



## OPEN Table tennis motion recognition based on the bat trajectory using varying-length-input convolutional neural networks

Jun Zhang<sup>1,4</sup>, Yuanshi Ren<sup>2</sup>, Liyue Lin<sup>3</sup>, Yu Xing<sup>4</sup> & Jie Ren<sup>2</sup>✉

Action recognition has been applied in fields such as smart homes, gaming, traffic management, and security monitoring. Motion recognition is helpful for biomechanical analysis, auxiliary training systems, table tennis robots, motion-sensing games, virtual reality and other fields. In our study, we collected data on table tennis skill motion, created the TTMD6 dataset, and analyzed the characteristics of table tennis paddle trajectories. We propose a motion recognition algorithm to recognize paddle trajectories. Other research has used multijoint data to identify actions, while we use only the paddle trajectory to recognize table tennis skill motions, accelerating the speed of motion recognition. Therefore, it is feasible to use paddle trajectories to recognize table tennis skill motions.

A time series is a collection of data points arranged in chronological order. Since time series data can reflect dynamic changes, feature extraction and extraction algorithms for time series data have become important research topics in many fields, such as biomedical engineering<sup>1,2</sup>, speech detection<sup>3</sup>, finance and other fields<sup>4,5</sup>. With the successful application of machine learning and deep learning<sup>6,7</sup> in feature recognition, the problem of feature extraction based on time series has new vitality. Traditional methods usually use dynamic time warping (DTW)<sup>8</sup> or shapelet transforms<sup>9</sup> to extract the discriminant features of the original time series. These two methods separate the feature extraction module from the classification module, which limits algorithm accuracy. In addition, most of the existing methods do not solve the problem that time series may have different time scales.

Dynamic time warping (DTW) involves expanding and shortening two time series to obtain two time series of the same length. For example, in speech recognition, because the time series of the source speech and the target speech to be processed are different lengths, DTW can adjust the time series of different lengths to the same size<sup>10</sup>.

The shapelet transform can be used for time series classification<sup>11</sup> and data enhancement<sup>12</sup>. Cui et al.<sup>12</sup> applied the shapelet transformation method to all time series in a given dataset when studying the multiscale convolutional neural network model and regarded all data slices as independent training instances.

The weaknesses and advantages of the four methods can be found in Table 1.

The duration of table tennis motion differs due to differences in athlete motion ability and task goals, which can also cause problems with different durations of table tennis motion recognition. To solve the time series and data volume problems, we use an infrared motion capture system to collect the time series of six popular motions (forehand attack, forehand drive, forehand push, backhand attack, backhand drive, and backhand push) in table tennis and create the time series dataset TTMD6 (Table Tennis MOCAP Dataset). The TTMD6 dataset contains 9,000 strokes of 6 table tennis motions, independent human joint data and independent paddle center of gravity data. We used MATLAB to create a skeletal animation of the movement trajectory of the paddles, recruited table tennis players for recognition and achieved a recognition accuracy of 92.63%. In this study, we propose a varying-length-input convolutional neural network framework to recognize paddle trajectories and obtain a recognition accuracy of 99.78%, which is much greater than that of athletes.

<sup>1</sup>School of Exercise and Health, Shanghai University of Sport, Shanghai 200438, China. <sup>2</sup>China Table Tennis College, Shanghai University of Sport, Shanghai 200438, China. <sup>3</sup>School of Psychology, Shanghai University of Sport, Shanghai 200438, China. <sup>4</sup>School of Sport Communication and Information Technology, Shandong Sport University, Jinan 250102, Shandong, China. ✉email: renjie@sus.edu.cn

No	Method	Weaknesses and advantages
1	Truncation and zero filling	Easy to implement, but can cause data distortion
2	Dynamic time warping	A source object is needed, which is not suitable for motion recognition
3	Shapelets	Time series of different lengths can be sliced to the same length using window slicing. However, the size of the window slicing is crucial to the recognition result
4	Varying-length-input	Data preprocessing is divided into three cases, and our method preserves the original characteristics of the data to the greatest extent

**Table 1.** The weaknesses and advantages of the four methods.

## Convolutional neural networks

Neural networks are important methods for studying artificial intelligence. At present, one of the most popular neural networks is the convolutional neural network (CNN)<sup>13–15</sup>. CNNs were first used in the ImageNet Large-scale Visual Identification Challenge (LSVRC) in 2012 by Krizhevsky et al.<sup>16</sup>. By adopting ReLU+dropout technology, researchers achieved the best classification results at that time (the network structure called AlexNet), which made CNNs increasingly valuable to researchers. Compared with AlexNet, Szegedy et al.<sup>15</sup> greatly increased the depth of CNNs and proposed a CNN structure with more than 20 layers (the model called GoogLeNet). The main advantage of this structure is that it improves the utilization of computing resources and has greater accuracy.

The AlexNet, GoogLeNet and VGG models all achieved good results in the ImageNet competition. However, they can accept only a fixed-size input. In fact, the convolutional layer of a CNN does not require a fixed-size input; it can generate feature surfaces of arbitrary size, but its fully connected layer requires a fixed-length input, so the restriction of consistent input length of CNNs stems from its fully connected layer<sup>17</sup>. The input image needs to be cropped or scaled to obtain a fixed-size input, but such a transformation will destroy the aspect ratio and complete the information of the input image to affect the recognition accuracy. He et al.<sup>17</sup> proposed an SPPNet model in which a spatial pyramid pooling (SPP) layer is added between the last convolutional layer and the first fully connected layer of CNNs. The SPP layer can make inputs of different sizes produce outputs of the same size, which breaks the restriction that the input of the CNN model is fixed in size.

In the<sup>6,7</sup> network model with very deep layers, there are gradient diffusion problems and degradation problems. Batch normalization (BN) is an effective method for solving the gradient diffusion problem<sup>18</sup>. He et al.<sup>14</sup> proposed residual networks (ResNets) to solve the degradation problem.

As AIs have developed, CNNs have been used in speech recognition<sup>19</sup>, face recognition<sup>20</sup>, image recognition<sup>14</sup>, motion analysis<sup>7</sup>, motion analysis<sup>21</sup>, image segmentation<sup>22</sup>, natural language processing<sup>23</sup> and EEG analysis<sup>5</sup>.

## Table tennis MOCAP dataset

### Participants

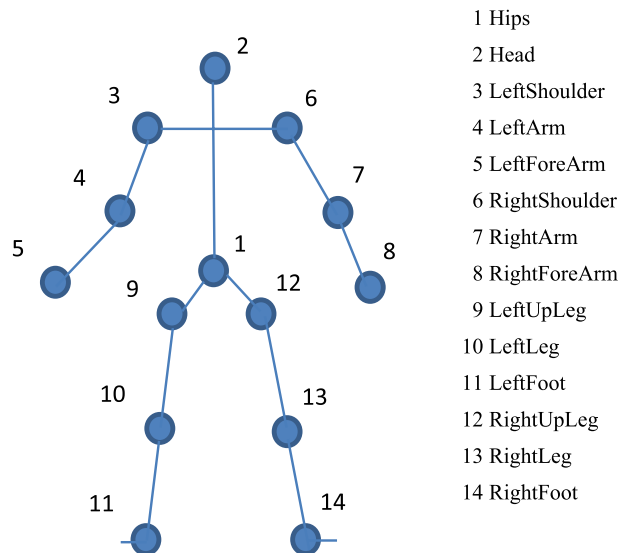
The thirty professional table tennis players (17 males and 13 females) who volunteered to participate in the data acquisition were all undergraduates of the China Table Tennis College of Shanghai University of Sport, and they were all national division I athletes, national division II athletes, or national master sportsmen who were right-handed. At least three months before the experiment, all participants had no injuries or deformities to the lower extremities. Moreover, they were required not to consume any coffee or food with related additives, such as caffeine, prior to 6 h of the experiment and signed a written informed consent. The experiment was approved by the Ethics Committee of Shanghai University of Sport (ethics approval reference: 102772019RT030). All methods were performed in accordance with ethical approval.

### Apparatus

The experiment was carried out on a standard-sized table. The table tennis ball was sent to the participants by a ball projection machine (Y&T® V-989H, Zhongshan, China) on the opposite side of the table. The serving frequency was set at 30 Hz, and the ball speed was 6.5 m/s. All the subjects used the same paddle to complete the task during the experiment. All motion was captured by both the full HD camera and the MOCAP system. The full HD camera is situated behind the ball projection machine. The camera is 2.30 m above the ground, and the sampling frequency is 25 Hz, which is used to record the frontal shot of the participant. The MOCAP system (CMTractor 2.0; Shanghai Qingtong Vision Co., Ltd., China) consists of 20 infrared cameras fixed on the ceiling of the laboratory with a capture frequency of 120 Hz. A camera array was used to track the 3D positions of 38 reflective marks placed on the participants' bodies.

Each participant needed to perform six skill motions. The experiment was divided into two stages: 1) the participant's adaptation stage before the experiment and 2) the motion capture stage. Before the test, the participants were told that the table tennis ball would be launched from the ball projection machine on the opposite side and that they would have to hit the cross-court ball. After the adaptation stage (5 serves), the participants complete the data acquisition of 55 shots without further instructions.

The 14-joint skeleton diagram is similar to the 15-joint human body model of S. Litvak<sup>24</sup>. The only difference is that the former does not have the shoulder center, which is located between the left shoulder and the right shoulder, compared with the latter. The shoulder center can be calculated. We collect 14 joint time series data and the center of gravity of the table tennis paddle to construct the TTMD6 dataset. The 14 joints used by TTMD6 are shown in Fig. 1.



**Figure 1.** Skeleton diagram of 14 joints.

### Data processing

The complete table tennis motion includes four stages: backward swing, stroke, follow-through, and recovery. The motion capture system recorded 55 shots of the participants at a time, and the 55 shots were saved in a file. Each file saves nearly 20,000 frames of data, and a single motion needs to be extracted. First, the analyzer software of the motion capture system was used to manually mark the starting frame of each motion, and then the motion extraction tool written in C# was used to extract a single motion. The motion visualization tool written in MATLAB is an alternative tool to the analyzer software of the motion capture system. This tool can display motions in the form of skeleton diagrams.

### Athletes recognize table tennis motions

Table tennis is one of the most popular paddle sports in the world; however, with more than 100 years of development, many different types of skill motions, such as attack, drive, long push, short push, loop, push-and-block, drop shot, off-table chop, smash, and lift, have been developed<sup>25</sup>. Each of these pongs hit a different position on the incoming ping pong. Accurate recognition of these motion patterns is highly important for the development of table tennis robots, daily training of athletes and somatosensory games.

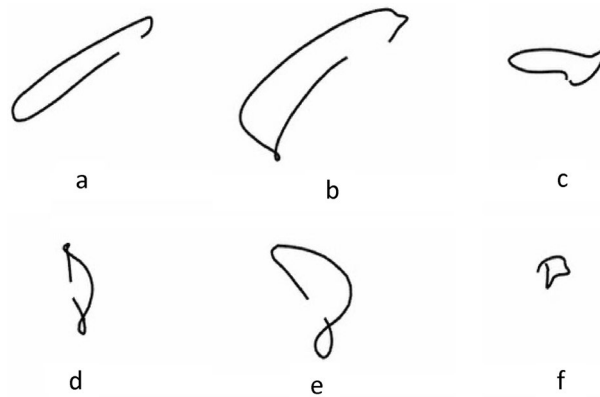
Acquiring table tennis skills depends on the guidance of coaches. Experienced coaches can determine the type of motion according to the type of body movement performed by the athlete. Athlete recognition was introduced to test the recognition accuracy of professional athletes. Our research purpose was to identify an algorithm that exceeds the recognition accuracy of traditional methods for athletes to prepare for the next step of constructing an auxiliary training system.

Unlike other motion recognition methods, in table tennis, players hit table tennis with a paddle. The movement trajectory of a paddle is the embodiment of the coordination of all joints, and an athlete can control the incoming ping pong through the movement of the paddle. The paddle was located on the extension line of the distal end of the upper limb. Compared with the shoulder, elbow and wrist, the paddle has a greater range of motion and a faster movement speed, which can better reflect the movement characteristics of table tennis. The paddle trajectories of the six skill motions are shown in Fig. 2. Experienced coaches and athletes can predict the type of motion, the trajectory of the ping pong ball and its landing point from the movements of the opponent's body and the trajectory of the paddle. The purpose of this experiment was to demonstrate the feasibility of motion recognition through a paddle trajectory.

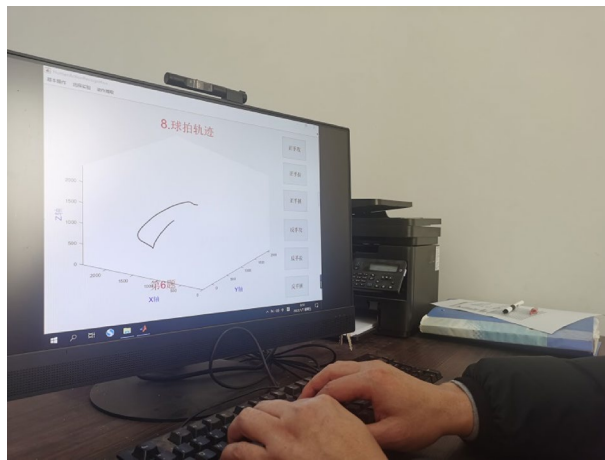
We selected 30 motions (5 for each skill motion) and used MATLAB software (version 2018A; MathWorks, Inc., USA) to generate animations, and the interval between motions was 2 s. Q, W, E, I, O, and P on the keyboard correspond to forehand attack, forehand drive, forehand push, backhand attack, backhand drive, and backhand push, respectively. After recognizing the motion, the athletes pressed the key on the keyboard as quickly as possible and entered the recognition result. The program records the recognition result of the athletes' input and the time it takes for them to recognize the motion (the first frame of the motion appears until the participant presses the key). We recruited 40 professional table tennis athletes from Shanghai University of Sport to participate in this study. Figure 3 shows the athlete's recognition of the paddle trajectory on the computer.

### Varying-length-input convolution neural networks

The basic CNN structure consists of an input layer, a convolutional layer, a pooling layer, a fully connected layer and an output layer. Generally, several convolutional layers and pooling layers are contained, convolutional layers and pooling layers are set alternately, and then a convolutional layer is connected after the pooling layer<sup>26</sup>.



**Figure 2.** Bat trajectory of table tennis motion (a) forehand attack (b) forehand drive (c) forehand push (d) backhand attack (e) backhand drive (f) backhand push.



**Figure 3.** The athlete recognizes bat trajectory on PC.

The convolutional layer of CNNs does not need a fixed-size input, but its fully connected layer does. Therefore, the limitation that the input sizes of CNNs must be consistent is due to the fully connected layer used<sup>17</sup>. The input image needs to be cropped or scaled in the field of image recognition to obtain a fixed-size input; however, such a transformation will destroy the aspect ratio and complete information of the image, thereby affecting the recognition accuracy. He et al.<sup>17</sup> proposed an SPP-net model, which adds a spatial pyramid pooling (SPP) layer between the last convolutional layer and the first fully connected CNN layer. The SPP layer can make different sizes of CNN inputs produce the same size of output, which breaks the previous limit of CNN model input fixed, and the improved CNN model has a faster training speed.

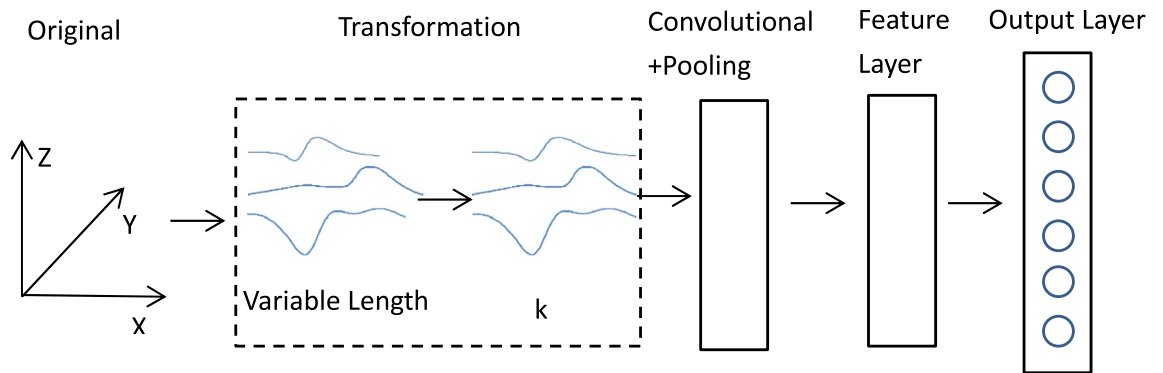
The duration of table tennis motion varies due to the athlete's ability to perform the motion and the target of the task. When training the model, DTW<sup>10,27</sup> and Shapelets<sup>9,28</sup> are used to truncate and adjust the time series to a fixed length, but such transformation will destroy the complete information of the time series. To solve these problems, we design a time series transformation layer in the input layer and convolution layer and integrate the transformation layer and CNN into a framework, as shown in Fig. 4.

Interpolation is a common method for data preprocessing. The most commonly used interpolation methods include nearest neighbor interpolation, linear interpolation and spline interpolation. Spline interpolation is a commonly used interpolation method for obtaining smooth curves. The cubic spline method is one of the most widely used interpolation methods. The basic idea of cubic spline interpolation is to divide  $n$  intervals within  $[a, b]$  and perform cubic spline interpolation fitting for each interval to generate a smoother curve.

In the time series transformation,  $l$  represents the actual length of the time series (the file name contains the actual length information), and  $k$  represents the length of the transformed time series. When dealing with a time series, we can divide it into three cases:  $k > l$ ,  $k = l$ , and  $k < l$ .

When  $l > k$ , we downsample the raw data to a fixed length. Assuming that the time series  $T = \{t_1, \dots, t_n\}$ , the downsampling ratio is  $l/k$ , the  $i$ th element of the transformed time series,  $t_i = x[\lceil i \cdot l/k \rceil]$ ,  $0 < i < k$ , and  $\lceil \cdot \rceil$  denoting the operation of the ceiling.

When  $l = k$ , there is no change in the time series.



**Figure 4.** Overview architecture for 3D time series classification.

When  $l < k$ , we use cubic spline interpolation<sup>29</sup> to interpolate the time series. Formula 1 is a mathematical formula describing cubic spline interpolation. In this article, the cubic spline interpolation is implemented using Python's SciPy toolkit.

$$\begin{cases} S(x) \in C^2[a, b] \\ S(x_i) = y_i \\ S(x) = a_3x^3 + a_2x^2 + a_1x + a_0 \end{cases} \quad (1)$$

In the extreme case, if  $k$  is less than the length of the shortest time series, all the time series are downsampled only. The time series length of a group of motions has a range. For example, a participant needs at most 212 frames and at least 169 frames to complete a forehand attack. When  $k < 169$ , we do not need to consider the first two cases but only the third case. However, if  $k$  is too small, it will affect the recognition accuracy. When all the time series are of the same length, the neural network can be trained.

## Discussion

To produce the TTMD6 dataset, we recruited 30 professional table tennis athletes from Shanghai University of Sport to participate in this test. The age range of the 30 participants was 14 to 24 years, which allowed more realistic changes in motion quality to occur. Since the infrared camera is fixed to the ceiling and the TTMD6 dataset is acquired in the laboratory, we evaluate environmental instability under various background conditions. The large differences in views and participants make cross-evaluation possible.

K-fold cross-validation<sup>8</sup> is a strategy for dividing datasets. When the dataset is small, using all the data to train the model will easily lead to overfitting. K-fold cross-validation can solve the problem of overfitting the small dataset training model to a certain extent. K-fold cross-validation can also be used for model evaluation and model selection. We used fivefold cross-validation to divide the TTMD6 dataset and adopted a CNN to test the stability of TTMD6, as shown in Table 2.

The stability of the TTMD6 dataset was tested by fivefold cross-validation. In our research, we divided the whole dataset into 5 parts, one of which was used as the test set and the other as the training set. The training set was used to train the model, and the test set was used to test the trained model. The whole process was performed five times. As shown in Table 2, the results of the five experiments are relatively stable, with the lowest accuracy rate of 95.83% and an average recognition accuracy rate of 98.38%. The test results show that the TTMD6 dataset has good stability.

Compared to body joints, the paddle is located on the extension line of the distal end of the upper limb, and the movement trajectory of the paddle can more effectively reflect the characteristics of each skill motion. Experienced coaches and athletes can predict the motion type, trajectory and landing point of the ping pong through the opponent's body movements and trajectory. We recruited 40 professional table tennis athletes from Shanghai University of Sport to perform the animation of paddle trajectories. Table 3 shows the confusion matrix of athletes who recognized table tennis motions.

A confusion matrix, also known as an error matrix, is a standard format for evaluating accuracy and is expressed in the form of a matrix with  $n$  rows and  $n$  columns. Each column of the confusion matrix represents

No	Part1	Part2	Part3	Part4	Part5	Accuracy
1	Train	Train	Train	Train	Test	96.16%
2	Train	Train	Train	Test	Train	98.44%
3	Train	Train	Test	Train	Train	98.94%
4	Train	Test	Train	Train	Train	99.56%
5	Test	Train	Train	Train	Train	95.83%

**Table 2.** Results of fivefold cross-validation.

Motion	FA	FD	FP	BA	BD	BP
FA	1332	160	0	8	0	0
FD	63	1429	0	0	8	0
FP	14	14	1464	0	0	8
BA	0	8	0	1218	250	24
BD	8	21	14	48	1409	0
BP	0	0	10	0	8	1482

**Table 3.** The confusion matrix of participants recognizing table tennis motions.

the predicted category, and the values in each column represent the number of data points predicted as the category. Each row represents the true category of the data, and the total number of data points in each row represents the number of data points in that category. Table 3 shows that the recognition accuracy of forehand push and backhand push is the highest, 97.7% and 98.6%, respectively, followed by forehand drive and backhand drive, and the lowest recognition accuracy is for forehand and backhand attacks. As it is easy to confuse forehand attacks with forehand drives and backhand attacks with backhand drives, most of the recognition errors made by athletes are mainly due to confusion from attacks and drives whose essentials are familiar. Figure 3 shows that the strength, speed, and movement range of the drives are greater than those of the attack. A drive with a slightly smaller movement range is easily recognized as an attack, while an attack with a slightly larger movement range is easily recognized as an attack. The average accuracy of athlete recognition by paddle trajectory was 92.63%. Therefore, it is feasible to use the paddle trajectory for motion recognition. Using the paddle trajectory to train the neural network greatly reduces the quantity of data and the time required to train the model.

With the development of AIs, human action recognition has made many breakthroughs in recent years. However, there is not much work on table tennis motion recognition, and the methods are mainly integrated into traditional machine learning. In addition, traditional CNNs generally use truncation and zero-filling methods to process time series of different lengths, which leads to data distortion and affects the recognition effect of the algorithm. The method we propose adjusts time series of different lengths to the same length, and the recognition accuracy of this method is higher than that of traditional processing methods. To solve the problems of time series length and data quantity, we use paddle trajectories to recognize table tennis motions, propose a varying-length-input CNN framework, and compare the results with those of traditional recognition algorithms. The comparison results are shown in Table 4.

We use six algorithms to recognize table tennis motions in the TTMD dataset.

The accuracy of our proposed algorithm in recognizing motions through paddle trajectories is 99.78%, which is 2.12% greater than that of standard three-layer CNNs. The confusion matrix is shown in Fig. 5. The LSTM algorithm accuracy in recognizing motions through the paddle trajectory is 76.6%. The accuracy of the FNN algorithm in recognizing motions through paddle trajectories is 98.4%. The accuracy of professional athletes in recognizing motions through paddle trajectories is 92.63%. The accuracy of our proposed algorithm is 99.78%, which is much greater than that of athletes. Both the dataset and the recognition algorithms in this paper are available for download.

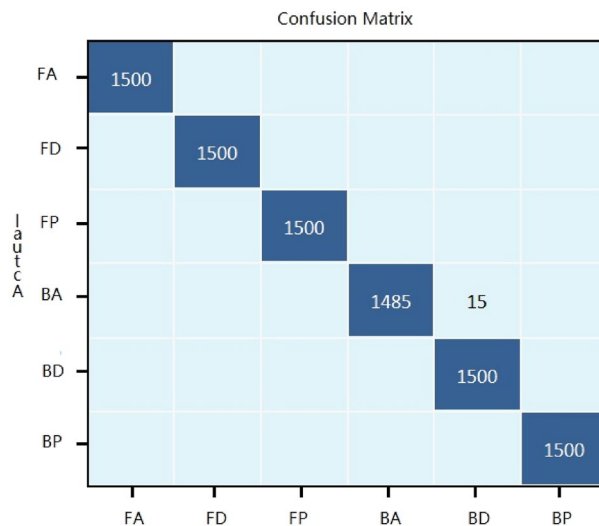
## Conclusions

The results of this study showed that the TTMD6 dataset that we created is stable and can be used for research on table tennis motion recognition and biomechanics. We obtained a recognition accuracy of 92.63% by using the paddle trajectory to create an animation and recruiting 40 professional table tennis athletes for motion recognition, indicating that it is feasible to use the paddle trajectory to recognize table tennis motions. We propose a varying-length-input CNN model, use our model to recognize the racket trajectory time series and obtain a recognition accuracy of 99.78%, which is higher than that of traditional recognition algorithms and athlete recognition. The trained model can be used in auxiliary training systems, table tennis robots, motion-sensing games, virtual reality and other fields.

No	Algorithm	Accuracy
1	Varying-length-input CNN(paddle)	99.78%
2	Standard CNN(paddle) <sup>30</sup>	97.66%
3	Standard CNN(joints) <sup>30</sup>	96.4%
4	LSTM <sup>31</sup>	76.6%
5	FNN <sup>32</sup>	98.4%
6	Athlete recognition	92.63%

**Table 4.** The performance of the six algorithms on the TTMD6 dataset.





**Figure 5.** Confusion Matrix of the varying-length-input convolution neural networks.

### Data availability

Data can be obtained from the URL: <https://pan.baidu.com/s/1lDdqULp6gsBovY7BaUJ9Q?pwd=epsk>

Received: 16 October 2023; Accepted: 8 February 2024

Published online: 12 February 2024

### References

- Junxiao, B., Culin, B., Xiang, Z. & Jinli, W. Deep learning algorithm in biomedical engineering in intelligent automatic processing and analysis of sports images. *Wirel. Commun. Mob. Comput.* **2022**, 1–10 (2022).
- Bao, J., Tuo, M., Hou, T. M., Li, Y. X. & Wang, Q. Research on intelligent medical engineering analysis and decision based on deep learning. *Int. J. Web Serv. Res. (IJWSR)* **19**, 1–9 (2022).
- Zhenyu, N. Voice detection and deep learning algorithms application in remote english translation classroom monitoring. *Mob. Inf. Syst.* **2022**, 1–10 (2022).
- Landa, V. & Reuveni, Y. Low-dimensional convolutional neural network for solar flares GOES time-series classification. *Astrophys. J. Suppl. Ser.* **258**, 12 (2022).
- Sablok, S., Gururaj, G., Shaikh, N., Shiksha, I. & Choudhary, A. R. in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*.
- Li, Y., Cao, J., Xu, Y., Zhu, L. & Dong, Z. Y. Deep learning based on Transformer architecture for power system short-term voltage stability assessment with class imbalance. *Renew. Sustain. Energy Rev.* **189**, 113913. <https://doi.org/10.1016/j.rser.2023.113913> (2024).
- Yang, L., Xueyi, W., Zelin, Z. & Fang, D. Deep learning based data augmentation for large-scale mineral image recognition and classification. *Miner. Eng.* **204**, 108411 (2023).
- Jacob, W. & Power, S. D. k-fold cross-validation can significantly over-estimate true classification accuracy in common EEG-based passive BCI experimental designs: an empirical investigation. *Sensors* **23**, 6077 (2023).
- Hills, J., Lines, J., Baranauskas, E., Mapp, J. & Bagnall, A. Classification of time series by shapelet transformation. *Data Min. Knowl. Disc.* **28**, 851–881 (2014).
- Sun, L., Li, K., Wang, H., Kang, S. & Meng, H. in *2016 IEEE International Conference on Multimedia and Expo (ICME)*. 1–6 (IEEE).
- Chen, J., Wan, Y., Wang, X. & Xuan, Y. Learning-based shapelets discovery by feature selection for time series classification. *Appl. Intell.* **52**, 1–16 (2022).
- Cui, Z., Chen, W. & Chen, Y. Multi-scale convolutional neural networks for time series classification. [arXiv:1603.06995](https://arxiv.org/abs/1603.06995) (2016).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- He, K., Zhang, X., Ren, S. & Sun, J. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.
- Szegedy, C. et al. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1–9.
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural. Inf. Process. Syst.* **25**, 1097–1105 (2012).
- He, K., Zhang, X., Ren, S. & Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 1904–1916 (2015).
- Ioffe, S. & Szegedy, C. in *International Conference on Machine Learning*. 448–456 (PMLR).
- Qayyum, A., Arefeen, A. & Shahnaz, C. in *IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICSCON 2019)*.
- Taigman, Y., Yang, M., Ranzato, M. A. & Wolf, L. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1701–1708.
- Ahmedt-Aristizabal, D., Nguyen, K., Denman, S., Sridharan, S. & Fookes, C. in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*.
- Sivanarayana, G. V., Naveen Kumar, K., Srinivas, Y. & Raj Kumar, G. V. S. Review on the Methodologies for Image Segmentation Based on CNN. In *Communication Software and Networks: Proceedings of INDIA 2019* (eds Satapathy, S. C. et al.) 165–175 (Springer Singapore, 2021). [https://doi.org/10.1007/978-981-15-5397-4\\_18](https://doi.org/10.1007/978-981-15-5397-4_18).
- Wróbel, K., Karwatowski, M., Wielgosz, M., Pietróń, M. & Wiatr, K. Compression of convolutional neural network for natural language processing. *Comput. Sci.* <https://doi.org/10.7494/csci.2020.21.1.3375> (2020).
- Litvak, S. Learning-based pose estimation from depth maps (Google Patents, 2013).

25. McAfee, R. *Table Tennis: Steps to Success* (Human Kinetics, 2009).
26. LeCun, Y., Kavukcuoglu, K. & Farabet, C. in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*. 253–256 (IEEE).
27. Cai, X., Xu, T., Yi, J., Huang, J. & Rajasekaran, S. Dtwnet: A dynamic time warping network. *Advances in neural information processing systems* **32** (2019).
28. Grabocka, J., Schilling, N., Wistuba, M. & Schmidt-Thieme, L. in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 392–401.
29. Liu, M., Zeng, A., Lai, Q. & Xu, Q. Time series is a special sequence: Forecasting with sample convolution and interaction. (2021).
30. Tabrizi, S. S., Pashazadeh, S. & Javani, V. Comparative study of table tennis forehand strokes classification using deep learning and SVM. *IEEE Sens. J.* **20**, 13552–13561 (2020).
31. Karim, F., Majumdar, S., Darabi, H. & Chen, S. LSTM fully convolutional networks for time series classification. *IEEE access* **6**, 1662–1669 (2017).
32. Dokur, Z. & Olmez, T. Classification of hyperspectral images by using spectral data and fully connected neural network. *arXiv e-prints* (2022).

### Author contributions

J.R. designed and supervised the study. J.Z. and Y.X. analyzed the data, manufactured the figures and wrote the manuscript. Y.S.R. and L.Y.L. collected the data. All the authors have read and approved the final manuscript. This study was supported by a grant (22010503700) from the Shanghai Science and Technology Committee.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-54150-5>.

**Correspondence** and requests for materials should be addressed to J.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024