Computer Vision Driven Film and TV Scene Modeling and Innovation

Jin He[1] and Haixia Zheng[2]

[1]Zhejiang Gongshang University Hangzhou College of Commerce, Hangzhou 310000, China, 1180268@zjhzcc.edu.cn
[2]Department of Art and Design, Shaanxi Fashion Engineering University, Xi'an 712046, China, 13992062402@163.com

Corresponding author: Jin He, 1180268@zjhzcc.edu.cn

Abstract. The expanding capabilities of computer vision technology now enable the extraction, analysis, and comprehension of high-level information from images and videos. This advancement opens up vast possibilities for innovation and technical enhancements in film and TV production. In the realm of film and TV scene modeling, computer vision algorithms offer significant improvements in accuracy and efficiency. Furthermore, these algorithms, coupled with data mining (DM) techniques, introduce refreshing artistic and visual components to scene design. This paper presents a novel approach that integrates Genetic Algorithms (GA) to optimize Convolutional Neural Networks (CNN), enhancing the effectiveness of computer vision algorithms in CAD modeling for film and TV scenes. The combination of GA and CNN leverages GA's global search capability to refine the network structure and parameter configuration of CNN, while GA's parallel computing power expedites the training of CNN models. The results show that GA-CNN can learn and adapt to different data distribution and noise conditions through GA optimization and has strong robustness and generalization ability. This means that GA-CNN can maintain high detection accuracy even in the face of scenes that have not appeared in the training set.

Keywords: Computer Vision; Film and TV Scenes; CAD Modeling; Data Mining
DOI: https://doi.org/10.14733/cadaps.2024.S19.82-96

1 INTRODUCTION

Based on the rapid growth of digital technology, the use of computer vision algorithms in the film, television, and digital media industry is gradually deepening, significantly improving film and TV production quality. The swift advancement of computer vision technology now enables individuals to extract, analyze, and comprehend sophisticated information embedded in images and videos. This technological progress has expanded the creative horizons and bolstered technical capabilities within the film and TV production industry. 3D thermal imaging technology is a method of non-destructive testing of objects using infrared thermal radiation. This technology measures the infrared radiation temperature of different parts of the building surface, combined with the internal structure and material characteristics of the object, to infer the internal structure and damage status of the building. Due to its non-contact and non-destructive characteristics, 3D thermal imaging technology has broad application prospects in the field of architectural heritage protection. Adamopoulos et al. [1] provided a detailed introduction to the application of computer vision algorithms in three-dimensional thermal imaging for rapid, non-destructive testing of architectural heritage. By analyzing the distribution and changes of thermal radiation data, computer vision algorithms can automatically detect the damaged parts of building structures and accurately locate them. This greatly improves the efficiency and accuracy of detection. By combining the material properties and structural characteristics of buildings, computer vision algorithms can evaluate the health status of building structures and predict potential damage development trends. Through computer vision algorithms, thermal imaging data can be transformed into three-dimensional models and visualized for presentation. This helps to have a more intuitive understanding of the internal condition and damage distribution of building structures. Especially in film and TV scene modeling, computer vision algorithms can not only improve the modeling accuracy but also provide innovative artistic and visual elements for scene design through DM technology. In the fields of 3D animation and virtual reality, the reconstruction of building facades is an important task. In order to achieve realistic animation effects, we need to accurately reconstruct the texture, color, shape, and other details of the building facade. In recent years, the development of deep learning technology has provided new solutions for this task. Bacharidis et al. [2] introduced how to use deep learning techniques for 3D animated building facade reconstruction. Using deep learning algorithms, extract features of building facades from images captured from multiple angles and perform matching to achieve multi-view reconstruction. Deep learning-based model reconstruction methods, such as autoencoders, can be used to recover three-dimensional structures from two-dimensional images. Deep learning can be used to automatically map textures onto reconstructed 3D models and perform detail restoration to improve model realism. It utilizes deep learning techniques to process these images and extract features and structural information of building facades. Next, by training deep learning models, reconstruction from 2D images to 3D models is achieved. Finally, the texture is automatically mapped onto the reconstructed 3D model and rendered to generate realistic 3D animation effects. In the past few decades, CAD modeling has become an indispensable part of film and TV production, which can help designers transform abstract imagination into concrete 3D models and provide a solid foundation for the production of movies, television, and cartoons.

With the rapid development of digital media, the demand and technology for film and animation production are constantly being updated and improved. Maya, as a widely used 3D animation software, provides a powerful platform for the development of visual-driven animation scene control technology in digital media. Cui and Sharma [3] delve into Maya-based digital media visual-driven animation scene control technology. Based on the extracted visual information and the creator's intention, use Maya for 3D scene modeling. This process needs to consider factors such as the layout of the scene, perspective relationships, lighting, etc. In Maya, realistic animation effects are created for characters through bone binding and skinning techniques. At the same time, according to the needs of the plot, set appropriate animations for the props. I use Maya's particle system, fluid dynamics, and other modules to add stunning special effects to film and television scenes, enhancing visual impact. Maya's rendering engine finely renders the completed scene and animation. Finally, use synthesis tools to integrate different elements and complete the final film and television work. However, traditional CAD modeling methods often rely on the manual operation and professional skills of designers. This is not only time-consuming and labor-intensive but also has some limitations in the fineness of the model. Computer vision algorithms have been widely applied in various fields, especially in industrial design, playing an important role. Through computer vision algorithms, designers can quickly and accurately construct three-dimensional models of products, enabling better design, evaluation, and optimization. Hu [4] provides a detailed introduction to the application of computer vision algorithms in three-dimensional modeling in industrial design. Based on two-dimensional image or video data, computer vision algorithms can automatically recognize and extract the features of objects, thereby establishing three-dimensional models. Common 3D reconstruction algorithms include stereo vision, structured light, etc. These algorithms can obtain geometric information about objects from multiple perspectives, improving the accuracy and completeness of the model. In order to make the 3D model more realistic, it is necessary to add appropriate materials and textures. Computer vision algorithms can automatically map textures to the surface of the model according to the needs of designers, achieving more realistic effects. Therefore, how to improve the automation and intelligence level of CAD modeling with the help of advanced computer vision algorithms is an urgent problem in the field of film and TV production.

In the virtual reality environment of film and television, the visual presentation quality of 3D models directly affects the perception and learning effectiveness of learners. High-quality models and textures can provide a more realistic environment, enabling learners to better understand and operate virtual objects. In addition, parameters such as field of view angle and resolution can also affect the visual experience and cognitive load of learners. Interaction design is a core element in virtual reality environments, especially crucial for learning 3D models. Effective interaction design can enable learners to better grasp knowledge through interaction with virtual environments. For example, simulating real physical interactions, such as collisions, gravity, etc., can help learners better understand the underlying working principles of 3D models. In the virtual reality environment of film and television, learners should be allowed to learn and explore independently based on their own learning needs and progress. By providing personalized learning paths and feedback, learners can improve their learning motivation and effectiveness. Meanwhile, this autonomous and personalized learning approach also helps cultivate learners' innovative abilities [5]. The combination of 3D reality technology (VR) and CAD (computer-aided design) is playing an increasingly important role in the production of film and television animation scenes. This combination not only improves the modeling efficiency of animation scenes but also provides new possibilities for data mining. Jing and Song delved into the impact and application of this technology in the production of film and television animation scenes. In the production of film and animation scenes, this combination allows designers to see the final effect of the scene more intuitively so that problems can be discovered and corrected in the early stages. In addition, through VR devices, directors and producers can conduct a comprehensive review of the scene before shooting, providing more accurate data support for decision-making. Through data mining, designers can understand the audience's preferences for characters and props, thereby designing more popular styles. Data mining can help designers understand the audience's preferences for scene layout and color in order to optimize scene design [6]. The powerful feature detection ability and efficient processing speed of CNN make it possible for us to apply CNN to the automation stage of CAD modeling of film and TV scenes. When it comes to processing intricate video scene data, the standard CNN model might encounter certain obstacles related to the model's generalization capabilities, training convergence rates, and handling of large-scale information. To enhance the computer vision algorithm's performance in CAD modeling for film and TV scenarios, this article introduces a novel approach based on GA-CNN. GA, an optimization algorithm, mimics the principles of natural selection and genetics. Through simulating biological evolution processes like reproduction, crossover, and mutation, it aims to locate near-optimal solutions within complex problem spaces. By integrating GA with CNN, we can harness GA's global search capability to refine the network structure and parameter configuration of CNN. Additionally, the parallel computing power of GA accelerates the training process of the CNN model. The primary contributions and innovations of this article are outlined below:

(1) In this article, GA and CNN are combined and applied to the field of film and TV scene CAD modeling, which makes full use of their advantages in global optimization and local fine processing.

(2) Aiming at the particularity of video scene data, this study designed the optimization strategy of the CNN model based on GA, including automatic design of network structure, intelligent adjustment of superparameters, and dynamic monitoring of the training process.

(3) Through DM technology, this research deeply digs into the artistic elements and visual features in film and TV scenes and provides designers with rich creative inspiration and design materials.

In the following chapters, this article will introduce the design principle and implementation details of the GA-CNN model in detail and verify its effectiveness in CAD modeling of film and TV scenes through experiments. By comparing it with traditional CAD modeling methods, we will show the advantages of the GA-CNN model in modeling accuracy, efficiency, and innovation. In addition, this article will also discuss the possible challenges and future development direction of the proposed method in practical application.

Finally, the goal of this study is to build a highly integrated and intelligent CAD modeling system for film and TV scenes. The system can automatically process and analyze video scene data, generate high-quality 3D models, and provide rich creative support for designers.

2 THEORETICAL BASIS

Immersive film and television VR has become a new way of entertainment and information dissemination. Among them, 3D user interface design plays a crucial role, directly affecting the satisfaction and immersion of the user experience. This article will explore how to use computer vision algorithm-driven 3D user interface design to improve the usability of immersive film and television VR. Computer vision algorithms play a crucial role in 3D user interface design. These algorithms can help identify user actions and gestures and convert them into interactive instructions in a virtual environment. In immersive film and television VR, the user's attention is mainly focused on the content, so the interface design should be as simple as possible to avoid interfering with the user's immersive experience. Meanwhile, the design should be intuitive and easy to understand, making it convenient for users to understand and operate quickly. Through natural and intuitive interaction, personalized customization, real-time feedback, and optimization in adaptability and scalability, Kharoub et al. [7] have improved user satisfaction and immersion. In the field of film and television production, scene construction is crucial. Traditional scene modeling methods often rely on manual modeling, which is time-consuming and difficult to ensure accuracy. With the advancement of technology, CAD film, and television scenes, reverse modeling technology based on multi-sensor measurement has gradually become a new solution. Li et al. [8] provided a detailed introduction to the principles, advantages, and applications of this technology in film and television production. CAD film and television scene reverse modeling based on multi-sensor measurement refers to the use of reverse engineering principles to convert data collected through sensors into a 3D CAD model. For scenes that are difficult to construct through traditional methods, such as large natural landscapes, historical buildings, etc., reverse modeling technology can quickly reconstruct realistic scene models. In special effects production, reverse modeling technology can help create realistic virtual objects or environments, improving the realism of special effects. In game development, reverse modeling technology can be used to create complex game scenes and props, enriching game content. Fully Convolutional Networks (FCN), as an important branch of deep learning, provide new solutions for high-quality texture reconstruction of 3D scenes. Liu et al. [9] provided a detailed introduction to the application of fully convolutional networks driven by computer vision algorithms in high-quality texture 3D scene reconstruction. Unlike traditional convolutional neural networks, fully convolutional networks can accept input of any size and output semantic information of the same size as the input. This characteristic gives fully convolutional networks great advantages in processing tasks such as image segmentation and detail reconstruction. Using fully convolutional networks for pixel-level prediction of depth maps can accurately estimate the depth information of each point in the scene. Through depth information, we can further calculate the surface geometric shape of the object, thereby completing three-dimensional reconstruction. Based on the results of depth estimation, we can use surface reconstruction algorithms such as Poisson surface reconstruction or Marching Cubes to generate 3D surface models from depth data. High-quality texture mapping is an important aspect of 3D scene reconstruction. Full convolutional networks can help us accurately map texture information from the original image to the reconstructed 3D model, thereby achieving realistic visual effects.

Immersive film and television virtual reality (VR) training, as an emerging training method, has demonstrated its unique value and potential in many fields. This training method not only provides a highly simulated environment, giving participants a sense of immersion but also enables a more comprehensive understanding of the training effect through multiple evaluation mechanisms. Makransky et al. [10] explored how immersive film and television VR training based on multiple assessments affects the motivation and cognitive benefits of participants. With the help of computer vision technology and sensors, immersive film and television VR training can achieve real-time interaction between humans and virtual environments. This interactivity allows participants to feel a greater sense of participation and control, further enhancing their training motivation. Immersive film and television VR training can be personalized according to the individual differences and needs of participants. This personalized learning approach can enable participants to participate more autonomously in training and enhance their intrinsic motivation. The 3D modeling and scene production of paper has been a highly focused direction in recent years. Through computer vision algorithms, we can quickly and accurately obtain the geometric information of the paper surface and convert it into a three-dimensional model for further use in scene production. Paczkowski et al. [11] provided a detailed introduction to the application of computer vision algorithms in 3D modeling and scene production of paper. Use high-resolution cameras and appropriate lighting equipment to capture images of the surface of the paper. The quality of collected images directly affects the accuracy of subsequent modeling. By using computer vision algorithms, information such as texture, lines, and text on the surface of paper is extracted, which will be used to establish a 3D model. Common feature extraction algorithms include SIFT, SURF, etc. Based on the extracted features, it utilizes computer vision algorithms for 3D reconstruction to generate a 3D model of the paper surface. Common 3D reconstruction algorithms include stereo vision, structured light, etc.

Pepe et al. [12] introduced an effective pipeline for generating high-quality 3D models from 3D point clouds to meet the needs of HBIM and structural analysis. To make a 3D model look more realistic, it is usually necessary to add textures and details to it. This step can be achieved by interpolating, upsampling, or using deep learning techniques on the original point cloud data. Texture mapping is the process of mapping 2D images (such as photos or scanned images) to the surface of a 3D model so that the model has a similar appearance to the actual object. It uses feature extraction algorithms such as PCA, FAST, SURF, and SIFT to identify and locate key features from point clouds, such as their boundaries, corners, and planes. The extracted features will be used for the construction of a 3D model, which typically involves techniques such as surface reconstruction and model smoothing. The application of visual algorithms in virtual film and television scenes mainly includes scene reconstruction, character animation, special effects production, and other aspects. Through visual algorithms, real-world images and video data can be transformed into three-dimensional models, textures, animations, and other elements in virtual scenes, achieving highly realistic visual effects. Reski and Alissandrakis [13] explored the comparative study of different input technologies in open data exploration of virtual film and television scenes driven by visual algorithms. Open data refers to publicly available and machine-readable datasets. Open data has important value in virtual film and television scenes. Open data can provide rich materials and resources for film and television production, reducing production costs by analyzing and mining open data, allowing new ideas and inspirations to be discovered, and promoting innovation and development in the film and television industry. By capturing human movements and postures and converting them into digital signals, more natural interactions can be achieved. This technology has been widely applied in the fields of virtual reality and gaming, but its application in virtual film and television scenes still needs further exploration.

The direct operation driven by visual algorithms has great potential for modeling film and television scenes in immersive virtual reality. It combines intuitive operation methods and efficient modeling processes, enabling users to create high-quality 3D models quickly. Through direct operation, users can have intuitive interaction in the virtual environment, thereby quickly and accurately creating the desired scene. Visual algorithms play a crucial role in implementing this process. Tastan et al. [14] explored how to use visual algorithm-driven direct operations to achieve efficient modeling of film and television scenes in immersive virtual reality. Direct operation allows users to interact in a virtual environment through simple gestures and actions without the need to learn complex commands or tools. This intuitive operating method greatly reduces learning costs and makes it easy for nonprofessionals to get started. The direct operation can quickly capture the user's intentions and transform them into models in a virtual environment in real time. This allows users to complete a large amount of modeling work in a short period of time, improving work efficiency. The establishment of scene models is crucial in film and television production. In order to construct complex film and television scene models quickly and accurately, a semiautomatic structural film and television scene model generation method based on CAD construction drawings has gradually attracted attention. This method combines CAD construction drawings and semiautomatic modeling technology, providing an efficient and accurate solution for scene model generation in film and television production. CAD construction drawings are detailed drawings used to describe building or scene design. Based on CAD construction drawings, scene modeling can ensure that the scene model is consistent with the actual design and improve the accuracy and reliability of the model. For the film and television production of historical buildings or historical sites, this method can quickly and accurately reconstruct realistic scene models. In order to improve shooting efficiency and control costs, this method can be used to create virtual shooting scenes for photographers to preview and adjust in the early stages. In game development and virtual reality experiences, this method can be used to create high-quality scene models, providing immersive gaming and VR experiences [15].

CAD vision algorithm is a method of using computer vision technology to identify and analyze graphic and symbol information in construction drawings. By training deep learning models, this algorithm can automatically recognize elements such as lines, annotations, and legends in drawings and extract their geometric and semantic information. In BIM modeling, the application of CAD visual algorithms is mainly reflected in structural recognition, element positioning, and parameter extraction. It uses manual and semiautomatic modeling methods to generate structural BIM models for multiple construction drawings. The experimental results show that the semiautomatic modeling method based on the CAD vision algorithm is superior to traditional manual modeling methods in terms of modeling efficiency and accuracy. Meanwhile, this method can significantly reduce manual intervention and error rates and improve the reliability of the model. Before creating a 3D animation scene, designers need to first conduct conceptual design. Through computer-aided graphic design software, sketches and concept drafts can be quickly drawn to better explore and express the designer's creativity and intentions. Zhao and Zhao [16] can use computer-aided graphic design software to create the necessary planar elements for 3D animation scenes directly, such as backgrounds, floors, walls, etc. These elements can be finely adjusted and modified in the software to meet subsequent 3D modeling and animation production needs. Textures and textures are essential elements in 3D animation scenes. Through computer-aided graphic design software, designers can create various textures and apply them to 3D models to increase the realism and delicacy of the scene. Color and lighting are key factors that affect the atmosphere of 3D animation scenes. Through computer-aided graphic design software, color matching and lighting effects in the scene can be pre-designed and adjusted to ensure that the presentation effect in virtual reality meets expectations. Layout and layout are crucial for the visual effects of 3D animation scenes. By using computer-aided graphic design software, various elements in the scene can be reasonably laid out and typeset, resulting in the best visual effect in three-dimensional space.

3 OPTIMIZATION OF CAD MODELING OF FILM AND TV SCENES BASED ON GA-CNN

CNN is a pivotal network structure in DL, particularly for computer vision tasks. It operates by simulating the human visual cortex's processing methods, utilizing a convolution kernel to convolve the input image and extract local features. In contrast to traditional image processing algorithms, CNN excels in automatically learning image feature expressions, sidestepping the intricacies and uncertainties associated with manual feature design.

The fundamental structure of CNN comprises convolution, pooling, and fully connected layers. The convolution layer is tasked with extracting image features, the pooling layer diminishes feature dimensions while bolstering model robustness, and the fully connected layer maps the extracted features to the ultimate output space. By stacking multiple convolution and pooling layers, CNN progressively extracts abstract image features from low to high levels, facilitating a profound comprehension of image content. GA mimics the biological evolution process to find optimal solutions by leveraging natural selection and genetic principles. In GA, the solution to a problem is conceptualized as an individual within a population, with each individual possessing specific genetic information. By employing selection, crossover, and mutation operations, the fittest individual is ultimately identified.

The core operations of GA comprise selection, crossover, and mutation. Selection involves picking out outstanding individuals to proceed to the next generation based on their fitness values. Crossover entails exchanging genetic material between two individuals to generate new offspring. Mutation introduces random changes to an individual's genes, thereby enhancing population diversity. Through repeated iterations of this evolutionary process, GA is capable of approximating optimal solutions within intricate problem spaces.

In this article, GA and CNN are combined to form a new optimization strategy (GA-CNN). This strategy makes use of GA's global search ability and parallel computing ability to optimize the structure and parameters of CNN, aiming to improve CNN's performance in processing complex video scene data.

The optimization process of GA-CNN includes the following steps:

(1) Coding: The network structure and parameters of CNN are coded as individual genes in GA to form an initial population.

(2) Fitness evaluation: using the training data set to evaluate each individual in the population and calculate its fitness value.

(3) According to the fitness value, select excellent individuals to enter the next generation population.

(4) crossover and mutation: crossover and mutation are performed on the selected individuals to generate new offspring populations.

(4) Repeat the operations of fitness evaluation, selection, crossover, and mutation until the termination conditions are met.

Through GA's global search and parallel computing capabilities, GA-CNN can quickly find excellent network structure and parameter configuration in a huge search space, which significantly improves the performance of the CNN model. At the same time, GA-CNN also has the adaptive ability to automatically adjust the network structure and parameters according to different tasks and data sets and realize the real automatic modeling process.

In the CAD modeling of film and TV scenes, DM technology can help us to deeply explore the potential laws and characteristics in the scene data and provide valuable reference information for the modeling process.

Through the analysis and mining of a large number of video scene data, the classification and clustering of scenes can be realized. This helps us to understand the characteristics and laws of different types of scenes and provides targeted guidance and reference for modeling. These features can be used as important reference information in the modeling process, helping us to restore the details and style of the scene more accurately. Film and TV scenes often contain rich creative elements and artistic inspiration. Through DM technology, these creative elements can be excavated, analyzed, and refined, and innovative design ideas can be provided for designers. This information can be fed back to the designer for targeted optimization to improve the quality of the model. In this article, we introduce an optimization method for CAD modeling of film and TV scenes based on the integration of GA and CNN. This approach leverages the global search capabilities of GA and the local fine-tuning abilities of CNN to achieve an automated and intelligent modeling process for film and TV scenes. Initially, the raw video scene data undergoes preprocessing, which includes techniques such as image denoising, color correction, and illumination equalization to ensure the quality and consistency of the input data. In addition, according to the specific task requirements, data can be enhanced, such as rotation, scaling, cropping, etc., to increase the diversity and generalization ability of data. In the GA-CNN model, the structure and parameters of CNN are encoded as individual genes in GA. Through the selection, crossover, and mutation of GA, the network structure and parameter configuration are continuously optimized to find the most adaptive CNN model.

In this study, the network structure and parameters of CNN are coded into a long vector by multi-level coding. In GA, a vector serves as an individual gene that undergoes iterative evolution through various stages such as population initialization, assessment of fitness, selection, crossover, and mutation. During the fitness evaluation phase, each individual is appraised using the training dataset to determine the efficacy of its corresponding CNN model for a particular task. Individuals with high fitness values are chosen to advance to the next generation, where crossover and mutation operations are executed to generate new offspring. The process of film and TV scene image recognition based on GA-CNN is shown in Figure 1.

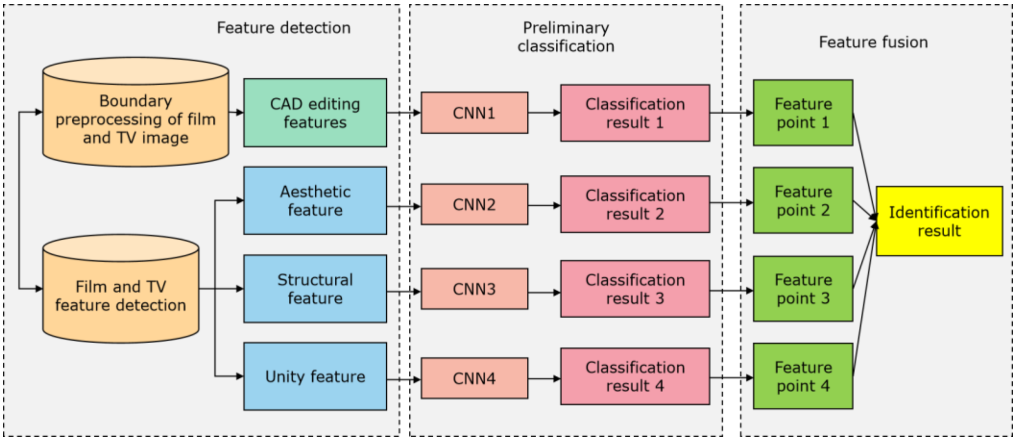

Figure 1: Image recognition process of film and TV scene.

After optimization of GA, an excellent CNN model configuration is obtained. Next, these models are trained with training data sets to learn the feature expression and classification rules of scene data. During the training phase, we employ the backpropagation algorithm and gradient descent method to adjust the network weights and bias terms. This process aims to minimize the loss function and enhance the model's overall performance. A crucial aspect of this research is feature detection in film and TV scenes, which involves extracting representative features that accurately depict the scene's characteristics from the original film and TV scene images. The information contained in film and TV scenes is rich and varied, so it is necessary to choose appropriate features to describe the scenes. In the process of feature detection, the algorithm needs to be optimized and adjusted according to the actual needs. For example, the effect of feature detection can be optimized by adjusting the network structure of CNN, increasing or decreasing convolution layers, and changing activation functions. Through effective feature detection, we can accurately describe the characteristics and styles of film and TV scenes. See Figure 2 for the feature detection process of film and TV scenes.

The connection between the camera and world coordinate systems can be expressed through the rotation matrix R and the translation vector T. The transformation relationship of a certain point P in space between the world coordinate system and the camera coordinate system is as follows:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{1}$$

In this context, $X_w, Y_w, Z_w, 1^T$ and $X_c, Y_c, Z_c, 1^T$ are the homogeneous coordinates of a spatial point P in the global and camera coordinate systems, respectively. Meanwhile, R denotes a 3×3 orthogonal matrix and T signifies a 3D translation vector.

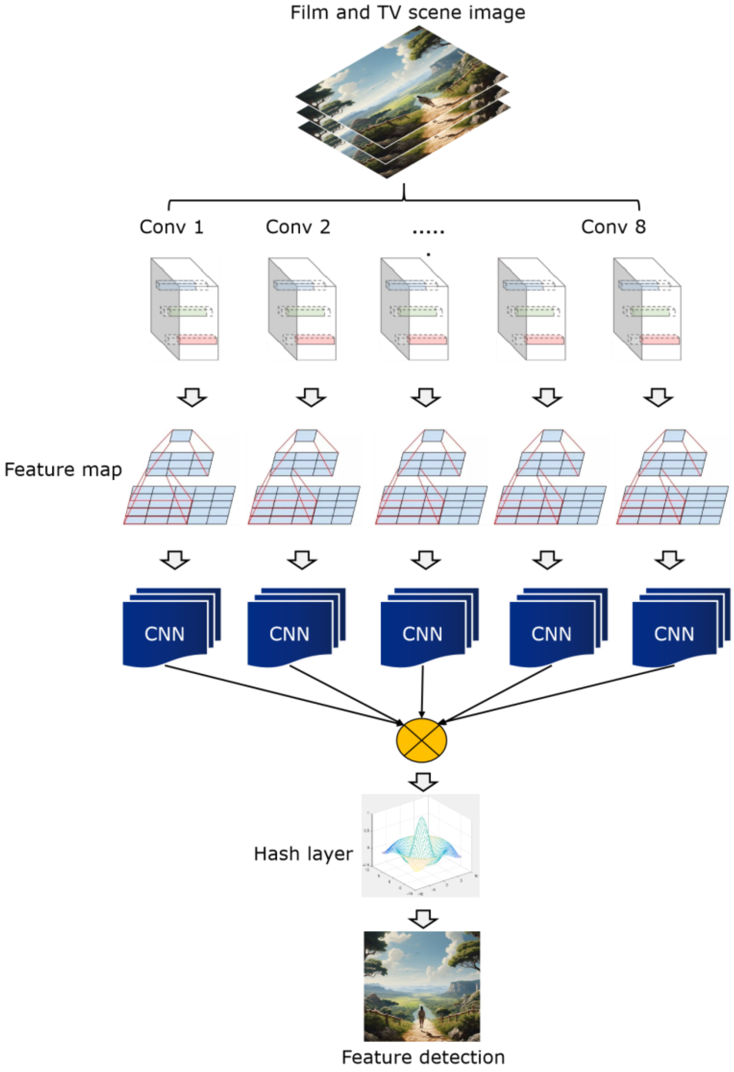

Figure 2: Feature detection process of film and TV scene.

By learning and training a large number of video scene data through the GA-CNN model, the scene can be automatically classified and labeled. This is helpful to quickly identify different types of scenes and provide targeted guidance for subsequent modeling work. Through the powerful feature detection ability of the GA-CNN model, we can accurately capture the detailed information in the scene and restore it to the CAD model. In addition, by using DM technology to analyze and mine the scene data deeply, we can also find the potential laws and characteristics in the scene and provide more valuable reference information for modeling. For chromosomes k possessing fitness f_k, the selection probability s_k is determined as follows:

$$s_k = \frac{r_k}{\sum_{i=1}^{pop\_size} f_i} \tag{2}$$

Proceed to compute the total fitness values of all chromosomes within the population:

$$F = \sum_{i=1}^{pop\_size} f_i \tag{3}$$

For every chromosome, the selection probability c_k is determined:

$$s_k = \frac{r_k}{F} \tag{4}$$

For every chromosome, compute the cumulative probability t_k :

$$t_k = \sum_{i=1}^{pop\_size} s_i \tag{5}$$

Film and TV scenes contain rich creative elements and artistic inspiration. Through the combination of the GA-CNN model and DM technology, these creative elements are excavated, analyzed, and refined. These creative elements can provide innovative design ideas and material support for designers. Designers can refer to the excavated creative elements such as color matching schemes and texture style to create a unique and artistic film and TV scene. Through the combination of the GA-CNN model and DM technology, the data of the established CAD model can be analyzed and mined. This information can be fed back to designers for targeted optimization. This data-based optimization strategy can align the CAD model with the audience's aesthetic needs and market trends.

By incorporating the notion of a kernel function, nonlinear SVM has the capability to transform the original dataset into a higher-dimensional feature space. This enables the identification of a linear hyperplane within this elevated dimension that facilitates classification. This mapping allows the algorithm to capture the nonlinear relationship in the data so that the data that was originally linearly inseparable in the original space becomes linearly separable in the mapped high-dimensional space. In the process of CAD modeling of film and TV scenes, the complexity and diversity of data is a problem that cannot be ignored. Different scenes, lighting conditions, object materials, and other factors will cause the distribution of data in the feature space to become extremely complicated. Nonlinear SVM can take advantage of the powerful ability of kernel function to deal with this complexity. By selecting a suitable kernel function, nonlinear SVM is able to locate the best classification hyperplane within the high-dimensional feature space, allowing for precise categorization and modeling of intricate video scene data. The principle of the SVM nonlinear classifier is shown in Figure 3.

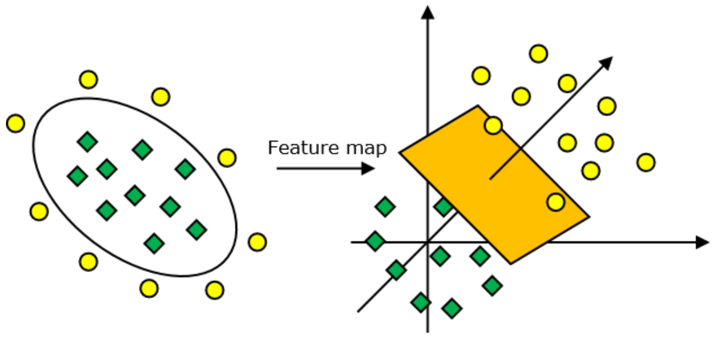

Figure 3: Principle of SVM nonlinear classifier.

The computation of the texture distribution set for the fuzzy regional scene image is outlined below:

$$w\ i,j\ = \frac{1}{Z\ i}\exp\left(-\frac{d\ i,j}{h^2}\right) \tag{6}$$

Which hare the first- and second-order texture distribution operators? The associated parameters are substituted and transformed into:

$$W^{'} = \frac{1}{2} f\ x^{'}, y^{'}, z^{'}\ + E \tag{7}$$

Where $x^{'}, y^{'}, z^{'}$ is the 3D coordinate with a visual constraint applied?

In film and TV production, the diversified styles of scenes are very important for creating atmosphere and conveying emotions. From realism to abstraction, from bright to dark, each style brings a unique visual experience to the audience. Kurtosis is a statistic describing the distribution of data, which can reflect the kurtosis and skewness of data distribution. In this study, a series of coefficients are obtained by using the basis function to respond to the video scene images. Then, the kurtosis distribution of these coefficients is calculated as a feature to distinguish different styles.

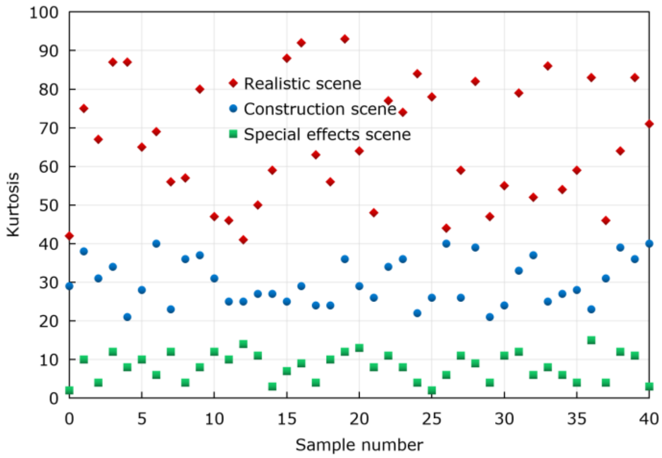

Figure 4: Whistle distribution map.

The results show that different styles of film and TV scenes show obvious differences in the kurtosis distribution of the basis function response coefficient (Figure 4). This difference provides an effective means to distinguish and identify different film and TV scene styles. When the test image with the same style as the static scene animation works is used, the kurtosis value is the largest. This means that kurtosis distribution can not only distinguish different styles but also quantitatively evaluate the similarity between images and specific styles.

In modeling film and TV scene spaces, consider the normal vector of each triangle within the associated triangle group, having vertex v_i, as n_k n_k. Additionally, let's x_k represent the center and a_k represent the area. The plane created by this normal vector and center, as defined below, is termed the average plane of the vertex:

$$N = \frac{\sum n_k a_k}{\sum a_k} \tag{8}$$

$$n = \frac{N}{|N|} \tag{9}$$

$$x = \frac{\sum x_k a_k}{\sum a_k} \tag{10}$$

The displacement from point P in the 3D space to the mesh model TM is delineated as:

$$d\ \ P,TM\ \ = \min\ d\ \ P,X \tag{11}$$

Where $d\ \ P,X$ is the Euclidean distance between point P and point X. In order to render the triangle mesh shadow model on a computer, it is necessary to acquire the normal vector for each vertex of the triangles.

4 RESULT ANALYSIS AND DISCUSSION

Google's open-source DL framework, TensorFlow, is utilized to construct the model and execute the algorithm. TensorFlow is a robust DL platform that offers an extensive range of tools and libraries, facilitating the creation of diverse and intricate DL models. The experimental core involves merging the content image with the style image through algorithmic manipulation to produce a novel image embodying a distinctive style. Initially, the features of both the content and style images are extracted, segregating their respective content and style characteristics. Subsequently, a designated loss function calculates the disparity between the content and style features, serving as the foundation for model training and optimization. Ultimately, the model attains the ability to transform the style of the content image into the desired target style, generating a fresh image imbued with that aesthetic. Figure 5 illustrates the successful integration of the content image and style image achieved by the model. The generated new image not only retains the basic structure and detailed information of the content image but also effectively incorporates the characteristics and elements of the target style.

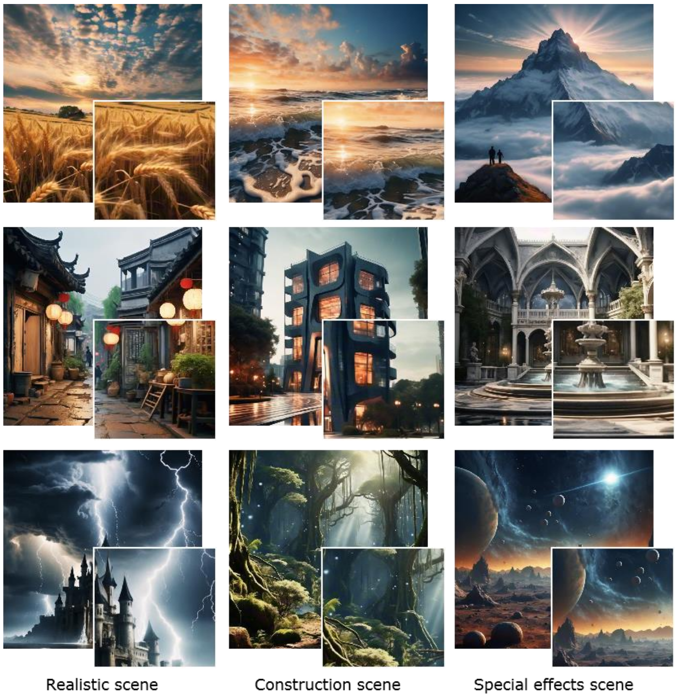

Figure 5: Partial results display.

From the comparison of the feature detection results between the given GA-CNN algorithm (Figure 6) and the traditional CNN algorithm (Figure 7), we can observe the superior performance of GA-CNN algorithm in feature detection. Compared with the traditional CNN algorithm, the output results of GA-CNN have a higher degree of coincidence with the actual results.

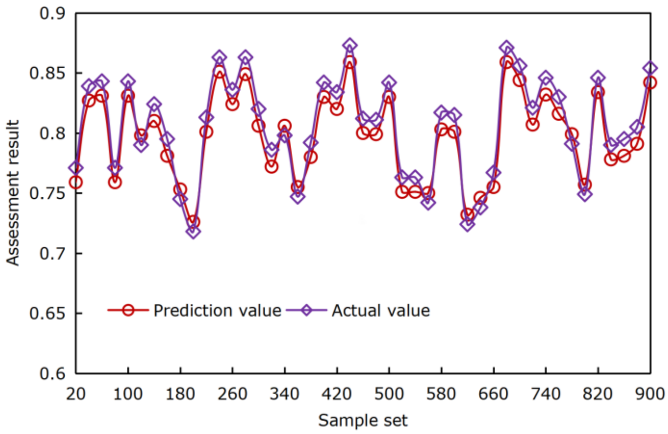

Figure 6: Detection performance of ga-CNN algorithm.

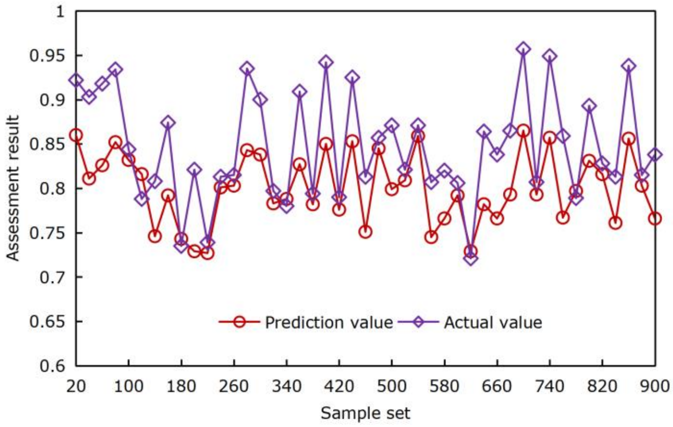

Figure 7: Detection performance of CNN algorithm.

Because GA-CNN combines GA's global search ability with CNN's local fine processing ability, it can extract key features in movie scenes more accurately. When optimizing the network structure and parameters, GA can adaptively adjust to different scene features, thus improving the accuracy of feature detection. Traditional CNN uses fixed network structure and parameters, which may not perform well in dealing with complex and changeable film and TV scenes, resulting in a large error in feature detection. Through the optimization of GA, GA-CNN can learn and adapt to different data distribution and noise conditions and has strong robustness and generalization ability. This means that GA-CNN can maintain high detection accuracy even in the face of scenes that have not appeared in the training set. Although GA-CNN is superior to traditional CNN in feature detection accuracy, its computational complexity and resource consumption may be relatively high due to the need to run GA for optimization. In practical application, it is necessary to choose the appropriate algorithm according to the specific needs and resource constraints.

5 CONCLUSION

As the film and TV industry expands rapidly, the need for sophisticated scene modeling technology becomes increasingly pressing. This paper introduces an enhanced approach to CAD modeling for film and TV scenes, leveraging the GA-CNN algorithm. This method aims to streamline the modeling process by harnessing the global search capabilities of GA and the precise local processing power of CNN. We detail the principles, implementation, and practical applications of the GA-CNN algorithm in film and TV production, highlighting its notable strengths in feature detection, classification, labeling, and optimization. In contrast to traditional CNN methods, GA-CNN offers improved accuracy in key feature extraction, greater adaptability to complex and varied film and TV scenes, and enhanced robustness and generalizability. Our findings demonstrate that GA-CNN achieves high precision and consistency in feature detection for film and TV scenes, thereby boosting the efficiency and quality of the modeling process. In addition, this method can also be applied to scene style discrimination, creative element mining, and other aspects, providing more valuable reference information and creative support for film and TV production.

However, the optimization method of film and TV scene CAD modeling based on GA-CNN still has some limitations. For example, when dealing with extremely complex or noisy scenes, the performance of the algorithm may be affected. At the same time, this method needs to be further optimized in terms of computational efficiency and resource consumption. Future research can further improve the performance and efficiency of the algorithm, explore more application scenarios and expansion possibilities, and pay attention to the interpretability and reliability of the algorithm.

Jin He, https://orcid.org/0000-0002-4097-1085
Haixia Zheng, https://orcid.org/0009-0004-0971-4293

REFERENCES

[1] Adamopoulos, E.; Volinia, M.; Girotto, M.: 3D Thermal mapping from IRT images for rapid architectural heritage NDT, Buildings, 10(10), 2020, 187. https://doi.org/10.3390/buildings10100187
[2] Bacharidis, K.; Sarri, F.; Ragia, L.: 3D building façade reconstruction using deep learning, ISPRS International Journal of Geo-Information, 9(5), 2020, 322. https://doi.org/10.3390/ijgi9050322
[3] Cui, Q.; Sharma, A.: Digital media animation control technology based on Maya, Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering), 14(7), 2021, 735-743. https://doi.org/10.2174/2352096514666210805160749
[4] Hu, L.: Application OF AutoCAD's 3D modeling function in industrial modeling design, Computer-Aided Design and Applications, 18(S1), 2020, 33-42. https://doi.org/10.14733/cadaps.2021.S1.33-42
[5] Huang, H.; Lee, C.-F.: Factors affecting usability of 3D model learning in a virtual reality environment, Interactive Learning Environments, 30(5), 2022, 848-861. https://doi.org/10.1080/10494820.2019.1691605
[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] Kharoub, H.; Lataifeh, M.; Ahmed, N.: 3D user interface design and usability for immersive VR, Applied Sciences, 9(22), 2019, 4861. https://doi.org/10.3390/app9224861
[8] Li, J.; Liu, J.; Chang, W.: CAD reverse modeling based on multi-sensor measurement, Computer-aided Design and Applications, 18(S1), 2020, 144-154. https://doi.org/10.14733/cadaps.2021.S1.144-154

[9] Liu, Z.-N.; Cao, Y.-P.; Kuang, Z.-F.; Kobbelt, L.; Hu, S.-M.: High-quality textured 3D shape reconstruction with cascaded fully convolutional networks, IEEE Transactions on Visualization and Computer Graphics, 27(1), 2019, 83-97. https://doi.org/10.1109/TVCG.2019.2937300

[10] Makransky, G.; Borre, G.-S.; Mayer, R.-E.: Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments, Journal of Computer Assisted Learning, 35(6), 2019, 691-707. https://doi.org/10.1111/jcal.12375

[11] Paczkowski, P.; Dorsey, J.; Rushmeier, H.: Papercraft3d: paper-based 3D modeling and scene fabrication, IEEE Transactions on Visualization & Computer Graphics, 25(4), 2019, 1717-1731. https://doi.org/10.1109/tvcg.2018.2820068

[12] Pepe, M.; Costantino, D.; Restuccia, G.-A.: An efficient pipeline to obtain 3D model for HBIM and structural analysis purposes from 3D point clouds, Applied Sciences, 10(4), 2020, 1235. https://doi.org/10.3390/app10041235

[13] Reski, N.; Alissandrakis, A.: Open data exploration in virtual reality: a comparative study of input technology, Virtual Reality, 24(1), 2020, 1-22. https://doi.org/10.1007/s10055-019-00378-w

[14] Tastan, H.; Tuker, C.; Tong, T.: Using handheld user interface and direct manipulation for architectural modeling in immersive virtual reality: An exploratory study, Computer Applications in Engineering Education, 30(2), 2022, 415-434. https://doi.org/10.1002/cae.22463

[15] Yang, B.; Liu, B.; Zhu, D.; Wang, Z.; Lei, K.: Semiautomatic structural BIM-model generation methodology using CAD construction drawings, Journal of Computing in Civil Engineering, 34(3), 2020, 04020006. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000885

[16] 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