WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 3, Number 3, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v3n3.45
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Body Recognition Sports APP Based on Android System

Gaorui Zhang [a], Han Wen [b], Huiqin Sun [c], Sitong Chen [d], Lingling Luo [e], Xiaohong Wang [f, *]

School of Information Engineering, Wuhan Business University, Wuhan 430056, China
[a] 2486846653@qq.com, [b] 1256948509@qq.com, [c] 3498065134@qq.com, [d] 3400918243@qq.com, [e] 360386367@qq.com, [f] 241544488@qq.com

ABSTRACT

With the enhancement of health awareness, people are more and more inclined to improve the quality of life through exercise, but the wrong way of exercise may lead to muscle strain and other problems. To solve this problem, this study designed a motion aid APP integrated with posture recognition technology, aiming at guiding users to correct movements through real-time feedback. The APP's core function, "AI online guidance," combines computer vision and image recognition technology to capture user movement data through the camera and analyze it using deep learning algorithms to identify and correct the user's movement posture. In this study, we first discuss three main methods of human behavior recognition: based on deep neural networks, based on 3D residual model and based on 3D attention mechanism. The application and structure design of 3D shallow feature extraction module and 3D attention mechanism in human behavior recognition are introduced. Finally, this study describes the functional implementation of the APP, including action analysis, error point identification, improvement suggestion generation, and user interaction design. The results of this study show that the APP can effectively improve the accuracy of users' movements during sports, reduce the risk of sports injuries, increase the interest of sports, and improve users' subjective initiative.

KEYWORDS

Image recognition; CNN Convolutional neural network; 3D residual model; Human behavior recognition research; AI online guidance function

1. INTRODUCTION

In recent years, as people pay more and more attention to their own health problems, the subjective initiative of sports is enhanced, and people are more willing to achieve the purpose of relieving pressure and relaxing their body and mind through sports. However, the ensuing, is because the action is not standard enough, the exercise did not stretch in time, etc., resulting in muscle strain, pain, weakness after exercise, this result runs counter to the original intention of our movement. At the same time, with the deepening of research on the human body, some people have proposed the use of human bone points to analyze the dynamics of the human body, which is the so-called "posture recognition".

In recent years, researchers have done a lot of research on intelligent methods of body health diagnosis. Gao et al. have developed a body health detection system named SIMIMMOTION [1], which includes a camera to record human posture and a posture assessment software to capture and analyze body posture images to understand students' physical health. Meanwhile, relevant foreign studies have developed many software for Posture evaluation, such as Body style, PostureSprint, Posture Screen, etc. [1]. They all use different instruments and equipment, such as depth camera and infrared imaging camera, to collect, analyze and evaluate human body posture. The pictures obtained in the

instrument and equipment are input into the designed software for processing, and the system analysis function of the software is used for evaluation.

Therefore, we propose to integrate posture recognition technology with sports APP, called "AI online guidance" function. In the process of people's movement, the user's bone point is positioned, compared with the original movement, to find out the error of the movement, provide guidance in time, find the force point for the user, improve the accuracy of the movement, so as to obtain satisfactory sports results.

2. FUNCTIONAL REQUIRED TECHNOLOGY

"AI online guidance" function is to import the user's movement data into the database through the camera, and then use the posture recognition technology to analyze the user's movement posture, and generate guidance and suggestions for the user combined with AI.

2.1. Research on Human Behavior Recognition

There are three kinds of human behavior recognition applied to this function: [2-3] Human behavior recognition based on deep neural network, [4-5] Human behavior recognition based on 3D residual, [6].

2.1.1. Based on deep neural networks

At present, [deep neural network] [7-9] in the field of video human behavior recognition is mainly divided into two branches: First, 2D CNN is used for feature extraction, which is represented by two-stream CNN. It uses Two 2D CNN models to extract and classify the spatio-temporal features of RGB images and optical Stream images, and then fuses the respective results through SVM classifier. The second is to extract the spatio-temporal features of the video frame sequence directly through 3D CNN[10], such as the classical C3D model, and then use the Softmax classifier for classification. However, the existing work still has some shortcomings, which are mainly reflected in the following aspects: (1) The extraction of spatio-temporal features of Two-Stream CNN model is independent, which is easy to ignore its internal relations, thus affecting the recognition performance; (2) When using the Two-Stream network to identify human behavior, it is generally impossible to use a single RGB picture. It is also necessary to preprocess RGB images to extract time sequence features, such as optical flow extraction and motion trajectory representation, which indirectly increases the cost of model training. (3) Although 3D CNN can extract spatiotemporal features of video sequences at the same time, the increase of model parameters in 3D convolutional operation is exponential compared with 2D convolutional operation, which makes it difficult to train and transfer the model, thus affecting the actual use effect of the model. (4) The 3D convolution operation cannot distinguish the background features from the human behavior subject features during feature extraction, and does not capture the associated features before and after the sequence data. As a result, the recognition of the model is susceptible to environmental factors, thus reducing the recognition performance of the model. Therefore, based on 3D CNN, we improve the existing 3D CNN model by combining residual structure and attention mechanism, and propose three categories of human behavior recognition models. They first use the 3D shallow feature extraction module to extract shallow features from video sequences, and then extract richer deep features through the 3D residual module, 3D attention mechanism or the deep feature extraction module of the two fusion, and finally classify and output the recognition results.

2.1.2. Based on 3D residual

The model strengthens the extraction of deep features by deepening the number of layers of the network, but it leads to the problem that the model parameters are too large and difficult to train. In order to give full play to their respective advantages, and restrain their respective disadvantages. In

this section, a 3D shallow feature extraction module is proposed and designed to extract the shallow feature information of video sequences, and according to the characteristics and properties of the residual structure, a residual module suitable for 3D convolution is designed to extract richer deep behavioral features. 3D Deeper Residual Model (R3D) is formed by combining 3D shallow feature module and 3D residual module. The overall structure of the corresponding model is shown in Figure 1.

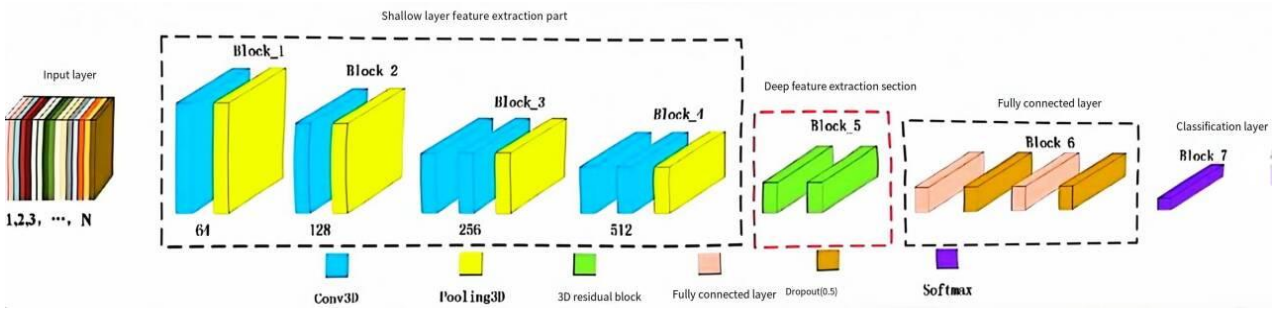

Figure 1. Schematic diagram of R3D overall structure

As can be seen from Figure 1, the R3D model consists of input layer, feature extraction layer, full connection layer and classification layer respectively. Where the input layer is composed of n1,2,3,... ,N consecutive RGB (or other format) image composition; The feature extraction layer includes two parts: shallow feature extraction module and deep feature extraction module. The full connection layer consists of two parts: full connection operation and neuron deactivation operation (Dropout). The introduction of Dropout is mainly to increase the probability of random deactivation of the model and reduce the number of connections between neurons, so as to prevent the risk of overfitting the model. The last is the classification layer, which is mainly for the classification of human behavior and corresponds to multi-category output, so Softmax classifier is used in the classification layer. Let's introduce the 3D shallow feature extraction module in the R3D model.

The 3D shallow feature extraction module refers to the structural design of the C3D model, and takes the original picture sequence of consecutive frames as input, and then extracts the shallow feature information of human behavior through its internal convolution layer and pooling layer in turn, and the final output is used as the input of the deep feature extraction module. The corresponding structure is shown in Figure 2.

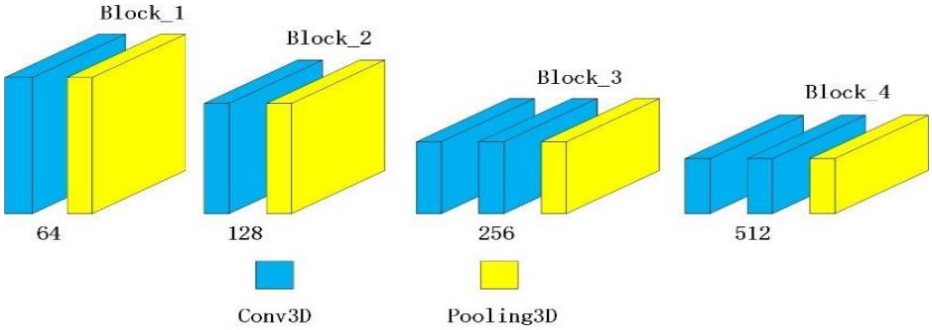

Figure 2. Structure of 3D shallow feature extraction module

For the designed 3D shallow feature extraction module, two operations are mainly carried out: 3D convolution and 3D pooling. Therefore, the next step is to introduce in detail how to extract shallow spatio-temporal features. 3D convolution operation refers to the use of three-dimensional convolution kernel (the size of convolution kernel adopted in this paper is 333) to extract features from the input continuous multi-frame image sequence, mainly because the input image sequence not only contains features of spatial dimension, but also contains timing information on time series. Different from 2D convolution operation, 3D convolution adds feature extraction operation in time dimension, and its schematic diagram of time sequence feature extraction is shown in Figure 3.

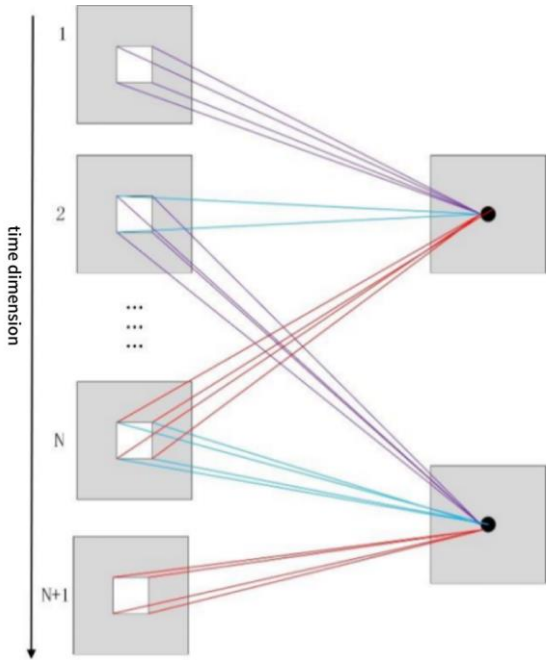

Figure 3. 3D convolution sequence feature extraction

In Figure 3, it is assumed that there are N +1 frames of pictures in the time dimension, the time dimension of 3D convolution operation is N (that is, convolution operation is performed on consecutive N frames of pictures), and the number of 3D convolution cores is 1. Then a 3D local feature is obtained after the 3D convolution operation, see equation (1-1). From the color of the lines in the figure, it can be seen that the same convolution kernel can only extract one type of feature, and its corresponding weights are shared (the same color represents the same weight).

$$C_{seq}^{i}=\mathrm{f}(w_{seq}x_{i:i+n-1}+b_{seq}) \tag{1-1}$$

Here, W_{seq} is the weight parameter matrix of the three-dimensional convolution kernel, $b_{seq}\in$ R is the corresponding offset value, $X_{i:i+n-1}$ is the sequence of consecutive n-frames of the corresponding input, and f is the corresponding activation function. The activation function used in the convolutional layer in this paper is ReLU. For all possible picture frame sequences {x1: n, x2: n+1, ... When xN−n+1: N} is convolved, the corresponding eigenmatrix C_{seq} is obtained, as shown in equation (1-2).

$$C_{seq}=[C_{seq}^{1},C_{seq}^{2},\ldots.,C_{seq}^{N-n+1}] \tag{1-2}$$

All convolution operations adopt the filling mode of "SAME". Compared with another "VALID" mode of convolution], a border filling operation of P size is added to the input feature graph. Suppose that the dimension corresponding to the input feature matrix is (B, T_{in}, H_{in}, W_{in}, C_{in}), and the number of convolution cores is N. If the size of the convolution kernel is K$\in R^{x*y*z}$, and the convolution step is S$\in R^{x*y*z}$, then the output dimension of the corresponding feature graph obtained after the convolution operation is (B, T_{out}, H_{out}, W_{out}, C_{out}), The specific mathematical expressions corresponding to the size of each dimension are shown in equations (1-3) to (1-6).

$$T_{out}=\left\lfloor\frac{T_{in}-d[0](K[0]-1)+P[0]-1}{S[0]}+1\right\rfloor \tag{1-3}$$

$$H_{out}=\left\lfloor\frac{H_{in}-d[1](K[1]-1)+P[1]-1}{S[1]}+1\right\rfloor \tag{1-4}$$

$$W_{out}=\left\lfloor \frac{W-d[2](K[2]-1)+P[2]-1}{S[2]}+1 \right\rfloor \quad (1\text{-}5)$$

$$C_{out}=N \quad (1\text{-}6)$$

There, the symbol "$\lfloor\ \rfloor$" is rounded down, B is the size of the input data batch, T is the size of the time dimension, H is the height, W is the width, C is the number of channels, and d is a constant (the default value is 1). The 3D pooling operation is to compress the feature map output of the convolution layer. The degree of compression is determined by the size of the pool core, the filling mode of the pool and the step size of the pool operation. Due to the need to retain more texture information of shallow features, this paper adopts the maximum pooling method to pool the feature graph, C_{seq} and the resulting pooling feature is C_{seq}, as shown in equation (1-7).

$$\widehat{c_{seq}}=max(c_{seq}) \quad (1\text{-}7)$$

Note: The specific calculation method of each dimension size of the output feature graph after pooling operation is basically the same as that of the convolutional calculation method, which can be referred to equations (1-3) to (1-6). The only difference is that the calculation result of each output dimension size of the pooling operation is rounded up.

2.1.3. Based on 3D attention mechanism

In video-based human behavior recognition, since the input processed is generally a continuous multi-frame image sequence, compared with the input recognition of a single image, in addition to the characteristics of the behavior itself, there may be a certain relationship between the continuous multiple frames, as shown in Figure 4.

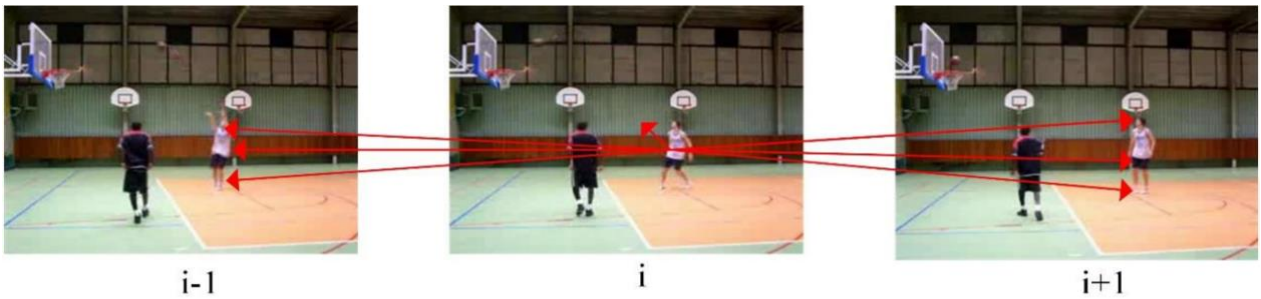

Figure 4. Schematic diagram of the connection between successive frames

In Figure 4, i represents the current action frame, I-1 represents the previous frame of the current frame, and i+1 represents the subsequent frame of the current frame. From the figure, we find that the current action behavior is not only affected by its own state, but also may be affected by its predecessor and subsequent behavior. Therefore, while extracting the spatio-temporal features of the image sequence, if the connection between the frames of the video sequence can also be captured, it is possible to improve the effect of human behavior recognition.

Since the only difference between A3D model and R3D model is that the deep feature extraction module adopts 3D attention mechanism module, the structure of 3D attention mechanism is mainly introduced next.

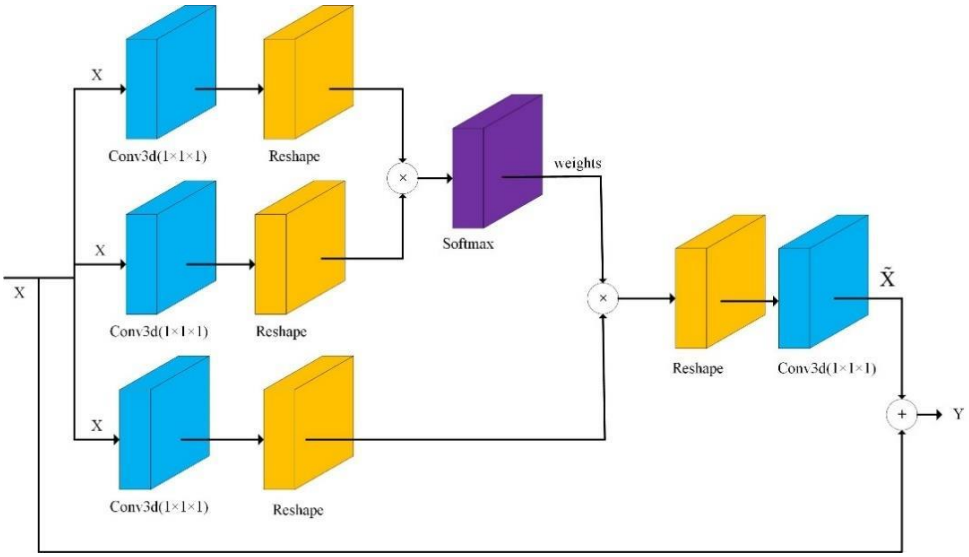

Figure 5. Structure of 3D attention mechanism

As can be seen from Figure 5, the 3D attention mechanism structure consists of three convolution branches and one join addition operation. The convolution branch can be divided into two different paths (from top to bottom) : the first and second two convolution branches constitute the attention weight extraction part; The third convolution branch is the feature extraction part.

For the attention weight extraction, first, perform1×1×1 convolution on the input feature graph to compress the channel feature size, and reduce parameters and computation amount. Then, Reshape the feature graph on the two-dimensional matrix through the reshape layer, and multiply the matrix to calculate the similarity and get the corresponding weight values. Finally, Softmax function is used for normalization operation to convert the calculated result into a probability value between [0,1]. The probability reflects the allocated attention. Embedded Gaussian Function is used in the calculation of similarity in this paper, which is a simple extension on the basis of ordinary Gaussian function. For the definition of ordinary Gaussian function, see equation (1-8). The Embedded Gaussian function is defined in equation (1-9).

$$G(x_i, x_j) = e^{x_i{}^T x_j} \tag{1-8}$$

$$G_{emd}(x_i, x_j) = e^{\theta(x_i)^T \varphi(x_j)} \tag{1-9}$$

There, i represents the output position, j represents all possible positions associated with i, x represents the received input, and $x_i^T x_j$ represents the dot multiplication operation used to calculate the similarity between the corresponding positions $x_i$and x_j.

For the feature extraction branch, first, a 1×1×1convolution layer is applied to the input feature map to extract features and reduce its dimensions, and then Reshape the feature map into a two-dimensional matrix through the reshape operation. Finally, multiply the obtained results with attention weights and output the allocated attention feature map.

In order to embed the 3D attention mechanism structure into any network structure, it is necessary to ensure the consistency of input and output dimensions. Therefore, before the final attention feature x is output, we added the Reshape layer and 1×1×1 convolution layer to restore the dimension of the feature map.

The operation branch is connected, and the original input x received by the 3D attention block is directly added to the attention feature graph x of the final output by referring to the design of the 3D residual structure, and then y is output as the input of the next module of the neural network. The

purpose of this design is mainly due to the phenomenon that the gradient may disappear in the addition of this module in the deep network.

To sum up, the specific implementation of the 3D attention mechanism module can be abstracted and expressed by the following formula, as shown in equations (1-10) to (1-13).

$$\delta' = \text{Softmax}((\text{Ro}(\text{Convox})) \otimes (\text{Ro}(\text{Convox}))) \tag{1-10}$$

$$\delta'' = \text{R}\circ(\text{Conv}\circ\text{x}) \tag{1-11}$$

$$\tilde{x} = Conv(R\circ(\delta' \otimes \delta'')) \tag{1-12}$$

$$\text{y}=\text{w}\tilde{x}\oplus\text{x} \tag{1-13}$$

Among them, "o" represents the operation, "Conv" represents the convolution operation, "R" represents the Reshape operation, "⊗" represents the matrix multiplication operation, and "⊕" represents the matrix addition operation.

2.2. Image Recognition

In recent years, with the wide application of [deep learning algorithm] [3], image recognition technology has made great progress. By building a multi-layer neural network model, deep learning can automatically learn higher-level abstract features from data, thus improving the accuracy of image recognition. By learning image recognition and computer vision, we can identify whether the approximate skeleton and motion trajectory of the exerciser are standard. (As shown in Figure 6)

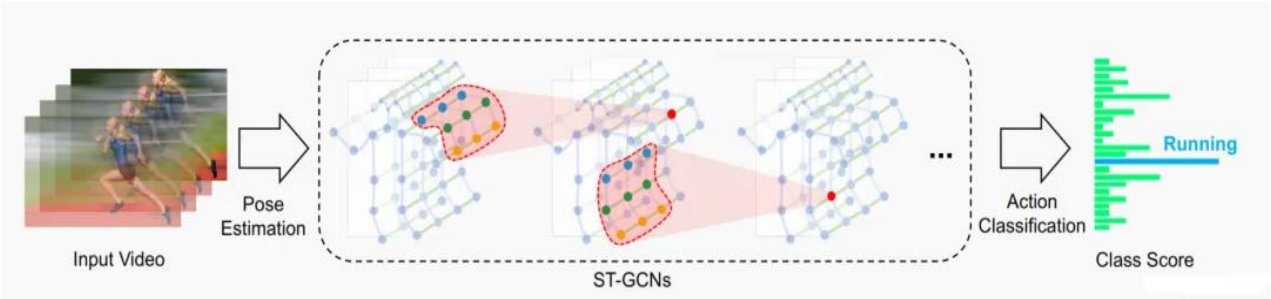

Figure 6. Identifying motion trajectories by deep learning

Skeleton-based data can be obtained from motion capture devices, as well as pose estimation algorithms from video. Usually the data is a sequence of frames, each with a set of joint coordinates. The body joint coordinates are given in 2D or 3D form. We construct a space-time graph with joints as graph nodes and the natural connection of human structure and time as graph edges.

2.3. Computer Vision

Computer Vision [11] is a field of artificial intelligence (AI) that aims to enable computers and systems to take meaningful information from images, videos, and other visual inputs and to act on or make recommendations based on that information. Computer vision works similarly to human vision, except that it is done in short bursts using cameras, data and algorithms, rather than relying on the human retina, optic nerve and visual cortex. It involves extracting the most representative and distinguishing feature information from the massive image or video raw data. Feature extraction technology has been widely used in many fields such as image retrieval, 3D reconstruction, face recognition, etc. With the development of deep learning technology, new feature extraction methods such as convolutional neural network (CNN) are also constantly improving the performance and generalization ability of computer vision algorithms.

3. "AI ONLINE GUIDANCE" FUNCTION IMPLEMENT

This function first uses computer vision and image recognition, identifies the body frame through the user camera, carries out dynamic tracking, and imports the input data into the background. 3D CNN is used to directly extract the spatiotemporal features of the video frame sequence, attach points to the skeletal parts of the human body (as shown in Figure 7), [skeleton diagram] [12], and take the 3D shallow feature extraction module as input to extract the shallow feature information of human movements through its internal convolution layer and pooling layer. This process can determine the user's current action and classify the user's action data with the Softmax classifier [13].

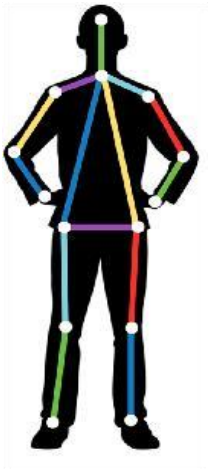

Figure 7. Recognition of bone points in human movement

Before human pose recognition, the APP first needs to pre-process the image data captured from the camera. This includes steps such as resizing the image, normalizing pixel values, and de-noising to improve the accuracy of model recognition. In addition, in order to increase the generalization ability of the model, we use data enhancement technology to generate more training samples by applying rotation, scaling, cropping and other transformations to the image.

Then it passes through the convolutional layer and the normalization layer in the 3D residual module in order to extract more rich deep feature information, and the final output is used as the input of the fully connected layer. The combination of these two modules can analyze and process human movements from shallow to deep. After processing, the action in the tutorial is decomposed in the same step, and the force point is determined by different parts of the exercise. The computer vision is used to analyze the user's action, determine the standard degree of the action, find out the error point, and generate improvement suggestions.

The use of computer vision technology for human posture recognition is the core of this APP. Through the image data captured by the user's camera, we use the deep learning model to estimate the human pose in real time. These models, often based on convolutional neural networks (CNNS), are trained on large amounts of pose data to accurately identify key points of the human body in images (such as the head, shoulders, elbows, wrists, hips, knees, and ankles). Once these key points are identified, the APP can track the movement of these points in the video sequence, forming a dynamic capture of human movement.

Spatiotemporal feature extraction from captured video frame sequences using 3D CNN is a key step in human motion analysis. 3D CNNS are capable of processing data in both spatial and temporal dimensions, making them particularly suitable for processing video data. By passing data in the 3D convolutional layer and the pooling layer, the model is able to effectively learn shallow features of human movements from video.

After shallow feature extraction, the data is passed to the 3D residual network. Through its deep structure, 3D residual networks can extract more complex and abstract deep features. These

characteristics are crucial for distinguishing between different modes of action. The deep features are then passed to the fully connected layer, and finally the action is classified by the Softmax classifier to achieve accurate recognition of the user's actions.

Once the system has identified and classified the user's movements, it compares them to a preset library of standard movements to find out where the user's movements fall short. Using image processing and analysis techniques, the APP is able to generate targeted improvement suggestions and present them to the user in an intuitive way, such as by showing the correct actions through graphics or animations.

When the system detects a deviation in the user's actions, it not only points out the error, but also provides suggestions for improvement. These recommendations are based on best practices in exercise science and physical therapy, and are algorithmically generated to provide targeted guidance. For example, if the user's knee is too far forward when performing a squat, the system will prompt the user to adjust the knee position and may recommend some additional exercises to strengthen the muscles in the back of the thigh.

In the function, we also designed: friends sharing function, at the same time of exercise, you can discuss the movement with friends, encourage each other, increase the fun of the exercise process; Forwarding function, users can share their own identification information to friends, and share AI online suggestions with friends; Background setting function, users can set the length of each online guidance, the page decoration of the guidance function, privacy information, camera length and width, page font size, etc., so that users can use more comfortable. The functional interface diagram of the APP is shown in Figure 8.

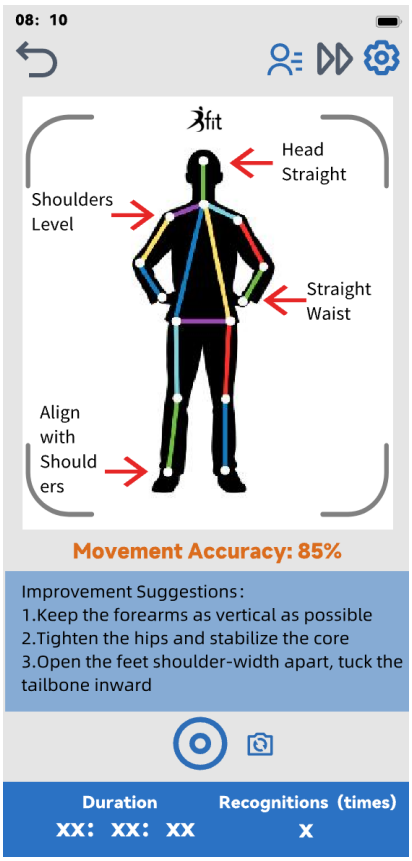

Figure 8. Functional UI interface diagram

In terms of user interaction, we designed an interactive user interface that allows users to get immediate feedback when using the AI online guidance function. When the system detects a large deviation from the standard movement of the user's posture, it will immediately display corrective

suggestions and demonstration videos. In addition, users can record their exercise journey, including the amount of exercise completed, history of improvement, and future training plans.

To enhance the user experience, we also have a dynamic difficulty adjustment mechanism. The system will gradually adjust the difficulty and complexity of the exercise according to the user's performance, ensuring that the user can maintain the challenge while avoiding injury.

In order to increase user engagement and motivation, the APP has designed an achievement system. Achievement badges and rewards can be unlocked when users complete specific exercise tasks or reach certain fitness goals. In addition, the APP may offer virtual items and health points, which can be redeemed for in-kind rewards or discounts to encourage users to stay active.

In terms of social features, we built a platform that allows users to create profiles and share their sporting achievements and experiences. Users can choose to share their sports data and progress with friends or social groups, thus increasing the transparency and competition of sports. In addition, users can participate in exercise challenges and events set by the APP operations team or third-party fitness trainers, which not only motivates users to continue exercising, but also promotes communication and mutual assistance within the community.

We also pay special attention to the security and privacy of user data. All user data is stored encrypted, and users can control how much of their data is shared at any time. We strictly comply with data protection regulations to ensure that users' personal information is not accessed by unauthorized third parties.

4. CONCLUSION

This paper discusses people's health problems and exercise habits, and designs a posture recognition sports APP based on Android system. The APP mainly reflects body recognition by the "AI online guidance" function. During the use of the app, users can obtain preliminary action analysis through computer vision and image recognition, and then conduct in-depth analysis of user actions through deep neural network and 3D residual, so as to obtain guidance and suggestions, improve the accuracy of user actions and reduce muscle strain. At the same time, it also improves the utilization rate of time, increases the interest of sports, and enhances the subjective initiative of users.

The body recognition sports APP based on Android system developed in this study can effectively promote the adoption of healthy lifestyle. By combining modern computer vision technology, deep learning and human-computer interaction design, our "AI online Guide" function not only helps users improve the efficiency and effectiveness of movement, but also enhances the motivation and enjoyment of movement through social interaction.

In the future, we will continue to optimize algorithms, enhance user experience, and explore more personalized and scenario-based functions, such as providing immersive sports experiences with virtual reality (VR) technology, and collecting physiological data with wearable devices to provide users with more comprehensive health management solutions. We believe that as technology advances, the APP will play an increasingly important role in driving people towards a healthier life."

ACKNOWLEDGEMENTS

This paper was funded and supported by the 2024 School-level Innovation and Entrepreneurship Training Program of Wuhan Business University (No.202411654184).

REFERENCES

[1] Li Zhuoran. Posture recognition based on skeleton point marker research [D]. Shenyang university of technology, 2023. The DOI: 10.27322 /, dc nki. Gsgyu. 2023.001254.

[2] Li Jingjing, Huang Zhangjin, Zou Lu. Human motion recognition based on Convolutional networks of motion guidance graphs [J/OL]. Journal of Computer Aided Design and Graphics, 1-10 [2024-07-30]. HTTP: / / http://kns.cnki.net/kcms/detail/11.2925.TP.20240716.1755.002.html

[3] Cui Yanxin. Research on Human behavior Recognition Algorithm based on Deep Learning [D]. Hebei: Hebei Normal University,2023.

[4] An improved method of 3D residual neural network video pedestrian action classification [J]. Journal of Dalian Minzu University,2019,21(3):225-229. (in Chinese) DOI:10.3969/j.issn.1009-315X.2019.03.007.

[5] Li Xusheng. Remote sensing tree species recognition based on 3D residual Convolutional neural network algorithm [D]. Xinjiang: Xinjiang Agricultural University,2020.

[6] Wu Lijun, Li Binbin, Chen Zhicong, et al. Behavior recognition under 3D Multiple attention mechanism [J]. Journal of Fuzhou University (Natural Science Edition),2022,50(1):47-53. DOI:10.7631/issn.1000-2243.20496.

[7] Liu H, Tu J H, Liu M Y. Two-stream 3D convolutional neural network for skeleton-based action recognition[EB/OL]. [2022-08-10]. http://arxiv-org-s.webvpn.wbu.edu.cn:8118/ftp/arxiv/papers/1705/1705.08106.pdf

[8] Kim T S, Reiter A. Interpretable 3D human action analysis with temporal convolutional networks[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos: IEEE Computer Society Press, 2017: 1623-1631.

[9] Wang P C, Li Z Y, Hou Y H, et al. Action recognition based on joint trajectory maps using convolutional neural networks[C] //Proceedings of the 24th ACM International Conference on Multimedia. New York: ACM Press, 2016: 102-106

[10] Wu Jiahui, Zhou Tao, Luo Mingxin, et al. Design and development of Facial expression Recognition system based on C3D CNN [J]. Information & Computer,2022,34(14):104-107. (in Chinese) DOI:10.3969/j.issn.1003-9767.2022.14.032.

[11] Lu Hongtao, Zhang Qinchuan. A review on the application of deep convolutional neural networks in computer vision [J]. Data Acquisition and Processing, 2016, 31 (01): 1-17. DOI:10.16337/ J.1004-9037.2016.01.001.

[12] Yan S J, Xiong Y J, Lin D H. Spatial temporal graph convolutional networks for skeleton-based action recognition[C] //Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2018: Article No.912

[13] Wan Lei, Tong Xin, Sheng Mingwei, et al. Application of Softmax classifier for deep learning image classification [J]. Navigation and Control, 2019, 18 (06): 1-9+47. (in Chinese)