Visual Effect Generation Via Deep Convolutional Generative Adversarial Network

Cheng Min Yang
Bachelor of Computer Science
Asia Pacific University of
Technology and Innovation
Kuala Lumpur, Malayisa
TP064404@mail.apu.edu.my

Murugananthan Velayutham
School of Computing
Asia Pacific University of Technology
and Innovation
Kuala Lumpur, Malayisa
murugananthan@apu.edu.my

Abstract—Visual effect is important as it can make a film/video look more realistic. However, generating visual effect is costly and time consuming. In this technology era, the used of neural network become more and more. A lot of expertise starts to use neural networks to generate images, videos, and others. There are many types of artificial neural networks having different purpose and function. From these neural networks, it is found that generative adversarial neural network (GAN) has the best performance to generate data and images. In addition, applying convolution idea make GAN becomes more reliable and stable. Thus, in this research a new model of deep convolutional generative adversarial network (DCGAN) to generate visual effect is proposed.

***Keywords*—Deep Learning, DCGAN, Visual Effect, GAN**

I. INTRODUCTION

In 21st century, more and more application and company start working with artificial intelligence. With the development of artificial intelligence, the used of machine learning and deep learning is continues to rise. Nowadays, most of the company will use machine learning as data driven decision. Besides that, this AI methodology can also help industries in financial services, automotive accelerate innovation, reduce cost, healthcare, and others (NetApp, 2019). In this research, we will focus on deep learning, understanding the usage of deep learning and how we use to solve problem using it.

The main core of deep learning is the artificial neural network (Greta, 2023). There are many uses of deep learning such as text generation, object and face recognition, robotics, and AI assistant. All of this deep learning application have their own neural network and our task is to train them to make sure they can successfully done the task. From the research of Le in year 2020, there are many difference types of neural network being used in difference domains. The main neural network we will focus on this research is called generative adversarial network (GAN). Generative adversarial network is mainly used for generation approach (Brownlee, 2019). Todays, people mostly use GAN to generate data, examples, or pictures. There are many applications involves GAN such as generating cartoon characters, face frontal view generation, photos to emoji, clothing translation, generate realistic photograph an even deep fake is done by using GAN. However, in this research we are trying to figure out a GAN model that can be used to generate visual effect.

Visual effect also known as VFX is used for visual storytelling (Team, 2023). From the research of Maio (2021), an imagery created, manipulated, or enhanced for any film is also a visual effect. From their report, visual effect is used to generate realistic looking environments through the computer-generated imagery (CGI) and use a particular VFX software to combine with actual footage. To create an ideal effect, it takes a group of people working together, spending a lot of time and effort, making it very costly to produce the visual effect. Thus, using GAN model can help to reduce the cost and time, increasing the efficiency of generating visual effect.

II. LITERATURE REVIEW

Generative adversarial network can be separate into two parts to be explained which is generative algorithm and adversarial idea. From the research of Gui et al. (2020), a generative algorithm is an algorithm based on fully probabilistic model of the observed data.Adversarial idea had been applied to many areas such as machine learning, natural language processing, computer vision and others. The adversarial idea in machine learning is to train twomodel and make them compete. To applyadversarial idea to GAN, there will have two modelin GAN which are the generator G and the discriminator D (Seff et al., 2017). In the research of Luo et al. (2020), G and D are independent to each other. In this case, G will generate fake data and the fake data will combine with real data to form a dataset. Then, the mission of D is to indicatethe data is real or fake. The general formula of GAN is shown in Fig.1. Nowadays, there are manytypes of GAN developed by machine learning scientist. For example, to make the generation databecome more accurate, conditional GAN (CGAN)had been developed by adding a condition y (Fig.2). Besides that, from the research of Radfordet al. in year 2015, they had proposed a new modelwhich is deep convolutional generative network (DCGAN). DCGAN is a model combining theconvolutional neural network (CNN) with generative adversarial network (GAN). Compared to traditional GAN, DCGAN is more stable architectures in training GAN. The theory of DCGAN is to replace all the

layers having strided convolutions (the discriminator) and fractional- strided convolutions (the generator). In this case, DCGAN will provide a lower error rate when doing classification.

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{x\sim p_z(z)}[\log(1 - D(G(z)))]$$

Fig1. General Formula of GAN

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}}[\log D(x,y)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z,y),y))]$$

Fig2. General Formula of CGAN

In addition, in the report of Hang (in press),he was trying to use conditional DCGAN to generate anime avatar. To generate anime avatar, he will need to collect a lot of data to train the model. Before training, he was preprocessing the picture by adding tags with the color of hair and eyes. This is used for the generator to generate the anime character according to the color provided. In short, the program proposed by Hang can successfully generate anime avatar after adjusting the parameters of the program source code by carpedm. However, the model will still generate some poor-quality sample due to the model is not fully trained. Furthermore, from the research of Wang et al. (2020), they proposed a new conditional GAN as ImaGINator to generate video sequences with given image and motion class. The generator of ImaGINator will first encoder input image, then, it will start decoding and producing image based on a motion and a random vector. Thevideo is generated based on one image and a condition (motion, facial expression…). Besides that, they had compared ImaGINator with another video generator neural network such as VGAN and MoCoGAN. Besides than these two neural networks, ImaGINator is having shorter training time.

From the report of Hong et al. (2019), theyproposed a new model use to generate game sprite. From their proposed solution, they tried to create anew model called multiple discriminant generativeadversarial net (MDGAN). Besides than general GAN, MDGAN having two discriminator which isshape discriminator and color discriminator. Firstly, there will be two encoders will try to combine bone (animation information) and base (shape and color information) and map into a new output image. Then one of the discriminators will be used to identify the shape and another one will be identified the color. Thus, the final output will be more stable, and the image generated by the model will be more reliable. In short, MDGAN cancreate image-to-image translation of 2D game sprite sets and the elements of image can be controlled individually by the users. Moreover, from the research of Kim et al. (2023), they proposed game effect sprite generative adversarialnetwork (GESGAN) to generate game sprite effect.In GESGAN, the generator will be trained to combine object with structures. For example, the object is sword, and the structures can be fire,smoke, or water. Then, the generator will try to combine and make the sword become fire sword, smoke sword and so on. Thus, the users just need to input the object and structures, then the model will combine them to become game effect. This model can generate a lot of effect with limited data.

III. PROBLEM STATEMENT

Generating visual effect is costly and time consuming. To generate visual effect that can be used in film, for some of the frame in the film, theymight need to take up to 12 hours to done rendering. In addition, generating VFX will also need to generate CGI. A good CGI is also expensive and need to take a very long time to create. The main issues of training a model togenerate visual effect will be the limited data. Fromthe report of Hong et al. (2019), they had mentioned that some when they were training the neural network, there is not that much source data that can be used to train the model. Furthermore, the visual effect needed by the market is keepchanging. This might cause an issue that when we train the model to generate a specific effect for a long time, but then the effect can only use for one time or in one film only. This might cause that theprofit is not proportional with the cost.

IV. RESEARCH METHODOLOLY

The aim of the research is to train a GAN model to generate visual effect that can be used in film/videos. As mentioned, generating visual effect is costly and time consuming, thus, this research is trying to train a neural network that can help to generate visual effect with small amount of data in a short time.

A. RESEARCH OBJECTIVES

- To train a neural network that can generate specified effect.
- To examine the time needed for generating effect.
- To generate more sample effect for future research uses.

B. RESEARCH QUESTION

1. From those GAN models, which one is best suit for visual effect generation?
2. How should the generator and discriminator be trained to efficiently generate visual effects?
3. Is the facility will affect the generation time, cost, and result?

C. RESEARCH SIGNIFICANCE

From the research of Mitchell (2004), an effect can be categorized into special effect and visual effect. The purpose of special effect is to modify the real world like you can splash water manually to make a raining scene while a visual effect is to modify the virtual world. For example, we can set the background on fire by replacing a photographic element. In the research, Mitchell had mentioned the main purpose of effect is to present a reality view that doesn't exist in the real world. Besides that, the visual effect can also apply when it is too dangerous to take the scene like the actress need to deal with explosion scene. However, making visual effect is expensive and time consuming. From the report of Miyatovich (2022), producing a CGI per minute might take upto 1000 to 1500 dollars while producing a VFX cost more and could go up to 2000 to 5000 per minute. For more example, Miyatovich had listed out some of the films which are Jurassic World, King Kong, and Transformers. In these films, the Jurassic World having estimated cost of $170 Million, King Kong with $207 Million budget which most of the money is spent for creating the main animated character and visual effect, and thelast film, Transformer had spent for $217 Million for the CGI at the end of production. From these examples, we can easily understand that how expensive is for creating visual effect. Miyatovichalso mentioned that reason of expensive cost for generating CGI and VFX, which is because togenerate these effects, we will need to have some top-level skills, and also it might take hours and hours to get perfection. Besides that, facility is alsoa problem, to generate high quality VFX, a high- configured computer is required.

This research is significant that it can generate the visual effect automatically, saving a lot of time and cost. Besides that, the visual effect generated will not only be used one time in one filmor video only. Those visual effect could be a huge sample data for further research and approach.Furthermore, using GAN to generate VFX willhave another advantage. To train a neural network,we will need to collect a lot of sample data, however, some of the neural network will need to pre-processing data manually but GAN is unsupervised learning neural network, thus, data labelling is not required, and it will save a lot of time and cost (Sharma, 2022). Moreover, the adversarial idea makes GAN having the ability to generate the sharpest image. In addition, DCGAN applying the concept of CNN, that will make the final output become more realistic. For now, generating VFX is still a very time-consumingtask, thus, using this proposed model will help to reduce the work and also the time spending. Furthermore, using neural network to generate VFX can produce a large amount of related data, that's mean we can choose the most suitable or perfect one from the dataset.

D. RESEARCH METHODOLOGY

In this research, we will need a lot of data about the existing visual effect, the trained model, and also the algorithm or source code. Thus, most of the data will try to be obtained from online sources, journal, article, research paper/report, andbooks. Moreover, to train the model, we will try tofind some references source codes from GitHub. To train the model, we will need a lot of CGIs andVFX samples which will be obtained from online sources or existing films. This research will be technological applied research since that we had identified the problems and trying to find a solution (Indeed Editorial Team, 2022). For the research methodology, according to Kothari (2004), we will also conduct library research, analysis of historicalrecords and documents will be the main research methodology. By using this research method, we will get enough information about the formula,algorithm, and different models of GAN. The mainobjective of this research is to train a model that can help to generate visual effect easily, thus, sampling data is not required.

V. PROPOSED SYSTEM OVERVIEW

In this study, we will try to propose a new model of DCGAN to generate visual effect with small amount of data. There are two mainstructures in a DCGAN model, which are the generator and discriminator. To make sure the finaloutput more realistic, the discriminator of this model will be using the technique of convolution neural network. The visual effect of this proposed model is to transform somethings into other thingsor done prosthetics. For example, this model can make a small ball become a planet and the surrounding of it become the galaxy. Otherwise, like the film Planet of Apes, this model will auto- transform people into apes and showing their facialexpression according to the human.

For the generator of proposed model, it willreceive two images, which is the effect sample andthe base. According to the effect sample, the generator will try to combine the effect with the base image to generate a real image which is having the visual effect as shown in Fig.3. The discriminator for this model will be using CNN technique and it will be trained with real image. After the generator had generated image, the image will then send to the discriminator and the task of discriminator is to identify whether the image is real of fake. Once most of the image generated by the generator is identify as real by discriminator, the model is then done training.

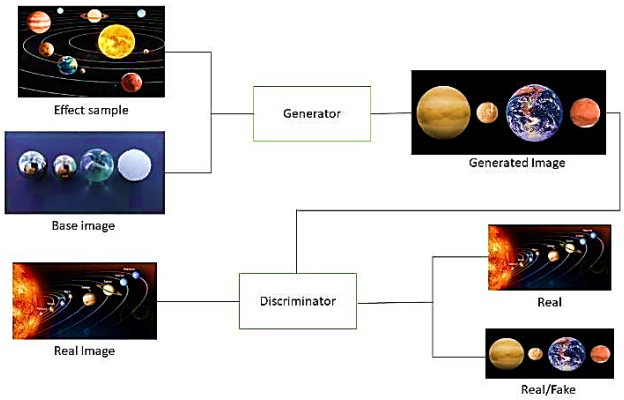

Fig3. Proposed DCGAN for Visual Effect Generation

VI. CONCLUSION

In this study, we had found out that GAN ismore suitable for generating purpose compared to another artificial neural networks. In addition, the adversarial idea used in GAN make the output datamore realistic. Using GAN to generate visual effectcan reduce a lot of works, save time, and also reduces spending on finding effects specialists.Furthermore, applying the theory of convolution neural network into GAN will make the discriminator of GAN become more stable, thus, the final output of DCGAN will be more reliable and high-quality. In short, using DCGAN to generate visual effect is a possible and efficiency way.

VII. REFERENCES

Brownlee, J. B. (2019, July 19). A Gentle Introduction to Generative Adversarial Networks (GANs). Machine Learning Mastery. https://machinelearningmastery.com/what- are-generative-adversarial-networks-gans/

Greta, G. (2023, January 26). What Is Deep Learning and How Does It Work? A Deep Dive. Hoory. https://www.hoory.com/blog/what-is- deep-learning

Gui, J. G., Sun, Z. S., Wen, Y. W., Tao, D. T., & Ye, J. Y. (2020). A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. JOURNAL OF LATEX CLASS FILES, 14(8), 2001.06937v1.

Anime Avatar Generation. Journal of SHANGHAI JIAO TONG University. https://www.cs.sjtu.edu.cn/~wang-xb/wireless_new/material/Final2018/EE/% E7%8E%8B%E8%88%AA-report.pdf

Hong, S., Kim, S., & Kang, S. (2019). Game Sprite Generator Using a Multi Discriminator GAN. KSII Transactions on Internet and Information Systems, 13(8). https://doi.org/10.3837/tiis.2019.08.025

Indeed Editorial Team. (2022, September 20). 19 Types of Research (With Definitions and Examples). Indeed. https://www.indeed.com/career- advice/career-development/types-of- research

Kim, J., Jin, K., Jang, S., Kang, S., & Kim, Y. (2023). Game effect sprite generation with minimal data via conditional GAN. Expert Systems With Applications, 211, 118491. https://doi.org/10.1016/j.eswa.2022.11849 1

Kothari, C. R. (2004). Research Methodology: Methods and Techniques. New Age International (P) Limited.

Le, J. (2020, May 11). The 10 Neural Network Architectures Machine Learning Researchers Need To Learn. Medium. https://data-notes.co/a-gentle-introduction- to-neural-networks-for-machine-learning- d5f3f8987786

Luo, J., Huang, J., & Li, H. (2020). A case study of conditional deep convolutional generative adversarial networks in machine fault diagnosis. Journal of Intelligent Manufacturing, 32(2),407–425. https://doi.org/10.1007/s10845-020- 01579-w

Maio, A. (2021, February 22). What is VFX? Defining the Term and Creating Impossible Worlds. StudioBinder. https://www.studiobinder.com/blog/what-is-vfx/

Matt, M. (2022, September 24). How Much Does A VFX Video Cost Per Minute? - Austin Visuals. Austin Visuals 3D Animation VR Studio. https://austinvisuals.com/how- much-does-a-vfx-video-cost-per-minute/

Mitchell, M. (2004). Visual Effects for Film and Television (Media Manuals) (1st ed.). Focal Press.

Miyatovich, K. M. (2022, May 25). How Much Does CGI Cost Per Second: Real-Time Analysis. The Motion Tree. https://themotiontree.com/how-much- does-cgi-cost-per-second/

NetApp. (2019). What Is Machine Learning – ML - and Why Is It Important? | NetApp. https://www.netapp.com/artificial- intelligence/what-is-machine-learning/

Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. ArXiv: Learning. http://export.arxiv.org/pdf/1511.06434

Seff, A. S., Beatson, A. B., Suo, D. S., & Liu, H. L. (2017). Continual Learning in Generative Adversarial Nets. Journal of Public and International Affairs, 2. https://doi.org/10.48550/ARXIV.1705.083 95

Sharma, A. (2022, December 2). Introduction to Generative Adversarial Networks (GANs). LearnOpenCV – Learn OpenCV, PyTorch, Keras, Tensorflow With Examples and Tutorials. https://learnopencv.com/introduction-to-generative-adversarial-networks/

Team, N. (2023, January 3). VFX – Everything You Need To Know. NFI. https://www.nfi.edu/vfx/

Wang, Y., Bilinski, P., Bremond, F., & Dantcheva, A. (2020). ImaGINator: Conditional Spatio-Temporal GAN for Video Generation. 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). https://doi.org/10.1109/wacv45572.2020.9 093492.