Journal of Engineering, Project, and Production Management
2026, 16(3), 2025-220

Virtual Reconstruction and Interactive Design of Prehistoric Creatures Using 3D Modeling and VR Technology

Yuexin Tang
Instructor and Deputy Head, Department of Journalism and Communication, College of Communication Science and Art, Chengdu University of Technology, Chengdu, 610059, China, E-mail: yuexintangyx@outlook.com

Project Management
Received October 11, 2025; revised December 8, 2025; accepted December 29, 2025
Available online April 8, 2026

Abstract: The reconstruction of prehistoric creatures plays an important role in scientific research, public education, and cultural entertainment. However, traditional reconstruction methods often involve long production cycles, high costs, and limited flexibility for model modifications. This study proposes a virtual reconstruction model for prehistoric creatures by integrating 3D modeling and Virtual Reality (VR) technology. Based on this, the study further develops an interactive reconstruction model using separable convolution and bidirectional long short-term memory networks. Experimental results show that the proposed virtual reconstruction model achieves a reconstruction accuracy of over 94% on two datasets when the number of iterations reaches 35. Additionally, when the reconstruction time is set to 10 seconds, the structural similarity index exceeds 0.94. Furthermore, the model achieves a gesture recognition accuracy above 0.93 for five different gestures, significantly outperforming the baseline models. And the model also achieved a maximum interaction rate of 70.62% and a highest re-experience rate of 72.85%. These results demonstrate that the proposed virtual reconstruction and interaction model enables accurate reconstruction of prehistoric creatures and ensures the precise execution of commands within the virtual reality system. It also enhances the immersive experience and promotes the development of public education on prehistoric life.

Keywords: 3D modeling, virtual reality technology, prehistoric creatures, virtual reconstruction, interactive design.

DOI 10.32738/JEPPM-2025-220

1. Introduction

Prehistoric creatures serve as living records of Earth's history. Their existence and extinction not only shaped modern biodiversity but also offers humans a glimpse into an evolutionary journey of life through fossils that span hundreds of millions of years (Yan and Xin, 2022). Reconstructing prehistoric creatures plays an important role in uncovering the mysteries of evolution and understanding ecological systems, making it highly valuable for public education (Hou et al., 2024). Traditional reconstruction methods rely heavily on the completeness of fossils and the subjective interpretations of researchers. These methods often suffer from long development time, limited model flexibility, and high replication costs (Yu et al., 2024). Three-Dimensional (3D) is widely applied across many fields, from industrial manufacturing to cultural entertainment, and serves as a fundamental technology in the digital age (Wahba et al., 2024). Virtual reality (VR) technology enables immersive interaction through real-time rendering and motion tracking. It is broadly applied in gaming, healthcare, and educational training (Hsiang et al., 2022). The framework of Convolutional Neural Network (CNN) combined with Long Short-Term Memory (LSTM) can extract spatial features through convolutional layers and pooling layers and capture temporal relationships using Recurrent Neural Networks. This makes it suitable for tasks that involve both spatial and temporal information (Gheisari et al., 2023). Based on the above, this study proposes a virtual reconstruction model for prehistoric creatures that integrates 3D modeling and VR technology. Therefore, the research proposes a virtual image restoration model for ancient creatures that combines 3D modeling, VR technology, and designs a gesture recognition algorithm for VR interaction based on CNN and LSTM. The innovation of this research lies in the combination of 3D point clouds and deep learning, which breaks through the traditional modeling's reliance on complete fossils. On this basis, the study employs deep separable convolution to optimize model parameters. It combines LSTM to capture the bidirectional temporal features of gestures, thereby addressing the issues of poor generalization and high latency in traditional interaction systems. This approach provides reusable technology for the cross-fusion of digital paleontology and popular science education.

2. Related Works

3D modeling overcomes the limitations of two-dimensional planes by creating stereoscopic digital models that can present object details in multiple dimensions. It has been widely applied in production, daily life, and scientific research. For example, Ge et al. proposed a 3D reconstruction method based on deep supervised neural radiance fields for the restoration of ancient architecture in cultural heritage protection. Their experimental results showed that the model could express important details more effectively than traditional modeling (Ghosh et al., 2022). Abdulredah and Al-Jawad (2024) addressed the interference of carbonate reservoirs in oilfield development by constructing a 3D geological model of the reservoir using a non-uniform network. Test results demonstrated the model's ability to accurately measure water saturation and porosity in the oil field, leading to a re-evaluation of the reservoir's capacity in Garraf. VR technology, with its immersive experience, breaks through limitations of time and space and plays an important role in entertainment, education, healthcare, and industry. Many scholars have conducted research on VR. Wulandari et al. (2025) proposed a VR-based interactive multimedia educational method to improve language skills in young children. Experimental results indicated that their method achieved higher average scores in language ability tests compared to traditional picture-based education, effectively enhancing children's language skills. Lai and Chen (2023) investigated an English vocabulary learning approach that combines VR technology with visual novel games. Their experiments showed that the test scores for this method were significantly higher than those of PC-based learning.

The network model combining CNN and LSTM (CNN-LSTM) possesses the spatial feature extraction advantages of CNN and the time-series dependency processing capabilities of the long short-term memory network. It has been widely applied in fields such as video analysis and speech recognition. Hou et al. (2022) addressed the issue that traditional temperature prediction models struggle to make precise predictions due to nonlinear relationships and proposed a temperature prediction model based on the CNN-LSTM network. Its average absolute error in the meteorological dataset can reach 0.82, which is significantly better than the comparison models and highly consistent with the measured values. Virtual image reconstruction uses computer graphics, 3D modeling, and other technologies to repair and enhance blurred or damaged images, reproducing their original states. It has been widely applied in the preservation of historical and cultural heritage. Bartolini and Rook (2023) proposed a paleontological reconstruction method combining computed tomography scanning and 3D printing for specimens in the Florence and Faenza museums. Their results showed excellent 3D mesh quality and fidelity of the digital models. Sun et al. (2023) developed an image restoration model based on adaptive curve fitting for damaged silk artifacts, aiming to maintain structural continuity and restore the original patterns. Experimental results demonstrated that the model attained a mean structural similarity index of 0.977, along with an average peak signal-to-noise ratio reaching 39.16. Nemoto et al. (2023) and colleagues introduced a virtual restoration approach for wooden ships with multiple variable components, utilizing non-rigid 3D shape assembly to reconstruct the original structure. Experiments verified the feasibility of the method, and the restored physical models were used for museum exhibitions. Gesture recognition enables users to control virtual scenes through natural hand gestures, allowing them to interact with virtual objects, greatly enhancing immersion and ease of operation. This has garnered the attention of many scholars. Miah et al. (2023) proposed a gesture recognition model based on a multi-head attention-optimized graph CNN to address inefficiency and poor generalization of traditional methods. Experiments showed their model achieved over 92% recognition accuracy across three different datasets. Alonazi et al. (2023) addressed gesture recognition in complex environments by proposing a CNN and a deep belief network-based gesture tracking and recognition model. Tests demonstrated an overall accuracy of 90.73% and 89.33% on two datasets.

In summary, scholars have made considerable progress in virtual image reconstruction and gesture recognition. However, traditional virtual image reconstruction requires high fossil completeness and incurs high costs. Traditional VR gesture recognition methods face challenges, including low recognition efficiency and poor generalization. Therefore, the research proposes a virtual image restoration model for ancient creatures that combines 3D point clouds, VR technology, and designs a gesture recognition model based on CNN and LSTM for interactive restoration of the images. The contribution of this research lies in proposing a fragmented fossil intelligent completion mechanism and a low-latency interaction method, which breaks through the reliance of traditional 3D modeling on complete fossils and the performance bottleneck of VR interaction systems, providing a reusable technical framework for digital paleontology. At the same time, this VR system, through immersive gesture interaction in museum scenarios, transforms the ancient practice of biological popularization from passive viewing to active exploration, achieving the democratization and personalization of educational outreach, which is conducive to museums fulfilling their core mission of disseminating science and preserving cultural heritage. Moreover, the proposed system can serve as the core content of the museum's distinctive interactive exhibition area, significantly enhancing the exhibition's appeal and driving the consumption of cultural products. It reduces the cost of fossil replication and model maintenance in traditional physical exhibitions, providing new operational paths for museums and promoting the transformation and upgrading of cultural heritage display and dissemination models.

3. Construction of Virtual Reconstruction and Gesture Interaction Models based on 3D Modeling and VR Technology

3.1. Virtual Image Reconstruction Model for Prehistoric Creatures

The common methods for collecting three-dimensional modeling data of ancient organisms have long subsequent data processing procedures, and the cost of scanning equipment is also high. However, this process is time-consuming due to scanning and data processing, and the equipment costs are high. In contrast, 3D point cloud technology collects the spatial coordinates of an object's surface through laser scanning, enabling more efficient and cost-effective data acquisition (Xiao

et al., 2023). The calculation equation of the coordinates is shown in Eq. (1).

$$\begin{cases} X = S\cos\beta\cos\alpha \\ Y = S\cos\beta\sin\alpha \\ Z = S\cos\beta \end{cases} \tag{1}$$

In Eq. (1), X, Y, and Z represent the horizontal coordinate, vertical coordinate and vertical coordinate of the point cloud data, respectively, and S represents the distance from the target object to the device. β and α represent the vertical and horizontal angles of the target object, respectively. After acquiring the point cloud data of the target object's surface via laser scanning, preprocessing is necessary. The preprocessing workflow is shown in Fig. 1.

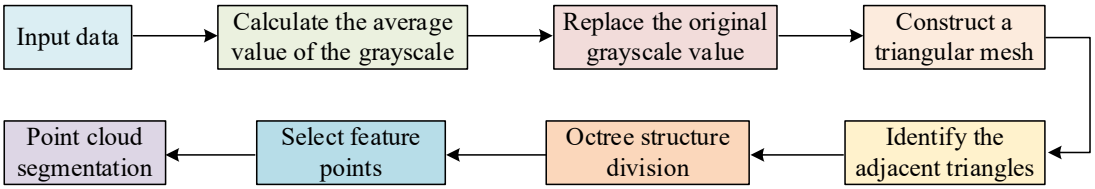

Fig. 1. Point cloud data preprocessing flowchart

In Fig. 1, the input point cloud data first undergoes bilateral filtering, which replaces the original grayscale values with the average grayscale values to eliminate noise caused by object reflection and device errors. Next, the octree algorithm simplifies the point cloud data to reduce the computational complexity of spatial queries and collision detection. Finally, adjacent points with similar feature sets are merged into regions, and the point cloud data is segmented to support the subsequent 3D modeling. The expression for bilateral filtering is shown in Eq. (2).

$$p_q' = p_q + \lambda n_q \tag{2}$$

In Eq. (2), p' represents the new data point obtained after bilateral filtering. p represents the original data point. q indicates the index of each point in the point cloud data. λ is the filtering factor, and n is the total number of data points involved in the processing. Afterward, the octree structure divides the point cloud by comparing the standard deviation of the normal vector and the capacity difference within the grid, which reduces the size of the point cloud data. The equation is shown in Eq. (3).

$$\delta_j = \sqrt{\frac{\sum_{a=1}^{m} | k_a - \overline{k}_j |^2}{m-1}} \tag{3}$$

In Eq. (3), δ is the standard deviation of the normal vectors. a indicates the data index, which is the total number of points in the grid. k represents the normal vectors of the points in the grid, and $\overline{k}$ is the mean normal vector of all points in the grid. Then, feature extraction is performed to calculate the dispersion of features. By comparing with a threshold, boundary points are identified, and segmentation is achieved. The equation is shown in Eq. (4).

$$\begin{cases} u = \dfrac{1}{l}\sum_{l=1}^{l} \arccos | P_l \cdot \overline{k} | \\ w = \sqrt{\dfrac{1}{k-1}\sum_{k=1}^{k} \arccos | l(P_l) - u |} \end{cases} \tag{4}$$

In Eq. (4), u indicates the expected value of the normal, l is the number of neighborhood points, P represents the neighborhood point set, and w is the mean normal vector of the point cloud. After preprocessing, the point cloud data is input into the VR system for 3D modeling. The modeling process is shown in Fig. 2.

As shown in Fig. 2, the preprocessed point cloud data is converted into solid Computer-Aided Design (CAD) data via vector transformation. The geometric contour of the object is extracted and modeled using 3DMax software. Texture and geometric information are combined to generate the 3D model. During reconstruction, VR imaging obtains the grayscale histogram of the object and renders contours and textures. The texture distribution is expressed in Eq. (5).

$$trace(x, y, \sigma_I) = \sigma_I (L_{xx}(x, y, \sigma_I) + L_{yy}(x, y, \sigma_I)) \tag{5}$$

In Eq. (5), $trace(x, y, \sigma_I)$ represents the texture distribution at coordinate (x, y). σ_I indicates the contour scale, and L_{xx} and L_{yy} are the cross-correlation functions.

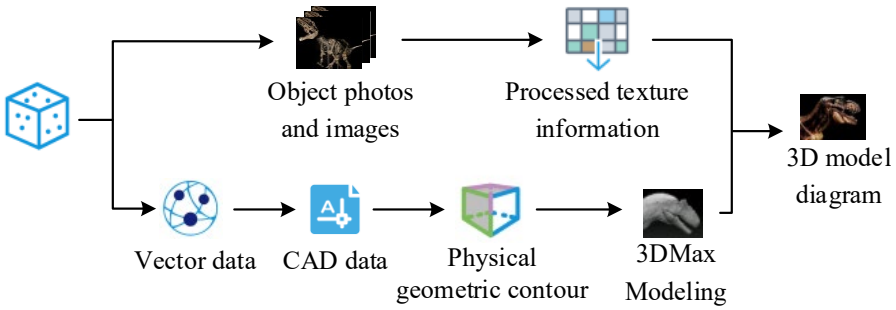

Fig. 2. Schematic diagram of the 3D modeling process of the target object (Icon source from: www.1001freedownloads.com)

Then, singular value decomposition is used to extract the edge contour features and normalize the results to obtain the degraded color feature, as shown in Eq. (6).

$$M = \sigma_D^{\,2} G(\sigma_1) \cdot \begin{bmatrix} L_x^{\,2}(x, y, \sigma_D) \\ L_y^{\,2}(x, y, \sigma_D) \end{bmatrix} \cdot (x, y, \sigma_1, \sigma_D) \tag{6}$$

In Eq. (6), M is the measure of color feature degradation, σ_D represents spatial scale, G is the decay factor, where L_x and L_y are the associated feature components. Afterward, the target color feature is matched, and the scale segmentation line of the object is obtained, as shown in Eq. (7).

$$Ncut(A, B) = \frac{assoc(A, V)}{assoc(B, V)} \tag{7}$$

In Eq. (7), $Ncut$ indicates normalized segmentation. A and B represent the two parts being segmented in the image. V is the entire image region, and $assoc(A,V)$ and $assoc(B,V)$ are the brightness values of the primary colors. Then, the feature matching matrix is obtained, as shown in Eq. (8).

$$J(x, y, \sigma) = \begin{pmatrix} \dfrac{\partial K}{\partial x} \\ \dfrac{\partial K}{\partial y} \end{pmatrix} = \begin{pmatrix} L_x(x, y, \sigma_D) \\ L_y(x, y, \sigma_D) \end{pmatrix} \tag{8}$$

In Eq. (8), J is the feature matching matrix, and K represents the data point to be reconstructed. With the feature matching matrix, the target data point can be obtained and input into the VR system to achieve virtual reconstruction of the prehistoric creature's image. The workflow of the 3D-VR virtual reconstruction model is shown in Fig. 3.

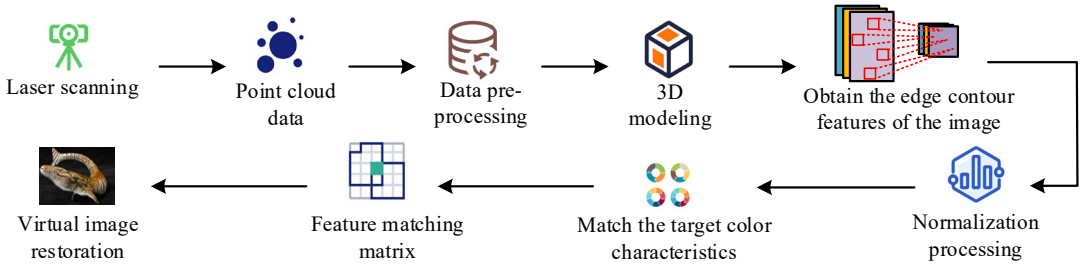

Fig. 3. 3D-VR ancient biological virtual image restoration model workflow (Icon source from: www.1001freedownloads.com)

As shown in Fig. 3, the point cloud data of prehistoric creatures are collected using laser scanning to accurately capture spatial coordinate information on the object surface. Next, the collected data is preprocessed to remove noise and redundancy, improving data quality. Then, 3D modeling converts the discrete point cloud into a structured 3D mesh model, building the framework for virtual biological forms. The image edge contour features are extracted to identify key boundary information for morphological restoration. After normalization to unify data scale and eliminate differences, color features are matched to assign reasonable coloring based on paleontological studies, enhancing visual realism. Finally, the original data is matched with the reconstruction target via the constructed feature matching matrix and input into the VR system to

complete the virtual image reconstruction.

3.2. Construction of Virtual Reconstruction and Interaction Model for Prehistoric Creatures based on the 3D-VR Model

The constructed 3D-VR virtual image reconstruction model of prehistoric creatures addresses the issues of high cost and long duration in traditional reconstruction methods. However, the reconstructed image is virtual, and conventional methods cannot support interaction with it. Gesture recognition, as a key natural interaction method in VR models, allows users to control virtual scenes and interact with virtual models through natural gestures. This greatly enhances immersion and ease of operation. In recent years, deep learning networks, such as CNN-LSTM, have performed well in handling time-series and spatial-sequence data (Wu et al., 2023). However, traditional CNN algorithms require a large number of parameters and are prone to overfitting. Therefore, this study uses Depthwise Separable Convolution (DSC) to optimize CNN and reduce its computational load. The structure of DSC is shown in Fig. 4.

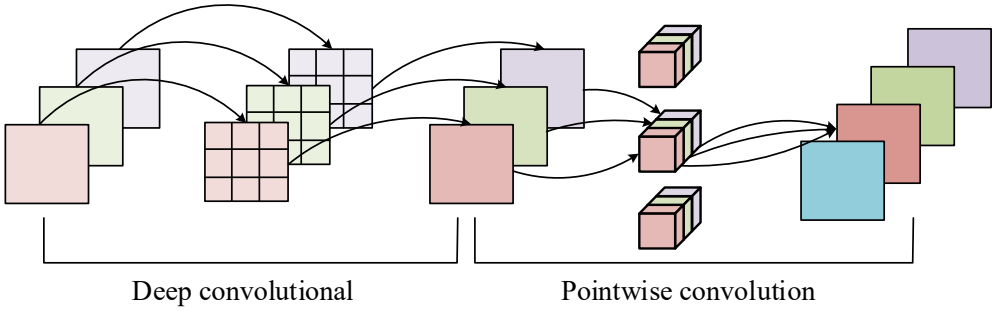

Fig. 4. Schematic diagram of the structure of DSC

Fig. 4 shows how DSC divides standard convolution into two operations: depthwise convolution and pointwise convolution. The former substantially decreases both the number of parameters and computational cost, whereas the latter maintains the feature map dimensions while performing linear combinations of depthwise outputs across channels. This enhances the model's expressive power. The total computation of DSC is shown in Eq. (9).

$$F = H \times W \times (C_{in} \times r^2 + C_{in} \times C_{out}) \tag{9}$$

In Eq. (9), F represents the total computation of DSC. H and represents the input feature map size. C_{in} and C_{out} are the input and output channels. r is the kernel size. Compared to standard convolution, DSC significantly reduces the parameter quantity and computational complexity, making the network more lightweight and suitable for mobile devices. LSTM can selectively memorize key information in a sequence through cell states and gating mechanisms. It performs well in handling time-series data (Singla et al., 2022). However, gestures often consist of multiple continuous actions with bidirectional dependencies, whereas traditional LSTMs can only process unidirectional temporal data and lack contextual awareness. Bidirectional LSTM (BiLSTM) processes forward and backward sequences simultaneously and uses bidirectional context to align key frames better (Zhang et al., 2023). Therefore, this study combines DSC and BiLSTM to build a DSC-BiLSTM gesture recognition model, as shown in Fig. 5.

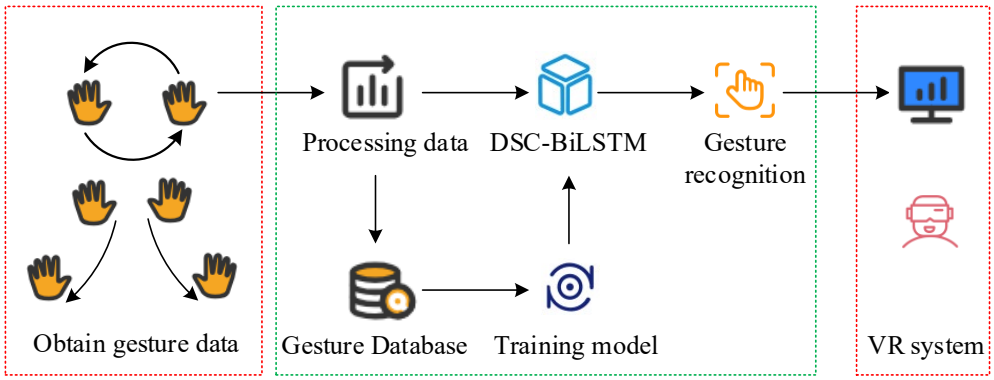

Fig. 5. DSC-BiLSTM gesture recognition model workflow (Icon source from: www.1001freedownloads.com)

As shown in Fig. 5, the system first collects multiple datasets for each type of gesture and builds a gesture database. Then, it extracts features from the sample data and trains the DSC-BiLSTM model. After this, the input gesture data is classified and recognized, and standardized control instructions in a unified format are output. The VR engine receives the instructions and parses them into specific spatial parameters and action parameters. These parameters are then displayed to the user through visual and tactile feedback provided by the VR device. BiLSTM primarily consists of four components:

the forget gate, input gate, memory cell, and output gate. The forget gate determines the retention of the previous information and blocks irrelevant inputs, ensuring that later neurons are not disrupted by unrelated data. The calculation of the forget gate is shown in Eq. (10).

$$f_t = \sigma'(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{10}$$

In Eq. (10), f_t is the forget gate output with a value range in [0, 1]. σ' represents the sigmoid activation function. W_f is the forget weight, h_{t-1} is the output at time $t-1$, x_t is the input at time t, and b_f is the bias vector. The input gate controls how much data should be updated into the memory cell. The calculation of the input gate is shown in Eq. (11).

$$i_t = \sigma'(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{11}$$

In Eq. (11), i_t is the input gate output. W_i is the input weight, and b_i is the bias vector of the input gate. The memory cell selectively stores neuron state information based on the signals from the forget and output gates. The memory cell state at the current time is calculated as shown in Eq. (12).

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c_t} \tag{12}$$

In Eq. (12), c_t is the memory cell state. c_{t-1} is the memory cell state at time $t-1$, and $\tilde{c_t}$ represents the retained information at time t, calculated as shown in Eq. (13).

$$\tilde{c_t} = \tanh*(w_c \cdot [h_{t-1}, x_t] + b_c) \tag{13}$$

In Eq. (13), b_c and w_c represent the bias vector and weight of the candidate vector, respectively. The output gate determines which parts of the memory cell should be retained and passed to the next output. The output gate calculation is shown in Eq. (14).

$$h_t = \tanh*\{\sigma(w_o[h_{t-1}, x_t] + b_o)\} \tag{14}$$

In Eq. (14), h_t is the model output at time t. w_o is the output gate weight, and b_o is the output gate bias vector. In summary, this study develops a virtual image reconstruction and interaction model for prehistoric creatures by combining a 3D-VR model and a DSC-BiLSTM model. The model is named 3D-VR-DBLSTM, and its workflow is shown in Fig. 6.

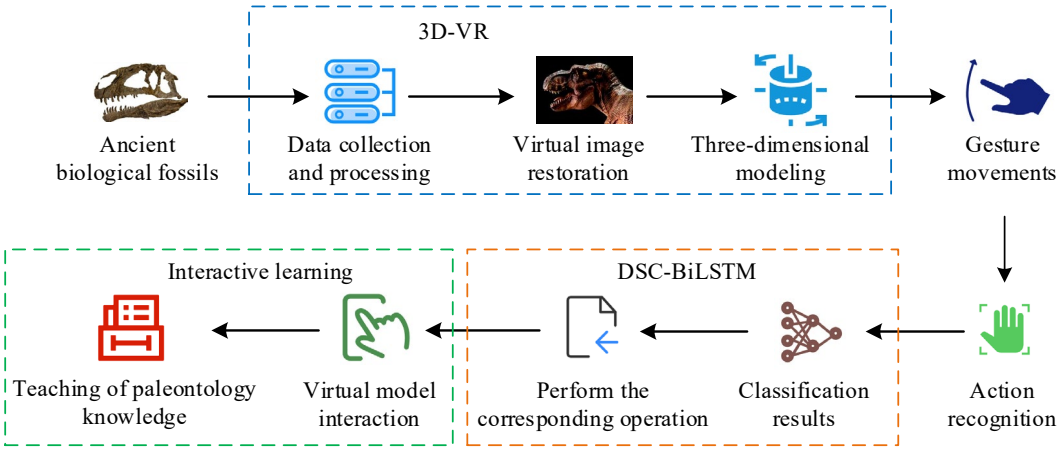

Fig. 6. 3D-VR-DBLSTM restoration and interactive model workflow (Icon source from: www.1001freedownloads.com)

As shown in Fig. 6, the model utilizes 3D-VR technology to collect and process fossil data of prehistoric creatures, providing fundamental support for subsequent virtual reconstruction. Then, the creature is virtually reconstructed and presented as a virtual image. The DSC-BiLSTM model classifies and recognizes user gestures, and corresponding instructions are executed based on the classification results. Viewers can actively adjust the observation angle, scale size and movement state of the virtual ancient creatures, completing the interaction between the operator and the virtual image, transforming passive knowledge acquisition into active exploration, thereby enhancing the participation in learning about ancient creatures.

4. Performance Validation of Virtual Image Reconstruction and Interaction Model based on 3D Modeling and VR Technology

4.1. Performance of the 3D-VR Virtual Reconstruction Model for Prehistoric Creatures

To validate the performance of the 3D-VR prehistoric creature virtual image reconstruction model, the study compared it with the 3D-Generative Adversarial Network (3D-GAN) and Convolutional Neural Network-Support Vector Machine (CNN-SVM) models. The experimental system ran on Windows 10, using the PyTorch deep learning framework and the Adam optimizer, with Python 3.10.12 as the programming language. The system had 32GB of RAM, an NVIDIA RTX3080

GPU with 10GB of memory, and an Intel Core i7-10700K@3.8GHz CPU. This study utilized high-precision CT scanning to examine ancient biological fossils from the Cambrian and Carboniferous periods, which served as the test dataset and included complete fossils of 60 ancient organisms. The samples were divided into the training set and test set in a 7:3 ratio. The experiment used the Faro Focus S70 laser scanner to collect point cloud data, with a scanning resolution of 0.5mm and a sampling frequency of 1 million points per second. The 3D-VR model optimizer was Adam (β_1=0.9, β_2=0.999), with an initial learning rate of 0.001 and a batch size of 16. Keep the training parameters of the comparison model consistent, and only replace the core network structure. To evaluate reconstruction performance, the study compared the reconstruction accuracy of the three models. The test results are shown in Fig. 7.

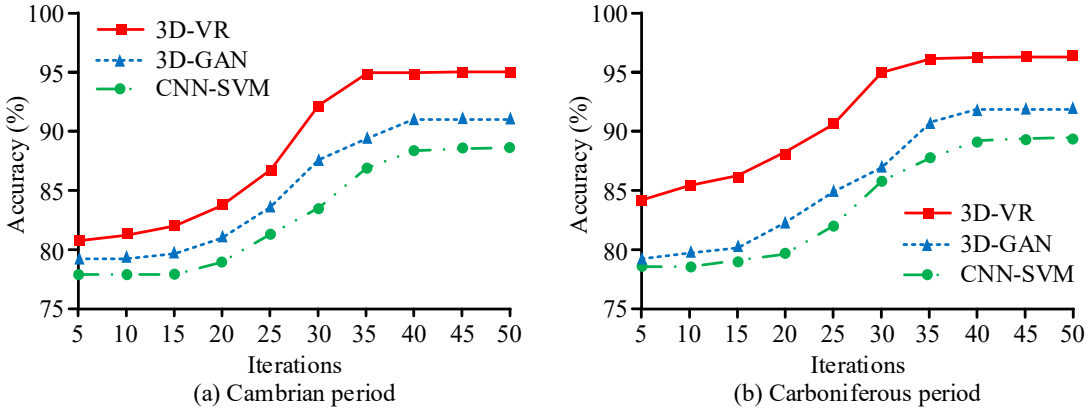

Fig. 7. Comparison of restoration accuracy results

As shown in Fig. 7(a), on the Cambrian dataset, the 3D-VR model achieved a reconstruction accuracy of 94.8% at 35 iterations. The experimental results demonstrate that the virtual images have achieved a level of similarity to real fossils in terms of appearance and verifiable anatomical structure. This enables ordinary viewers to have a more intuitive and accurate understanding of ancient creatures, avoiding the formation of incorrect perceptions due to model errors. As shown in Fig. 7(b), the 3D-VR model achieved 95.4% accuracy on the Carboniferous dataset after 45 iterations. To further assess performance, the study tested the reconstruction speed of all three models across 20 trials each. The results are shown in Fig. 8.

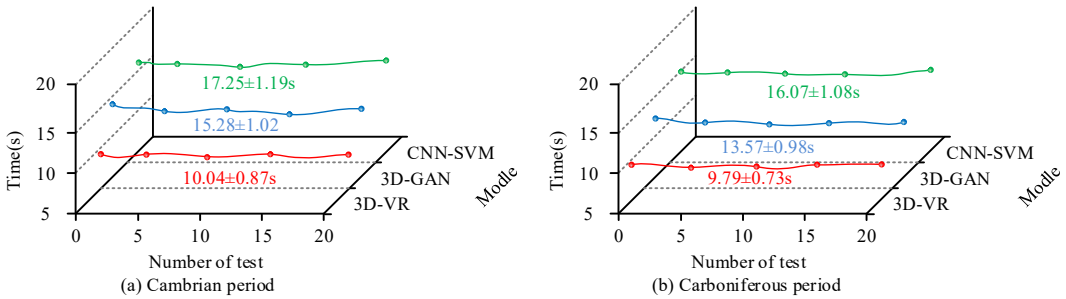

Fig. 8. Comparison of recovery speed test results

As shown in Fig. 8(a), in the Cambrian dataset, the average restoration speed of the 3D-VR model was 10.04 ± 0.87 seconds, which was 5.24 seconds faster than the CNN-SVM model. As shown in Fig. 8(b), the average restoration times for the 3D-VR and 3D-GAN models in the Carboniferous dataset were 9.79 ± 0.73 seconds and 13.57 ± 0.98 seconds, respectively. The restoration speed fluctuations of the 3D-VR model in different datasets were smaller compared to those of the comparison models, demonstrating its consistent performance. In conclusion, the 3D-VR model, which optimizes the computational load using DSC, has a faster restoration speed and avoids the long processing time caused by redundant feature extraction in traditional models. To further test reconstruction efficiency, the study compared the SSIM values over time for the three models. The results are shown in Fig. 9.

As shown in Fig. 9(a), when the reconstruction time reached 15s, the SSIM of the 3D-VR model was 0.95, while the 3D-GAN and CNN-SVM models achieved 0.86 and 0.82, respectively. As shown in Fig. 9(b), the 3D-VR model achieved an SSIM of 0.94 at 10s, which significantly outperformed baseline models. In conclusion, the 3D-VR model, which combines VR imaging technology with grayscale histogram rendering, can precisely match the color characteristics and edge contours of fossils. It can achieve high-precision virtual image restoration of ancient creatures within a short period, thereby avoiding the problem of texture stitching discontinuities, and can present the key appearance features of ancient creatures more clearly.

4.2. Performance Validation of the 3D-VR-DBLSTM Virtual Reconstruction and Interaction Model for Prehistoric Creatures

After validating the performance of the 3D-VR model, the study analyzed the performance of the 3D-VR-DBLSTM reconstruction and interaction model by comparing it with the 3D-VR-LSTM, CNN-SVM-Transformer, and 3D-GAN-ResNet models. The number of hidden units in 3D-VR-DBLSTM is 256, the dropout rate is set to 0.5, the learning rate is 0.001, and the batch size is 32. The test datasets consisted of historical data from two museums (A Museum and B Museum). The average training loss for each model was measured, and the results are shown in Fig. 10.

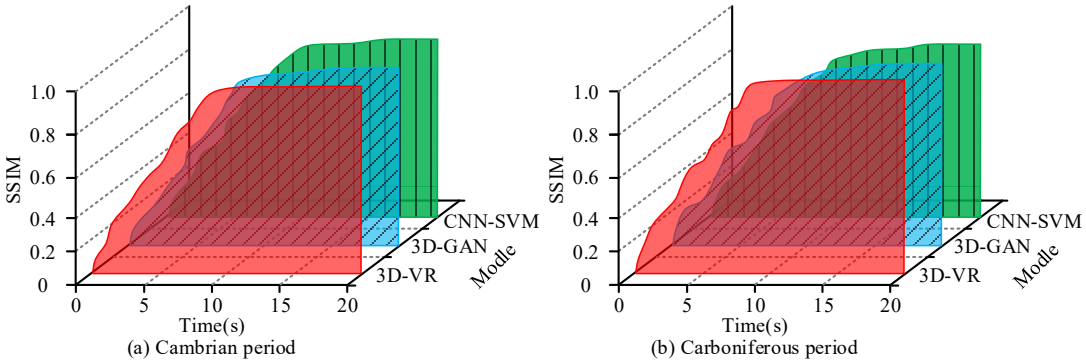

Fig. 9. Comparison of SSIM index test results

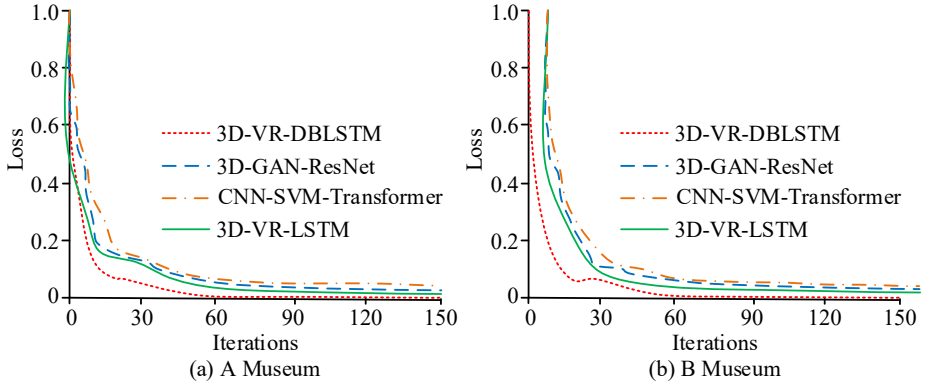

Fig. 10. Comparison of average loss training and testing curves

From Fig. 10(a), it can be seen that when the number of iterations is 90, the loss value of the 3D-VR-DBLSTM model is 0.014, while the loss values of the 3D-VR-LSTM and 3D-GAN-ResNet models are 0.022 and 0.026, respectively. From Fig. 10(b), it can be observed that when the number of iterations is 120, the loss value of the 3D-VR-DBLSTM model is 0.013, which has a significant advantage compared to the CNN-SVM-Transformer and 3D-GAN-ResNet models. In conclusion, the 3D-VR-DBLSTM model has the fastest convergence speed and the lowest loss value. It can effectively prevent the frustration caused by the failure of instructions to be responded to by the audience and enhance their willingness to interact. The accuracy of gesture recognition has a significant impact on the efficiency of interaction. To further verify the performance of the 3D-VR-DBLSTM model, the study selected five core interaction gestures (forward, backward, amplify, minification, and rotate), covering all the core interaction requirements of the virtual ancient creature images. Forward and backward enable adjustment of the observation distance, amplification and magnification meet the needs of detailed viewing and overall overview. Rotation enables 360° perspective switching. The five gestures together form a complete interaction system of distance, perspective, scale, and start/stop. The study conducted recognition tests on five gestures with four models each, and the test results are presented in Table 1.

As shown in Table 1, the 3D-VR-DBLSTM model achieved recognition accuracy rates of over 93% for all five gestures. Among them, the recognition accuracy rate for backward movement reached 97.08 ± 1.45%. The 3D-GAN-ResNe model had an accuracy rate of 88.16 ± 2.47% for forward movement. The 3D-VR-LSTM model had the highest accuracy rate for forward movement, reaching 85.17 ± 3.68%. The above test results indicate that the 3D-VR-DBLSTM model, by relying on BiLSTM to capture the bidirectional action sequence and DSC to efficiently extract key features, can accurately recognize different types of gestures. The audience does not need to repeat the operation, which can significantly enhance the sense of immersion. Compared with the contrast model, it has obvious advantages. To further assess its practical performance, the study evaluated the interaction rate and revisit rate of all four models. The interaction rate = (Number of viewers who performed gesture interaction / Total number of visitors entering the exhibition area) * 100%, and the revisit

rate = (Number of viewers who participated in the same model experience again / Number of viewers who first experienced the model) * 100%. The results are shown in Table 2

Table 1. The test results of the recognition of 5 types of gestures

Type of gesture	3D-VR-DBLSTM	3D-VR-LSTM	3D-GAN-ResNet	CNN-SVM-Transformer
Forward	95.12±2.08 %	91.15±3.38 %	88.16±2.47 %	85.17±3.68 %
Backward	97.08±1.45 %	92.07±2.85 %	87.23±3.18 %	83.22±4.05 %
Amplify	94.25±1.76 %	89.13±3.56 %	86.19±2.75 %	82.19±3.87 %
Minification	93.37±2.25 %	88.21±3.97 %	84.26±3.29 %	80.24±4.46 %
Rotate	96.15±1.68 %	94.18±3.05 %	85.14±2.98 %	84.16±3.47 %

Table 2. Interaction rate and secondary experience rate test results

Test site	Model	Interaction rate (%)	Revisit rate (%)
A Museum	3D-VR-DBLSTM	70.62	72.85
	3D-VR-LSTM	62.43	53.41
	3D-GAN-ResNet	54.36	50.64
	CNN-SVM-Transformer	57.64	44.75
B Museum	3D-VR-DBLSTM	68.33	62.91
	3D-VR-LSTM	54.94	50.27
	3D-GAN-ResNet	50.47	46.72
	CNN-SVM-Transformer	48.62	41.36

As shown in Table 2, the 3D-VR-DBLSTM model achieved an interaction rate of 70.62% in the Museum A tests, which was 12.98% and 16.26% higher than those of the CNN-SVM-Transformer and 3D-GAN-ResNet models, respectively. In the Museum B tests, the 3D-VR-DBLSTM model achieved an interaction rate of 68.33% and a revisit rate of 62.91%. In conclusion, the 3D-VR-DBLSTM model has achieved high interaction rates and secondary experience rates in various museums, verifying the effectiveness of its accurate gesture recognition and natural interaction design in enhancing user engagement and retention. Subsequently, in order to evaluate the educational value and long-term operational value of the 3D-VR-DBLSTM model, each model was deployed in the museum for one month. The study compared the sales of cultural products in the museum during the operation of each model. At the same time, the research used a questionnaire survey method to ask 500 visitors questions and calculated the accuracy rate of the answers. The questions were about the knowledge related to ancient creatures displayed in the museum. The final test results are shown in Table 3.

Table 3. The test results of the educational value and long-term operational value of each model

Model	Accuracy rate of responses (%)	Sales revenue (¥)
3D-VR-DBLSTM	84.72	89634
3D-VR-LSTM	76.32	67802
3D-GAN-ResNet	70.58	53274
CNN-SVM-Transformer	65.81	48537
Traditional sculpture	52.47	31906

As shown in Table 3, in terms of educational value, the 3D-VR-DBLSTM model achieved the highest accuracy rate for tourists' knowledge responses, reaching 84.72%. The 3D-VR-LSTM model had an accuracy rate of 76.32%, the 3D-GAN-ResNet model had 70.58%, the CNN-SVM-Transformer model had 65.81%, while the accuracy rate of traditional sculpture display forms was only 52.47%. This gap confirms the enhancing effect of immersive interactive experiences on the efficiency of knowledge absorption, that is, the high accuracy rate of the model's gesture recognition and natural interaction design can drive tourists to shift from passive viewing to active exploration, deepening the memory and understanding of paleontological knowledge. Furthermore, in terms of long-term operational value, the 3D-VR-DBLSTM model led to increased sales of cultural and creative products, reaching 89,634 yuan. The 3D-VR-LSTM model followed closely with 67,802 yuan. The 3D-GAN-ResNet model and CNN-SVM-Transformer model were 53,274 yuan and 48,537 yuan, respectively. In contrast, traditional sculptures were only 31,906 yuan. This result indicates that the 3D-VR-DBLSTM model, with a higher interaction rate and secondary experience rate, significantly enhances the attraction of the exhibition area. This effectively drives cultural consumption and opens up a more promising operational path for museums. Therefore, the traditional physical display form has a relatively limited impact on consumption. The above experimental data fully validate the dual advantages of the 3D-VR-DBLSTM model in enhancing the effectiveness of scientific popularization education and empowering long-term museum operations, providing practical support for the transformation and upgrading of cultural heritage display and dissemination models.

5. Conclusion

To address the limitations of traditional prehistoric creature reconstruction methods, such as long reconstruction cycles and high replication costs, this study proposed a prehistoric virtual image reconstruction model based on 3D modeling and VR technology. On this basis, a virtual image reconstruction and interaction model was further constructed by integrating deep learning networks. Experimental results showed that the 3D-VR prehistoric virtual image reconstruction model achieved reconstruction accuracy rates of 94.8% and 95.4% on two datasets when the number of iterations reached 35. In addition, when the reconstruction time reached 10 seconds, the SSIM index exceeded 0.94 in both datasets. Furthermore, the 3D-VR-DBLSTM model achieved a loss value below 0.015 at 90 iterations. The recognition accuracy for five different types of gestures was above 0.93. When tested in two different museums, the model maintained an interaction rate above 68% and a revisit rate exceeding 62%. In conclusion, the 3D-VR-DBLSTM model effectively reconstructed various types of prehistoric creatures, enabling accurate interaction within the VR system. Although the proposed virtual reconstruction and interaction model demonstrated excellent performance, the experimental scope remained limited, and its generalizability requires further verification. Future work will involve testing the model on prehistoric creatures from different periods and in various exhibition settings to continually improve its overall performance.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

Declaration of Artificial Intelligence (AI) Tools

The author confirms that no AI tools were used in the preparation of this manuscript.

Reference

Abdulredah, S. K., and Al-Jawad, M. S. (2024). Building 3D geological model using non-uniform gridding for Mishrif reservoir in Garraf oilfield. *Petroleum Science and Technology*, 42(7), 809–827. https://doi.org/10.1080/10916466.2023.2305678

Alonazi, M., Ansar, H., Al Mudawi, N., Alotaibi, S. S., Almujally, N. A., Alazeb, A., and Min, M. (2023). Smart healthcare hand gesture recognition using CNN-based detector and deep belief network. *IEEE Access*, 11(5), 84922–84933. https://doi.org/10.1109/ACCESS.2023.3285604

Bartolini, L. S., and Rook, L. (2023). Nurturing Italian geo-palaeontological heritage with virtual palaeontology: Preliminary report of its application in two natural history museums. *Geoheritage*, 15(2), 40–53. https://doi.org/10.1007/s12371-023-00754-y

Gheisari, M., Hamidpour, H., Liu, Y., Saedi, P., Raza, A., Jalili, A., Rokhsati, H., and Amin, R. (2023). Data mining techniques for web mining: A survey. *Artificial Intelligence and Applications*, 1(1), 3–10. https://doi.org/10.47852/bonviewAIA2202290

Ghosh, R. M., Jolley, M. A., Mascio, C. E., Chen, J. M., Fuller, S., Rome, J. J., and Whitehead, K. K. (2022). Clinical 3D modeling to guide pediatric cardiothoracic surgery and intervention using 3D printed anatomic models, computer aided design and virtual reality. *3D Printing in Medicine*, 8(1), 11-27. https://doi.org/10.1186/s13018-022-02960-6

Hou, J., Wang, Y., Zhou, J., and Tian, Q. (2022). Prediction of hourly air temperature based on CNN–LSTM. *Geomatics, Natural Hazards and Risk*, 13(1), 1962-1986.

Hou, Y., Canul Ku, M., Cui, X., and Zhu, M. (2024). Super‐resolution reconstruction of vertebrate microfossil computed tomography images based on deep learning. *X‐Ray Spectrometry*, 53(5), 405–414. https://doi.org/10.1002/xrs.3375

Hsiang, E. L., Yang, Z., Yang, Q., Lai, P. C., Lin, C. L., and Wu, S. T. (2022). AR/VR light engines: Perspectives and challenge. Advances in Optics and Photonics, 14(4), 783–861. https://doi.org/10.1364/AOP.447992

Lai, K. W. K., and Chen, H. J. H. (2023). A comparative study on the effects of a VR and PC visual novel game on vocabulary learning. *Computer Assisted Language Learning*, 36(3), 312–345. https://doi.org/10.1080/09588221.2022.2030017

Miah, A. S. M., Hasan, M. A. M., and Shin, J. (2023). Dynamic hand gesture recognition using a multi-branch attention-based graph and a general deep learning model. *IEEE Access*, 11, 4703–4716. https://doi.org/10.1109/ACCESS.2023.3243875

Nemoto, T., Kobayashi, T., Kagesawa, M., Oishi, T., Kurokochi, H., Yoshimura, S., and Taha, M. (2023). Virtual restoration of ancient wooden ships through non-rigid 3D shape assembly with ruled-surface FFD. *International Journal of Computer Vision*, 131(5), 1269–1283. https://doi.org/10.1007/s11263-022-01693-4

Singla, P., Duhan, M., and Saroha, S. (2022). An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network. *Earth Science Informatics*, 15(1), 291–306. https://doi.org/10.1007/s12145-021-00763-0

Sun, X., Jia, J., Xu, P., Ni, J., Shi, W., and Li, B. (2023). Structure-guided virtual restoration for defective silk cultural relics. *Journal of Cultural Heritage*, 62, 78–89. https://doi.org/10.1016/j.culher.2023.01.015

Wahba, D. G., Abu El-Kheir, G. A., Tantaawy, A. A., and Abdel-Gawad, M. K. (2024). Extraction and restoration of the dinosaur bones of the Quseir Formation, Western Desert, Egypt. *New Valley University Journal of Basic and Applied Sciences*, 2(1), 32-36. https://doi.org/10.21608/nujbas.2023.251841.1019

Wu, J. M. T., Li, Z., Herencsar, N., Vo, B., and Lin, J. C. W. (2023). A graph-based CNN-LSTM stock price prediction algorithm with leading indicators. *Multimedia Systems*, 29(3), 1751–1770. https://doi.org/10.1007/s00530-022-00965-2

Wulandari, F., Widyaningrum, N., Sa'ida, N., and Masturoh, U. (2025). Meningkatkan kemampuan bahasa anak usia dini melalui pembelajaran multimedia interaktif berbasis AR dan VR. *Academicus: Journal of Teaching and Learning*, 4(1), 61–70. https://doi.org/10.59373/academicus.v4i1.86

Xiao, A., Huang, J., Guan, D., Zhang, X., Lu, S., and Shao, L. (2023). Unsupervised point cloud representation learning with deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9), 11321–11339. https://doi.org/10.1109/TPAMI.2023.3235689

Yan, G., and Xin, H. (2022). Practical research on artificial intelligence algorithms, paleontology, data mining, and digital restoration of public information. *Computational Intelligence and Neuroscience*, 2022(1), 3068686–3068694. https://doi.org/10.1155/2022/3068686

Yu, C., Qin, F., Watanabe, A., Yao, W., Li, Y., Qin, Z., and Xu, X. (2024). Artificial intelligence in paleontology. *Earth-Science Reviews*, 4(6), 104765–104788. https://doi.org/10.1016/j.earscirev.2024.104765

Challa, S. K., Kumar, A., and Semwal, V. B. (2022). A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data. *The Visual Computer*, 38(12), 4095-4109. https://doi.org/10.1007/s00371-021-02283-3

Yuexin Tang earned her master's degree in Design from Tongji University, Shanghai, in 2016. She previously served as a UI Designer at Huawei Shanghai Research Institute and Lenovo Chengdu Research Institute. She is currently the Deputy Head of the Department of Journalism and Communication in the College of Communication Science and Arts at Chengdu University of Technology. With extensive experience in both design development and higher education, her research focuses on Digital Media Design and Technology, specifically in 3D Virtual Reconstruction, Immersive Interactive Experiences (VR/AR), and Information Visualization. In her teaching, she adheres to the principle of integrating art and technology. She has led teams to complete over a dozen interactive media development projects and has independently spearheaded the design of multiple digital cultural and creative products for various brands.