Deep Learning Algorithm based Wearable Device for Basketball Stance Recognition in Basketball

Lan Jiang[1], Dongxu Zhang[2]*
Basic Teaching Department, Sichuan Film and Television University, Chengdu, 610000, China[1]
College of Modern Services, Mianyang City College, Mianyang, 621000, China[2]

***Abstract*—With the continuous improvement of technology, modern sports training is gradually developing towards precision and efficiency, which requires more accurate identification of athletes' sports stances. The study first establishes a classification structure of basketball stance, then designs a hardware module to collect different stance data by using inertial sensors, thus extracting multidimensional motion stance features. Then the traditional convolutional neural network (CNN) is improved by principal component analysis (PCA) to form the PCA+CNN algorithm. Finally, the algorithm is simulated and tested. The outcomes demonstrated that the average discrimination error rate of the improved PCA+CNN algorithm in the Human 3.6M dataset was 3.15%, which was a low error rate. In recognition of basketball sports pose, the wearable based on the improved algorithm had the highest accuracy of 99.4% and took the quietest time of 18s, which was better than the other three methods. It demonstrated that the method had high discrimination precision and recognition efficiency, which could provide a reliable technical means to improve the science of basketball sports training plan and training effect.**

Keywords*—Deep learning; wearable devices; basketball; sports pose; CNN

I. Introduction

When basketball players training and contesting program is to be devised then developing a scientific and rational training program based on individual circumstances is the basis for improving their skills[1]. Considering player's actual ability, traditional training methods use experience and theory as a reference for the development of the program. However, this model is highly subjective and requires much time to analyze athletes' gestures, which is hard to meet the requirements of modern sports training. When coaches control the various postures of different athletes accurately, the training effect will be greatly raised, so it is important to collect and analyze posture data to achieve accurate identification. Basketball posture is one of the human postures, and today's human posture recognition is based on image acquisition and inertial sensor-based posture recognition [2]. The image acquisition stance recognition is achieved through camera image, video acquisition or different classifiers, whose technical maturity is high. But it has drawbacks that it needs a large amount of equipment and it is difficult to be applied widely. On the other hand, inertial sensor recognition is achieved by wearing a data acquisition sensor, which transmits the collected data in real time. Finally, the processing terminal completes the recognition. Its high recognition efficiency and low requirements for the use of the environment have become a research hotspot in this field [3]. In addition, with the increasing penetration of deep learning in various fields, the CNN, it contains, is used in discrimination widely. Therefore, in order to improve the recognition accuracy of the deep learning algorithm in basketball posture, wearable devices are applied in basketball training. This paper studies the recognition of basketball posture based on inertial sensors and convolution neural network. At the same time, the convolution neural network is improved by principal component to provide more high-performance auxiliary training technology for basketball.

II. Related Works

Deep learning algorithms are highly capable of learning, cover a wide range of areas, and can exhibit greater stability as more data is available. Through deep learning algorithms, the recognition of motion gestures has attracted huge attention from many professionals recently. And a series of forward-looking and practical research outcomes have been achieved. Hu B et al. developed a gesture recognition system based on UAV control for dynamic gesture recognition instead of human-machine interaction. This system used 8-layers CNN, 5-layers fully connected network and 2-layers fully connected network that could convert 4-D spatio-temporal data into 1-D and 2-D matrices to model the data of gesture sequences. The experimental outcomes demonstrated an average accuracy of 89.1% in the unscaled dataset [4]. Azad R's team designed a multi-level temporal sampling method based on depth sequence key frames for gesture recognition in computer vision. And they combined it with weighted depth motion maps to extract spatio-temporal information in the sequences by accumulating weighted absolute differences in successive frames. The outcomes demonstrated that the method had great precision [5]. Sun Y et al. made a posture discrimination processing framework based on a radar system with a shallow CNN whose input is a feature cube. And it could feed gesture contours into the CNN through lower latency. Experimental outcomes demonstrated that the framework achieved 92.08% accuracy in performing real-time classification of 12 gestures [6]. Zhang Y's team addressed the human-computer interaction in wearable devices. A benchmark dataset called EgoGesture was proposed to address the problem of human-computer interaction in wearable devices. In static and dynamic gesture recognition of different scenarios, it performed well with sufficient variability and realism when training deep neural networks [7]. The outcomes demonstrated that the method could achieve the discrimination of different body movements in basketball with a high accuracy rate [8]. Pan T Y et al. developed a gesture discrimination way based on multiple inertial measurement unit transducers, which could record the rotation information and acceleration of hand joints. And the average precision in the discrimination of basketball referee signals was 90.02% [9].

Zhao Y et al. designed a 3D position estimation method with an integrated ankle sensing device consisting of a magnetometer, accelerometer, gyroscope and barometer. And they added vertical variables to the adaptive multimodal stride in combination with it. Experimental outcomes demonstrated that it raised the applicability and precision of pedestrian horizontal position estimation [10]. Neethu P S's team applied a convolutional neural network classification method to human hand gesture detection and recognition, which extracted fingertip features in hand images by a connected component analysis algorithm. The outcomes demonstrated that the method had good operational performance [11]. Zhang W et al. applied deep learning networks to gesture recognition for human-computer interaction in hand gestures. It learned short-term and long-term features in the video input and feeded into a CNN for feature extraction. Experiments on the Jester and Nvidia datasets demonstrated its high accuracy [12]. Lian C et al. designed an IoT wristband to facilitate quantitative shooting action guidance for basketball players, which used a miniature inertial measurement unit sensor and was capable of collecting. The outcomes demonstrated that it was able to achieve 98.0% accuracy for layups, free throws, positioning shots and jump shots. And the overall accuracy was nearly 97.5% for 18 out of 18 shooting motions [13]. Kamel A's team developed a way for discriminating depth maps and pose data through convolutional neural networks, which used two inputs to describe the action representation. That is, a depth motion image that accumulates human movements and a motion joint descriptor that represents body joints over time. And the action prediction outcomes are generated from three CNN channels, which are experimentally demonstrated to be around 6.84% higher than general recognition methods [14]. Adithya V et al. applied CNN in deep learning to solve the problem of automatic sign language recognition by capturing gesture images and deriving complex feature descriptors. The outcomes demonstrate that it has good recognition accuracy on the hand gesture dataset [15].

To sum up, most researchers had proposed corresponding recognition methods for motion, gesture and human posture recognition. And these methods had improved the convolutional neural network in deep learning, which had achieved good application results. However, few people combined convolution neural network with inertial sensors, and there was a lack of research on the application of wearable devices in basketball motion recognition. And there was little research on the recognition of basketball posture. Therefore, the research first improves the action recognition effect of convolutional neural network, and combines it with wearable devices to apply it in basketball training, so as to improve the level of basketball training.

III. Basketball Pose Recognition based on Deep Learning Algorithms

A. *Basketball Posture Recognition Model and Wearable Device Design*

There are very complex and variable body movements during basketball. To carry out effective and accurate posture recognition, it is necessary to establish a scientific and comprehensive posture classification [16]. Based on the various limb states of the basketball player, the first two categories are static and athletic [17]. The athletic state is athletes' performance of different basketball actions, where the limbs remain in motion. While the resting state is the state in which the athlete's limbs do not perform any action and are strictly static. To effectively discern sporting gestures, the study has implemented a progression of two levels. In the first level, there are two types of movement gestures, namely transient and continuous, based on whether they are cyclical or not. And in the second level, it is further divided into seven stances, namely running, jumping, shooting, catching, passing, walking and dribbling, according to whether the movement state is lower or upper limb. The recognition of basketball movement stance actually becomes a recognition of the movement stance proposed by the automatic recognition. The classification of basketball sports stance is demonstrated in Fig. 1.

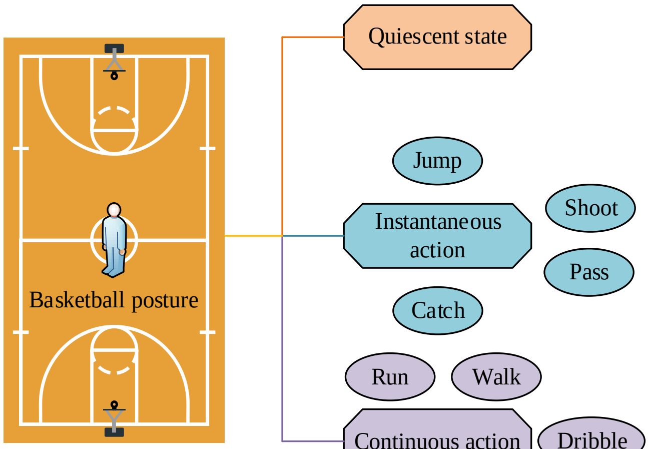


Fig. 1. Classification structure of basketball posture.

After determining the individual basketball postures, the wearable device is optimized. Timely and accurate gesture data acquisition is essential for accurate identification, so the research designs a basketball gesture data acquisition module based on inertial sensors in the wearable device. The data is first collected by using angular velocity, magnetic and acceleration sensors fixed to the basketball player. Then it is transmitted to the terminal device using wireless sensors for gesture discrimination. The hardware of the data acquisition includes data acquisition and transmission, containing a base station for data transmission and four nodes for information collection. Together form the data acquisition points which contains the acceleration and palstance information of the human body, the tri-axial accelerometer, tri-axial gyroscope MPU3050 and magnetometer LSM303DLH are collected. The wireless transceiver nRF24L01 is the core component of the data transmitting site and is for receiving information from the nodes and transmitting this data to the data terminals via the wireless network. For the information collection module, the microcontroller STM32F103 is for the core processing functions. At the same time, a 3.7V lithium-ion battery completes the energy supply for this module, the hardware structure is shown in Fig. 2.

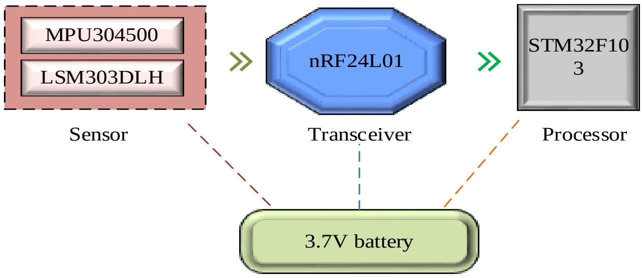


Fig. 2. Hardware structure diagram of basketball player posture data acquisition module.

In the data acquisition module, the signal transmission is carried out containing two parts. Firstly, the information conveying site receives the human posture data from the sensor nodes. Secondly, it is transmitted to the processing terminal. This part needs to minimize data collision rates and retain data to prevent large amounts of data loss, thus improving the accuracy of the collected data. The star topology network is the basis for the signal conveying between the processing terminal and the information conveying station, which is carried out via a time division multiplexing protocol. And the calibration of the clock deviations of the different nodes to maintain time uniformity is key to this. In conjunction with the classification of the basketball stance, four sensor nodes are set up in the legs and arms to obtain accurate data on the legs and arms of the athlete to collect magnetic field strength, acceleration and palstance data. The acceleration vector addition and the palstance vector addition for the n sampling point are calculated as demonstrated in equation (1).

$$\begin{cases} a_n = \sqrt{(a_n^x)^2 + (a_n^y)^2 + (a_n^z)^2} \\ p_n = \sqrt{(p_n^x)^2 + (p_n^y)^2 + (p_n^z)^2} \end{cases} \tag{1}$$

In Eq. (1), a_n^x , a_n^y and a_n^z correspond to the axial accelerations of x , y and z at the first n point, respectively, while p_n^x , p_n^y and p_n^z represent the corresponding angular velocities. a_n and p_n are the combined motion of the accelerations and palstance, respectively. The vector sums and three vectors of acceleration and angular velocity are then combined to form an 8-dimensional feature matrix. When the sampling point is N , each sample owns a feature matrix of $N \times 8$. The mean and variance are the time domain characteristics identified, and the mean and variance of each point is calculated as demonstrated in Eq. (2).

$$\begin{cases} \delta^2 = \frac{1}{N}\sum_{n=1}^{N}(a_n - \mu_a)^2 \\ \mu_a = \frac{1}{N}\sum_{n=1}^{N} a_n \end{cases} \tag{2}$$

In equation (2), δ^2 represents the mean value and μ_a is the variance. The time domain data is then transformed into the frequency domain data by the Fourier transform, which is calculated as demonstrated in Eq. (3).

$$S(n) = \sum_{i=0}^{N-1} a_i e^{-j\frac{2\pi}{N^n}} \tag{3}$$

In Eq. (3), $S(n)$ represents the value of adoption point n , where the frequency domain is located. The time domain features required for basketball gesture discrimination is actually the peaks of the Fourier transform, as demonstrated in equation (4).

$$f = \frac{Kf_s}{N} \tag{4}$$

In equation (4), K represents the quantity of frequency domain sampling point and f_s represents the sampling frequency. After the motion stance features are extracted, a 32-dimensional feature parameter set is obtained. For these parameters as a whole, there are some features with low or even no correlation with basketball motion pose. And there are also some features with redundant information between them, which seriously affect the classification efficiency and performance. Therefore, the study needs to make feature dimensionality further being reduced to achieve better recognition outcomes.

B. *Motion Pose Recognition based on Convolutional Neural Networks*

After obtaining the basketball pose data features, a CNN is used to downscale and identify the features. A CNN includes convolutional layers, pooling layers, activation functions and fully connected layers [18]. In general, the upper layers are convolutional layers, and then one or more fully connected layers are cascaded. All fully connected and convolutional layers are followed by an activation function, which is used as a non-linear transform. At the same time, pooling layer is after convolutional layer, which serves to reduce the amount of data contained in the intermediate outcomes. In a convolutional neural network, the convolutional layer is the most basic unit [19]. Under the action of mapping, a well-trained convolutional layer automatically and efficiently extracts features from the data and transfers the original data to the hidden feature space. The fully connected layer is the classifier in CNN, which maps the learned features to the sample data labeling space. At the same time, the essence of the operation in this layer is the multiplication of vectors and matrices. If it is transformed into vector form after the straightening operation, only the convolutional layer feature map can be input to the fully connected layer [20]. The pooling layer differs significantly from the fully connected and convolutional layers. Because it has no bias or weight parameters and is mostly used to lowering feature map. Thereby, the redundancy is eliminated significantly in the feature map. Among the pooling approaches, maximum and average pooling are more widely used. Similar to the convolutional layer, the pooling mode slides through a box n_1 high and n_2 wide, over all the input channels of the feature map. And it finds the maximum or average value in this box and the size of the box are usually consistent with the step size of the slide. The activation function acts as the key step in the convolutional neural network to complete the non-linear mapping. And almost all fully connected and convolutional layers must go through it. The missing activation function will cause the stack of fully connected layers and multiple volume layers. It is equivalent to the multiplication of multiple matrices to finally get a matrix, resulting in the original data not being activated to obtain a strong fit. The basic structure of CNN is demonstrated in Fig. 3.

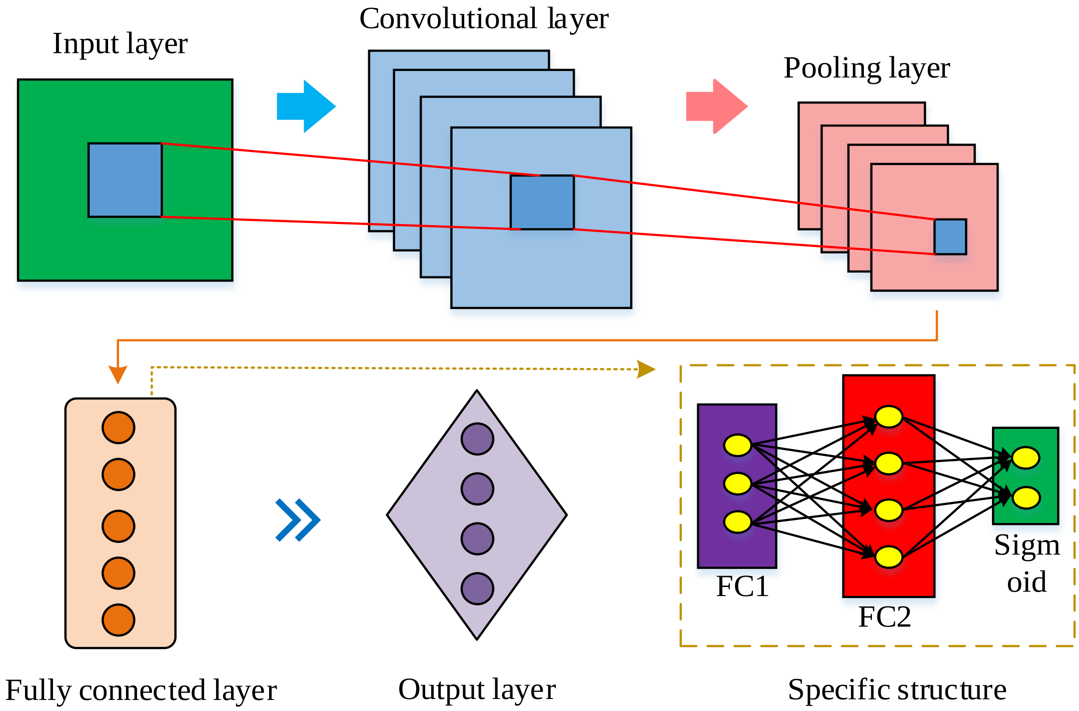


Fig. 3. Basic structure of convolutional neural network.

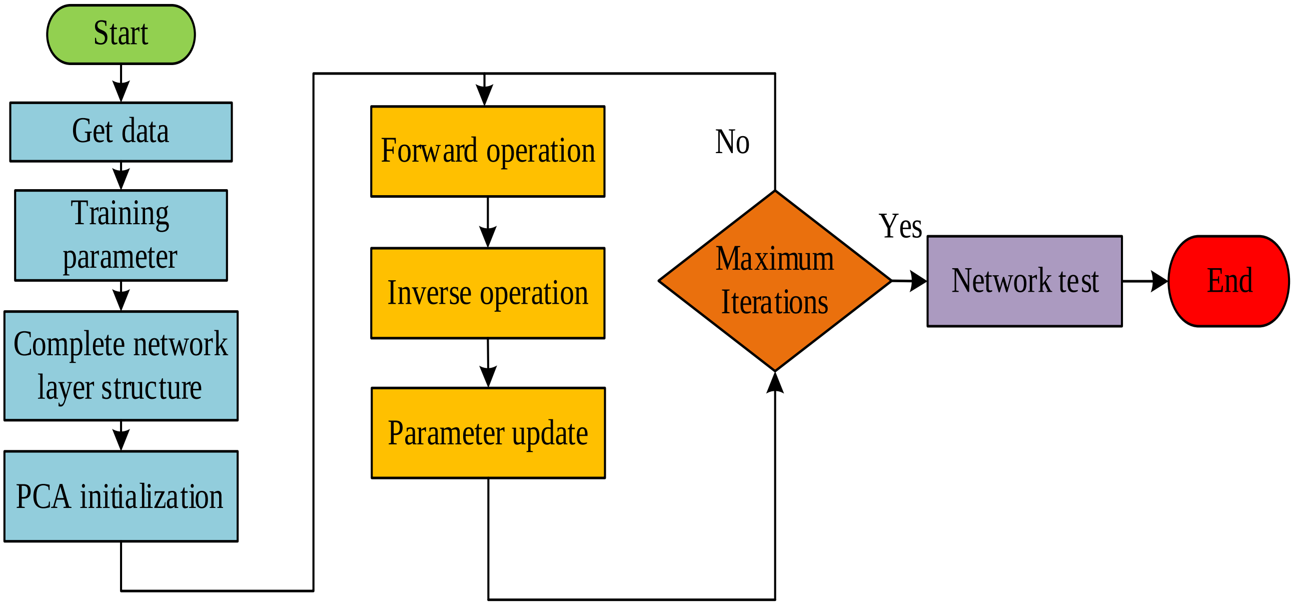


Fig. 4. Algorithm flow of CNN with principal component analysis.

The multi-layer perceptron is the basic structure of a fully connected layer in CNN, including input and output layer, with a weight matrix connecting the layer to the layer [21]. The input layer is formed by the upper layer feature map after vectorization. After the inner product of the input vector and the weight matrix, and mapped by the activation function, the outcome is obtained for the output layer. The calculation is demonstrated in Eq. (5).

$$y_{pj}^{l} = f\left(\sum_{i=0}^{N^{l-1}} X_{pj}^{l-1} \cdot b_j^l + w_{ji}^l\right) \tag{5}$$

In Eq. (5), f represents activation function, N is the quantity of neurons, b represents the neuron bias, w is the weights. l is the convolutional layer sequence, j is the feature map sequence, and p is the training sample sequence. As the number of layers in the network increases, the extracted features become more abstract and discriminative, and these features facilitate the classification of basketball poses. At the same time, the features give feedback to the extraction of shallow features [22]. The main problem with convolutional neural networks is the network parameters training and updating. To raise the ability of basketball posture recognition, principal component analysis is introduced to integrate with convolutional neural networks. Principal component analysis is a statistically based feature extraction method that can filter features with a high correlation, i.e. selecting the optimal sign. At the same time, the principal component analysis method has a high similarity with the learning outcomes of self-coding neural networks. Therefore, the study uses this method to perform multiple calculations on all convolutional kernel sets in layers. The convolutional kernel parameters are initialized to optimize the convolutional neural network performance. The optimized algorithm is demonstrated in Fig. 4.

In Fig. 4, the improved convolutional neural network first completes all network layer structures based on training and test data, then the training parameters can be defined. From there, the initialized network parameters are analyzed by principal component analysis, followed by forward and backward operations for network training. Then parameter updates are implemented. Finally the network is tested and concluded by determining whether iterations has been reached to complete the network. The principal component analysis method calculates the input layer feature vector set V and then orients the first L of the vector set. These can be used as the principal component feature vector of the input sample data set, thus forming the convolutional kernel set $V1$ of the $C1$ layer, as demonstrated in Eq. (6).

$$\begin{cases} \min \left\| Y - V_1 V_1^T Y \right\|_F^2 \\ V_1^T V_1 = I \end{cases} \tag{6}$$

In Eq. (6), I represents the unit matrix and V_1 represents the training data fed into the convolutional neural network. In the convolutional layer $C2$, the same method is used to treat the feature maps from the previous layer as the sample set for the principal component analysis. So it outcomes in the convolutional group $V2$ of the $C2$ layer. According to the convolution and size, and used as the initialization values for the corresponding layer's convolution kernel, all the column vectors contained in all the computed feature vectors are arranged, which finally completes the convolution kernel initialization. The optimized convolutional neural network model is in Fig. 5.

Finally, the optimized convolutional neural network is applied to the wearable device classifier to obtain different basketball motion pose types, thus completing the output of the recognition outcomes.

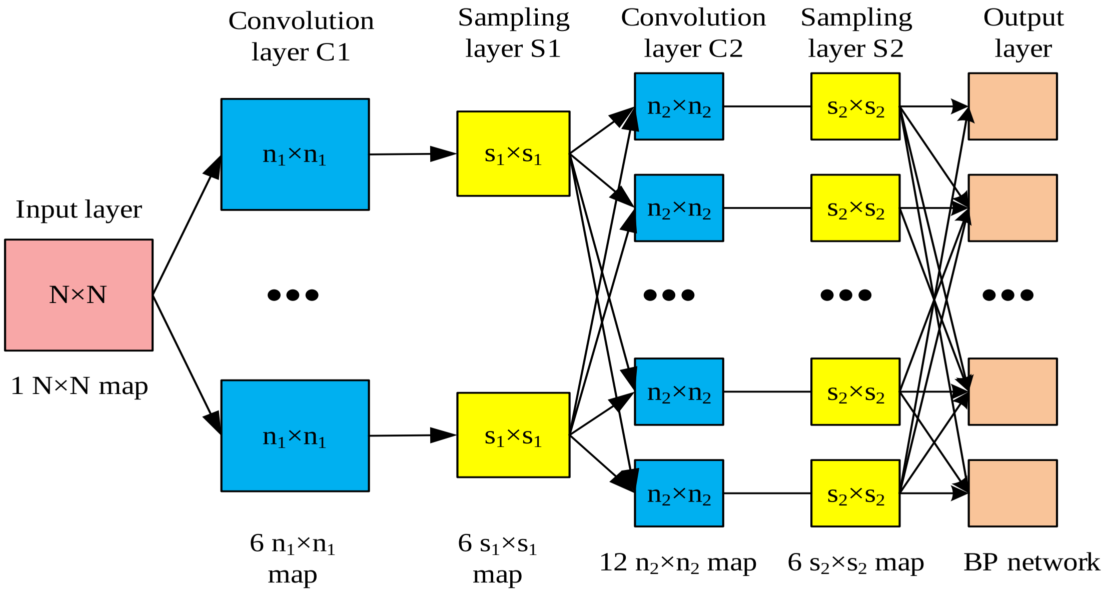


Fig. 5. CNN optimized by introducing principal component analysis.

IV. Analysis of Basketball Posture Recognition Outcomes based on Improved CNN

The function of the PCA-improved CNN is first examined. The simulation environment is a discrete graphics card GeForce MX150, 7th generation Intel i5 processor, Matlab (R2016b), 2 GB GDDR5, 256 GB PCIe SSD, memory 8 GB DDR4, Deep Learning-Toolbox. The PCA+CNN is compared with the classical CNN algorithm and the selected dataset is Human3.6M dataset. This dataset is a large public dataset for 3D human pose estimation, containing 3.6 million human poses and corresponding images. A comparison of the outcomes of the two algorithms is in Table I.

Table I demonstrates the comparison outcomes of the error rate and time between CNN algorithm and PCA+CNN for five runs on the Human3.6M dataset. In Table I, in terms of running time, the average time of the PCA+CNN and the classical CNN over the five experiments was 291.60s and 289.11s respectively, which shows a small difference. In terms of error rate, the average error rate of PCA+CNN algorithm and classical CNN were 3.15% and 4.68% respectively. The former had a lower error rate and the error rate of each experiment was lower than that of the traditional CNN. Then the number of iterations with mean square error outcomes were taken from one of the five experiments. And the outcomes of the other four experiments were graphically similar to the selected experimental outcomes. The comparison of the two algorithms was demonstrated in Fig. 6.

In Fig. 6, the red line was the experimental run of the classical CNN algorithm, while blue line represented the run of PCA+CNN. The horizontal coordinate was 6000 samples after 5 iterations with 50 training samples input each time. And the vertical coordinate was the mean square error for each input training sample number of 50. From Fig. 6, in the initial period of iteration, the mean square error obtained by the CNN algorithm was greater than that produced by PCA+CNN. The difference was most significant when the number of iterations was between 145 and 1000. While after the number of iterations exceeds 1000, the mean square error of both algorithms changes less. But the error convergence curve of PCA+CNN always stayed below the CNN experimental outcomes. At the same time, when the mean square error was 0.0479, iterations of PCA+CNN was only 4261. While the number of iterations of conventional CNN was 5893, which is 1632 times higher than the former. So the error rate obtained by PCA+CNN algorithm was smaller than that of the classical CNN algorithm. The two algorithms were then simulated for the MPII Human Pose dataset, evaluated by loss function curves and accuracy curves. 410 human activities were included in the MPII Human Pose dataset, and each image was clearly labelled with an activity, making it a dataset for human pose recognition. For the accuracy metric, it was the ratio of the number of correctly recognised poses to the total. The loss function was an important metric for evaluating how well the deep learning training was done. Here, the loss function was mainly for the test set and was judged on the basis that the loss functions of both the test and training sets converge and the difference was small. The loss function and precision curves are in Fig. 7.

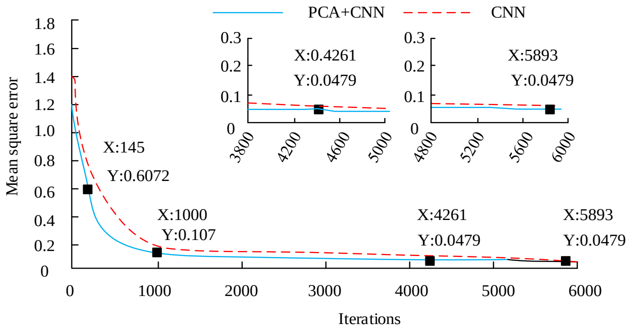
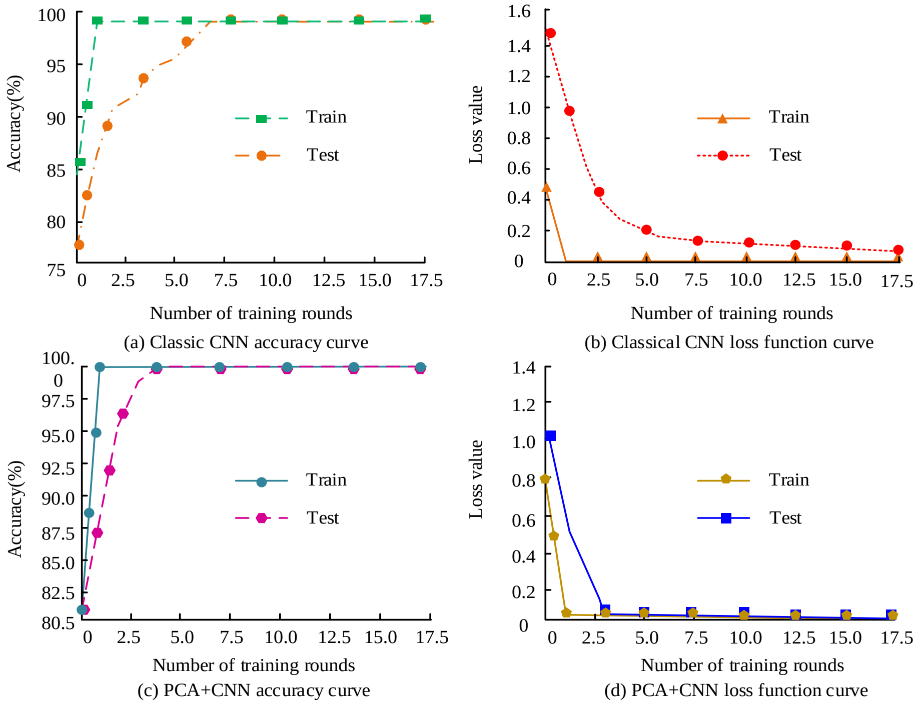


Fig. 6. Error convergence curve outcomes of CNN algorithm and PCA+CNN algorithm.

TABLE I. Comparison of Running Outcomes of PCA+CNN and Classical CNN Algorithms on Human 3.6M Dataset

Algorithm	Number of experiments	Running time/s	Error rate/%
PCA+CNN	1	291.53	3.04
	2	290.14	3.22
	3	293.52	3.13
	4	292.37	3.28
	5	290.45	3.10
CNN	1	289.12	4.35
	2	287.56	4.47
	3	289.73	4.92
	4	290.51	5.01
	5	288.64	4.63



(a) Classic CNN accuracy curve

(b) Classical CNN loss function curve

(c) PCA+CNN accuracy curve

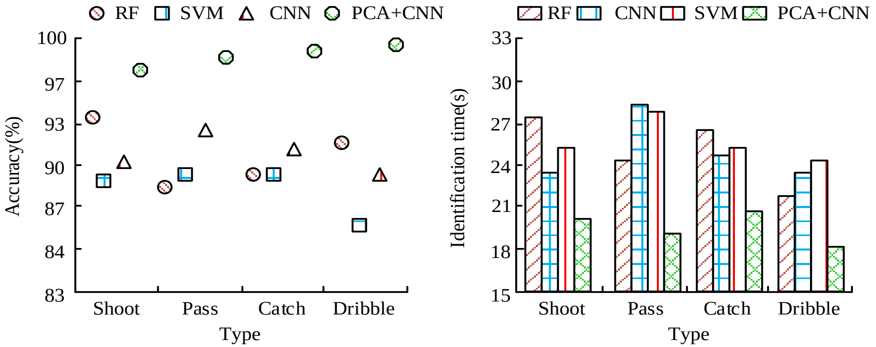
(d) PCA+CNN loss function curve

Fig. 7. Loss function and accuracy curve outcomes of CNN and PCA+CNN algorithms.

Fig. 7(a) and Fig. 7(b) demonstrated the accuracy curve and loss function curve of the classical CNN algorithm, respectively, while Fig. 7(c) and Fig. 7(d) demonstrated the precision and loss-function curve of the PCA+CNN algorithm. From Fig. 7(a), the classical CNN achieved a fit at round 7, where the accuracy was 99%. In Fig. 7(b), the loss function curve starts to smooth out and converge to 0 in round 8. In Fig. 7(c) and Fig. 7(d), the PCA+CNN converged near round 5 with 100% accuracy, which was faster and more accurate than the CNN algorithm, and the loss function also converged faster than the CNN algorithm, which has better performance. Finally, for example validation, sample data was collected from 10 athletes in eight stances: dribbling, walking, shooting, no movement, jumping, passing, catching and running. 80 sets of repeated data were collected for each movement, and a total of 6400 data samples can be obtained. The data was collected by performing a series of pre-determined basketball stances in conjunction with the participant's own exercise habits. All postures included both upper and lower limb movements, and the lower and upper limb movements were also analyzed separately for identification purposes. For checking the validity of the posed way, Support Vector Machine (SVM) and Random Forest (RF) were for comparison in the example validation. The outcomes of the four methods for the recognition of upper limb basketball postures are demonstrated in Fig. 8.

Fig. 8(a) and Fig. 8(b) the upper limb recognition accuracy and recognition time outcomes for the four algorithms respectively, with the horizontal coordinates all representing the types of upper limb basketball sports poses, i.e. shooting, catching, passing and dribbling. From Fig. 8(a), it can be seen that among the four types of upper limb sports pose recognition, RF, SVM and CNN are stable between 88~94%, 86~89% and 89~93% respectively, while the accuracy of PCA+CNN is above 97% for all of them, and the recognition accuracy for passing is as high as 99%. From Fig. 8(b), among the four methods, the PCA+CNN algorithm recognized the four types of upper limb movement poses in 20s, 19s, 21s and 18s, respectively, all of which were faster than the other three methods, with a maximum lead of 10s and high operational efficiency. The recognition outcomes of the four methods for the three lower limb motion postures of jumping, walking and running are demonstrated in Fig. 9.

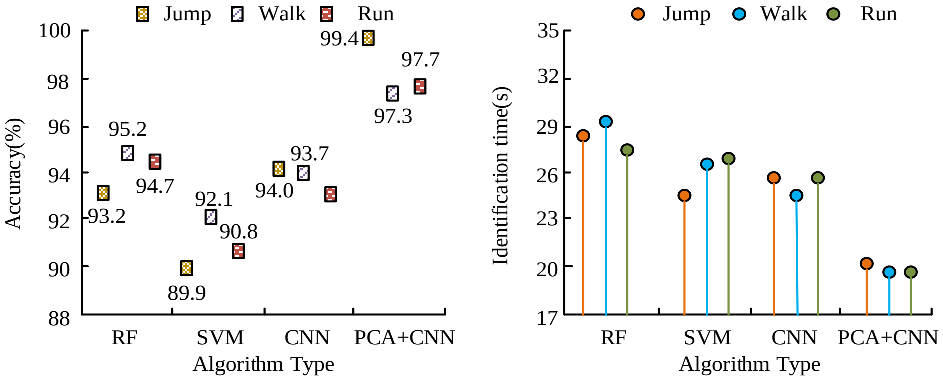
Fig. 9(a) and Fig. 9(b) represent the accuracy and time outcomes of the four methods for lower limb motion pose recognition, respectively, and the horizontal coordinates represent the four algorithms compared. From Fig. 9(a), the accuracy of the proposed PCA+CNN among the four methods is as high as 99.4% in the recognition of jumping, and the accuracy of 97.7% and 97.3% for running and walking respectively, which is higher than the other three methods. From Fig. 9(b), the recognition time of the Random Forest algorithm was above 28s, which was the worst performance among the four methods. In contrast, the PCA+CNN algorithm was stable at around 20s, which was more efficient than the other three methods and had better recognition performance. To sum up, the method proposed in the study can identify the basketball posture with high accuracy and maintain high recognition efficiency, which can provide more scientific and effective methods for basketball training.



(a) Accuracy of upper limb basketball posture recognition

(b) Recognition time of upper limb basketball posture

Fig. 8. Recognition outcomes of upper limb basketball movement posture using four methods.



(a) Accuracy of the legs basketball posture recognition

(b) Recognition time of the legs basketball posture

Fig. 9. Four methods of lower limb basketball posture recognition.

V. Conclusion

Traditional basketball training is based on the coach's personal training experience and theory, making it difficult to evaluate the training effect objectively. With the continuous improvement of deep learning algorithms, neural networks are increasingly widely used in data reduction and classification processing. The study proposes a scientific sports stance classification structure for basketball's complex and variable stance characteristics and thus establishes a data information collection module based on inertial sensors. The convolutional neural network is then improved using principal component analysis, and finally, the improved algorithm is applied to the recognition of basketball postures. Experimental outcomes demonstrate that the proposed PCA+CNN algorithm has an average recognition error rate of 3.15% in the Human3.6M dataset. Compared to 4.68% of the traditional CNN, the difference is 1.53%, and the error convergence curve of PCA+CNN is consistently below the CNN outcomes. In the MPII Human Pose dataset, the PCA+CNN converged in only the 5th round achieved 100% accuracy with better performance. In recognition of the upper limb pose for basketball, the method performed above 97% accuracy for all four types of dribbling, passing, catching, and shooting. Its running time is 10s faster and the accuracy for the lower limb motion pose is 99.4%. It has higher running efficiency and accuracy. However, the performance of CNN algorithm in the case of increased learning rate has not been analyzed in the study, and the power consumption caused by wearable devices has not been optimized. Therefore, it is necessary to further reduce the power consumption and increase the learning rate to achieve better results.

References

[1] Zhao B, Liu S. Basketball shooting technology based on acceleration sensor fusion motion capture technology[J]. EURASIP Journal on Advances in Signal Processing, 2021, 2021(1): 1-14.

[2] Pfitscher M, Welfer D, Do Nascimento E J, Cuadros M A D S L & Gamarra D F T. Article users activity gesture recognition on kinect sensor using convolutional neural networks and fastdtw for controlling movements of a mobile robot.Inteligencia Artificial, 2019, 22(63): 121-134.

[3] Li B, Xu X. Application of artificial intelligence in basketball sport.Journal of Education, Health and Sport, 2021, 11(7): 54-67.

[4] Hu B, Wang J. Deep learning based hand gesture recognition and UAV flight controls.International Journal of Automation and Computing, 2020, 17(1): 17-29.

[5] Azad R, Asadi-Aghbolaghi M, Kasaei S & Escalera S.Dynamic 3D hand gesture recognition by learning weighted depth motion maps.IEEE Transactions on Circuits and Systems for Video Technology, 2018, 29(6): 1729-1740.

[6] Sun Y,Fei T,Li X,Warnecke A,Warsitz E & Pohl N.Real-time radar-based gesture detection and recognition built in an edge-computing platform.IEEE Sensors Journal, 2020, 20(18): 10706-10716.

[7] Zhang Y, Cao C, Cheng J, & Lu H.EgoGesture: A new dataset and benchmark for egocentric hand gesture recognition.IEEE Transactions on Multimedia, 2018, 20(5): 1038-1050.

[8] Li J, Gu D. Research on basketball players' action recognition based on interactive system and machine learning.Journal of Intelligent & Fuzzy Systems, 2021, 40(2): 2029-2039.

[9] Pan T Y, Chang C Y, Tsai W L & Hu M C. Multisensor-based 3D gesture recognition for a decision-making training system.IEEE Sensors Journal, 2020, 21(1): 706-716.

[10] Zhao Y, Liang J, Cui Y, Sha X & Li W J.Adaptive 3D position estimation of pedestrians by wearing one ankle sensor.IEEE Sensors Journal, 2020, 20(19): 11642-11651.

[11] Neethu P S, Suguna R, Sathish D. An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks.Soft Computing, 2020, 24(20): 15239-15248.

[12] Zhang W, Wang J, Lan F. Dynamic hand gesture recognition based on short-term sampling neural networks.IEEE/CAA Journal of Automatica Sinica, 2020, 8(1): 110-120.

[13] Lian C, Ma R, Wang X,Zhao Y, Peng H, Yang T & Li W J. ANN-enhanced IoT wristband for recognition of player identity and shot types based on basketball shooting motion analysis. IEEE Sensors Journal, 2021, 22(2): 1404-1413.

[14] Kamel A, Sheng B, Yang P.Deep convolutional neural networks for human action recognition using depth maps and postures.IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2018, 49(9): 1806-1819.

[15] Adithya V, Rajesh R. A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition.Procedia Computer Science, 2020, 171(12):2353-2361.

[16] Han C. RETRACTED ARTICLE: Urban air pollution resolution and basketball training optimization based on time convolution network.Arabian Journal of Geosciences, 2021, 14(18): 1-7.

[17] Rast F M, Labruyère R. Systematic review on the application of wearable inertial sensors to quantify everyday life motor activity in people with mobility impairments.Journal of NeuroEngineering and Rehabilitation, 2020, 17(1): 1-19.

[18] Mangiarotti M, Ferrise F, Graziosi S, TAMBURRINO F & Bordegoni M. A wearable device to detect in real-time bimanual gestures of basketball players during training sessions.JOURNAL OF COMPUTING AND INFORMATION SCIENCE IN ENGINEERING, 2019, 19(1): 1-13.

[19] Murao K,Yamada H,Terada T & Tsukamoto M.Estimating Timing of Specific Motion in a Gesture Movement with a Wearable Sensor.Sensors and Materials, 2021, 33(1): 109-126.

[20] Pai N S, Chen Y H, Hung C P, Chen P Y, Kuo Y C & Chen J Y.Dual-input Control Interface for Deep Neural Network Based on Image/Speech Recognition.Sensors and materials, 2019, 31(11):3451-3463.

[21] Jain S, Rustagi A, Saurav S, Saini, R & Singh S.Three-dimensional CNN-inspired deep learning architecture for Yoga pose recognition in the real-world environment.Neural Computing and Applications, 2021, 33(12): 6427-6441.

[22] Hu Y, Lu M, Lu X. Driving behaviour recognition from still images by using multi-stream fusion CNN. Machine Vision and Applications, 2019, 30(5): 851-865.