

Research on AI Generative Video Principles from a Roland Barthes Semiotic Perspective

Xinxue Liu¹, Lu Peng^{2*}, Huairen Zhang³

1347822398@qq.com¹, 772596833@qq.com^{2*}, zhanghuairen1999@126.com³

Beijing Institute of Fashion Technology, Beijing, China

Abstract. In the field of AI generative art, the public usually focuses on algorithms, data, and technology, but less on interpreting its essence from a semiotic perspective. Based on Roland Barthes' semiotic theory, the article takes AI video as a case study to delve into the semiotic system and decoding process therein, and to verify the applicability and explanatory power of semiotic analysis in understanding AI video. The article also explores the future development direction of human designers in this field, providing them with new thinking paths and methods. By focusing on semiotics, the meaning and expression of AI artworks and the interaction between technology and humanity can be better revealed.

Keywords: Semiotics, Ai Video, Sign Systems, Roland Barthes, Generative Art

1 Introduction

From the beginning of the 21st century, computing, data analytics, and artificial intelligence have gradually entered the field of aesthetics. Traditional video production requires the creativity and skill of the artist and consumes a lot of manpower and resources. But now, with techniques such as style migration and video diffusion modelling, computers are also able to generate high-quality artworks quickly, triggering thoughts about the nature of art creation and the role of the artist. The article introduces Barthesian semiotics to the field of AI video, exploring its semiotic system and the rule patterns of sign manipulation, helping the public to better understand the principles of the technology and revealing its creative process.

"Artificial Intelligence Art" refers to the field of artistic creation and expression through the use of Artificial Intelligence technologies and algorithms, the type of art that humans are unable to create due to the limitations of their bodies, brains, and other constraints [1]. It encompasses techniques such as deep learning and generative adversarial networks, which enable computers to generate, adapt or mimic the process of a work of art. Artificial intelligence art is created through rule-based algorithmic generation, data-driven learning models, or machine learning algorithms that learn from a large number of artworks to create new artworks in the category of "contemporary art".

The earliest attempts at AI art date back to 1950, when Labowski first showed distorted wavy lines on an oscilloscope in 1952, a work of art that failed to convey any readable information and was named Electronic Abstraction. From the beginning of the 21st century, computation, data analysis, and artificial intelligence gradually entered the field of aesthetics, with computer

scientists inviting art historians to evaluate images generated by neural networks mimicking the style of a particular artist [2]. In September 2022, the China Academy of Information and Communications Technology (CAICT) and the Jingdong Exploration Research Institute (JERI) jointly released the White Paper on Artificial Intelligence Generated Content (AIGC), which defines AIGC as "a category of content that is both classified from the viewpoint of a content producer, a method of content production, and a collection of technologies used for the automated generation of content [3].

Roland Barthes, a renowned French philosopher and semiotician, introduced influential concepts in semiotics that explore the relationship between signs, meanings, and cultural contexts. His theories delve into the structure of signs, their components, and the layers of meaning they convey. This research will apply Barthes' semiotic theory to dissect the various layers of meaning embedded in AI-generated videos, thus shedding light on how these videos construct and convey their narratives.

2 Semiotic theory

Semiotics is a discipline that studies symbols and how they operate, and Roland Barthes is considered one of the key theorists of semiotics. Applying semiotic theories to the field of AI-generative art can help the public gain a deeper understanding of the principles and creative process of AI videos. Through the perspective of semiotics, the symbolic system and the rules of symbolic operation in AI-generative artworks can be analysed to reveal the artistic elements and expressions. In semiotics, a symbol is defined as a form of representation that represents a concrete thing, concept or meaning. Symbolic systems, through algorithms, transform input data into symbolic representations and use symbol manipulation rules to create, transform and combine symbols, which can be combinations of natural language, images, music and other forms, to produce works of art. For example, GAN can be used to learn the features of real artworks. Decoding of symbols is the process of transforming the surface form of a symbol into meaning. Different cultures and languages assign different meanings to symbols. Symbol decoding requires understanding the relationship between different symbols based on common cultural, linguistic and social backgrounds, taking into account the context and context of the symbols, as well as the interrelationships between symbols.

Generative AI, represented by ChatGPT, is a major change in the field of artificial intelligence, and as an AIGC-specific application, it will not only make life easier and richer, but will also change the mode of production and innovation and improve social productivity [4]. At present, the biggest difference between AIGC and "professional content" and "user-produced content" is that new technologies have created a new type of content creation [5]. It is also a core tool in the Web 3.0 era, assisting designers in generating original design sketches, optimising the image processing process, and automatically generating visualisation effects. Major technology companies around the world have also actively embraced AIGC and continue to launch related technologies. For example, Phaida face mapping software StyleGAN, Runway Gen-2 text-generated video technology, and so on (Table 1).

Table 1. Some famous AI painting products

Time	Name	Use	Company	parameter scale	Related Literature
2018.12	StyleGAN	Face Mapping	NVIDIA	481 billion	1.555 million
2022.09	Dall-E2	Text to Images	OpenAI	12 billion	/
2022.03	Mjdney	Text to Images	Mjdney	/	0.03 million
2022.05	IMAGEN	Text to Images	Google	20 billion	/
2022.10	Phenaki	Text to Video	Google	1.8 billion	/
2023.02	Gen-1	Video to Video	Runway	Unknown	/
2023.03	Gen-2	Text to Video	Runway	1.7 billion	0.003 million

To improve the quality and diversity of generative paintings, researchers have also attempted to improve models and techniques to move from blurry, low-resolution images to highly realistic and aesthetically pleasing images, contributing to the rise of synthetic media. Large language models and models that share the latent space of textual images also support new ways of interacting with software and synthesising media, such as text-guided generative diffusion models unlocking powerful image creation and editing tools. Patrick Esser et al. propose a structure- and content-guided video propagation model that edits video based on the visual or textual descriptions of the output, and the use of a model with different levels of detail of the Training monocular depth estimation can provide control over structure and content fidelity [6]. Fundamentally, the essence of a generative algorithm is to model the distribution of the training samples and then draw new samples based on the model [7].

Elgammal A et al. proposed a Creative Adversarial Network (CAN) model that generates artful images by learning deviations from artistic styles and norms [8]. In the study by Ma L et al. a pose-guided character image generation method is presented, in which character images with artistic effects are generated by learning symbolic representations of human poses and gestures [9]. Vasilenko E et al. provide an overview of research on the use of AI to generate artworks, covering different AI techniques and algorithms for art creation and discussing concepts related to semiotics, such as meaning and symbol systems [10]. Ma Lixin et al. proposed that the Law of Artistic Reality should be taken as the fundamental guideline for all artistic creations, integrating and compatibilising the aesthetic advantages of the first two generations of digital art and all atomic art, and realising the computability of the Law of Artistic Reality [11].

Numerous scholars have explored the connotation and relationship between AI and art creation from the theoretical perspectives of culture and art, however, the research on the semiotic application of AI under the combination of practice cases is still relatively weak, mostly generalised, lacking empirical research and substantive exploration of the integration path and development mechanism. Semiotics provides a new perspective for the public to understand AI painting. By studying symbolic systems, symbolic manipulation and symbolic decoding, designers can explore the principles and creative processes of AI video in greater depth, and apply semiotic theories to improve and develop AI painting techniques and explore the integration of AI and art.

3 The Principle of AI Video in the Semiotic Perspective of Roland Barthes

3.1 Symbolic system

Barthes introduces the concepts of "signifier" and "signified" to explain levels of symbols, emphasizing implicit meanings. The referent is the symbol's material aspect, while denotation is its represented thing. AI-generated videos use symbols like pixels, colors, text, audio, etc., combining to express concepts, emotions. Meaning arises from the symbols and viewer's interpretation, influenced by culture. Symbols' choice shapes AI video meaning. Like film metaphors, AI video symbols can be metaphorical. Analyzing symbol combinations reveals themes, emotions, and intentions in videos. In addition, the analysis of symbol systems can focus on AI algorithms and programming codes. These codes guide the way elements are generated and presented, similar to the technical means of cinema. A deeper understanding of how these codes affect the generation process and the final work can help reveal the inner mechanisms of AI video creation (Table 2).

Table 2. Symbol system for AI videos

Symbol	Signifier	Denotation	Connotation
Ai generates the objective content of the video	Video Subject	Ai generates video visualisations	Ai generates the hidden message (emotion) behind the video

3.2 Symbolic generation

conveying concepts and expressions. AI learns from image data to establish symbol relationships through algorithms and data processing. This complex process encompasses understanding video data, feature extraction, and algorithm tuning. AI models are trained on diverse video data, learning object features and converting them into numerical values. AI adds animation effects and controls symbol combinations by adjusting algorithm parameters, aiming to achieve the designer's intent or a specific style. Human designers play a crucial role in guiding AI-generated video creation, as AI's creativity and outcomes depend on dataset and algorithm variations.

3.3 Symbolic decoding

Based on the pioneering ideas of semiotics and film semiotics, such as Saussure, Roland Barthes, Pierce, etc., Metz has become a master in the field of film semiotics. He synthesised the theories of these pioneers, and also integrated Lacan's "mirror theory", injecting a deeper insight into film semiotics. AI-generated video creation uses algorithms and data to generate images and sounds, but the symbolic system and decoding of meanings behind it can also be examined from the perspective of film semiotics. In *Film Semiotics and Aesthetic Thought*, Lotman's perspective emphasizes the key position of symbols in human culture, especially as manifested in the realm of art. The process of interconverting symbols becomes the basis of human understanding of the world, and this conversion is particularly significant in AI video. Just as art as part of a cultural system requires different types of symbols for expression, AI video as a modern art form also requires symbols to convey meaning.

The content of a traditional narrative film is likened to a set of axes, a set of vertical axes (v_1, v_n), and a set of horizontal axes (h_1, h_n), which dynamically intersect at any position between h_1, \dots, h_n and v_1, \dots, v_n [12]. Each of these horizontal axes represents a different plot line, and each vertical axis represents a different theme or element. Through these intersections, the narrative and plot of the film intertwine on multiple levels to create a colourful story world (Figure 1).

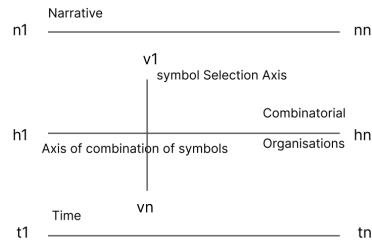


Fig. 1. Traditional narrative film approach

Currently the mainstream ai video there are two main forms of video generated video and text generated video.

video generated video. Utilize AI to transform images through text-generated modifications. Combine text and video clips for novel visual experiences in four methods: adapt video frames with text styles, alter video style with input images, modify subjects with text prompts, and change subjects while retaining the background for narrative information conveyance.

Text Generated Video.

Composition involves crafting expressive visuals from symbols. In literature, linguistic elements like speech, grammar, and fonts become key in conveying meaning due to their pictorial nature. Phrases like "sad girl on a rainy day, clay animation" transform into videos through computer language, using diffusion for high-quality images and transformers for text comprehension. This process turns user descriptions into realistic visuals. (Figure 2).

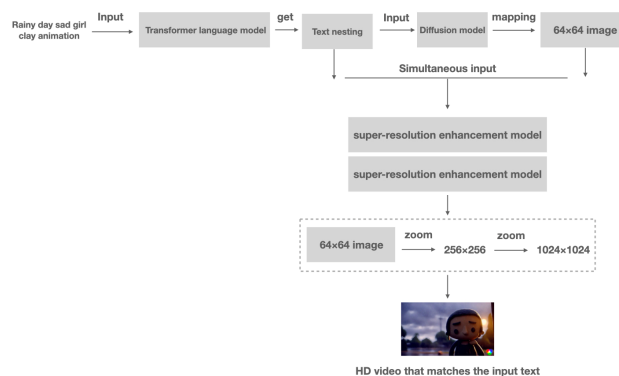


Fig. 2. AI video model in computer language

The sentence "Sad girl on a rainy day, clay animation" is divided into four parts: "rainy day", "sadness", "girl", and "clay animation" under semiotic language. "rainy day", "sadness", "girl", and "clay animation". The human designer inputs the above linguistic symbols into the computer, and the AI provides additional information for the phrases and devises various combinations, such as "drizzle", "blues", "twisted braid", "stone braid", "stone braid", "sadness", "girl", and "clay animation". The AI provides additional information to the phrases and designs various combinations, such as "drizzle", "blues", "twisted braid", and "stoneware clay", and predicts the user's video preferences, generating a huge number of AI videos. This process requires designers to intervene and choose the best of the best, constituting the most basic semiotic structure of image narratives (Figure 3).

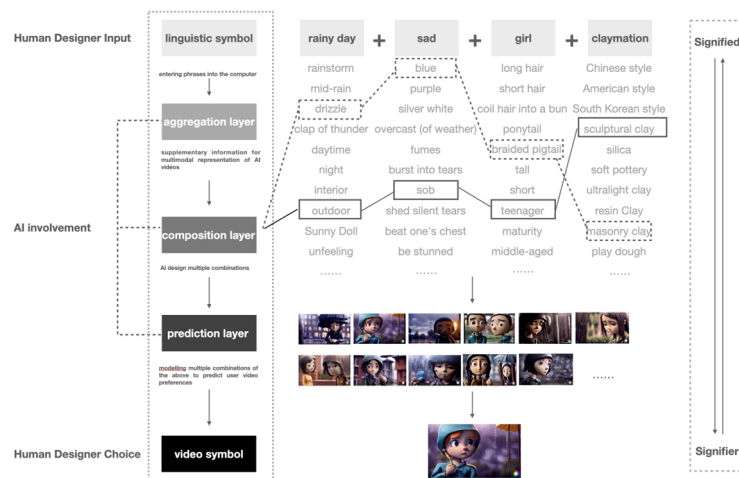


Fig. 3. AI video model in semiotic language

Traditional video production needs to go through a complex process of script creativity, sub-shot drawing, character setting and production, props and scene production, frame-by-frame shooting, post-synthesis, etc., which is time-consuming and costly, and humans can't make a large number of styles in a short period of time, but AI can quickly generate countless styles, which greatly improves the efficiency of video production (Figure 4).

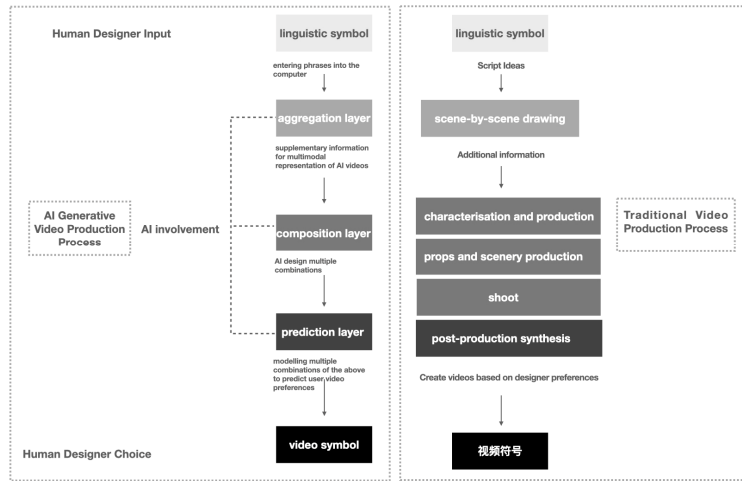


Fig. 4. Comparison between AI video and traditional video production

In a symbol system, there are relationships and connections between symbols. These relationships can be semantic, such as the spatial relationships of objects in images and the semantic associations of words in text; they can also be visual connections, such as the transitions between frames in animation and the rhythmic changes of notes in audio. These relationships and connections between symbols constitute the structure and coherence in AI videos. Barthes believes that symbols are polysemous, i.e., they can have different meanings in different contexts. In AI video generation, algorithms learn to create symbols to generate videos based on input text or images. Since AI may introduce multiple interpretations when processing semantics and images, the generated video may present diverse meanings depending on the input. Consider the interaction between the AI video creator and the generation algorithm. The creator actually influences the generation of symbols when selecting training data, adjusting parameters and guiding the generation process. Analysing this creator-algorithm interaction can reveal how the creator shapes the symbol generation process by intervening and guiding it to achieve specific intentions and expressions.

3.4 The stylistic dimension of symbolic representation

Third Semiotics is a concept conceived by Barthes, which is mainly concerned with the "stylistic" dimension of symbolic representation, emphasizing that symbolic expression is not only limited to transmitting information, but also encompasses creative artistry and unique visual presentation. The "stylistic" dimension of symbolic representation refers to the unique artistic style and visual characteristics of video works generated through AI technology. This concept draws on Barthes and the third semiotic theory proposed by film theorist Wang Zhimin, which emphasizes that symbolic expression is not only limited to transmitting information, but also encompasses the artistry of creation and unique visual presentation [13]. In recent years, the concept of "language-image co-rule" has been put forward in the field of semiotics in China, which provides a basic assumption for re-understanding the film image as a symbolic perspective.

In AI video creation, the "style" dimension focuses on how to create a unique artistic style through symbols, images and sounds, etc. AI technology can learn and imitate the creative styles of different artists and film directors, so as to generate video works with specific styles. These styles can cover a wide range of aspects such as the use of colours, picture composition, camera movement, sound effect processing, etc., forming a unique and recognizable visual and auditory sign. AI-generated video works can be realised to simulate different film styles in a short period of time, ranging from black-and-white classic film styles to modern visual effects, from realism to abstract expressionism, and so on. These stylistic shifts go beyond simple visual changes and involve the expression of emotions, sentiments and themes. By linking AI-generated videos to the dimension of "style" in film semiotics, we can gain a deeper understanding of how AI video creation can use symbols to create diverse artistic styles, enriching the possibilities of artistic creation.

4 The future thinking of AI painting and human design

Although AI has a certain degree of instrumentality in artistic creation, there are limitations and limited subject participation. For example, at this stage, AI text generation videos are not fine enough to accurately portray close-up views of hand movements, stiff facial expressions, bizarre colour tones, and other defects. AI is usually based on the extension of existing data, and it is difficult to actively create styles, and it still requires the human artist to create styles, which also involves relevant copyright and ethical issues. In addition, AI has limitations in understanding metaphorical level expressions and symbolic connotations, involving complex elements such as culture, emotion and subjective understanding, thus still requiring humans to inject unique emotions and meanings into their works.

Humans can't be fully replaced by technology due to unchangeable values shaped by culture, generations, and consciousness. Generational shifts bring unique creativity, needs, and self-awareness that AI can't mimic. Innovation spans imagination, problem-solving, user experience, and more, offering fresh solutions and behavior changes. Human designers can use AI drawing technology to assist in their creations, yet the final work of art still requires a unique expression of the artist's creativity and personal style. Technology can provide more possibilities for creation, but the essence of creation still stems from human creativity and emotional expression. Regardless of the development of AI technology, the unique inspiration and human emotion behind the artwork will always remain the irreplaceable core.

5 Conclusion

Based on the theoretical foundation of Barthesian semiotics, the article digs deeper into the symbolic system and symbolic decoding behind AI-generated videos. AI videos are based on designers' inputs and analyses of large-scale data, from which patterns and trends are extracted to rapidly generate a variety of artistic styles, forms, and expressions, which provide artists with more creative choices and inspirations, speeding up the speed and efficiency of the design process, and broadening their creative boundaries. In the future, designers will work closely with AI and emerging technologies to create more innovative and impactful design solutions, while also focusing on social and environmental sustainability.

Acknowledgments. This study was supported by a grant from the 2023 Graduate Student Research and Innovation Programme of Beijing Institute of Fashion Technology (Grant No. X2023-064).

References

- [1] Lev Manovich, et al. Lev Manovich: Art and Aesthetics of Artificial Intelligence (AI)[J]. World Cinema,(03):4-24 (2023)
- [2] Artsiom Sanakoyeu,Dmytro Kotovenko,Sabine Lang,Björn Ommer, "A Style-Aware Content Loss for Real-time HD Style Transfer, "https://arxiv.org/abs/1807.10201.
- [3] China Academy of Information and Communication Research, Jingdong Discovery Institute. Artificial Intelligence Generated Content (AIGC) White Paper [R/OL].(2022-09-02)[2022-11-05].http://www.caict.ac.cn/sytj/202209/P020220913580752910299.pdf. China Academy of Information and Communication,Jingdong Exploration Research Institute.AIGenerated Content(AIGC)White Paper [R/OL]. (2022-09-02) (2022-11-05)
- [4] Yongwei Chen. Beyond ChatGPT: Opportunities, Risks and Challenges of Generative AI [J].Journal of Shandong University(Philosophy and Social Sciences), No.258(03):127-143 (2023)
- [5] Baiyang Li, et al. The technical Features and Aromorphosis of Artificial Intelligence Generated Content (AIGC)[J]. Documentation,Information & Knowledge,40(01):66-74 (2023)
- [6] Esser P, Chiu J, Atighehchian P, et al. Structure and content-guided video synthesis with diffusion models[J]. arXiv preprint arXiv:2302.03011 (2023)
- [7] Bond-Taylor S.,et al. "Deep Generative Modelling:A Comparative Review of Vaes,Gans, Normalizing Flows, Energy-based and Autoregressive Models", IEEE Transactions on Pattern Analysis and Machine Intelligence (2021)
- [8] Elgammal A, Liu B, Elhoseiny M, et al. Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms[J]. arXiv preprint arXiv:1706.07068, (2017)
- [9] Ma L, Jia X, Sun Q, et al. Pose guided person image generation[J]. Advances in neural information processing systems, 30 (2017)
- [10] Vasilenko E, Vasilenko P, Saenko N, et al. Engineering aspect of modern concept of professional education of artists and designers in academic figure[J]. International Journal of Engineering Research and Technology, 13(11): 3625-3630 (2020)
- [11] Lixin Ma, Shaohui Tu. AI art creation mechanism research[J]. Art Research, No.204(06):82-86 (2022)
- [12] Kiran Tomaselli, et al. Semiology of Film, Space-time Theory, and the Semiotics of Virtual Reality[J]. Journal of Shanghai Jiaotong University(Philosophy and Social Sciences),30(04):146-159 (2022)
- [13] Zhimin Wang. Principles of Film Aesthetic Analysis [M]. Beijing: China Film Press, 7:150-156 (1997)