IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 41, No. 2, April, 2024, pp. 979-987
Journal homepage: http://iieta.org/journals/ts

3D Image Modeling and Visual Presentation Technologies for Education

Lei Wang*, Boyan Yin, Mengwei Zhu, Shuang Hao

School of Fine Arts and Design, Hebei Normal University, Shijiazhuang 050010, China

Corresponding Author Email: 15533918645@163.com

https://doi.org/10.18280/ts.410238

Received: 6 December 2023
Revised: 11 March 2024
Accepted: 26 March 2024
Available online: 30 April 2024

Keywords:
3D image modeling, visual presentation technologies, educational technology, inter-layer feature learning, visual rendering optimization, deep learning, real-time interaction, immersive learning

ABSTRACT

With the extensive application of digital media and interactive technologies in the field of education, 3D image modeling and visual presentation technologies have become key tools for enhancing the learning experience. These technologies can concretize abstract educational content, assisting students in understanding complex knowledge through intuitive means. Although current 3D modeling and rendering technologies are widely applied in education, existing methods still need improvement in feature extraction, model expressiveness, and rendering efficiency, particularly in meeting the demands for real-time interaction and providing immersive learning experiences. Addressing this issue, this paper delves into 3D educational image modeling methods based on inter-layer feature learning and explores visual rendering optimization strategies for 3D educational image scenes. By improving the capability of deep learning models to process educational content and optimizing the rendering process to support real-time interaction in large-scale scenes, this research aims to provide more accurate and efficient teaching tools, thereby enhancing the efficiency and quality of teaching and learning. The results indicate that the proposed methods significantly improve model detailing and rendering speed while ensuring visual effects, offering substantial practical significance in enhancing interactivity and immersion in educational technology.

1. INTRODUCTION

With the rapid development of computer technology, the application of 3D image modeling and visual presentation technologies in the field of education has become an important means to improve teaching quality and learning efficiency [1-3]. 3D educational images can provide students with an intuitive and interactive learning experience, helping them to better understand complex concepts and processes [4]. Especially in fields such as science, medicine, and engineering, 3D visual materials make abstract theories more visual, enhancing students' spatial cognition abilities, thus showing great application potential and value in the field of education [5-7].

However, the educational application of 3D images not only requires high-quality image modeling but also demands that visual presentations accurately convey educational content [8-10]. Moreover, to meet the needs of different teaching scenarios, related research work must be optimized and improved for the specific issues of the education field [11, 12]. This includes how to better capture and understand the key features of teaching content, and how to achieve efficient, real-time visual rendering to ensure the interactive experience in the educational environment [13, 14]. Therefore, in-depth research on the application of 3D image modeling and visual presentation technologies in education is of significant practical importance.

Although existing studies have proposed various 3D modeling and visual rendering methods, there are still some deficiencies [15-17]. For example, existing modeling technologies often overlook the learning of inter-layer features, which limits the accuracy and expressiveness of the model in capturing complex teaching content. At the same time, traditional visual rendering technologies often encounter inefficiency issues when dealing with large-scale educational scenes, preventing real-time interaction, affecting the continuity of teaching, and the immersion of learners [18-20].

To address the above problems, the main content of this paper is divided into two parts: Firstly, it studies the 3D educational image modeling technology based on inter-layer feature learning, optimizing the feature extraction and representation capabilities of the model through deep learning methods to better adapt to the complexity and diversity of educational content. Secondly, it discusses visual rendering optimization strategies for 3D educational image scenes, aiming to improve rendering efficiency through algorithm optimization to achieve smoother and more realistic visual presentation effects, thereby enhancing the interactivity of teaching and the experience of learners. Through these two aspects of research, this paper aims to advance the modeling and rendering technologies of 3D educational images, provide more efficient and vivid learning tools for the field of education, and pave new paths for the progress and application of related technologies.

2. 3D EDUCATIONAL IMAGE MODELING BASED ON INTER-LAYER FEATURE LEARNING

2.1 Overall network architecture

In the field of education, especially when it involves the design of intelligent facilities and gerontechnology education, a deep understanding of complex structures and processes often requires students to observe continuous changes and microscopic details. This means that 3D images for educational purposes not only need to have high clarity to ensure all details are clearly presented but should also be able to show these details' continuous changes in real environments, allowing students to understand the entire process of change smoothly, like watching a movie. This is particularly important in teaching smart home design, as students need to understand how devices operate continuously under different conditions and how they adapt to the specific needs of elderly users. However, due to cost and technological limitations, especially in the field of gerontechnology, it is often difficult to directly obtain a sufficient number of continuous images to meet this need. For example, when teaching how to design smart devices suitable for the elderly, continuous dynamic presentations can help students understand how different devices and user interfaces can adaptively adjust as the user's behavior and health status change.

For this reason, the 3D educational image modeling method based on inter-layer feature learning proposed in this paper uses inter-layer interpolation technology, as shown in Figure 1. Inter-layer interpolation generates intermediate images between two existing adjacent images, not only greatly enriching the image sequence but also enhancing the learning experience of students, helping them understand complex 3D structures and dynamic changes in a more continuous and intuitive way. Specifically, this method first analyzes two adjacent 3D images input, extracting the features and deformation trends of the images, and then estimates the transition features between layers through the generative model obtained by training. In this way, the generated intermediate images maintain a high degree of visual consistency with the original images and can reveal more subtle detail changes that are difficult to observe. Assuming the generation of inter-layer images is represented by U_{MID}, the formula is as follows:

$$U_{MID} = d\left(U_0, U_2\right) \tag{1}$$

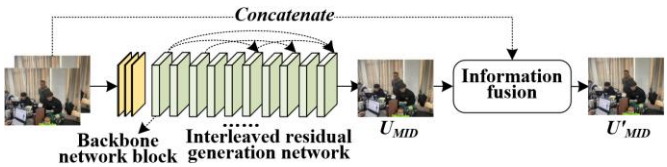

Figure 1. Principle of 3D educational image modeling method based on inter-layer feature learning

The appropriate function d for generating inter-layer images is crucial for achieving high-quality 3D educational image layer interpolation. For this challenge, this paper proposes an inter-layer interpolation network for 3D educational images. This network consists of three main parts: 1) The first part is preprocessing and feature extraction, that is, using a three-layer convolutional network to process the two input 3D educational images, aiming to extract the basic features and texture information of the images. These features provide the necessary raw information for the subsequent network's inter-layer generation task, ensuring that the generated images maintain visual coherence with the original images and retain the key visual information of the educational content. 2) The second part is the interleaved residual generation network. After feature extraction, the extracted features are fed into an interleaved residual network. An intermediate result is generated by learning the residuals between the input images. This intermediate result represents the predicted changes between layers, enhancing the original data. 3) The third part is the inter-layer information fusion module. By integrating the initially extracted features and the intermediate generation results, the effective information between image layers is strengthened, and the overall visual effect of the images is enhanced, ensuring that the generated inter-layer images are more accurate in anatomical structure and functional presentation, thus providing better visual support for understanding complex 3D structures in the field of education. Specifically, assuming the intermediate result output by the interleaved residual generation network is represented by U'_{MID}, the image features extracted by the three convolutional layers are represented by a, and the interleaved residual generation network is represented by $Y(.)$, the formula is:

$$U_{MID}' = Y\left(a\right) \tag{2}$$

Assuming the inter-layer information fusion module is represented by $D(.)$, using $D(.)$ to integrate U'_{MID} can generate the final inter-layer image, represented by U_{MID}, as follows:

$$\begin{aligned} U_{MID} &= d\left(U_0, U_2\right) = D\left(Y\left(x\right), U_0, U_2\right) \\ &= D\left(U_{MID}^{'}, U_0, U_2\right) \end{aligned} \tag{3}$$

2.2 Residual and inter-layer information fusion

In the field of modeling and visual presentation of educational 3D images, the application of deep neural networks can greatly enrich the feature expression of images, improving the accuracy of modeling and the realism of visual effects. However, simply increasing the depth of neural networks often encounters problems of gradient vanishing or exploding, leading to poor training effects, a phenomenon known as network degradation. This is particularly prominent for complex educational 3D images, as these images often require deeper networks to extract more subtle structural information. To enhance network performance and overcome the degradation problem, this paper designs a new network architecture that can effectively extract rich hierarchical features while avoiding network degradation. The necessity of this architecture lies in its ability to allow the network to deeply learn the complex structures of 3D images while maintaining training stability, thus better serving educational purposes, such as providing more accurate and vivid visual teaching materials. Specifically, an interleaved residual generation network structure is adopted, utilizing skip connections to achieve identity mapping, allowing gradients to directly flow through some layers. These skip connections help alleviate the problem of gradient vanishing because they provide a shortcut for gradients, maintaining good information and gradient flow even in very deep networks. Moreover, through carefully designed interleaving patterns, the information transfer between lower and higher layers is

optimized, achieving effective propagation of information in deep networks.

Assuming the input of the residual network block is represented by u, the number of layers by m, and the stacked network layers by $G(u)$, the formula is as follows:

$$u_m = G(u_{m-1}) + u_{m-1} \tag{4}$$

The backbone network block for processing image feature information is set with five convolution layers $Z0$, $Z1$, $Z2$, $Z3$, $Z4$. To fully utilize feature information and better preserve image details, convolution layers use small kernels of size 3*3, with three interleaved skip connections $T0$, $T1$, $T2$. Based on the above formula, the inputs for $Z2$ and $Z4$ can be derived:

$$u_2 = G(u_1) + u_0 \tag{5}$$

$$u_4 = u_0 + G(u_1) + G(u_3) \tag{6}$$

This paper constructs a deep interleaved residual generation network composed of 10 network blocks. This network architecture uniquely introduces skip connections within and between network blocks, which not only enhances feature reuse but also preserves key information of the image and effectively mitigates the degradation problem during network training. Specifically, skip connections are set from block 0 to block 9, from block 0 to block 5, from block 2 to block 7, and from block 5 to block 9. Such a tiered design allows feature information to be effectively transmitted and fused between layers of different depths, helping to capture both the detailed information and overall contour of the image simultaneously. In the actual construction of this network structure, special attention was paid to how to deal with the dimension matching problem when concatenating feature maps of different dimensions in skip connections. For this purpose, a special convolution layer is set at each concatenation point of the skip connection to adjust and match the dimensions of the feature maps output by different network blocks. This design ensures the seamless fusion of features from different levels, optimizing information flow, and providing a more refined feature representation for the task of inter-layer interpolation of 3D images.

In the current state of application of 3D educational image modeling and visual presentation, traditional methods of inter-layer interpolation often rely on direct pixel-level comparison between adjacent images in time or space dimensions. This approach tends to produce inaccurate motion estimation when dealing with images that have complex movements or structural changes. This is particularly disadvantageous in the field of education, where teaching content often requires precise and coherent visual representation so that students can clearly understand and master knowledge. Therefore, to provide a more stable and coherent visual experience in continuous image sequences, this paper proposes an inter-layer information fusion module, as shown in Figure 2. This module specifically processes the inter-layer structural information and change patterns output by the interleaved residual generation network that generates the intermediate result U_{MID}. The core of the module is composed of *ConvGRU* units, combining the capabilities of Convolutional Neural Networks (CNNs) in extracting spatial features from images with the advantages of Gated Recurrent Units (GRUs) in processing sequential data. Such a design allows the network to better simulate continuous motion and changes in the real world when processing educational 3D images, providing a highly realistic and coherent visual environment for teaching.

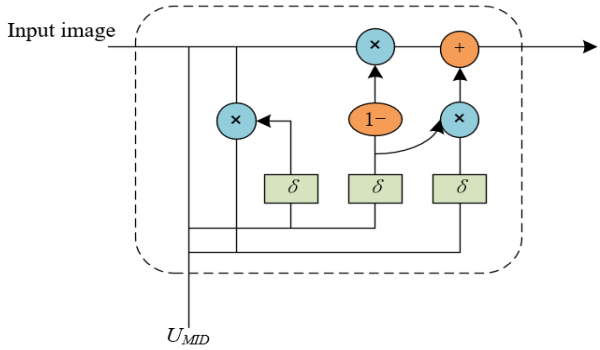

Figure 2. Structure of the inter-layer information fusion module

Assuming the input hidden state is represented by g, the activation function by δ, the current input to the *ConvGRU* unit by t_s, and the network-learned weight matrices by Q and I. The *ConvGRU* unit internally has a reset gate E and an update gate C, with the formulas as follows:

$$E = \delta(Q_E t_s + I g_{s-1}) \tag{7}$$

$$C = \delta(Q_C t_s + I g_{s-1}) \tag{8}$$

Assuming the current input to the *ConvGRU* unit is represented by U_{MID}, the previous hidden state by g_{s-1}, and the intermediate quantity that includes the current state U'_{MID} by g'. E is used to selectively fuse the inter-layer deformation features implied in the two consecutive input images U_0 and U_2 onto U_{MID}, that is, adding the information of g_{s-1} to t_s, with the calculation formula as follows:

$$g_s = (1-C) g_{s-1} + C g' \tag{9}$$

2.3 Weighted fusion loss function

Considering the roles and importance of the interleaved residual generation network and the inter-layer information fusion module during the generation process, this paper opts to introduce a weighted fusion loss function. This allows for adjusting the influence of these two modules on the final output according to different educational content and visual requirements, aiming to produce superior 3D visual effects.

Inter-layer interpolation of educational 3D images requires pixel-level generation, and both local loss functions use the L2 loss function, with the calculation formula as follows:

$$M_2(H,O) = \sum_{u=1}^{v} (H_u - O_u)^2 \tag{10}$$

Let the loss functions corresponding to the interleaved residual generation network and the inter-layer information fusion module be represented by M_Y and M_D, respectively. The number of image pixels is represented by v, the ground truth by H, and the generated image by O. Thus, the fusion loss function M is defined as:

$$M = \beta M_Y + \alpha M_D \tag{11}$$

3. VISUAL RENDERING OPTIMIZATION FOR 3D EDUCATIONAL IMAGE SCENES

In the field of education, especially in teaching intelligent facility and gerontechnology design, 3D image modeling and visual presentation technologies are indispensable tools. They enable complex intelligent systems and design concepts, such as automation controls, user interfaces, and assistive devices, to be presented to learners in an intuitive form. However, since most 3D models were initially designed to achieve general visual effects, their hierarchical structures often do not match the layer-by-layer detailed presentation required for educational purposes. This discrepancy could hinder students' learning efficiency and cognitive processes in the fields of intelligent facility design and gerontechnology applications. Therefore, researchers in the fields of gerontechnology and smart home education are working on developing visual hierarchy rendering optimization techniques that can effectively reconstruct 3D educational image scenes. The goal is to develop a universal spatial division method to reorganize the hierarchical structure of 3D models, ensuring that models are not only visually appealing but also accurately reflect the key functions and operating principles in intelligent facility design and gerontechnology. Thus, research on visual hierarchy rendering optimization for 3D educational image scenes holds significant research value. It aims to develop a universal spatial division method to restructure the hierarchical structure of 3D models, making them more aligned with teaching needs.

3.1 Visual perception selection criteria

In the educational domain, the current application of 3D image modeling and visual presentation indicates that learners typically require visual materials to acquire and understand complex concepts and processes. However, due to the limitations of human visual perception, learners' demand for details changes when observing 3D educational images from different distances and angles. At long distances, high-precision details cannot be perceived, rendering their visualization not only unnecessary but also uneconomical, potentially causing resource wastage and processing delays, affecting the smoothness of the learning experience. As the observation point gets closer, the sensitivity to detail perception increases, necessitating higher precision models to meet the demands of visual perception. Therefore, constructing a visual hierarchy perception selection criterion for 3D image scenes that adapts to learners' visual perception can effectively balance resource use and visual quality, enhancing educational outcomes.

This paper starts from simulating the mechanism of human visual perception. Firstly, it introduces a dynamic detail level switching mechanism based on viewing distance, i.e., automatically adjusting the detail level of the model based on the distance between the user and the model; more distant objects use simplified models, while closer objects use high-precision models. Secondly, the influence of viewing angle is considered, with algorithms determining the user's viewing angle to provide more details when the user is directly facing the model, and reducing detail rendering when observing from an edge view. Finally, the size of the model is also a consideration factor, with larger models displaying more details to utilize their space in the visual field. Integrating these factors, a rendering system can be designed that is highly adaptive, dynamically optimizing detail levels based on the observation behavior of the user and the spatial characteristics of the model, thus achieving efficient 3D image rendering without sacrificing visual quality, and enhancing educational effectiveness.

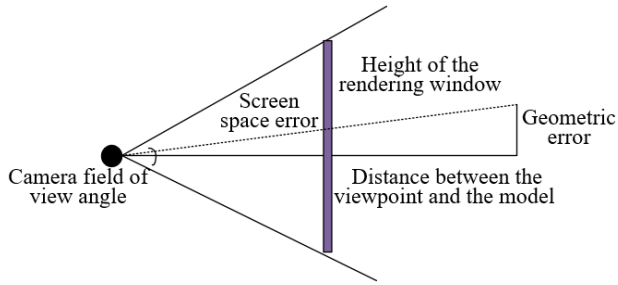

Figure 3. Principle of screen space error calculation

Screen space error in 3D image modeling and visual presentation for education is a key concept, measuring the deviation between the pixel size of a 3D object's projection on the screen and its expected pixel size. The calculation principle is shown in Figure 3. In the educational domain, optimizing this error is crucial because excessive error may lead to learners' misunderstanding of the teaching content. For instance, in biology or geography education, learners frequently observe different parts of a model to understand complex structures or topographic features. Optimizing screen space error ensures the correct display of model details at different observation distances, allowing learners to obtain accurate visual information even when zooming or rotating the model. Geometric error, on the other hand, describes the error that arises when using discrete methods like triangular meshes to approximate continuous geometric surfaces. In educational applications, controlling geometric error is vital for ensuring that learners can correctly understand the spatial and structural properties of 3D models. For example, in engineering or architectural education, precise geometric representation can help students better understand key concepts such as force distribution and material properties. Geometric error is typically quantified by the maximum distance deviation between each triangle mesh of the model and the real surface it represents. In 3D educational image scenes, controlling geometric error means selectively simplifying model complexity while maintaining model visual recognizability and the accuracy of teaching content, thus optimizing rendering performance and enhancing interaction response speed, enhancing educational effectiveness.

In Level of Detail (LOD) algorithms, assuming the camera's field of view angle is represented by ϕ, the screen space error by r_t, the geometric error by r_h, the height of the rendering window by g, and the distance between the viewpoint and the model by f, then the calculation formula is:

$$\frac{r_t}{r_h} = \frac{g / \left(2TAN \frac{\varphi}{2} \right)}{f} \tag{12}$$

Transforming the above equation yields:

$$r_t = \frac{r_h g}{2 f TAN \frac{\varphi}{2}} \tag{13}$$

In 3D image modeling and visual presentation for education, the purpose of using LOD technology is to optimize the rendering process, enhancing the rendering efficiency and visual quality of models at different observation distances. To achieve this goal, allowing users to customize visual perception selection criteria based on educational needs before generating the hierarchical structure is crucial, as illustrated in Figure 4 for the principle of quantifying visual selection criteria. This is because different educational scenarios may require different levels of visual precision. Therefore, providing a visual slider control for users to set the rendering precision of models under specific teaching content and visual importance autonomously can make the 3D models more aligned with educational objectives while also being more resource-efficient.

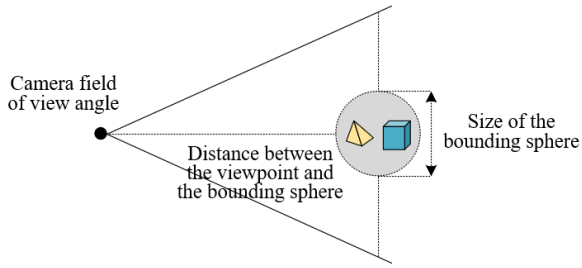

Figure 4. Principle of quantifying visual selection criteria

In the user interface design, the slider should be clear, intuitive, and easy to manipulate, allowing users to select a percentage value by dragging, which will determine the application threshold of LOD technology. In terms of interaction logic, the control is connected to the rendering engine's LOD management system, dynamically adjusting the behavior of the rendering engine based on the user-set percentage (e.g., a preset of 60%). During program execution, this percentage value will be used to determine which level of the model to render. If the user sets a higher threshold, the system will tend to render high-precision models; if the threshold is lower, it will render lower-precision models or switch to rendering a higher-level parent node, thus optimizing the use of memory and processor resources while ensuring visual quality. Assuming the size of the bounding sphere in the world coordinate system is represented by t, and the distance between the viewpoint and the bounding sphere by f. The camera's field of view angle is represented by ϕ. The value of relative height is represented by e_g. The calculation formula for the above visual selection criteria is given by:

$$e_g = \frac{t}{2fTAN\frac{\varphi}{2}} \tag{14}$$

Assuming the height of the rendering window is represented by g, and the geometric error by r_h. Combining the two formulas yields:

$$e_g = \frac{r_t}{r_h}\frac{t}{g} \tag{15}$$

Clearly, since g is fixed and the size of t for the same object is also fixed, with a given r_h, e is directly proportional to r_t.

3.2 LOD method

Educational scenarios often involve complex 3D models and dynamic scene changes. Axis-Aligned Bounding Box (AABB) provides a predictable and consistent method for handling the spatial relationships of 3D objects, making it an ideal basis for constructing a LOD hierarchy. In educational applications, it ensures that only those details important to the current visual teaching objectives are rendered, while less important details are appropriately simplified or culled. This paper opts for an AABB-based LOD method for visual hierarchy rendering optimization. However, when using AABB for hierarchical merging of 3D meshes, disproportionally scaled bounding boxes can emerge, leading to a waste of resources for high-precision rendering from afar, as excessive details are not necessary from an educational perspective. This issue also hinders the flexible application of models in different educational scenarios. This paper proposes introducing an AABB shape constraint factor to modify the bounding box merging cost function, primarily to address the dimension stretching issue, optimizing visual presentation and rendering performance. The introduction of a shape constraint factor ensures that the merged bounding boxes maintain a more balanced proportion, closer to the proportion sense of real-world objects, enhancing the visual accuracy of educational content and the intuitiveness of learning while considering computational efficiency. Figure 5 provides an illustration of the ideal AABB merging.

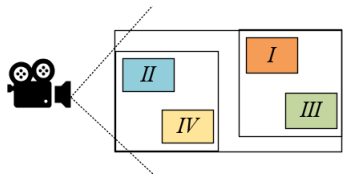

Figure 5. Ideal AABB merging illustration

The AABB shape constraint factor in this paper is defined as the difference between the maximum and minimum sizes across three dimensions plus a positive factor, aimed at quantifying the "cubicness" of a bounding box. The closer to a cube, the smaller the shape constraint factor, and the lower the merging cost accordingly. In practice, this factor will act as a weight in the calculation of merging costs, guiding the merging process to favor the generation of balanced bounding boxes. In this way, when automatically merging meshes, the algorithm tends to create bounding boxes with more uniform dimensions, thus avoiding excessive stretching in one dimension, ensuring that 3D educational images are not only more visually natural and harmonious but also technically optimized for rendering. Assuming the merging cost for generating a new AABB is represented by Z, the volume of the newly generated AABB by N, the proportion rate of the newly generated AABB by O, the dimensions of the newly generated AABB on three axes by SI_a, SI_b, SI_c, and the positive factor by γ, the following formula provides the calculation for the AABB merging cost after introducing the AABB shape constraint factor:

$$Z = \frac{N}{O}\left[MAX\left(SI_a, SI_b, SI_c\right) - MIN\left(SI_a, SI_b, SI_c\right) + \gamma\right] \tag{16}$$

4. EXPERIMENTAL RESULTS AND ANALYSIS

An analysis of the comparison of inter-layer information fusion results under different weights as shown in Figure 6 indicates that using 3D educational image modeling technology based on inter-layer feature learning, in conjunction with the inter-layer information module network and interleaved residual generation network, can effectively enhance image quality. From the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) metrics, it is observed that as the first weight in the weight arrangement increases from 0.1 to 0.5, PSNR remains relatively stable while SSIM slightly decreases. This suggests that the overall brightness and contrast of the image have been optimized to some extent, but there is a slight loss in structural fidelity. As the first weight in the weight arrangement increases from 0.6 to 0.9, both PSNR and SSIM show a downward trend, indicating that within this weight range, the model's ability to retain image details weakens, possibly due to an overemphasis on the first feature. A comprehensive analysis of experimental data shows that when the weights of the two features are relatively balanced (e.g., a weight arrangement of (0.5, 0.5)), the model can better balance overall brightness/contrast and structural fidelity, resulting in a more balanced outcome. However, when the weight is biased towards the features of the inter-layer information module network, the image's detail representation capability declines. This demonstrates that the weighted fusion loss function proposed in this paper can better adapt to the complexity and diversity of educational content under specific weight distributions, especially when the weights of the two features are relatively balanced. Therefore, the 3D educational image modeling method based on inter-layer feature learning proposed in this paper, by adjusting the weight distribution appropriately, can effectively enhance the quality of 3D educational images, especially in achieving a good balance between maintaining image structural fidelity and overall brightness/contrast.

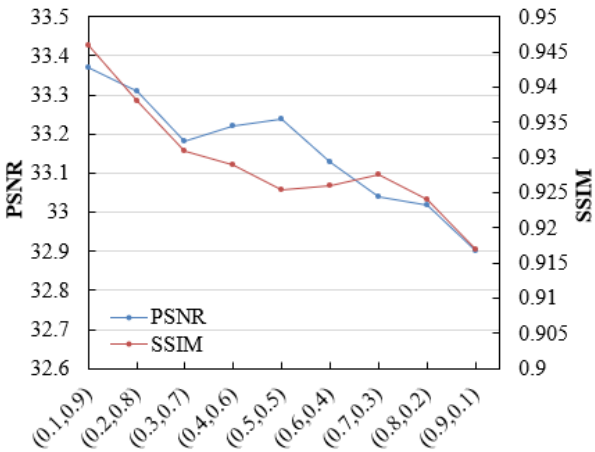

Figure 6. Comparison of inter-layer information fusion results under different weights

Based on the ablation study comparison results presented in Table 1, the effectiveness of the proposed 3D educational image modeling method can be comprehensively assessed. The model discussed in this paper shows the best performance across three key evaluation metrics: PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index), and RMSE (Root Mean Square Error), achieving 32.5614, 0.9154, and 0.0465, respectively. By sequentially removing key components for ablation experiments, the impact of each component on model performance can be observed. After removing the inter-layer information module network, the model's PSNR and SSIM decreased, and RMSE increased, indicating the critical role of the inter-layer information module network in enhancing image quality. Further removal of the interleaved residual generation network led to a further decline in all metrics' performance, especially in PSNR and RMSE, emphasizing the importance of this network in improving image details and reducing reconstruction errors. Finally, when the weighted fusion loss function is not used, the model's performance drops again, particularly in the SSIM metric, highlighting the loss function's crucial role in optimizing model structure fidelity. Integrating these experimental results, it can be concluded that the 3D educational image modeling method based on inter-layer feature learning proposed in this paper achieved excellent results across the three main performance metrics. This achievement is attributed to the collaborative effect of the inter-layer information module network, interleaved residual generation network, and weighted fusion loss function. The inter-layer information module network optimized feature extraction and fusion, the interleaved residual generation network enhanced the model's detail capturing ability, and the weighted fusion loss function balanced the importance of different features during model training, making the model better adapted to the complexity and diversity of educational content. Therefore, the integration of these components significantly improved the modeling quality of 3D educational images, validating the effectiveness of the proposed method.

Table 1. Ablation study comparison results of 3D educational image modeling methods

Methods	PSNR	SSIM	RMSE
The Proposed Model	32.5614	0.9154	0.0465
- Inter-Layer Information Module Network	31.2569	0.9124	0.0524
- Interleaved Residual Generation Network	30.2114	0.9034	0.0647
- Weighted Fusion Loss Function	30.5477	0.8947	0.0771

Table 2. Performance comparison results of different 3D educational image modeling methods

Methods	PSNR	SSIM	RMSE
VoxNet	31.2564	0.9231	0.04562
ProGANs	31.5689	0.9124	0.0578
DAEs	31.2487	0.9265	0.0568
ConvLSTM	28.9654	0.9236	0.0625
PointNet++	32.6589	0.9214	0.0421
DGCNN	31.2548	0.9256	0.0425
The Proposed Model	32.6985	0.9365	0.0412

From the performance comparison results of 3D educational image modeling methods in Table 2, the proposed model outperforms other comparative methods in three key performance indicators: PSNR, SSIM, and root mean square error (RMSE). Compared with current popular 3D image modeling methods such as VoxNet, ProGANs, DAEs, ConvLSTM, PointNet++, and DGCNN, the proposed model achieves the highest PSNR at 32.6985, reaches 0.9365 in SSIM, and lowers RMSE to 0.0412. Especially when compared with DGCNN, which has the closest SSIM value, the proposed model shows significant improvement in both PSNR and RMSE. These results demonstrate that the proposed

model performs excellently in terms of reconstruction quality, image structural similarity, and error minimization. It reflects the effectiveness of inter-layer feature learning in 3D image modeling and highlights the important contribution of the introduced inter-layer information module network and interleaved residual generation network in enhancing modeling performance.

Table 3. Data statistics for a basic classroom environment scene

Configuration	Texture Size	Vertex Count	Mesh Count	Space Occupancy
Default	1024×1024	0.81*M*	936	51.26*MB*
Bounding Sphere	1024×1024	1.69*M*	1356	158.2*MB*
AABB	1024×1024	1.56*M*	1235	145.8*MB*

Table 4. Data statistics for a complex laboratory environment scene

Configuration	Texture Size	Vertex Count	Mesh Count	Space Occupancy
Default	2048×2048	1.87*M*	1569	215.2*MB*
Bounding Sphere	2048×2048	3.18*M*	3157	478.2*MB*
AABB	2048×2048	2.89*M*	2568	458*MB*

Table 5. Data statistics for a highly detailed campus scene

Configuration	Texture Size	Vertex Count	Mesh Count	Space Occupancy
Default	4096×4096	14.89*M*	6125	1.1*GB*
Bounding Sphere	4096×4096	22.36*M*	9858	2.4*GB*
AABB	4096×4096	22.78*M*	9125	2.3*GB*

Based on the data statistics from Tables 3-5, it can be observed that the LOD technology plays a key role in the visual rendering optimization of 3D educational image scenes. In the comparison among basic classroom, complex laboratory, and highly detailed campus scenes, different configurations show varying degrees of improvement in texture size, vertex count, mesh count, and space occupancy. This improvement reflects the effectiveness of LOD technology in increasing scene details and enhancing rendering quality. Especially in the highly detailed campus scene, when using a 4096x4096 texture size, compared to the default configuration, the bounding sphere and AABB configurations show significant growth in vertex and mesh counts, with space occupancy increasing from 1.1GB to 2.4GB and 2.3GB, respectively. This indicates that ensuring higher visual quality correspondingly requires more computing resources and storage space to process these details.

From the perspective of visual rendering optimization strategies, the AABB configuration demonstrates higher data efficiency across all scenes compared to the bounding sphere configuration. AABB achieves a more compact hierarchy, reducing the number of nodes in the hierarchy tree, thus decreasing the mesh and vertex counts and correspondingly reducing space occupancy. Although the AABB and bounding sphere configurations are very close in rendering performance, the optimized AABB configuration's savings on disk space occupancy highlight its superiority in rendering optimization strategies. Therefore, it can be concluded that for efficient visual rendering while maintaining a high-quality educational interactive experience, adopting the AABB configuration of LOD technology is an optimized strategy that balances resource management and visual effects. This strategy can effectively reduce the demand for storage resources without sacrificing rendering quality, making it an ideal choice for rendering 3D educational image scenes.

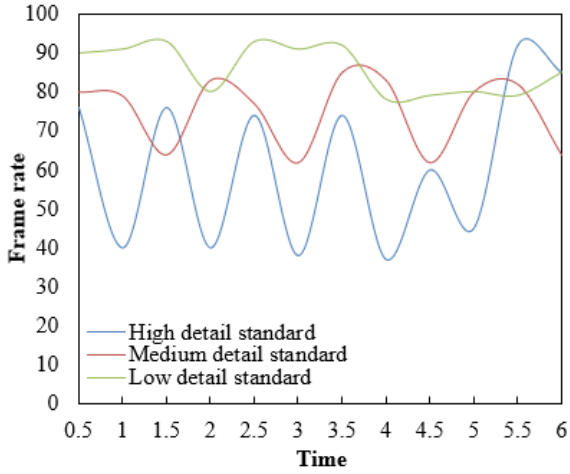

Figure 7. Visual rendering curve of 3D educational image scenes at different levels of detail

The experimental data shown in Figure 7 indicates significant differences in the rendering frame rates of 3D educational image scenes under various detail standards. Under high detail standards, the fluctuation in rendering frame rates is more significant due to all meshes being loaded into memory, leading to high memory usage and affecting rendering efficiency. In this case, the instability of rendering frame rates increases with scene roaming and model precision switching, reflecting the system's performance bottleneck when handling large volumes of data. Under medium detail standards, memory usage is controlled, only loading the meshes that need to be rendered and unloading them when no longer needed. This strategy not only reduces the memory burden but also improves the stability of rendering frame rates. At low detail standards, the highest rendering frame rates and stability are achieved, thanks to the application of a preloading streaming mechanism. This mechanism prepares for upcoming model switches by loading all child node meshes of the current hierarchy level, effectively reducing rendering wait times and optimizing frame rate performance.

The comprehensive experimental results conclude that for rendering optimization in 3D educational image scenes, the appropriate LOD technology should be chosen based on the specific needs of the scene and resource limitations. High detail standards, although capable of providing richer visual effects, demand higher hardware resources, which may affect rendering efficiency and system stability. Conversely, low detail standards, through effective resource management and preloading mechanisms, can maintain high frame rates while also providing stable rendering performance, suitable for situations with limited resources.

5. CONCLUSION

This paper focuses on two aspects of 3D educational images: modeling and rendering, primarily optimizing the precision and adaptability of 3D image modeling through inter-layer

feature learning technology. This method, utilizing a deep learning framework, enhances the extraction and representation of model features, making them better match the needs of educational content. Secondly, the paper delves into visual rendering optimization strategies, improving rendering efficiency through algorithmic enhancements to achieve smoother and more realistic visual experiences. These optimization measures help to enhance the interactivity of the teaching process and improve the learner's experience.

Integrating the research content and experimental results of the entire paper, the 3D image modeling technology based on inter-layer feature learning shows excellent performance in improving the complexity and diversity adaptability of educational images. Moreover, by introducing LOD technology, rendering efficiency is significantly enhanced without sacrificing visual quality, especially under low detail visual perception standards, achieving high frame rates and stability in rendering output. This provides an effective pathway for real-time interaction and high-quality visual presentation of 3D educational images.

Despite this, the research also has limitations, such as the high demand for hardware resources during high-detail rendering, which may limit its application on lower-end devices. Future research could focus on further reducing the resource requirements for high-quality 3D rendering or developing more efficient algorithms to lessen the dependence on high-end hardware. Moreover, integrating the technologies from this research with emerging educational technologies, such as augmented reality or virtual reality, could open new chapters in the application of 3D educational images, offering more immersive and interactive learning experiences for future education.

FUNDING

This study was supported by the 2023 National Social Science Foundation's Art Program (Grant No.: 23BG111).

REFERENCES

[1] Abu-Haifa, M., Lee, S.J. (2023). Image-based 3D modeling-to-simulation of single-wythe masonry structure via reverse descriptive geometry. Journal of Building Engineering, 76: 107125. https://doi.org/10.1016/j.jobe.2023.107125

[2] Guo, T., Dong, K. (2024). Research on 3D geometric modeling of urban buildings based on airborne lidar point cloud and image. In Fourth International Conference on Geology, Mapping, and Remote Sensing (ICGMRS 2023), Wuhan, China, pp. 623-635. https://doi.org/10.1117/12.3019457

[3] Chen, Z., Agarwal, D., Aggarwal, K., Safta, W., Balan, M.M., Brown, K. (2023). Masked image modeling advances 3D medical image analysis. In 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, pp. 1970-1980. https://doi.org/10.1109/WACV56688.2023.00201

[4] Liu, C., Cheng, Y., Tamura, S. (2023). Masked image modeling-based boundary reconstruction for 3D medical image segmentation. Computers in Biology and Medicine, 166: 107526. https://doi.org/10.1016/j.compbiomed.2023.107526

[5] Hong, D., Lee, S., Kim, T., Baek, J.H., Lee, Y.M., Chung, K.W., Sung, T.Y., Kim, N. (2019). Development of a personalized and realistic educational thyroid cancer phantom based on CT images: An evaluation of accuracy between three different 3D printers. Computers in Biology and Medicine, 113: 103393. https://doi.org/10.1016/j.compbiomed.2019.103393

[6] Di Tore, S., Todino, M.D., Campitiello, L. (2021). Lab-H: A laboratory to develop 3D printable inclusive open educational resources. In International Workshop on Higher Education Learning Methodologies and Technologies Online, Pisa, Italy, pp. 233-247. https://doi.org/10.1007/978-3-030-96060-5_17

[7] Liu, C., Cao, L. (2023). Automatic detection of sports injuries based on multimedia intelligent 3D images. Advances in Multimedia, 2023: 4180887. https://doi.org/10.1155/2023/4180887

[8] Zhang, S. (2021). Research on operation characteristics of UHV converter valve hall based on intelligent image processing and 3D modeling technology. Journal of Physics: Conference Series, 1871(1): 012135. https://doi.org/10.1088/1742-6596/1871/1/012135

[9] El Hazzat, S., El Akkad, N., Merras, M., Saaidi, A., Satori, K. (2020). Fast 3D reconstruction and modeling method based on the good choice of image pairs for modified match propagation. Multimedia Tools and Applications, 79: 7159-7173. https://doi.org/10.1007/s11042-019-08379-2

[10] Wang, M., Li, M. (2021). Research on 3D digital modeling and virtual simulation technology of ancient architecture based on image sequence. In 2021 International Conference on Aviation Safety and Information Technology, Changsha, China, pp. 651-655. https://doi.org/10.1145/3510858.3511351

[11] Han, J., Shen, S. (2019). Scalable point cloud meshing for image-based large-scale 3D modeling. Visual Computing for Industry, Biomedicine, and Art, 2(1): 10. https://doi.org/10.1186/s42492-019-0020-y

[12] Chen, P., Chen, Y., Yang, D., Wu, F., Li, Q., Xia, Q., Tan, Y. (2021). I2uv-handnet: Image-to-uv prediction network for accurate and high-fidelity 3D hand mesh modeling. In 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, pp. 12929-12938. https://doi.org/10.1109/ICCV48922.2021.01269

[13] Wang, X. (2021). Application of network protocol improvement and image content search in mathematical calculus 3D modeling video analysis. Alexandria Engineering Journal, 60(5): 4473-4482. https://doi.org/10.1016/j.aej.2021.02.030

[14] Zheng, J., Liu, Q. (2021). Design of 3D scene visual communication modeling based on virtual reality graphics rendering framework. Journal of Physics: Conference Series, 1982(1): 012183. https://doi.org/10.1088/1742-6596/1982/1/012183

[15] Peresunko, P., Mamatin, D., Antamoshkin, O., Peresunko, E., Nikitin, A. (2021). Models of experts for shaders estimation of rendering complex 3D scenes in real time. In 2021 3rd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA), Lipetsk, Russian Federation, pp. 895-897. https://doi.org/10.1109/SUMMA53307.2021.9632071

[16] Lu, L., Ma, J., Qu, S. (2020). Value of virtual reality

technology in image inspection and 3D geometric modeling. IEEE Access, 8: 139070-139083. https://doi.org/10.1109/ACCESS.2020.3012207

[17] Yoo, J H., Chae, D., Park, J.H., Ko, J.H. (2019). Effective 3D modeling method using indirect information of targets for SAR image prediction. Target and Background Signatures V, 11158: 55-60. https://doi.org/10.1117/12.2532483

[18] Ying, H., Yu, M., Jiang, G., Peng, Z., Chen, F. (2020). Perceived depth quality-preserving visual comfort improvement method for stereoscopic 3D images. Signal Processing, 169: 107374. https://doi.org/10.1016/j.sigpro.2019.107374

[19] Zhang, Y., Zhu, X. (2021). Visual nondestructive rendering of 3D animation images based on large data. In Advanced Hybrid Information Processing: 4th EAI International Conference, ADHIP 2020, Binzhou, China, pp. 409-420. https://doi.org/10.1007/978-3-030-67874-6_38

[20] Whang, J., Polys, N.F. (2020). DeepCinema: Adding depth with X3D image-based rendering. In Proceedings of the 25th International Conference on 3D Web Technology, Virtual Event, Republic of Korea, pp. 9. https://doi.org/10.1145/3424616.3424713