The Application of Virtual Technology Based on Posture Recognition in Art Design Teaching

Qinyan Gao
College of Communication and Art Design
University of Shanghai for Science and Technology, Shanghai 200093, China

***Abstract*—With the development of virtual technology, posture recognition technology has been integrated into virtual technology. This new technology allows users to further understand and observe the activities carried out in life scenes based on their original observation of the external world. And it enables them to make intelligent decisions. Existing posture recognition cannot meet the requirements of precise positioning in virtual environments. Therefore, a two-stage three-dimensional pose recognition model is proposed. The experiment illustrates that the three-dimensional gesture recognition performance is excellent. In addition, under the ablation experiment, the error accuracy of the research model decreased by more than 5 mm, and the overall error accuracy decreased by 10%. In the P-R curve, the accuracy rate of the model reaches 0.741, and the recall rate reaches 0.65. When conducting empirical analysis, the virtual posture disassembly action is complete; the disassembly error is less than 5%, and the disassembly error accuracy is good. The fit degree of the leg bending amplitude reaches over 96%, and the fit degree of the arm bending amplitude reaches over 95%. When the model is applied to actual teaching, the overall satisfaction score of teachers and students reaches 94.6 points. This has effectively improved the teaching effect of art design and is of great significance to the development of education in China.**

***Keywords*—Posture recognition; deep learning; art design; time convolutional network; VR**

I. INTRODUCTION

5G technology has gradually integrated into all aspects of life. Virtual reality (VR) technology is being applied in various fields such as engineering design, game entertainment, and teaching applications as a new emerging technology. In a highly informational society, bringing virtual reality technology into teaching experiments and learning training can bring better teaching experiences. And its rich sensory stimulation and immersive teaching experience provide teachers and students with new teaching paths [1-3]. With the improvement of people's requirements for virtual technology, the accuracy of gesture recognition systems and the requirements for visual tasks are also increasing in response. The animation generated by two-dimensional gesture recognition systems has gradually been unable to meet the needs of high accuracy positioning of virtual space for doormen. Therefore, three-dimensional attitude recognition systems have been developed. However, 3D pose recognition systems have problems such as complex motion capture machines, and time-consuming multi camera and multi perspective solutions. Moreover, there are problems such as demanding camera operating environments, and the degree of occlusion affecting the accuracy of dynamic capture [4-6]. In the process of generating three-dimensional poses from two-dimensional images, there are currently problems such as insufficient joint point capture accuracy, occlusion affecting the range of joint and limb movements, discontinuous movement of the bone model within a single frame image, joint point change faults, and motion sequence generation jitter. These issues are not conducive to building 3D animation. However, with the publication of various pose datasets and the design and generation of new recognition models, research on capturing and generating three-dimensional poses using common cameras has gradually stepped onto the right track. In the sequential task of single frame image generation, temporal convolutional neural networks perform well in terms of energy. Moreover, the network has time correlation, and is applied to three-dimensional pose recognition in combination with high-precision calculation of angle vectors. This can effectively improve the accuracy of 3D pose recognition. And when it is applied in art design teaching, it can effectively solve the use of VR technology and recognition technology in online art teaching. This can also provide better technical support for the development of the art industry [7-9]. Therefore, by combining time convolutional neural networks and high-precision angle vector calculation, we can compensate for the current problems of insufficient node capture accuracy and the impact of occlusion on joint movement range, thereby effectively improving the accuracy of 3D pose recognition. By optimizing the temporal task process of single frame image generation, problems such as discontinuous movement of bone models, changes in joint points and faults, and jitter in action sequence generation are solved, laying the foundation for building high-quality 3D animations. Apply the optimized 3D pose recognition system to art and design teaching, improving the practical application value of VR technology in online art teaching.

The research content mainly includes four parts. The second part is a review of the research status of pose recognition technology and online teaching. The third part proposes a two-stage three-dimensional pose joint positioning and temporal image processing system model. The first section constructs the image to process the model, and the second section constructs the pose recognition model. The fourth part verifies the improved system performance and application effectiveness. The results indicate that the attitude recognition system has good application effects. The fifth Section concludes the research.

II. Related Works

The principle of posture recognition technology is to calculate human joint points from images and link them as a whole and output them in the form of images. And two-dimensional attitude recognition technology has a mature system and has been applied in various industries. The technology of three-dimensional gesture recognition is also developing rapidly. S Li et al. proposed a DS (Dempster Shafer) based evaluation algorithm to detect students' attention state in class by measuring students' facial posture. The detection of facial pose angle can be implemented in low pixel surveillance video. The experiment showcases that using DS theory to integrate each student's attention state, the overall classroom attention score of students changes and improves over time. This keeps the whole class in a good state of attention. Meanwhile, it proves that the design of the algorithm is feasible and effective. Compared to the average score of the questionnaire given by the reviewers, the accuracy of the proposed algorithm exceeds 85% [10]. To solve the inaccurate face recognition results caused by factors such as illumination, noise intensity, affine, and projection transformation, SWS et al. proposed scale invariant feature transformation (SIFT). Its research incorporates a SIFT algorithm based on principal component analysis (PCA) dimensionality reduction to reduce computational complexity and improve the efficiency of the algorithm. The experiment indicates that the dimension of SIFT has been reduced to 20 dimensions through experiments on open databases. This improves the efficiency of face extraction; Comparative analysis of several experimental results has verified the superiority of the improved algorithm [11]. MGR Alam et al. proposed an improved loMT emotion recognition system for studying and recognizing human emotional states. Experimental results show that the performance of the proposed method using benchmark dataset analysis has a high classification accuracy in judging human emotional state [12]. Hong Zhen et al. proposed a collaborative solution based on the artificial intelligence Internet of Things to solve the high rate of misjudgment in motion capture in healthcare. This scheme proposes an offline algorithm for multi posture recognition implemented on wearable hardware for posture recognition based on multidimensional data. The results show excellent performance in terms of accuracy and reliability [13].

In online teaching, diversified teaching methods reflect the optimization and reform of students' teaching concepts. In terms of improving online teaching, Z Fen proposed a biological immune algorithm framework using GBDT algorithm coding to improve the efficiency of online English teaching. Then, a flow feature selection algorithm based on bag learning is proposed to solve the problem that redundant information between features can reduce the accuracy of the framework. The research results show that the model constructed in the study has high reliability [14]. A study by Doligan et al. identified the relationships between specific variables, teaching experience, professional development, and teaching support, and self-learning among teachers transitioning to online teaching during the pandemic. The results showed that higher online teaching effectiveness scores were related to participation in online additional qualification courses and professional development courses. The highest online teaching effectiveness score is positively correlated with traditional learning management systems and the use of virtual technology support [15]. With the development of language research and teaching, M Li studies virtual reality technology based on artificial intelligence and machine learning. This technique is then applied to immersive contextual teaching to improve students' English learning abilities. Through a comparative teaching experiment between two classes of freshmen in a university, it is found that in traditional teaching classes, teachers occupy most of the time and students passively receive information. Therefore, there are insufficient channels for information exchange and the expression of ideas in target languages. The overall English level of immersion teaching is better than that of the control class, with an average score of 2.8 points higher. This indicates that immersive contextual teaching of college English combining constructivism theory with VR technology can indeed improve students' English proficiency [16]. For college art and design majors, virtual scenes have a more natural interactive way to promote learning. Y Zhang proposed a new interactive intelligent virtual reality (VR) paradigm. Aiming at the problem that traditional image-based rendering cannot meet the needs of artistic style virtual environments, nonrealistic rendering technology is introduced into virtual scene construction. Therefore, this study proposes a virtual environment generation method based on nonrealistic rendering. The experiment illustrates that the research model can not only simulate artistic images of linear wave motion, but also realizes the construction of artistic style virtual environments. This makes immersion teaching excellent [17].

In summary, research by domestic and foreign scholars on art design teaching and human gesture recognition shows that the teaching mode of art design is still relatively traditional. More curriculum reforms tend to focus on multimedia digital teaching, while virtual reality technology is less applied. There is still much room to accurately identify and apply the optimization of online education. Therefore, this study integrates temporal convolutional networks and angle vector computation into attitude recognition, and optimizes the gradient disappearance and gradient explosion problems of temporal convolutional networks. Then it improves the posture recognition effect and simplifies the model application operation. Finally, this study explores the optimization and development of art design teaching.

III. Construction of a Virtual Reality System for Art Design Teaching based on Gesture Recognition

A. Two-Stage Three-Dimensional Posture Joint Positioning and Timing Image Processing

In the teaching of art design, the state that reflects the mechanical structure of the human body is the posture of the human body. Common human postures are divided into three-dimensional models and skeletal models. A three-dimensional model is a virtual modeling model based on the human body, while a skeleton model is a tree structure diagram showing the changes in human posture. Therefore, the evaluation skeleton model has the advantage of being more concise and clearer than the virtual model. This is shown in Fig. 1.

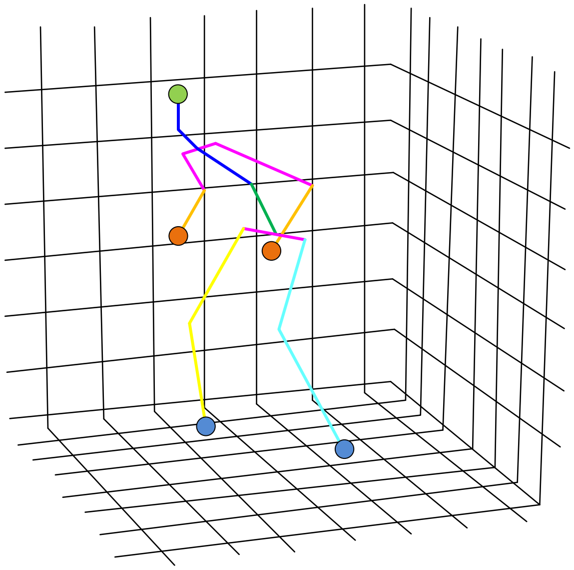


Fig. 1. 3D Tree structure diagram of skeleton model.

In the skeleton model, the research uses joint points as coordinates to represent human posture recognition and predict joint points; it expresses information such as the rotation angle of the human body with different posture changes. The stereoscopic model has strong picture representation and more simulation. However, due to factors such as computational complexity and estimated cost, it is not widely used in virtual costume changing and virtual animation design in art design teaching. The skeleton model has the advantages of greater flexibility and abundant joint information points, which makes it more efficient in motion tracking and other aspects. Therefore, this study selects a skeleton model for 3D pose recognition. Common human posture recognition methods include end-to-end posture recognition and two-stage posture recognition. End-to-end pose recognition methods that need to span data dimensions during training can lead to inaccurate posture prediction. Two-stage pose recognition is to extract important joint composition information from videos and images of the two-dimensional human body and reconstruct the three-dimensional pose in a virtual three-dimensional space. The study selected OpenPose for preliminary positioning of skeleton joint points. In the OpenPose algorithm, this study uses convolutional neural networks as feature extractors to extract feature points and calculate affinity and confidence, assemble joint points, and human posture features. The confidence algorithm uses neural network features to extract the probability value of a certain point coordinate joint point in the image. The personal confidence calculation formula is shown in Formula (1).

$$S_{i,j}^{*}(p)=e^{\left(\frac{-\left\|p-x_{i,j}\right\|^{2}}{\sigma^{2}}\right)} \tag{1}$$

In formula (1), $x_{i,j}$ is the corresponding bone joint point in the corresponding body; p is the coordinate point. The lower the confidence level when the coordinate point is away from the position of the bone joint point. The calculation of multi person attitude confidence is shown in Formula (2).

$$S_{i}^{*}(p)=\max_{j} S_{i,j}^{*}(p) \tag{2}$$

In formula (2), $S_{i}^{*}(p)$ is the multi person confidence level; And the final result is a normal distribution. When the coordinate points coincide with the bone joint points, the confidence level reaches the maximum value. In the skeleton model, to ensure that joint points correctly connect the direction of the limb, the study selected an affinity algorithm for prediction. The definition of the affinity algorithm is to predict the possible joint connection directions for each coordinate position at different positions of the body. When x_{j1} and x_{j2} are set as different joint points of the human body, v is the unit vector from the first joint point to the second joint point, and v_r is the vector perpendicular to the unit vector. Whether there is a point in the limb can be calculated. The calculation formula is shown in Formula (3) and Formula (4).

$$0 \le v\cdot(p-X_{j}) \le l_{f} \tag{3}$$

$$\left|v_{r}\cdot(p-X_{j})\right| \le l_{f} \tag{4}$$

In formulas (3) and (4), the length of the limb is l_f ; the limb width is w_j . Therefore, the distance between the point and the torso is calculated as a confidence vector. The affinity of pixel points is shown in Formula (5).

$$A_{f}(p)=\begin{cases}1\ , & \text{P On the limb}\\ 0\ , & \text{P is not on the limb}\end{cases} \tag{5}$$

When you select Openpose to generate a skeleton frame, there are a total of 25 joint points, as shown in Fig. 2.

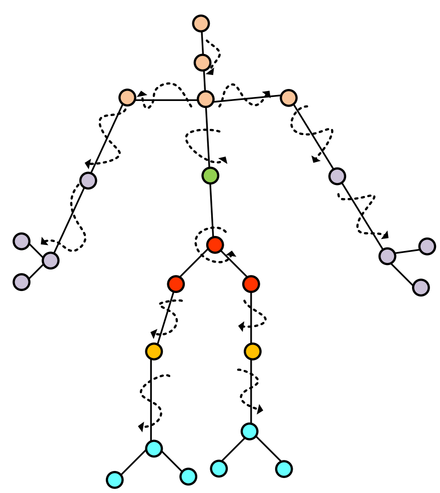


Fig. 2. Skeleton frame diagram.

In the skeleton frame diagram, the limbs have three different frames to determine joint points. However, when calculating the affinity of a point on a limb, it is not possible to determine whether the connection is correct, so the confidence level is recalculated using a definite integral method. The calculation of vector confidence at two different joint points is shown in Formula (6).

$$E_c = \int A_c(p(u)) \frac{X_{j2} - X_{j1}}{\left\| X_{j2} - X_{j1} \right\|} du \quad (6)$$

When the calculated confidence level is higher, it shows that the joint connection direction is unified, and the joint connection probability is higher. When sufficient joint point information is obtained, all joint points are rearranged and combined to distinguish different human bodies. Then, this study selects the Hungarian algorithm to re match the joint point distance. When there are three joint points, the corresponding joint points have only one edge actively connected. The constraint principle can be expressed as follows: if there is an interconnection relationship between type A joint points and type B joint points, the sum of the confidence levels of type A joint points n and all type B joint points is less than 1; Otherwise, the end of joint point n and B type joint point columns exceeds 1. After determining the backbone network under the three algorithms, temporal image processing was selected for this study. Attitude recognition based on two-dimensional key point sequences is expressed in multiple frames of images. Therefore, it can be viewed as a sequential task that utilizes time convolutional networks for parallel processing and adjusting the size of convolutional receptive fields. When there is an input sequence L, define the sequence as in Formula (7).

$$L = \{l_1, l_2, l_3, \ldots l_M\} \quad (7)$$

In sequence (7), M is the sequence length; the target sequence can be defined as Formula (8).

$$Y = \{y_1, y_2, y_3, \cdots y_N\} \quad (8)$$

In sequence (8), N is the target sequence length. Therefore, the sequence model can be defined as formula (9) through a mapping function.

$$\hat{Y} = f(x_1, x_2, x_3, \cdots x_T) \quad (9)$$

In the sequence (9), $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \hat{y}_3, \cdots \hat{y}_N\}$ is the output prediction result. For ease of calculation, this study assumes that the input sequence and the target sequence have the same length. The model satisfies constraints when the input and event information are correlated. The temporal image processing model should meet the principles of equal output and input, and the prediction results should not be correlated with delayed future information. Therefore, a linear convolution structure is selected in the model to ensure that the input layer and the output layer have the same length. And the convolution kernel selects causal convolution optimization replacement.

Fig. 3 is a schematic diagram of causal convolution. Each location in the input layer corresponds to a time step, and there are multiple dimensions of data in the practice department. The convolutional kernel moves through the first layer and passes on to obtain a causal effect. However, the time series of causal convolution is too long, resulting in excessively high model complexity. As a result, gradient disappearance and gradient explosion occur, and their selective cavity convolution increases the size of the receptive field and the amount of compression parameters. Therefore, when there is an input sequence, the calculation method for whole convolution processing a certain frame is shown in Formula (10).

$$F(t) = (X * f_d)(t) = \sum_{i=0}^{k-1} f(i) \cdot x_{t-d \cdot i} \quad (10)$$

In formula (10), d is the expansion coefficient; k is the number of convolution kernels; $t - d \cdot i$ is the time slice size

B. Construction of a Two-Stage Three-Dimensional Attitude Recognition Model

To improve the accuracy of 3D human posture prediction, research has been conducted to extract joint information from images or videos, and reconstruct 3D posture from 2D joint points. After generating 2D coordinates through OpenPose and generating complete 2D pose recognition joint point coordinates using a time convolution network, the 3D pose recognition reconstruction is completed using angle vectors. The overall structure of the model is shown in Fig. 4.

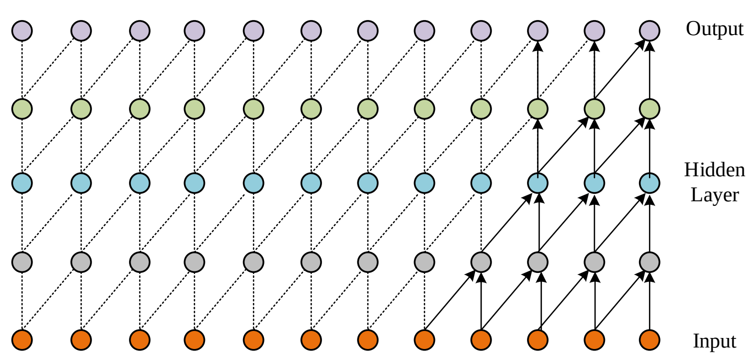


Fig. 3. Causal convolution diagram.

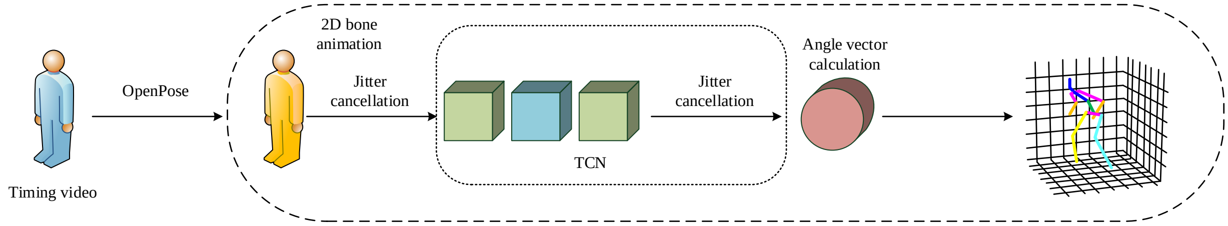


Fig. 4. Overall simplified flow chart of the model.

In Fig. 4, after capturing the original human posture and generating a timing video, two-dimensional bone animation is performed. Then it uses a time convolution model to eliminate jitter and calculates the angle vector to generate a three-dimensional human posture. The purpose of angle vector calculation is to reconstruct three-dimensional coordinates. Taking the head joint point as an example, it determines the length ratio of different joint points to determine the angle between the subject's body trunk and the photographing instrument; When the included angle is not 0, the transformation matrix normalizes the plane and coordinate axis, providing convenience for three-dimensional coordinate calculation. The normalization diagram is shown in Fig. 5.

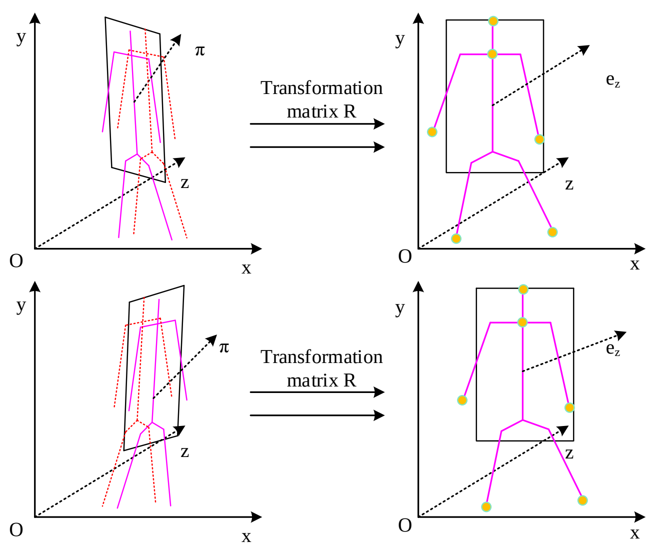


Fig. 5. Schematic diagram of human body normalization and front view.

During the normalization, the conversion matrix is obtained by taking the plane of the normal line of the head joint point as the normal vector. The rotated normal is the normal vector plane. The formula for calculating the rotation vector is shown in Formula (11).

$$A=(a_x,a_y,a_z)=\frac{\pi'\times e_z}{\left\|\pi'\times e_z\right\|} \tag{11}$$

In formula (11), e_z is the normal; (a_x,a_y,a_z) is the point coordinate; π is the plane normal vector of the head joint point. The formula for calculating the rotation angle is shown in Formula (12).

$$\theta=\cos^{-1}\left(\frac{\pi\cdot e_z}{\left\|\pi\right\|\left\|e_z\right\|}\right) \tag{12}$$

Therefore, the calculation formula for the rotation matrix is shown in (13).

$$R=\hat{A}+con\theta\cdot(I-\hat{A})\sin\theta\cdot A^{*} \tag{13}$$

In formula (13), $\hat{A}$ is the joint point matrix; The calculation formula for A^{*} is shown in matrix (14).

$$A^{*}=\begin{bmatrix}0 & -a_z & a_y\\ a_z & 0 & -a_z\\ -a_y & a_z & 0\end{bmatrix} \tag{14}$$

After obtaining the front coordinates of the photograph, measure the length of the joint points associated with each other and perform three-dimensional pose recognition. The schematic diagram of converting 2D joint points to 3D joint points is shown in Fig. 6.

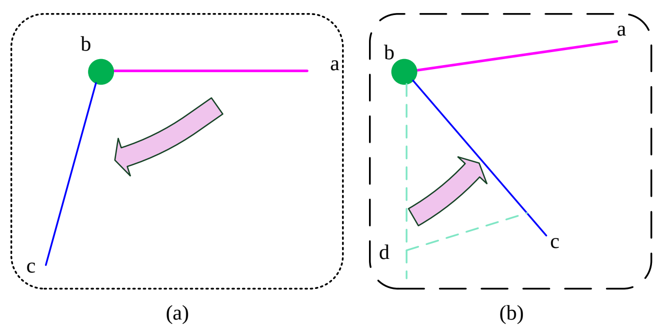


Fig. 6. Schematic diagram of 2D joint conversion 3D joint points.

In a two-dimensional plane, the angle between joint points (a, b) and joint points (b, c) is known. Then, by calculating the included angle of the vector, it can be concluded that when the joint point c moves, the two-dimensional output vector that changes is L_{bd}. The changed c_1 is the mapping of the joint point c. Therefore, the length of the vector before and after the change can be obtained from the image distance, and the included angle between the change vectors can be calculated. The included angle calculation formula is shown in Formula (15).

$$\theta_{bc} = \cos^{-1} \frac{L_{bc}}{L_{bd}} \quad (15)$$

In formula (15), θ_{bc} is the angle between the change vectors. By repeating this operation for each adjacent joint point, all three-dimensional coordinate positions can be obtained. Finally, the joint points are connected in a tree structure to obtain a human skeleton feature map. After obtaining the human bone feature map, evaluate the actual data and predicted data of predicted joint points, and the evaluation formula is shown in Formula (16).

$$MPJPE = \frac{1}{N} \sum_{i=1}^{N} \left\| J_i^* - J_i \right\|_2 \quad (16)$$

In formula (16), J_i^* is the i coordinate point corresponding to the predicted coordinate; J_i is the coordinate corresponding to the real coordinate point.

IV. Performance Verification of Three-Dimensional Coordinate Motion Recognition and Empirical Analysis of Motion Capture System

A. *Performance Verification and Action Testing of Action Recognition Algorithm*

In the performance analysis experiment of a two-stage 3D pose recognition model, in order to ensure the effectiveness of the model, a public dataset was used as the test set, consisting of Human 3.6M and Human Eva-I. The Human 3.6M dataset contains over 3 million three-dimensional poses, and the Human Eva-I dataset contains various motion process datasets. When testing the performance of the model, we first analyze the accuracy comparison between different algorithms and research algorithms in the Human Eva-I dataset. The comparison results are shown in Fig. 7.

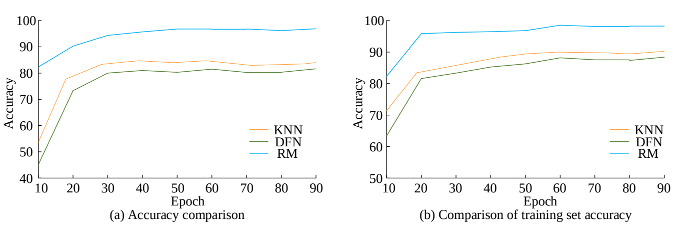


Fig. 7. Accuracy comparison.

Fig. 7(a) shows the comparison of accuracy between the DFN model and the research model. It can be seen that in training sets with a large number of people, the improvement in accuracy of the DFN model after 40 iterations is relatively limited, while the research model performs relatively well after 60 iterations. Additionally, overall, the performance of the model is quite good. In the training set with fewer people in Fig. 7(b), the accuracy of the DFN model increases with the increase of iterations, reaching a maximum accuracy of 89.57%. The accuracy of the research model has increased to 95.85% after 20 iterations, and since then, the number of iterations continues to increase, the rate of improvement is slow, and the accuracy gradually converges to 98%. The study selected methods such as ablation experiment analysis and estimation accuracy comparison for testing. The results of the ablation test are shown in Fig. 8.

In Fig. 8, in the comparison of ablation experiments, the red line represents the accuracy of the traditional OpenPose and angle vector combined pose recognition model under different error estimates; The blue lines are used to increase the accuracy of research models with residual structures under different error estimates. It can be seen that the MPJPE of the traditional model reaches 61mm, and the PA-MPJPE error calculation still reaches 50.5mm. The error accuracy of the research model with the addition of a residual structure rapidly decreased, and the overall error decreased by 5 mm when calculated through MPJPE. The optimal error accuracy is a model with three residual structures, and the error accuracy is only 51.3mm. Under the PA-MPJPE error calculation, the overall accuracy of the research model with a residual structure has a significant downward trend. And the optimal error accuracy is still the research model with three residual structures, only 43.2 mm. It shows that using time convolution network to eliminate video sequence jitter is effective and obvious. However, there is a gradient disappearance problem with too many modules. In the estimation accuracy comparison, it selects eight different actions in the Human3.6M dataset for testing, and selects DFN model, BKFC model, and KNN model for comparative testing.

From the scatter diagram in Fig. 9, it can be seen that the DFN model has the largest error. The action with the largest error among the eight actions is photo, reaching an error of 73 mm; the action with the smallest error is a walk, reaching 48.4mm. In the BKFC model, the action with the largest error is also a PHOTO action, reaching 57.2mm. The KNN model is the most effective model among the comparison models. The overall model error accuracy is improved by 5%, and the minimum error action is walk, only 41.8 mm. The overall error accuracy of the model in this study is 10% higher than the other three models, and the error is reduced to within 40 mm. The maximum error action photo is 38.1mm, and the minimum error action is walk, only 30.9mm. Even for the smooth purchase action, the accuracy improvement reaches about 8%. This indicates that using angular vector constraints to generate human posture has the best effect.

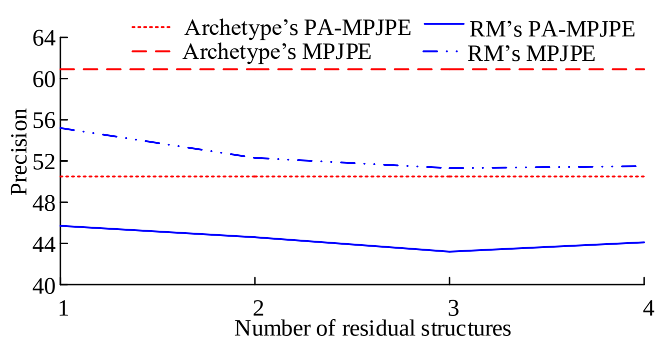


Fig. 8. Comparative results of ablation experiments.

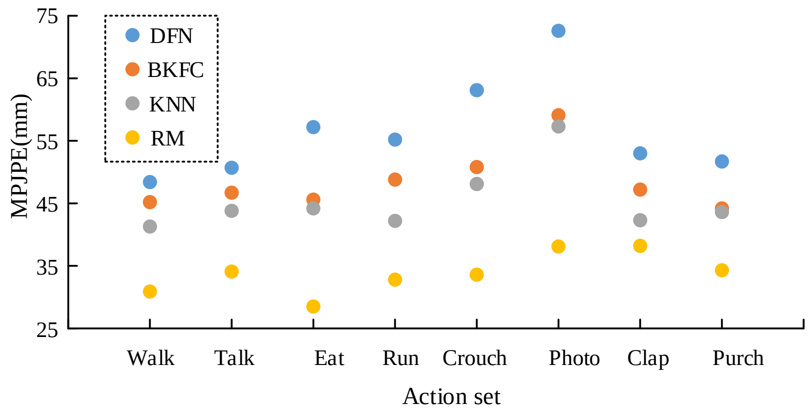


Fig. 9. Comparison of error tests for different models in data sets.

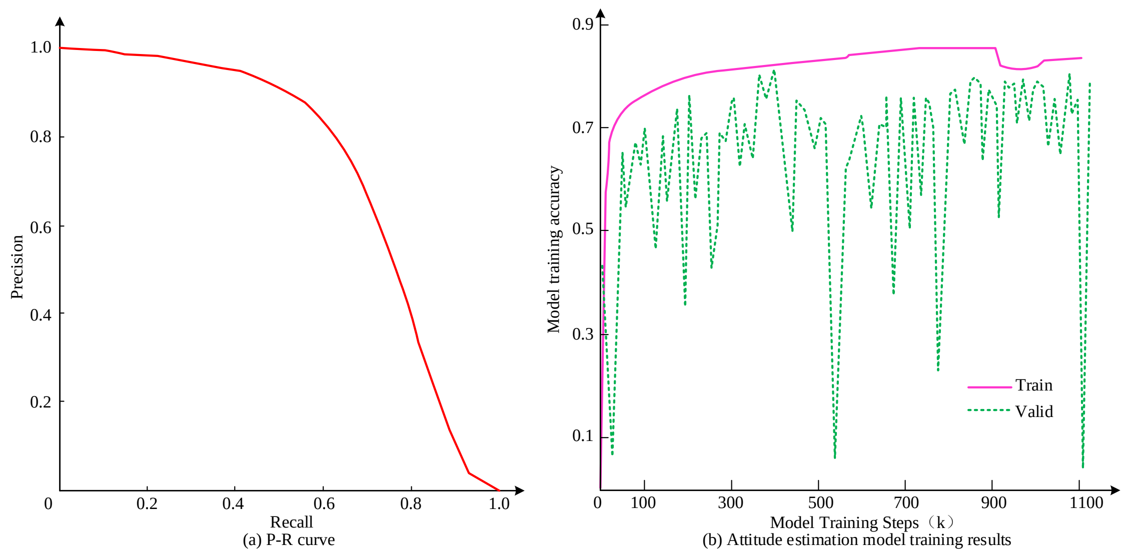


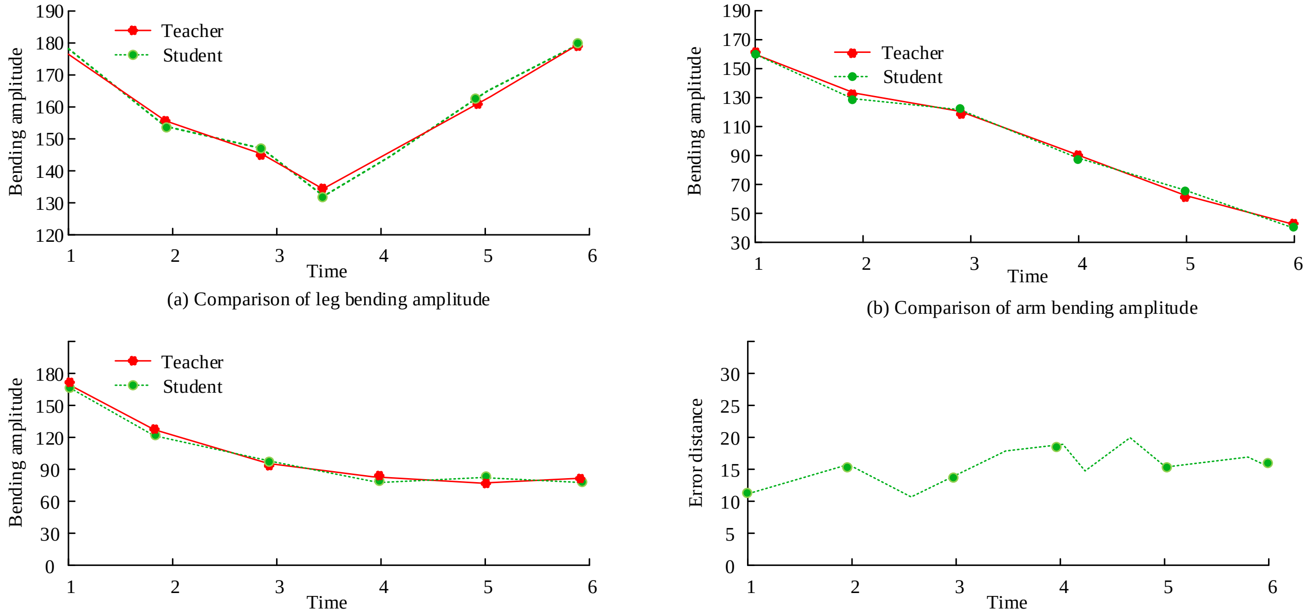
Fig. 10. Model training result accuracy and P-R curve.

Fig. 10(a) shows the model training results; During 300 iterations, the accuracy rate of test results on the validation set reached 0.741, and the recall rate reached 0.65. In the attitude estimation model training results in Fig. 10(b), the final accuracy of the model reaches 0.87, and the mAP verified and evaluated on the test set reaches 0.65.

B. *Empirical Analysis of a Two-Stage Three-Dimensional Attitude Capture System*

In the teaching of art design, this study utilizes research and improved three-dimensional posture capture system. This can enable students to understand the key design principles and multi-dimensional design expression techniques of certain actions in art design teaching from a comprehensive and three-dimensional perspective and at a data and three-dimensional level. This is different from traditional teaching in which students observe things or use digital courseware to learn and think about design norms and design concepts. Therefore, in the teaching of motion capture system, the research selects the leg disassembly teaching action and arm disassembly teaching action of three-dimensional animation football shooting action as capture demonstration. The experimental results are shown in Fig. 11.

The difference analysis and comparison of teacher and student motion capture in Fig. 11 shows that after disassembling the shooting posture in a three-dimensional manner, students can highly imitate the teacher's design skills. This study uses the accuracy of limb tracking as a performance indicator of the algorithm, and the bending amplitude of the leg and arm joints represents the tracking accuracy. In Fig. 11(a), the change in leg bending amplitude shows that the three-dimensional movements of the teacher's legs are clearer and more accurate, allowing students to observe the design from multiple angles, making it easier to learn and understand. The initial action designed by the students is slightly higher than that of the teacher, and the bending amplitude is gradually adjusted during the shooting process. Finally, an agreement was reached with the teacher to design the action after the shooting action was completed. The overall design model fits more than 96%. In the variation of arm bending amplitude in Fig. 11(b), there is body part occlusion, but the overall design fit still exceeds 95%. At the beginning of designing the action, the student's arm design angle is slightly lower than the teacher's design angle. In designing follow-up actions, the overall error is smaller and the arm bending amplitude is more easily detected. To sum up, the research on improved two-stage three-dimensional gesture recognition can improve the anti-interference ability of the sensor by separating the magnitude of the motion. Motion posture error recognition is more accurate and motion amplitude changes are sensitive. This not only has a higher accuracy, but also has a lower error angle than traditional gesture recognition methods to achieve accurate motion capture.



(a) Comparison of leg bending amplitude

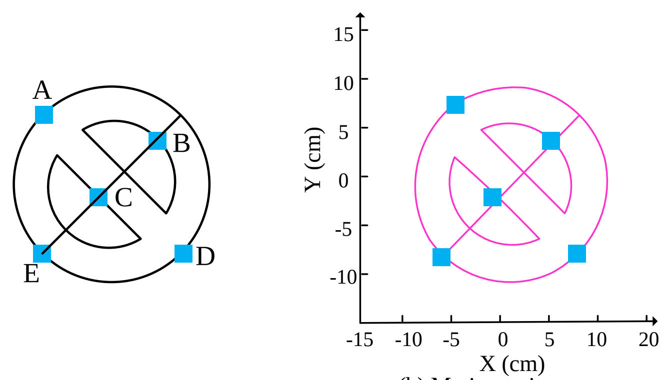
(b) Comparison of arm bending amplitude

(c) Feet bending range

(d) Overall movement deviation error

Fig. 1. Analysis of the difference in action capture between teachers and students.

Fig. 12 shows whether the studied human posture recognition system can correctly position the human body and whether the positioning error of joint points is within an acceptable range. Therefore, regular circular route experiments were conducted on the model. In Fig. 12(a), a walking route map is specified for the tester, and timing is performed at five checkpoints. Fig. 12(b) is a simulation diagram of the torso motion trajectory of the human body model by the human posture recognition system. The figure shows that the overall planning path of the model is good, and the actual travel range conforms to the specified travel path. The only point where a significant offset occurs is at point A. Overall, the research model can not only correctly identify human movements, but also effectively identify paths within the range of movement, with good motion capture accuracy. The position changes described meet the requirements of art design teaching. In addition to empirical analysis of the human posture recognition system, the system has also been applied to teaching. Based on this, a teaching satisfaction table is designed to compare with traditional design teaching and multimedia courseware-based design teaching. It compares the learning satisfaction, learning effectiveness, and teaching efficiency of the three teaching methods based on scores. The higher the score, the better is the effect. The comparison results are shown in Table I.



(a) Planning path map

(b) Motion trajectory simulation diagram

Fig. 11. Planning path simulation experiment results.

TABLE I. TEACHING SATISFACTION EVALUATION FORM

/	/	**Research model**	**Offline teaching**	**Traditional online teaching**
First group	Concept understanding	96	74	81
	Learning effect	93	76	86
	Teaching effectiveness	97	72	82
Second group	Concept understanding	89	80	89
	Learning effect	98	79	83
	Teaching effectiveness	96	71	81

In the evaluating teaching effectiveness, this study selected art and design majors from a certain university to conduct an experimental comparative evaluation of three teaching modes. When the research model is applied in the teaching process, the learning effect, concept understanding, and teaching situation have obvious advantages compared to traditional teaching and multimedia courseware teaching. In the score of concept understanding, the average score of the research model reached 92.5. In terms of learning effectiveness, the average score reached 95.5. In the teaching effectiveness score, the average score of the research model reached 96.5. Compared to the other two teaching modes, the average score exceeds 5 points. The score indicates that the improved posture recognition system in this study has achieved good teaching results when applied to design teaching. Moreover, students have a sufficient understanding of artistic design concepts, and can achieve a high degree of artistic space, with excellent teaching effects.

V. Conclusion

Efficient and accurate gesture recognition technology is essential for achieving human-computer interaction and applying interaction systems to the field of artistic design or production. Measuring human body data and calculating the movement amplitude of joint points pose a higher requirement for the refinement of gesture recognition technology. To improve the teaching effect of art design, a two-stage three-dimensional gesture recognition model is studied and designed. The experiment demonstrates that using open data sets as test sets and performing ablation experiments are compared with the estimation accuracy; the accuracy of the model error decreases rapidly when a residual structure is added to the ablation experiment. The overall MPJPE calculation error decreased by 5 mm to 51.3 mm. Under PA-MPJPE error calculation, the error accuracy is only 43.2mm, and the overall optimization is three modules. In the estimation accuracy comparison, the study selected eight different actions from the Human3.6M dataset for testing. The overall error accuracy of the research model was improved by 10%, and the error range was within an acceptable range. The accuracy rate of the model reached 0.741, and the recall rate reached 0.65. When conducting empirical analysis, the accuracy of motion disassembly error is good. The fit degree of the leg bending amplitude reaches over 96%, and the fit degree of the arm bending amplitude reaches over 95%. When the model is applied to actual teaching, the overall satisfaction score of teachers and students exceeds 94 points, which can effectively improve the teaching effect of art design. The research deficiency lies in the insufficient application of complex multi-level structure design. This is also the direction of future in-depth research.

References

[1] S. Wang, H. Zhang, X. Meng, “Design of Video Teaching System Based on Virtual Reality Technology”. Electronics Science Technology and Application, 2021, 7(4):72-78.

[2] L. Chen, D. Hu, X. Han, “Study on forearm swing recognition algorithms to drive the underwater power-assisted device of frogman”. Journal of Field Robotics, 2022, 39(1):14-27.

[3] T. Stone, S. Schumacher, E. Belilos, V. Kottamasu, “Virtual Teaching Kitchen within the Shared Medical Appointment Model”. American journal of lifestyle medicine. 2021,15(6s):4-5.

[4] T. Y. Pan, C. Y. Chang, W. L. Tsai, M. C. Hu. “Multisensor-Based 3D Gesture Recognition for a Decision-Making Training System”. IEEE Sensors Journal, 2021, 21(1):706-716.

[5] R. Salvador, P. C. Naval. “Towards a Feasible Hand Gesture Recognition System as Sterile Non-contact Interface in the Operating Room with 3D Convolutional Neural Network”. Informatica: An International Journal of Computing and Informatics, 2022,46(1):1-12.

[6] X. Liu, H. Shi, X. Hong, H. Chen, G. Zhao, “3D Skeletal Gesture Recognition via Hidden States Exploration”. IEEE Transactions on Image Processing, 2020, 29:4583-4597.

[7] B. Javidi, F. Pla, J. M. Sotoca, et al. “Fundamentals of automated human gesture recognition using 3D integral imaging: a tutorial.” Advances in Optics and Photonics, 2020, 12(4):1237-1299.

[8] Q. Liu, Human motion state recognition based on MEMS sensors and Zigbee network. Computer Communications, 2022, 181:164-172.

[9] Y. Peng, H. Tao, W. Li, H. Yuan, T. Li. “Dynamic Gesture Recognition Based on Feature Fusion Network and Variant ConvLSTM”. IET Image Processing, 2020,14(11):2480-2486.

[10] S. Li, Y. Dai, K. Hirota, Z. Zuo. “A Students' Concentration Evaluation Algorithm Based on Facial Attitude Recognition via Classroom Surveillance Video”. 2020,24(7 TN.147):891-899.

[11] W. S. Shun, Y. W. Sun., L. J. Xu. “Research on will-dimension SIFT algorithms for multi-attitude face recognition”. High Technology Letters, 2022, 28(3):280-287.

[12] M. Alam, S. F. Abedin, S. I. Moon, A Talukder, C. S. Hong, “Healthcare IoT-Based Affective State Mining Using a Deep Convolutional Neural Network. IEEE Access, 2019(7):75189-75202.

[13] Z. Hong, M. Hong, N. Wang, Y. Ma, X. Zhou, W. Wang, “A wearable-based posture recognition system with AI-assisted approach for healthcare IoT”. Future generations computer systems: FGCS, 2022(127):286-296.

[14] Z. Fen. “Efficiency improvement of English online teaching system based on bagging learning flow feature selection”. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 2021,4(40):6695-6705.

[15] T. Dolighan, M. Owen. “Teacher Efficacy for Online Teaching During the COVID-19 Pandemic”. Brock Education Journal, 2021, 30(1):95-116.

[16] M. Li. “An Immersive Context Teaching Method for College English Based on Artificial Intelligence and Machine Learning in Virtual Reality Technology”. Mobile Information Systems, 2021, 2021(2):1-7.

[17] Y. Zhang, “Application of Intelligent Virtual Reality Technology in College Art Creation and Design Teaching”. Journal of Internet Technology, 2021,22(6):1397-1408.