www.cad-journal.net
Computer-Aided Design AND APPLICATIONS

Automatic Generation of Animation Special Effects Based on Computer Vision Algorithms

Yang Liu[1], Lei Li[2] and Xiang Lei[3]

[1]International College, Jiangxi College of Applied Technology, GanZhou 341000, China, 15170219146@126.com
[2]International College, Jiangxi College of Applied Technology, GanZhou 341000, China, 18879770182@139.com
[3]International College, Jiangxi College of Applied Technology, GanZhou 341000, China, xsleixiang2@jxyy.edu.cn

Corresponding author: Xiang Lei, xsleixiang2@jxyy.edu.cn

Abstract. In the dawn of the "Internet plus" era, advancements in science and technology have unlocked immense potential for animation special effect design. As a prime exemplar of the fusion between creativity and technology, the animation industry is steadily progressing toward a more intelligent and automated future. This article introduces a groundbreaking algorithm that merges CAD (computer-aided design), reinforcement learning (RL), and computer vision (CV) algorithms to revolutionize the automatic generation of animation special effects. Initially, CAD models are employed to build the animation scenes and character models. Subsequently, RL is utilized to learn and mimic character actions and behaviours. Lastly, CV algorithms are leveraged to identify and track scene elements, seamlessly generating corresponding special effects. This automated and intelligent workflow significantly enhances the efficiency and quality of animation special effects generation. Experimental outcomes demonstrate that our proposed algorithm, which integrates CAD, RL, and CV techniques, can produce remarkably realistic animation special effects. This innovation offers fresh perspectives and techniques that are poised to propel the animation industry forward.

Keywords: CAD; Reinforcement Learning; Computer Vision Algorithms; Animation Special Effects; Automatic Generation
DOI: https://doi.org/10.14733/cadaps.2024.S23.69-83

1 INTRODUCTION

Animation special effects production is a fascinating technology that is based on computer technology and creates a virtual world through specific software systems. Non-paired sports style transition from video to animation. This technology allows us to extract motion data from videos and convert it into an animation style, thereby creating a brand-new visual effect. Aberman et al. [1] introduced the principle, application, and importance of this technology in the field of computer vision. The traditional animation production process usually requires a lot of manual drawing and frame-by-frame processing, which is not only time-consuming and laborious but also makes it difficult to achieve complex and natural motion effects. However, through computer vision algorithms, we can automatically extract motion data from videos and convert it into animation styles, greatly improving the efficiency and flexibility of animation production. The non-paired motion style conversion technology from video to animation has broad application prospects. It can be used in fields such as game development, virtual reality, and movie special effects, providing creators with richer and more diverse visual expression methods. In this world, creators can freely design various objects, scenes, and characters and give them vivid motion trajectories and special effects. With the development of computer graphics, the application of 3D graphics engines in animation design is becoming increasingly widespread. Traditional animation design methods often require manual adjustment of parameters, creation of keyframes, etc., which is time-consuming and inefficient. In recent years, the application of intelligent algorithms in 3D graphics engine animation design has gradually emerged, bringing revolutionary changes to animation design. In 3D graphics engine animation design, intelligent algorithms can be applied to character animation, physical simulation, scene rendering, and other aspects to achieve automated and intelligent animation design. Character animation is an important component of 3D graphics engine animation design. Traditional character animation design requires manual adjustment of parameters such as bones and muscles, creating keyframes, which are labor-intensive and prone to errors. The deep learning-based expression generation algorithm can automatically generate corresponding expressions based on the emotional state of the character, improving the realism and expressiveness of the animation [2]. This technology not only immerses the audience in a three-dimensional visual feast but also provides infinite creative space for animation creators. To ensure product quality, automatic detection technology has become crucial. In recent years, Ben et al. [3] have successfully developed an efficient automatic detection method for mechanical components by combining 3D CAD models with real 2D images. The core of this method lies in using 3D CAD models as standards to accurately match with real 2D images. Firstly, we obtain a 3D CAD model of the mechanical components to be tested, which provides us with an accurate and complete geometric description. Then, we use high-resolution cameras to capture 2D images of mechanical components, which contain information such as the actual shape, size, and surface quality of the components. Next, we will use computer vision and image processing techniques to process and analyze 2D images. Through steps such as feature extraction and edge detection, we extract key information from the image and compare it with the 3D CAD model. This comparison not only considers the overall shape and size of the components but also focuses on details and surface quality.

The key to animation special effects production lies in digital art modelling. Through precise modelling techniques, creators can construct realistic objects and scenes and combine computer graphics, physics engines, and other technologies to simulate real-world effects such as light, shadow, materials, and collisions. In terms of material prediction and recommendation, CAD technology combines advanced machine learning and graph learning algorithms, such as hierarchical graph learning, providing unprecedented convenience and accuracy for designers and engineers. Bian et al. [4] explored the application of hierarchical graph learning in material prediction and recommendation in computer-aided design. Traditional methods rely on a combination of rules and experience, but this appears inefficient and error-prone when dealing with large amounts of data. As an advanced machine learning method, hierarchical graph learning can effectively process graph-structured data and discover complex patterns and relationships in the data. By analyzing the chemical structure, microstructure, and past performance data of materials, hierarchical graph learning can predict the performance of new materials under specific conditions. This is crucial for accelerating the development of new materials and reducing experimental costs. By constructing a graph model of materials, hierarchical graph learning can identify materials that are similar in performance, composition, or structure. This is very useful for material substitution and expanding material databases. These special effects not only elevate the animation's visual appeal but also immerse the audience in a profound emotional and narrative experience. With the continuous progress of technology, animation special effects production is also constantly developing. Bodini [5] reviewed methods for extracting facial features in 2D images and videos using deep learning and explored how big data and cognitive computing bring new opportunities and challenges to this field. The application of deep learning in facial feature extraction is mainly based on convolutional neural networks (CNNs). By training a large amount of labelled facial image data, CNN can learn the precise position of facial landmarks. In recent years, with the expansion of datasets and optimization of model structures, the accuracy and speed of facial feature extraction have been significantly improved. Big data provides rich training data for deep learning models, enabling them to learn more facial features and changes. By collecting and analyzing large-scale facial image and video data, we can establish more accurate and robust facial feature extraction models. At the same time, big data also provides us with more diverse facial data, including facial images of different races, ages, genders, and expressions, which helps to improve the model's generalization ability.

New algorithms, technologies, and tools are constantly emerging, providing creators with more choices and possibilities. With the rapid development of deep learning and computer vision technology, image sentiment analysis has become a highly focused research field. Among them, the emotional classification of cartoon images is particularly important because cartoon images usually have strong emotional colours and visual impact. Cao et al. [6] explored how to classify cartoon images with a "moe" (meaning "cute") style based on deep learning techniques for bipolar emotions (positive and negative). Moe-style cartoon images are deeply loved by people for their unique cute elements and visual effects. However, for sentiment analysis of a large number of cartoon images, traditional methods often find it difficult to accurately capture the emotional information in the images. Deep learning models perform well in image sentiment classification tasks, especially in models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). CNN can capture spatial features in images, while RNN can process sequence data and capture temporal dependencies. For sentiment classification of cartoon images, we can use CNN models to extract features from the images and then use fully connected layers or classifiers to classify these features. At the same time, the audience's expectations for animation special effects are constantly increasing, which requires creators to constantly learn and explore new technologies to meet the needs of the audience. Animation special effects production is a technology full of challenges and opportunities. With the rapid development of computer vision and animation technology, 3D pose recognition in animation has become a hot research topic. Traditional methods are often limited by computational complexity and recognition accuracy, and the rise of deep convolutional neural networks (DCNN) has brought new opportunities in this field. Ding and Li. [7] explored how to improve deep convolutional neural networks to achieve high speed and accuracy in 3D pose recognition for animations. Animation 3D pose recognition refers to extracting and recognizing pose information in 3D space from animated videos. This is of great significance for fields such as animation production, human-computer interaction, and virtual reality. However, due to the complexity and diversity of animated videos, traditional pose recognition methods often struggle to achieve ideal results. Deep convolutional neural networks provide a new solution for animation 3D pose recognition by automatically learning feature representations in images. By training a large amount of animated video data, DCNN can learn the three-dimensional pose features of animated characters and accurately classify and recognize them. It can not only bring stunning visual effects to the audience but also provide creators with vast creative space [8]. With the swift progression of society and advances in computer technology, artificial intelligence (AI) has emerged as a leading force in the new wave of technological and industrial transformations. As a pivotal branch of AI, computer vision (CV) plays an increasingly significant role in various scientific and engineering domains, presenting both challenges and opportunities for innovation. Particularly within the animation industry, the application of CV holds immense importance.

Animation special effects constitute a crucial aspect of animation works, enriching them with deeper emotions and bolstering their visual impact. Nevertheless, traditional methods of creating animation special effects often rely heavily on manual design and adjustment, a process that is not only time-consuming and labour-intensive but also challenging to perfect. Consequently, the quest for automated and intelligent techniques to generate animation special effects has become an urgent priority. This article introduces an innovative algorithm that blends CAD, RL, and CV techniques for the automated generation of animation special effects.

The key innovations of this article are highlighted below:

(1) The successful integration of three distinct technological fields: CAD, RL, and CV algorithms. This interdisciplinary approach offers fresh perspectives and methodologies for the automated creation of animation special effects, fostering cross-application and technological advancement.

(2) The proposed algorithm achieves automated and intelligent special effects generation through RL, surpassing the limitations of traditional manual design and adjustment methods. This intelligent approach enables agents to learn and optimize special effects generation strategies independently within a virtual environment, significantly enhancing both efficiency and quality.

(3) The versatility and adaptability of the proposed algorithm allow it to cater to a wide range of animation scenes and special effects requirements. By adjusting algorithm parameters and strategies, it can accommodate diverse animation needs, providing valuable technical support for the industry's continued diversification and growth.

This article commences with a discussion of the research background and significance, followed by an exploration of the application of CAD, RL, and CV algorithms in animation special effects design. Subsequently, a comprehensive description of the algorithm's structure and process is presented, covering input, processing steps, and output. The effectiveness and feasibility of the algorithm are then substantiated through experimental validation. In the concluding section, the main findings and innovations of this article are summarized, along with recommendations for future research directions.

2 RELATED WORK

With the rapid development of computer vision and graphics technology, it has become possible to learn detailed 3D facial models that can be animated from wild images. This technology not only provides us with realistic and vivid virtual characters but also brings revolutionary changes to fields such as movies, games, and virtual reality. Feng et al. [8] explored how to learn detailed animated 3D facial models from wild images and discussed their application prospects. In the digital age, 3D facial models have become an important digital asset. By learning from wild images, we can obtain facial models with high realism and details, thereby achieving various animated effects. This technology not only requires the model to have high geometric accuracy but also needs to consider factors such as texture, lighting, and facial expressions to achieve realistic animation effects. Next, we need to use these features to reconstruct a 3D face model. This can be achieved through methods such as multi-view stereo vision and structured light. Among them, multi-view stereo vision uses multiple images from different angles to reconstruct 3D models, while structured light projects specific light patterns onto the surface of the face to obtain its geometric shape. Virtual reality (VR) technology and machine vision are gradually becoming innovative forces in the fields of architecture and animation. The combination of these two not only provides unlimited possibilities for architectural animation creation but also brings a new perspective and analytical method to architectural decoration aesthetics. Gong [9] explored how to conduct aesthetic analysis of architectural animation based on virtual reality technology and machine vision. Machine vision allows computers to "understand" images and videos, extracting useful information from them. In architectural animation, virtual reality technology can create realistic architectural scenes, and machine vision can conduct detailed aesthetic analyses of these scenes. Virtual reality technology provides a rich means of expression for architectural animation. Designers can use virtual reality technology to present the design concept, structure, and decorative details of a building in a three-dimensional form, allowing the audience to appreciate the beauty of the building from all angles and angles. In addition, virtual reality technology can also simulate building effects under different lighting conditions, providing more choices for animation creation. Machine vision technology can automatically analyze and process images and videos in architectural animation, extracting features related to aesthetics. For example, machine vision can analyze the texture, colour, and lighting effects of building surfaces to evaluate their coordination with the overall architectural style.

Animation image depth estimation has become a hot research direction in the field of computer vision. In recent years, the combination of machine learning and binocular stereo vision has brought new breakthroughs in this field. Poggi et al. [10] investigated the synergistic effect of machine learning and binocular stereo vision in-depth estimation of animated images and conducted a comprehensive survey. Animation image depth estimation refers to the restoration of depth information of a three-dimensional scene from a two-dimensional image. This is crucial for many application scenarios, such as virtual reality, augmented reality, autonomous driving, etc. Traditional depth estimation methods mainly rely on manually designed features and algorithms, but in recent years, with the rapid development of machine learning, especially the rise of deep learning, new ideas and methods have been provided for animation image depth estimation. Meanwhile, binocular stereo vision, as a classic technology in the field of computer vision, also provides an important source of information for depth estimation. Traditional animation production methods often require artists to manually draw each frame, which is both time-consuming and laborious. However, now, by utilizing advanced computer vision algorithms, filmmakers are able to create realistic virtual worlds and characters in unprecedented ways. Computer vision algorithms are the core of the film animation revolution. These algorithms can parse and understand information in images and videos, thereby generating realistic virtual scenes and characters. Reddy et al. [11] analyzed deep learning algorithms to train models for recognizing and generating faces, body movements, and expressions, thereby achieving highly realistic character animations. In addition, computer vision algorithms can also be used for scene reconstruction and special effects production, such as creating realistic digital cities, explosions, and flames. The introduction of AI technology makes animation production more efficient and realistic. By utilizing AI technology, filmmakers can automate many tedious tasks, such as background drawing, character animation, and special effects production. AI can also help artists quickly generate multiple design solutions, thereby accelerating creative iteration and optimization. Most importantly, AI technology can automatically adjust and optimize animation parameters to generate the most realistic visual effects. 3D animation scene design has become a core component of modern multimedia and entertainment industries. In order to meet the growing needs of users and complex visual experiences, it is particularly important to combine deep learning and information security technologies in the design of 3D animation scene graphics. Tang [12] discussed how to use deep learning to improve the design efficiency and quality of 3D animation scenes while ensuring information security during the design process. Deep learning is a branch of machine learning that automatically extracts useful features from a large amount of data by simulating the workings of human brain neural networks, enabling efficient learning and inference. In 3D animation scene design, deep learning can be applied to texture generation, object recognition, scene optimization, and other aspects, greatly improving the automation and intelligence level of design. In the process of designing 3D animation scenes, deep learning and information security technologies can be combined to jointly improve design efficiency and quality.

The combination of wearable technology and 3D printing technology is gradually changing our way of life. Among them, wearable 3D machine weaving technology, with its unique advantages, is becoming an important direction for the development of smart textiles in the future. This technology not only allows for personalized customization based on human body shape but also utilizes animation automatic generation technology to present dynamic effects while covering real-world objects. Wu et al. [13] analyzed the combination of wearable 3D machine weaving technology with 3D printing and textile technology. Through precise mechanical control, fibre materials are stacked layer by layer to form textiles with specific shapes and structures. Automatic animation generation technology utilizes computer algorithms to dynamically simulate and render the woven fabric, enabling it to present a coherent animation effect when covering objects. The advantages of wearable 3D machine weaving technology lie in its high degree of personalization and customization ability, as well as the dynamic visual effects brought by animation automatic generation technology. However, this technology also faces some challenges, such as reducing material costs, improving printing speed, and optimizing animation generation algorithms.

Virtual animation video production involves complex scene construction, character action planning, expression rendering, and other multiple stages. Traditional animation video production methods often require manual participation, which is inefficient and difficult to adapt to diverse needs. Based on deep reinforcement learning, virtual animation video-driven adaptive planning can automatically adjust and optimize the generation process of animation videos through intelligent algorithms, improving production efficiency and quality. Wu et al. [14] explored the principles, applications, and future development directions of virtual animation video-driven adaptive planning based on deep reinforcement learning. Deep reinforcement learning combines the advantages of deep learning and reinforcement learning and achieves perception and decision-making in complex environments by constructing deep neural network models. In virtual animation video production, deep reinforcement learning models can learn to extract useful features from raw video data and generate corresponding animation effects based on these features. Meanwhile, by interacting with the environment, the model can continuously optimize its action planning strategy and achieve adaptive animation video driving. Computer-aided graphic design (CAD), as an important tool in the design field, combined with virtual reality technology, has brought revolutionary changes to the creation of 3D animation scenes. Zhao and Zhao [15] explore how to use computer-aided graphic design technology to optimize and create 3D animation scenes for virtual reality. Virtual reality technology provides users with a completely different visual experience from traditional two-dimensional screens, thanks to its immersive experience. In the process of creating 3D animation scenes, computer-aided graphic design technology plays a crucial role. CAD technology can not only improve design efficiency but also ensure the accuracy and precision of design, providing high-quality animation scenes for virtual reality applications. Using CAD technology, designers can quickly and accurately create 3D scene models. These models can include elements such as terrain, architecture, vegetation, etc., providing a foundation for subsequent animation rendering and interactive design.

The development of animation special effects automatic generation algorithms cannot be separated from the support of computer technology, machine learning (ML), and AI technology. By combining CAD, RL, and CV algorithms, we can achieve automated and intelligent generation of animation special effects, bringing new opportunities and challenges to the development of the animation industry. In the future, with the continuous progress of technology and the expansion of application scenarios, the automatic generation algorithm of animation special effects will play a greater role in the animation industry, promoting the sustained prosperity and development of the animation industry.

3 CAD AND RL

3.1 CAD

With the continuous development of the Internet and computer technology, the level of animation production and the quality of special effects production have significantly improved, promoting the continuous development of animation technology, as shown in Figure 1.

CAD, which stands for Computer-Aided Design, employs software and technology to enhance the design process. This technology finds widespread application in multiple design domains, such as architecture, machinery, electronics, and aerospace. By leveraging CAD, designers can effortlessly create, modify, analyze, and refine design solutions digitally, thereby enhancing efficiency, precision, and quality. Furthermore, in the realm of automated animation special effects generation, CAD holds significant importance. Specifically, it enables the creation of precise 3D models and animated scenes. Designers can utilize CAD software to fashion realistic virtual environments encompassing terrains, buildings, props, and more, thus providing a solid foundation for the generation of animation special effects. These 3D models can then be seamlessly integrated into special effects generation software, serving as the basis for a range of special effects applications.

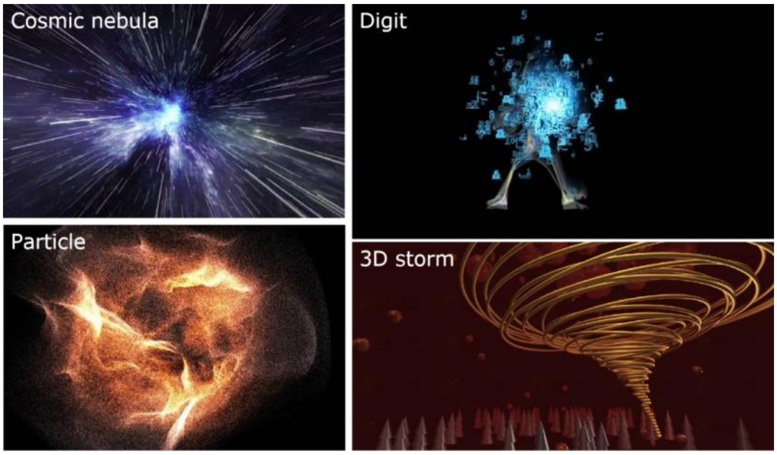

Figure 1: Example of animation special effects.

Secondly, CAD technology can be used to define and describe the structure and motion of animated characters. Through CAD software, designers can create precise models of skeletal systems, muscles, and skin, thereby defining the motion patterns and animation effects of characters. These data can be used as input for animation special effects generation software to generate special effects that match character settings and storylines. In addition, CAD technology can also be used to simulate and analyze the effects of animation special effects. Through CAD software, designers can simulate and test special effects to predict and adjust their effects and performance. This simulation and analysis process can help designers optimize the parameters and strategies of special effects to improve the quality and realism of special effects. In summary, CAD technology has broad application prospects in the field of automatic generation of animation special effects. By utilizing CAD technology, designers can create and optimize animation special effects more efficiently and accurately, bringing new opportunities and challenges to the development of the animation industry.

3.2 RL

RL is a special ML method whose core feature is that agents learn how to achieve a certain goal or maximize a certain benefit through interaction with the environment. Its basic structure is shown in Figure 2.

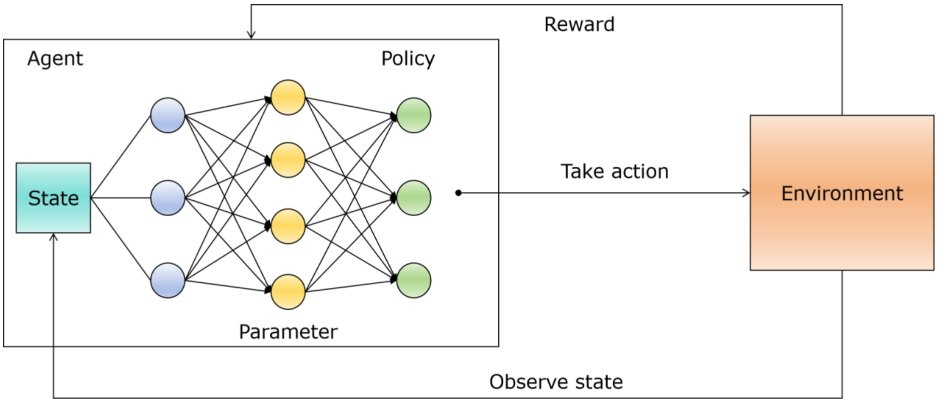

Figure 2: RL structure.

Compared with other ML paradigms, such as supervised learning and unsupervised learning, RL does not require pre-annotated data but learns and improves its behavioural strategies through trial-and-error processes between agents and the environment. In RL, the agent executes a series of actions and observes the response of the environment to these actions, which is usually fed back to the agent in the form of a numerical reward. The goal of an intelligent agent is to find a strategy that guides it in selecting actions to maximize cumulative rewards over a long period. This process typically involves balancing immediate rewards with potential future rewards, which is typically achieved through a discount factor that determines the importance of future rewards in current decisions. A key component of RL is the balance between exploration and utilization. Exploration means trying new, unproven actions or strategies to discover possible better solutions, And utilization refers to selecting the optimal action based on the current knowledge. Effectively balancing these two is crucial for the success of RL. In addition, RL algorithms are usually divided into two categories: value iteration and policy iteration. Value iteration approximates the optimal strategy by continuously updating the state value function, and strategy iteration is to directly iterate and improve the strategy until it converges to the optimal strategy. RL has a wide range of applications, including but not limited to robot control, game AI, natural language processing, financial transactions, etc. In these fields, RL demonstrates its powerful learning and adaptive capabilities, enabling agents to achieve efficient goal-oriented behaviour in complex and uncertain environments.

4 ALGORITHM DESIGN

4.1 The Application of CAD, RL, and CV Algorithms in Animation Special Effects

Animation, as a medium that integrates technology and art, constantly seeks innovation and breakthroughs at the intersection of technology and art. 2D animation and 3D animation, as the two main branches of animation, both rely on advanced technological support to achieve high-quality visual effects. 2D animation has the advantages of being simple, intuitive, and vivid, while 3D animation has more complex and realistic characteristics, which can better represent the real state of things. With the development of society, 3D animation is increasingly favoured by people. Computer digital technology is the primary foundation for 3D animation, which better presents the simulation process by fully utilizing digital technology. With the continuous updates of modern computer software and the emergence of new technologies, the technical content of animation special effects in the film and television industry is increasing, and the technical requirements for production personnel are also increasing. In this context, the application of technologies such as CAD, RL, and CV is particularly important. The application of CAD technology in animation special effects, especially in modelling and scene design, provides powerful tools for animators. Through CAD software, animators can accurately create 3D models and animated scenes, ensuring the realism and accuracy of special effects. This not only improves the efficiency of animation production but also reduces the cost of artistic design.

The application of the RL algorithm in animation special effects generation makes the process of special effects generation more intelligent and automated. By defining an appropriate reward function, the RL algorithm can guide agents to autonomously learn and optimize special effects generation strategies in virtual environments. This method can not only improve the efficiency of special effects generation but also generate more natural and realistic special effects, thereby enhancing the viewing and appeal of animation. The application of the CV algorithm in animation special effects is mainly reflected in the recognition, tracking, and synthesis of special effects. By utilizing the CV algorithm, automatic analysis and evaluation of special effects can be achieved, ensuring the quality and realism of the effects. In addition, the CV algorithm can also be used to achieve real-time tracking and synthesis of special effects, making them more accurately fit into animated characters and scenes, further improving the realism and viewing quality of the animation. The application of technologies such as CAD, RL, and CV in animation special effects not only improves the efficiency and quality of animation production but also provides animators with more creative possibilities and artistic expression. With the continuous development and improvement of these technologies, I believe that future animation works will be more exciting.

4.2 Algorithm Selection

Before creating a 3D animation, it is necessary to use 3D modelling techniques to construct a suitable scene. Building a flat shape that can accommodate two-dimensional images is an important step in this process. In order to ensure the accurate position and deformation of two-dimensional images in three-dimensional space, it is indeed necessary to detect the distribution of all nodes in the plane where the animated image is located and record the node data well. When selecting the location of 3D testing points and targets, multiple factors need to be considered. Firstly, there should be enough test points to ensure sufficient overlap and reconstruction areas on each test node. This helps to improve the accuracy and stability of the model. Secondly, there should be at least four commonly used targets between every two test points, which can increase the correlation and consistency between the models. By selecting appropriate testing points and target positions, an accurate and stable 3D model can be constructed, providing a solid foundation for subsequent animation production. The process of detecting nodes is as follows:

$$S = f^{m+k} * \; f-1 \; * Q \, / \; f^{n} -1 \tag{1}$$

The formula S represents the 3D detection node, f represents a three-dimensional scene map, Q represents continuity among different recognition nodes, m represents node shape, k represents the number of recognized nodes, n and the inline relationships of nodes at different levels in 3D imaging.

According to equation (1), use x, y as the spatial coordinate axis of the planar image to construct a two-dimensional plane of the animated image. The construction process is as follows.

$$\lim_{x,y \to 0}\left(\frac{1}{x} - \frac{1}{y^{x}}\right) = \frac{x,y}{2} \tag{2}$$

The formula x represents the horizontal axis of the two-dimensional plane, y which represents the vertical axis. Complete the construction of the 2D animation plane based on the above calculation formula.

The motion module plays a crucial role in animation special effects design, as it directly affects the smoothness, realism, and audience viewing experience of the animation. The character motion and camera motion modules are the two core components of the motion module. The character motion module is responsible for handling the actions and postures of characters in animations. In order to enhance the motion effects of animation, it is necessary to carefully design the character's movement trajectory and rhythm based on their personality, emotions, and plot needs. This includes setting animation node coordinates, which define the position and posture of the character at keyframes to ensure the natural and smooth movement of the character. The camera motion module is responsible for controlling the camera's motion trajectory and angle to present the best visual effect. The movement of the camera can guide the audience's gaze, highlight important scenes and characters, and enhance the rhythm and expressiveness of the animation. Similarly, by setting the coordinates of animation nodes, the camera's motion trajectory and angle can be precisely controlled to achieve the desired visual effect. In the design process of the sports module, the rhythm of movement is a key evaluation criterion. A reasonable rhythm of movement can make animations more vivid and interesting, enhancing the audience's viewing experience.

The point P_l is used in the animation image where the sub-pixel point is u, v , and its corresponding single-world 3D coordinate is X_w, Y_w, Z_w . In the animation special effects, the camera coordinates in the radial and tangential distortions obtained from the sub-pixel point are normalized. The formula is:

$$\begin{bmatrix} x_d(1) \\ x_d(2) \\ 1 \end{bmatrix} = \begin{bmatrix} a_x & 0 & u_0 \\ 0 & a_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{3}$$

In the formula, the scale factor of the animation special effects image on the X axis is represented by a_x; The scale factor within the Y axis is represented by a_y; The coordinates of the animation special effects center are represented by u_0, v_0; The points on the left side of the X, Y axis are represented by $x_d(1), x_d(2)$; x_n represents the normalized animation effect coordinates. To achieve radial and tangential distortion compensation, multiple iterations are required. When the initial value is x_d, the corresponding normalized animation effect coordinates are:

$$x_d = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_d(1) \\ x_d(2) \end{bmatrix} \tag{4}$$

The return function plays a crucial role in RL. It is a reward signal obtained during the interaction between the intelligent agent and the environment, used to guide the learning direction and strategy optimization of the intelligent agent. The reward function defines the reward that an intelligent agent receives after performing an action in a specific state, which can be positive (representing a successful or good outcome) or negative (representing a failed or bad outcome).

$$R_t = w^I R_t^I + w^G R_t^G \tag{5}$$

Among them, R_t^I and R_t^G represent imitation rewards and adaptive penalties, respectively; w^I and w^G are the corresponding weights for both.

Let X_e, Y_e, Z_e be the on-site coordinate of the camera, and the real-world coordinate system can be converted to the camera coordinate system using the following formula:

$$Z_e \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = Z_e \begin{bmatrix} X_e / Z_e \\ Y_e / Z_e \\ 1 \end{bmatrix} = R \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + t \tag{6}$$

In the formula, the external translation vector of the camera is represented by t, and the selection matrix corresponding to this vector is represented by R. The depth coordinates of the candidate network nodes for the final animation special effects can be calculated using the formula (6).

In complex animation scenes, the CV algorithm can be used to recognize and track the position and direction of objects, and this information can be used for subsequent special effects generation. SiamFC models visual tracking as a similarity learning problem, using a twin network structure to extract image features. By using a convolutional neural network with weight sharing to extract image features, two feature maps of different resolutions can be obtained, and a cross-correlation operation is used to calculate the similarity between the template image and the search area features, ultimately outputting a dense response map:

$$f(x, z) = \phi_\theta(x) * \phi_\theta(z) + b \tag{7}$$

Among them $*$ represents cross-correlation and b represents bias term.

The network model is usually trained end-to-end using logical loss, which is formalized as:

$$L_{\log ist}(y, f) = \sum_{x_i \in \chi} \log\left(1 + e^{-y[x_i] \cdot f[x_i]}\right) \tag{8}$$

Where χ represents the set of positive and negative instances.

5 RESULT ANALYSIS AND DISCUSSION

To assess the efficacy of the algorithm proposed in this article, it is imperative to conduct experiments. We employ the OTB-100 dataset, a publicly available tracking dataset, to evaluate the performance of our algorithm. Figures 3 and 4 depict the distance accuracy and overlap success rate, respectively, of various algorithms on the OTB-100 dataset. Distance accuracy is typically gauged by the center position error, which computes the Euclidean distance between the target position predicted by the algorithm and the true position. On the other hand, the overlap success rate is evaluated using the overlap rate of bounding boxes. This is typically calculated as the ratio of intersection to union between the predicted bounding boxes generated by the algorithm and the actual bounding boxes.

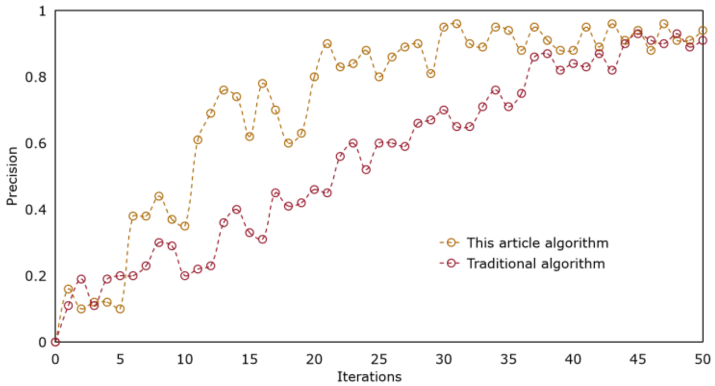

Figure 3: Comparison of precision of different algorithms.

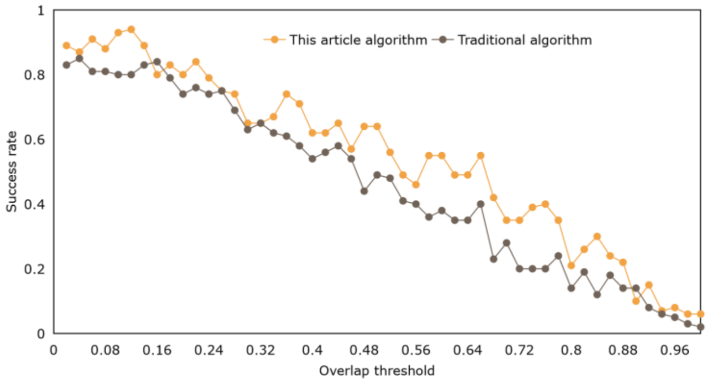

Figure 4: Comparison of success rates of different algorithms.

Upon analyzing Figures 3 and 4, it becomes evident that the algorithm introduced in this paper exhibits noteworthy improvements in both distance accuracy and overlap success rate when compared to conventional algorithms. This enhancement signifies that the proposed algorithm is more adept at guiding search strategies, enabling precise target position localization and mitigating the likelihood of erroneous tracking.

Figure 5 presents a comparison of recall rates among various algorithms on the dataset. Recall rate, a crucial metric for evaluating target tracking algorithms, quantifies the algorithm's ability to correctly track targets. Specifically, it measures the proportion of successfully tracked frames by the algorithm relative to the total number of frames. Upon inspection of Figure 5, it becomes evident that the algorithm introduced in this article exhibits a significant boost in recall rate compared to traditional algorithms. This implies that the proposed algorithm is superior in precisely identifying and tracking targets during tracking tasks, thereby minimizing target loss or tracking failures. This enhanced performance can be attributed to the introduction of a robust motion model, which offers more precise predictions of the target's position and motion trajectory. In addition, the algorithm may also adopt an accurate model update strategy to avoid model drift issues and improve tracking stability.

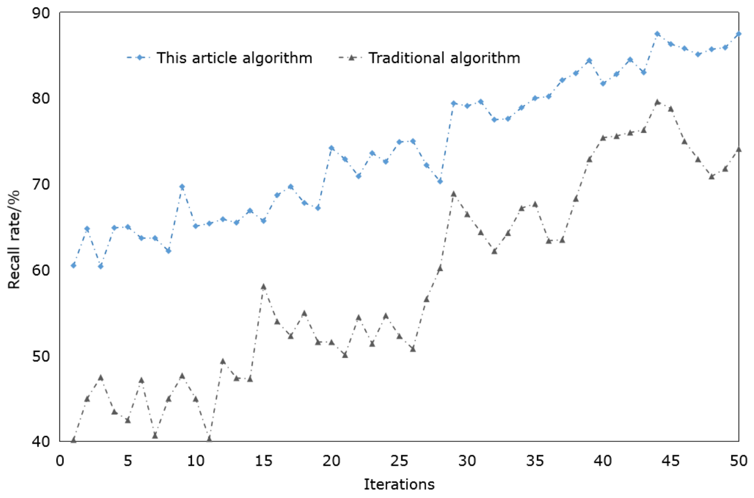

Figure 5: Comparison of recall rates.

Figure 6 shows a comparison of the cumulative return growth curves of different algorithms during the training process. The cumulative return growth curve is an important indicator for measuring the learning effectiveness and convergence speed of algorithms. In RL tasks, agents maximize cumulative returns by interacting with the environment, so the growth trend of cumulative returns can reflect the learning effectiveness and performance of the algorithm.

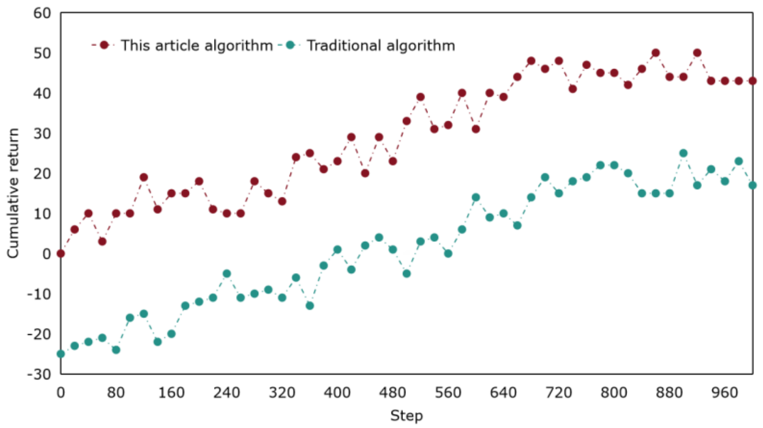

Figure 6: Comparison of cumulative return growth curves.

In Figure 6, it can be observed that there are significant differences in the cumulative return growth curves of different algorithms. The algorithm proposed in this article shows a rapid growth trend in the early stages of training, which means that the algorithm can quickly learn and optimize its strategies from the environment. As the training progresses, the cumulative return of the algorithm in this article continues to increase and gradually converges to a higher level. In contrast, the cumulative return growth curve of traditional algorithms is relatively flat, with a slower convergence speed, and the final convergence return value is also lower. The comparative results of the cumulative return growth curve validate the effectiveness of the algorithm proposed in this paper in RL tasks. The faster convergence speed and higher return value indicate that our algorithm can learn the optimal strategy faster and achieve better performance.

To verify the practical value of the algorithm in digital art production, we selected a newly recorded video as the test object. This video content has rich colours and dynamic scenes, making it suitable for evaluating the performance of algorithms in handling complex scenes. According to the evaluation results, the video clarity obtained by the algorithm in this article ranges from 70% to 80%, with an average value of 75%. The video clarity obtained by traditional algorithms ranges from 60% to 70%, with an average value of 65.5%. From the data, it can be seen that the clarity obtained by the algorithm in this article is about 9.5 percentage points higher than that obtained by traditional algorithms. This result validates the practical value of the algorithm proposed in this paper in digital art production. In practical applications, higher image clarity can bring a better visual experience and enhance the quality and viewing value of digital artworks. In addition, the algorithm in this article can also maintain high performance when dealing with complex scenes, indicating its strong robustness and adaptability. Figure 7 shows the comparison results between the traditional algorithm and the algorithm proposed in this paper.

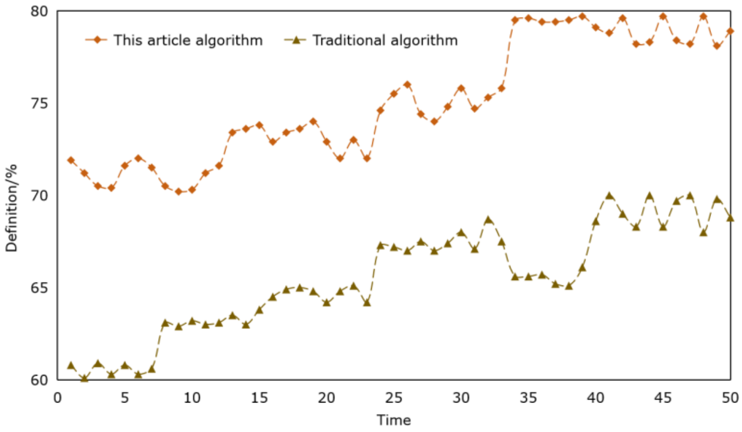

Figure 7: Clarity comparison.

6 CONCLUSION

With the rapid progress of modern science and information technology, the position of digital animation design in the film production industry is increasingly prominent. As a crucial aspect of animation design, special effects art not only enriches the visual expression of animation but also enhances the overall artistic expression effect. In the current technological context, combining CAD, RL, and CV algorithms, this paper proposes an innovative algorithm for the automatic generation of animation special effects, aiming to promote sustained innovation and development in the animation industry. The organic combination of CAD, RL, and CV algorithms has achieved automatic generation and real-time adjustment of special effects through data-driven and intelligent optimization. This algorithm not only lowers the threshold for animation production, allowing more creators to participate in the animation industry but also brings more innovative possibilities and business opportunities to the animation industry.

Although this study has achieved certain results, we still need to recognize that there are many challenges and problems in the field of automatic generation of animation special effects that we need to face and solve. Although our algorithm can generate satisfactory special effects in most cases, its performance may still need improvement in some complex or special scenarios. The efficiency and stability of algorithms are also areas that we need to focus on and improve in the future. In practical applications, how to ensure that algorithms can still maintain stability and efficiency when processing large-scale data or real-time rendering is a problem worth studying.

7 ACKNOWLEDGEMENT

This work was supported by Science and Technology Research project of Jiangxi Provincial Department of Education, Application Research of Virtual Reality (VR) Technology in Jiangxi Nuo Culture Brand Creative Industry, Project number: GJJ214912.

Yang Liu, https://orcid.org/0009-0007-3182-962X
Lei Li, https://orcid.org/0009-0000-8834-6628
Xiang Lei, https://orcid.org/0009-0000-6595-9691

REFERENCE

[1] Aberman, K.; Weng, Y.; Lischinski, D.; Cohen, O.-D.; Chen, B.: Unpaired motion style transfer from video to animation, ACM Transactions on Graphics (TOG), 39(4), 2020, 64:1-64:12. https://doi.org/10.1145/3386569.3392469

[2] Bao, W.: The application of intelligent algorithms in the animation design of 3d graphics engines, International Journal of Gaming and Computer-Mediated Simulations, 13(2), 2021, 26-37. https://doi.org/10.4018/IJGCMS.2021040103

[3] Ben, A.-H.; Jovančević, I.; Orteu, J.-J.; Brèthes, L.: Automatic inspection of aeronautical mechanical assemblies by matching the 3D CAD model and real 2D images, Journal of Imaging, 5(10), 2019, 81. https://doi.org/10.3390/jimaging5100081

[4] Bian, S.; Grandi, D.; Liu, T.; Jayaraman, P.-K.; Willis, K.; Sadler, E.; Li, B.: HG-CAD: hierarchical graph learning for material prediction and recommendation in computer-aided design, Journal of Computing and Information Science in Engineering, 24(1), 2024, 1-14. https://doi.org/10.1115/1.4063226

[5] Bodini, M.: A review of facial landmark extraction in 2D images and videos using deep learning, Big Data and Cognitive Computing, 3(1), 2019, 14. https://doi.org/10.3390/bdcc3010014

[6] Cao, Q.; Zhang, W.; Zhu, Y.: Deep learning-based classification of the polar emotions of" moe"-style cartoon pictures, Tsinghua Science and Technology, 26(3), 2020, 275-286. https://doi.org/10.26599/TST.2019.9010035

[7] Ding, W.; Li, W.: High speed and accuracy of animation 3d pose recognition based on an improved deep convolution neural network, Applied Sciences, 13(13), 2023, 7566. https://doi.org/10.3390/app13137566

[8] Feng, Y.; Feng, H.; Black, M.-J.; Bolkart, T.: Learning an animatable detailed 3D face model from in-the-wild images, ACM Transactions on Graphics (ToG), 40(4), 2021, 1-13. https://doi.org/10.1145/3450626.3459936

[9] Gong, M.: Analysis of architectural decoration esthetics based on VR technology and machine vision, Soft Computing, 25(18), 2021, 12477-12489. https://doi.org/10.1007/s00500-021-05986-w

[10] Poggi, M.; Tosi, F.; Batsos, K.; Mordohai, P.; Mattoccia, S.: On the synergies between machine learning and binocular stereo for depth estimation from images: a survey, IEEE Transactions on

Pattern Analysis and Machine Intelligence, 44(9), 2021, 5314-5334. https://doi.org/10.1109/TPAMI.2021.3070917
[11] Reddy, V.-S.; Kathiravan, M.; Reddy, V.-L.: Revolutionizing animation: unleashing the power of artificial intelligence for cutting-edge visual effects in films, Soft Computing, 28(1), 2024, 749-763. https://doi.org/10.1007/s00500-023-09448-3
[12] Tang, J.: Graphic Design of 3D animation scenes based on deep learning and information security technology, Journal of ICT Standardization, 11(3), 2023, 307-328. https://doi.org/10.13052/jicts2245-800X.1135
[13] Wu, K.; Tarini, M.; Yuksel, C.; McCann, J.; Gao, X.: Wearable 3D machine knitting: automatic generation of shaped knit Sheets to cover real-world objects, IEEE Transactions on Visualization and Computer Graphics, 28(9), 2021, 3180-3192. https://doi.org/10.1109/TVCG.2021.3056101
[14] Wu, Y.; Zhang, Z.; Qiu, D.; Li, W.; Su, Z.: Video driven adaptive grasp planning of virtual hand using deep reinforcement learning, Multimedia Tools and Applications, 82(11), 2023, 16301-16322. https://doi.org/10.1007/s11042-022-14190-3
[15] Zhao, J.; Zhao, X.: Computer-aided graphic design for virtual reality-oriented 3D animation scenes, Computer-Aided Design and Applications, 19(S5), 2022, 65-76. https://doi.org/10.14733/cadaps.2022.S5.65-76