like convolution and self-attention and hence the kind of operations thesis have on a GAN is one of the mechanisms that defines them. This separation in functioning of the GANs permitted the processing of more definition complimentary images, ultimately leading to the good quality and more diverse images that were generated. The problem herein is the cost-intensive requirement for high computational demand to process the more complex models with advanced architectures.

Hence, the calculated results reveal the presence of an optimal balance among quality, diversity, and complexity of generative processes. These conclusions imply that incorporation of Fs may amplify the performance of GANs; however, it necessitates more computational power and training duration. It is, therefore, important for the system developer to find a good correlation between the figures in the duration and the quality of images for GAN systems to perform as expected with minimal resource usage.

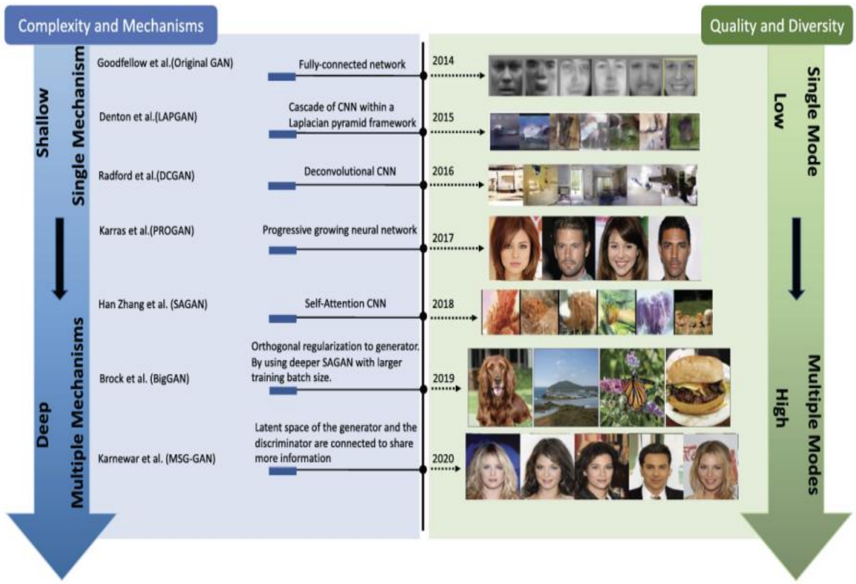

Fig. 2. Shows the GANs flow diagram about the quality, diversity, and the complexity of the mechanisms [12].

This research examines the comparative advantages and limitations of several image creation techniques. GANs have achieved rapid progress in producing images of remarkable quality. Nevertheless, non-detailed design features of GAN models may threaten ensuring the issue of high complexity and diversity of generated results. For example, Deep Convolution Generative Adversarial Networks (DCGAN) can produce high-quality images but only as long as the exact details, like complex scenes that are hard to handle, are taken into account.

On the other hand, StyleGAN revolutionizes by integrating much more complex network architecture and the technique of style controls [7]. Consequently, the outcome of this experiment are pictures, which not only consist of unique characteristics but also showcase a wide range of image complexity. The research in next phase should strive for better balance of quality with diversity; that is, quality should not be sacrificed for the sake of diversity. One possible direction of research is to develop a multimodal GAN, which will be capable of generating various kinds of outputs simultaneously and thus will meet the needs of different applications. In addition, depiction of GANs can be achieved through the usage of relevant techniques that ensure controllability so as to manipulate the positions of the given images to attain desired attributes by generators. Therefore, this remains one of the most promising research avenues [4].

The potential usage of GANs in art creation, creation of advertising designs as well as medical images production are the areas in which progress may be seen. At the same time, the latest GAN models still have flaws, such as undistinguished details and features or the blurriness and distortion of the texture of the objects. For instance, one way of addressing this could be the use of inter-grade integrating FLM as well as a mechanism using advanced classification techniques to make the images more real and add more details [5]. And also, it is necessary to engage in a process of researching those fields sharing interest in GAN (including utilizing them as educational aids) to foster the utility of GAN, and open the door to new application aspects.

4 Conclusion

This research investigates image generation technology, emphasizing the theoretical foundations and practical implementations of how computers create images. The primary objective of this study is to elucidate the concepts and mechanisms involved in advanced image generation techniques, particularly GANs and MirrorGANs, and to assess their applications in fields such as intelligent design, image search, and virtual reality. GANs and MirrorGANs were used to understand and generate useful images by transferring input text descriptions to semantic feature vectors. There are four major steps of image generation process: data preprocessing, network initialization, adversarial training, and result evaluation. These methods reflect tile switchover from the traditional image generation methods to deep learning-based ones, by which they expect to upgrade the control over and preciseness of the generated images. Many experiments were run to evaluate the proposed approaches, with the outcomes showing a marked improvement both in image quality and semantic consistency through the use of GANs and MirrorGANs, respectively. This proves the adaptation of AI technology in service ecosystems and their various uses. For the future study, more research will target how to further enrich the interaction between people and robots by image generation technology. The next-generation will focus on refining the practical and communicational aspects of intelligent contraptions, promoting the upsurge in this sphere, and heightening its tempo of nowadays information technology integration.

References

1. F.J. Chen, F. Zhu, Q. X. Wu, Y. M. Hao, E. D. Wang & Y. G. Cui, A review of generative adversarial networks and their applications in image generation, Journal of Computer science, 44(2), 347-369 (2021)
2. Q. Shi & R. L. Ruo, Research progress of clothing image generation based on generative adversarial network, Modern textile technology, 31(2), 36 (2023)
3. L.F. Li & R.Z. Sun, Bridge crack detection algorithm based on image processing under complex background, Laser & Optoelectronics Progress, 56(6), 061002 (2019)
4. B. Zhao, L. Meng, W. Yin & L. Sigal, Image generation from layout. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8584-8593 (2019)
5. T. Qiao, J. Zhang, D. Xu & D. Tao, Mirrorgan: Learning text-to-image generation by redescription. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 1505-1514 (2019)
6. B. Li, X. Qi, T. Lukasiewicz & P. Torr, Controllable text-to-image generation. Advances in neural information processing systems, 32 (2019)
7. J. Oppenlaender, The creativity of text-to-image generation. In Proceedings of international academic mindtrek conference, 192-202 (2022)
8. J. Johnson, A. Gupta & L. Fei-Fei, Image generation from scene graphs. In Proceedings of the IEEE conference on computer vision and pattern recognition, 1219-1228 (2018)
9. E. Lin, Comparative Analysis of Pix2Pix and CycleGAN for image-to-image translation. Highlights in Science, Engineering and Technology, 39, 915-925 (2023)
10. M. Li, J. Lin, Y. Ding, Z. Liu, J.Y. Zhu & S. Han, Gan compression: Efficient architectures for interactive conditional gans. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 5284-5294 (2020)

11. A. Grebo, L. Krstulović-Opara & Z. Domazet, Thermal to digital image correlation image to image translation with CycleGAN and Pix2Pix. Materials today: proceedings, 93, 752-760 (2023)
12. Z. Wang, Q. She & T.E. Ward, Generative adversarial networks in computer vision: A survey and taxonomy. ACM Computing Surveys (CSUR), 54(2), 1-38 (2021)