2024 International Conference on Artificial Intelligence, Deep Learning and Neural Networks (AIDLNN) | 979-8-3315-2081-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/AIDLNN65358.2024.00035

The Application of AI Video Generation Technology in Virtual Reality (VR) and Augmented Reality (AR)

Biqin Deng
Guangzhou Institute of Science and Technology, Guangzhou 510540, China
51935959@qq.com

***Abstract*-AI has become one of the key forces driving social progress, particularly in the fields of VR and AR, where the application of AI video generation technology is spearheading a technological revolution. This article delves into the specific applications, potential impacts, and future development trends of AI video generation technology in VR and AR. VR technology enables users to fully immerse themselves and interact with virtual environments. AR technology, on the other hand, overlays virtual information onto the real world, integrating virtual objects and information into the user's actual surroundings. AI video generation technology plays a crucial role in these two domains. In the VR sector, AI video generation technology, leveraging deep learning techniques, can learn and mimic human creative styles, automatically generating video content. Through image processing and machine learning technologies, this article explores how AI can generate high-quality virtual objects and scenes, making them more realistic and seamlessly integrated with the real world. Additionally, this article analyzes the challenges and opportunities faced by AI video generation technology in VR and AR applications, such as technological bottlenecks, data privacy, and security issues. AI video generation technology is poised to play an increasingly significant role in the VR and AR sectors.**

Keywords-Deep Learning, AI, Virtual Reality, Augmented Reality, Video Generation

I. Introduction

VR and AR technologies, with their unique interactivity and immersive experience, are transforming the way people interact with the world. The application of AI video generation technology in VR and AR is increasingly gaining attention. VR technology creates a three-dimensional virtual environment through computer simulation, allowing users to interact with the virtual environment by wearing VR devices. The implementation of VR technology relies on multiple technologies, including computer graphics, sensor technology, audio processing, and human-computer interaction. Scholars' research on VR technology mainly focuses on the following aspects: Yang, Y. et al. studied the implementation principles and methods of VR technology, including the construction, rendering, and interaction of virtual environments [1-2]. Wang, Q. et al. studied the application scenarios of VR technology, such as education, healthcare, entertainment, etc [3]. Zhang, H et al. studied the user experience of VR technology, including comfort, immersion, and interactivity [4]. These studies provide important theoretical support and practical guidance for the development of VR technology.

AR technology is a technique that superimposes virtual information onto the real environment. Through devices such as smartphones, tablet computers, or specialized glasses, AR technology can integrate virtual objects and information into the user's real environment. The core of AR technology lies in enhancing users' perception of the real world and providing additional information and interaction. Scholars' research on AR technology mainly focuses on the following aspects: Smith, J. et al. has studied AR technology's image recognition, three-dimensional modeling, and positioning technology [5]. Liu, Z. et al. has researched AR technology's application scenarios, such as navigation, education, entertainment, etc [6]. These studies provide important theoretical support and practical guidance for the development of AR technology.

AI video generation technology utilizes deep learning, computer vision, and other technologies to automatically generate realistic video content. The application of AI video generation technology in VR and AR provides users with richer and more vivid virtual experiences. Scholars' research on AI video generation technology mainly focuses on the following aspects. Kim, H. et al. studied the principles and methods of AI video generation technology, including the selection and optimization of deep learning models, the generation and editing of video content, etc [7]. Harvey, R. et al. studied the application scenarios of AI video generation technology, such as virtual reality movies, games, education, etc [8-9]. Vondrick, C. et al. studied the performance evaluation and optimization of AI video generation technology, including the quality, speed, and stability of generated videos [10].

Although scholars have made significant progress in the research of VR, AR, and AI video generation technologies, there are still some shortcomings. Firstly, issues such as comfort, resolution, and latency of hardware devices limit the widespread application of VR and AR technologies and the enhancement of user experience. Meanwhile, AI video generation technology also faces some technical challenges, such as the realism and fluency of generated videos. These technological bottlenecks require further research and solutions. Secondly, VR and AR technologies require the collection and use of users' personal information, such as location, direction, and actions. These data privacy and security issues have always been a focal point. With the continuous development of technology, how to protect users' privacy and data security has become an urgent problem to be solved. Meanwhile, AI video generation technology also involves data privacy and security issues, requiring further strengthening of data protection measures.

In response to the deficiencies in existing research, this paper adopts an interdisciplinary research perspective, combining VR, AR, and AI video generation technologies for in-depth exploration and discussion. This interdisciplinary approach helps break down barriers between fields and promotes the cross-integration and innovative development of

979-8-3315-2081-6/24/$31.00 ©2024 IEEE
DOI 10.1109/AIDLNN65358.2024.00035

technologies. This paper not only delves into the principles and methods of VR, AR, and AI video generation technologies but also conducts case studies and empirical research in conjunction with practical application scenarios. This combination of theory and practice helps verify the feasibility and effectiveness of the technologies, providing strong support for their promotion and application. This article aims to delve into the application of AI video generation technology in VR and AR, providing strong support for the optimization and improvement of the technology. Through interdisciplinary research perspectives, a combination of theoretical and practical research methods, user experience research, as well as innovative and forward-looking research methods, this article will offer valuable references and insights for the future development of VR, AR, and AI video generation technology.

II. AI Video Generation Technology Model

A. Theoretical Model Architecture

AI video generation technology utilizes deep learning, computer vision, and other technologies to automatically generate high-quality video content based on given text, image, or video data. In the field of VR and AR, this technology can be used to generate realistic virtual scenes, dynamic elements in augmented reality experiences, etc., greatly enhancing users' immersion and interactive experience. The architecture of the AI video generation technology model is shown in Figure 1.

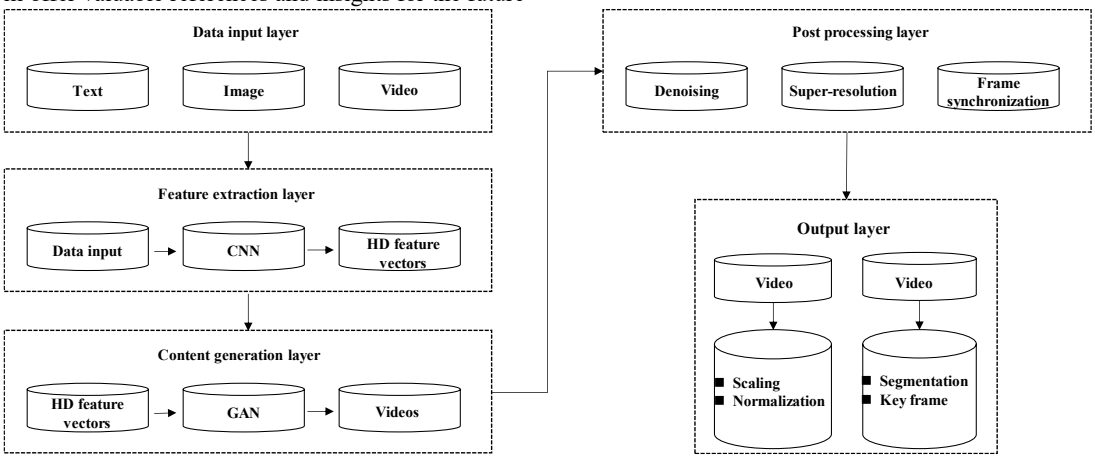

Figure 1 Process of AI-based model construction

In Figure 1, the data input layer includes various modalities of data input such as text, images, and videos, which are used to provide raw materials for AI video generation. The feature extraction layer utilizes deep learning models (such as CNN) to extract features from input data and generate high-dimensional feature vectors. The forward propagation calculation process of CNN is shown in Equation 1 below.

$$f = \alpha(W^T\theta + \beta) \tag{1}$$

Among them, θ is the input feature vector, W is the convolution kernel matrix, β is the bias term, and α is the activation function (such as ReLU). The content generation layer is based on the extracted feature vectors and uses Generative Adversarial Networks (GANs) or other generation models to generate video content that meets the requirements. The calculation of the loss function of GAN is shown in Equation 2 below.

$$L_{GAN}(S,P) = E_{r\sim r_d}[logP(r)] + E_{s\sim z_d}[\log(1 - P(S(z)))] \tag{2}$$

Among them, S is the generator, P is the discriminator, r is the real data, z is the random noise vector, and r_d and z_d are the distributions of the real data and noise vector, respectively. The post-processing layer performs post-processing on the generated video, including denoising, super-resolution, frame synchronization, etc., to improve video quality. The loss calculation for video frame generation is shown in Equation 3 below.

$$L_f = \sum_{i=1}^{I-1}\left\|Z(X_i, X_{i+1}) - \hat{Z}(X_i, X_{i+1})\right\|_2^2 \tag{3}$$

Among them, $Z(X_i, X_{i+1})$ is the real optical flow field, $\hat{Z}(X_i, X_{i+1})$ is the predicted optical flow field, and X_i and X_{i+1} are adjacent frames of images, respectively. The loss calculation of video super-resolution is shown in Equation 4.

$$L_{SR} = \|G(X_{DR}) - D_{GR}\|_2^2 \tag{4}$$

Among them, G is the super-resolution model, X_{DR} is the low-resolution image, and D_{GR} is the high-resolution image. The calculation process of frame synchronization loss is shown in Equation 5 below.

$$L_{syrn} = \sum_{i=1}^{N}\|B(v_i) - v_{i+1}\|_2^2 \tag{5}$$

Among them, B is the time transformation function, v_i and v_{i+1} are the feature vectors of adjacent frames, and N is the frame number. The final output layer outputs the processed video content to VR or AR devices for users to watch and interact with.

B. Construction Process of AI Video Generation Model

Firstly, in terms of data preparation, collect and preprocess input data such as text, images, and videos, including steps such as data cleaning and normalization. At the model training

level, use large-scale datasets to train deep learning models, including feature extraction models and generative models. During the training process, it is necessary to continuously optimize the model parameters to improve the quality and diversity of the generated videos. Secondly, input the preprocessed input data into the trained model to generate video content that meets the requirements. This step may require multiple iterations and optimizations to ensure that the generated video content meets expectations. Post process the generated video to improve its quality and realism. This step may include various techniques such as denoising, super-resolution, frame synchronization, etc. Finally, in terms of output and interaction dimensions, it is necessary to design a reasonable interaction mechanism during the process of outputting processed video content to VR or AR devices, so that users can naturally interact with the generated virtual scene.

C. Application and Optimization of Model

In order to verify the effectiveness of the theoretical model proposed in this article, further analysis is needed on the realism and interactivity of dynamic elements in virtual scenes and augmented reality experiences generated using AI video generation technology. By optimizing model parameters and post-processing steps, the quality and diversity of generated videos can be further improved. During the verification process, potential issues and challenges were identified in the model. For example, due to data sparsity and noise, the training process of deep learning models may be disrupted. In addition, the generated video content may have issues such as artifacts or distortion in some cases. To address these issues, it is necessary to further optimize the model architecture and training strategy, and explore new post-processing techniques and interaction mechanisms.

At the optimization level of the data input layer, the input data is enhanced, such as rotation, scaling, flipping, etc., to increase the diversity and generalization ability of the data. The use of data augmentation techniques can generate more training samples, thereby helping the model better learn the inherent patterns of the data. Remove noise and redundant information from input data to improve data quality. Normalize the data to meet the input requirements of the model. At the optimization level of feature extraction, more advanced deep learning models such as ResNet, EfficientNet, etc. are selected to improve the accuracy and efficiency of feature extraction. Select the appropriate model architecture and parameters based on the specific requirements of the task. In addition, using pre trained models for transfer learning can save training time and computational resources. By fine-tuning the pre trained model to better adapt to video generation tasks in VR and AR.

III. Application of AI Video Generation Technology in VR and AR

A. Application of AI Video Generation Technology in VR

As an advanced computer simulation system, VR can create and immerse users in a highly realistic virtual world. The application of AI video generation technology in VR mainly relies on deep learning algorithms and computer graphics technology to generate realistic virtual scenes and interactive content. At the level of 3D scene generation, AI can learn from a large amount of real-world data to generate realistic 3D scenes. These scenes can include natural environments, architectural models, character roles, etc. Secondly, at the level of interactive content generation, AI can generate corresponding virtual content in real-time based on user interaction behavior. For example, when a user interacts with a virtual character, AI can generate the character's reactions and conversations. In addition, at the level of real-time rendering and synthesis, AI technology can achieve efficient real-time rendering and synthesis, ensuring the smoothness and realism of VR scenes. The application process of AI video generation technology in VR is shown in Figure 2.

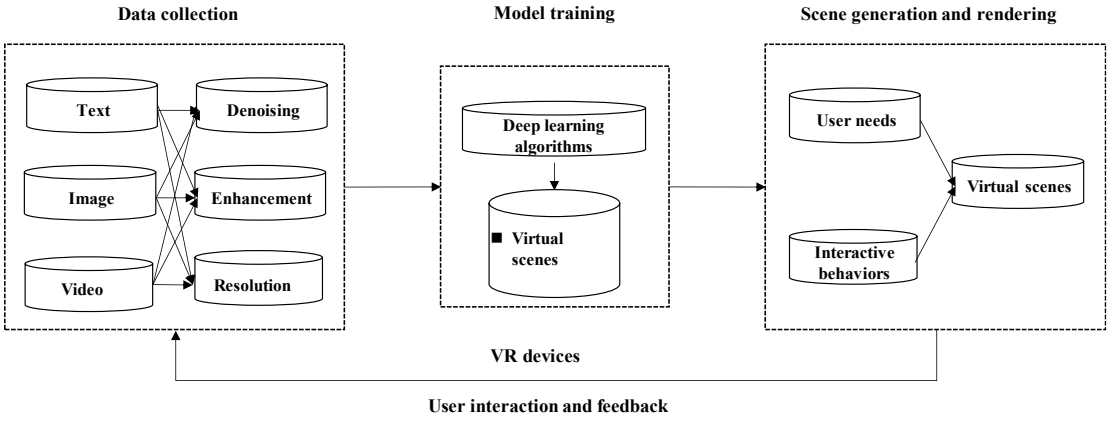

Figure 2 Application process of AI video generation technology in VR

As shown in figure 2, in the dimension of data collection and preprocessing, a large amount of real-world image, video, and audio data is collected and preprocessed, including denoising, enhancement, etc. Using deep learning algorithms to train preprocessed data and generate models capable of generating virtual scenes. In the dimension of scene generation and rendering, virtual scenes are generated based on user needs and interactive behaviors, and real-time rendering and

synthesis are performed. Users enter virtual scenes through VR devices, interact with virtual content, and AI adjusts and optimizes based on user feedback.

B. Application of AI Video Generation Technology in AR

As a technology that superimposes virtual information onto the real environment, the application of AI video generation technology in AR is mainly achieved through computer vision and deep learning algorithms to perceive the real world and superimpose virtual information. AI can perceive the user's environment in real-time through cameras and sensors, including location, direction, and the shape and size of surrounding objects. Based on the perceived environmental information, AI can accurately overlay virtual information onto the real world, achieving the effect of augmented reality. Users can interact with virtual information through AR devices, and AI adjusts and optimizes based on user feedback. Collect environmental information using cameras and sensors, and process and analyze it through AI algorithms. At the level of virtual information generation and overlay, corresponding virtual information is generated based on processed environmental information and overlaid onto the real world. Finally, users interact with virtual information through AR devices, and AI adjusts and optimizes in real-time based on user feedback. The schematic process of AR user interaction is shown in Figure 3.

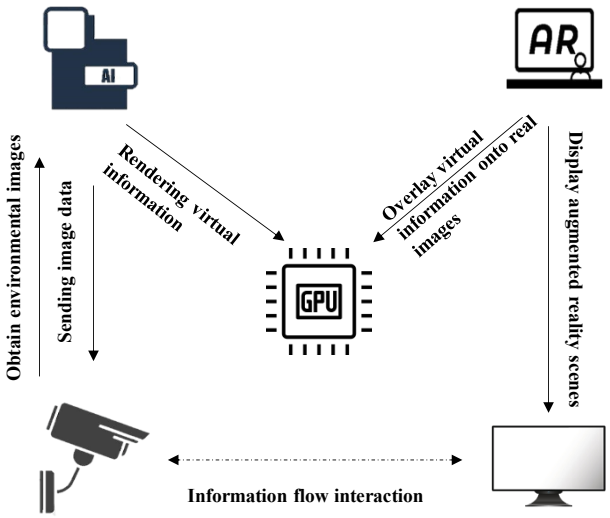

Figure 3 The schematic process of AR user interaction

The lightweighting and portability of hardware devices are important directions for the improvement of VR and AR devices. Being more lightweight, comfortable, and affordable will greatly promote the popularization and application of VR and AR technology. VR and AR devices with higher resolution and lower latency, enhancing visual immersion and smooth interaction, combined with technologies such as 5G, will significantly improve data transmission speed and reduce latency issues between devices. AI technology will be deeply integrated with VR and AR to achieve more intelligent understanding of virtual characters and scenes, enhancing users' interactive experience.

IV. Application Effect of AI Video Generation Technology

A. Application Effect of AI Video Generation in VR

VR technology simulates computer-generated 3D images, allowing users to immerse themselves in digital environments for interaction. The application of AI video generation technology in VR not only enhances user experience, but also provides more creative possibilities for content creators. 360 degree panoramic video is an important component of VR applications, allowing users to freely choose their viewing angle and immerse themselves in the virtual world. AI video generation technology can efficiently generate high-quality 360 degree panoramic videos, greatly improving the efficiency and quality of content production. The performance data of VR scene generation and scene rendering are shown in Table 1.

Table 1 Performance data of VR scene generation and scene rendering

Data Type	Data Volume	Preprocessing Time	Frame Rate	Rendering Time
Image	8000	1.5h	40	8ms
	9000	1.8h	50	12ms
	10000	2.0h	60	16ms
Audio	500	1.0h	15	32ms
	800	1.6h	20	45ms

	1000	2.0h	30	64ms
	200	1.0h	15	64ms
Video	400	2.0h	20	128ms
	600	3.0h	30	256ms

From Table 1, it can be seen that the data volume and preprocessing time of the video are significantly higher than those of the image and audio, and the frame rate and rendering effort of the data have also been improved. AI video generation technology can also simulate virtual theaters and live streaming activities, allowing users to feel the atmosphere and interactive experience of the scene. In scenarios such as concerts and sports competitions, VR devices combined with AI generated video content can allow users to immerse themselves in the lively atmosphere of the scene. In addition, with the popularity of smart homes, consumers' demand for smart home products is also increasing. AI video generation technology can generate operation guides and usage tutorials for smart home products, helping consumers better understand and use the products.

B. Application Effect of AI Video Generation in AR

AI video generation technology can create interactive AR short video content, where viewers can interact with AR elements in the video through their phone screens, such as clicking, dragging, or rotating virtual objects. This interactivity not only increases user engagement, but also provides new creative inspiration and possibilities for creators. Interactive short video content data includes interaction types, user engagement, creative efficiency, application scenarios, and technology providers. In tourism short videos, creators can add virtual tour guides to guide viewers to explore the history and culture of the attractions; In educational short videos, AR elements can help students better understand abstract concepts. This interactive content not only increases user engagement, but also enhances the fun and educational value of the content. The performance data of AR environment perception and virtual information overlay under the support of AI video generation technology is shown in Table 2.

Table 2 Performance data of AR scene generation and scene rendering

	Data Type	Data Volume	Processing Time	Rendering Time
Image	Sensor Data	3000	3.0s	8ms
	position information	5000	5.0s	16ms
Audio	Sensor Data	800	2.5s	15ms
	position information	1000	3.0s	20ms
Video	Sensor Data	500	2.5s	18ms
	position information	1000	5.0h	36ms

From Table 2, it can also be seen that the data volume and preprocessing time of the video are significantly higher than those of the image and audio, and the frame rate and rendering effort of the data have also been improved. AR technology combined with AI video generation technology can create virtual makeup and fitting experiences. Users can preview different makeup and clothing effects through AR filters and special effects without actually applying makeup or trying on clothes. The application of AI video generation technology in AR can also be extended to the fields of virtual meetings and social interaction. Through AR technology, users can experience more realistic interactive experiences in virtual meetings, such as eye contact, gesture recognition, etc. In remote work and online education, the combination of AR technology and AI video generation technology can create more realistic interactive experiences, improve communication efficiency and teaching quality. In the field of social entertainment, AR technology can also bring users more interesting interactive experiences and enhance user stickiness.

C. Application Optimization of AI Video Generation Technology

The application optimization of AI video generation technology in VR and AR mainly includes the optimization of data input layer, feature extraction layer, content generation layer, post-processing layer, overall architecture and training strategy, and interaction mechanism. The application optimization architecture of AI video generation technology is shown in Figure 4. Among them, the optimization of the data input layer is divided into two dimensions: data augmentation and data cleaning. The optimization of the feature extraction layer includes model selection, transfer learning, and other aspects. The optimization of the content generation layer mainly includes the selection and improvement of the generation model, multimodal fusion, and other aspects.

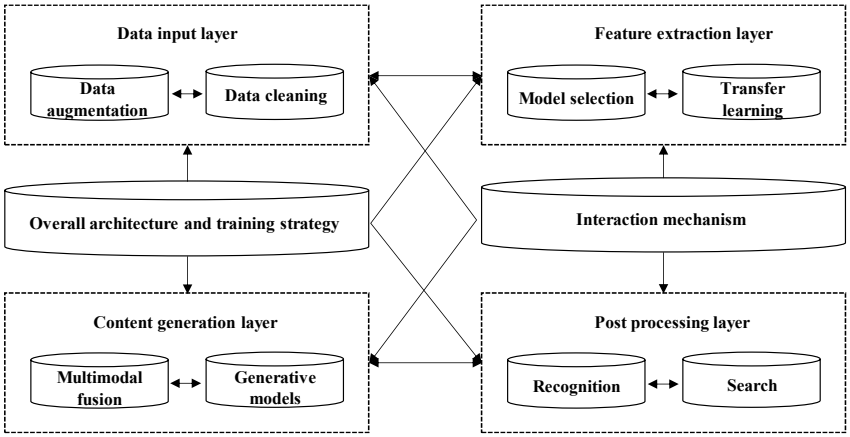

Figure 4 The optimized architecture of AI video generation

In addition, the optimization of the overall architecture and training strategy includes several dimensions such as distributed training, model compression and quantization, adaptive learning rate, and early stopping method. The optimization of interaction mechanisms mainly focuses on two aspects: natural interaction and real-time feedback. In summary, by optimizing the data input layer, feature extraction layer, content generation layer, post-processing layer, as well as the overall architecture and training strategy, the application effect of AI video generation technology in VR and AR can be further improved. Meanwhile, by optimizing the interaction mechanism, users can be provided with a more natural, intuitive, and rich interactive experience.

V. Conclusions

This paper delves into the application effects of AI video generation technology in the fields of VR and AR. Through detailed analysis of multiple application examples and data tables, it fully demonstrates how AI technology can bring revolutionary changes to VR and AR experiences. Firstly, the application of AI video generation technology in the VR field has greatly improved the efficiency and quality of generating 360-degree panoramic videos. Secondly, in the field of AR, AI video generation technology has also played an important role. By combining AR technology, AI can generate interactive short video content, allowing viewers to interact with virtual elements through their mobile screens, thereby increasing user engagement and satisfaction. In addition, AI technology is also applied in virtual makeup and fitting scenarios, providing users with a more convenient and authentic shopping experience. These applications not only enhance the shopping experience for users, but also increase the conversion rate and user stickiness of e-commerce platforms. In summary, the application effect of AI video generation technology in the VR and AR fields is significant, not only improving user experience and participation, but also providing more creative possibilities for content creators.

References

[1] Yang, Y., Yu, M., & Jiang, G. Research on 3D mobile phone stereo image synthesis algorithm and virtual viewpoint synthesis algorithm based on AIV. Journal of Computer-Aided Design & Computer Graphics, 2019, 21(3), 337-343.

[2] OpenAI. Video Generation Models as World Simulators. Technical Report, OpenAI:2024, 1409.1556.

[3] Wang, Q., Pan, D., Li, D., & Liu, H. Study on cross-talk and evaluation method of autostereoscopic displays. Optoelectronics·Laser, 2020, 21(5), 767-771.

[4] Zhang, H., Xu, H., & Liu, H. AI-Generated Videos: Techniques, Applications, and Challenges. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2023, 19(2), 1-23.

[5] Smith, J., & Owens, A. Deep Learning for Video Synthesis and Editing: A Survey. ACM Transactions on Graphics (TOG), 2019, 38(6), 1-34.

[6] Liu, Z., Ye, X., & Liu, Y. Video Synthesis with Spatio-Temporal Transformer Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 44(11), 7257-7270.

[7] Kim, H., & Cha, M. Neural Video Synthesis: From Static Images to Dynamic Videos. ACM Transactions on Graphics (TOG), 2021, 40(6), 1-12.

[8] Harvey, R., & De Rose, T. The Role of AI in Film and Television Production. Journal of Media Practice, 2020, 21(1), 5-20.

[9] Su, J., Zhu, X., & Zhang, C. AI-Driven Video Creation: Techniques, Challenges, and Future Directions. ACM Multimedia,In Proceedings of the IEEE conference on computer vision and pattern recognition 2024, (pp. 225-234).

[10] Vondrick, C., Pirsiavash, D., & Torralba, A. Generating Videos with Scene Dynamics. Advances in Neural Information Processing Systems, 2019, 29, 613-621.