www.cad-journal.net
Computer-Aided Design and Applications

Immersive Experience of Movie Scenes Based on Convolutional Neural Network

Jian Yuan[1] and Tingting Peng[2]

[1]Visual Computing and Virtual Reality Key Laboratory of Sichuan Province, Sichuan Normal University, Chengdu 610066, China, yuanjian222@sicnu.edu.cn
[2]College of Movie and Media, Sichuan Normal University, Chengdu 610066, China, 20100101@sicnu.edu.cn

Corresponding author: Tingting Peng, 20100101@sicnu.edu.cn

Abstract. This paper discusses how to use virtual reality (VR) and computer aided design (CAD) technology to enhance the immersive experience of movie scenes. Firstly, DL technology, especially the model of Convolutional Neural Network (CNN), is used to classify and identify the interactive objects in the virtual movie scene. Then, the deep reinforcement learning algorithm is used to simulate and experiment these objects to determine the best interaction mode. Finally, the best interactive effect is generated by combining the input information of the audience and the results of the deep reinforcement learning algorithm. CNN is obviously superior to the 3D modeling algorithm based on wavelet neural network (WNN) model in accuracy and recall. This improvement shows that CNN can more accurately identify the key elements and details in VR movie scenes, and has better scene layout optimization ability. The focus of the research is to create a highly realistic virtual movie scene by combining deep learning (DL) technology and computer graphics, so that the audience can interact with the VR scene more naturally. This interactive design can improve the immersive experience of the movie scene and enhance the audience's viewing experience.

Keywords: Virtual Reality; Computer Aided Design; Movie Scenes; Immersive Experience
DOI: https://doi.org/10.14733/cadaps.2024.S12.189-204

1 INTRODUCTION

With the continuous development of science and technology, the film industry has entered a new digital era. Traditional film production methods can no longer meet the audience's demand for visual experience, so the immersive experience of movie scenes based on VR and CAD technology came into being. Traditional anomaly detection methods for surveillance videos are usually based on image processing and computer vision technology, but due to the complexity and diversity of surveillance video data, these methods often face many challenges. In recent years, the application of deep learning and H265 technology in the field of anomaly detection in surveillance videos has received increasing attention. Abbas and Al-Ani [1] discussed the application background, research status, technical principles, experimental methods, and application prospects of H265 and deep learning in anomaly detection in surveillance videos. Araiza et al. [2] explored the advantages and applications of immersive VR tools in problem-solving skills learning. In the traditional problem-solving skills learning process, students often rely on theoretical learning and case analysis. However, this approach often fails to provide real practical scenarios, making it difficult for students to truly master problem-solving methods. With the development of VR technology, more and more educators and researchers are exploring the application of immersive VR in problem-solving skills learning. Taking the supply chain optimization problem in the management discipline as an example, using immersive VR tools for problem-solving skills learning has a significant effect. Students can simulate the entire process of supply chain management in a highly simulated virtual environment, from supplier selection, procurement planning, to logistics operations, and comprehensively experience and learn how to optimize supply chain management. Through interactive operations and diversified solution design, students can gain a deeper understanding of the essence of supply chain problems and the criticality of solutions. Compared to traditional theoretical learning, immersive VR tools provide students with a real and immersive learning environment, enabling them to better master problem-solving skills in practice. By using the VR helmet, the audience can experience the movie scene in an immersive way. They can feel the movement of the body, the change of the environment and the interaction of roles. This first-person perspective makes the audience seem to be in the scene of the movie, which greatly enhances the immersion of the viewing experience. Bilgin and Thompson [3] introduced the definition, application background, and significance of self-supervised pose adaptation in cross domain image animation, as well as relevant methods and technologies. Their effectiveness was verified through experiments, challenges faced were analyzed, and corresponding solutions were proposed. It is necessary to establish a highly realistic game world, including elements such as scenes, props, characters, etc. At the same time, it is necessary to ensure the openness and scalability of the game, facilitating subsequent updates and extensions. Adopting advanced graphics rendering techniques such as ray tracing and volume rendering to achieve more realistic visual effects. At the same time, attention should be paid to the details of the environment, including light and shadow, color, texture, etc. Through precise positioning technology, players' body movements can be synchronized with game characters. In addition, it is necessary to design the character's movement and behavioral logic, so that players can naturally control the character. This experience transcends the limitations of traditional games, making players feel like they are in the game world, interacting directly with characters and the environment. Through VR helmet or other equipment, the audience can freely choose the viewing angle, observe the movie scene from different perspectives, and feel the reality of being in the movie scene.

With the popularity of mobile devices and the explosive growth of mobile data demand, Mobile Edge Networks (MENs) have become an important development direction for the next generation network. In mobile edge networks, the application of Augmented Reality (AR) technology is becoming increasingly widespread, bringing unprecedented convenience to people's lives and work. However, due to the limited computing and storage resources of mobile devices, how to achieve energy-saving task offloading and resource allocation in augmented reality has become an urgent problem to be solved. Chen and Liu [4] introduced a method of implementing energy-saving task offloading and resource allocation for augmented reality in mobile edge networks through deep reinforcement learning (DRL). Mobile edge network is a network architecture that migrates computing and storage resources from the cloud to the edge of the network, with advantages such as low latency, high bandwidth, and low energy consumption. Augmented reality is a technology that integrates virtual information with the real world, providing people with a richer visual experience. In mobile edge networks, augmented reality technology can help offload complex computing and storage tasks to edge servers, thereby improving device computing and storage capabilities while reducing energy consumption. This is a brand-new breakthrough to the traditional movie viewing experience. VR technology can realize the interaction between audience and movie content. Through gesture control and voice interaction, the audience can participate in the development of the movie plot, changing the traditional passive viewing mode and increasing the sense of participation and immersion. Driven by digital technology, the digital reconstruction of historical buildings and the virtual integration of murals have become possible, bringing new possibilities for virtual reality (VR) experiences. Digital reconstruction technology has enabled the rebirth of historical buildings, while virtual fusion of murals has revitalized these buildings. Elgewely et al. [5] delved into the application of these two technologies in virtual reality and how they can create immersive experiences. Virtual fusion of murals refers to the combination of digital technology with ancient murals, enabling them to be vividly displayed in a virtual environment. This technology includes digitizing mural elements and integrating them with a virtual environment to create an immersive experience that feels like a tourist is immersed in the mural. The application of digital reconstruction of historical buildings and virtual integration of murals in virtual reality has opened up a new path for the protection and inheritance of historical and cultural heritage. This technology not only allows us to appreciate disappearing historical buildings and murals in a virtual environment, but also provides interactive and immersive experiences, allowing us to have a deeper understanding and experience of historical and cultural heritage. Through CAD technology, filmmakers can design and construct various complex scenes and props more accurately. For example, producers can accurately model 3D through CAD software, and then import these models into the virtual environment through VR technology, which makes the construction of movie scenes more efficient and real. By constructing a virtual three-dimensional scene, this technology enables users to feel the reality of the movie, thus improving the audience's viewing experience.

With the continuous development of technology, virtual reality technology is becoming more and more mature, and the application fields are also becoming more and more widespread. Among them, the bridge system, as a ship command and control center, has extremely high requirements for safety and reliability. In order to improve the operational efficiency of the bridge system and ensure the safety of ships, virtual reality technology can be used to enhance the configuration design of the bridge system. Favi et al. [6] will use the bridge system as an example to illustrate the design of virtual reality enhanced configurations for customized workplaces. When designing virtual reality augmentation configurations, it is necessary to follow a certain design process. Firstly, in the requirements analysis stage, it is necessary to have a deep understanding of the actual requirements of the bridge system, including command and control, navigation management, communication, and other aspects. By communicating with crew members to understand their workflow and requirements, further clarify the goals of virtual reality augmentation configuration design. This paper will discuss how to use DL technology for 3D modeling and interactive design of VR scene. In addition, the application of CNN model in image recognition of movie scenes will be introduced. For all kinds of objects in the movie scene, this paper will use DL technology to automatically identify key information such as their types, shapes, sizes and positions, and generate three-dimensional models according to this information. The advantage of this method is that it can greatly reduce the time and cost of manual modeling and improve the accuracy and consistency of modeling. At the same time, it will also study how to optimize and improve the details of the 3D model by using other technologies such as texture mapping and lighting model, so as to make it more truly reflect the objects in the movie scene. The input information of the audience is naturally integrated with the VR scene, so that the audience can interact with the VR scene in a more intuitive and natural way. The research method is mainly based on DL and CNN model. By training and learning a large number of movie scene images, the model can automatically identify various objects in the scene, and automatically generate a three-dimensional model according to the attributes and position information of the objects. At the same time, it will also study how to use DL technology to design VR scenes interactively, so that viewers can interact with VR scenes more naturally.

With the continuous progress of technology, people's demand for visual experience is also constantly increasing. From passively receiving visual information to actively exploring and selecting visual information, people's demand for visual experience continues to escalate. In an immersive 360 real-world environment, active vision has become a new way of visual experience, pushing visual technology to a new level. Haskins et al. [7] analyzed active vision in immersive 360 real-world environments. The implementation principle of active vision mainly involves bringing users into a virtual 360-degree panoramic environment through devices such as headsets and controllers. Users can freely observe and explore the surrounding environment, while interacting with devices such as controllers. Capture images or videos through a 360-degree camera or multiple cameras to obtain complete information about the surrounding environment. Process and compress the collected images or videos, and then transmit them to the head-worn display. The head mounted display decodes and presents the transmitted images or videos, allowing users to see a 360-degree panoramic environment. Users can operate and interact with virtual environments through devices such as controllers. The significance of this study is that automatic recognition and 3D modeling of movie scenes through DL technology can greatly improve the efficiency of movie production, reduce the production cost, and bring more realistic viewing experience to the audience. At the same time, by studying how to use DL technology to design VR scenes interactively, the audience can interact with VR scenes more naturally, and the audience's participation and viewing experience are improved. Through the research of this paper, I hope to provide an efficient, accurate and innovative scene modeling and interactive design method for film production, so that the audience can feel the movie plot more truly and naturally and improve the viewing experience. At the same time, I hope to provide a general, efficient and accurate reference for other fields in 3D modeling and interactive design. The innovation of this paper is mainly reflected in the following aspects:

(1) This paper applies DL technology to 3D modeling and interactive design of movie scenes. Through a lot of training and learning, DL model can automatically identify various objects in the scene, and automatically generate a three-dimensional model according to the attributes and position information of the objects.

(2) In this paper, CNN model is applied to film scene image recognition. Through training and learning, CNN model can accurately identify various objects in the movie scene, and classify and locate them. This method can help filmmakers to build scenes and design props more efficiently, thus improving the efficiency and quality of film production.

(3) This paper proposes a new interactive design method for VR scenes. The interactive design of VR scene through DL technology can make the audience interact with VR scene more naturally, and improve the audience's participation and viewing experience. This interactive design method can not only be applied to movie scenes, but also be extended to other fields, such as games and animation.

Firstly, this paper introduces the theoretical basis of VR and CAD and their applications in film production. Then, an algorithm of VR movie scene modeling based on CNN is proposed. Finally, the performance of the algorithm is verified and summarized by experiments.

2 RELATED TECHNICAL BASIS

With the continuous development of technology, virtual reality (VR) and three-dimensional panoramic environments have become innovative tools in many fields. In the construction industry, the combination of these two technologies provides new possibilities for safety training, which helps improve employees' safety awareness and ability to respond to risks. Jeelani et al. [8] explored the application prospects and development methods of virtual reality and three-dimensional panoramic environment in building safety training. Virtual reality and three-dimensional panoramic environment have broad application prospects and display effects in building safety training. By developing virtual reality and three-dimensional panoramic environments for the construction industry, students can be provided with an immersive and interactive learning experience, improving learning effectiveness and safety awareness. With the continuous development of technology, this new type of training method will be increasingly valued and applied. With the continuous progress of technology, virtual reality technology has become an important tool in many fields, especially in the fields of education and training. The demand for learning forensic chemistry is increasing among future chemists. In order to help these learners better understand and apply forensic chemistry knowledge, Kader et al. [9] aimed to construct an interactive immersive virtual reality crime scene. This crime scene not only allows learners to experience the practical application of forensic chemistry firsthand, but also enhances their learning interest and motivation. In the design process of virtual reality crime scenes, the first thing we need to consider is the background and environment of the crime scene. We chose an abandoned factory in a city as the crime scene, and the environment and facilities of this factory provide rich clues for forensic chemistry analysis. Learners can personally collect these clues and identify key evidence through chemical analysis. Model construction requires processing the collected data and converting it into a printable 3D model, taking into account the complex biological structure of the object. Post processing includes model optimization, printing settings, etc. to ensure that the printed objects meet design requirements. Letov et al. [10] proposed a series of solutions to address these challenges. Firstly, for data collection, advanced computer vision and reverse engineering technologies are used to achieve accurate acquisition of multi-dimensional data of objects. Secondly, during the model construction process, advanced modeling software and algorithms are utilized to achieve accurate modeling of complex biomimetic 3D objects. Finally, in the post processing stage, through refined optimization and printing settings, ensure that the printed objects meet the design expectations. Mora et al. [11] aim to explore how virtual reality (VR) experiences can enhance awareness of occupational hazards at construction sites. Firstly, briefly introduce the VR experience and the understanding of occupational hazards on construction sites. Next, elaborate on the application of automation factors in VR experience and their impact on occupational hazard awareness. To create a VR experience, it is first necessary to collect detailed data on the construction site environment, including various sounds, dust particles, etc. Then, use these data to simulate the environment, and use devices such as headsets to immerse users in the environment. In addition, safety training and operation guidance can be added during the experience process, allowing workers to learn how to avoid occupational hazards while experiencing. However, there are also some limitations to the VR experience, such as high device costs and the need to improve penetration rate. Moreover, current VR technology is not yet able to fully simulate all occupational hazards on construction sites. Schnack et al. [12] used a combination of literature review and experimental research methods. Firstly, by reviewing previous literature, relevant research on shopper personality, VR technology, and the relationship between the two was organized and analyzed. Secondly, an experiment was designed using immersive virtual reality technology to observe the behavior of shoppers with different personalities in a VR store environment. The experimental results indicate that the personality of shoppers has a significant impact on behavior in an immersive virtual reality store environment. Specifically, extroverted shoppers are more willing to actively explore and try new things in VR environments, while introverted shoppers are more inclined to act in familiar environments. In addition, feeling seekers and risk averse individuals are more likely to try to find new shopping experiences in VR stores. Responsible and organized shoppers place greater emphasis on maintaining order and efficiency in virtual environments. The research results of this article reveal the impact of shopper personality on behavior in an immersive virtual reality store environment. Digital reconstruction technology has enabled the rebirth of historical buildings, while virtual fusion of murals has revitalized these buildings.

Soto et al. [13] delved into the application of these two technologies in virtual reality and how they can create immersive experiences. Digital reconstruction is the use of computer technology to create three-dimensional models of historical buildings, digitally preserving and presenting their original appearance. Through technologies such as laser scanning and photo modeling, digital reconstruction can capture various details of buildings, including shape, structure, materials, etc., and create a digital model that can be explored in a virtual environment. The reconstructed historical building can be observed from various perspectives in a virtual environment, and can even simulate human walking and visiting experiences. This technology not only allows us to appreciate the disappearing historical buildings in a virtual environment, but also helps experts in building protection and restoration work. In the field of virtual reality (VR), through tactile feedback technology, users can feel information such as texture, temperature, and texture on the surface of virtual objects. In addition, in the medical field, using enhanced tactile perception technology, doctors can conduct surgical simulations in a virtual environment and improve surgical skills. Enhanced tactile perception and tactile feedback loops as human-machine interfaces can provide users with a more realistic and efficient user experience. Sun et al. [14] analyzed enhanced tactile perception and tactile feedback loops as human-machine interfaces, aiming to achieve immersive interaction. By using physics engines and algorithms, tactile information such as force, temperature, and humidity generated when virtual objects interact with users can be calculated. Transmit tactile perception information to users through a feedback loop, enabling them to feel the realism of the virtual environment. By real-time transmission and processing of tactile data, smooth and real-time tactile feedback can be achieved. Takao et al. [15] explored the impact of immersive experience on blink rate in virtual reality games from aspects such as visual output, visual pressure, and neural response. In order to reduce the blink rate of players, game developers can take corresponding measures from aspects such as technology, visual presentation, and neuroscience. By optimizing the quality of game graphics, reducing visual pressure, and reducing neural reactions, players can enjoy an immersive experience while reducing the number of blinks, thus obtaining a more comfortable and natural gaming experience. The object-oriented immersive virtual reality practice game application framework aims to provide an open and scalable framework. Facilitate developers to easily create VR game applications while providing users with a more realistic gaming experience. Wang et al. [16] created a highly realistic virtual movie scene that allows viewers to interact more naturally with VR scenes. This system is responsible for processing user input, such as input from devices such as handles, touch screens, keyboards, etc., and passing the input to the object system for processing. At the same time, the system is also responsible for rendering the output of the object system to VR devices to provide an immersive visual experience. This system is responsible for handling communication between clients and servers, including the transmission of user data, scene information, and other data. With the advancement of technology, virtual reality and immersive experience technology are gradually integrating into our daily lives, and they have wide applications in many fields, including cognitive health and happiness enhancement.

Yan and Lv [17] explored the impact of immersive virtual reality systems on online social applications. Immersive virtual reality system is a technology that can provide users with an immersive experience. The development of immersive virtual reality systems has gone through a long process from the earliest text mud to the current virtual reality technology. With the improvement of hardware device performance and the decrease in price, immersive virtual reality systems have gradually become popular and important tools in various fields such as gaming, education, and healthcare. To better illustrate the impact of immersive virtual reality systems on online social applications, we can choose a typical case for analysis. For example, the virtual reality social platform "VRChat" allows users to create their own virtual images and communicate with other users in a virtual environment. With the improvement of people's living standards, the demand for interior design is also increasing. In order to meet people's aesthetic needs and improve the quality of interior design, many designers have begun to use 3D computer-aided simulation (3D CAF) technology for interior design optimization. Yang [18] aims to introduce the application of 3D CAF in interior design optimization, helping students master the technology and improve their design skills. This teaching adopts a combination of theory and practice. Firstly, it introduces the basic principles and advantages of 3D CAF, and then demonstrates how to use 3D CAF for interior design optimization through specific cases. In the teaching process, we will highlight the practical application of 3D simulation technology in the field of interior design, with a focus on explaining how to use 3D CAF to solve practical problems and improve design efficiency. With the continuous development of technology, the combination of virtual reality (VR) and building information modeling (BIM) technology provides new solutions for emergency response training in highway tunnels. By combining virtual reality and building information modeling technology, a more realistic simulation environment for emergency response in highway tunnels can be created, improving training effectiveness and trainees' response capabilities. Yu et al. [19] will explore how to combine virtual reality with building information modeling to improve the level of emergency response training for highway tunnels. By utilizing virtual reality and building information modeling techniques, a highly realistic emergency response simulation environment for highway tunnels can be constructed. Zhang et al. [20] analyzed the research on 3D architectural scene construction technology based on augmented reality. Augmented reality technology is a technology that can integrate virtual information with the real world. It inserts virtual objects into the user's actual environment through real-time rendering algorithms, enabling users to interact with these virtual objects, thereby creating a virtual reality fusion experience. In the field of architecture, augmented reality technology provides designers with a new way to present their designs, while also providing engineers with an effective tool for understanding and manipulating complex architectural scenes. The construction technology of 3D architectural scenes involves a series of complex technologies and methods, including 3D model establishment, data collection, processing, and display. These technologies help us transform the designer's imagination into concrete 3D models, and then present these models to users through augmented reality technology. Designers can use this technology to present their design proposals for better communication and modification.

3 METHODOLOGY

3.1 3D Modeling of VR Scene

3D modeling of VR scene based on DL is an important aspect of this paper. Through high fidelity and immersive display, users can better feel the plot and atmosphere of the movie. At present, although there have been some researches on VR scene modeling and interactive design, these methods often have some limitations, such as unable to automatically identify various objects in the scene and unable to describe the scene in detail. In order to solve these problems, this paper puts forward a method of 3D modeling and interactive design of VR scene based on DL to solve the shortcomings in the existing technology. With the development of DL, more and more researchers began to try to use DL to model the VR scene. This method can learn from a large number of images, automatically identify various objects in the images, and model them in three dimensions, which greatly improves the efficiency and accuracy of modeling. In terms of immersive experience of movie scenes, this modeling method can bring more realistic scene effects and make the audience feel better immersed. In the immersive experience of movie scenes, the 3D modeling method of VR scenes based on DL can automatically identify various objects in movie scenes, conduct accurate 3D modeling, and generate realistic virtual environment. This can not only improve the visual experience of the audience, but also let the audience participate in the plot of the film through interactive design, thus improving the immersive experience of the film.

In this paper, a 2×2 convolution kernel is used, which can slide on the image and perform convolution calculation at each position. For each convolution kernel position, the convolution kernel is multiplied by the input image window at that position pixel by pixel, and then these products are added up to get an output value. This output value is called a pixel of the feature map. After getting each pixel of the feature map, it needs to be input into an activation function to introduce nonlinear characteristics. In order to reduce the dimension of feature mapping and the risk of over-fitting, a pooling layer is added after convolution calculation. The pooling layer uses a window size of 2×2, and takes the maximum value of each 2×2 region of feature mapping as the output. After the last pool layer, we added a fully connected layer. The fully connected layer connects the pixel points of each feature map to the output layer of the neural network. Finally, an output layer is added, which maps the output of the fully connected layer to the probability distribution of the actual label. In this paper, softmax activation function is used to transform the output of neural network into the probability distribution of actual tags. The convolution calculation process is shown in Figure 1.

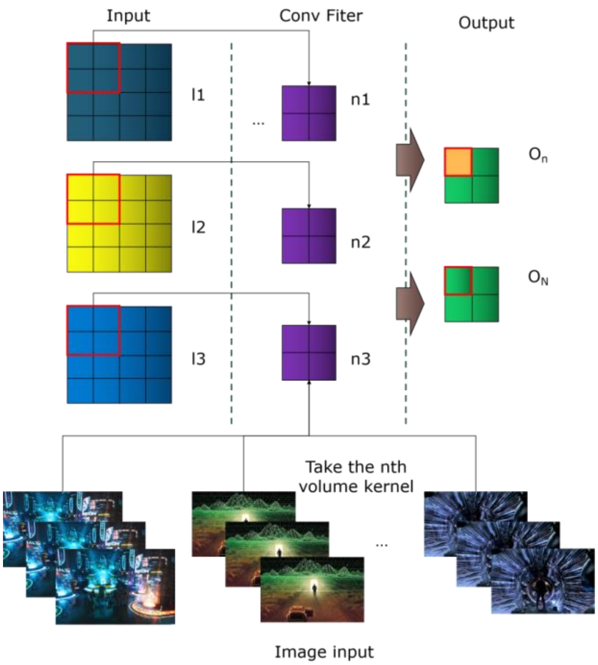


Figure 1: Convolution calculation process.

Using CNN technology to interactively design VR scenes can improve the quality and fidelity of immersive experience of movie scenes. By identifying interactive objects, using CNN model to classify and identify them, and then using deep reinforcement learning and other technologies to simulate and experiment these objects, a highly realistic virtual environment can be realized, and viewers can interact with VR scenes more naturally. Convolution is the most important part of CNN, which is used to extract feature information from data:

$$H(x,y)=A\otimes k(x,y)=\sum_{M,N}A(m,n)k(x-m,y-n) \tag{1}$$

Assuming that the l-th layer is a fully connected layer, the weight matrix is W^l and the bias is b^l, the calculation process of the fully connected layer is:

$$Z_j^l=f\left(W^l X^{l-1}+b^l\right) \tag{2}$$

Because the expressive ability of linear model is not enough, in order to obtain nonlinear model, a nonlinear activation function is added to CNN to produce nonlinear response to input. Use ReLU as activation function:

$$f(x)=\max(0,x) \tag{3}$$

Through the above methods, a large number of 3D models of movie scenes with strong sense of reality can be quickly constructed. These models can be applied to the VR environment to provide a more immersive viewing experience. For example, the audience can freely walk and explore in the virtual environment through the head-mounted device, and observe the details in the movie scene through the first-person perspective; The audience can also interact with objects in the virtual environment, such as picking up props and climbing buildings. This immersive film experience not only enhances the audience's viewing experience, but also provides a broader innovation space for film production.

3.2 Interactive Design of VR Scene

In VR scene, different objects have different interactivity. For example, some objects can only be observed by the audience, while others can be manipulated or interacted by the audience. Therefore, it is very important to determine the position and type of interactive objects in the scene. CNN model can learn the characteristics and patterns of objects, and automatically identify interactive objects in the scene according to the input information of the audience. CNN model usually consists of several convolution layers, pooling layers and fully connected layers. The convolution layer can capture the local features of the image, the pooling layer can reduce the dimension of the data, and the fully connected layer is used to map the features of the previous layer to the output space. Once the interactive objects are identified, the deep reinforcement learning technology can be used to simulate and experiment these objects.

Enhanced 3D model (E3D) is an advanced graphics processing technology, which is often used in film production and game development. The main function of E3D is to obtain the three-dimensional model and texture data of the object by scanning the object in multi-angle and multi-illumination conditions, so as to create realistic visual effects in the virtual environment. First of all, it is necessary to obtain high-precision and high-resolution image data, such as images taken by drones or scanned point cloud data. These data can provide rich details and accurate object shape information. Preprocessing the acquired image data, including denoising and image enhancement, to improve the identifiability of the boundary line. Using E3D, the image boundary line is transformed into a spatial boundary line, and the spatial boundary line is closed, so that a more accurate feature boundary line can be obtained. The technical process is shown in Figure 2.

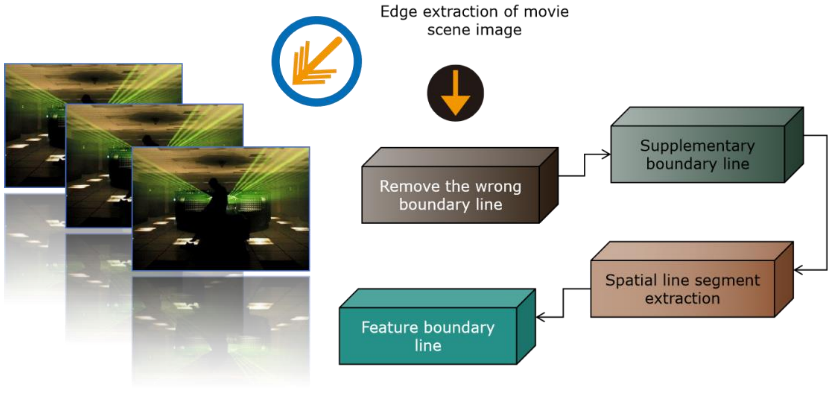


Figure 2: Algorithm flow of feature boundary line extraction.

When the spatial boundary line is closed to obtain a more accurate characteristic boundary line, the preliminary spatial boundary line is optimized to remove redundant or false detection boundary points, so that the boundary line is more concise and accurate. By fitting the boundary line in space and repairing the discontinuous part, the boundary line forms a closed curve in space. Representative feature information, such as shape, scale and direction, is extracted from the closed boundary line. According to the extracted feature information, a more accurate feature boundary line is generated.

Suppose that the projection of the image of the three-dimensional space point $P(X,Y,Z)$ on the left and right image planes $p_1(x_1,y_1)$ and $p_2(x_2,y_2)$ respectively. Superimpose the first camera on the world coordinate system, and the X coordinates of P point can be obtained as follows:

$$X_1 = \frac{x_1}{f}\left(Z_1 - f\right) \tag{4}$$

Where X_1 and Z_1 indicate that the first camera moves to the origin of the world coordinate system, and the second camera and P point will also move with it while maintaining the relative geometric relationship. If you move the second camera to the origin of the world coordinate system, then:

$$X_2 = \frac{x_2}{f}\left(Z_2 - f\right) \tag{5}$$

Because the baseline length is L, and the Z coordinate of P point is the same for the two camera coordinate systems, there are:

$$X_2 = X_1 + L, Z_2 = Z_1 = Z \tag{6}$$

Substituting Formula (6) into Formula (4) and Formula (5), we can get:

$$X_1 = \frac{x_1}{f}\left(Z_1 - f\right), X_1 + L = \frac{x_2}{f}\left(Z_2 - f\right) \tag{7}$$

From this, the spatial coordinates of point P can be obtained as follows:

$$Z = f + \frac{fL}{x_2 - x_1} \tag{8}$$

Define parallax:

$$d = x_2 - x_1 \tag{9}$$

The depth value of point P is:

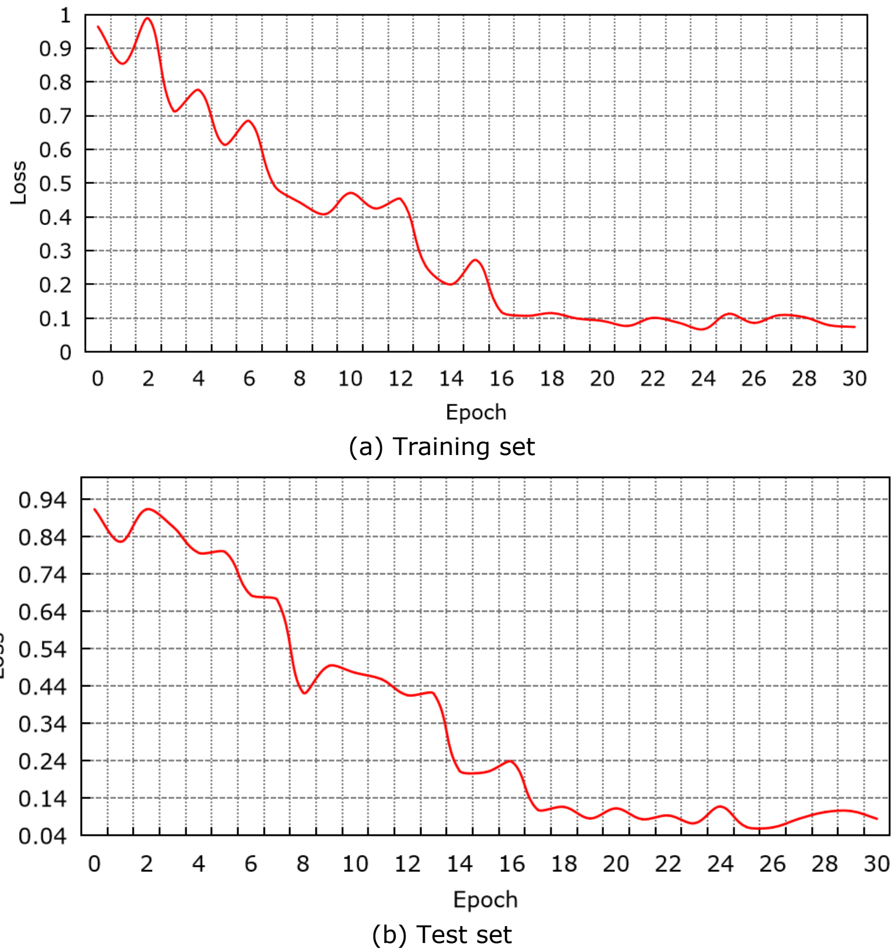
$$Z = f + \frac{fL}{d} = f\left(1 + \frac{L}{d}\right) \tag{10}$$

Through training and learning a large number of movie scene images, the DL model can automatically identify various objects in the scene, and automatically generate a three-dimensional model according to the attributes and position information of the objects. In order to achieve this goal, CNN model is adopted to identify the movie scene images. In the training process, a large number of movie scene images are used as training data, including images with different angles, different lighting conditions and different object postures. We input the training data into the CNN model, and train and optimize the model, so that it can automatically identify various objects in the image and determine its position, size, shape and other parameters.

Finally, according to the audience's input information and the results of the deep reinforcement learning algorithm, the best interactive effect is generated. The audience's input information may include gestures, voices, eye movements, etc., which will be converted into instructions by CNN model, thus realizing the interaction with VR scenes. For example, the audience can move objects in the virtual environment through gestures, or change the layout of the VR scene through voice. According to the results of the deep reinforcement learning algorithm, I can generate the best interactive effects, such as moving objects and changing the scene layout. These interactive effects can be simulated and experimented in virtual environment to ensure their accuracy and stability.

4 RESULT ANALYSIS AND DISCUSSION

The experiment used 160 groups of pictures as data sets, including 125 groups as training sets and 35 groups as test sets. This division method can make the model learn more patterns on the training set and evaluate the generalization ability of the model on the test set. The training process is divided into two groups, one group contains the training of movie scenes, and the other group does not contain the training of movie scenes. Each group carries out 20,000 iterative trainings, each of which takes 1 hour. This shows that the training time of the model is long, and the training process of the model is relatively stable. As shown in Figure 3.



(a) Training set

(b) Test set

Figure 3: Loss curve of two trainings.

As can be seen from the Loss curve, the Loss curves of the two trainings have shown a gradual downward trend, which shows that the model is constantly optimized during the training process and more characteristics of movie scene pictures are learned. At the same time, the Loss curve of training without movie scenes decreases faster, which may be because the training data of this group is richer and more diverse, which is helpful for the model to learn the characteristics of various movie scenes better. After training, the generator loss of the two models is 0.091 and 0.061 respectively, and the curves converge. This shows that the generator part of the model has learned how to generate a more realistic CAD model of the movie scene. At the same time, the generator loss without the training of movie scenes is low, which shows that the model is more accurate and true when generating the CAD model of movie scenes.

In this experiment, a reasonable data set division method is adopted, which makes the model learn more patterns on the training set and evaluates the generalization ability of the model on the test set. The training process of the model is relatively stable, and the training time is long, which shows that the model can learn more characteristics of movie scene pictures. The Loss graph shows that the model is constantly optimized in the training process, and more characteristics of the movie scene pictures are learned. At the same time, the Loss curve of the training without movie scenes decreases faster, indicating that the training data of this group is more abundant and diverse. The analysis of generator loss shows that the generator part of the model has learned how to generate a more realistic CAD model of the movie scene. At the same time, the generator loss without the training of movie scenes is low, which shows that the model is more accurate and true when generating the CAD model of movie scenes.

In the experiment, the running efficiency of different scale movie scenes with different nodes is tested (Figure 4).

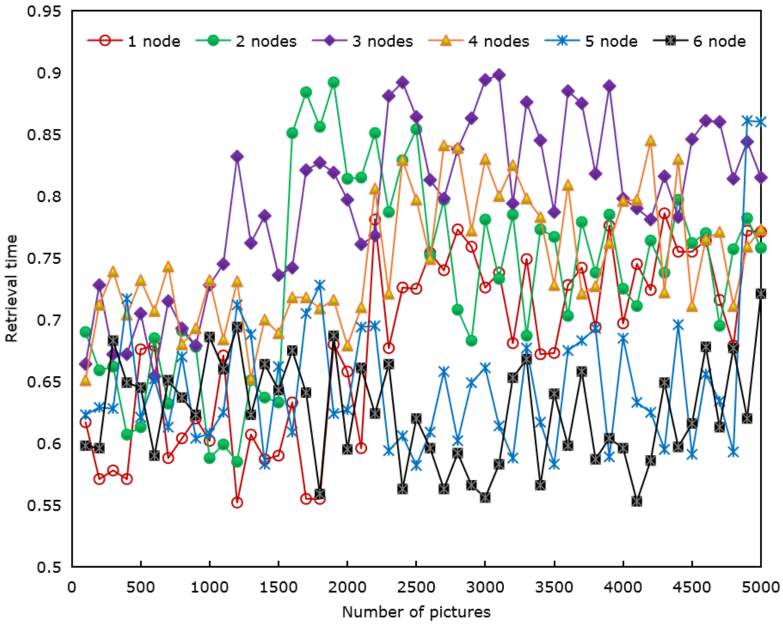


Figure 4: Time-consuming of image recognition with different number of nodes.

From the experimental results, it can be found that with the increase of the scale of the movie scene, the number of nodes needed to identify the image also increases. This trend is reasonable, because larger scenes usually need more nodes to process and render details. When the scale of the movie scene is not large, increasing the number of nodes does not obviously improve the recognition efficiency. This may be because the details and complexity in the scene are relatively low, so fewer nodes have been able to deal with it effectively. When the image scale becomes larger and larger, the recognition efficiency of multi-nodes shows an obvious upward trend. This may be because as the scale of the scene increases, the details and complexity increase accordingly, and more nodes are needed to accurately render and process these details. In addition, larger scenes also need more computing resources for high-precision image processing, which may also be one of the reasons for improving efficiency. In the process of increasing the number of nodes, the improvement of recognition efficiency is not linear. In other words, adding more nodes may not bring corresponding improvement in recognition efficiency. This is because the increase of the number of nodes will also lead to the increase of management and coordination expenses, thus limiting the improvement of efficiency.

These findings have important guiding significance for practical application. For example, when dealing with large-scale movie scenes, we can consider using more nodes to improve the recognition efficiency. However, it is also necessary to consider the management and coordination costs brought by the increase in the number of nodes. In order to achieve the best efficiency, it may be necessary to properly weigh and optimize the number of nodes.

Compare the time-consuming of CNN model and WNN model in film scene processing, as shown in Figure 5. As can be seen from Figure 5, the time consumption of CNN model in processing movie scene images is shorter than that of WNN model. This result may be influenced by many factors.

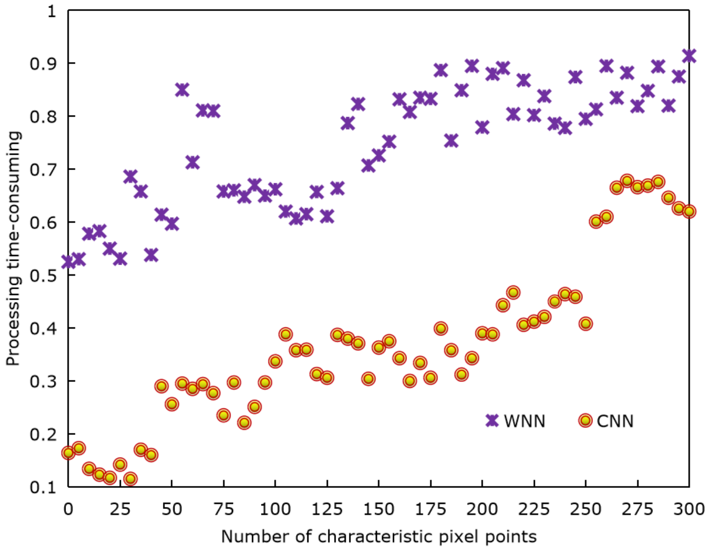


Figure 5: Time-consuming test of image recognition processing.

As can be seen from the figure, the WNN method takes longer to digitize the movie scene. CNN model can usually use GPU for parallel computing, while WNN model may only use CPU for computing. On large data sets, this difference in computational efficiency may lead CNN model to be faster than WNN model in training and reasoning. In addition, GPU usually has higher computing density and memory bandwidth than CPU, which further improves the processing efficiency of CNN model. CNN model usually has fewer parameters and lower model complexity, which means that CNN model can complete the training and reasoning process faster with the same computing resources. On the contrary, WNN model may have more parameters and more complex structure, which may lead to longer training and reasoning time. CNN model has convolution layer and pooling layer, which can automatically extract features from images. This automatic feature representation learning ability can make CNN model more efficient and accurate in processing image data. In contrast, WNN model may need to design features manually, which may increase processing time and complexity. Film scene images may need preprocessing (for example, scaling, cropping, normalization, etc.) to meet the input requirements of the model. CNN model can usually accept original images or simply preprocessed images as input, while WNN model may need more complicated data processing and preparation steps. This data processing and preparation time may also lead to a longer time-consuming WNN model.

From the experimental results of comparing the recall and accuracy of the algorithm for VR movie scene pattern optimization (as shown in Figure 6 and Figure 7), the algorithm has higher accuracy in 3D modeling of VR movie scenes, which is more than 20% higher than the 3D modeling algorithm based on WNN model.

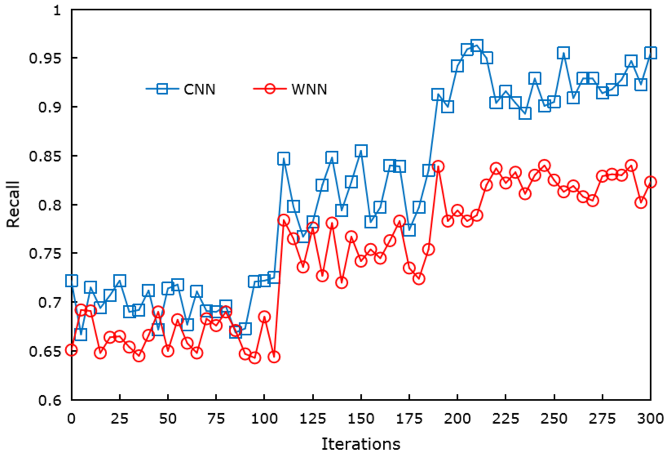


Figure 6: Comparison of recall rate of VR movie scene optimization.

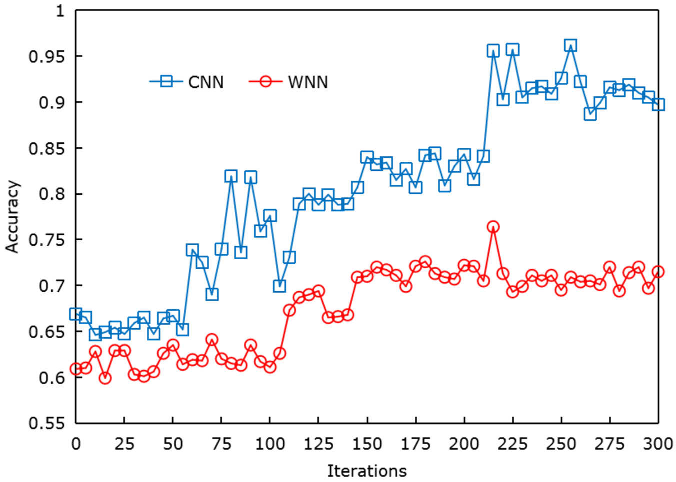


Figure 7: Comparison of accuracy of VR movie scene optimization.

As can be seen from Figure 6 and Figure 7, CNN is obviously superior to the three-dimensional modeling algorithm based on WNN model in accuracy and recall. This improvement shows that CNN can more accurately identify the key elements and details in VR movie scenes, and has better scene layout optimization ability. In the field of VR, 3D modeling with high accuracy and recall has important application value for scene layout optimization. In addition, CNN can accurately locate the edge contour of the optimized space, which provides an important reference for the layout optimization of movie scenes.

5 CONCLUSION

VR technology can realize the interaction between audience and movie content. Through gesture control and voice interaction, the audience can participate in the development of the movie plot, changing the traditional passive viewing mode and increasing the sense of participation and immersion. Through CAD technology, filmmakers can design and construct various complex scenes and props more accurately. In this paper, the immersive experience of movie scenes based on VR and CAD technology is studied, and a method of 3D modeling and interactive design of VR scenes based on DL is proposed. This method combines CNN model to realize image recognition of movie scenes, which can accurately identify various objects in movie scenes, and classify and locate them. The interactive design of VR scene through DL technology can make the audience interact with VR scene more naturally, and improve the audience's participation and viewing experience. The experimental results show that the algorithm proposed in this paper has high accuracy in 3D modeling of VR movie scenes, which is more than 20% higher than that based on WNN model. This shows that the algorithm has a wide application prospect and can provide useful reference for the layout optimization of VR movie scenes. To sum up, the research results of this paper can further improve the immersive experience of movie scenes, enhance the audience's viewing experience, and provide new ideas and methods for the development of VR movies, which has important theoretical and application value.

Compared with WNN model, CNN model takes less time to process movie scene images. This may be because CNN model has higher computational efficiency, lower model complexity, stronger learning ability of automatic feature representation and more optimization. However, it should be pointed out that there may be differences between the two methods in dealing with movie scenes of different scales. Therefore, future research can conduct more comprehensive experiments on movie scenes of different scales, so as to better compare the performance of the two methods.

6 ACKNOWLEDGEMENT

This work was supported by Department of Education Humanities and Social Sciences Project, Research on Urban Children's Recreation Landscape Design from the Perspective of Place Spirit, (No. JGYQ2022012); Sichuan Normal University 2022 experimental technology and management project, Design and implementation of immersive augmented reality technology integrated into drama performance teaching system.

Jian Yuan, https://orcid.org/0009-0005-0078-1198
Tingting Peng, https://orcid.org/0009-0006-7107-6477

REFERENCES

[1] Abbas, Z.-K.; Al-Ani, A.-A.: Anomaly detection in surveillance videos based on H265 and deep learning, International Journal of Advanced Technology and Engineering Exploration, 9(92), 2022, 910-922. https://doi.org/10.19101/IJATEE.2021.875907
[2] Araiza-Alba, P.; Keane, T.; Chen, W.-S.; Kaufman, J.: Immersive virtual reality as a tool to learn problem-solving skills, Computers & Education, 164(2), 2021, 104121. https://doi.org/10.1016/j.compedu.2020.104121
[3] Bilgin, C.; Thompson, M.: Processing presence: how users develop spatial presence through an immersive virtual reality game, Virtual reality, 26(2), 2022, 649-658. https://doi.org/10.1007/s10055-021-00528-z
[4] Chen, X.; Liu, G.: Energy-efficient task offloading and resource allocation via deep reinforcement learning for augmented reality in mobile edge networks, IEEE Internet of Things Journal, 8(13), 2021, 10843-10856. https://doi.org/10.1109/JIOT.2021.3050804
[5] Elgewely, M.-H.; Nadim, W.; ElKassed, A.; Yehiah, M.; Talaat, M.-A.; Abdennadher, S.: Immersive construction detailing education: building information modeling (BIM)–based virtual reality (VR), Open House International, 46(3), 2021, 359-375. https://doi.org/10.1108/OHI-02-2021-0032
[6] Favi, C.; Moroni, F.; Manieri, S.: Virtual reality-enhanced configuration design of customized workplaces: a case study of ship bridge system, Computer-Aided Design and Applications, 16(2), 2019, 345-357. https://doi.org/10.14733/cadaps.2019.345-357

[7] Haskins, A.-J.; Mentch, J.; Botch, T.-L.; Robertson, C.-E.: Active vision in immersive, 360 real-world environments, Scientific Reports, 10(1), 2020, 14304. https://doi.org/10.1038/s41598-020-71125-4
[8] Jeelani, I.; Han, K.; Albert, A.: Development of virtual reality and stereo-panoramic environments for construction safety training, Engineering, Construction and Architectural Management, 27(8), 2020, 1853-1876. https://doi.org/10.1108/ECAM-07-2019-0391
[9] Kader, S.-N.; Ng, W.-B.; Tan, S.-W.-L.; Fung, F.-M.: Building an interactive immersive virtual reality crime scene for future chemists to learn forensic science chemistry, Journal of Chemical Education, 97(9), 2020, 2651-2656. https://doi.org/10.1021/acs.jchemed.0c00817
[10] Letov, N.; Velivela, P.-T.; Sun, S.: Challenges and opportunities in geometric modeling of complex bio-inspired three-dimensional objects designed for additive manufacturing, Journal of Mechanical Design, 143(12), 2021, 121705. https://doi.org/10.1115/1.4051720
[11] Mora, S.-J.; Muñoz, R.-F.; Valero, I.: Factors for the automation of the creation of virtual reality experiences to raise awareness of occupational hazards on construction sites, Electronics, 10(11), 2021, 1355. https://doi.org/10.3390/electronics10111355
[12] Schnack, A.; Wright, M.-J.; Elms, J.: Investigating the impact of shopper personality on behaviour in immersive Virtual Reality store environments, Journal of Retailing and Consumer Services, 61(1), 2021, 102581. https://doi.org/10.1016/j.jretconser.2021.102581
[13] Soto, M.-O.; Fuentes, P.-A.; Martin, G.-J.: A digital reconstruction of a historical building and virtual reintegration of mural paintings to create an interactive and immersive experience in virtual reality, Applied Sciences, 10(2), 2020, 597. https://doi.org/10.3390/app10020597
[14] Sun, Z.; Zhu, M.; Shan, X.; Lee, C.: Augmented tactile-perception and haptic-feedback rings as human-machine interfaces aiming for immersive interactions, Nature Communications, 13(1), 2022, 5224. https://doi.org/10.1038/s41467-022-32745-8
[15] Takao, M.: Immersive experience influences eye blink rate during virtual reality gaming, Polish Psychological Bulletin, 50(1), 2019, 49-53. https://doi.org/10.24425/ppb.2019.126018
[16] Wang, Y.; Ijaz, K.; Yuan, D.; Calvo, R.-A.: VR‐Rides: An object‐oriented application framework for immersive virtual reality exergames, Software: Practice and Experience, 50(7), 2020, 1305-1324. https://doi.org/10.1002/spe.2814
[17] Yan, Z.; Lv, Z.: The influence of immersive virtual reality systems on online social application, Applied Sciences, 10(15), 2020, 5058. https://doi.org/10.3390/app10155058
[18] Yang, J.: Teaching optimization of interior design based on three-dimensional computer-aided simulation, Computer-Aided Design and Applications, 18(S3), 2021, 72-83. https://doi.org/10.14733/cadaps.2021.S3.72-83
[19] Yu, X.; Yu, P.; Wan, C.; Wang, D.; Shi, W.; Shou, W.; Wang, X.: Integrating virtual reality and building information modeling for improving highway tunnel emergency response training, Buildings, 12(10), 2022, 1523. https://doi.org/10.3390/buildings12101523
[20] Zhang, X.; Yao, J.; Dong, L.; Ye, N.: Research on 3D architectural scenes construction technology based on augmented reality, Journal of Computational Methods in Sciences and Engineering, 21(2), 2021, 311-327. https://doi.org/10.3233/JCM-204390