Journal of Multimedia Information System VOL. 11, NO. 1, March 2024 (pp. 45-56): ISSN 2383-7632 (Online)
https://doi.org/10.33851/JMIS.2024.11.1.45

Deep Learning Driven Human Posture Location in Physical Education Teaching

Shaohua Wang[1*], Wanli Shi[2]

Abstract

The study of human posture is widely applied in physical education teaching, human motion recognition, and other aspects. With the rise of online teaching, the lack of convenient physical education teaching methods has been able to improve. However, due to the complex structure of human body, the study of human posture is a hard problem of consciousness problem in the area of computer vision. This article mainly studies human posture research algorithms based on deep learning. It uses 101-layer network of ResNet to detect the key points of human body in the image and obtains the categories and coordinates of these key points. In this article, a 101-layer network of ResNet model is constructed to fully learn the visual features of key points in human posture. Secondly, the key point location loss function is improved, and the human posture research is realized by using huber loss function instead of mean square error (MSE) loss function. Finally, experimental analysis shows that compared to traditional integral pose regression (IPR) and location adaptive integral pose regression (LAIPR), the use of ResNet based human posture estimation method for human posture recognition improves precision. It has practical significance for physical education teaching applications.

Key Words: Human Posture, Physical Education Teaching, Deep Learning, ResNet.

I. INTRODUCTION

For a long time, the study found that the physical exercise of primary and secondary school students after class and during holidays tapered. With the expansion of online education and tutoring, the time for students to exercise after class has been greatly crimped [1]. However, with the expansion of modern information technology, the promotion of Internet plus education and the continuous development of intelligent physical education, physical education distance learning has become one of the inevitable trends of future development [2]. Artificial intelligence and other technologies not only smash the restrictions of space, time, region and other factors on the traditional sports classroom, but also provide guarantees for schools to provide teaching feedback links in extracurricular sports teaching. Educators redesigned physical education teaching, and made physical education teaching present a new learning space and environment by changing the traditional way of teacher-student interaction. However, there are a series of issues in the online physical education teaching procedure, such as a single layout of content, insufficient recognition of human posture, and an objective evaluation system of actions [3].

The progress of physical education also relies on artificial intelligence technology. At present, experts and scholars have clearly pointed out that fully tapping the potential of artificial intelligence and developing artificial intelligence sports education with more comprehensive and safer functions will be a vital research guide in the future [4-5]. Therefore, the combination of artificial intelligence and physical education has received continuous attention from researchers. However, according to literature research, the current program of artificial intelligence technology in the area of physical education is mostly concentrated in the collection of motion sensor information, data analysis, and other aspects [6]. These works are mostly auxiliary to existing education models, without substantive optimization of teaching elements such as learning pathways and evaluation methods. In other words, the integration of artificial intelligence technology and physical education is as usual in its infancy. It is inevitable to deeply analyze the characteristics of artificial intelligence technology and the demand of physical education, break the knowledge boundary between the two, and study the deep integration of artificial intelligence technology and physical education in specific sports projects [7]. Physical education teaching has its own

Manuscript received December 20, 2023; Revised January 27, 2024; Accepted February 01, 2024. (ID JMIS-23M-12-050)
Corresponding Author (*): Shaohua Wang, +86-15143209986, wangsh_2020jl@163.com
[1]Jilin Technology College of Electronic Information, Jilin, China, wangsh_2020jl@163.com
[2]Jilin University of Agricultural Science and Technology, Jilin, China, shiwanli@jlnku.edu.cn

characteristics. Students not only need to learn basic theories, but also need to practice basic movements, and combine the actual analysis of action learning results to achieve accurate and intelligent feedback and scientific evaluation of basic movement learning, forming a practical action evaluation process.

The artificial intelligence boom triggered by deep learning technology has swept through many fields and achieved fruitful results. In the areas of machine vision and so on, deep learning technology shows its advantages in processing big data, security and multi-source heterogeneous data [8-9]. Based on virtual reality technology, a movement teaching platform with a more immersion feel can simulate human motion scenes and achieve realistic action simulation effects. The artificial intelligence technology based on deep learning has achieved excellent results in the area of human posture research [10]. This kind of method can achieve precise analysis of human posture in time and space, with strong practicality and wide applicability. Human motion posture recognition technology refers to the use of computer vision technology to recognize and analyze the posture of the human body during the movement process, thereby providing real time feedback and guidance.

Human posture research is a fundamental issue in computer vision, serving as the foundation for multi person posture research and motion object analysis. It can be universally applied in numerous fields such as human-computer interaction, pedestrian behavior recognition, behavior analysis, human segmentation, and physical education teaching [11]. The goal of the one-person pose study is to find the coordinates of a person's different nodes from an image containing a human body. The study of individual poses in images is facing daunting challenges due to the influence of shooting angles, scenes, lighting, and clothing. However, the research of single person attitude has made rapid development. Since the methods in the attitude research are mainly based on the depth convolutional neural network, and a few measures are based on the generative adversarial network. Simultaneously, we have developed the multi person attitude research based on the single person attitude research. The final multi person attitude can be obtained by accurately detecting the target person in the input RGB video, predicting 2D key points, and eventually predicting accurate 3D key points through our network. On the basis of existing hardware, we can completely achieve real-time multi motion object analysis, which provides us with many conveniences in our daily lives.

The process of posture matching based on skeleton features requires the use of human bone point data information. Human bone points are the joint parts that connect limbs into a steel hinge system. The correlation of bone point data can be used to determine the specific state of bone points and the action forms of each limb. With the development of hardware devices and breakthroughs in visual algorithms, especially pose research algorithms based on deep learning, obtaining bone point data has become more convenient [12-13]. The main methods for obtaining human bone point data include direct acquisition from devices and pose estimation extraction from images. This article combines practical examples of physical education teaching and focuses on the study of human posture. The main work and structural arrangement of this article are as follows.

(1) Firstly, the direction and significance of human posture research in the context of deep learning were introduced, as well as the corresponding elaboration of the research on physical education teaching applications and model algorithms in artificial intelligence. Finally, the major research content and organizational structure of this article were introduced.
(2) By researching on human posture, human posture data processing is analyzed, the ResNet network principles are introduced, and the human posture estimation method based on ResNet is proposed.
(3) A platform was built based on relevant requirements, and the accuracy and performance of the experimental results were evaluated and analyzed. Compared with traditional IPR and LAIPR, the ResNet based human posture estimation method for human posture recognition improved accuracy, which to some extent verified the effectiveness and practicality of the application in physical education teaching.

The remaining part of this article consists of four parts, and the second part is related literature related to the work of this article. The third part provides detailed introduction to dance recognition based on deep algorithm models and the design of human-machine interaction platforms for data centers. The fourth part analyzes the proposed method and its effectiveness through experiments and indicators. Finally, the main research contents and conclusions of this paper are summarized.

II. RELATED WORK

Big data technology and other emerging computer technologies had made continuous progress in physical education teaching and field applications. The study of human posture emerged in 1980, and early methods used model-based methods for estimating human posture. Ren et al. [14] first used segmentation methods to receive the features of various portion of the human body, and then matched the model using constraints such as relative position, scale consistency, and shape consistency between joint points to obtain the human posture. Hua et al. [15] used Markov networks to model the position of human joint points, inferring

human posture through information such as shape, edge, and color in the image. Mori and Malik [16] obtained human posture by matching shapes, which not only obtained the position of joint points, but also achieved tracking of joint points during motion. In addition, there was the Body Plan (BP) method, which referred to a series of human features learned from image data under the limitations of color, texture, and geometric attributes. Using body maps could achieve segmentation and recognition of human bodies in complex environments.

The study of human posture based on deep learning had now become the mainstream research method for three-dimensional posture research. Martinez et al. [17] proposed the classic two-stage method baseline based on deep learning, which proved its feasibility through experiments and provided a foundation for subsequent research. Zhao et al. [18] also proposed an improved method based on Baseline, which used two-dimensional skeleton coordinates to predict the three-dimensional pose skeleton through its network. Pavllo et al. [19] proposed a two-stage method based on baseline, which obtained three-dimensional human body coordinates based on two-dimensional skeleton coordinates, with high precision and real-time behavior. The innovation of the method proposed by Yang et al. [20] was the use of an adversarial learning framework to supervise two-dimensional skeletons, which could also be understood as a combination of two types of methods. The method proposed by Oberweger et al. [21] was also applied to supervise the predicted three-dimensional skeleton in order to improve accuracy. Xiang et al. [22] proposed a three-dimensional vector called part orientation fields (POF), which was trained through a network to obtain end joints such as hands and feet. However, due to the lack of available datasets, its universality was not strong. In contrast, the top-down method first found all the people in the image, and then performed pose estimation to find each person's key points, which could be directly achieved using single person pose estimation. Pavlakos et al. [23] proposed a novel idea of outputting three-dimensional skeleton coordinates based on images, which innovatively divided the three-dimensional space into grids. Park et al. [24] suggested using two-dimensional skeleton information to estimate three-dimensional human skeleton. Kanaza et al. [25] suggested restoring global 3D human posture in a real environment. Although these methods had good performance in predicting 3D poses directly from input images in real-time, it was often difficult to measure their remaining errors. Pavlakos et al. [26] estimated the 3D heat map U-net network for each joint by extending it.

The method based on hidden markov model (HMM) was a stochastic model based on transition probability and transmission probability. It consisted of two parts, state and observation. The probability of the current state of the system was only related to the state at the previous time, and was independent of other historical state conditions. Yamato et al. [27] first trained the HMM for matching in human posture recognition, and used Baum Welch algorithm to obtain HMM training parameters. Nguyen [28] proposed a hierarchical HMM, which had a multi-layer model structure and could clearly express the details of posture and behavior in human movement. Natarajan et al. [29] used the hierarchical variable HMM to model the human body in three layers. The top layer modeled the human comprehensive action, and contained a separate Markov chain. The middle layer and bottom layer modeled the primitive behavior and body posture respectively. Ren et al. [30] proposed primitives to address human gesture matching in theme-specific behavior recognition. Primitives were composed of features that described contextual information. Deep belief network (DBN) improved the efficiency and robustness of matching by integrating different weak information features and enhancing their functionality. Luo et al. [31] first introduced DBN into human posture behavior recognition for posture matching, indicating that due to the inclusion of further hidden nodes and observation points. Weinland et al. [32] proposed motion history volumes (MHV) templates to describe human behavior based on free perspective. Fourier transformed the templates in the cylindrical coordinate system, and finally used Fourier features to describe human behavior posture. Park et al. [33] used a network to match behavior and deep neural networks for multi part pose matching. On account of the demand for vast feature parameters, the dynamic Bayesian network method has a higher computational complexity. The method extracted useful features related to the target task from a given sequence of images or video frames, and then converted the feature sequence into a set of static templates. The test sequence was matched with the pre stored standard pose template. The template based matching algorithm had the advantages of less computational complexity, but was more sensitive to sequence spacing and noise. Bobick et al. [34] proposed the spatio-temporal template method, which utilized the accumulated binary images of moving image sequences and matched templates based on Hu moments. The similarity between templates in this method was measured using Markov distance.

III. HUMAN POSTURE BASED ON DEEP LEARNING IN PHYSICAL EDUCATION TEACHING

3.1. Processing of Human Posture Data in Physical Education Teaching

The artificial intelligence technology based on deep

learning has achieved excellent results in the area of human posture recognition. This method offers accurate fundamental data for action recognition and posture matching based on the video human posture feature information in sports teaching applications. And it can achieve precise analysis of human posture in time and space.

In obtaining the position coordinate data of human posture bone points in a two-dimensional image, a pose estimation model is employed to handle the image containing the human body to obtain the coordinate data of the bone points. Human pose research is an important research direction in the field of computer vision, aiming to infer the pose information of the human body, including positions and pose angles, through two-dimensional image or video data. It is mainly achieved through image processing and learning algorithms. Generally speaking, it can be divided into two main steps, namely human keypoint detection and pose estimation. Human keypoint detection refers to the detection of key points of human bones in two-dimensional images, that is the position of the human body.

The pixel position coordinate data in the image is based on the fixed point in the upper left corner of the image as the origin, and posture matching is performed based on this coordinate data. When the relative position of the human body in the image is different, even if the action standard to be matched is not the same. Due to the different coordinate data of its bone points, the action is judged to be non-standard. Therefore, it is necessary to reposition the coordinate origin, select suitable bone points as the coordinate origin, reconstruct the relative coordinate system. It solves the problem of inconsistent bone point coordinate information caused by different positions and resolution scales of the human body in the image. The joint point of the midpoint of the shoulder joint is took as the origin (0, 0). The coordinate system used for the raw data has the right horizontal direction as the x-axis, the vertical downward direction as the y-axis, and the upper left corner of the image as the coordinate origin. The new skeleton space coordinate system is reconstructed. The center point of the shoulder joint is taken as the coordinate origin. The coordinate system referencing the human skeleton space is formed, and the coordinates of other bone points are recalculated and determined.

$p_{i0} = (x_{i0}, y_{i0})$ is used to represent the original coordinate data of the *i*-th joint point, and $p_{i1} = (x_{i1}, y_{i1})$ is used to represent the coordinate data of the *i*-th joint point in the new coordinate system. Therefore, the original coordinate data of the shoulder joint midpoint with No. 1 is $p_{10} = (x_{10}, y_{10})$, and the coordinate of the shoulder joint midpoint in the new coordinate system is $p_{11} = (0,0)$. Taking the nose joint point with serial number 0 as an example, the nose coordinate data obtained is represented by $p_{00} = (x_{00}, y_{00})$. Then the coordinate data after the original data conversion in the new coordinate system is $p_{01} = (x_{01}, y_{01})$, where $x_{01} = x_{00} - x_{10}, y_{01} = y_{00} - y_{10}$. Therefore, for 25 joint points, the coordinate data after the *i*-th joint point conversion is shown in equation (1).

$$p_{i1} = p_{i0} - p_{10} = (x_{i0} - x_{10}, y_{i0} - y_{10}) = (x_{i1}, y_{i1}). \quad (1)$$

Finally, the human posture joint points are standardized to achieve consistent pixel scale. And equation (2) is used to transform it.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} z & 0 \\ 0 & z \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \quad (2)$$

Among them, (x', y') is the coordinate value of the human bone point in the standardized image. (x, y) is the coordinate value of the human bone point in the source image. $z = x_{o,j1}/x_{o,j2}$ is the scaling ratio. $x_{o,j1}$ is the coordinate data of nose in the standardized image, and $x_{o,j2}$ is the edge length of the source image, all measured in pixels.

3.2. ResNet Network Principles

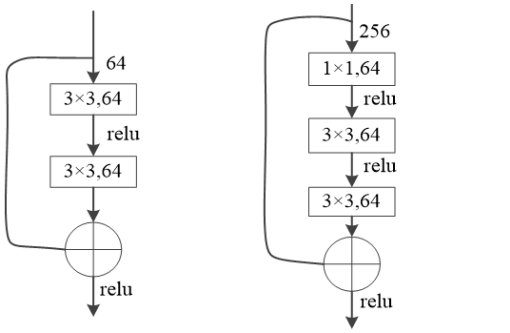
ResNet model has deeper layers than general convolutional neural network model. ResNet, with its deep layers, provides a wider range of local receptive area, which is more conducive to processing tasks for instance detecting local key points of human posture. Of course, related safety needs to be considered on the other hand [35]. For deep learning, as the number of convolutional layers increases, the extracted features gradually become more abstract and contain more semantic information from low to high dimensions. But if we simply stack the convolution layer to increase the depth, the result is that the convolutional neural network has vanishing gradient problem. The biggest advantage of ResNet model is to figure out the issue of vanishing gradient problem when the network level is very deep.

The ResNet network alleviates the problem of gradient vanishing by introducing cross layer connections. In the process of backpropagation, traditional deep neural networks multiply the gradient layer by layer by the weight matrix, causing the gradient to continuously shrink and eventually disappear. The cross-layer connections in the ResNet network can directly transfer the input gradient to the subsequent layers, thus avoiding the problem of gradient vanishing.

Generally, the expression of the loss function of the neural network is shown in equation (3).

$$Loss = F(X_L, W_L, b_L). \quad (3)$$

Among them, X_L represents the input data of the L-th layer, W_L represents the weight of the L-th layer, and b_L represents the bias of the L-th layer. The core content of ResNet is the residual structure block, and a residual block



(a) A residual block of 34 layer network of ResNet structure (b) A residual block of 50 and 101 layer network of ResNet structure

Fig. 1. Residual block comparison graph of ResNet.

for 34-layer network of ResNet structure is displayed in Fig. 1(a) [36]. For obtaining a deeper model, the 2-layer modules in the 34-layer network are replaced with the 3 layer "bottleneck" modules, which are 1×1, 3×3 and 1×1 convolution. The first 1×1 is used to reduce dimensions, the third 1×1 is used to restore dimensions. The remaining 3×3 layers reduce the dimensions of input and output. The residual blocks for 50-layer and 101-layer network of Res Net are displayed in Fig. 1(b) [37].

The pose image in the 101-layer network of ResNet passes through one convolutional layer, one maximum pooling layer, four residual modules, finally the pooling layer and fully connected layer. In order to adapt to the research task of key points, the original structure for 101-lay er network of ResNet is modified by removing the final pooling layer and fully connected layer. The seventh module is changed to a prediction layer, which outputs 14×2 dimensions of key point category prediction and 14 dimensions of key point coordinate prediction.

3.3. Method Driven Human Posture Research in Physical Education Teaching

With the popularization of the Internet, the core of physical education teaching is the effective detection of human posture. The quality of human posture analysis plays a vital role in the learning process of students, contributes to enhance the overall strength of physical education, and has an important impact on the application of physical education. Therefore, the algorithm of human posture detection based on deep learning is studied, and the method of human posture estimation based on ResNet is proposed. The loss function of key point positioning is improved, which is of tremendously significance to the progress of physical education and teaching in the new era.

3.3.1. ResNet Human Key Point Detection Model

The ResNet human posture estimation model is composed of two parts. The first is the training model of the ResNet human key point detection network, and the second is the testing model of the ResNet human key point detection network. The training model is divided into the following modules. The preprocessed human posture database corresponds to 14 key point annotations for each human body. It first passes through a convolutional layer, then passes through a pooling layer, and enters the residual block to obtain the category prediction and coordinate prediction of the 14 key points of this human body. Cross entropy loss is calculated for category prediction, and positioning loss is calculated for coordinate prediction. Then, a random gradient descent algorithm is adopted to iteratively optimize the weights and parameters of the network. After the training model converges, it is tested using unlabeled single person pose images. After passing through a convolutional and pooling layer, it enters the residual block to obtain 14 key point category predictions and coordinate predictions for this human body. The predicted key points are visualized and the coordinates of the key points are saved.

The two core issues of human key point detection are the probability that the key points belong to each of the 14 categories, and the *X* and *Y* coordinates of the key points in the image. The key point detection model adopts the 101-layer depth residual network of ResNet, and the loss function as the optimization goal consists of two parts. The one part is the classification error of human key points, which is defined as cross entropy loss. The other part is the positioning coordinate error of human key points, defined as the distance between predicted human key points and real human key points on the training dataset.

The distance positioning is divided into three loss function, namely mean square loss, smoothed L_1 loss, and huber loss Compared to the mean square loss, smoothed L_1 loss is less sensitive to outliers and can prevent gradient explosion problems. The 101-layer depth residual network of ResNet is trained using random gradient descent, and there are two prediction results for human body static frame images. The one is the category of human body key point coordinates, and the other is the position of human body key point coordinates. The total loss function is the total of the key point positioning loss and the key point coordinate regression loss, as shown in equation (4). Among them, $L_{classify}$is the regression loss of keypoint coordinates, and $L_{location}$ is the keypoint localization loss.

$$L_{total} = L_{classify} + L_{location}. \tag{4}$$

3.3.2. Loss Model for Positioning Key Points of Human Posture

For the research of human posture in the application of physical education teaching, two commonly used positioning loss function are set first. The first is the mean squared

error loss L_{mse}, and the second is the smooth L_1 loss, called $L_{smoothL1}$. At present, the location loss function that has been widely used is the smooth L_1 loss. SSD, other target detection frameworks all use the smooth L_1 loss function as the location loss of target detection.

Mean squared error loss function L_{mse} is the mean of the euclidean distance square of all sample estimates and predictions. It is defined as the sum of euclidean distances between predicted human key points and real human key points on each key point of each image on a batch of samples. x_i and y_i refer to the true coordinate values. x_i' and y_i' refer to the predicted coordinates. The number of samples are set to N, where i represents the i-th sample. A total of C keys is set for an image, as each key belongs to a different class. So, there are a total of C categories. The loss function is shown in equation (5).

$$L_{mse} = \frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{14}\left\{\left(x_i^{c\prime} - x_i^{c}\right)^2 + \left(y_i^{c\prime} - y_i^{c}\right)^2\right\}. \quad (5)$$

The smooth L_1 loss is proposed in the Fast RCNN network structure, and when the predicted value differs significantly from the target value, the gradient is prone to explosion. Compared with L_2's loss, L_1's loss is more robust to outliers in the data sample and less susceptible to the influence of sample data that may be noisy. Because the loss of L_2 is the main component of the loss when there are outliers. If the actual value is 1 and predicted 10 times, one time the predicted value is 1,000, and the other times the predicted value is around 1. It is obvious that the loss value is primarily controlled by 1,000. The original definition of smooth L_1 loss is shown in equation (6), where x is the imbalance between the predicted value and the true value.

$$smoothL1loss = \begin{cases} 0.5x^2 & if|x| < 1 \\ |x| - 0.5 & otherwise \end{cases}. \quad (6)$$

In the issue of human posture estimation and positioning, the smooth L_1 loss function is defined as the $L_{smoothL1}$ loss between the predicted human key points and the actual human key points on each key point of each image on a batch of samples. First, the difference is calculated between the predicted coordinates and the actual coordinates, and then incorporated this difference into the smooth L_1 loss function. Let the predicted key point position be expressed in vector form as $P_{predicti}$, and the key point position in the real labeled data be expressed in vector form as P_{labeli}. The difference between the two is denoted as $diff$, where i represents the i-th key point of a sample. $diff$ calculates the sum of the absolute values of the prediction errors of all key points on a sample, which is a scalar, as shown in equation (7). The loss function $L_{smoothL1}$ is displayed in equation (8).

$$diff = \sum_{i=1}^{14}\left|\vec{P}_{labeli} - \vec{P}_{predicti}\right|. \quad (7)$$

$$L_{smoothL1} = \begin{cases} 0.5(diff)^2 & if|diff| < 1 \\ |diff| - 0.5 & otherwise \end{cases}. \quad (8)$$

3.3.3. Improved Loss Model for Positioning Key Points of Human Posture

In the physical education teaching, human body movements are a continuous and time-varying signal, and the corresponding observation posture sequence is also continuous. Although there are various vector quantization coding methods to discretization continuous signal, human motion is a complex continuous time-varying signal, which may lead to a large amount of loss of effective information after vector quantization coding. So continuous hidden markov model (CHMM) is selected for human motion recognition. The description of human body posture still uses the relative distance between specific joints of the 3D skeleton, while the CHMM algorithm is used for human motion recognition.

Huber loss is used as the loss for target localization. It is commonly used in regression problems. Huber loss L_{huber}, the original function definition of huber loss, where the difference between the actual value and the predicted value is x, as shown in equation (9).

$$huberloss_k(x) = \begin{cases} 0.5(x)^2 & if|x| \le k \\ k|x| - 0.5k^2 & otherwise \end{cases}, \quad (9)$$

where k is a set parameter, usually taken as 1, 2 ,3. The impact of three parameters on the accuracy of key point detection is set to analyze. In the problem of human posture estimation and localization, L_{huberk} is defined as the huber loss between predicted and real human key points on each key point of each image on a batch of samples. First the difference is calculated between the predicted coordinate and the real coordinate, and then this difference is brought into the huber loss function. The prediction error uses equation (7), and the loss function L_{huberk} is shown in equation (10).

$$L_{huberk}(diff) = \begin{cases} 0.5(diff)^2 & if|diff| \le k \\ k|diff| - 0.5k^2 & otherwise \end{cases}. \quad (10)$$

The impact of the value of k on the final key point detection effect is analyzed. When k is taken as 1, the specific equation of the loss function $L_{huberk1}$ is shown in equation (11).

$$L_{huberk1}(diff) = \begin{cases} 0.5(diff)^2 & if|diff| \le 1 \\ |diff| - 0.5 & otherwise \end{cases}. \quad (11)$$

Comparing equation (10) and (8), it was found that when k=1, the form of huber loss degenerates into a smooth L_1 loss. When k=2, the loss function $L_{huberk2}$ is shown in equation (12). When k=3, the loss function $L_{huberk2}$ is shown in equation (13).

$$L_{huberk2}(diff) = \begin{cases} 0.5(diff)^2 & if|diff| \le 2 \\ 2|diff| - 2 & otherwise \end{cases}. \quad (12)$$

$$L_{huberk3}(diff) = \begin{cases} 0.5(diff)^2 & if|diff| \le 3 \\ 3|diff| - 4.5 & otherwise \end{cases}. \quad (13)$$

In the random gradient descent algorithm, the momentum coefficient is added. The prominent advantages of the momentum method are firstly that it enables the network to converge more optimally and stably. Secondly, it reduces the oscillation process. The essence of the momentum method is equivalent to accelerating the gradient descent by adding a momentum coefficient to the original velocity vector in the direction of gradient descent. The updated equation is shown in equation (14). Where $param$ represents the weight parameter of the neural network, v is the velocity, and $g(param)$ represents the derivative. The mu is the momentum coefficient, and it is set to 0.9 in the model. The lr is the learning rate.

$$\begin{matrix} v = mu \times v - lr \times g(param) \\ param = param + v \end{matrix}. \quad (14)$$

IV. EXPERIMENTS AND RESULTS

4.1. Evaluation of Human Posture Testing for Models

The operating system used in this experiment is Wi ndows or a higher version of the system. In the hardware configuration, the CPU model is Intel i7-8700k, the GPU model is RTX2080ti, the memory is 64 GB, and the hard drive is 1 TB. The development language supports Python 3.7 and JDK 17 or higher versions. The database platfor m uses MySQL 5.6.

On the dataset of physical education teaching, 300 pieces of data are randomly selected for testing. The evaluation index for human posture estimation uses the general indicator percent correct keypoints (PCK), which is defined as the amount of estimated key points in right way, accounting for the proportion of all key points in the entire test data, as shown in equation (15).

$$PCK = \frac{N_{keypointcorrect}}{N_{keypointsum}}. \quad (15)$$

$N_{keypointcorrect}$ represents the number of correctly estimated key points. $N_{keypointcorrect}$ represents the amount of all key points in the entire test data. Correct definition of key point estimation is that the euclidean distance between the estimated and the real coordinates is less than a specific threshold, which is set to 17×17. It means that the key point coordinates must be estimated correctly in the circular box with the radius of 17 as the real key point coordinates. If the threshold value is exceeded, the key point estimation is considered to be estimated incorrectly. Fig. 2 shows the PCK corresponding to the loss models.

Quantitative analysis of the experimental results was carried out to compare the positioning loss function. The positioning accuracy at key points was convenient. The best model was the huber loss model for key point positioning, followed by the smooth L_1 loss model for key point positioning, and finally the mean square error loss model for key point positioning. The reason for this result is huber loss. The smooth L_1 loss model is less sensitive to outliers in the data than the mean square error loss. Huber loss and smooth L_1 loss have higher robustness than mean square error loss. The accuracy of the model with three parameters k=1, 2, 3 are compared for using of huber loss as a loss function. The accuracy of the model corresponding to k=1 is 78.3%, the accuracy of the model corresponding to k=2 is 79.2%, and the accuracy of the model corresponding to k=3 is 77.4%. Therefore, the optimal huber parameter is k=2.

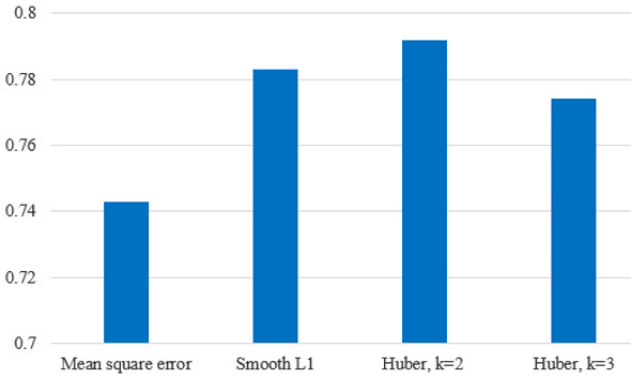


Fig. 2. PCK corresponding to four loss models.

4.2. System Experiment Results

The detection performance indicator used in the experimental evaluation section is *mAP*, which is the average accuracy mean and the average value of each category of *AP*. Its value range is between 0 and 1. The calculation is shown in equation (16).

$$mAP = \frac{1}{m} \times \sum_{i=1}^{m} AP_i. \quad (16)$$

Among them, *AP* is the average accuracy, which is the integral of the *P*-*R* curve. *P* represents the accuracy, and *R* represents the recall rate, as shown in equations (17) and (18).

$$P = \frac{TP}{TP+FP}. \quad (17)$$

$$R = \frac{TP}{TP+FN}. \quad (18)$$

TP is the true sample, representing the number of successful predictions of positive classes as positive classes. *FP* is the false positive sample, representing the amount of errors in predicting negative classes as positive classes. *FN* is the false negative sample, representing the amount of errors in predicting positive classes as negative classes. And *TN* is the amount of successful predictions of negative classes as negative classes. In addition, this experiment selected

IPR [38] and LAIPR [39] to conduct precision detection and evaluation experiments on the deep learning model on the dataset of physical education teaching. The IPR considers the characteristics of personalized web pages and conveys corresponding value based on differences in permissions. By comparing the evaluation results, they were analyzed. The experimental results recorded using MATLAB are shown in Fig. 3, Fig. 4 and Fig. 5 respectively.

Through the contrast of the deep learning model in Fig. 3, it can be seen that with the increase of the human posture dataset in physical education teaching, the corresponding accuracy gradually increases. During the training process of the dataset, the average accuracy of the ResNet based human posture estimation method is higher than traditional IPR and LAIPR in analysis, indicating that the prediction results of the ResNet based human posture estimation method are close to the actual situation.

Through the application comparison in Fig. 4, the ResNet based human posture estimation method has good manifestation in the recall rate of the test dataset and is in a high position in the training set. The accuracy of the ResNet based human posture estimation method steadily increases with the increase of the sample set. And compared to traditional IPR and LAIPR, the research results on human posture in physical education teaching are more accurate.

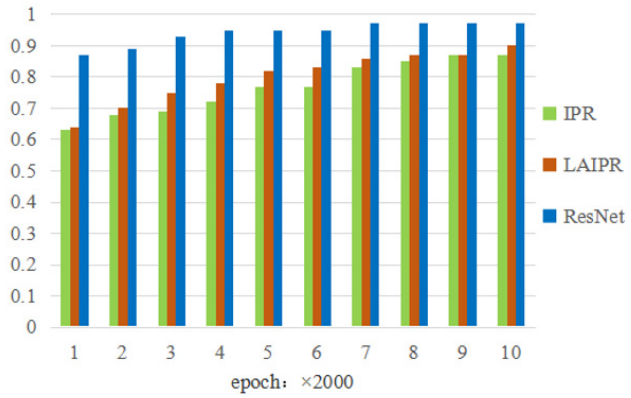


Fig. 3. Comparison of *P* changes among IPR, LAIPR, and ResNet.

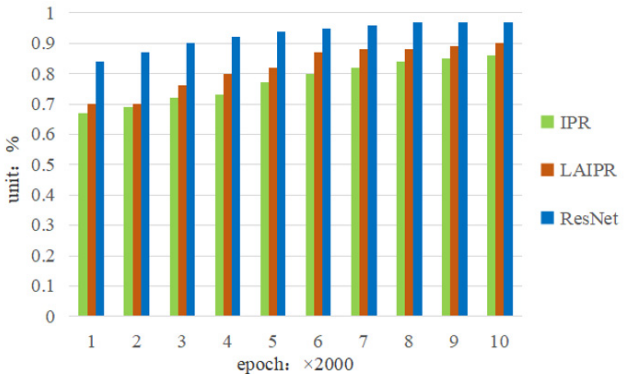


Fig. 4. Comparison of *R* changes among IPR, LAIPR, and ResNet.

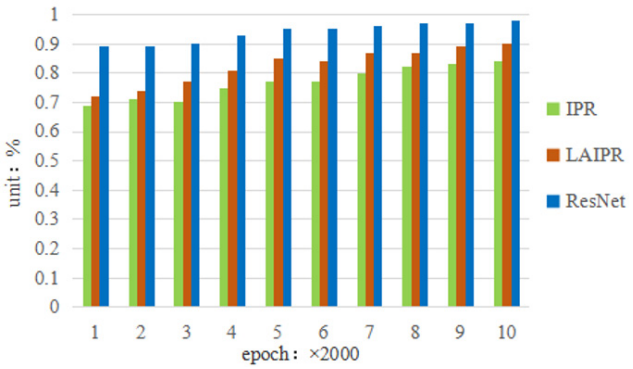


Fig. 5. Comparison of *mAP* changes among IPR, LAIPR, and ResNet.

The human posture estimation method based on ResNet has proved its superiority compared with other methods. From the application comparison in Fig. 5 that in physical education teaching, the human posture estimation method based on ResNet has great advantages in recognition speed and accuracy with the increase of human posture sample set, so it is recommended to use the human posture estimation method based on ResNet in physical education teaching.

In summary, in human posture research, ResNet based human posture estimation method perform best among various indicators. Therefore, in the teaching mechanism of physical education, it is recommended to use ResNet based human posture estimation methods for human posture recognition, which has practical significance for physical education teaching and application.

V. CONCLUSION

Based on the perspective of the intelligent era and traditional physical education teaching, this article utilizes deep learning technology in the area of artificial intelligence to study human posture. Human pose esti-mation is usually the basis for accurate recognition of human movements. The objective of human posture research is to discover different parts of the human body and estimate the coordinates of key points of joints. In response to the problem of low accuracy in traditional algorithms, this paper studies a deep learning based human posture detection algorithm, using the 101-layer network of ResNet, and proposes a human posture estimation method based on ResNet. And an improved key point location loss function is proposed. Huber loss fun- ction is employed to compute the location loss. Three different key point location loss function are set, and the impact of different location loss function on the accuracy of the model is elaborated. Experiments show that huber error loss is used to replace mean square error loss to improve the accuracy of the model. Finally, based on experimental consequences, it is shown that the ResNet based human posture estimation method has higher recognition ac-

curacy and better robustness compared to traditional IPR and LAIPR. This proves the effectiveness of the ResNet based human posture estimation method, which has practical significance for the deep learning driven human posture research in physical education teaching.

REFERENCES

[1] J. Aurini, R. Missaghian, and R. P. Milian, "Educational status hierarchies, after-school activities, and parenting logics: Lessons from Canada," *Sociology of Education*, vol. 93, no. 2, 173-189, 2020.

[2] Q. Jiang and T. Mao, "Research on future education development under the trend of information technology and artificial intelligence in the sixth scientific and technological revolution," in *2021 2nd International Conference on Artificial Intelligence and Education (ICAIE),* IEEE, 2021, pp. 591-595.

[3] J. R. Thomas, P. Martin, J. L. Etnier, and S. J. Silverman, *Research Methods in Physical Activity.* Human Kinetics, 2022.

[4] Z. Yang, "Research on basketball players' training strategy based on artificial intelligence technology," in *Journal of Physics: Conference Series. IOP Publishing*, 2020, vol. 1648, no. 4, p. 042057.

[5] J. Miao, Z. Wang, X. Miao, and L. Xing, "A secure and efficient lightweight vehicle group Authentication protocol in 5G networks," *Wireless Communications and Mobile Computing*, vol. 2021, pp. 1-12, 2021.

[6] Q. Li, P. M. Kumar, and M. Alazab, "IoT-assisted physical education training network virtualization and resource management using a deep reinforcement learning system," *Complex & Intelligent Systems*, pp. 1-14, 2022.

[7] S. Wang and M. N. Bin Nazarudin, "Research on the application of virtual reality technology in physical education in colleges and universities," in *International Conference on Information Systems and Intelligent Applications: ICISIA 2022,* Cham: Springer International Publishing, 2022, pp. 371-379.

[8] V. Sorin, Y. Barash, E. Konen, and E. Klang, "Deep learning for natural language processing in radiology—fundamentals and a systematic review," *Journal of the American College of Radiology*, vol. 17, no. 5, pp. 639-648, 2020.

[9] J. Miao, Z. Wang, X. Ning, N. Xiao, and R. Liu, "Practical and secure multifactor authentication pro-tocol for autonomous vehicles in 5G," *Software: Practice and Experience*, 2022.

[10] A. Dhillon and G. K. Verma, "Convolutional neural network: A review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence*, vol. 9, no. 2, pp. 85-112, 2020.

[11] H. S. Fang, J. Li, H. Tang, C. Xu, H. Zhu, and Y. Xiu, et al, "Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[12] K. Hu, J. Jin, F. Zheng, L. Weng, and Y. Ding, "Overview of behavior recognition based on deep learning," *Artificial Intelligence Review*, vol. 56, no. 3, pp. 1833-1865, 2023.

[13] J. Miao, Z. Wang, M. Wang, X. Feng, N. Xiao, and X. Sun, "Security authentication protocol for massive machine type communication in 5G networks," *Wireless Communications and Mobile Computing*, vol. 2023, pp. 1-13, 2023.

[14] X. Ren, A. C. Berg, and J. Malik, "Recovering human body configurations using pairwise constraints between parts," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1.* IEEE, pp. 824-831, 2005.

[15] G. Hua, M. H. Yang, and Y. Wu, "Learning to estimate human pose with data-driven belief propagation*," in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05).* IEEE, 2005, pp. 747-754.

[16] G. Mori and J. Malik, "Estimating human body configurations using shape context matching," in *ECCV,* vol. 3*,* 2002, pp. 666-680.

[17] J. Martinez, R. Hossain, J. Romero, and J. J.Little, "A simple yet effective baseline for 3D human pose estimation," in *Proceedings of the IEEE International Conference on Computer Vision,* 2017, pp. 2640-2649.

[18] L. Zhao, X. Peng, Y. Tian, M. Kapadia, and D. N. Metaxas, "Semantic graph convolutional networks for 3D human pose regression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recog-nition*, 2019, pp. 3425-3435.

[19] D. Pavllo, C. Feichtenhofer, D. Grangier, and M. Auli, "3D human pose estimation in video with temporal convolutions and semi-supervised training," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7753-7762.

[20] W. Yang, W. Ouyang, X. Wang, J. Ren, H. Li, and X. Wang, "3D human pose estimation in the wild by adversarial learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2018, pp. 5255-5264.

[21] M. Oberweger, M. Rad, and V. Lepetit, "Making deep heatmaps robust to partial occlusions for 3D object pose estimation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 119-134.

[22] D. Xiang, H. Joo, and Y. Sheikh, "Monocular total capture: Posing face, body, and hands in the wild," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10965-10974.

[23] G. Pavlakos, X. Zhou, K. G. Derpanis, and K. Daniilidis,"Coarse-to-fine volumetric prediction for single-image 3D human pose," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7025-7034.

[24] S. Park, J. Hwang, and N. Kwak, "3D human pose estimation using convolutional neural networks with 2D pose information," in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands,* Springer International Publishing, 2016, pp. 156-169.

[25] A. Kanazawa, J. Y. Zhang, P. Felsen, and J. Malik, "Learning 3D human dynamics from video," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 2019, pp. 5614-5623.

[26] G. Pavlakos, X. Zhou, K. G. Derpanis, and K. Daniilidis, "Coarse-to-fine volumetric prediction for single-image 3D human pose," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7025-7034.

[27] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden Markov model," in *Proceedings of the Computer Vision and Pattern Recognition*, 1992, pp. 379-385.

[28] N. T. Nguyen, D. Q. Phung, S. Venkatesh, and H. Bui, "Learning and detecting activities from movement trajectories using the hierarchical hidden Markov model," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 955-960.

[29] P. Natarajan and R. Nevatia, "Online, real-time tracking and recognition of human actions," in *Proceedings of the IEEE Workshop on Motion and Video Computing*, 2008, pp. 1-8.

[30] H. Ren, G. Xu, and S. C. Kee, "Subject-independent natural action recognition," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, 523-528.

[31] Y. Luo, T. D. Wu, and J. N. Hwang, "Object-based analysis and interpretation of human motion in sports video sequences by dynamic Bayesian networks," *Computer Vision and Image Understanding*, vol. 92, no. 2, pp. 196-216, 2003.

[32] D. Weinland, R. Ronfard, and E. Boyer, "Free viewpoint action recognition using motion history volumes," *Computer Vision and Image Understanding*, vol. 104, no. 2, pp. 249-257, 2006.

[33] S. Park and J. K. Aggarwal, "A hierarchical Bayesian network for event recognition of human actions and interactions," *Multimedia Systems*, vol. 10, no. 2, pp. 164-179, 2004.

[34] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257-267, 2001.

[35] J. Miao, Z. Wang, X. Xue, M. Wang, J. Lv, and M. Li, "Lightweight and secure D2D group communication for wireless IoT*," Frontiers in Physics*, vol. 11, p. 433, 2023.

[36] X. Li, N. Li, C. Weng, X. Liu, D. Su, and D. Yu, et al, "Replay and synthetic speech detection with res2net architecture," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 6354-6358.

[37] F. Liu, J. Liu, J. Fu, and L. U. Hanqing, "Improving residual block for semantic image segmentation," in *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM).* IEEE, 2018, pp. 1-5.

[38] A. Sprake and C. A. Palmer, "PE to Me: A concise message about the potential for learning in physical education," *Journal of Qualitative Research in Sports Studies*, vol. 13, no. 1, pp. 57-60, 2019.

[39] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 2019, pp. 5693-5703.

Journal of Multimedia Information System VOL. 11, NO. 1, March 2024 (pp. 45-56): ISSN 2383-7632 (Online)
https://doi.org/10.33851/JMIS.2024.11.1.45

AUTHORS

Shaohua Wang received his master degree from Jilin Normal University. He is currently working at Jilin Technology College of Electronic Information as a Associate Professor. His main research interests include sport education, AI, etc.

Wanli Shi received his master degree from Beijing Sport University. He is currently working at Jilin University of Agricultural Science and Technology as a Senior Lecture. His main research interests include sport education, AI, etc.