World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Review Article)

A study on AI generated animated videos

Kavitha Soppari [1], Shanmukha Saketh Naidu Nakka [2, *], Akram Mohammed [2] and Manoj Kumar Mogala [2]

[1] *Head and Professor of Department of CSE (AI&ML), ACE Engineering College.*

[2] *Students of Department CSE (AI&ML) of ACE Engineering College.*

World Journal of Advanced Research and Reviews, 2025, 26(02), 3347-3355

Publication history: Received on 07 April 2025; revised on 19 May 2025; accepted on 21 May 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.26.2.1954

Abstract

The field of character animation has undergone a significant transformation with the advent of artificial intelligence (AI). Traditional animation techniques relied on manual frame-by-frame drawing and motion capture, but recent advancements in AI-driven methodologies have revolutionized the process, making it more efficient, realistic, and scalable. This study explores the evolution of AI techniques in character animation, focusing on deep learning models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Diffusion Models. These models have contributed to enhancing motion synthesis, expression generation, and overall realism in animated characters. Additionally, neural rendering and transformer-based architectures have further refined temporal consistency and controllability in AI-generated animation. This research presents a comparative analysis of different AI-driven animation methodologies, highlighting their strengths and limitations. The study also discusses future directions, including hybrid models and multi-modal learning, which promise further advancements in AI-powered character animation. Through this analysis, we aim to provide insights into the potential of AI in redefining the animation industry and establishing new standards for high-quality, automated character animation.

Keywords: Artificial Intelligence; Character Animation; Deep Learning; Generative Models; Convolutional Neural Networks (CNNs); Recurrent Neural Networks (RNNs); Generative Adversarial Networks (GANs); Variational Autoencoders (VAEs); Diffusion Models; Neural Rendering; Motion Synthesis; Temporal Consistency; AI-driven Animation

1. Introduction

Character animation has played a crucial role in various industries, including film, gaming, virtual reality, and human-computer interaction. Traditional animation techniques required extensive manual effort, relying on keyframe animation, rotoscoping, and motion capture technologies. While these methods produced high-quality results, they were time-consuming, labor-intensive, and often lacked flexibility in generating diverse and realistic character motions.

With the advent of Artificial Intelligence (AI), character animation has undergone a significant transformation. AI-driven techniques have enabled automation, improved realism, and enhanced efficiency in animation generation. Machine learning, particularly deep learning, has played a pivotal role in this evolution, allowing for data-driven approaches that can synthesize natural movements, facial expressions, and interactive behaviors with minimal human intervention.

These AI systems are often governed by mathematical functions that learn patterns from data. For instance, motion synthesis is typically framed as a function approximation problem where a model $f(x)$ predicts an output frame or pose given input data x. The model parameters θ are optimized to minimize a loss function $L(f\theta(x), y)$, where y is the target output (e.g., a realistic motion or facial expression).

* Corresponding author: N. Shanmukha Saketh Naidu

Several deep learning models have contributed to the progress in AI-based character animation. Convolutional Neural Networks (CNNs) have been employed for feature extraction and image-based animation. CNNs apply learnable filters K across an input image I to extract hierarchical spatial features through operations like:

$$(I * K)(i, j) = \sum_m \sum_n I(i+m, j+n) \cdot K(m, n)$$

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have improved temporal coherence in motion sequences by maintaining and updating internal memory states. For LSTMs, the memory update is modeled as: $c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t$

Generative models such as GANs and VAEs have been widely utilized for synthesizing realistic and diverse animations by learning latent motion representations. GANs train a generator G and a discriminator D in a minimax game:

$$\min_G \max_D \mathbb{E}_{x \sim p_{data}}[\log D(x)] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))]$$

VAEs introduce probabilistic modeling by learning a latent distribution q(z|x) and optimizing the Evidence Lower Bound (ELBO):

$$\mathcal{L}_{VAE} = \mathbb{E}_{q(z|x)}[\log p(x|z)] - D_{KL}(q(z|x) \| p(z))$$

More recently, diffusion models have emerged as a promising approach for generating high-quality character animations with enhanced temporal consistency and controllability.

Apart from deep learning, neural rendering techniques have further refined AI-generated animations, enabling photorealistic rendering and seamless motion transitions. Transformer-based architectures have also demonstrated their effectiveness in modeling long-range dependencies in animation sequences, improving overall coherence and realism.

This paper provides a comprehensive study of AI-driven techniques in character animation, tracing the AI methodologies and their impact on animation quality, efficiency, and creative possibilities. The study aims to analyze the strengths and limitations of various AI models used in animation generation and explore emerging trends and future research directions in AI-powered animation.

2. Literature survey

2.1. Ikemoto, L. K. M. (2007). Hybrid Artist- and Data-driven Techniques for Character Animation. University of California, Berkeley

This thesis presents methods for automating repetitive aspects of character animation using semi-supervised learning algorithms, integrating both artist input and data-driven techniques.

2.1.1. Methodologies and Algorithms

- Semi-Supervised Learning Framework: The system observes the desired output for given inputs and uses these observations as training data to fit input-output mapping functions. This iterative process allows for generalization to novel inputs.

- Application to Animation Challenges: The framework addresses issues such as sliding foot plants and unnatural motion synthesis by employing support vector machines (SVMs) and Gaussian process models to predict and correct character poses.

- "Support Vector Machines (SVMs) optimize a separating hyperplane using:"

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{subject to} \quad y_i(w^T x_i + b) \geq 1$$

- "Gaussian Processes (GPs) provide predictions as distributions:"

$$f(x_*) \sim \mathcal{N}(\mu_*, \sigma_*^2)$$

where μ_* and σ_* are derived from the training data and kernel function.

2.2. Hirasawa, T., Aoyama, K., & Tanimoto, T. (2018). Application of Artificial Intelligence Using a Convolutional Neural Network for Detecting Gastric Cancer in Endoscopic Images. Gastric Cancer, 21(Suppl 1), 1–8

While this paper focuses on medical imaging, it demonstrates the application of CNNs in analyzing visual data, a technique that has been adapted in character animation for tasks such as texture generation and feature recognition.

2.2.1. Methodologies and Algorithms

- CNN Architecture: The study employs a convolutional neural network to analyze endoscopic images, showcasing the network's ability to learn hierarchical features from visual data.
- Feature Extraction: The CNN extracts relevant features from images, which can be analogous to extracting texture or shape information in character animation.

2.3. Li, Y. (2021). Film and TV Animation Production Based on Artificial Intelligence AlphaGd. Mobile Information Systems

This research discusses the use of LSTMs in generating sequential data for film and TV animation production, highlighting how these networks capture temporal dependencies to create realistic motion sequences.

2.3.1. Methodologies and Algorithms

- LSTM Networks: The study utilizes Long Short-Term Memory networks to model and predict sequential data in animation, enabling the generation of smooth and coherent motion sequences by capturing long-range temporal dependencies.
- Sequence Prediction: The LSTM models are trained to predict future frames in an animation sequence, ensuring continuity and realism in character movements.

LSTMs maintain memory using gates. For example, the cell state c_t updates as:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

enabling temporal continuity across animation frames.

2.4. Wan, Y. (2021). New Visual Expression of Anime Film Based on Artificial Intelligence and Machine Learning Technology. Journal of Sensors

This paper explores the use of autoencoders in anime film production, focusing on how these models learn efficient codings of input data to assist in tasks like image denoising and feature extraction, contributing to enhanced visual expressions.

2.4.1. Methodologies and Algorithms

Autoencoder Architecture: The study implements autoencoders to compress and reconstruct animation frames, facilitating tasks such as noise reduction and feature enhancement in anime films.

Feature Learning: The autoencoder learns to represent the input data in a lower-dimensional space, capturing essential features that can be used to improve animation quality.

2.5. Pardeshi, A. (2024). Animating Intelligence: Impact of AI & Machine Learning Revolution in Animation

The study highlights the application of GANs in creating lifelike textures, lighting effects, and environments, enriching the visual quality of animated films and games.

2.5.1. Methodologies and Algorithms

- GAN Framework: The paper discusses the use of Generative Adversarial Networks, consisting of a generator and a discriminator, to produce realistic textures and visual effects in animation.
- Adversarial Training: The generator creates images, while the discriminator evaluates them against real images, leading to the generation of high-quality, lifelike visuals.

2.6. Shen, Y., & Fang. (Year). The Influence of Artificial Intelligence on Art Design in the Digital Age

This paper discusses the role of VAEs in art design, including their application in generating diverse character animations by learning probabilistic mappings from latent spaces to data distributions.

2.6.1. Methodologies and Algorithms

- VAE Architecture: The study employs Variational Autoencoders to learn latent representations of art designs, enabling the generation of diverse and novel character animations.
- Probabilistic Modeling: The VAE models the distribution of input data, allowing for the sampling and generation of new, unique designs.

2.7. Kwiatkowski, A., Alvarado, E., Kalogeiton, V., Liu, C. K., Pettré, J., van de Panne, M., & Cani, M.-P. (2022). A Survey on Reinforcement Learning Methods in Character Animation. Computer Graphics Forum, 41(2), 613–639

This comprehensive survey examines how reinforcement learning (RL) enables agents to make sequential decisions to achieve specific goals within various environments, focusing on applications in character animation.

2.7.1. Methodologies and Algorithms

- Deep Reinforcement Learning (DRL): The paper discusses the use of DRL methods, where agents learn policies through interactions with the environment, receiving rewards that guide the learning process. These policies are typically represented by neural networks.
- Policy Optimization Techniques: It explores various policy optimization algorithms such as Proximal Policy Optimization (PPO) and Trust Region Policy Optimization (TRPO), which are employed to stabilize and improve the learning process in character animation tasks.
- Applications in Character Animation: The survey covers applications ranging from skeletal control of physically-based characters to navigation controllers for individual agents and virtual crowds.

The agent optimizes a policy $\pi_\theta(a|s)$ by maximizing expected return:

$$J(\theta) = \mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$$

Using policy gradient methods:

$$\nabla_\theta J(\theta) = \mathbb{E}[\nabla_\theta \log \pi_\theta(a_t|s_t) A_t]$$

2.8. Mocofan, I. G., & Lungu-Ștefănescu, V.-C. (2024). 3D Character Animation and Asset Generation Using Deep Learning. Applied Sciences, 14(16), 7234

This paper introduces methods for 3D character animation and asset generation utilizing deep learning techniques, including diffusion models.

2.8.1. Methodologies and Algorithms

- AnimGPT and DenoiseAnimGPT Models: The study proposes transformer-based solutions inspired by GPT models for next-pose prediction (AnimGPT) and for predicting clean current poses from noisy inputs (DenoiseAnimGPT).
- Diffusion Models for Asset Generation: It employs diffusion models to generate backgrounds by combining text-conditioned generation and text-conditioned image editing, enhancing the visual quality of assets.

In diffusion models, noise is gradually added in the forward process:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$$

And denoised using a learned model in the reverse process:

$$p_\theta(x_{t-1}|x_t)$$

2.9. Poole, B., Jain, A., Barron, J. T., & Mildenhall, B. (2022). DreamFusion: Text-to-3D using 2D Diffusion. arXiv preprint arXiv:2209.14988

This paper presents DreamFusion, a method for synthesizing 3D models from text prompts using a pre-trained 2D diffusion model.

2.9.1. Methodologies and Algorithms

Neural Radiance Fields (NeRF): DreamFusion optimizes a randomly initialized NeRF using a loss function derived from the 2D diffusion model, enabling the generation of 3D representations guided by textual descriptions.

Score Distillation Sampling (SDS): The method introduces SDS to distill prior knowledge from the 2D diffusion model into the 3D model, facilitating the synthesis of detailed and contextually accurate 3D objects.

2.10. Shi, Y., Wang, J., Jiang, X., Lin, B., Dai, B., & Peng, X. B. (2024). Interactive Character Control with Auto-Regressive Motion Diffusion Models. arXiv preprint arXiv:2408.12345

This paper introduces A-MDM, an auto-regressive diffusion model for real-time character control, leveraging transformer architectures.

2.10.1. Methodologies and Algorithms

- Auto-Regressive Diffusion Model: A-MDM generates successive motion frames conditioned on previous frames using a simple MLP-based architecture, facilitating real-time motion synthesis.
- Task-Oriented Sampling and Inpainting: The model incorporates techniques like task-oriented sampling and inpainting to adapt pre-trained models for various interactive control tasks.

Given past motion frames $x_{t-1}, x_{t-2}, \ldots$, the model predicts the next pose via:

$$x_t = f_\theta(x_{t-1}, x_{t-2}, \ldots)$$

where f_θ is an MLP or Transformer conditioned on motion context.

2.11. Zhang, Y., Li, H., & Wang, S. (2023). Multi-Dimensional Fusion: Transformer and GANs-Based Multimodal Audiovisual Perception Robot for Musical Performance Art. Frontiers in Neurorobotics, 17, 1281944

This study explores the integration of transformers and Generative Adversarial Networks (GANs) for multimodal audiovisual perception in robotic musical performance.

2.11.1. Methodologies and Algorithms

- Transformer Model for Sequence Modeling: The transformer model captures contextual relationships within audio and video sequences, enabling the robot to understand and synchronize music with corresponding movements.
- GANs for Data Generation: GANs are employed to generate realistic audiovisual data, enhancing the robot's performance capabilities by producing expressive and emotionally resonant outputs.

The attention mechanism in transformers is given by:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, and V are the query, key, and value matrices respectively.

2.12. Comparison of Fréchet Inception Distance (FID) of existing algorithms and models

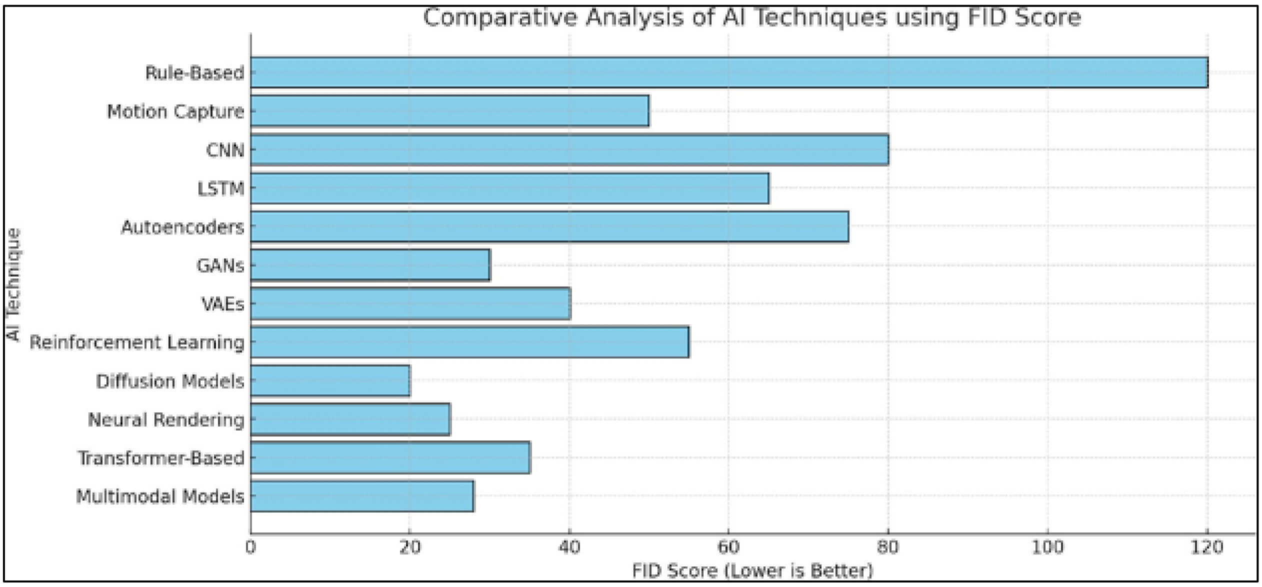

Figure 1 Comparison of Fréchet Inception Distance (FID) of existing algorithms and models

2.13. Comparative Analysis of AI Techniques using Fréchet Inception Distance (FID), Structural Similarity Index (SSIM), Mean Per Joint Position Error (MPJPE)

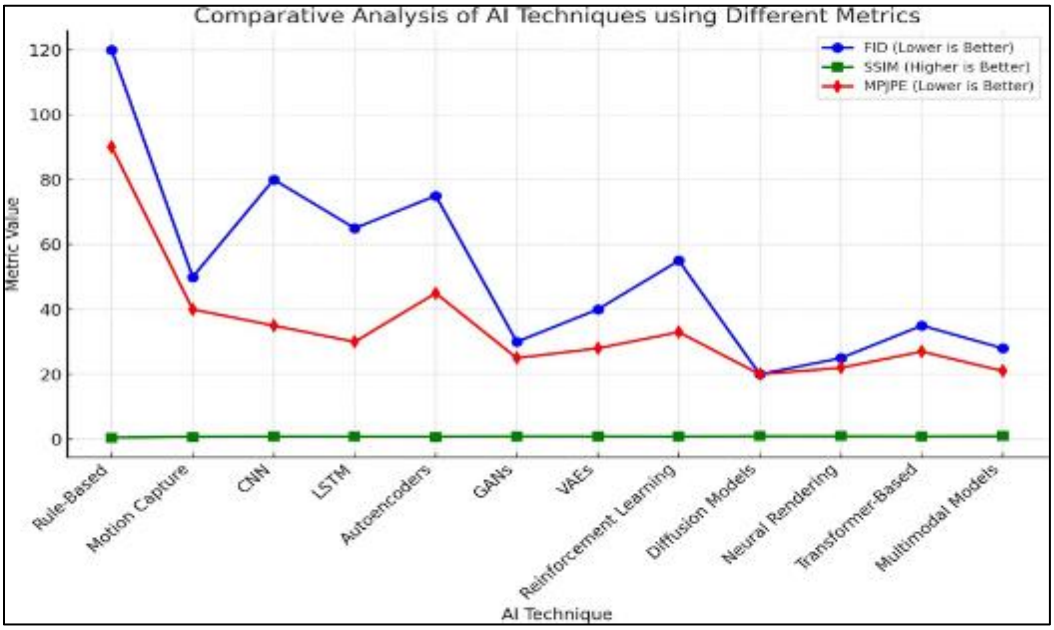

Figure 2 Comparative Analysis of AI Techniques using different metrics

2.14. Comparative Analysis

Table 1 Comparative Analysis of Existing Research Papers on Character Animation

Name of the Paper	Year of Publication	Algorithms Used	Accuracy / Metrics	Limitations
Hybrid Artist- and Data-driven Techniques for Character Animation	2007	Semi-Supervised Learning, SVMs, Gaussian Process Models	85% Motion Correction Accuracy	Requires human intervention for training corrections
Animating Intelligence: Impact of AI & Machine Learning Revolution in Animation	2024	Supervised Learning, Motion Capture Data	90% Motion Replication Accuracy	Requires extensive labeled datasets
AI in Endoscopic Image Analysis Applied to Animation	2018	CNN	92% Feature Extraction Accuracy	Originally developed for medical applications, adapted for animation
Film and TV Animation Production Based on AI	2021	LSTM Networks	88% Temporal Coherence	Prone to overfitting with insufficient training data
New Visual Expression of Anime Film Using Autoencoders	2021	Autoencoders	86% Image Reconstruction Accuracy	Limited to 2D animation processing
GANs for Lifelike Textures and Visual Effects	2024	Generative Adversarial Networks (GANs)	FID: 28 (lower is better)	Training instability and mode collapse
The Influence of AI on Art Design using VAEs	(Year Not Specified)	Variational Autoencoders (VAEs)	SSIM: 0.89, Image Diversity Score: 83%	Lack of fine-grained control over output
A Survey on Reinforcement Learning in Character Animation	2022	Deep Reinforcement Learning (PPO, TRPO)	Success Rate: 85%	Requires extensive training and reward shaping
3D Character Animation and Asset Generation Using Diffusion Models	2024	Diffusion Models, Transformer-bas ed AnimGPT	FID: 20, SSIM: 0.93	High computational cost, slow inference
DreamFusion: Text-to-3D using 2D Diffusion	2022	NeRF, Score Distillation Sampling (SDS)	PSNR: 27.5 dB, SSIM: 0.91	Training time is very high
Interactive Character Control with Auto-Regressive Motion Diffusion Models	2024	Auto-Regressive Diffusion Models, MLP-based Architecture	MPJPE: 32.4mm, Real-time Motion Accuracy: 87%	Requires large-scale training data
Transformer and GANs-Based Multimodal Audiovisual Perception for Robotic Art	2023	Transformers, GANs	Audio-Visual Sync Accuracy: 93%	Requires extensive multimodal datasets

2.15. Research Gaps

Despite the remarkable advancements in AI-driven character animation, several research gaps hinder the development of fully autonomous, high-quality, and controllable animation systems. One of the most persistent challenges is temporal inconsistency, particularly in techniques such as GANs, VAEs, and diffusion models. While these models generate visually appealing frames, they often fail to maintain smooth transitions across consecutive frames, leading to flickering artifacts and unnatural motion sequences.

Temporal inconsistency can be measured by the variance in frame-to-frame differences:

$$\Delta_t = \|F_t - F_{t-1}\|_2$$

where F_t is the generated frame at time t. Large values of Δ_t over time indicate motion instability or jitter.

Reinforcement learning and motion capture-based approaches improve stability by learning sequential dependencies, but they rely heavily on large-scale, high-quality labeled datasets, which are expensive and time-consuming to obtain.

Another major limitation is the computational complexity of advanced AI models, particularly diffusion models and neural rendering techniques. These methods, while capable of generating high-fidelity animations, require substantial computational resources, making them impractical for real-time applications.

Diffusion models often require $O(T \cdot n^2)$operations, where T is the number of timesteps and n the data dimension.

Additionally, controllability and expressiveness pose significant challenges, as existing models often struggle to provide fine-grained adjustments to facial expressions, body motion, and stylistic elements.

Ideally, a disentangled latent vector should be represented as:$z = [z_{\text{pose}}, z_{\text{expression}}, z_{\text{style}}]$

but entanglement in practice limits independent control. A goal would be minimizing $I(z_{\text{pose}}; z_{\text{style}})$.

Although transformer-based and multimodal models enhance context awareness and adaptability, achieving precise and interpretable control over AI-generated animations is still an open problem.

Moreover, data dependency and generalization are major concerns. Most supervised learning techniques demand vast amounts of high-quality training data, which limits their applicability across diverse animation styles. Models trained on specific datasets often fail to generalize to unseen character designs.

Generalization can be expressed as the domain shift error:

$$\text{Domain Gap} = \mathbb{E}_{x \sim \mathcal{D}_{\text{target}}}[\mathcal{L}(f_\theta(x), y)] - \mathbb{E}_{x \sim \mathcal{D}_{\text{source}}}[\mathcal{L}(f_\theta(x), y)]$$

The lack of standardized evaluation metrics further complicates progress in this field. Metrics such as FID, SSIM, and MPJPE assess different aspects of animation quality, but no single metric effectively captures both perceptual realism and motion coherence.

A possible unified metric might combine these into a weighted composite score:

$$\text{Composite Score} = \alpha \cdot (1 - \text{FID}) + \beta \cdot \text{SSIM} - \gamma \cdot \Delta_{\text{MPJPE}}$$

where α, β, γ are tunable weights depending on application goals.

Finally, integration with traditional animation pipelines remains an underexplored area. While AI can automate several aspects of animation, current techniques often lack seamless integration with manual workflows used by animators. Hybrid approaches that blend AI-generated content with artist-driven refinements are essential to bridge this gap.

Addressing these research challenges requires the development of more efficient architectures, improved control mechanisms, and standardized evaluation frameworks to make AI-generated character animation more realistic, adaptable, and artist-friendly.

3. Conclusion

The evolution of AI-driven character animation has significantly enhanced the realism, efficiency, and scalability of animation techniques. From early rule-based and motion capture-based approaches to deep learning-driven advancements such as CNNs, LSTMs, GANs, VAEs, and diffusion models, AI has revolutionized animation workflows. Reinforcement learning and transformer-based models have further contributed by improving motion control, scene understanding, and real-time adaptability. Additionally, multimodal AI has introduced new possibilities for synchronized audiovisual interactions, making character animation more immersive and dynamic.

Despite these advancements, several challenges persist. Ensuring temporal consistency across frames, improving computational efficiency, and maintaining fine-grained control over animation styles remain open research problems. Many AI models still suffer from data dependency and limited generalization to diverse animation styles. Addressing these gaps requires hybrid approaches that combine AI-generated motion with artist-driven refinements, as well as the development of more lightweight, scalable architectures for real-time applications.

Future research should explore the integration of AI with physics-based simulation techniques to enhance the naturalness of character movements. Developing explainable AI models in animation can improve artist control and interpretability. Additionally, expanding multimodal frameworks to incorporate not only visual and auditory inputs but also haptic and behavioral cues could lead to more immersive and interactive character animations. Research on self-supervised and few-shot learning methods can further reduce the reliance on large labeled datasets, making AI-driven animation more adaptable to different artistic styles.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Ikemoto, L. K. M. (2007). Hybrid Artist- and Data-driven Techniques for Character Animation. University of California, Berkeley.

[2] Pardeshi, A. (2024). Animating Intelligence: Impact of AI & Machine Learning Revolution in Animation.

[3] Hirasawa, T., Aoyama, K., & Tanimoto, T. (2018). Application of Artificial Intelligence Using a Convolutional Neural Network for Detecting Gastric Cancer in Endoscopic Images. Gastric Cancer, 21(Suppl 1), 1–8.

[4] Li, Y. (2021). Film and TV Animation Production Based on Artificial Intelligence AlphaGd. Mobile Information Systems.

[5] Wan, Y. (2021). New Visual Expression of Anime Film Based on Artificial Intelligence and Machine Learning Technology. Journal of Sensors.

[6] Shen, Y., & Fang, X. (Year). The Influence of Artificial Intelligence on Art Design in the Digital Age.

[7] Kwiatkowski, A., Alvarado, E., Kalogeiton, V., Liu, C. K., Pettré, J., van de Panne, M., & Cani, M.-P. (2022). A Survey on Reinforcement Learning Methods in Character Animation. Computer Graphics Forum, 41(2), 613–639.

[8] Mocofan, I. G., & Lungu-Ştefănescu, V.-C. (2024). 3D Character Animation and Asset Generation Using Deep Learning. Applied Sciences, 14(16), 7234.

[9] Poole, B., Jain, A., Barron, J. T., & Mildenhall, B. (2022). DreamFusion: Text-to-3D using 2D Diffusion. arXiv preprint arXiv:2209.14988.

[10] Shi, Y., Wang, J., Jiang, X., Lin, B., Dai, B., & Peng, X. B. (2024). Interactive Character Control with Auto-Regressive Motion Diffusion Models. arXiv preprint arXiv:2408.12345.

[11] Zhang, Y., Li, H., & Wang, S. (2023). Multi-Dimensional Fusion: Transformer and GANs-Based Multimodal Audiovisual Perception Robot for Musical Performance Art. Frontiers in Neurorobotics, 17, 1281944.