IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 April 2023, accepted 5 May 2023, date of publication 10 May 2023, date of current version 17 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3275012

RESEARCH ARTICLE

A Hybrid Deep Learning-Based Intelligent System for Sports Action Recognition via Visual Knowledge Discovery

LEI ZHAO
Physical Education College, Jiaozuo Normal College, Xinxiang, Henan 454002, China
e-mail: kaul2002@jzsz.edu.cn

This work was supported by the Jiaozuo Normal College.

ABSTRACT The intelligent recognition systems for sports actions have been a more general demand, so as to facilitate technical analysis of health management. This highly relies on deep analysis towards frame-level image data from the perspective of visual knowledge discovery. In recent years, the rapid development of deep learning technology has well boosted a number of technical breakthrough in computer vision. In this context, this work takes aerobics as the main object, and proposes a hybrid deep learning-based intelligent system for sports action recognition via visual knowledge discovery. Specifically, the human skeleton is represented as a graph based on the physical structure of the human body in this paper, and the selective hypergraph convolution network is selected to adaptively extract the multi-scale information in the skeleton. And the selective-frame temporal convolution is specially selected for the situation to construct recognition model. Upon the basis of proper feature extraction, a triple loss-based error measurement method is employed to construct objective function, and a recurrent neural network structure is further developed to model dynamic action sequence characteristics. The data source of this article is mainly the private data compiled by the research group. Finally, experiments are carried out on the CMU motion capture dataset, and the effectiveness of the proposed algorithm is verified by comparing the experimental results with those of the existing algorithms.

INDEX TERMS Hybrid deep learning, intelligent systems, action recognition, visual knowledge discovery.

I. INTRODUCTION

In recent years, artificial intelligence technology changes with each passing day, a variety of intelligent devices emerge in endlessly [1]. And the intelligent processing of all kinds of information in all kinds of life has shown a trend of diversification [2]. In daily life, people's communication is not limited to language, body language is also a very direct and efficient way of communication [3]. Thus, it is of great significance to recognize aerobics [4].

For the exercise practice of aerobics, its theoretical research is relatively backward [5]. The degree of integration with artificial intelligence is low [6]. And the improvement of athletes relies on traditional empirical exercises, lacking research on the essential characteristics of sports, the basic laws of sports technology development and the main factors affecting sports performance [7]. Since physical training is the basis of athletes' training, the relationship between athletes' physical fitness and sports technology is mutual promotion and mutual influence [8]. Only when athletes' physical training is done well can they ensure stronger sports skills. The physical training of track and field athletes must ensure the combination of theory and practice. Theoretical knowledge serves as a guide and leads practical training activities, so as to ensure the best effect of physical training [9].

Therefore, in the sport of aerobics, carrying out the analysis of the human body's behavioral posture and establishing an action recognition model based on convolutional neural networks is the basis of scientific training in the sport of

The associate editor coordinating the review of this manuscript and approving it for publication was Laura Celentano.

aerobics [10]. The progress of scientific research and the development of sports programs is a dynamic process that promotes each other, and the lack of theoretical scientific research will certainly become a barrier that restricts the development of sports technology [11]. Therefore, the quantification of the indicators of aerobics and the integration of new technologies into the sports program will be the future development trend [12].

From a business value point of view, it lies in the fact that if the human pose in a given image can be quickly and accurately acquired [13]. It can be used in a real-time platform to analyze the real-time human state based on the corresponding behavior obtained from pose recognition, thus playing a role of human monitoring. With the application of the recognition and analysis system based on deep learning algorithms, although the embedded dimension value level of aerobics dynamic poses detected by the system host has a certain upward trend, no matter whether the poses are simple or not, the embedded dimensions. The numerical level can be well controlled, and its average value is always lower than the detection value of the spatio-temporal weight gesture motion feature recognition method, which has a strong promoting effect on accurately capturing the dynamic posture of aerobics [14]. In terms of recognition time, as the value of the joint angle increases, the deep learning algorithm can always effectively control the time required to accurately identify the pose data of aerobics. This plays an important role in many commercial reality scenarios [15]. Main contributions of this paper can be summarized as three aspects:

- The research issue of computation intelligence-based aerobics action recognition is discussed and put forward.
- A hybrid deep learning-based intelligent system via visual sensing is proposed for this purpose.
- Some experiments are conducted to evaluate the proposed recognition method.

II. RELATED WORK

At present, the theoretical research on aerobics is mostly discussed from the aspects of aerobics teaching and the influence of aerobics on human body [16]. In addition to the basic teaching research, the academic research on aerobics focuses on the research of aerobics on morphological function, human psychological quality, female form and other physical aspects, on the other hand, it is the research on the rules and choreography creation of aerobics itself, most of these researches are qualitative analysis, which are relatively superficial, and the research on the assessment of aerobics special athletic ability and movement technology is very rare [17]. In the area of action recognition, many scholars have made a lot of theoretical discussions on action recognition of aerobics in the last decade. The research methods are divided into two types: traditional methods and deep learning methods [18]. The traditional behavior recognition methods are generally designed manually through manual observation and design, and feature extraction methods that can characterize the action features are designed manually because they require manual setting of parameters, consume a lot of human and material resources, and have a low accuracy rate.

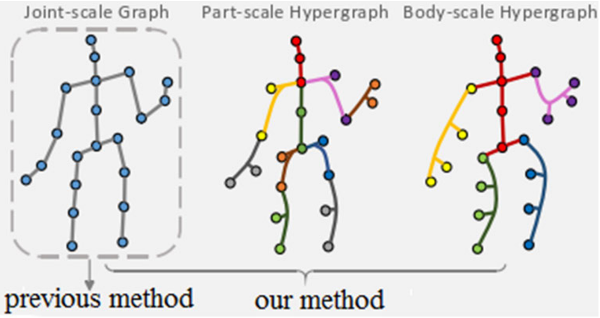


FIGURE 1. Human skeleton based on graphs and hypergraphs.

Deep learning methods can accomplish end-to-end motion recognition with high accuracy without a lot of manual labor [19]. Applying deep learning to video motion target detection can effectively describe the visual features such as appearance, structure, and color of the target to achieve target localization. In the early stage of research, artificial feature-based motion recognition methods process feature information into feature vectors, which are then input to a classifier for learning and training. Artificial features have two major drawbacks: one is that it is difficult to recognize complex sequential features, and the other is that they cannot adequately reflect the spatial and temporal characteristics of aerobics. In recent years, deep learning methods have been gradually applied to the detection and classification of images and videos, so the field of motion recognition has been developing rapidly and vigorously in recent years [20].

The traditional spatio-temporal weight gesture motion feature recognition method can determine the time-series modeling relationship between dynamic nodes according to the spatio-temporal features of human body pose video frames, and then use big data technology to realize the on-demand recognition of associated nodes. However, this method has poor accuracy in capturing the dynamic poses of aerobics, and due to the relatively large amount of calculation, it is easy to cause infinite prolongation of pose data recognition time. The essence of the deep learning network is the simulation of the human brain nervous system. During the application process, it contains multiple hidden layer structures at the same time. In recent years, deep learning techniques have shown powerful modeling capabilities in computer vision and natural language processing, and with the availability of large amounts of skeletal data, they have attracted the attention of many scholars [21]. According to deep learning techniques, they are classified into the following categories: recurrent neural networks or long short-term memory RNNs, convolutional neural networks LSTMs, and graph convolutional neural networks GCNs.

TABLE 1. Action indicators for evaluating aerobics athletes technical level.

Action code	Difficulty action name	Degree of difficulty	Weight factor
A123	Vincent Push Ups	0.2	0.043
A334	squat jump	0.3	0.032
B103	Raise the buttocks and turn the body 180 degrees	0.8	0.011
B332	Split legs support 360 degree rotation	0.2	0.012
C234	540 degree swivel body jump	0.4	0.072
C554	Rotate 180 degrees to jump, then rotate 180 degrees into a push-up	0.4	0.112
D123	Scissor transformation jump body 360 degrees	0.7	0.123
D643	Rotate 360 degrees Cossack jump and then turn 180 degrees	0.5	0.101
E212	Right-angle split leg combination support swivel 720 degrees	0.3	0.023
E767	Unsupported splits	0.8	0.033

III. AEROBICS MOVEMENT RECOGNITION MODEL BASED ON CONVOLUTIONAL NEURAL NETWORK

A. PROBLEM STATEMENT

In the aerobics competition, it can be divided into power strength, jumping, kicking and static strength according to different action strength. In this paper, we selected the most suitable aerobics movements from all the difficult movements of competitive aerobics for the construction of the movement recognition model [22]. Taking aerobics as an example, the movements of the human body can be regarded as a series of pose data that appear over time. Compared with other methods, the special kinematic feature model of the human skeleton has the ability to describe the state of posture changes. 100 test athletes performed 30 difficult movements of aerobics, and five experts in the field of aerobics scored the complexity of the movements, and for the nth movement of the mth athlete, the scores of the five experts from the highest to the lowest order were al a2 a3 a4 a5, removing one of the highest and lowest scores, then the gymnast's movement score can be expressed as: hj = (a2 + a3 + a4)/3.

The overall score Z for this aerobic gymnast can be calculated as follows:

$$Z = \sum_{1}^{11} l_i h_i \tag{1}$$

where l_i is weight factor, which can be calculated as follows:

$$l_i = e_i / \sum_{1}^{11} e_i \tag{2}$$

The final aerobics movement assessment system determined according to the expert scoring is shown in Table 1, and these 10 movements are selected as the main objects of aerobics movement identification evaluation in this paper.

B. ACTION RECOGNITION ALGORITHM BASED ON SELECTIVE HYPERGRAPH CONVOLUTIONAL NETWORK

In terms of skeletal motion recognition, previous methods treat the skeleton as a false image or sequence, and then use convolutional neural nets or recurrent neural nets to further extract motion characteristics [23]. This paper represents the human skeleton as a diagram based on the physical structure of the body to make it more natural. The method based on key feature description can better identify continuous and interactive actions with strong robustness.

The left one in Figure 1 indicates that the human skeleton is modeled as a simple graph with joint-scale information. The two diagrams on the right are the human skeleton diagrams used in this paper. This method introduces hypergraphs containing multi-scale information to compensate for the lack of representational capability of behaviors by epistemic features and motion features. The method uses convolutional neural networks to automatically extract features of the target in the image, which eliminates the instability of manually labeled features and also extracts deep features of the target, improving the accuracy of recognition. In this chapter, selective hypergraph convolutional network is chosen to extract the multiscale information in the skeleton adaptively. In addition to this this model can also selectively aggregate temporal keyframe features, thus compensating for the shortcomings of keyframe dropout, and its structure is shown in Figure 3.

In this paper, we propose to represent the human skeleton as a hypergraph, which can be done without destroying the inherent spatial properties between joints, and also maintain the higher order correlation of the human skeleton in the model built. In a selective hypergraph convolutional network, a hyperedge connects more than two joints, and the advantage of this hypergraph is that the higher-order correlations between joints can be easily captured. Using the hypergraph convolution operator, graph neural networks can be easily extended to other models and applied to handle various non-pairwise relations [24].

The hypergraph convolutional network architecture diagram is divided into the following parts. There are eight selective spatio-temporal hypergraph convolutional blocks and a fully connected classifier with global average pooling. Every convolutional layer is connected with a batch normalization layer and a ReLU activation layer. Because multi-scale information is an essential aspect in the process of skeleton recognition, and most spatio-temporal graph convolutional networks perform poorly in this aspect, focusing only on single-scale information. Therefore, this paper uses

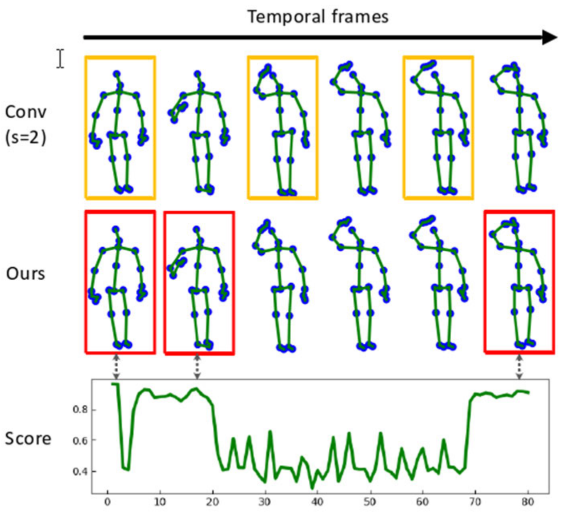


FIGURE 2. Graph-based and hypergraph-based human skeletons.

the proposed Supplemental Health Care (SHC) network to capture multi-scale contexts as well as to aggregate information from multiple scales through adaptive sensory fields. As shown in Figure 4, this automatic selection mechanism of the scale-selective hypergraph convolution network consists of three stages.

The process is as follows, first input $\mathrm{Y} \in \mathrm{S}^{\mathrm{C}\times\mathrm{T}\times\mathrm{V}}$,, input the value to the three branches, the calculation is as:

$$\begin{cases} Y_{joint} = \sigma\left(\sum_{h=1}^{H} Y_h X \Lambda_h\right) \\ Y_{part} = \sigma\left(\sum_{h=1}^{H} Y_{part} X H_{part}\right) \\ Y_{body} = \sigma\left(\sum_{h=1}^{H} Y_{part} X H_{body}\right) \end{cases} \tag{3}$$

where σ represents the error parameter, X represents the input matrix, and H represents the height matrix. In the case of this paper, after performing the hypergraph convolution operation, the number of nodes in the hypergraph does not produce a numerical change, so that the hypergraph convolution features obtained at different scales can be aggregated smoothly [25]. In this paper, when aggregating the three branch features, the element level summation is taken, which is calculated as:

$$\mathrm{V} = \mathrm{Y}_{\text{joint}} + \mathrm{Y}_{\text{part}} + \mathrm{Y}_{\text{body}} \tag{4}$$

In this paper, the global contextual information is collected by globally averaging the pooling in both spatial and temporal dimensions in a sequential manner, so that all information in the feature graph is averaged without losing too much critical information. A fully connected layer with nonlinearity is then used to make the selective weights more adaptive and to reduce the feature dimensionality. After multi-layer convolutional pooling of the combined features, valid features are selected by a fully connected layer to construct a mapping relationship with the output.

$$A = \delta(C(GD(\text{Avg Pool}\,(V)))) \tag{5}$$

Then, this paper uses soft attention on the three channel dimensions to adaptively select information at different scales. The features of each channel are reallocated to complete the feature rescaling on the channel dimension, giving the model the ability to better identify the features of each channel. The soft attention in the three branches does not share weights and is implemented through a fully connected layer with softmax normalization, which indicates the selective weights assigned to the feature maps at different scales. The specific computational expression is given as:

$$\begin{cases} \mathrm{b}_{\text{joint}} = \mathrm{Softmax}(GD(a)) \\ \mathrm{b}_{\text{part}} = \mathrm{Softmax}(GD(a)) \\ \mathrm{b}_{\text{body}} = \mathrm{Softmax}(GD(a)) \end{cases} \tag{6}$$

Finally, the selective feature maps are obtained by calculating selective weights at multiple scales, as expressed as:

$$\mathrm{Y}_{\text{select}} = \mathrm{b}_{\text{joint}} \cdot \mathrm{Y}_{\text{joint}} + \mathrm{b}_{\text{part}} \cdot \mathrm{Y}_{\text{part}} + \mathrm{b}_{\text{body}} \cdot \mathrm{Y}_{\text{body}} \tag{7}$$

In addition, the Nonlocal module is integrated in this paper to obtain long range information. The feature map $Y_{\text{in}} \in S^{C\times T\times V}$ is given first, and then the scale selection context information is obtained, and then the Non-local module is used to obtain the long range confidence. By this method, the model can obtain more semantic information, which leads to better performance of the model. The algorithm incorporates multiple features to ensure the accuracy of the algorithm and improves the efficiency of the algorithm by using multi-scale pictures and filtering process [26]. In addition, the original input is first fed into two embedding functions (e.g., λ and ξ) to obtain the encoded features, and then the element-level product is used to obtain the attention matrix, which is calculated as:

$$N_{\text{att}} = \mathrm{Softmax}\left(Y_{\text{in}}^{U} X_{\lambda}^{U} W_{\xi} Y_{\text{in}}\right) \tag{8}$$

Higher-order features are obtained using an aggregation function, calculated as:

$$Y_{\text{nonlocal}} = \mathrm{X}\mathrm{Y}_{\text{in}} N_{\text{att}} \tag{9}$$

This paper proposes frame selection time convolution, which is based on a selective convolution mechanism of key frames to replace stepwise convolution and adaptively select more important frames. A comparison of traditional step-time convolution and frame-selective time-convolution is shown in Figure 5.In this paper, the frame selection time convolution consists of three branches: the frame importance calculation branch, the frame feature aggregation branch, and the residual concatenation branch. To be consistent with previous methods, using only the frame feature aggregation branch and the residual connection branch is the original time convolution, which can selectively pool the time series frames after adding the frame importance calculation branch. Combining the advantages of time series models

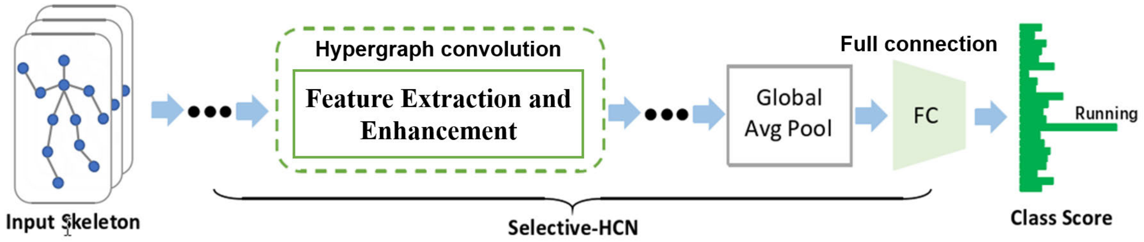


FIGURE 3. Selective Hypergraph Convolutional Network Architecture.

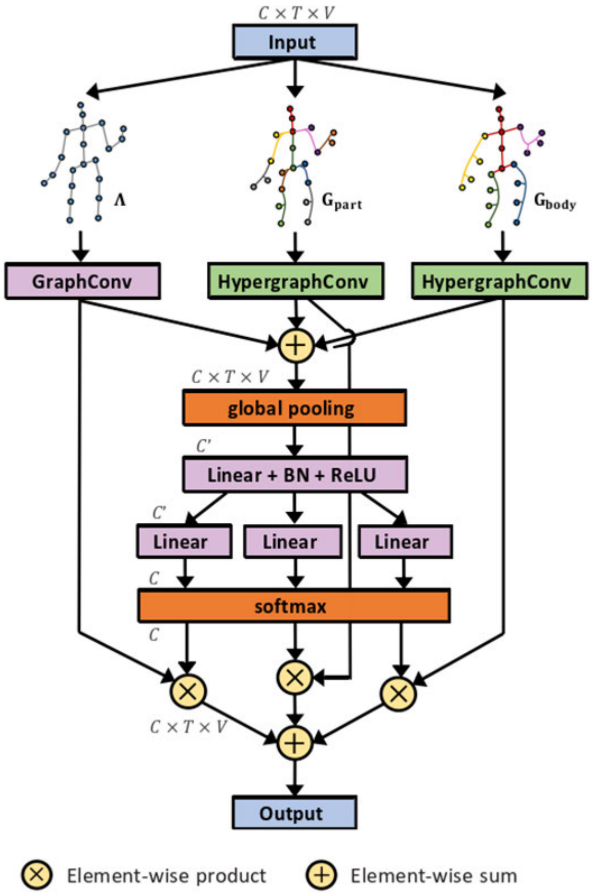


FIGURE 4. Scale-selected hypergraph convolution module.

to construct a combined model can improve the prediction accuracy.

Next, the entire selection mechanism is described in Algorithm 1. It is worth noting that the original input features are also selectively pooled to obtain the final residual branching results. The whole process is computed as:

$$\begin{cases} \lambda = \mathrm{Sigmoid}\left(\mathrm{NLQ_u}\left(\mathrm{Conv}_t\left(\mathrm{Conv}_d\left(Y_{\mathrm{in}}\right)\right)\right)\right) \\ \mathrm{idx}\ , \mathrm{imp} = \mathrm{Top}-\mathrm{l}(\lambda) \\ \mathrm{Y_{tcn}} = \mathrm{imp} \cdot \mathrm{Y_{tcn}}\left(:, i\,dx, :\right) \\ \mathrm{Y_{res}} = \mathrm{Y_{res}}\left(:, i\,dx, :\right) \end{cases} \tag{10}$$

Algorithm 1 Selective-Frame Temporal Convolution

```
Input: Feature$Y_{in}$ Step size T
Parameters:
TCN; Residual; MLP;$Conv_d$ $Conv_t$
Output: Feature $Y_{out}$
$Y_{tcn}$=TCN($Y_{in}$)
$Y_{res}$=Residual($Y_{in}$)
if t==3 then
    $\lambda = Sigmoid(NLQ_u(Conv_t(Conv_d(Y_{in}))))$
    $idx, imp = Top - l(\lambda)$
    $Y_{tcn} = imp \cdot Y_{tcn}\ (:, idx, :))$
    $Y_{res} = Y_{res}\ (:, idx, :)$
end
$Y_{out}$=$Y_{tcn}$+$Y_{res}$
Go back to $Y_{out}$
```

IV. AEROBICS MOVEMENT EVALUATION MODEL BASED ON DEPTH METRIC LEARNING

Aerobics often change in different sequences, which means that a certain pose will appear on the sequence representing the same movement at different moments [27]. When comparing body postures, it is an interesting question how to evaluate the similarity of two postures and action sequences. As shown in Figure 6, the first two lines of action are both walking, but both the L2 and DTW metrics consider the unrelated "standing" sequence (bottom) to be more similar than the semantically related "walking" sequence (top).

A. LOSS FUNCTION

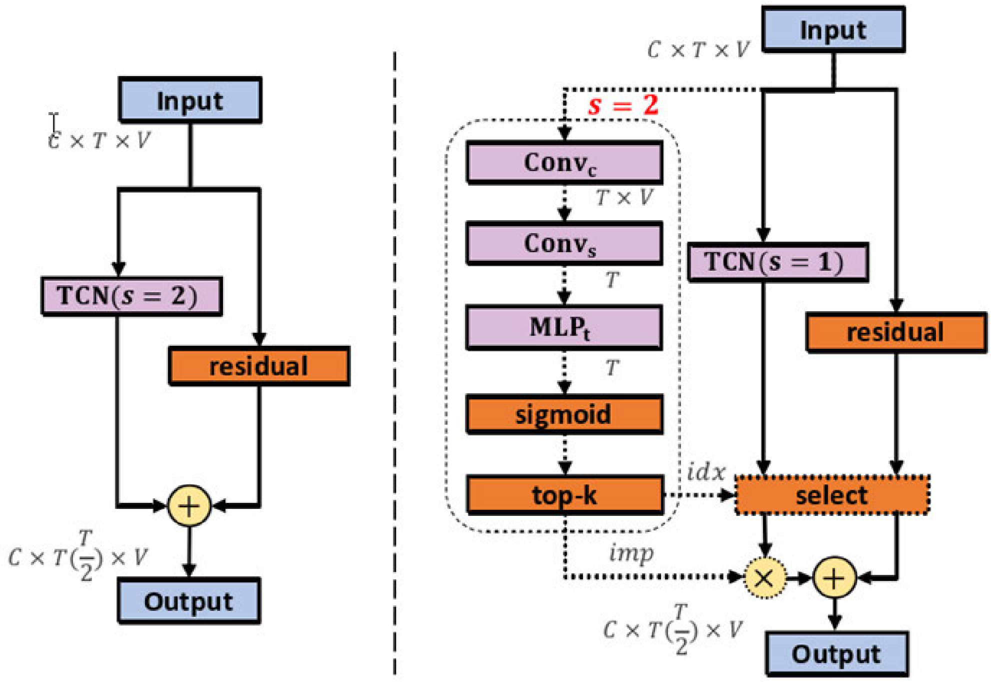
A loss function is used in the similarity test with the goal of using the network to reduce the distance between the anchor and the positive sample distribution when increasing to negative values [28]. The conventional cubic loss function is obtained from the same type of sampling and is given as the boundary distance. Denoting the aerobics sequence by Y^1, Y and Y^{-1}, respectively, the loss function is shown as:

$$M_{tri} = \max\left(0, \Omega - \Delta + b_{m\arg in}\right) \tag{11}$$

where

$$\Omega = \left\| f\left(Y\right) - f\left(Y^1\right) \right\|^2 \tag{12}$$

$$\Delta = \left\| f\left(Y\right) - f\left(Y^{-1}\right) \right\|^2 \tag{13}$$



(a) Traditional Strided Convolution (b) Selective-frame Temporal Convolution

FIGURE 5. Comparison of (a) traditional strided convolution and (b) selective-frame temporal convolution.

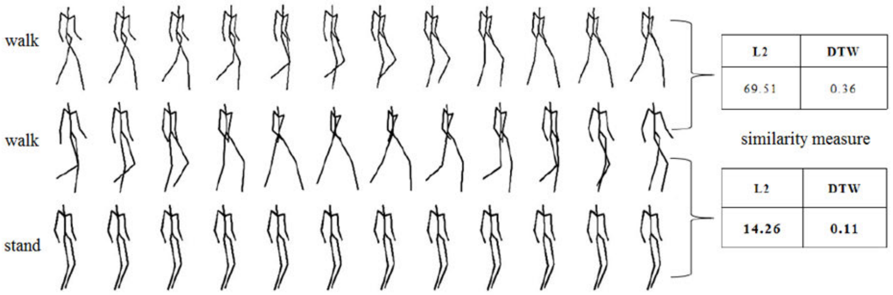


FIGURE 6. Traditional Similarity Measurement Methods.

and b_{margin} is not easily measurable and this problem can be solved by NCA (Neighbourhood Components Analysis) with the loss function shown as:

$$M_N = \frac{\exp\left(-\left\|f(Y) - f\left(Y^1\right)\right\|^2\right)}{\sum_{Y-\epsilon D} \exp\left(-\left\|f(Y) - f\left(Y^{-1}\right)\right\|^2\right)} \tag{14}$$

where D denotes all categories except positive samples. Ideally, when iterating over three sets of samples, it is expected that samples from the same category are grouped in the same cluster in the corresponding embedding space. However, finding all possible triples is very time-consuming and impractical, and the model should be trained on only a small number of meaningful triples.

Metric learning using triplet networks became more popular due to Google's FaceNet, which uses a triplet loss to learn the embedding space of images of faces so that embeddings of similar faces are closer together and embeddings of different faces are closer together. Far. For face recognition, positive images are images from the same person in the anchor image, while negative images are images of people randomly selected from the mini-batch. However, our case does not have a classification that allows easy selection of positive and negative instances. Although it would be costly to find hard-to-score samples in the dataset, it is possible to select them as negative samples. As the parameters are continuously updated, the current positive and negative samples will get further apart and will also get closer to other types that can move the full cluster, which can be shown as:

$$\begin{aligned} MMD[h, q, r]^2 = F_{y,y'}\left[h\left(y, y'\right)\right] - 2F_{y,z}[h(y, z)] \\ + F_{z,z'}\left[h\left(z, z'\right)\right] \end{aligned} \tag{15}$$

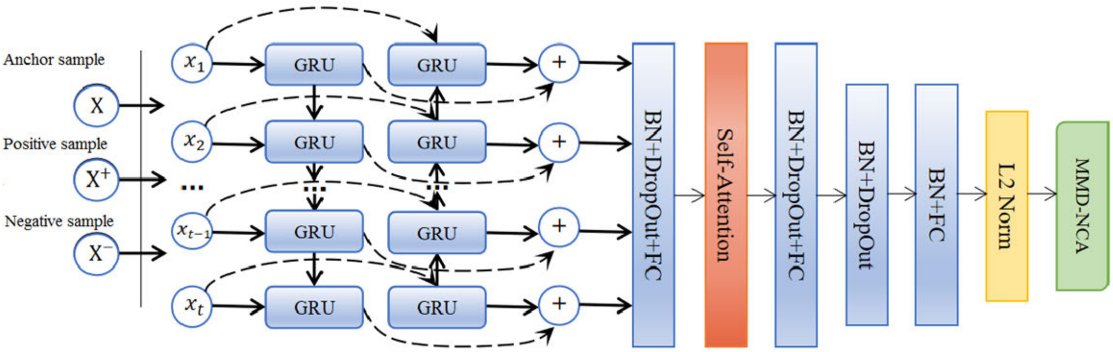


FIGURE 7. Recurrent Neural Network with Attention Mechanism.

where y and y' are independently and identically distributed sequences drawn from q, z and z' are independently and identically distributed sequences drawn from r, and h denotes the kernel function as:

$$h\left(y, y'\right) = \sum_{r=1}^{H} k_{\sigma_r}\left(y, y'\right) \tag{16}$$

Varying the expected value of a given sample gives as:

$$MMD[k, Y, Z]^2 = \frac{1}{n^2}\sum_{j=1}^{n} h\left(y_i, y_j'\right) - \frac{2}{no}\sum_{j=1}^{o} h\left(y_i, z_j\right) + \frac{1}{o^2}\sum_{j=1}^{o} h\left(z_i, z_j'\right) \tag{17}$$

where $Y = \{y_1, y_2, \cdots, y_n\}$ is the sample set drawn from the q-distribution and $Z = \{z_1, z_2, \cdots, z_n\}$ is the sample set drawn from the r-distribution, and the distance between the two sets of distributions can be measured as:

$$M_{M-N} = \frac{\exp\left(-MMD\left[h, f(Y), f\left(Y^1\right)\right]\right)}{\sum_{j=1}^{M} \exp\left(-MMD\left[h, f(Y), f\left(Y_{dj}^{-}\right)\right]\right)} \tag{18}$$

B. TWO-LAYER RECURRENT NEURAL NETWORK STRUCTURE

Based on the triple loss function of Formula 15, a convolutional recurrent neural network structure incorporating a self-attentive mechanism is designed and implemented in this section, and the structure is shown in Figure 7. The data captured by aerobics is a time series, which has some correlation with the current data in time and space [29]. Bidirectional GateRecurrent Unit (GRU) structure learns to model the implicit relationship of time dimension, SA_MMD_NCA consists of two layers of bidirectional GRU units, every layer has t GRUs corresponding to the dimension of the input data, the structure of GRU units is shown in Figure 8.

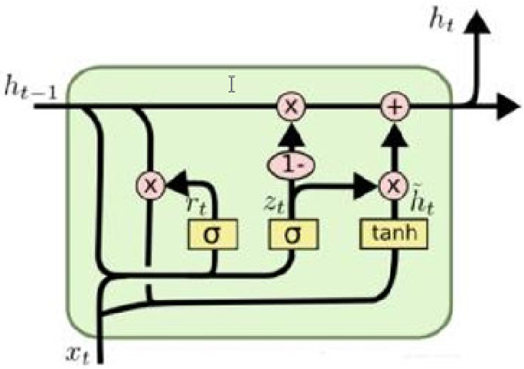


FIGURE 8. Sketch map for GRU model.

Suppose that given a sequence of motions of length t, then for an input of x_t, each GRU cell contains an update gate z, and a reset gate r_t, and each GRU state update as:

$$s_t = \lambda\left(X_s \cdot [\mathrm{i}_{t-1}, y_t]\right) \tag{19}$$

$$a_t = \lambda\left(X_z \cdot [i_{t-1}, y_t]\right) \tag{20}$$

$$\tilde{\mathrm{i}}_t = \tanh\left(X_{\tilde{\mathrm{i}}} \cdot [s_t \times h_{t-1}, y_t]\right) \tag{21}$$

$$\mathrm{i}_t = (1 - a_t) \times \mathrm{i}_{t-1} + a_t \times \tilde{\mathrm{i}}_t \tag{22}$$

$$z_t = \lambda\left(X_0 \cdot \mathrm{i}_t\right) \tag{23}$$

The weight matrix is represented by the corresponding weight matrices X_s, X_a, X_i X_p, which means:

$$\begin{aligned} X_s &= X_{sy} + X_{si} \\ X_a &= X_{ay} + X_{ai} \\ X_{\tilde{i}} &= X_{\tilde{i}y} + X_{\tilde{i}i} \end{aligned} \tag{24}$$

The SA_MMD_NCA pseudo-code is shown in Algorithm 2.

C. CMU DATASET EXPERIMENTAL RESULTS

To prove the effectiveness of the algorithm, experiments are conducted on the CMU motion capture dataset in this paper. Eleven classes are selected as the training set and 10 classes are selected as the test set in the CMU motion capture data, respectively. Since training with different sizes slows down the training process, this experiment uses a fixed sequence

Algorithm 2 SA_MMD_NCA Pseudo-Code

```
Input:
Y: training set input action data
Z: training set action labels
MMD-NCA-Group Num :batch size
Category batch size :The number of samples of each type in a single batch
Negative batch size: The number of negative samples in a single batch
$epoc$:number of training iterations
$lr$:learning rate
Output:
$l2\_norm$:sample distance
$Y$:Prediction sample label
1. set the parameter of optimizer
2. for e = 1 to epoch do
3. Select anchor samples and positive and negative samples under every batch
4. Extract sequence time series features T
5. Calculate the score B at every moment and the encoded GRU output is F
6. Compute the variance loss function that measures the samples
7. Clear the previous gradient
8. Back propagation, calculate the current gradient
9. Update network parameters
11. end for
```

TABLE 2. Action indicators for evaluating aerobics athletes' technical level.

Model	TPR-90	TPR-80	TPR-70
DTW	44.31	42.63	36.65
MDDTW	42.52	38.91	34.52
CTW	46.03	39.21	39.11
GDTW	45.56	38.63	35.32
DCTW	40.53	38.32	27.92
Triplet	39.34	34.45	27.22
SA_MMD_NCA	32.82	26.07	21.02

length for training, dividing the action sequence into 120 consecutive frames and leaving 20 frame gaps. This experiment uses the false positive rate with different percentage recall for the evaluation of model performance. The results on the CMU dataset compared with different models are shown in Table 2.

As seen in Table 2 and Figure 9, SA_MMD_NCA obtained lower FPR rates at different percentages of TPR. With a TPR rate of 80%, the method in this chapter has nearly 19% improvement in FPR compared to the first four methods and about 5% improvement compared to the three deep learning models. The results show that the modeling accuracy of this paper's method is more and the performance is better, which illustrates the effectiveness and superiority of this paper's method by comparing with the selected methods.

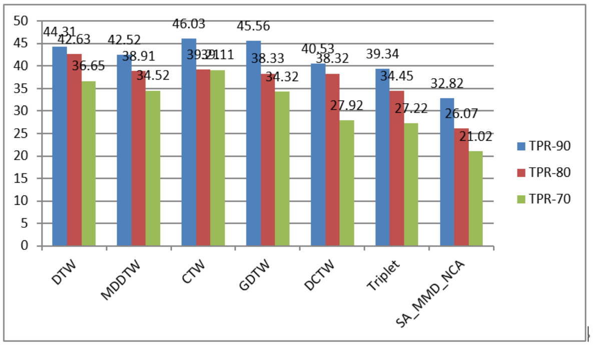


FIGURE 9. Comparing algorithm accuracy comparison.

V. CONCLUSION

In this paper, the calisthenics action evaluation system is constructed, and the calisthenics action recognition model based on convolutional neural network is established. Selective-frame temporal convolution frame selection time convolution is proposed. Frame selection time convolution consists of three branches, including frame importance calculation branch, frame feature aggregation branch and residual connection branch. In addition, this paper establishes an aerobics action evaluation model based on deep metric learning. A measurement method based on triple loss and maximum average difference is proposed, and the action sequence measurement model based on MMD-NCA recurrent neural network architecture is described in detail. Finally, experiments are carried out on the CMU motion capture dataset, and the effectiveness of the proposed algorithm is verified by comparing the experimental results with those of the existing algorithms.

The aerobics dynamic posture recognition and analysis system designed in this paper can control the adjustment ability of RBM coefficients by integrating dynamic posture data, and then define a specialized human skeleton model according to the principle of feedback fine-tuning. Due to the existence of the behavior recognition data set structure, the signal feature extraction results can directly affect the pose data capture ability matched with the system host.

REFERENCES

[1] Q. Liu, "Aerobics posture recognition based on neural network and sensors," *Neural Comput. Appl.*, vol. 34, no. 5, pp. 3337–3348, Mar. 2022, doi: 10.1007/s00521-020-05632-w.

[2] L. Jia and L. Li, "Retraction note: Research on core strength training of aerobics based on artificial intelligence and sensor network," *EURASIP J. Wireless Commun. Netw.*, vol. 2022, no. 1, p. 123, Dec. 2022, doi: 10.1186/s13638-022-02202-7.

[3] S. Yue, "Image recognition of competitive aerobics movements based on embedded system and digital image processing," *Microprocessors Microsystems*, vol. 82, Apr. 2021, Art. no. 103925, doi: 10.1016/j.micpro.2021.103925.

[4] G. Yan and M. Woźniak, "Accurate key frame extraction algorithm of video action for aerobics online teaching," *Mobile Netw. Appl.*, vol. 27, no. 3, pp. 1252–1261, Jun. 2022, doi: 10.1007/s11036-022-01939-1.

[5] F. Huang, "Design of diversified teaching platform of college aerobics course based on artificial intelligence," *J. Comput. Methods Sci. Eng.*, vol. 22, no. 2, pp. 385–397, Mar. 2022, doi: 10.3233/JCM-215668.

[6] Z. Guo, K. Yu, N. Kumar, W. Wei, S. Mumtaz, and M. Guizani, ‘‘Deep-distributed-learning-based POI recommendation under mobile-edge networks,’’ *IEEE Internet Things J.*, vol. 10, no. 1, pp. 303–317, Jan. 2023.

[7] J. Zhou, J. Zhang, A. C. McLain, W. Lu, X. Sui, and J. W. Hardin, ‘‘Semi-parametric regression of the illness-death model with interval censored disease incidence time: An application to the ACLS data,’’ *Stat. Methods Med. Res.*, vol. 29, no. 12, pp. 3707–3720, Dec. 2020.

[8] Q. Li, L. Liu, Z. Guo, P. Vijayakumar, F. Taghizadeh-Hesary, and K. Yu, ‘‘Smart assessment and forecasting framework for healthy development index in urban cities,’’ *Cities*, vol. 131, Dec. 2022, Art. no. 103971.

[9] Z. Guo, Y. Shen, S. Wan, W.-L. Shang, and K. Yu, ‘‘Hybrid intelligence-driven medical image recognition for remote patient diagnosis in Internet of Medical Things,’’ *IEEE J. Biomed. Health Informat.*, vol. 26, no. 12, pp. 5817–5828, Dec. 2022.

[10] Z. Guo, K. Yu, A. K. Bashir, D. Zhang, Y. D. Al-Otaibi, and M. Guizani, ‘‘Deep information fusion-driven POI scheduling for mobile social networks,’’ *IEEE Netw.*, vol. 36, no. 4, pp. 210–216, Jul. 2022.

[11] Z. Guo, K. Yu, Z. Lv, K.-K.-R. Choo, P. Shi, and J. J. P. C. Rodrigues, ‘‘Deep federated learning enhanced secure POI microservices for cyber-physical systems,’’ *IEEE Wireless Commun.*, vol. 29, no. 2, pp. 22–29, Apr. 2022.

[12] B. Markham, E. Brush, and R. Connick, ‘‘Fitness centers in mixed-use development: Examples from practice,’’ *J. Acoust. Soc. Amer.*, vol. 144, no. 3, pp. 1787–1788, Sep. 2018.

[13] Q. Zhang, Z. Guo, Y. Zhu, P. Vijayakumar, A. Castiglione, and B. B. Gupta, ‘‘A deep learning-based fast fake news detection model for cyber-physical social services,’’ *Pattern Recognit. Lett.*, vol. 168, pp. 31–38, Apr. 2023.

[14] Q. Zhang, K. Yu, Z. Guo, S. Garg, J. J. P. C. Rodrigues, M. M. Hassan, and M. Guizani, ‘‘Graph neural network-driven traffic forecasting for the connected Internet of Vehicles,’’ *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 5, pp. 3015–3027, Sep. 2022.

[15] M. H. Davenport, A. P. McCurdy, M. F. Mottola, R. J. Skow, V. L. Meah, V. J. Poitras, A. Jaramillo Garcia, C. E. Gray, N. Barrowman, L. Riske, F. Sobierajski, M. James, T. Nagpal, A.-A. Marchand, M. Nuspl, L. G. Slater, R. Barakat, K. B. Adamo, G. A. Davies, and S.-M. Ruchat, ‘‘Impact of prenatal exercise on both prenatal and postnatal anxiety and depressive symptoms: A systematic review and meta-analysis,’’ *Brit. J. Sports Med.*, vol. 52, no. 21, pp. 1376–1385, Nov. 2018.

[16] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, ‘‘Subject independent facial expression recognition with robust face detection using a convolutional neural network,’’ *Neural Netw.*, vol. 16, nos. 5–6, pp. 555–559, Jun. 2003.

[17] U. R. Acharya, H. Fujita, O. S. Lih, M. Adam, J. H. Tan, and C. K. Chua, ‘‘Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network,’’ *Knowl.-Based Syst.*, vol. 132, pp. 62–71, Sep. 2017.

[18] H. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. B. H. Hassen, L. Thomas, A. Enk, and L. Uhlmann, ‘‘Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists,’’ *Ann. Oncol.*, vol. 29, no. 8, pp. 1836–1842, 2018.

[19] F.-C. Chen and M. R. Jahanshahi, ‘‘NB-CNN: Deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion,’’ *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4392–4400, May 2018.

[20] R. Cang, H. Li, H. Yao, Y. Jiao, and Y. Ren, ‘‘Improving direct physical properties prediction of heterogeneous materials from imaging data via convolutional neural network and a morphology-aware generative model,’’ *Comput. Mater. Sci.*, vol. 150, pp. 212–221, Jul. 2018.

[21] A. Rikhtegar, M. Pooyan, and M. T. Manzuri-Shalmani, ‘‘Genetic algorithm-optimised structure of convolutional neural network for face recognition applications,’’ *IET Comput. Vis.*, vol. 10, no. 6, pp. 559–566, Sep. 2016.

[22] X. Jun, J. Wang, J. Zhou, S. Meng, R. Pan, and W. Gao, ‘‘Fabric defect detection based on a deep convolutional neural network using a two-stage strategy,’’ *Textile Res. J.*, vol. 91, nos. 1–2, pp. 130–142, Jan. 2021.

[23] M. Zhu, S. Bin, and G. Sun, ‘‘Lite-3DCNN combined with attention mechanism for complex human movement recognition,’’ *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–9, Sep. 2022.

[24] G. Liu, Z. Yin, Y. Jia, and Y. Xie, ‘‘Passenger flow estimation based on convolutional neural network in public transportation system,’’ *Knowl.-Based Syst.*, vol. 123, pp. 102–115, May 2017.

[25] H. Y. Khaw, F. C. Soon, J. H. Chuah, and C. Chow, ‘‘Image noise types recognition using convolutional neural network with principal components analysis,’’ *IET Image Process.*, vol. 11, no. 12, pp. 1238–1245, Dec. 2017.

[26] S. Bin and G. Sun, ‘‘Matrix factorization recommendation algorithm based on multiple social relationships,’’ *Math. Problems Eng.*, vol. 2021, pp. 1–8, Feb. 2021.

[27] S. Wen, J. Chen, Y. Wu, Z. Yan, Y. Cao, Y. Yang, and T. Huang, ‘‘CKFO: Convolution kernel first operated algorithm with applications in memristor-based convolutional neural network,’’ *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 40, no. 8, pp. 1640–1647, Aug. 2021.

[28] M. Ibrahim, J. Sagers, and M. Ballard, ‘‘A convolutional neural network applied to Arctic acoustic recordings to identify soundscape components,’’ in *Proc. Meetings Acoust. ASA*, 2020, vol. 42, no. 1, Art. no. 070005.

[29] C. Liu, C. Tang, M. Xu, and Z. Lei, ‘‘Binarization of ESPI fringe patterns based on an M-net convolutional neural network,’’ *Appl. Opt.*, vol. 59, no. 30, pp. 9598–9606, 2020.

LEI ZHAO was born in Shandong, China, in 1982. She received the bachelor’s degree from Zhengzhou University, Henan, in 2004. Since 2004, she has been with the Physical Education College, Jiaozuo Normal College. She published three articles and one book. Her research interests include big data and imaging processing.