Development and Research of Generative Animation Based on AIGC

Li Peng*

College of Art and Design, Hubei University of
Technology, Wuhan, Hubei, China
19950027@hbut.edu.cn

Bihan Fu

College of Art and Design, Hubei University of
Technology, Wuhan, Hubei, China
102201264@hbut.edu.cn

ABSTRACT

AIGC is the automatic generation of relevant content from a large number of models and databases based on the input of relevant instructions. This paper mainly discusses the development of artificial intelligence generation technology in animation in the form of images. With the rapid development of artificial intelligence technology, generative animation content has become an important way of animation production and production in modern society. Artificial intelligence-based generative animation is taking up an increasing proportion of the film, game, advertising and other industries, AIGC has been rapidly developed by virtue of its lower production cost and higher production efficiency, which constantly impacts the development of the animation industry and attracts the widespread attention of animation creators. This paper analyses the Generative Adversarial Network and Diffusion Model through the deep learning principle of AIGC technology, and explores the important impact of AIGC technology on the form of generative animation. Using computer discipline thinking to analyse the type of AIGC model for animation creation, and according to the model simulation results of AIGC technology used in generative animation production of technology, market aspects of its development prospects are summarised.

CCS CONCEPTS

• **Applied computing** → Arts and humanities; Media arts; Information systems; Information systems applications; Multimedia information systems; Multimedia databases; Human-centered computing; Human computer interaction (HCI); Interaction techniques; Text input; Computing methodologies; Computer graphics; Animation; Procedural animation; Computing methodologies; Computer graphics; Image manipulation; Image processing.

KEYWORDS

AIGC, animation, generative, database

Table 1: As of 2023, the classification of AIGC related models and products is shown in the legend.

Category	Product
Language generation	BERT
	MT-NLG
	ChatGPT
	LLaMa
	ERNIE Bot
	Flamingo
Code Generation	Codex
	AlphaCode
	Midjourney
Image Generation	IMAGEN
	Stable Diffusion
	DALL-E2
	Dreamfusion
	Phenaki
Video Generation	Make-A-Video
Audio Generation	AudioLM

1 INTRODUCTION

The rapid development of Artificial Intelligence is constantly impacting changes in various industries, and the creation of a series of generative simulators such as Midjourney and Stable Diffusion is impacting the development of animation, film, games and other industries. AIGC, also known as AI Generated Content, is a form of technology that automatically generates images, audio, and video using a computer to input the appropriate conditions. It is considered a new way of content creation after PGC and UGC [4]. Table 1 shows the classification of AIGC related models and products through 2023.

AIGC technology is impacting the development of the painting, animation and film industries with its high efficiency and low cost, and is one of the hotspots of artificial intelligence research in today's society [5]. With the gradual maturation of this technology, generative text, images, audio, animation has been increasingly intense. The emergence of such generative simulators in the auditory, visual and other aspects of the development has gradually exceeded the human form of processing. AIGC technology which greatly improves the speed of production of animation short films, and promotes the development of the animation industry. In the

Figure 1: Screenshot of the animated short film "The Crow".

Figure 2: Screenshot of the painting "Space Opera House".

field of animation, the commonly used generative models are Generative Adversarial Networks and Diffusion Model. In July 2022, The Crow, an animated short film produced by AIGC, won the Best Short Film Award at the Cannes Short Film Festival [8]. Nowadays, AIGC technology has already achieved significant success in the animation industry. Figure 1 is a screenshot of the animated short film "The Crow".

2 TECHNICAL PRINCIPLES AND MODELS OF AIGC DEEP LEARNING

AIGC is an automatic content generation method based on Generative Adversarial Network (GAN), Generative Diffusion Model, etc. It is a kind of animation production method based on artificial intelligence technology, which combines a large number of databases such as image, audio, and video. Through data learning and pattern recognition, it realizes the integration between various databases and automatically generates image contents according to the user's input instructions [7]. In recent years, the popularity of "AI paintings", "digital people" and "chatbots" are all products of the development of AIGC technology. Compared with traditional animation production methods, generative animation forms can quickly generate a large number of animation materials, greatly reducing the animation production of labor and time costs. "Dogs and Boys" is the first animated film to uses AIGC technology to assist in creation.

This short film uses AIGC technology to draw complete animation scenes, which is a new experiment for animators and an innovation in animation creation methods. With the increasing maturity of AIGC technology, the resulting new mode of animation production will gradually develop.

2.1 Generative Adversarial Networks

The generative adversarial network consists of two neural networks: generator and discriminator [4]. The function of the generator is to generate new data similar to the input data, while the discriminator is used to distinguish the authenticity of the input data [2]. The two networks confront each other, and through repeated training, adjustment, and optimization, they ultimately enable the generator to generate very realistic forged data, and the discriminator cannot

distinguish between truth and false, reaching the point of "confusing truth with false" [8]. The application of Generative Adversarial Networks in animation mainly includes: automatic generation of animation and generation of animation keyframes.

In terms of automatic animation generation, Generative Adversarial Networks are mainly used for generating object dynamics, and character actions. The establishment of this model first involves collecting relevant information such as the shape, style, and character motion effects of the animation to form a database [11]. Second, train the collected database to generate adversarial network models, and generate animations based on the animation shape of the generative adversarial network using motion engines and skeletal animation techniques; Finally, continuously adjust the parameters of the generative adversarial network model to improve the authenticity of the generated animation and achieve the best results. In terms of generating animation keyframes, the generating adversarial networks is mainly used to generate keyframes during the character's motion process. Based on the relationship between keyframes in the database, attempts are made to find intermediate frames in the image. Generative animation keyframes are usually needed to provide the first and last frames for the generative adversarial network so that it can recognize the style type of the generated animation.

2.2 Diffusion Model

The Diffusion Model refers to the gradual restoration of a noise free image by adding noise to the data, which is divided into two types: forward diffusion and reverse diffusion. Firstly, it is necessary to gradually add noise through forward diffusion until the image is completely destroyed by noise, and then convert noise through reverse diffusion to gradually restore the image without noise and restore the original image [4]. Through training, the model can use relevant denoising methods to generate new data models based on random input instructions [10]. Up to now, breakthroughs have been made in the use of Diffusion Modeling on Midjourney, Stable Diffusion. In August 2022, a painting created by Midjournal, "Space Opera House," won the Digital Art category at an art fair in Colorado [8]. Figure 2 is a screenshot of the painting "Space Opera House".

On January 30, 2023, Netflix officially released an animated short film "The Dogs & The Boys". The background in this short film is drawn using a Diffusion Model, using which hundreds of images

Figure 3: Screenshot of the animated short film "The Dogs & The Boys".

Figure 4: The process of generating a stable diffusion pattern.

Figure 5: For example, enter the text 'a house in the forest', the basic code is as above.
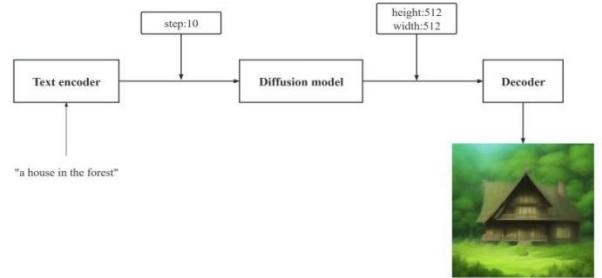
can be generated in a short period of time according to the relevant instructions, generating a number of images that far exceeds the scope of the animation creator's hand-drawn drawings in the same time [9]. Figure 3 is a screenshot of the animated short film "The Dogs & The Boys".

Until now, the models often used by animation creators are Midjourney, Stable Diffusion. Among them, Midjournal can generate highly unique style images based on instructions, with a low threshold [9]. Stable Diffusion is enhanced and improved on the basis of Midjourney, the generation speed is amazing, and a large number of pictures can be generated in a short period of time, and the user can also train his own model according to his own painting characteristics, and create as he pleases, the threshold is high. The application of diffusion model in Stable Diffusion mainly includes text-generated graph, graph-generated graph and image in painting. Among these three items, the most basic and commonly used function is the text-generated graph, which can generate corresponding images based on user input instructions [6].
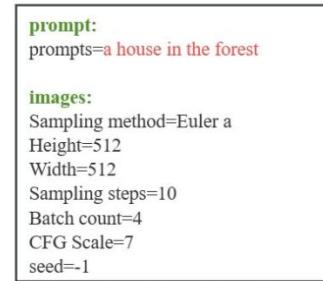
3 ANIMATION GENERATION MODEL OF AIGC

Animation is an image activity formed by continuous playback of solid-state images captured in a frame-by-frame manner, while video is a continuous sequence of images, so animation is also a video technology [7]. Disco Diffusion is an AI image generation program that generates images based on keywords entered by the user, a form of non-linear feature extraction based on the analysis of textual data, and supports the generation of more creative and visually appealing videos from different images when generating images in a multimodal manner.

On September 29, 2022, Meta-AI launched the artificial intelligence system Make-A-Video, which successfully achieved the generation function from text to video and from images to video. The development of such generation models laid the foundation for the birth of generative animation. Users can generate short videos based on input keywords or images. In addition to short film generation, AI technology can help users create desired characters and scene designs, greatly reducing the original painter's drawing time [1]. In unit time, the number of AI generated images far exceeds the limit of human painting, and the generated characters and scenes are also unique in style [9]. During the process of generating images

and videos, users need to adjust the values appropriately to ensure the quality of the short film. Users can use palette technology to change the color of the image, use motion blur to simulate more realistic state of the object movement, and add lighting, projection, and other technologies to the image to make it more realistic and vivid. At present, but short film generation technology is still in the development stage, and the generation technology is not yet mature. There is still a gap compared to the traditional image effect.

Next, taking the application of diffusion model in Stable Diffusion as an example, we will demonstrate the steps of creating a scene in an animated short film using text-generated graph.

Firstly, the user enters the prompt word, the number of image steps is set to 10, the image corresponds to a height of 512 and a width of 512 (the height and width refer to the noise in the diffusion model, and the image size of 512x512 corresponds to a noise dimension of 64x64x4). Finally, it is fed into the generation to get the text-generated graph. Figure 4 is the process of generating a stable diffusion pattern. Figure 5 shows the basic code for entering the text "a house in the forest". Figure 6 shows the final 4 images generated after adjusting the noise in the diffusion model for the desired number of skins.

The commonly used random seed is -1. When the random seed is -1, the generated image is completely random. After each click to generate, the correlation between the generated images is small and irregular. However, when the random seed is 0 or larger, the generated images only show slight changes each time. Controlling

Figure 6: After the adjustment of the noise in the diffusion model, the number of skins is adjusted according to the desired number and finally generate 4 images.

Figure 7: Adjusting the image to 256x256, the resulting effect is as above.

random seeds can better help users improve the performance, quality, and color saturation of images without changing the specific shape.

The image size can be determined by changing the height and width. The previous figure is the value originally determined in Stable Diffusion, we can according to the desired image size. However, when the size is smaller than 512x512, the image quality will be reduced. Figure 7 shows the effect of adjusting the image to 256x256. Figure 8 shows the result when the image size is larger than 512x512 and the image will be repeated or stretched and deformed.

Although the diffusion model Stable Diffusion allows for arbitrary resizing of images and supports generating various sizes. But through extensive simulation and practice, it can be concluded that: The size of 512x512 is the best image effect. Users can adjust the image according to their needs to achieve the expected effect. Figure 9 shows the adjustment of the second parameter sampling step, changing the sampling step to "10, 30, 50, 80".

As the value of Sampling steps increases, the quality of the image improves. The following images show the effect of only adjusting the Sampling steps without changing other values. Figure 10 shows the images generated when the sampling steps are 10, 30, 50 and 80 respectively.

Figure 8: When the size is greater than 512x512, the image will be repeated or stretched and deformed. Adjusting the image to 512x768, the result is as above.

Figure 9: The second adjustable parameter is Sampling steps. Change Sampling steps to "10, 30, 50, 80", the basic code is as above.

It is concluded that when the value of Sampling steps is larger, the clarity and richness of the image is higher, but the higher the value of Sampling steps, the longer the computer computing time. Users do not have to pursue the high value of Sampling steps, we can adjust it according to the need. When the value of Sampling steps is 50, it can generate relatively stable and good quality images. Figure 11 shows the adjustment of the third parameter, CFG Scale, which has an initial value of 7. Change the CFG Scale to "1, 5, 7, 10".

As the value of CFG Scale increases, the color saturation of the image increases. The following images show the effect of adjusting only the CFG Scale without changing any other values. Figure 12 shows the images generated when the CFG Scale are 1, 5, 7 and 10 respectively.

Figure 10: Sampling steps=10, Sampling steps=30, Sampling steps=50, Sampling steps=80.

Figure 11: The third parameter that can be adjusted is CFG Scale, the initial value of CFG Scale is 7, change CFG Scale to "1, 5, 7, 10", the basic code is as above.

Figure 12: CFG Scale=1, CFG Scale=5, CFG Scale=7, CFG Scale=10.

It is concluded that when the value of CFG Scale is higher, the color saturation of the image is higher. However, there is no need for the user to pursue high saturation values; when the color saturation is above a certain value, the image has an unnatural picture effect.

Figure 13: Enter the text "a house in the forest, with a stone paved road in front of it, and a small stream, butterflies dancing in the air", the result is as above.

Users can experiment around the initial value of 7 to get a higher image result.

In addition to the basic text input, you can also add prompt words on this basis, each prompt word separated by ",". As more text is input, the more the image generation meets the expected requirements. Figure 13 shows the image generated by typing the text "a house in the forest, with a stone paved road in front of it, and a small stream, and butterflies dancing in the air".

The commonly used Sampling method are Euler a、DPM++2S a Karras、DDIM. The difference between them is the number of sampling steps, with DDIM having the highest number of sampling steps. In the generation process, the speed comparison: Euler a>DPM++2S a Karras>DDIM, the picture quality comparison: DDIM>DPM++2S a Karras>Euler a. Figure 14 shows the images from left to right are: Euler a, DPM++2S a Karas, DDIM.

4 PROSPECTS FOR THE DEVELOPMENT OF AIGC ANIMATION

With the rapid development of AEGI technology, the computer algorithms supported by AEGI technology still need to be improved to be truly applied to the animaion industry. Attention must be paid to the challenges faced in development, whether it is the maturity of the technology, the stability of the data, or the regulation of the market, all of which need to overcome certain difficulties. Further improve the stability, security and effectiveness of AEGI technology, so that generative animation has been applied at a deeper level, reducing the workload of creators and improving the quality of animation creation. This section will point out the challenges faced by AEGI technology in order to better facilitate the development of AEGI technology.

4.1 Technological maturity

AI was used to draw the backgrounds of the animated short film "The Dogs & The Boys", which was created by Netflix, rinna, and WIT STUDIO. The animation creator inputs information into Netflix's database, the AI generates the corresponding images, and the animation creator adjusts and integrates the AI-generated images, demonstrating the innovative results of AIGC technology. Through the animated short film "The Dogs & The Boys", the great potential of generative animation supported by AIGC technology

Figure 14: The following images, from left to right are: Euler a, DPM++2S a Karras, DDIM.

has been brought to light. AIGC technology can help generate the background, music and sound effects of the image, which greatly reduces the repetitive nature of the work of animation creators and saves costs and production time. However, the technique has some shortcomings besides shortening the production cycle and reducing the cost, such as the quality of the generated images, the stability of the generation algorithm, and the availability of the data information. In terms of the quality of the generated images, there is still a gap between the quality of the AI generated images and the traditional 2D drawing, with rough image quality. In terms of the stability of the generation algorithm, AI generation technology is now in an experimental stage, generating images that are sometimes good and sometimes bad, with an uncontrollable nature[9]. In terms of availability of data information, the image generated by AI rely on the information in the data set, but most of the data sets are biased, resulting in the data not being better presented on the image.

4.2 Market Regulation

AIGC technology belongs to the emerging field in the direction of artificial intelligence, and industries are gradually shifting their development direction to seek market shares supported by AIGC technology. AIGC technology is a data set assembled from a large number of images, sounds, and videos that imitate a large number of people's appearances, behaviors, voices, and integrates and assembles the various collected information [3]. While the technological enhancement and generative development aspects of AIGC have many positive effects on modern life, they also pose some challenges. With the improvement of intelligence, the risk of information leakage is increasing, which brings great harm to people's information and property security. Therefore, it is necessary to adopt corresponding laws and regulations and corresponding market supervision to cope with this challenge.

At the level of market supervision, market regulators should build an intelligent monitoring system centered on AI technology according to the development needs of generative artificial intelligence, to protect the personal interests of users and maintain a good market order. For example, regulators can quickly identify market anomalies and find relevant countermeasures in a timely manner by means of technological regulation and big data analysis, to reduce market risks as well as infringement of users' interests [9].

At the level of laws and regulations, the country needs to formulate corresponding laws and regulations to have higher requirements on data processing and privacy protection for relevant industries using AEGI technology. Related industries must comply with laws and regulations on intellectual property and security management in the use of copyright. In the use of the market, the laws and regulations on fair competition must be followed. In the creation of datasets, the relevant data protection laws and regulations must be observed.

5 CONCLUSION

This article through the AIGC technology in the animation industry development prospects for in-depth study, pointed out that the AIGC technology in the animation industry has a huge development potential and a wide range of application potential [9]. The research in this article mainly emphasizes the use of Generative Adversarial Networks and Generative Diffusion Models in Stable Diffusion based on deep learning principles, which gradually enriches and adjusts the generative pictures supported by AIGC technology through forward diffusion and backward diffusion in the diffusion model. The proposed diffusion model is of great help to the adjustment of generative pictures and the form of animation production, and the development of this technology will change the traditional image drawing and animation production form, to promote the development of the animation industry. However, the AIGC multimodal algorithm still needs to overcome certain difficulties in order to truly promote the development of the animation industry, and we need to continue to strengthen the maturity of AIGC technology and market regulation while utilizing AIGC technology. With the continuous improvement of technology, the form of animation creation empowered by AEGI technology is rapidly advancing.

REFERENCES

[1] Jiang, R., Wang, L., & Tsai, S.-B. (2022)., "An Empirical Study on Digital Media Technology in Film and Television Animation Design," Mathematical Problems in Engineering, 2022.

[2] Luo, C. 2022., "Understanding Diffusion Models: A Unified Perspective," arXiv.

[3] Wang, Q., Ma, H., Wei, W., Li, H., Chen, L., Zhao, P., Zhao, B., Hu, B., Zhang, S., Zheng, Z., & Wu, B. 2023. Attention Paper: How Generative AI Reshapes Digital Shadow Industry," arXiv.

[4] Su, N. 2023., "Research on Multiparty Participation Collaborative Supervision Strategy of AIGC," ICEIEC 2023 - Proceedings of 2023 IEEE 13th International Conference on Electronics Information and Emergency Communication, 268–272.

[5] Xu, C. 2023., "Immersive animation scene design in animation language under virtual reality," SN Applied Sciences, 5(1).

[6] Wu, J., Gan, W., Chen, Z., Wan, S., & Lin, H. 2023., "AI-Generated Content (AIGC): A Survey," arXiv.

[7] Aysolmaz, B., & Reijers, H. A. 2021., "Animation as a dynamic visualization technique for improving process model comprehension," Information and Management, 58(5).

[8] Ouyang Chunxue., "Exploration of multimodal AIGC animation based on deep learning [J]," Modern Film Technology, 2023 (01): 41-47.

[9] Gao Rui., "From the animated film "Dog and Youth" to the Metaverse: The Potential, Applications, and Challenges of AIGC [J]," Modern Film Technology, 2023 (05): 24-28+17.

[10] R. Rombach, A. Blattmann, D. Lorenz, P. Esser and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 10674-10685, doi: 10.1109/CVPR52688.2022.01042.

[11] Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. S., Gontijo-Lopes, R., Salimans, T., Ho, J., Fleet, D. J., & Norouzi, M. 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. Advances in Neural Information Processing Systems, 35.