

**International Journal of Computational Systems Engineering**

ISSN online: 2046-3405 - ISSN print: 2046-3391

<https://www.inderscience.com/ijcsyse>

---

**The application of VR-based fine motion capture algorithm in college aerobics training**

Hui Wang

**DOI:** [10.1504/IJCSYSE.2027.10066368](https://doi.org/10.1504/IJCSYSE.2027.10066368)

**Article History:**

Received:	13 November 2023
Last revised:	02 February 2024
Accepted:	08 February 2024
Published online:	01 April 2025

---

# The application of VR-based fine motion capture algorithm in college aerobics training

---

Hui Wang

School of Physical Education,  
Yan'an University,  
Yan'an 716000, China  
Email: wanghui19800517@163.com

**Abstract:** In response to the problems of noise and incompleteness in the motion capture of VR technology in college aerobics training, this study first built a fine motion capture model based on an improved iterative nearest point algorithm, and then constructed an action recognition model based on an improved spatiotemporal graph convolutional neural network. The outcomes denote that the Top-1 of the action capture model is 36.5%, while the Top-5 is 59.4%. The Top-1 and Top-5 accuracy peaks of the action recognition model are 90.1% and 99.0%, respectively. The classification accuracy on the two datasets is 0.914 and 0.983, respectively. The standardisation level of the experimental group is 9.4 points higher than that of the control group. In summary, the model constructed through research has good application effects in fine motion capture and recognition, which helps to improve the teaching effect of efficient aerobics.

**Keywords:** action capture algorithm; virtual reality; VR; aerobics training; iteration closest point.

**Reference** to this paper should be made as follows: Wang, H. (2025) 'The application of VR-based fine motion capture algorithm in college aerobics training', *Int. J. Computational Systems Engineering*, Vol. 9, No. 6, pp.1–10.

**Biographical notes:** Hui Wang is currently working as an Associate Professor at the School of Physical Education, Yan'an University. She mainly engaged in aerobics, yoga and rope skipping teaching. She obtained her Bachelor's degree in Physical Education from Shaanxi University of Technology in 2003 and Master's degree in Physical Education and Training from Xi'an Sport University in 2008.

This paper was originally accepted for a special issue on 'AI and Cognitive Computing for Next Generation Mobile Networks' guest edited by Dr. Arvind Dhaka, Dr. Amita Nandal, Dr Edmar Candeia Gurjão and Dr. Dijana Capeska Bogatinoska.

---

## 1 Introduction

As the quick advancement of technology, various intelligent technologies have been utilised to the process of sports training, playing a significant role in improving the effectiveness of sports training. Virtual reality (VR) technology is an emerging product that integrates multiple information technologies and can bring users a sense of immersion and realism. It has been widely applied in fields such as medicine, sports training, and entertainment (Chang et al., 2020). However, in current VR-based motion capture, motion capture data is prone to noise and missing data, resulting in low quality of motion capture data and unsatisfactory subsequent motion recognition accuracy. The iterative closest point (ICP) algorithm is an iterative calculation method mainly used for precise stitching of depth images in computer vision. It achieves precise stitching by continuously iterating to minimise the corresponding points between source and target data, which has a certain impact on raising the accuracy of action capture (Zhang et al., 2021; Kurobeet al., 2020). spatiotemporal graph convolutional networks (ST-GCN)

can extract complex spatiotemporal dependencies by integrating graph neural networks and various temporal learning methods (Xiao et al., 2021; Li et al., 2021). Aerobics training requires attention to many fine body movements, and traditional VR-based motion capture methods have certain shortcomings in capturing fine movements. In this context, research has built an improved ICP algorithm-based fine motion capture model and an improved ST-GCN-based action recognition model, in order to help college aerobics teachers better understand the training situation of students using professional VR motion acquisition equipment. The innovation of this study mainly includes two points. The first point is to address the issue of time-consuming traversal of the ICP algorithm. KD tree and Levenberg Marquaret methods are used to improve the ICP algorithm, in order to achieve more accurate registration of point cloud data and improve the speed and accuracy of VR motion capture. The second point is to optimise ST-GCN based on the perturbation mechanism of Gaussian noise, so that ST-GCN can extract rich features between non-adjacent joint points. The main structure of the study includes four

parts. The first part is an analysis of the current relevant research status. The second part is to build an improved ICP algorithm-based fine motion capture model and an improved ST-GCN-based action recognition model. The third part is to analyse the application performance of the proposed model. The final part is a summary of the entire study.

## 2 Related works

VR technology can not only create virtual computer system environments, but also provide users with virtual experiences that share the same senses as the real world. Abich et al. (2021) stated that although many evaluations have been conducted on the effectiveness of VR technology, the evidence of training results has not been summarised and classified. To evaluate the potential benefits of these new technologies to trainees, research on the training effectiveness of VR technology was reviewed to determine the ability to effectively train or enhance through the use of VR from a domain agnostic perspective (Abich et al., 2021). Mäkinen et al. (2022) stated that the growth of VR technology has allowed the creation of the most comprehensive user experience. To analyse the use of different VR technologies in learning and the user experience of these technologies in medical practice and education, eight international databases were searched (Mäkinen et al., 2022). Tang et al. (2021) applied interactive VR technology to supplement traditional methods for the prevention of adverse events in hospitals, promoting program training, and conducted partial least squares modelling to study the correlation between each pair of measurement variables. The results indicated that the proposed model had good reliability for various measurement factors, and there was a significant correlation between each pair of measurement variables (Tang et al., 2021). Fussell and Truong (2022) developed a technology acceptance model that combined factors related to education and VR technology in training environments to address the issue of limited research on students' perception of the technology and their intention to use it for training in the exploration of learning availability related to VR. This model helped to identify factors that affect students' willingness to use VR in dynamic learning environments (Fussell and Truong, 2022). McGovern et al. (2020) investigated the usage of VR technology in education to better understand how VR technology can help students improve their communication skills in effective speeches and public speaking activities. The results indicated that the adoption of VR technology was very beneficial for business educators to help students improve their presentation skills, help students evaluate their presentation skills, and gain more confidence (McGovern et al., 2020). Han et al.'s (2022) discussion on the potential risks of VR technology remained largely limited to usability challenges, with only a few studies reflecting the social, psychological, and physical impacts exposed by this immersive technology. They critically reviewed literature on escapism to discuss issues

in the design and use of VR consumer experience escapism (Han et al., 2022).

Action capture refers to the collection of target posture information through a certain means, and the reproduction of the information into digital information that can be processed. It has been widely used in fields such as entertainment, health records, and sports training. Shimada et al. (2020) proposed the first algorithm for physically real-time and unlabelled human 3D motion capture using a monochromatic camera at a speed of 25 frames per second, addressing the significant limitations of unlabelled 3D human motion capture. They also used convolutional neural networks to infer 2D and 3D joint positions. The results indicated that the proposed method could capture global 3D human motion that is reasonable and stable in time from the video (Shimada et al., 2020). Mehta et al. (2020) raised a real-time method for multi person 3D motion capture using a single RGB camera at speeds exceeding 30 fps, and used convolutional neural networks to estimate 2D and 3D pose features and identity assignments for all visible joints of all individuals. The results indicated that the proposed system operated at a faster speed and realised state-of-the-art accuracy (Mehta et al., 2020). Chatzitofis et al. (2021) introduced an end-to-end differentiable marker pose model to address the high cost of motion capture methods based on optical markers, to address constrained position estimation, noise input data, and spatial configuration invariance. They also introduced a new spatial 3D coordinate regression technique under the concepts of multi view rendering and monitoring to estimate the 3D pose (Chatzitofis et al., 2021). Kadirvelu et al. (2023) addressed the issue of using simple behavioural assessments and requiring a longer trial time in the current gold standard clinical scale. They used wearable sensors to capture patients' whole-body movements, defined digital behavioural features based on data, and combined these features using machine learning. The results indicated that the proposed method had certain potential in achieving behavioural transcriptomics (Kadirvelu et al., 2023). Gao et al. (2023) designed a hand pose estimation method with hand biological constraints in a single RGB image and a dexterous robotic arm teleoperation framework with hand motion capture and video calling tools to address the problem of monocular hand motion capture. The results indicated that the proposed hand posture estimation method performed better than existing techniques on the hand posture estimation dataset and had a certain degree of effectiveness (Gao et al., 2023). Wierschem et al. (2020) introduced a manufacturing time and motion research method based on motion capture, which used motion capture technology to collect, transform, store, and analyse data on repetitive body movements performed by manufacturing workers. The results indicated that this method has developed the ability to identify repetitive motion patterns and statistically significant deviations from these patterns (Wierschem et al., 2020).

In summary, although many previous scholars have conducted extensive research on motion capture and affirmed the role of VR technology in healthcare, education,

and motion capture, the accuracy and speed of motion capture based on VR technology still need to be improved. To this end, the research will build an improved ICP algorithm-based fine motion capture model and an improved ST-GCN-based action recognition model to improve the efficiency and accuracy of motion capture.

### 3 Fine motion capture algorithms for aerobics based on VR

It is crucial to strengthen physical education teaching in universities to cultivate high-quality talents with comprehensive development. In response to the noise and incompleteness issues of VR technology in motion capture during aerobics training, the study will build an improved ICP algorithm-based fine motion capture model and an improved ST-GCN-based action recognition model.

#### 3.1 Building a fine motion capture model based on improved ICP algorithm

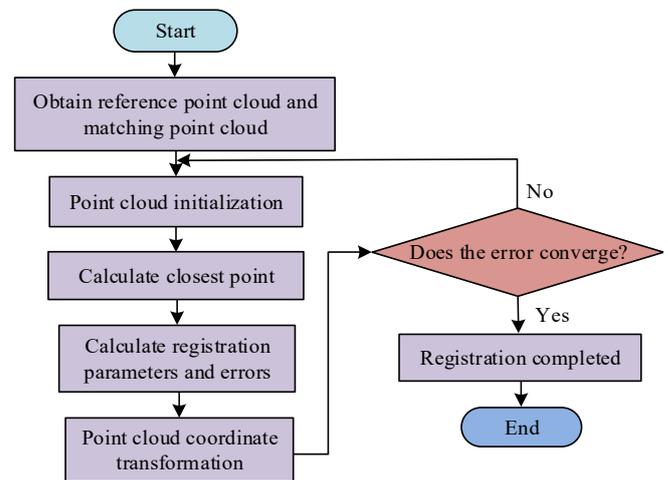
Wearable VR devices can effectively capture the actions of teachers and students during aerobics teaching and training, and export them as visual animations for aerobics video teaching. Students can learn by watching the teacher's aerobics movements captured by VR devices, and also analyse their own shortcomings by watching the movements captured by VR devices, thereby improving the training effect of efficient aerobics. The advantage of this training method is to help students have a more intuitive understanding of the correct movement posture and techniques, better grasp the correct movement methods, and reduce the possibility of misunderstandings and misinterpretations. You can also compare and analyse your own actions with standard actions, in order to identify your shortcomings and make improvements. However, the capture speed and registration accuracy still cannot meet the practical application requirements. The use of high-precision motion capture equipment, high-resolution camera technology, sensor fusion technology, as well as machine learning and artificial intelligence technology can help ensure the capture of fine motion in a realistic and reliable manner in the virtual world. To better capture the fine movements of aerobics and ensure that the capture speed and registration accuracy meet the needs of practical applications, more accurate registration of point cloud data is also needed to ensure effective improvement of registration accuracy. Obtaining point cloud data using VR devices mainly involves three steps. Firstly, use VR devices to track the position and movements of the head and hands during aerobics training. Then, install point cloud capture software such as Kinect SDK to convert the actions captured by VR devices into point cloud data. Finally, export the captured point cloud data for subsequent analysis

and processing. The ICP algorithm is based on data registration and utilises the nearest point search method to address an algorithm based on free form surfaces. Its essence is the optimal matching algorithm based on the least squares method. The ICP algorithm is mainly calculated through the optimisation method of the least squares method, where the minimisation function is shown in formula (1).

$$F(R, T) = \sum_{i=1}^N \|Q_i - (RP_i + T)\|^2 \quad (1)$$

In formula (1),  $R$  represents the rotation matrix;  $T$  means the translation vector;  $Q_i$  denotes the point closest to  $P_i$  in the target data point;  $P_i$  expresses the point set corresponding to the initial data. The ICP algorithm has advantages such as high accuracy and is widely used in 3D reconstruction, autonomous driving, VR, and augmented reality. The specific process of the algorithm is denoted in Figure 1.

**Figure 1** Flow chart of ICP algorithm (see online version for colours)



However, the classic ICP algorithm is prone to bottlenecks during the iteration process, resulting in slow convergence speed, low accuracy, and poor universality, making it difficult to obtain point cloud data. Therefore, to more accurately and comprehensively reflect the environmental features, the use of normal vectors and curvature are used to match feature points, and on the basis of considering Euclidean distance, vector angle difference is utilised to filter out mismatched points, to improve the ICP algorithm. The improved algorithm is called the IICP algorithm. Before point cloud matching, it needs to preprocess the point cloud data. The centre coincidence method utilises the reference point cloud data to be matched for centre coincidence, which can effectively reduce the translation error of the two frame point cloud data and improve the accuracy of subsequent point cloud matching, as shown in formula (2).

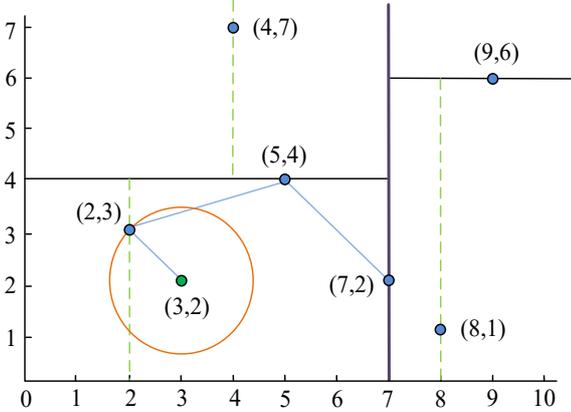
$$\begin{cases} u_x = \frac{1}{n} \sum_{i=1}^n x_i \\ u_p = \frac{1}{n} \sum_{i=1}^n p_i \\ w = u_p - u_x \\ u'_x = x_i - w \end{cases} \quad (2)$$

In formula (2),  $u_x$  and  $x_i$  represent the centroid and the coordinates of the point cloud to be matched, respectively;  $p_i$  refers to the coordinates of the reference frame point cloud;  $w$  expresses the translation vector;  $u'_x$  denotes the coordinates of the centre of gravity point cloud. KD-tree is an algorithm that divides data points into a  $k$ -dimensional space to form a data structure, which can quickly find data points that meet the conditions and reduce computational complexity. The central idea is to divide the entire space into specific parts and perform search and comparison operations in that specific space (Zhang et al., 2022). Firstly, the KD Tree is constructed using the idea of segmentation operation. Six feature samples are selected for the study, namely (2, 3), (4, 7), (5, 4), (7, 2), (8, 1), and (9, 6). Then, the variance of the samples is calculated, as shown in formula (3).

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (3)$$

In formula (3),  $N$  refers to the amount of samples;  $x_i$  stands for the coordinates of the samples;  $\bar{x}$  represents the mean of the samples. KD-Tree can divide laser point cloud data into a certain data structure space. When searching for the nearest point, there is no need to traverse every point, but instead, it searches in a specific subspace, greatly reducing the time for searching for the nearest point and improving the point cloud matching speed. When there are points (3, 2), the schematic diagram of using KD-Tree to find nearest neighbour points is shown in Figure 2.

**Figure 2** Schematic diagram of KD-Tree searching for neighbouring points (see online version for colours)



Due to the nonlinear least squares problem involved in the ICP algorithm, it needs to be solved using the Levenberg-Marquardt method, and the iterative process

converges to obtain the desired solution. In the process of finding point matches, normal vectors and curvature are taken into account, which largely ruled out some incorrect matches. While considering Euclidean distance, it also needs to consider the distance between normal vectors. The distance from the matching point to the surface is shown in formula (4).

$$\begin{cases} I^{Y_k}(x) = \sum_{y_i \in Y} W_i(x) ((x - y_i) n_i) \\ W_i(x) = e^{-\|x - y_i\|^2 / h^2} \end{cases} \quad (4)$$

In formula (4),  $I^{Y_k}(x)$  represents environmental information;  $Y_k$  represents reference frame;  $W_i(x)$  represents  $y_i$  matching degree;  $n_i$  represents  $y_i$  normal vector;  $h$  represents corresponding point distance. The IICP algorithm utilises real surface features from environmental maps to filter out erroneous point matches, taking into account the angle difference of the corresponding point normal vector, which makes the solution of the rotation matrix more accurate. But rotation can cause significant translation errors, while translation does not cause rotation errors. The error function is shown in formula (5).

$$\begin{cases} e_{ji}(T) = (\tilde{p}_i^c - T \oplus \tilde{p}_i^r) \\ T \oplus \tilde{p}_i \begin{bmatrix} R p_i + t \\ R n_i \end{bmatrix} \end{cases} \quad (5)$$

In formula (5),  $e_{ji}(T)$  represents the error function;  $T$  represents the transformation matrix;  $p_i$  represents the point coordinates;  $n_i$  represents the normal vector;  $\tilde{p}_i$  serves as the point to be matched;  $T \oplus \tilde{p}_i$  represents the mathematical description of the rotation transformation and normal vector;  $\oplus$  represents the operator;  $R$  represents the radius. The objective function is shown in formula (6).

$$\begin{cases} \sum_C e_{ij}(T)^T \tilde{\Omega}_{ij} e_{ij}(T) \\ \tilde{\Omega}_{ij} = \begin{bmatrix} \Omega_i^s, 0 \\ 0, \Omega_i^n \end{bmatrix} \end{cases} \quad (6)$$

In the process of solving curvature and normal vectors, point cloud data is considered as a Gaussian distribution, and the normal vector is defined as the feature vector corresponding to the minimum eigenvalue. The mean of all points  $v_i$  within the radius  $R$  range around point  $p_i$  is shown in formula (7).

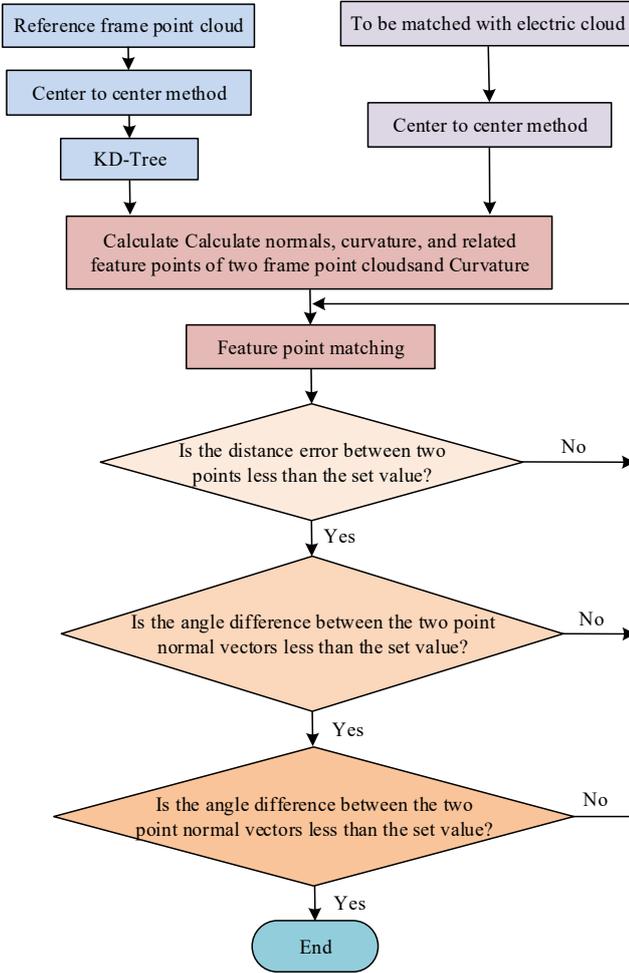
$$\mu_i^s = \frac{1}{|v_i|} \sum_{p_j \in v_i} p_j \quad (7)$$

The variance and curvature are shown in formula (8).

$$\begin{cases} \sum_i^s = \frac{1}{|v_i|} \sum_{p_j \in v_i} (p_j - \mu_i)^T (p_j - \mu_i) \\ \sigma_i = \frac{\lambda_1}{\lambda_1 + \lambda_2} \end{cases} \quad (8)$$

The matching rules in the above steps filter out a large number of error points. If the distance between two points is greater than the set threshold, it will be filtered out. If the difference in curvature is greater than the threshold, it will be filtered out. If the angle difference of the normal vector is greater than the threshold, it will also be filtered out. In contrast to the traditional ICP algorithm, the IICP algorithm solves the objective function through the LM method, taking into account the curvature and normal vectors in the environment surface. It eliminates some error points before matching, shortens the calculation time, and improves the accuracy of matching (Chen et al., 2022). The specific process of IICP algorithm is denoted in Figure 3.

**Figure 3** Flow chart of IICP algorithm (see online version for colours)

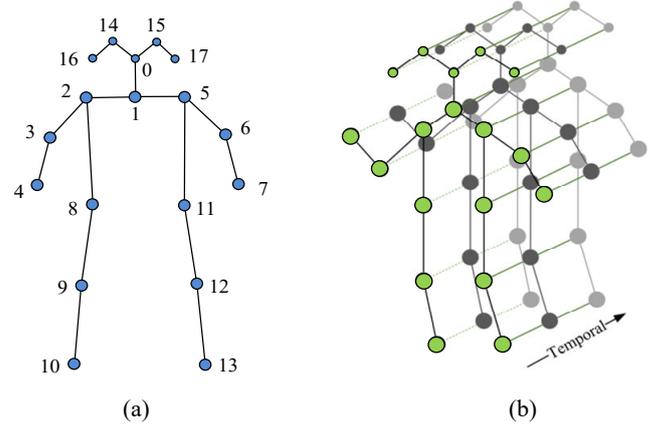


### 3.2 Building an action recognition model based on improved ST-GCN

After capturing the fine movements of aerobics training using the IICP algorithm, it is also necessary to recognise the training movements. Action recognition is based on human pose estimation, extracting human skeleton information from each frame of the video image, and combining the human skeleton from the time dimension to obtain action sequence information. To better achieve human head and body positioning, commonly used action

recognition datasets add neck key points between the left and right shoulders to the annotation of posture estimation datasets, forming an 18 degree of freedom human skeleton structure. Each frame of the human skeleton is correlated from the time dimension to obtain the skeleton spatiotemporal sequence input to the human action recognition model, as shown in Figure 4.

**Figure 4** Action recognition skeleton structure and skeleton space-time diagram, (a) action recognition skeleton structure (b) skeleton space-time series (see online version for colours)



GCN can model and learn complex human motion trajectory data, effectively capturing the dependency relationships between joints, thus having certain advantages in human motion analysis and modelling. ST-GCN applies graph convolutional neural networks to human action recognition, learning the spatial and temporal patterns of human actions from data by integrating spatial graph convolution and temporal convolution. It can extract complex spatiotemporal dependencies, and the feature expression effect is more prominent (Wang et al., 2022). Under the condition that the rules for constructing spatiotemporal maps are known, the ST-GCN model first dynamically models the human skeleton sequence in space and time, and then convolutions the spatiotemporal features of the modelled actions. Finally, the features obtained after convolution will go through the global average pooling layer and the SoftMax classifier to output the probability vector for each action category. The category with the highest probability value corresponds to the human action category. On the spatial dimension, the graph convolution on a certain joint point  $v_i$  is shown in formula (9).

$$f_{out}(v_i) = \sum_{v_j \in \beta_i} \frac{1}{Z_{ij}} f_{in}(v_j) \cdot \omega(l_i(v_i)) \quad (9)$$

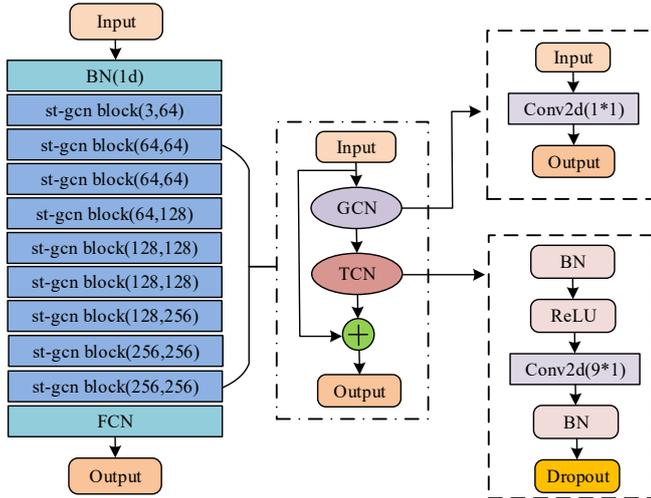
In formula (9),  $f_{out}$  and  $f_{in}$  represent the output and input feature maps;  $\beta_i$  represents the sampling area during  $v_i$  convolution;  $Z$  represents the cardinality of the  $v$  subset where the vertex HH is located;  $\omega$  represents the weight function that provides the weight vector. In order to use the weight function correctly, ST-GCN designs a mapping function that can divide the neighbourhood  $B$  into  $x$  subsets, assign different labels to each subset, and map the nodes in

each neighbourhood to a certain subset  $B$  of  $x_i$  according to the corresponding mapping rules. Converting formula (9) can obtain the graph convolution of ST-GCN, as shown in formula (10).

$$\begin{cases} f_{out} = \sum_k^{k_v} \omega_k (f_{in} A_k) \oplus M_k \\ A_k = \Lambda_k^{-\frac{1}{2}} \bar{A}_k = \Lambda_k^{-\frac{1}{2}} \end{cases} \quad (10)$$

In formula (10),  $k_v$  represents the size of the convolution kernel;  $\bar{A}_k$  means the adjacency matrix of  $N \times N$ ;  $\Lambda_k^{-\frac{1}{2}}$  represents the normalised diagonal matrix;  $M$  represents a learnable weight matrix;  $\omega$  represents the  $1 \times 1$  convolution operation;  $\oplus$  represents the dot product operation. In summary, ST-GCN extracts higher-level spatiotemporal features after convolution operations in both spatial and temporal dimensions. The network framework of ST-GCN is shown in Figure 5.

**Figure 5** The network framework of ST-GCN (see online version for colours)



However, ST-GCN cannot generate new connections between non-adjacent joint points during the training process using point multiplication operations, which limits its learning ability. Such adjacency matrices are not the optimal choice. To address this issue, a disturbance mechanism based on Gaussian noise is proposed to optimise ST-GCN, which introduces Gaussian noise into the spatial image convolution layer of the original ST-GCN image convolution layer. Gaussian noise is a type of noise and its probability density function is in line with a Gaussian distribution. This noise is added to the original adjacency matrix to form a new adjacency matrix. The disturbance of Gaussian noise enables the creation of connection relationships between non-adjacent joint points, optimising the adjacency matrix and obtaining a skeleton space structure that is more suitable for describing action samples, thus enabling better action recognition (Wang et al., 2021; Khmag, 2023). Gaussian noise is shown in formula (11).

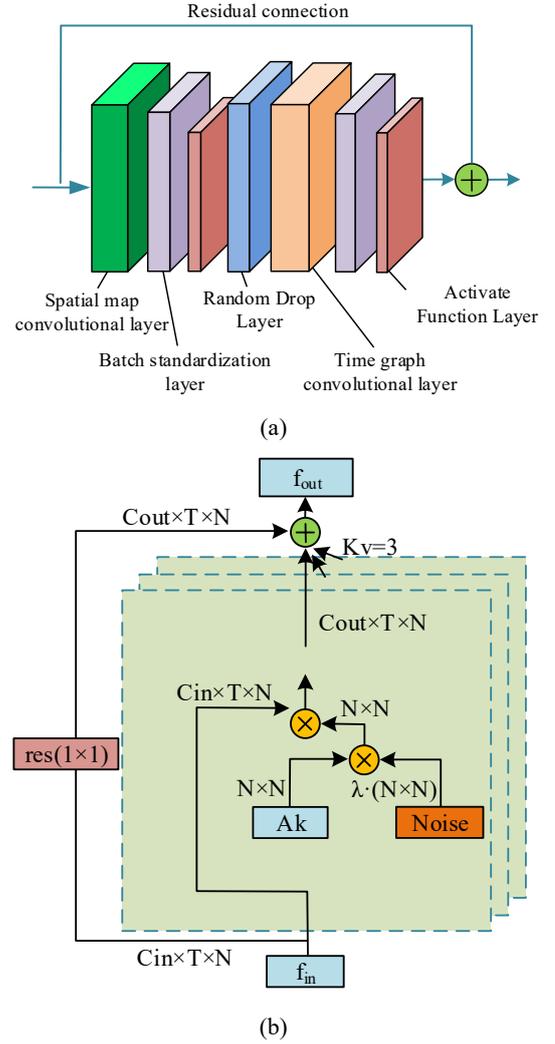
$$Noise = \lambda \cdot \varphi \quad (11)$$

In formula (11),  $\lambda$  represents a trainable parameter, and  $\varphi$  represents Gaussian noise. After adding noise, formula (10) can be transformed into formula (12).

$$f_{out} = \sum_k^{k_v} \omega_k f_{in} (A_k + (\lambda \cdot \varphi)) \quad (12)$$

During the entire optimisation process, prior knowledge  $A_k$  is retained and Gaussian matrix Noise is added to  $A_k$ . The schematic diagram of noise addition is expressed in Figure 6.

**Figure 6** Basic units and noise addition diagram of ST-GCN, (a) basic units of ST-GCN (b) schematic diagram for adding noise in convolutional layers (see online version for colours)



#### 4 Analysis of the effect of fine motion capture algorithm in aerobics based on VR

To promote the application effect of VR technology in efficient aerobics training, a fine motion capture model based on IICP algorithm and an improved ST-GCN-based action recognition model have been studied and built. However, its effectiveness still needs further verification. The research mainly analyses from two aspects. The first

part analyses the application effect of the fine motion capture model based on the IICP algorithm, and the second part verifies the improved ST-GCN-based effectiveness of the action recognition model.

#### 4.1 Effect analysis of fine motion capture model based on IICP algorithm

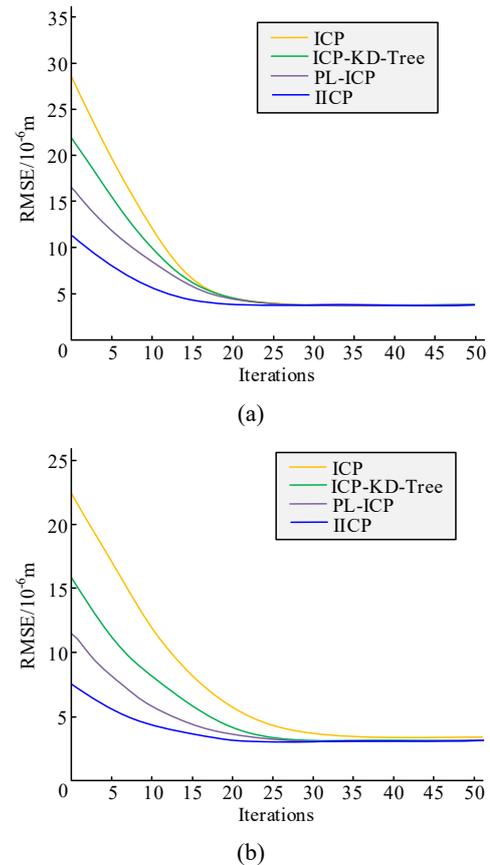
To evidence the effectiveness of the IICP algorithm, simulation analysis was conducted in the MATLAB environment using the Bunny dataset and complex surface dataset to obtain the root mean square error (RMSE) of the algorithm. The results were compared with the ICP algorithm, PL-ICP algorithm, and ICP-KD-tree algorithm, as shown in Figure 7. From the graph, among the four algorithms, the IICP algorithm had the fastest iteration speed on both datasets, while the ICP algorithm had the slowest iteration speed. In addition, when the number of iterations was 20, compared to the traditional ICP algorithm, the error of the IICP algorithm was reduced by 40.16% and 36.27%, respectively. The results showed that the proposed improved ICP algorithm could effectively improve the convergence speed of the ICP algorithm, and the IICP algorithm had a certain degree of effectiveness.

**Table 1** Accuracy comparison of five motion capture methods on the kinetics skeleton dataset

Model	Top-1/%	Top-5/%	Time consuming (min/epoch)
Artificial design features	14.8	25.6	None
LSTM	16.3	35.1	None
TCN	20.2	39.9	None
AGCN	33.7	55.9	45.6
Our	36.5	59.4	23.4

To prove the effect of the IICP algorithm-based fine action capture model, the Kinetics Skeleton dataset was used to validate the accuracy of the model. The mean of Gaussian noise and variance were set to 0.03 and 0.2, respectively. And compared with methods based on artificial design features, long short-term memory neural network (LSTM), temporal convolutional network (TCN), and adaptive graph convolutional networks (AGCN). LSTM can be used to model and predict continuous motion trajectories, and can effectively process action information at different time scales. TCN can effectively capture short-term and long-term dependencies in action sequences, and has good applicability for action capture. The accuracy comparison results of the five models are shown in Table 1. From the table, among the five motion capture methods, the model in this study had the highest accuracy, with Top-1 being 36.5% and Top-5 being 59.4%, and the time consumption was shorter, proving the superiority of the proposed model.

**Figure 7** Comparison of iteration errors of four algorithms, (a) Bunny dataset (b) complex surface dataset (see online version for colours)



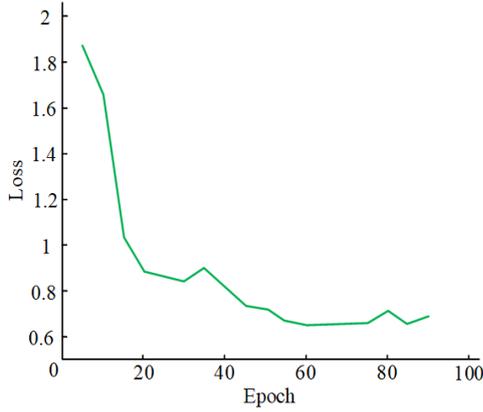
#### 4.2 Analysis of the effect of an improved ST-GCN-based action recognition model

To verify the feasibility of the improved ST-GCN-based action recognition model, the study used an X-Sub data subset of NTU-RGB D for training and validation. The momentum, batch size, weight attenuation, and the dropout were set to 0.9, 4, 0.0001, and 0.5, respectively. The performance change curve of the model is expressed in Figure 8. From the graph, the accuracy of the model in this study was relatively high. Around the 60th epoch, the model's loss attenuation began to stabilise, and around the 90th epoch, the Top-1 and Top-5 accuracy of the model reached their peak, reaching 90.1% and 99.0%, respectively.

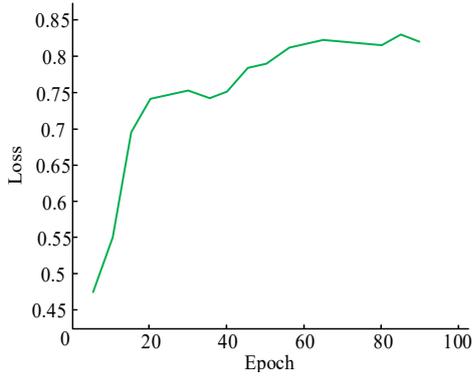
To verify the accuracy of the action recognition model based on improved ST-GCN, the NTU-RGB D and UMONS-TAICHI datasets were used for testing, and compared with AGCN, traditional ST-GCN, and LSTM. AGCN combines the advantages of graph convolutional networks, fully considering the relationships between joints, and has strong adaptability and robustness, demonstrating

good performance in processing human motion data. The results are shown in Figure 9. From the graph, among the four methods, the classification accuracy of the model in this study was the highest on both datasets, with values of 0.914 and 0.983, respectively. The outcomes indicated that the improved ST-GCN-based action recognition model had high accuracy in action recognition, and had certain feasibility and effectiveness.

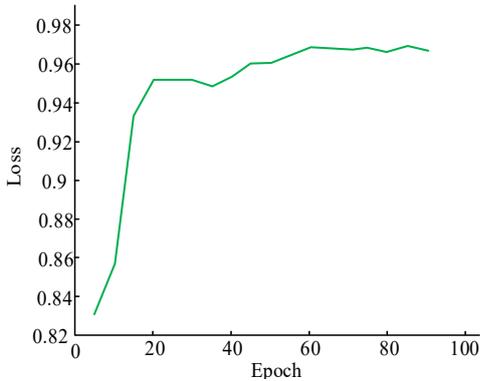
**Figure 8** Performance variation curve of the model, (a) loss value curve (b) Top-1 curve (c) Top-5 curve (see online version for colours)



(a)



(b)

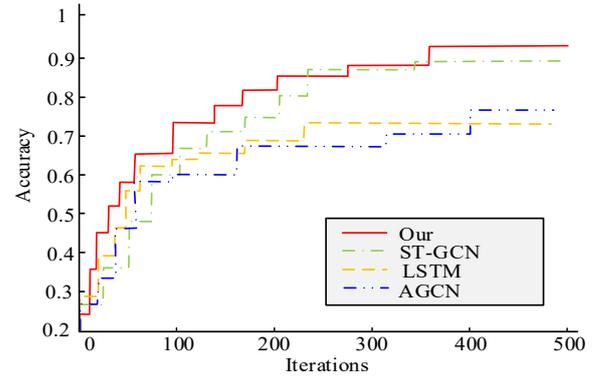


(c)

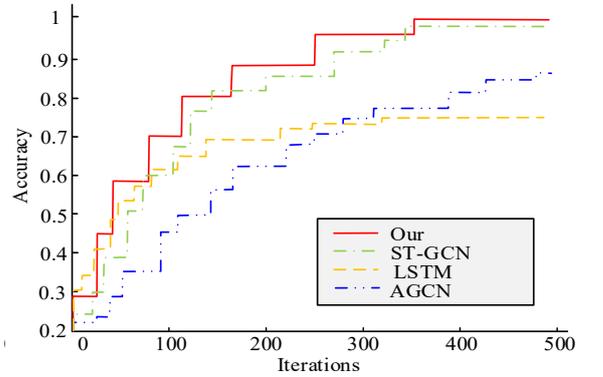
To further prove the effect of the proposed action recognition model, the NTU-RGB D dataset was applied to perform 50 recognitions on four methods, and the error rate of each recognition result was recorded. The results are

shown in Figure 10. From the graph, among the four methods, the recognition error rate of the model in this study was the lowest overall, with the highest and lowest recognition error rates of 0.09 and 0.01, and the average recognition error rate of 0.03. In addition, the average recognition error rate of AGCN was the highest, at 0.14, which was 0.11 higher than the model in this study. The findings expressed that the action recognition model based on improved ST-GCN could effectively recognise action capture data in VR and had good performance.

**Figure 9** Action classification accuracy of four methods on two datasets, (a) NTU-RGB D dataset (b) UMONS-TAICHI dataset (see online version for colours)

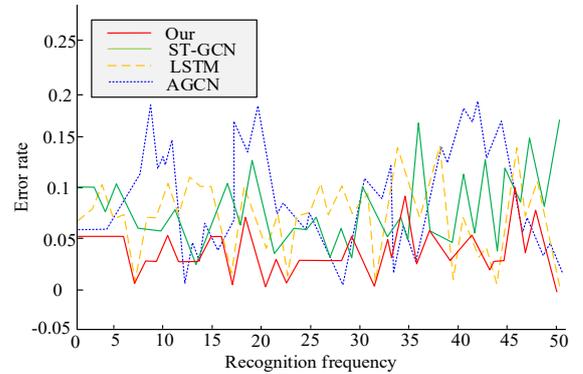


(a)



(b)

**Figure 10** Comparison of error rates of recognition results using four methods



**Table 2** Comparison results of average teaching scores

Group	Strength	Range of motion	Coherence	Degree of standardisation	Class participation/%
Experimental group	85.4	85.2	84.2	85.5	93.33
Control group 1	78.6	81.3	80.9	76.1	86.67
Control group 2	80.2	83.4	80.7	78.2	93.33

Finally, to evidence the application effect of the proposed motion capture model and motion recognition model in college aerobics training, 90 students from a certain university aerobics course were selected for a three-month research analysis. Broken into experimental group, control group 1, and control group 2, with 30 students in each group. The experimental group was trained using the model proposed by the research institute, control group 1 was trained using traditional methods, and control group 2 was trained using traditional VR motion capture technology. The comparison results of the average teaching scores between the three groups are indicated in Table 2. The experimental group achieved higher scores in four aspects: strength, action amplitude, coherence, and standardisation, especially in terms of standardisation, which was 9.4 points higher than the control group 1 and 7.3 points higher than control group 2. Therefore, compared to traditional training methods and VR training, VR based fine motion capture algorithms are more conducive to improving the accuracy of students' movements. At the same time, the higher the degree of standardisation of movements, the more effective it is to prevent injuries caused by inadequate movements. In addition, the classroom participation rate of students in the experimental group during aerobics training was 93.33%, which was higher than that of the control group 1 at 6.66%. The results indicate that the model proposed by the research institute has a promoting effect on the skill development of students, and can effectively improve the training effect of college aerobics, thereby improving the learning outcomes of students.

## 5 Conclusions

With the advent of the intelligent era, various intelligent technologies such as VR have been applied to sports training. In response to the noise and incompleteness issues of VR technology in motion capture in college aerobics training, a fine motion capture model based on IICP algorithm and an improved ST-GCN motion recognition model were studied and built. The results indicated that the IICP algorithm had the fastest iteration speed on both datasets, with an error reduction of 40.16% and 36.27% compared to the traditional ICP algorithm at 20 iterations. The precision of the fine motion capture model based on the IICP algorithm was the highest, with Top-1 being 36.5% and Top-5 being 59.4%, and the time consumption was relatively short. The action recognition model based on improved ST-GCN reached its peak accuracy of Top-1 and

Top-5 at around the 90th epoch, with 90.1% and 99.0%, respectively. The classification accuracy was the highest on both datasets, with 0.914 and 0.983, respectively. The recognition error rate was the highest at 0.09 and the lowest at 0.01, and the average recognition error rate was 0.03. The proposed model can effectively improve the training effect of college aerobics. The experimental group achieved higher scores in four aspects: intensity, action amplitude, coherence, and standardisation, especially in terms of standardisation, which was 9.4 points higher than the control group. In summary, the proposed method has certain feasibility and effectiveness. However, the proposed method has certain limitations for the cross-subject evaluation criteria in the NTU-RGB D dataset, which may affect the practical application effectiveness of action recognition models. Therefore, future research should focus on how to better optimise noise, and attempts can be made to improve the structure of GCN or introduce other deep learning techniques to further enhance the performance of the model, in order to adapt to various fitness levels, exercises, or individual specific requirements.

## Fundings

The research is supported by: A Study on the Integration of Chinese Sports Spirit into the Ideological and Political Courses of Adult Higher Education Physical Education Major (Continuing Education Teaching Reform Project of Yan'an University). An Innovative Research on Post-match Management of Sports Venues and Stadiums in Shaanxi Province for the 14th National Games (Shaanxi Provincial Department of Science and Technology Project 2024ZC-YBXM-046).

## References

- Abich IV, J., Parker, J., Murphy, J.S. and Eudy, M. (2021) 'A review of the evidence for training effectiveness with virtual reality technology', *Virtual Reality*, Vol. 25, No. 4, pp.919–933.
- Chang, E., Kim, H.T. and Yoo, B. (2020) 'Virtual reality sickness: a review of causes and measurements', *International Journal of Human-Computer Interaction*, Vol. 36, No. 17, pp.1658–1682.
- Chatzitofis, A., Zarpalas, D., Daras, P. and Kollias, S. (2021) DeMoCap: low-cost marker-based motion capture', *International Journal of Computer Vision*, Vol. 129, No. 12, pp.3338–3366.

- Chen, L., Huang, Y., Lu, T., Dang, S. and Kong, Z. (2022) 'Metering equipment running error estimation model based on genetic optimized LM algorithm', *Journal of Computational Methods in Sciences and Engineering*, Vol. 22, No. 1, pp.197–205.
- Fussell, S.G. and Truong, D. (2022) 'Using virtual reality for dynamic learning: an extended technology acceptance model', *Virtual Reality*, Vol. 26, No. 1, pp.249–267.
- Gao, Q., Li, J., Zhu, Y., Wang, S., Liufu, J. and Liu, J. (2023) 'Hand gesture teleoperation for dexterous manipulators in space station by using monocular hand motion capture', *Acta Astronautica*, Vol. 204, No. 11, pp.630–639.
- Han, D.I.D., Bergs, Y. and Moorhouse, N. (2022) 'Virtual reality consumer experience escapes: preparing for the metaverse', *Virtual Reality*, Vol. 26, No. 4, pp.1443–1458.
- Kadirvelu, B., Gavriel, C., Nageswaran, S., Chan, J.P.K., Nethisinghe, S., Athanasopoulos, S., Ricotti, V., Voit, T., Giunti, P., Festenstein, R. and Faisa, A. (2023) 'A wearable motion capture suit and machine learning predict disease progression in Friedreich's ataxia', *Nature Medicine*, Vol. 29, No. 1, pp.86–94.
- Khmag, A. (2023) 'Additive Gaussian noise removal based on generative adversarial network model and semi-soft thresholding approach', *Multimedia Tools and Applications*, Vol. 82, No. 5, pp.7757–7777.
- Kurobe, A., Sekikawa, Y., Ishikawa, K. and Saito, H. (2020) 'Cornsnet: 3d point cloud registration by deep neural network', *IEEE Robotics and Automation Letters*, Vol. 5, No. 3, pp.3960–3966.
- Li, Y., Liu, Y., Guo, Y.Z., Liao, X.F., Hu, B and Yu, T. (2021) 'Spatio-temporal-spectral hierarchical graph convolutional network with semisupervised active learning for patient-specific seizure prediction', *IEEE Transactions on Cybernetics*, Vol. 52, No. 11, pp.12189–12204.
- Mäkinen, H., Haavisto, E., Havola, S. and Koivisto, J.M. (2022) 'User experiences of virtual reality technologies for healthcare in learning: an integrative review', *Behaviour & Information Technology*, Vol. 41, No. 1, pp.1–17.
- McGovern, E., Moreira, G. and Luna-Nevarez, C. (2020) 'An application of virtual reality in education: can this technology enhance the quality of students' learning experience?', *Journal of Education for Business*, Vol. 95, No. 7, pp.490–496.
- Mehta, D., Sotnychenko, O., Mueller, F., Xu, W., Elgharib, M., Fua, P. and Seidel, H.P. (2020) 'XNect: real-time multi-person 3D motion capture with a single RGB camera', *ACM Transactions on Graphics (TOG)*, Vol. 39, No. 4, pp.1–17.
- Shimada, S., Golyanik, V., Xu, W. and Theobalt, C. (2020) 'Physcap: physically plausible monocular 3d motion capture in real time', *ACM Transactions on Graphics (TOG)*, Vol. 39, No. 6, pp.1–16.
- Tang, Y.M., Ng, G.W.Y., Chia, N.H., So, E.H.K., Wu, C.H. and Ip, W.H. (2021) 'Application of virtual reality (VR) technology for medical practitioners in type and screen (T&S) training', *Journal of Computer Assisted Learning*, Vol. 37, No. 2, pp.359–369.
- Wang, X., Cheng, M., Eaton, J. et al. (2022) 'Fake node attacks on graph convolutional networks', *Journal of Computational and Cognitive Engineering*, Vol. 1, No. 4, pp.165–173.
- Wang, Y., Gui, G., Ohtsuki, T. and Adachi, F. (2021) 'Multi-task learning for generalized automatic modulation classification under non-Gaussian noise with varying SNR conditions', *IEEE Transactions on Wireless Communications*, Vol. 20, No. 6, pp.3587–3596.
- Wierschem, D.C., Jimenez, J.A. and Mendez Mediavilla, F.A. (2020) 'A motion capture system for the study of human manufacturing repetitive motions', *The International Journal of Advanced Manufacturing Technology*, Vol. 110, Nos. 3–4, pp.813–827.
- Xiao, G., Wang, R., Zhang, C. and Ni, A. (2021) 'Demand prediction for a public bike sharing program based on spatio-temporal graph convolutional networks', *Multimedia Tools and Applications*, Vol. 80, No. 15, pp.22907–22925.
- Zhang, H., Xu, Y., Liu, Q. and Li, Y. (2022) 'Solving Fokker-Planck equations using deep KD-tree with a small amount of data', *Nonlinear Dynamics*, Vol. 108, No. 4, pp.4029–4043.
- Zhang, J., Yao, Y. and Deng, B. (2021) 'Fast and robust iterative closest point', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 7, pp.3450–3466.