Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 2, No. 1, 2025

The Transformation and Prospect of Content Creation Driven by AIGC

Weiyan Yang, Tao Tan *

School of Computer and Software, Southwest Petroleum University, Nanchong, China

* **Corresponding author**: Tao Tan

Abstract: This paper delves into the technological evolution of generative AI and elaborates on its various application forms in the field of image creation, such as image-to-video generation, style transfer, digital human creation, and 3D model generation. On this basis, it further explores effective strategies for guiding AIGC to achieve efficient creation, aiming to provide theoretical and practical references for promoting the deep integration of generative AI and content creation. At the same time, it analyzes the key challenges faced by AIGC in the content creation process, including the lack of content rationality, the ambiguity of copyright and originality, and cultural and ethical risks. Taking the 3D animation industry as a typical case, it deeply analyzes its application status and industry pain points.

Keywords: AIGC; Content Generation; Technical Limitations; Industrial.

1. Introduction

With the rapid development of artificial intelligence technology, AIGC has caused a huge stir in the field of content creation. Traditional content creation relies on professional skills and a large amount of human and time investment. The emergence of AIGC has broken this pattern, bringing unprecedented opportunities and changes to image creation and becoming a key hotspot in the digital art and media fields. The state attaches great importance to the development of the artificial intelligence industry and has successively issued a series of policies. The "New Generation Artificial Intelligence Development Plan" clearly states that it is necessary to accelerate the integration and application of artificial intelligence in various industries and cultivate and expand the artificial intelligence industry. This plan provides strong policy guidance for the exploration and practice of AIGC in the field of content creation, prompting many enterprises and creators to actively engage in the application and innovation of AIGC technology.

2. Basic Principles of Generative AI

Any image is divided into pixels and stored in a computer, and then displayed on the screen. Each pixel contains hue and luminance information. If different pixels are randomly arranged, the result displayed will be similar to the state of an old TV without a signal, with a dense array of snowflakes. Each arrangement is regarded as a vector, and these vectors form a pixel space. A specific image is one of the pixel arrangement combinations, that is, a point in the pixel space. Translating most of the points in the pixel space will result in meaningless images. Generally speaking, natural images that conform to human cognition basically follow certain arrangement rules, such as the pixels in the main content part being similar and the overall tone of the picture being uniform, etc. These rule features all meet certain constraint conditions. How to find the corresponding points according to the requirements is the essence of generating images.

Currently, the mainstream image generation model is the Diffusion Model. This model is similar to simulating the diffusion and reverse diffusion in thermodynamics, rearranging disordered pixels into orderly images. Its training process can be roughly described as: gradually adding Gaussian noise of a certain intensity to the training set images to simulate the diffusion state until the images become standard Gaussian noise. Then, attempts are made to denoise the images. The denoising method utilizes the characteristic that the reverse process of diffusion also approximately follows a Gaussian distribution. During this process, additional association labels are made for the features and prompt words of the training data, and the model learns the pixel arrangement and combination rules of the training set images. It can be understood that the model has learned how to shape a randomly shaped piece of clay into a statue similar to the content of the training set.

Figure 1. Simulate diffusion and denoising process

Stable Diffusion is one of the earlier models that sparked the image generation craze in recent years. Later, Transformer demonstrated significant advantages in the field of text understanding and generation with features such as position encoding and attention mechanisms. By combining the aforementioned diffusion model, Transformer became a model capable of efficiently capturing dependencies in data and generating high-quality results - DiT (Diffusion Transformer) [1].

3. Main Forms of AIGC Content Generation

Apart from the common forms of text-to-image or text-to-video generation, AIGC also includes the following forms.

3.1. Image-to-Video Generation

Enabling static image content to move mainly involves the

following two technologies:

Depth Information Utilization: By using depth estimation methods to obtain the depth information of objects in the image, and then adjusting the movement and occlusion relationships of the objects based on the depth information. In a scene with multiple objects, AI can correctly handle the occlusion and appearance relationships of the objects during movement according to their depth order.

Optical Flow Estimation and Motion Interpolation: Optical flow refers to the motion information of objects in an image. By calculating the optical flow, the movement direction and speed of each pixel in the image can be determined. Through algorithmic analysis of the optical flow information in the input images, and then using motion interpolation techniques to generate intermediate frames between adjacent image frames, the movement of the images becomes smoother and more natural.

For example, when converting a static landscape image into a dynamic one, if there is a scene of wind blowing leaves in the picture, AI first analyzes the depth information, estimates the movement direction and amplitude of the leaves through optical flow, and then inserts appropriate intermediate frames between different time point image frames to show the dynamic effect of the leaves swaying in the wind.

3.2. Style Transfer

The common method is based on feature extraction and transformation. By using a pre-trained deep neural network model, the content features of the input image and the style features of the target style image are extracted, and then these features are recombined and mapped through specific algorithms to achieve the transformation of an image from one style to another. For example, by analyzing and processing the high-frequency and low-frequency features of the image, the details and overall atmosphere of the image are changed to achieve the effect of style transfer.

Figure 2. Behind the scenes production of rock, paper, scissors

If it is a case of converting a real person into an anime style, in the early stage, real person videos need to be shot and a model needs to be trained using images of the target style. However, it is inevitable that there will be unstable details in the picture and flickering problems. The animation shown in the picture was manually modified extensively in the later stage of production to achieve a relatively ideal effect [2].

3.3. Digital Humans

This is a type of comprehensive technology that includes audio feature extraction, facial expression recognition, and the establishment of motion mapping relationships. Generally, audio and reference image of the character are required. The uploaded or real-time audio is analyzed to extract information such as pitch, timbre, and loudness that reflect the voice and convert it into feature vectors. Facial landmark points are located to obtain the position and shape of the facial features, providing a basis for subsequent expression and mouth shape generation. The model used needs to be pre-trained to establish a universal mapping relationship between voice and mouth shape. Then, based on the subsequent provided voice, the mouth shape movement can be predicted, and the mouth shape can be fused with the reference image to generate the corresponding virtual digital human.

3.4. 3D Model Generation

In addition to 2D images, AI has gradually demonstrated certain capabilities in the field of 3D production. Some of the more advanced models include Rodin and others that are trained on a large amount of 3D model data, which covers the geometric shapes, topological structures, and texture information of various objects, scenes, and structures. During the training process, the model learns the feature patterns and spatial relationships of different objects. For example, for common furniture models, it can grasp the leg structure of a chair, the shape of the backrest, and the proportional relationship between them.

In addition to data-driven learning, some AI 3D model generation methods also incorporate rule-based and constraint-based mechanisms. For instance, in the generation of architectural models, the generation process is constrained and guided according to architectural design norms and principles, such as the rationality of room layout and structural stability, to ensure that the generated 3D models meet the requirements of practical applications.

4. Strategies on Guiding Efficient AIGC Creation

Most domestic AIGC platforms operate in a cloud-based manner, processing users' requests on the server side and returning the results, which often requires queuing. If sufficient hardware resources are available and faster results are desired, local computing platforms can be chosen, such as the open-source image generation platform ComfyUI. To provide users with greater editing space, AIGC platforms set numerous parameters for adjustment, which may vary across different platforms. Here, taking ComfyUI as an example, we will explore several aspects to enable more precise and efficient realization of creative intentions through AI.

4.1. Optimizing Prompt Words

For text-to-image generation, to produce images that meet expectations, the writing of prompt words needs to be comprehensive and detailed, with certain grammatical rules and techniques.

A complete prompt should at least include two types: content-based prompt words and standardized prompt words. Content-based prompt words mainly describe the features of characters and subjects in the picture, scene elements,

environmental lighting, and camera angles, while standardized prompt words mainly specify the quality and style of the picture, such as general high definition, illustration style, and realistic style. Start by describing a general image, then gradually add details and adjust the weights of different prompt words. The aforementioned are positive prompt words, and negative prompt words are usually added to indicate what the AI should not include in the picture, which can be directly applied using existing templates. If necessary, related prompt word generation tools can be used as an aid.

4.2. Low-Resolution Prototyping

A perfect image that fully meets expectations often requires dozens or even hundreds of generation attempts. If the resolution is set too high from the start, the generation efficiency will be very low. Therefore, the general approach is to first generate at a lower resolution until the desired general content appears, then correct parts of the image through local redrawing, such as correcting hand positions or adding elements, and then use the high-resolution refinement function to enlarge and redraw to optimize details. This approach is also applicable to video generation, where low-resolution and low-frame-rate segments are generated first, and then the quality is improved through enlargement and frame interpolation algorithms.

4.3. Video Stability and Continuity

In video generation, increasing the video length significantly increases the computational load, and poor continuity in long videos is a common issue [3]. Generally, image-to-video generation is adopted, which offers better controllability than text-to-video generation. First, conceive the video's plot script, and generate corresponding storyboards and key frames through text-to-image generation. Then, use AI to animate the key frames into a video. Sometimes, to ensure that modifications to the foreground subject and background do not affect each other, the two can be separated through the chroma keying function. To avoid defects and turn shortcomings into features, it is necessary to be familiar with camera language and use editing techniques to skillfully arrange the plot and storyboards [4].

4.4. Adaptation to Creation Scenarios

The current application environment of generative AI is in simple precise content or complex ambiguous content. That is, the more specific the creator's expectations, the less likely the generation results will meet expectations. Different generation methods have certain limitations, and models should be flexibly selected based on the creative concept to maximize the advantages of generative AI.

At present, AI content generation in static planar images is relatively mature, and the introduction of AI in the field of graphic design has become a common phenomenon. AI can combine and associate based on prompt words, and the generated content is often beyond normal logic, which helps break through conventional thinking and serves as a highly efficient inspiration library for designers in the initial concept stage when ideas are still vague [5].

In terms of dynamic videos, current AI can generate realistic short clips, but it is not yet ideal for scenes with strong narrative or high requirements for rhythm changes. This makes it more suitable for assisting in generating empty shots in films and television. Such shots do not appear frequently throughout the film, but sometimes they require a considerable amount of time to create special effects scenes in post-production. Additionally, science popularization or explanatory videos are also a good application scenario for AI. These videos do not have high requirements for the continuity of the picture, but rather focus on the relevance between the picture and the script. They require extensive collection of materials, so AI can be fully utilized to enhance the richness and vividness of the video. In terms of 3D models, compared to language or images, there are far fewer datasets of 3D models available for training. Therefore, there is still a need for accumulation in quantity. The models generated by AI can basically meet the simple initial models of the creators, and modifications can be made on this basis, which can also save the basic work in the early stage of creation. However, in fields with a large number of precise parameter requirements such as architecture or industrial design, the instability of AI temporarily makes it unsuitable for these tasks.

5. The Impact of AIGC on Content Creation

5.1. Enhancement of Creation Efficiency and Quality

AI has significantly shortened the image creation cycle. Complex image creations that previously took several days or even weeks can now be completed in a draft form within a few hours with the help of intelligent algorithms. In terms of quality, by learning from a large amount of high-quality image data, AI-generated images can reach a high level in composition, color matching, and other aspects, providing creators with high-quality starting points or material references for creation. This enables creators to focus more on creative conception and detail refinement, thereby enhancing the overall efficiency and quality of image creation.

5.2. Innovation and Integration of Artistic Styles

AI can rapidly learn and integrate various artistic styles, creating brand-new visual styles. It breaks the boundaries of traditional artistic styles, interweaving elements from different cultures and historical periods. For instance, by combining traditional Chinese meticulous painting with Western modern abstract art, it generates image works that possess unique oriental charm and modernity, providing new ideas and methods for artistic innovation and enriching the forms of expression and cultural connotations of image art.

6. Challenges Faced by AIGC in Content Creation

6.1. Conformity to Real-world Laws and Uncertainty

From the training process, it can be seen that AI generates content by understanding the surface features of images or the real world, but it does not understand the reasons behind the appearance of these features. This often leads to the generation of content that violates the physical laws of reality. Insufficient controllability is also one of the reasons why generative AI cannot be widely integrated into the standard animation production process. The production of conventional animation involves many stages, such as storyboarding, keyframe drawing, animation, and scene

creation, with the final image being composed step by step. AI, on the other hand, follows an end-to-end model, generating the final image in one step. The output is indeed related to the prompt, but it is not stable, making it difficult for creators to modify certain details to achieve the desired effect. This may force the entire production approach to be altered, with the idea of "what I want to do" being compromised for "what AI can do".

6.2. Copyright and Originality Issues

There is a gray area regarding the copyright ownership of AI-generated content. Currently, there is no well-established normative criterion to regulate the legality of training data. AI needs to learn based on a large amount of existing data, and in fact, most training sets are scraped without the knowledge of the owners for model training. The existing data belongs to the creative assets of content creators. The labor achievements of traditional content creators are not protected, which is also a blow to their originality enthusiasm. At the same time, over-reliance on AIGC may lead to the weakening of creators' original thinking. Some creators may only make simple modifications based on the content generated by AIGC rather than truly independent creation, which has a certain impact on the original content ecosystem [6].

6.3. Cultural and Ethical Issues

The content generated by AI stems from the analysis and reasoning of vast amounts of data. Nevertheless, AI itself lacks the ability to perceive and judge social norms and moral ethics, potentially giving rise to risks of bias and discrimination. During the data training process, if the data primarily originates from a specific cultural background or group, the generated content is likely to lack understanding and expression of other cultures and groups. In the event that the training set lacks sufficient breadth, it might lean towards a certain extreme. Take Google's AI model, Gemini, for instance. It has been pointed out that it is unable to generate correct images of white historical figures. When asked to provide an image of the Pope, Gemini produced a non-white image. When asked to provide images of "The Three Musketeers", the Russian Tsar, and German soldiers in 1943, it generated images featuring people of color, whereas these figures are generally regarded as white in common perception. This has triggered concerns regarding AI's distortion of correct reality.

7. Conclusion and Outlook

In the current situation, AI has made significant progress in the domain of content creation, particularly for content viewers and creators of smaller scale with lower requirements for content quality. AI has brought numerous positive effects, such as enhancing content creation efficiency, promoting the innovation of creative styles, and expanding the creative community. However, to genuinely address the pain points of creators within the industry, AI research needs to delve deeper into the creative process.

Take the traditional 3D animation industry as an example. It is a labor-intensive sector with a considerable amount of repetitive and cumbersome tasks, such as UV unwrapping, topology, weight adjustment, and bone binding. These tasks are more suitable for automation. By liberating creators from these repetitive labors, they can invest more energy in content design that reflects artistic value and provokes thinking. Nevertheless, the current generative AI still has considerable gaps in these targeted functions [7].

Concurrently, the development of generative AI confronts numerous challenges, such as copyright issues, cultural and ethical matters, and technical refinement. These problems cannot be disregarded. Only when technology and norms develop in concert can generative AI fully unleash its potential in the field of content creation, truly achieve the deep integration of technology and art, and enable image art to radiate even more resplendently in the digital era.

References

[1] Zhao Keao, Han Feilin Progress and Comparative Analysis of Cultural and Creative Video Applications for AIGC [J]. Modern Film Technology, 2024, (06):31-38.

[2] Liu Zheliang, Zhu Wei, Yuan Ziyang Combining fully convolutional networks and CycleGAN for image instance style transfer [J]. Chinese Journal of Image and Graphics, 2019,24 (08): 1283-1291

[3] Huang Dongjin, Liu Jinhua, Wang Qian, etc Innovation of Artificial Intelligence Generated Content in Film Creation: ECCV 2024 AIGC Technology Review [J]. Modern Film Technology, 2024, (11):25-33.

[4] Wang Junfei Sora and Ke Ling: Comparative Analysis of Large Models in Wensheng Video [J]. Film and Television Production, 2024, 30 (12): 42-46

[5] Zhao Huirong The Application and Challenges of AIGC Technology in Art and Design Majors [J]. Journal of Suzhou Arts and Crafts Vocational and Technical College, 2024, (04):50-53.

[6] Zhang Aijun, Guo Zhenyu The myth of technology under the "black box worship": Sora's technological demystification and cognitive approach [J/OL]. Theoretical Construction, 1-16 [2020-01-21] http://kns.cnki.net/kcms/detail/34.1136.D.20241224.0845.006.html.

[7] Jia Huoran The Impact of Generative Artificial Intelligence (AIGC) on Animation Creation [J]. Journal of Suihua University, 2024, 44 (11): 48-50.