Research on the Role Generation and Dynamic Expression Technology of Animation Short Play Based on Deep Learning

Jianzhao Sha[1]

[1] *Liaoning Communication University, Shenyang 110000, China*
[1] *Email: 493752795@qq.com*

ABSTRACT
Aiming at the pain points of high cost, long cycle and difficult to maintain the consistency of roles in the production of short animation plays, this article proposes a theoretical framework of role generation and dynamic performance based on deep learning (DL). Firstly, the decoupling and reconstruction mechanism of potential spatial features is analyzed, and the theoretical cornerstone of role generation is established. Then, an algorithm of style transfer and identity preservation is constructed, and the constraint optimization strategy is used to solve the problem of multi-view consistency, so as to realize efficient role customization under the condition of small samples. At the level of dynamic performance, the loss constraint of time sequence consistency is introduced to ensure the fluency of action and the accuracy of emotional expression. This article studies the establishment of a multi-dimensional evaluation system covering visual fidelity and time sequence stability, and designs a closed-loop ethical risk assessment process to define copyright and ensure content compliance. The purpose of this theoretical framework is to lower the production threshold, ensure the stability of the role identity and the accurate transmission of dynamic emotions, and at the same time provide methodological support for the industrial production of animation short plays.

Keywords: *Deep learning, Anime short play, Role generation, Dynamic performance, Ethical evaluation*

1. INTRODUCTION

With the popularity of mobile Internet and the increasing demand for fragmented entertainment, animation short plays have become an important growth point in the field of network audio-visual and the market scale continues to expand [1]. However, the traditional animation production process is cumbersome, long-term and expensive, and it is difficult to meet the content demand of high frequency and fast iteration in the short drama market [2]. Especially in the role design and animation performance, manual drawing and binding are not only inefficient, but also difficult to ensure the high consistency of the role images in the series [3]. The characteristic drift of role identity and the stiffness of dynamic performance are always the core bottlenecks restricting automatic production.

In recent years, the breakthrough of DL technology, especially the rise of generative artificial intelligence, has provided a new paradigm for the intelligent production of animation content. Through convolution neural network, generation countermeasure network and diffusion model, the machine has the ability to understand semantics, generate high-fidelity images and drive complex actions [4]. The existing generation models tend to focus on the quality of a single frame image, but lack the performance in the temporal coherence of video sequences [5]. In addition, the animation style is highly abstract, and it is difficult for the general model to directly capture its unique lines and color rules, which leads to the lack of artistic beauty in the generated results [6].

This study focuses on the technology of character generation and dynamic expression of animation short plays based on DL, aiming at

exploring how to use the algorithm model to solve key problems such as keeping the identity consistency of characters, generating multi-views and audio-driven facial animation. The research pays attention to the visual fidelity of the generated quality, and emphasizes the emotional communication and time sequence coherence of dynamic performance. By constructing the theoretical framework and technical path, this article tries to demonstrate the feasibility of DL in lowering the production threshold and improving production efficiency.

Although the generation technology of single mode has become more and more mature, the dynamic cooperation of roles across modes still faces challenges. This article will analyze the role generation mechanism and dynamic performance principle from the theoretical framework, and establish the corresponding evaluation system. It is expected that this study can provide theoretical support for the industrial production of animation short plays, promote the deep integration of technology and art, and realize the large-scale output of high-quality content. This study will also discuss the ethical issues behind technology, ensure that technology application conforms to copyright norms and social values, and promote the healthy and sustainable development of the industry. Through systematic theoretical combing, this article hopes to provide a clear direction for the subsequent engineering practice and fill the theoretical gap in the field of automatic generation of animation short plays.

2. ROLE GENERATION MECHANISM

In the field of automatic production of animation short plays, role generation occupies a core position. The core challenge in this process lies in how to maintain the stability of character identity and the consistency from multiple perspectives while ensuring high-fidelity image quality [7]. The decoupling and reconstruction of potential spatial features constitute the theoretical cornerstone of role generation. By mapping high-dimensional image data to low-dimensional potential space, the depth generation model achieves a structured presentation of character content.

Theoretically speaking, the ideal potential space should have decoupling characteristics, that is, the identity characteristics of the role are separated from the variable factors and placed in different subspace vectors. By manipulating the potential vectors in a specific dimension, the model can reconstruct the character images in different poses without changing the character identity. This decoupling mechanism greatly reduces the complexity of the generation process. Table 1 lists the performance of identity characteristics and variable factors in different dimensions.

Table 1. Decoupled Dimensions and Features of Latent Space Characteristics

Latent Space Dimension Classification	Specific Orientation of Subspace Vectors	Examples of Identity Features	Examples of Variable Factors
Facial Feature Dimension	Facial Feature Sub-vectors	Eye shape, nose contour	Facial expressions
Head Hair Dimension	Hairstyle Sub-vectors	Hair length, hairstyle type	Light reflection effects on hair
Body Shape Dimension	Body Figure Sub-vectors	Height proportion, body fatness/thinness	Posture
Clothing and Accessories Dimension	Apparel Sub-vectors	Clothing style, accessory type	Presentation of clothing colors under different lighting conditions

Style transfer and identity preservation algorithm effectively solve the problem of role consistency. In the process of making short plays, the characters need to maintain visual unity in multiple scenes. The identity preserving algorithm based on attention mechanism extracts the deep features embedded in the reference image and integrates them into the generating network as conditions. From a theoretical perspective, this constitutes a constrained optimization problem: minimizing the difference between the generated image and the target style while maximizing its similarity to the reference character in the feature space [8]. By introducing the identity loss function, the model can suppress the feature drift in the process of style transfer.

Multi-view consistency constraint theory gives the role a three-dimensional sense. Although cartoon characters are presented in two dimensions, they must follow three-dimensional geometric logic. The multi-view generation mechanism uses

epipolar geometry constraint or implicit neural representation to ensure the structural coherence of the role in different angles. Theoretically, this requires the generated model to learn an implicit three-dimensional manifold, so that it can converge to the same role entity when projecting from any perspective. By introducing a multi-view discriminator, the model can punish the generation results that do not conform to geometric logic, and then eliminate the flicker and deformation that may occur when switching perspectives.

The role customization mechanism under the condition of small samples improves the practical value of the technology. Aiming at the characteristics of rich and diverse characters but lack of data in short plays, the customization mechanism based on meta-learning or efficient fine-tuning of parameters is studied. This mechanism advocates that on the basis of universal prior knowledge, only a few samples are needed to quickly adapt to new roles. By freezing most of the model parameters and optimizing only the specific adaptation layer, the model can capture the unique characteristics of new characters at a very low cost, which provides a theoretically feasible technical way for the large-scale production of animation short plays.

3. DYNAMIC REPRESENTATION PRINCIPLE

The core of dynamic representation principle lies in cross-modal semantic mapping and sequential coherence control [9]. Firstly, audio-driven facial motion mapping is realized by establishing a nonlinear relationship between acoustic features and blendshape weights. Let the audio feature of t be A_t, the motion parameter be M_t, and the mapping function be defined as:

$$L_{temp} = \sum_t || M_t - M_{t-1} ||_2^2 \quad (1)$$

Where z is the potential vector of role identity, which ensures the decoupling of mouth shape and identity. Secondly, it is very important to smooth the time sequence. In order to prevent flicker between video frames, the constraint of timing consistency loss is introduced:

mechanism to prevent the recurrence of similar errors.

$$L_{temp} = \sum_t || M_t - M_{t-1} ||_2^2 \quad (2)$$

This constraint forces the motion change of adjacent frames to be smooth, which conforms to the law of physical motion inertia. In addition, emotional expression is realized by fine-tuning the mapping matrix through attention mechanism, which enhances the naturalness of micro-actions. Finally, the generator optimizes dynamic performance in confrontation training to ensure smooth movements and accurate emotions, realize the transformation from static characters to dynamic performance, and solve the problem of low efficiency of traditional key-frame animation.

4. EVALUATION SYSTEM AND ETHICAL GOVERNANCE

Building a scientific evaluation system is the premise of technology landing. This study proposes a multi-dimensional quality evaluation index covering visual fidelity, time sequence stability and identity consistency. In this system, an improved video Fréchet starting distance (FVD-Anim) and identity consistency score (ICS-Anim) are introduced, and the FVD threshold is set below 150 as the qualified line, and the ICS-Anim should be higher than 0.85. The weight distribution adopts the analytic hierarchy process (AHP) to ensure that the dynamic performance weight accounts for no less than 40%, and the static image quality and semantic alignment account for 30% respectively. Among them, time sequence stability is measured by optical flow variance to eliminate motion flicker, and identity consistency is calculated based on Euclidean distance of facial feature points. This quantitative standard effectively solves the problem of inconsistent subjective evaluation and provides clear guidance for model iteration.

At the level of technical ethics, this study designed a closed-loop risk assessment mechanism. As shown in Figure 1, this process ensures that each generation link can be traced, thus reducing legal risks. In particular, the mechanism includes user feedback loop, which re-incorporates the violation cases into the training set for negative sample optimization to form a self-correction

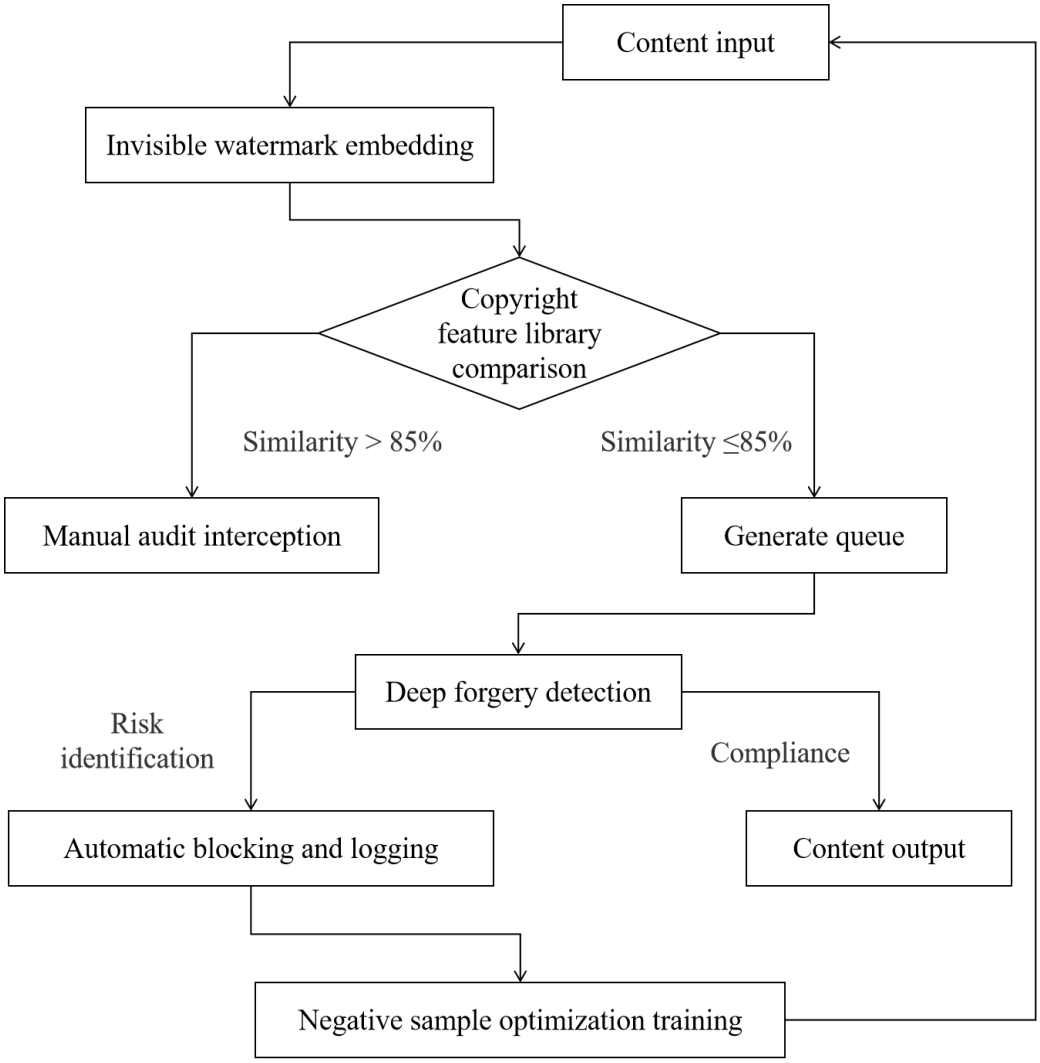


Figure 1 Ethical risk assessment process

Aiming at the core dispute of copyright ownership, this article studies and puts forward a mechanism based on blockchain, which makes clear the contribution ratio of algorithm and creator by recording the complete generated link. In order to prevent the abuse of deep forgery, it is suggested that invisible digital watermark should be forcibly implanted to achieve accurate traceability. In the future, technological evolution will tend to the man-machine cooperation mode, in which algorithms undertake repetitive work and human beings focus on creative decision-making. At the same time, the study emphasizes the need to pay attention to the elimination of model bias, ensure cultural diversity and avoid the generation of stereotypes. At the policy level, it is suggested to formulate a hierarchical management system and implement mandatory labeling for high-risk generated content.

Through the dual drive of quantitative evaluation and ethical constraints, we can promote the technology to be good, realize the sustainable development of the industry, and build a security barrier for the intelligent transformation of animation short plays. This study further suggests establishing industry alliance, sharing blacklist database, jointly maintaining the ecological health of content and ensuring the technology to operate within the legal and moral framework.

5. CONCLUSION

In this article, the theoretical framework of role generation and dynamic performance of animation short plays based on DL is systematically constructed. The decoupling and reconstruction mechanism of latent spatial features can effectively separate identity attributes from variable factors,

and combine style transfer and identity preservation algorithms to improve role consistency in multi-views. The introduction of audio-driven mapping mechanism and time sequence consistency loss constraint ensures the fluency of facial movements and the naturalness of emotional expression, and can realize the low-cost transformation from static roles to dynamic performances. The established multi-dimensional evaluation system and ethical risk evaluation process provide a standard basis for the quality quantification and safety compliance of generated content, and effectively prevent the risk of technology abuse; The quantitative indicators put forward in it set an executable acceptance standard for the industry. Future research will focus on the deployment of lightweight models to enhance on-device inference efficiency, enabling real-time interactive generation. It will also explore the deep integration of 3D implicit representations with 2D generation to strengthen the spatial dimensionality of characters. In terms of ethical governance, it is recommended to promote the establishment of industry standards and form a cross-platform copyright traceability alliance to ensure the responsible development of technology.

REFERENCES

[1] Zhang Ruisi, Pan Ye. Deep Learning Techniques for Character Expression Animation Generation[J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(05): 675-682.

[2] Fu Menghan, Fu Mengyuan. Exploration of the Generative Evolution and Cultural Characteristics of Chinese Animated Hero Characters[J]. Movie Literature, 2025, (15): 61-65.

[3] Wang Lei, Li Shujie, Xie Wenjun, et al. Skeletal Motion Data Generation Algorithm for Humanoid Character Animation[J]. Journal of Hefei University of Technology (Natural Science Edition), 2023, 46(01): 36-41+140.

[4] Zheng Ruikun, Liu Gengxin, Hu Ruizhen. Semi-supervised Character Animation Style Transfer[J]. Journal of Computer-Aided Design & Computer Graphics, 2025, 37(05): 821-831. DOI: 10.3724/SP.J.1089.2023-00385.

[5] Weng Jiankai, Zhao Na. Reshaping and Innovation of AI-Driven Animation Film Creation Process[J]. Movie Literature, 2025, (02): 94-98+103.

[6] Ding Zhuogun. Innovation Drive and Industrial Upgrading of AI Animation in the Era of "Human-Machine Co-creation"[J]. Media, 2025, (16): 54-56.

[7] Liang Wei, Huang Yifan, Shen Yuxin, et al. Instruction-Driven Stylized Motion Generation for Virtual Characters[J]. Transactions of Beijing Institute of Technology, 2024, 44(11): 1199-1207. DOI: 10.15918/j.tbit1001-0645.2024.017.

[8] Gao Xuan, Liu Dongyu, Zhang Juyong. A Review of Multimodal Digital Human Modeling, Synthesis and Driving[J]. Journal of Image and Graphics, 2024, 29(09): 2494-2512.

[9] Liu Shuyan, Ding Yubin, Zhu Yunfei. Crowd Path Planning in 3D Animation Based on Improved Particle Swarm Optimization[J]. Journal of Nanjing University of Science and Technology, 2022, 46(05): 594-599. DOI: 10.14177/j.cnki.32-1397n.2022.46.05.011.