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

Animation Scene Object Recognition and Modeling Based on Computer Vision Technology

Zhengzhong Shen[1] and Wei Zhang[2]

[1]Art College, Chongqing Technology and Business University, Chongqing 400060, China, 2007051@ctbu.edu.cn
[2]School of Animation And Digital Film, Chongqing Engineering Institute, Chongqing 400056, China, zhangwei18@ctbu.edu.cn

Corresponding author: Zhengzhong Shen, 2007051@ctbu.edu.cn

Abstract. Based on computer vision technology, this article proposes a computer-aided design (CAD) modeling method for animation scene objects based on recognition. It conducts a series of experiments to verify the effectiveness and practicality of this method. This method aims to automatically and accurately recognize objects from animation scenes and generate corresponding CAD models using object recognition algorithms and CAD modeling techniques. To achieve this goal, the article first uses the Convolutional Neural Network (CNN) algorithm for object recognition and obtains the boundary information of the object. Then, CAD modeling uses boundary information based on recognition to generate CAD models similar to real objects. The performance of the system was assessed through a series of experiments. The results show that this method has a high accuracy in object recognition, with an essential accuracy of over 93%, and can accurately extract boundary information of objects. The gap between the generated CAD model and the real object is small in CAD modeling, and the modeling accuracy is high. Moreover, the system's running time is within an acceptable range, proving the system's efficiency. This method can provide animators with an automated CAD modeling tool, reducing the time and effort costs of manual modeling and improving modeling accuracy.

Keywords: Computer Vision Technology; Animation Scenes; Object Recognition; Deep Learning; CAD Modeling
DOI: https://doi.org/10.14733/cadaps.2024.S15.16-34

1 INTRODUCTION

In the digital entertainment industry, animated films and game production have become important components. For animation and game production, the recognition and modeling of scene objects is one of the core links. In animation production, the recognition and modeling of scene objects determine the visual effect and realism of the animation. By identifying and modeling objects in the scene, real-world objects and environments can be transformed into animated scenes with artistic and dynamic effects. At the same time, the modeling of scene objects also involves the implementation of lighting, materials, textures, and other effects in animation production, which is crucial for improving the quality and realism of animation. Baimukashev et al. [1] used computer vision synthesis technology to recognize and extract objects in animation scenes. Utilize deep learning algorithms to automatically model the recognition results, and use CAD technology to achieve refined modeling and visualization of the model. Combining deep learning algorithms for automatic modeling of recognition results. By using techniques such as 3D reconstruction and reverse engineering, a 3D model of objects in the scene is automatically constructed based on the recognition results. At the same time, neural network algorithms are used to optimize and refine the model, in order to improve its accuracy and precision. Traditional methods often rely on manual operations and the involvement of professional designers, which is not only time-consuming but also costly. Traditional animation scene modeling requires manual work, which is labor-intensive and inefficient. Modelers need to manually create 3D models, perform a series of tedious tasks such as detail adjustments, material mapping, and lighting rendering, which consume a lot of time and effort. Manual modeling is difficult to achieve high-precision modeling, especially for complex scenes and works with high requirements for details, and the accuracy of modeling is often difficult to meet the requirements. Traditional animation scene modeling is usually done in static scenes, making it difficult to achieve dynamic modeling and real-time updates. Meanwhile, traditional methods also lack flexibility for modeling complex scenes and dynamic objects. To address this issue, Barreto et al. [2] utilized visual technology to recognize and extract objects in animation scenes, including feature information extraction such as color, shape, and texture. Using deep learning algorithms to automatically model recognition results, including techniques such as 3D reconstruction and reverse engineering, to achieve automated and high-precision modeling. Deep learning algorithms are a key technology for achieving automatic animation modeling, which requires selecting appropriate algorithms based on animation scenes and considering how to optimize algorithms to improve modeling efficiency and accuracy. The realism and immersion of virtual reality environments are key to user experience, and attention needs to be paid to the detailed presentation and immersive experience of the environment. With the growth of computer vision technology, automatic and semi-automatic object recognition and modeling methods have gradually become a research hotspot. Automatic and semi-automatic animation object recognition and modeling methods can improve the effectiveness of animation production. Computer vision technology can accurately recognize and extract object features in images or videos, and use them as a basis for modeling. This makes animation production more refined and realistic, improving the audience's viewing experience. In addition, with the continuous development of computer vision technology, its application in animation object recognition and modeling has also been further expanded. For example, object recognition methods based on deep learning can learn higher-level feature representations from a large amount of image data, thereby achieving more accurate object recognition. Meanwhile, by combining different feature extraction methods and classifiers, the accuracy and robustness of object recognition can be further improved. Gong [3] used visual technology to automate the evaluation of architectural decoration animation images, achieving more accurate extraction and recognition of target objects. By preprocessing the input image or video, including image denoising, enhancement, segmentation, and other operations, interference is reduced and recognition accuracy is improved.

Guo and Ma [4] utilize advanced image processing and pattern recognition algorithms to achieve accurate and fast recognition of various objects in animation scenes. Through this technology, computers can automatically analyze and understand different objects that appear in animation scenes, and can provide corresponding feedback and interaction based on user input. In this system, the computer will obtain animated image information of the scene through cameras or other sensors, and convert it into digital signals for processing. By utilizing advanced algorithms such as deep learning, computers can perform feature extraction, object detection, and classification operations on animated images, thereby achieving accurate identification of different objects. For example, in a virtual gaming world, when a cat appears in front of a player, the system can quickly identify it as a cat and respond accordingly. Computer vision technology provides new ideas and methods for human-computer interaction in animation scenes, enabling the creation of richer and more realistic animation scenes and interactive experiences. Computer sensor technology provides new ideas and methods for object recognition and capture in animation scenes, which can create richer and more realistic animation scenes and interactive experiences. In film and television production, this technology can be used to capture actors' actions, expressions, and other information, which can be used for the design of animation characters' actions and expressions. In game development, this technology can be used to capture information such as user actions and expressions, achieving a more realistic gaming experience. In virtual reality, this technology can be used to create realistic virtual scenes and objects, improving the immersion and interactivity of virtual reality. Herdog et al. [5] used computer vision technology and machine learning algorithms to recognize and classify processed sensor animation signals. Using computer graphics and CAD technology, extract the pose, shape, and other information of the object, and perform corresponding processing and transformation. Utilizing this technology to achieve more efficient 3D animation production, reducing manual involvement through automation or semi automation, and improving production efficiency and quality. Deep learning (DL) technology, especially CNN, has demonstrated its powerful ability in these tasks. In the aspect of object recognition in animation scene, the method based on DL has achieved some success, but it still faces challenges in recognition accuracy and complex scene processing. Computer aided design (CAD) and 3D reality technology have been widely applied in various fields. In animation design, the introduction of these technologies provides designers with new tools and methods, making the design of animated characters and the construction of animated scenes more convenient and efficient. Jing and Song [6] explore the application of 3D reality technology and CAD in animation design. By utilizing 3D reality technology, designers can accurately model animated characters as needed. By adjusting the shape, size, and muscle lines of characters, realistic animated characters can be created. 3D reality technology can capture and simulate the expressions and actions of characters in real-time, making their movements more natural and smoother and improving the viewing quality of animation. Through 3D reality technology, designers can endow characters with realistic materials and lighting effects, making animation scenes more realistic and trustworthy. Taking a 3D animated movie as an example, the designer uses 3D reality technology to model characters, design expressions and actions, and uses CAD software to bind bones and render special effects. The final presentation is an animated work with stunning visual effects and profound emotional expression.

In the aspect of CAD modeling, traditional modeling methods usually require the professional knowledge and experience of designers. In recent years, image-based modeling methods have attracted more and more attention. By extracting geometric and texture information of objects from single or multiple images, these methods can automatically generate or assist in generating CAD models. However, how to effectively transform the recognized objects into accurate CAD models is still an open problem. The goal of this study is to use computer vision technology to automatically and accurately identify the objects in the animation scene and generate the corresponding CAD model. In order to achieve this goal, the research contents include the automatic recognition of objects in animated scenes, feature extraction and representation of objects, design of transformation algorithm from recognition result to CAD model, and system integration and simulation verification. The innovations of this paper include the following aspects:

A. Applying computer vision technology to object recognition in animation scenes has achieved automatic and accurate object detection and positioning. Compared to traditional manual methods, this method has higher efficiency and accuracy.

B. This study not only focuses on object recognition but also further combines the recognition results with CAD modeling. Automatically generate or assist in generating corresponding CAD models by extracting object features and boundary information. This method can significantly reduce the workload of manual modeling and improve modeling efficiency.

C. The use of DL technology, especially CNN, has improved the performance of object recognition. By training deep neural networks, higher-level object features can be learned, thereby improving object recognition accuracy.

D. This article not only proposes a complete method for object recognition and CAD modeling in animation scenes, but also conducts systematic simulation research and performance assessment. By constructing a simulation environment and conducting experimental verification on the proposed method, its effectiveness and practicality were demonstrated.

This article is divided into seven sections. The first section is an introduction, introducing the background, significance, current status, and development trends of the research. The second section elaborates on the basic theories and methods of computer vision technology. The third section delves into the key technologies and methods of object recognition in animation scenes. The fourth section introduces the basic theory and technology of CAD modeling. The fifth section elaborates on the method of CAD modeling for animation scene objects based on recognition. The sixth section introduces the integration and simulation research of the system. The final section summarizes the entire text and proposes directions and suggestions for future research.

2 FUNDAMENTALS OF COMPUTER VISION TECHNOLOGY

In complex dynamic environments, traditional visual technology for object detection in animated scenes may face some challenges. Objects in complex dynamic environments may have various motion states and interactive behaviors, which increases the difficulty of object detection. Background noise in complex dynamic environments may also interfere with the results of object detection. The disturbance factors in the background may be mistaken for the target object, leading to deviation in the detection results. In order to improve the performance of object detection methods in animation scenes using visual technology in complex dynamic environments, Kapusi et al. [7] attempted to use adaptive color correction methods or use techniques such as deep learning to improve the accuracy of object detection. For example, deep neural networks can be trained to automatically adapt to image processing under different lighting conditions, thereby improving the accuracy of object detection. To address the interference problem of background noise, filtering techniques or deep learning-based background subtraction methods are used to reduce the impact of interference factors. The accuracy of object detection can be improved by using methods such as median filtering to eliminate noise in animated images or using deep neural networks to learn the distinction between background and foreground. In computer-aided animation design, curve modeling is an important technique for creating animation effects with naturally smooth motion trajectories. Curve modeling allows animators to control the shape and motion of the model by adjusting parameters, thereby creating realistic and natural animation effects. In curve modeling, parameter curves or spline curves are usually used to represent the shape and motion of the model. These curves can change the shape and motion trajectory of the model by adjusting their parameters. For character animation, Li [8] controls the character's actions and expressions by adjusting the parameters of the spline curve, thereby achieving various complex actions and expression changes. In short, curve modeling is a very important computer-aided animation design technique that can help animators create realistic and natural animation effects. By using different curve types and physics principles, animators can create various complex and interesting animation effects. Virtual reality CAD visual technology also has extensive applications in object recognition in animation design scenes. In character animation design, virtual reality CAD technology can be used to create realistic character models and expressions. By capturing the facial expressions and actions of real characters, virtual reality CAD technology can apply this information to animated characters, making them more realistic and natural in expression and action. This can help animators create high-quality character animations in a short period of time, improving production efficiency and quality. Li and Li [9] combine animation with interactive elements to provide users with a more immersive experience. In game development, virtual reality CAD technology can be used to create realistic game scenes and characters, while combining user input to control the game's progress and ending. This can help game developers improve the attractiveness and playability of their games. Virtual reality CAD visual technology has a wide range of applications in object recognition in animation design scenes, which can help animators create high-quality animation works in a short period of time, improving production efficiency and quality. At the same time, virtual reality CAD technology can also provide users with a more immersive experience, enhancing the attractiveness and playability of animation. Li et al. [10] analyzed a framework for object recognition and localization in animation scenes based on computer vision technology. This technology can be used to identify and simulate animated objects in the virtual reality world. By using sensors, animated images or point cloud data of virtual objects can be obtained, and then computer vision technology can be used for analysis and processing to achieve object recognition and positioning. This can help animators combine animation scenes with the real world to achieve more realistic animation effects. Finally, animation scene object recognition and CAD modeling based on computer vision technology can be combined with virtual reality technology to provide users with a more immersive experience. By combining virtual reality technology with physical object recognition and positioning frameworks, users can interact and operate with virtual objects in the real world, achieving more intelligent animation production and display. The deep learning animation target recognition and localization based on multi perspective image acquisition can be combined with virtual reality technology to provide users with a more immersive experience and interactive operations. By combining virtual reality technology with multi perspective image acquisition and deep learning technology, users can observe and operate target objects in a virtual environment, achieving more intelligent animation production and display. Lin et al. [11] used deep learning analysis methods to perform more intelligent analysis of target objects in three-dimensional animation environments. Using a large amount of multi view image data for training to improve the recognition accuracy and generalization ability of the model. Through multi perspective image acquisition, we can obtain more comprehensive, three-dimensional, and detailed target information. Compared to traditional single view image acquisition methods, multi view image acquisition can provide more diverse and robust data inputs, thereby enhancing the adaptability of target recognition and positioning systems in complex environments. In cross disciplinary fields, intelligent model construction recognition technology can be utilized to integrate the advantages of animation art design and CAD technology. By using machine learning and computer vision techniques, it is possible to automatically recognize and extract key elements in animation scenes, such as characters, props, backgrounds, etc., and use CAD technology for precise modeling and layout. At the same time, intelligent model construction recognition technology can be used to combine animation art design with CAD technology, achieving more intelligent and efficient animation production. Intelligent model construction recognition technology can also be used for the automatic generation and optimization of animation scenes. Liu and Liu [12] analyze a large amount of animation scene data to train machine learning models to generate animation scenes that meet user needs automatically. At the same time, intelligent model construction recognition technology can be used to optimize and improve the generated animation scenes in order to improve their quality and expressiveness.

Self-supervised learning is a machine learning method that can train on unlabeled data, thereby saving a lot of manpower and resources. Self-supervised learning is also widely used in object recognition and CAD modeling in animation scenes. Firstly, in terms of object recognition in animation scenes, self-supervised learning can be used to identify various objects in animation scenes. By utilizing unlabeled animation data for training, self-supervised learning can learn various features and patterns in animation scenes, and test and recognize them on new animation data. This can help animators quickly and accurately identify the objects that need to be edited, improving production efficiency and quality. Secondly, in CAD modeling, self-supervised learning can be used to automate the modeling of animation scenes and characters. By utilizing unlabeled CAD data for training, self-supervised learning can learn the laws and techniques of CAD modeling, and automatically establish models that meet the requirements. This can help animators quickly create realistic animation scenes and characters, improving production efficiency and quality [13]. CAD controllers play a crucial role in animation production. CAD technology can be used to create and edit 3D models, while controllers can be used to drive the motion and behavior of these models. In animation visual hybrid virtual debugging, CAD controllers can be used to identify and locate target objects, as well as adjust and control the visual effects of animation. Secondly, computer vision technology can be combined with CAD controllers to achieve more efficient and accurate object recognition and positioning. Computer vision technology can be used to analyze image and video data to identify and extract the features and contours of target objects. This information can be shared with CAD controllers for more precise control and adjustment. In hybrid virtual debugging, CAD controllers can be used to adjust and control the visual effects of animations. For example, by adjusting the direction and brightness of the light source, the brightness and color saturation of the scene can be changed. Meanwhile, computer vision technology can be used to identify and locate target objects to ensure the accuracy and consistency of animations [14]. Qiu [15] explores the computer program simulation design of ocean 3D animation. Ocean 3D animation is a computer program that simulates physical phenomena such as ocean morphology, water flow, waves, and lighting, generating highly realistic dynamic images. Its basic principles are based on disciplines such as physics, fluid mechanics, optics, etc. By establishing mathematical models, various physical phenomena in the ocean are numerically simulated. Using computer graphics technology, convert the data calculated by mathematical models into images. In this process, visual factors such as lighting, color, and texture need to be considered to achieve a highly realistic visual effect.

Virtual reality CAD simulation for enhancing 3D animation in biological systems is an advanced technology that combines computer-aided design (CAD), computer vision, and virtual reality technology to provide more realistic and accurate simulation and visualization for biological systems. Virtual reality technology can integrate CAD models with information extracted from computer vision, providing users with a more realistic and intuitive simulation environment. In this environment, users can observe and operate three-dimensional animated models of biological systems and understand their structure and function from various perspectives. Rahatabad et al. [16] combined object recognition algorithms with CAD modeling technology to achieve automatic and accurate recognition of objects in animation scenes. Firstly, using object recognition algorithms can automatically identify various objects in the image and extract their position and size information. Then, using CAD modeling technology, an accurate 3D model can be established based on this information, and the model can be precisely edited and adjusted. Finally, the model can be presented in animated scenes through rendering and other techniques to achieve automatic and accurate recognition.

Sami et al. [17] generate highly realistic images and videos by simulating the human visual system and can also perform various analyses and processing on images and videos to extract useful information. In construction progress monitoring, computer animation visual technology can be used to generate simulated images of the construction process and provide real-time information on construction progress. The automated construction progress monitoring system based on computer animation vision has the following advantages: firstly, it can provide real-time construction progress information to help management personnel discover and solve problems in a timely manner; Secondly, it can comprehensively track and analyze the construction process, providing comprehensive construction progress data; Finally, it can improve construction efficiency, reduce construction costs, and shorten construction cycles. The use of 3D synthetic visual technology pose sequences to simulate and predict the behavior of animated targets is a very effective method, which can help us better understand and control the actions and behaviors of animated targets. Firstly, 3D synthetic vision technology can be used to collect image data of animated targets in different poses, which can include information such as the shape, position, and direction of the target. By using these models, we can simulate the reactions and behaviors of animation targets in different situations, thereby better controlling and adjusting the effects and quality of the animation. Torres et al. [18] analyzed that 3D synthetic vision technology can also be used for real-time tracking and recognition of the behavior and actions of animated targets. By using algorithms to process and analyze real-time image data, the pose and behavior of animated targets can be detected and recognized in real time, and the actions and behaviors of animated targets can be adjusted and controlled based on the recognition results. Due to the diverse types and shapes of objects in animation scenes, it is necessary to choose appropriate feature extraction methods and classifier design strategies for object recognition in order to improve the recognition accuracy and generalization ability of the model. At the same time, it is also necessary to continuously update and optimize the model to adapt to the constantly changing needs of animation scenes. CAD visual technology for object recognition in animation scenes is a method that combines computer-aided design (CAD) technology and computer vision technology to identify and extract objects in animation scenes. CAD technology can be used to create and edit various 3D models, including information such as the shape, size, and position of objects. In animation scenes, CAD technology can be used to establish models of various objects, and they can be accurately edited and adjusted as needed. Computer vision technology can be used to analyze image and video data to identify and extract the features and contours of target objects. This information can be used to validate and improve CAD models, making them more accurately reflect actual animation scenes [19]. The object recognition of animation scenes based on CNN visual technology needs to be designed and implemented in conjunction with specific animation production requirements and application scenarios. Zhang et al. [20] trained and optimized the model using appropriate optimization algorithms to improve its accuracy and generalization ability. At the same time, it is also necessary to adjust the parameters and fine-tune the model to perform best in specific animation scene object recognition tasks.

3 OBJECT RECOGNITION IN ANIMATION SCENES

3.1 Overview of Object Recognition in Animation Scene

Computer vision is a science that studies how to enable computers to obtain information, understand content, and make decisions from images or videos. Image processing is a fundamental technology in computer vision, which involves performing various operations on images to enhance their quality or extract useful information. Image processing techniques include image denoising, enhancement, transformation, etc., which can improve the appearance of images and simplify subsequent analysis tasks. Feature extraction is one of the key technologies in computer vision, aiming to extract meaningful information from images for describing and distinguishing different objects or scenes. Common features include texture, shape, color histogram, etc. The methods of feature extraction can be divided into traditional methods and DL methods. Traditional methods are usually based on manually designed feature extractors, while DL methods automatically learn image feature representations through neural networks. Object detection refers to the process of locating and recognizing specific objects in an image or video. It involves finding interesting targets in complex scenes and determining their positions and boundaries. Object detection algorithms are usually based on strategies such as sliding windows, region proposals, or anchor boxes, combined with feature extraction algorithms and classifiers to achieve object detection. Target recognition is a higher-level task that requires not only detecting the position of the target, but also identifying the category of the target. Target recognition typically uses classifiers such as Support Vector Machine (SVM), Decision Tree, or Neural Networks in DL to classify and recognize the extracted features.

DL has made remarkable breakthroughs and successes in the field of computer vision. By using neural network model, it can learn hierarchical feature representation from a large number of image data, thus solving computer vision tasks more effectively. CNN is one of the most commonly used models in DL, and it has achieved excellent performance in image classification, object detection and semantic segmentation. In addition, DL is also applied to image generation, image super-resolution, attitude estimation and so on, which brings new ideas and methods for computer vision research. Object recognition in animation scenes is an important application in the field of computer vision, with the aim of automatically detecting objects from animation scenes and accurately recognizing and locating them. The research and application of object recognition in animation scenes have important value in the digital entertainment industry. It can improve the efficiency of animation production, reduce manual operations, and promote the industry's rapid growth. This task requires a comprehensive utilization of technologies such as image processing, feature extraction, object detection, and recognition to achieve accurate and efficient object recognition.

3.2 Design and Implementation of Object Recognition Algorithms for Animation Scenes

Feature extraction is one of the key steps in object recognition in animation scenes. Using computer vision technology, various object features, including textures, shapes, edges, colors, etc., can be extracted from animated scenes. These features should have good distinguishability and stability to accurately describe the differences between different objects. Moreover, in order to facilitate subsequent classification and recognition, it is also necessary to represent the extracted features appropriately, such as feature vectors, feature maps, etc. Based on the extracted features, algorithms for object recognition in animated scenes can be designed and implemented. Common algorithms include traditional machine learning methods such as SVM and random forest, as well as DL based methods such as CNN. These algorithms can establish a relationship model between object features and object categories through training and learning and achieve accurate recognition of objects in animation scenes. The design and implementation of algorithms need to consider requirements such as accuracy, real-time performance, and robustness. After careful consideration, this article selects the CNN algorithm to achieve object recognition in animation scenes. The model structure is shown in Figure 1.

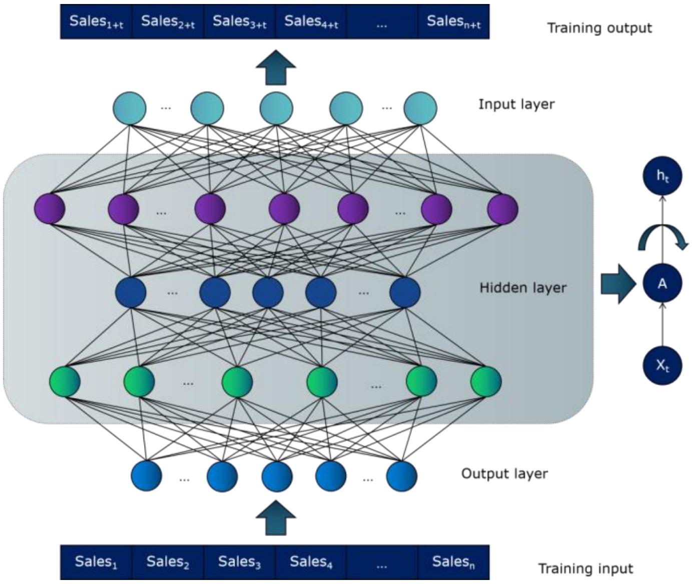

Figure 1: Model structure.

The first step is data collection and preprocessing: collecting images from various animation scenes to ensure that each category has a sufficient number of samples. For each object, label its bounding box and category. Moreover, the image is preprocessed and adjusted to 256x256 pixels, color space conversion (RGB to grayscale), and mean-variance normalization in this article. Then, select a pre-trained CNN model - ResNet-50 as the infrastructure. The pre-trained model has been trained on a large number of images, thus possessing strong feature extraction capabilities. Adjust the fully connected layer to match its output with the number of object categories. For example, if there are 10 object categories, the last layer outputs 10 neurons. In addition, to improve positioning accuracy, additional positioning branches (bounding box regression technology) can be considered. The CNN function in this article is defined as:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} \times k_{ij}^l + b_j^l\right) \tag{1}$$

In the formula, x_i represents the input feature map, k represents the convolution kernel, b represents the deviation term, and the output after convolution is the feature map x_j. The formula of the activation layer after the convolution layer is as follows:

$$F_j^{n+1} = f\ F_j^n \quad (2)$$

Where f is a point-by-point activation function.

The training of the model may adopt an early stop strategy, which means that when the performance of the validation set does not significantly improve in consecutive epochs, the training is stopped to prevent overfitting. Table 1 shows the accuracy, recall, and F1 scores of the algorithm.

Training times	*Accuracy (%)*	*Recall (%)*	*F1 score (%)*
1	87.04	93.89	93.26
2	89.73	96.15	96.06
3	92.03	94.48	96.93
4	83.49	95.67	96.16
5	92.97	96.92	94.06
6	88.90	95.54	96.91
7	86.83	90.22	94.68
8	91.59	96.75	95.59
9	91.67	90.51	94.91
10	89.74	96.79	96.18
Average value	89.399	94.692	95.474

Table 1: Accuracy, recall, and F1 score of the algorithm.

According to the data in Table 1, the accuracy, recall, and F1 score of the algorithm in this paper have all reached over 89%, indicating its good performance in object recognition tasks in animation scenes.

3.3 Experiment and Result Analysis

In order to verify the performance and effectiveness of the designed animation scene object recognition algorithm, experiments, and result analysis are required. Suitable experimental datasets, including samples of various animation scenes and objects, can be constructed. Then, use the designed algorithm to perform object recognition on the experimental dataset and record the recognition results. By qualitatively and quantitatively analyzing the experimental results, the performance of the algorithm can be assessed and compared with other methods. This article uses SVM and decision tree algorithms as comparative methods to analyze the accuracy and error of different algorithms. The same dataset as in the CNN algorithm experiment was used in the experiment to ensure data consistency and fairness. The dataset is divided into a training set and a testing set for training and testing these three algorithms. For SVM and decision tree algorithms, it is needed to extract relevant features from animated scene images. This article adopts traditional image processing methods to extract features such as texture, shape, and color histograms as inputs for these two algorithms. For the SVM algorithm, RBF kernel functions were used, and parameter C and γ Tuning for. For the decision tree algorithm, relevant parameters such as tree depth and splitting criteria were adjusted to achieve optimal performance. The specific experimental results are shown in Figures 2 and 3.

The CNN algorithm proposed in this paper outperforms SVM and decision tree algorithms in terms of accuracy and error. Among them, the SVM algorithm performs well in object recognition tasks in animated scenes, but its accuracy and error are slightly lower compared to the CNN algorithm. This is because the SVM algorithm has limited ability to represent complex image features and cannot automatically learn and extract advanced features like CNN. The decision tree algorithm performs relatively poorly on various assessment indicators. This is because decision trees are prone to overfitting when dealing with high-dimensional and complex image features, resulting in insufficient generalization ability.

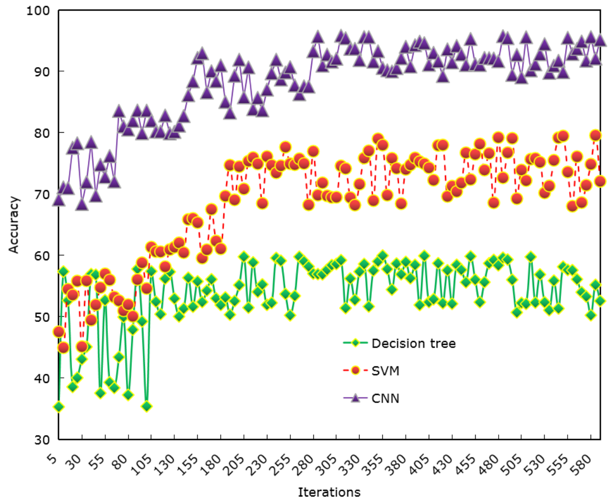

Figure 2: Accuracy of different algorithms.

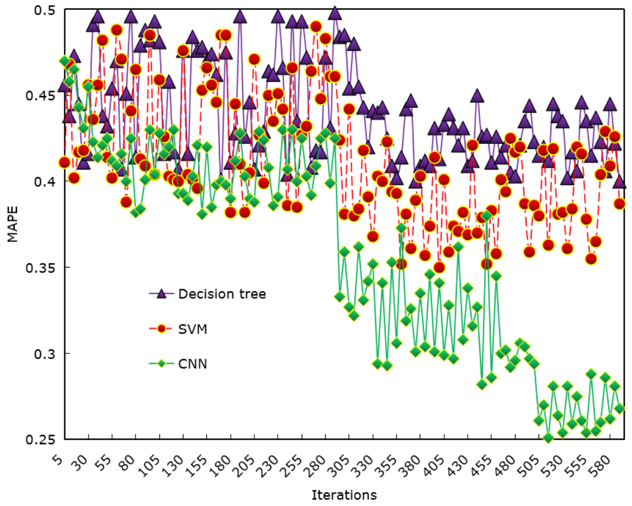

Figure 3: Error of different algorithms.

In summary, the advantages of the CNN based animation scene object recognition algorithm proposed in this paper have been further verified through comparative experiments with SVM and decision tree algorithms. The CNN algorithm can achieve better recognition performance by automatically learning and extracting advanced features from images. However, traditional machine learning algorithms are limited by the complexity of manually designed features and algorithms when processing image recognition tasks, resulting in relatively poor performance.

4 CAD MODELING OF ANIMATION SCENE OBJECTS BASED ON RECOGNITION

4.1 Process Design from Identification to CAD Modeling

CAD modeling of animation scene objects based on recognition is a comprehensive process that combines computer vision technology and CAD modeling methods, aiming to automatically convert the identified animation scene objects into accurate CAD models. The entire process from recognition to CAD modeling is a coherent and closely related sequence. Firstly, in the object recognition stage of animation scenes, the DL algorithm is used to analyze the images or video frames of the animation scene, accurately identify and locate the objects, and provide basic data for subsequent steps. Next, in the stage of object feature extraction and representation, a deeper analysis is conducted on the identified objects, extracting key features such as shape, texture, and color, and representing them with appropriate data structures. These features will serve as important basis for CAD modeling. Subsequently, during the CAD modeling algorithm selection stage, the most suitable one or more CAD modeling algorithms will be selected based on the characteristics and requirements of the object. This step is crucial for ensuring the accuracy and efficiency of modeling. This article assumes that all input is point cloud data. When the input is triangular mesh data, a new point cloud data can be formed by extracting only its triangular vertex information. First of all, by calculating the covariance matrix formed by the nearest k neighbor points p_i around the target point and itself, the specific definition is as follows:

$$M = \frac{1}{k}\sum_{i=0}^{k} (p_i - p')(p_i - p')^T \tag{3}$$

$$p' = \frac{1}{k}\sum_{i=0}^{k} p_i \tag{4}$$

Where p' is the center of mass of these points. Then three eigenvalues and eigenvectors can be obtained by eigenvalue decomposition. For the feature point p in the local support surface, the covariance matrix between it and other vertices p_i in the local support surface is calculated. Defined as follows:

$$M = \frac{1}{Z}\sum_{i:d_i \le r} (r - d_i)(p_i - p)(p_i - p)^T \tag{5}$$

$$Z = \sum_{i:d_i \le r} (r - d_i) \tag{6}$$

$$d_i = \| p_i - p \| \tag{7}$$

Where Z is a normalized term; d_i represents the distance from the vertex p to p_i. Here, Euclidean distance is used for the wider application of this method. Let V_1, V_2 and V_3 be three adjacent points in a patch, then the normal vector can be expressed by their cross product as follows:

$$m = (V_1 - V_2) \times (V_3 - V_2) \tag{8}$$

You can also normalize the obtained vector to a unit length. However, if the two vectors $V_1 - V_2$ and $V_3 - V_2$ are nearly parallel, the result of cross product will be very small, which will bring the problem of numerical instability. This paper deduces a robust method to solve the above problems. The components m_x, m_y, m_z of the vector m are calculated by the following equation:

$$\begin{cases} m_x = \sum_{i=0}^{N-1} (y_i - y_{next\ i})(z_i - z_{next\ i}) \\ m_y = \sum_{i=0}^{N-1} (z_i - z_{next\ i})(x_i - x_{next\ i}) \\ m_z = \sum_{i=0}^{N-1} (x_i - x_{next\ i})(y_i - y_{next\ i}) \end{cases} \tag{9}$$

Where N is the total number of vertices in a patch; x_i, y_i, z_i is the position coordinate of the $N-1$ vertex. Modular operation ensures that the last vertex of the $N-1$ vertex is 0. In the process of calculating each component of the vector, this method only needs to use the multiplication operation once for each edge, and does not need to do collinear test.

In this paper, it is defined that the positive side of the coordinate axis direction of the local reference system must contain more vertices in the supporting surface. Take the x axis as an example:

$$S_x^+ = \quad p_i - p \ \cdot x^+ \wedge d_i \leq r \tag{10}$$

$$S_x^- = \quad p_i - p \ \cdot x^- \wedge d_i \leq r \tag{11}$$

$$x = \begin{cases} x^+ & S_x^+ \geq S_x^- \\ x^- & otherwise \end{cases} \tag{12}$$

Where S_x^+ represents the number of vertices on the positive side of the current x axis in the local support surface; S_x^- represents the number of vertices on the negative side. The direction of z axis can be easily determined by the same method. Finally, the new y axis can be obtained by multiplying the z axis by the x axis.

In this article, the CAD model generation and optimization stage utilizes algorithms to generate the initial CAD model, and optimizes it according to specific needs, which may involve adjusting the details of the model, optimizing performance, and compressing file size to ensure that the final generated CAD model is both accurate and practical. Each step of the entire process is carefully designed to ensure the accuracy and smoothness of data during the conversion process from recognition results to CAD models, providing users with an efficient and reliable CAD modeling solution.

4.2 Parametric Modeling and Nonparametric Modeling of Animation Scene Objects

Parametric modeling is a method of generating or modifying a model by adjusting parameters. For objects in animation scenes, if their geometric shape and topological structure can be determined, parametric modeling can be used to create corresponding CAD models. For example, some common geometric bodies (such as boxes, cylinders, spheres, etc.) can be parameterized for modeling by setting parameters such as length, width, and height. In addition, for more complex objects, advanced parametric modeling techniques such as feature based modeling and constraint-based modeling can be used.

For objects that cannot be modeled through parameterization, non-parametric CAD modeling methods can be used. These methods are typically based on point cloud data, triangular meshes, or voxel representations to construct CAD models of objects. For example, point cloud data obtained from animation scenes can be used to generate surface mesh models of objects through surface reconstruction algorithms. Alternatively, voxel representations can be used to approximate the 3D shape of the object, and the model can be refined through voxel operations such as merging and cutting. Non-parametric modeling methods have flexibility and can meet the modeling needs of various complex shapes.

4.3 Experiment and Result Analysis

A series of experiments were conducted to verify the effectiveness and practicality of the recognition-based animation scene object CAD modeling method. Different animation scenes and objects were used in the experiment, and the results of manual and automatic modeling were compared. Multiple different animation scenes and objects were selected as experimental data to ensure the diversity and representativeness of the data. These animation scenes and objects cover various shapes, sizes, and complexities to verify the generalization ability of the method. Manual modeling was first conducted as a comparative experiment. Experienced modelers use professional

CAD software to manually model objects in animation scenes. This process requires the modeler to carefully observe the objects in the animation scene and use CAD tools to accurately model them. Moreover, the recognition-based animation scene object CAD modeling method proposed in this article is used for automatic modeling. Firstly, use the object recognition algorithm in this article to recognize objects from animation scenes. Then, based on the recognition results, automatically extract the features of the object and generate the corresponding CAD model. The output results of manual and automatic modeling are shown in Figure 4.

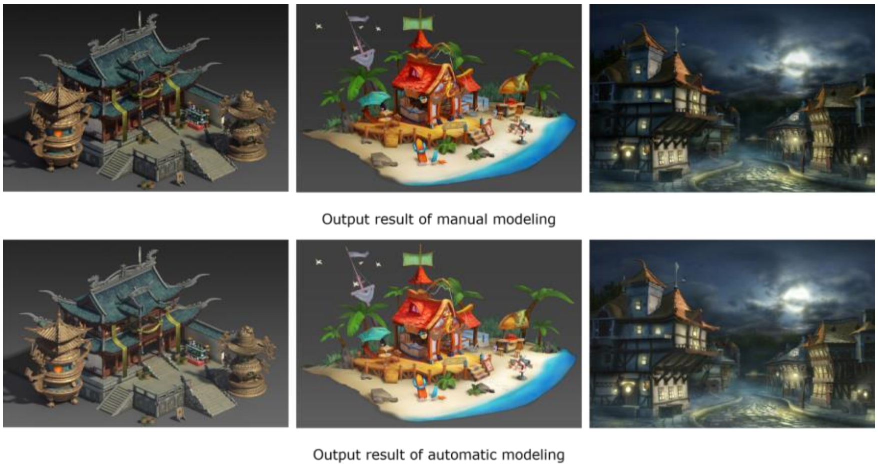

Figure 4: Output results of manual and automatic modeling.

Through experimental data and analysis, the performance of this method in modeling accuracy, modeling time, and user satisfaction can be assessed, as shown in Figure 5(A).

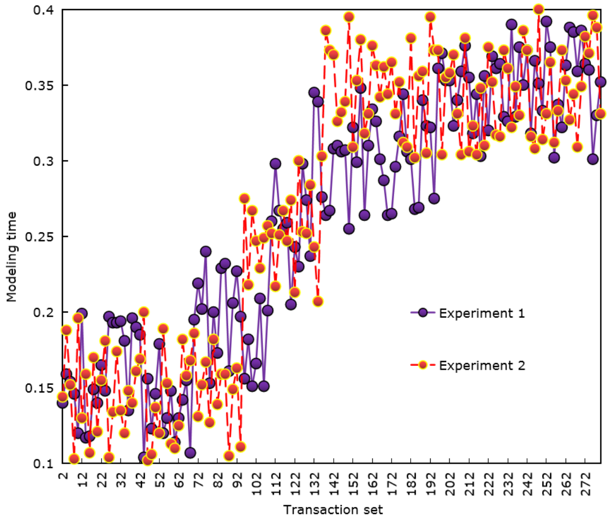

Figure 5(A): Modeling time.

From Figure 5(B), it can be seen that the results of automatic modeling and manual modeling have almost the same modeling accuracy. This indicates that the recognition-based animation scene object CAD modeling method proposed in this article can achieve accuracy similar to manual modeling. Meanwhile, compared to manual modeling, the time required for automatic modeling is greatly reduced. This is because the automatic modeling method eliminates the tedious manual operations in manual modeling and improves modeling efficiency through automated steps such as object recognition and feature extraction.

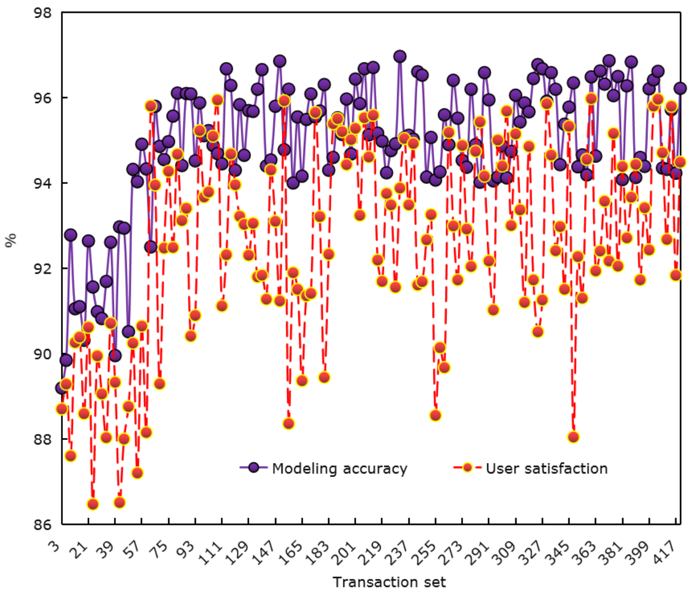

Figure 5(B): Modeling accuracy and user satisfaction results.

Through user satisfaction assessment, it can be found that users have a high level of satisfaction with the results of automatic modeling. Although automatic modeling may differ slightly from manual modeling in some details, overall users believe that the results of automatic modeling are acceptable and have practicality and application value.

This section verifies the effectiveness and practicality of the recognition-based animation scene object CAD modeling method proposed in this paper through comparative experiments with manual modeling. This method significantly reduces modeling time while maintaining high modeling accuracy, and has gained user recognition. This provides strong support for further promoting the automation and intelligence of animation scene object CAD modeling.

5 RESEARCH ON SYSTEM INTEGRATION AND SIMULATION

5.1 Integrated Realization of Object Recognition and CAD Modeling in Animation Scene

In order to integrate object recognition and CAD modeling in animation scenes, a system integration framework needs to be designed. The framework should include the following key components: (1) Animation scene object recognition module, responsible for identifying objects from animation scenes; (2) Feature extraction and representation module, used to extract the features of objects and generate corresponding representations; (3) CAD modeling module, generating CAD models based on object features and representations; (4) Integrated control module, responsible for coordinating data flow and control flow between various modules. Through the design of the framework, automated transformation from animation scenes to CAD models can be achieved.

Under the system integration framework, the integration of object recognition and CAD modeling in animation scenes involves multiple technical aspects. Firstly, it is needed to select appropriate computer vision algorithms and CAD modeling methods, and package and call them. Secondly, it is needed to establish a mapping relationship between object recognition results and CAD modeling methods to ensure that the identified objects can be correctly converted into CAD models. Finally, achieve data transfer and control logic between various modules to complete the workflow of the entire integrated system. By implementing this integration, the efficiency and accuracy of object modeling can be greatly improved.

5.2 System Simulation Environment and Performance Assessment

In order to verify the performance and effectiveness of the integrated system, it is needed to construct corresponding simulation environments and methods. Firstly, collect datasets of various animation scenes, preprocess and annotate them to provide reliable testing and validation data. Secondly, use simulation software or tools to simulate the object recognition and CAD modeling process of animation scenes, and generate corresponding simulation results. In order to assess the performance of the system, this article sets some assessment indicators: calculating recognition accuracy to assess the performance of object recognition, using CAD modeling accuracy to measure the gap between the generated CAD model and real objects, and recording runtime to assess the efficiency of the system. By comprehensively evaluating the simulation results, the feasibility and practicality of the integrated system can be determined, as well as whether there is room for improvement and potential application value.

In the experiment, multiple different animation scenes were selected and the objects to be recognized were annotated as test datasets. Meanwhile, in order to measure the accuracy of CAD modeling, this article obtained real CAD models of these objects as a reference. For object recognition, use the previously described CNN algorithm for recognition. For CAD modeling, the proposed recognition-based animation scene object CAD modeling method is used for automatic modeling. In runtime assessment, record the total time required from the input animation scene to the final generation of the CAD model. For each animation scene, first perform object recognition and record the recognition accuracy. Next, perform automatic CAD modeling on the identified objects. After the modeling is completed, calculate the accuracy difference between the generated CAD model and the actual CAD model, and record the running time. The recognition accuracy, CAD modeling accuracy, and runtime of the system are shown in Figure 6, Figure 7, and Figure 8, respectively.

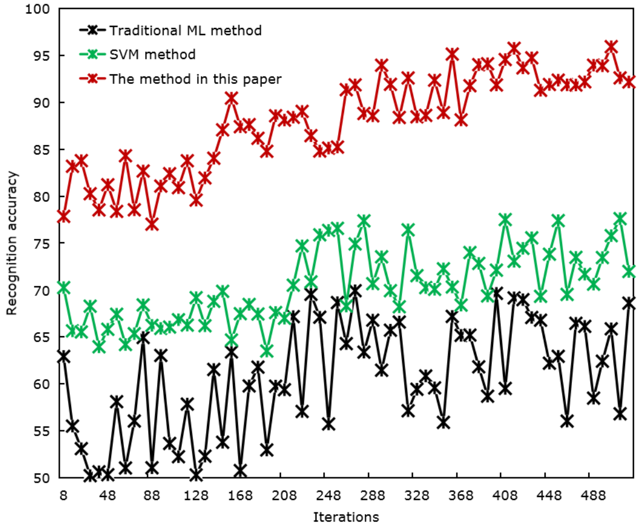

Figure 6: Recognition accuracy of the system.

From Figure 6, it can be seen that the recognition accuracy in most scenarios is relatively high, indicating that the system has good object recognition performance. However, in some complex or blurry scenes, the recognition accuracy has decreased due to factors such as lighting and occlusion. Figure 7 reflects the gap between the generated CAD model and the real object.

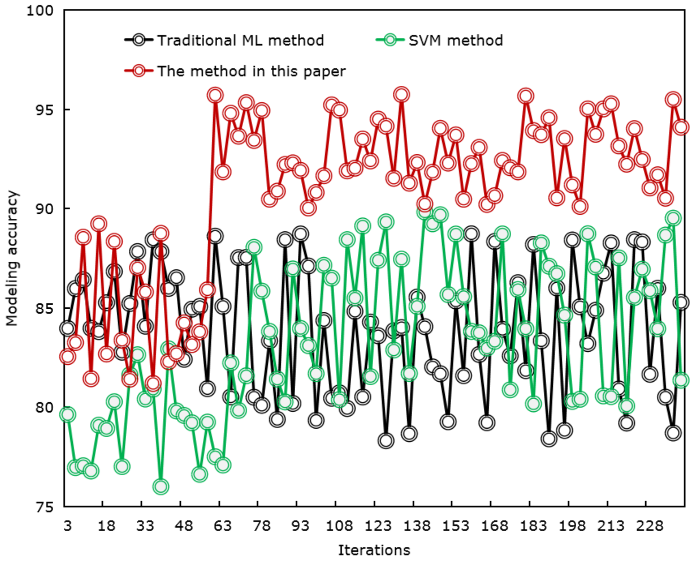

Figure 7: CAD modeling accuracy.

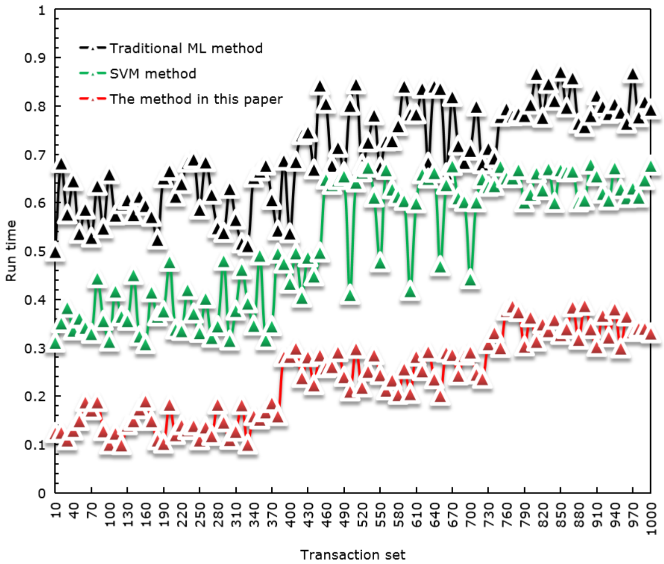

Figure 8: Running time of the system.

In most cases, the CAD modeling accuracy of this system is relatively high. However, in some objects with rich details or complex structures, the modeling accuracy may be slightly lower, which is caused by the error transmission from the recognition stage to the modeling stage. Figure 8 shows the running time of the system. As the dataset increases, the runtime also increases. But overall, the system's running time is within an acceptable range, proving that the system has a certain level of efficiency.

By conducting simulation experiments on integrated systems, a series of simulation results can be obtained. These results include recognized object boundary information, generated CAD models, etc. As shown in Figure 9.

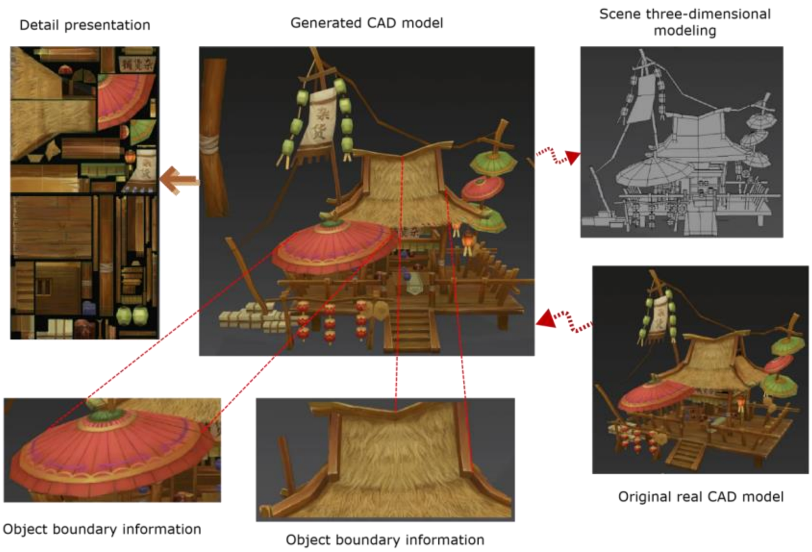

Figure 9: Results of a system simulation experiment.

Object boundary information: Through the recognition module, the system can accurately identify objects in the animation scene and obtain their boundary information. The accuracy of boundary information is crucial for subsequent CAD modeling, as it directly affects the geometric shape and detail representation of CAD models. From the results, it can be seen that the system can effectively extract boundary information of objects, providing a reliable foundation for subsequent modeling. Generated CAD model: Based on the identified boundary information, the system's CAD modeling module can automatically generate the corresponding CAD model. By observing and comparing the generated CAD model with the real CAD model of the original object, it can be seen that the generated model can restore the original object in terms of shape, structure, and details. This further verifies the effectiveness and accuracy of the system.

Through simulation experiments on the integrated system, this article successfully obtained the recognized object boundary information and generated CAD models. The results demonstrate the performance and capability of the system in object recognition and CAD modeling, providing valuable references for subsequent applications and research.

6 CONCLUSIONS

This article applies computer vision technology to the integration research of object recognition and CAD modeling in animation scenes, and combines DL technology for improvement and optimization. Moreover, systematic simulation research and performance assessment are conducted, providing new ideas and methods for research and practice in related fields. The article verifies the good performance of the CNN based animation scene object recognition algorithm through reasonable experimental settings and detailed assessment. All key indicators have reached a level of over 89%, and the algorithm has a low detection rate for objects in animation scenes. The results demonstrate the effective recognition ability of the algorithm for various objects in animation scenes, demonstrating its effectiveness and application potential in object recognition in animation scenes. Through a comprehensive assessment of recognition accuracy, CAD modeling accuracy, and runtime, it can be concluded that the system performs well in most scenarios, with high recognition accuracy and CAD modeling accuracy. Moreover, the runtime is also within an acceptable range and has been recognized by users. This verifies the feasibility and practicality of the integrated system.

Through this study, accurate recognition of objects in animated scenes and automatic CAD modeling can be achieved. This will greatly improve the efficiency of animation and game production, reduce manual participation and costs. In terms of application prospects, this method can be widely applied in the digital entertainment industry, such as animation film production, game development, and other fields. Moreover, it can also be extended to other fields, such as virtual reality and augmented reality, to provide technical support for creating more realistic and vivid 3D scenes. Future work can further optimize algorithms and improve system performance to expand application scenarios and explore more potential application value.

7 ACKNOWLEDGEMENT

This work was supported by 2019 Chongqing Art Science Research and Planning Project "Research on Animation Translation Strategy of Han Terra-Cotta Statue Art," 19YB02.

Zhengzhong Shen, https://orcid.org/0009-0008-0845-0387
Wei Zhang, https://orcid.org/0009-0006-2043-2363

REFERENCES

[1] Baimukashev, D.; Zhilisbayev, A.; Kuzdeuov, A.; Oleinikov, A.; Fadeyev, D.; Makhataeva, Z.; Varol, H.-A.: Deep learning-based object recognition using physically-realistic synthetic depth scenes, Machine Learning and Knowledge Extraction, 1(3), 2019, 883-903. https://doi.org/10.3390/make1030051

[2] Barreto, J.-C.-D.-L.; Cardoso, A.; Lamounier, J.-E.-A.; Silva, P.-C.; Silva, A.-C.: Designing virtual reality environments through an authoring system based on CAD floor plans: A methodology and case study applied to electric power substations for supervision, Energies, 14(21), 2021, 7435. https://doi.org/10.3390/en14217435

[3] 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

[4] Guo, Q.; Ma, G.: Exploration of human-computer interaction system for product design in virtual reality environment based on computer-aided technology, Computer-Aided Design & Applications, 19(S5), 2022, 87-98. https://doi.org/10.14733/cadaps.2022.S5.87-98

[5] Hercog, D.; Bencak, P.; Vincetič, U.; Lerher, T.: Product assembly assistance system based on pick-to-light and computer vision technology, Sensors, 22(24), 2022, 9769. https://doi.org/10.3390/s22249769

[6] Jing, Y.; Song, Y.: Application of 3D reality technology combined with CAD in animation modeling design, Computer-Aided Design and Applications, 18(S3), 2020, 164-175. https://doi.org/10.14733/cadaps.2021.S3.164-175

[7] Kapusi, T.-P.; Erdei, T.-I.; Husi, G.; Hajdu, A.: Application of deep learning in the deployment of an industrial scara machine for real-time object detection, Robotics, 11(4), 2022, 69. https://doi.org/10.3390/robotics11040069

[8] Li, L.: Application of cubic b-spline curve in computer-aided animation design, Computer-Aided Design and Applications, 18(S1), 2020, 43-52. https://doi.org/10.14733/cadaps.2021.S1.43-52

[9] Li, L.; Li, T.: Animation of virtual medical system under the background of virtual reality technology, Computational Intelligence, 38(1), 2022, 88-105. https://doi.org/10.1111/coin.12446

[10] Li, X.; Cao, R.; Feng, Y.; Chen, K.; Yang, B.; Fu, C.-W.; Heng, P.-A.: A sim-to-real object recognition and localization framework for industrial robotic bin picking, IEEE Robotics and Automation Letters, 7(2), 2022, 3961-3968. http://dx.doi.org/10.1109/LRA.2022.3149026

[11] Lin, H.-Y.; Liang, S.-C.; Chen, Y.-K.: Robotic grasping with multi-view image acquisition and model-based pose estimation, IEEE Sensors Journal, 21(10), 2020, 11870-11878. https://doi.org/10.1109/JSEN.2020.3030791

[12] Liu, L.; Liu, G.: Intelligent teaching method of interdisciplinary art design and CAD, Computer-Aided Design and Applications, 19(S8), 2022, 96-104. https://doi.org/10.14733/cadaps.2022.S8.96-104

[13] Nguyen, T.; Miller, I.-D.; Cohen, A.; Thakur, D.; Guru, A.; Prasad, S.; Kumar, V.: PennSyn2Real: Training object recognition models without human labeling, IEEE Robotics and Automation Letters, 6(3), 2021, 5032-5039. https://doi.org/10.1109/LRA.2021.3070249

[14] Noga, M.; Juhás, M.; Gulan, M.: Hybrid virtual commissioning of a robotic manipulator with machine vision using a single controller, Sensors, 22(4), 2022, 1621. https://doi.org/10.3390/s22041621

[15] Qiu, L.: Computer program simulation design of marine 3D animation, Journal of Coastal Research, 112(SI), 2020, 425-428. https://doi.org/10.2112/JCR-SI112-112.1

[16] Rahatabad, F.-N.; Mortazavi, S.-K.: Integrated poser+ MATLAB environment to enhance virtual reality toolbox capabilities for bio-system 3D animations, Frontiers in Biomedical Technologies, 9(2), 2022, 97-101. https://doi.org/10.18502/fbt.v9i2.8848

[17] Sami, U.-R.-M.; Shafiq, M.-T.; Ullah, F.: Automated computer vision-based construction progress monitoring: a systematic review, Buildings, 12(7), 2022, 1037. https://doi.org/10.3390/buildings12071037

[18] Torres, C.-W.; Roberts, D.; Golparvar, F.-M.: Synthesizing pose sequences from 3D assets for vision-based activity analysis, Journal of Computing in Civil Engineering, 35(1), 2021, 04020052. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000937

[19] Zhang, S.; Chen, F.: The effects of computer-aided animation technology in the teaching of hematological medicine, Computer-Aided Design and Applications, 18(S3), 2020, 58-69. https://doi.org/10.14733/cadaps.2021.S3.58-69

[20] Zhang, W.; Fu, X.; Li, W.: Point cloud computing algorithm on object surface based on virtual reality technology, Computational Intelligence, 38(1), 2022, 106-120. https://doi.org/10.1111/coin.12449