IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

**Traitement du Signal**
Vol. 38, No. 6, December, 2021, pp. 1801-1807
Journal homepage: http://iieta.org/journals/ts

# Image Recognition of Standard Actions in Sports Videos Based on Feature Fusion

Songjiao Wu

Chongqing University of Arts and Sciences, Chongqing 402160, China

Corresponding Author Email: 20060076@cqwu.edu.cn

https://doi.org/10.18280/ts.380624

**Received:** 15 July 2021
**Accepted:** 3 November 2021

***Keywords:***
*sports, action recognition, local feature extraction, time-space feature fusion*

**ABSTRACT**

Standard actions are crucial to sports training of athletes and daily exercise of ordinary people. There are two key issues in sports action recognition: the extraction of sports action features, and the classification of sports actions. The existing action recognition algorithms cannot work effectively on sports competitions, which feature high complexity, fine class granularity, and fast action speed. To solve the problem, this paper develops an image recognition method of standard actions in sports videos, which merges local and global features. Firstly, the authors combed through the functions and performance required for the recognition of standard actions of sports, and proposed an attention-based local feature extraction algorithm for the frames of sports match videos. Next, a sampling algorithm was developed based on time-space compression, and a standard sports action recognition algorithm was designed based on time-space feature fusion, with the aim to fuse the time-space features of the standard actions in sports match videos, and to overcome the underfitting problem of direct fusion of time-space features extracted by the attention mechanism. The workflow of these algorithms was explained in details. Experimental results confirm the effectiveness of our approach.

## 1. INTRODUCTION

On August 3, 2021, the State Council of China issued the *National Fitness Plan*, which lists the specific goals and main tasks of the national fitness campaign in 2021-2025 [1-4]. Under this incentive policy, the Chinese people attach greater importance to sports, and the public service level of national fitness improve significantly, and sports become an indispensable part of the daily life [5-10]. Sports actions are the fundamentals of sports. Standard actions are crucial to sports training of athletes and daily exercise of ordinary people. The recognition and analysis of whether sports actions are up to standard help athletes carry out scientific training and improve their competitiveness [11-19]. Therefore, it is very significant to study the recognition of sports actions.

The broadcast sports videos often face problems like poor image quality, unstable camera, and low resolution of the player. To address these problems, Liu et al. [20] carried out a motion analysis of the spatial distribution features of the optical flow field for the tracked persons, extracted the descriptive features of motions by grid classification, a local analysis method, and presented a recognition method based on motion descriptive features. Their method achieved better recognition effects than the existing approaches, which are based on performance features. To establish an action-based index of sports videos and to support coaches and athletes with kinematic movements, Li et al. [21] proposed an automatic detection and analysis system for athletes' complex actions in sports video series with a moving background, and designed a robust algorithm for the system. Their system can filter out outliers adaptively, make global estimation of motions, and realize target segmentation and automatic body tracking based on adaptive background constructs. The action recognition of dance videos has developed slowly, owing to the high complexity of dance actions, and the body occlusions in dance performance. Ni and Yao [22] introduced the research of sports dance action recognition systems facing body motion monitoring and perception, and deeply analyzed the excellent action recognition techniques. To improve the recognition accuracy of composite actions of sports training, Jiang and Tsai [23] put forward a hybrid recognition model for composite actions of sports training, based on sequential minimal optimization (SMO) algorithm and artificial intelligence (AI). By expanding the standard action data, they established a standard results comparison database, and designed the system structure and key acquisition modules based on three-dimensional (3D) data. During physical education (PE), the prevention and correction of wrong actions are not only the liability of the teacher, but also the key for the students to master the basic sports skills. Based on 3D modeling and detection, Zhai [24] developed a sports action recognition method for PE. Compared with the traditional recognition methods, their method can accurately and efficiently recognize wrong actions in time through the reconstruction and analysis of visual images, and satisfy the demand for detecting and correcting wrong sports actions.

There are two key issues in sports action recognition: the extraction of sports action features, and the classification of sports actions. The existing studies on human action recognition mostly tackle the preprocessing and optimization of video images, and the construction of action classification algorithms and models, trying to improve the accuracy and real-timeliness of action recognition. However, the resulting algorithms and models are not suitable for sports competitions. To solve the problem, this paper develops an image recognition method of standard actions in sports videos, which

merges local and global features. The main contents are as follows: Section 2 combs through the functions and performance required for the recognition of standard actions of sports, and introduces the research ideas in details: introducing the attention mechanism to the local image feature extraction algorithm of sports match videos, and specifying the flow of the algorithm. Section 3 designs a sampling algorithm based on time-space compression, and develops a standard sports action recognition algorithm based on time-space feature fusion. The two algorithms can fuse the time-space features of the standard actions in sports match videos, and overcome the underfitting problem of direct fusion of time-space features extracted by the attention mechanism. Finally, the effectiveness of our approach was demonstrated through experiments.

## 2. IMAGE FEATURE FUSION OF SPORTS MATCH VIDEOS

The existing human action recognition algorithms mainly target images containing relatively simple and slow actions like hand clapping, standing, walking, and jumping. The images can be recognized accurately, due to the large between class dispersion of the samples. However, the existing algorithms perform poorly on sports competitions, which feature high complexity, fine class granularity, and fast action speed. This calls for an action recognition algorithm capable of recognizing complex actions.

To complete a standard sports action, the athlete must complete continuous steps through the cooperation between hands, feet, and torso. The action recognition algorithm must have a strong recognition ability, in order to judge whether the sports action of the athlete is up to standard. For the same sport, the evaluation of a single technical action needs to recognize dozens of sub-actions. Despite belonging to the same sports, these sub-actions differ in posture, and fall into different classes, according to the rules of the sports. As a result, the recognition algorithm of standard sport actions must have a high granularity.

Many sports are implemented at a high speed. The athlete is required to complete the technical actions quickly. This poses a new challenge to the recognition of standard sports actions. For the above reasons, this paper introduces the attention mechanism to the local image feature extraction of sports match videos, and proposes a recognition algorithm for standard sports actions that fuses space-time features.

The technical actions of rhythmic gymnastics were selected for our theoretical and experimental research, including jumping, rotation, footwork, and swivel. In this sport. Among them, jumping and rotation require the highest level of skills. There are two difficulties concerning the recognition of these two actions: (1) extract fine spatial features of technical actions based on local body parts; (2) perceive the correlations between key start and end frames, and between adjacent frames of technical actions in the time domain.

Figure 1 shows the flow of feature extraction and fusion of sports action video frames. Based on the attention mechanism, the proposed algorithm selects the most relevant features out of various local features of technical actions of rhythmic gymnastics. The attention mechanism can learn which technical action features are related to the preset class, and rank them by correlation. The output of the mechanism is the weighted sum $B$ of the convolutional features extracted by the network. Let $g_i \in R^u$ be the local eigenvector of the i-th technical action of rhythmic gymnastics; $u$ be the size of the eigenvector; $\beta(g_i;\Psi)$ be the attention score function obtained by training the eigenvector; $\omega$ be the parameter of $\beta$; $Q \in R^{N\times u}$ be the weight of the fully-connected layer of the pretrained convolutional neural network (CNN). In our network, the cross entropy is taken as the loss function. Then, the output of our network can be calculated by:

$$b = Q\left(\sum_m \left(1+\beta\left(g_i;\omega\right)\right)\cdot g_i\right) \tag{1}$$

Let $b^*$ be the ground-truth of the one-hot code; $g_i^T$ be the transposition of $g_i$. Then, the cross-entropy loss function can be expressed as:

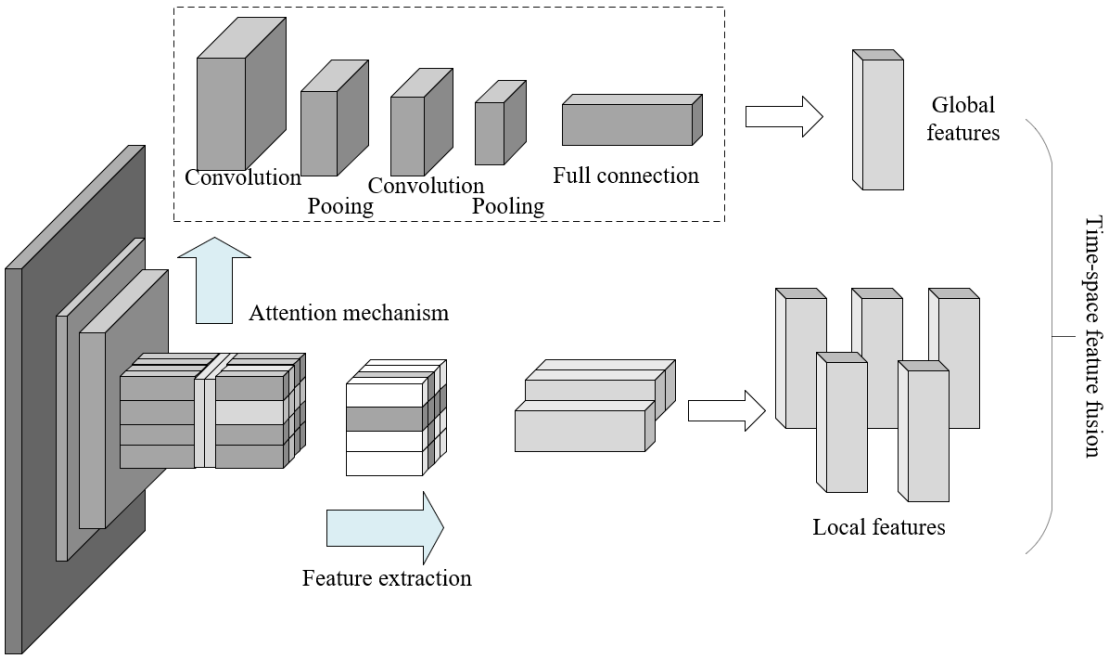


**Figure 1.** Flow of feature extraction and fusion of video frames

$$LOSS = -b^* \cdot log\left(\frac{e^b}{g_i^T \cdot e^b}\right) \tag{2}$$

The parameters in the attention score function $\beta$ are trained based on backpropagation. Then, the gradient can be calculated by:

$$\frac{\partial LOSS}{\partial \omega} = \frac{\partial LOSS}{\partial b}\sum_m qg_i \frac{\partial \beta_i}{\partial \omega} \tag{3}$$

Formula (3) shows that the backpropagation of the output score $\beta_i \equiv \beta(g_i;\Psi)$ is the same as that of standard artificial neural network (ANN).

The local features after each match video frame are selected as the key feature points: $A=\{a_1,a_2,a_3,...,a_n\}$. Through k-means clustering (KMC), the high-score features are clustered to form $w$ cluster centers $V=\{v_1,v_2,v3,...,v_w\}$, with $v_j \in R_u$. The clustering operation aims to make high-score features to participate in the residual calculation of all cluster centers. Then, the triangulation is embedded. Let $v_j$ be the center of the j-th cluster; $s_j(a)$ be the residual vector obtained through embedding. Then, the embedding operation can be calculated by:

$$s_j(a) = \left\{\frac{a - v_j}{\|a - v_j\|}\right\}, j = 1...l \tag{4}$$

Figure 2 shows the principle of local feature fusion. Following the extraction of local features, the extracted features are fused with the global features of the video frames. After the embedment of local features, the max pooling is implemented. Here, the kernel matrix is adopted to judge the similarity between two feature descriptors: the feature descriptor $A=\{a_1, \ldots, a_n\}$ of the target image, and the feature descriptor $B=\{b_1 \ldots, b_n\}$ of the reference image of standard action. Let $\Psi$: $VS^{LO} \rightarrow VS^{HO}$ be the mapping from low-dimensional vector space to high-dimensional vector space; $l(a,b)=(\psi(a)|\psi(b))$ be the matching kernel for the inner product of image features after embedment. Then, the kernel matrix can be expressed as:

$$L(A,B) = \sum_{a\in A}\sum_{b\in B} l(a,b) = \delta(A)^T \delta(B) \tag{5}$$

$$\delta_r(A) = \sum_{a\in A} \Psi(a) \tag{6}$$

Let $\psi(b)$ be a local feature of a video frame after embedment; $\psi(b)^T$ be the transposition of $\psi(b)$. The kernel matrix can be expressed as:

$$L(A,B) = \sum_{a\in A}\sum_{b\in B} \psi(a)\psi(b)^T \tag{7}$$

To effectively weigh the importance of local and global features in match video frames, the role of independent feature in the matching between the target image and the reference image must be highlighted. To realize feature fusion, this paper multiplies the eigenvector of each technical action of rhythmic gymnastics with the attention score. After feature fusion, the kernel matrix can be expressed as:

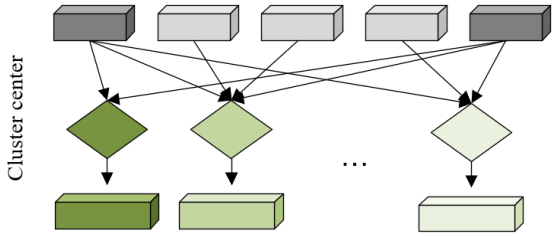


**Figure 2.** Principle of local feature fusion

$$L(A,B) = \sum_{a\in A}\sum_{b\in B} \mu_A(a)\mu_B(b) l(a,b) \tag{8}$$

where, $\mu_A(a)=\beta(g_i;\Psi)$, $a \in A$; $\mu_B(b)=\beta(g_i;\Psi)$, $b \in B$; $\mu_A(a)$ and $\mu_B(b)$ are always greater than zero.

## 3. ACTION RECOGNITION

### 3.1 Sampling algorithm

Despite their simple fusion operations, traditional feature fusion approaches, namely, addition, multiplication, splicing, two-dimensional (2D) convolution, and 3D convolution, fail to fully consider the interaction of effective information between features. To fuse the time-space features for standard sports action recognition, this paper presents a sampling algorithm based on time-space compression.

There are two parts in the sampling algorithm based on time-space compression: data dimensionality reduction, and data fusion. The input image features are dimensionally reduced with the CM-Sketch projection function. Let $d \in R^n$ be the features of a single input image for dimensionality reduction; $n$ and $m$ be the feature dimensions before and after the reduction, respectively; $q$ be the initial zero vector; $f$ and $r$ be parameter vectors. The two parameter vectors are initialized randomly as values from $\{1, ..., m\}$ and $\{-1, +1\}$, respectively. Then, the mapping function can be given by:

$$q(f(i)) = q(f(i)) + r(i)\cdot d(i) \tag{9}$$

where, $i \in \{1, ..., n\}$. During data fusion, the sampling algorithm is simplified, using the convolutional transform formula. Let $\{d_j \in R^n_j\}^l_{j=1}$ be $l$ input features; $\{q_j \in R^m\}^l_{j=1}$ be the data after dimensionality reduction. Then, the output of the proposed sampling algorithm can be calculated by:

$$\phi\left(\{d_j\}^l_{j=1}\right) = F^{-1}\left(\otimes^l_{j=1} F(q_j)\right) \tag{10}$$

where, $\varphi$ is the fusion function; $F$ is the fast Fourier transform (FFT); $F^{-1}$ is the inverse FFT; $\otimes$ is the multiplication of corresponding elements.

### 3.2 Feature information fusion

The proposed algorithm can fuse image features well. However, a serious underfitting may occur, if the algorithm is directly applied to fuse the space-time features of images extracted by the attention mechanism. Facing the time-space continuous sports actions, this paper designs a standard sports action recognition algorithm based on space-time information fusion. The algorithm consists of a fusion module that merges

space features with time features, and a fusion module that merges compressed features. Figure 4 shows the number of channels for the sampling algorithm and standard action recognition algorithm.

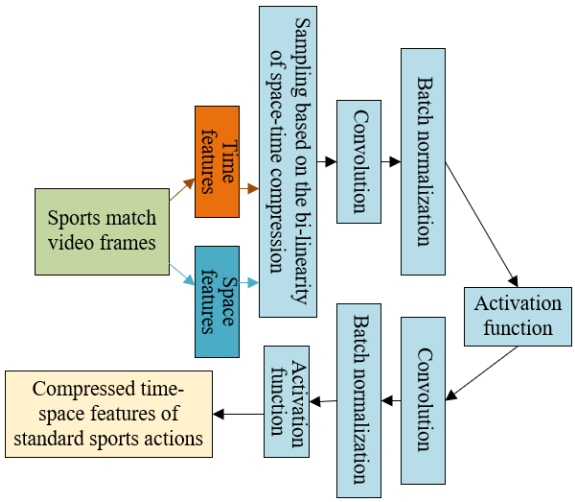


**Figure 3.** Flow of the sampling algorithm based on time-space compression

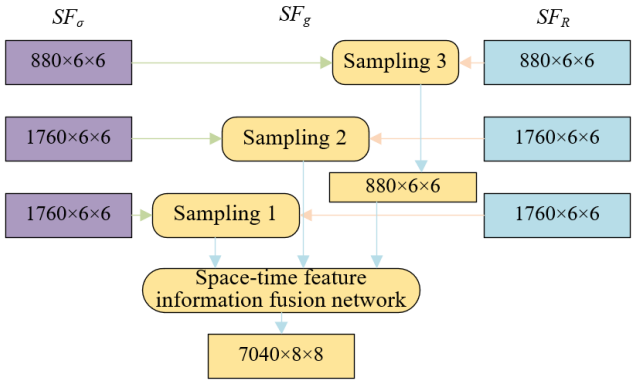


**Figure 4.** Number of channels for the sampling algorithm and standard action recognition algorithm

Let $A_{OR}$ be the image data of the input match video frame; $SF_R$ be the space flow network; $JG_R$ be the recognition result of the space flow network; $CH_R^k$ be the output feature of the k-th module of the space flow network, $k$=1,2,3. Then, the space flow network can be expressed as:

$$\left[JG_R, CH_R^1, CH_R^2, CH_R^3\right] = SF_R\left(A_{OR}\right) \tag{11}$$

Let $A_{GL}$ be the optical flow data of the input match video frame; $SF_\sigma$ be the time flow network; $JG_\sigma$ be be the recognition result of the time flow network; $CH_\sigma^k$ be the output feature of the k-th module of the time flow network. Then, the time flow network can be expressed as:

$$\left[JG_R, CH_R^1, CH_R^2, CH_R^3\right] = SF_\sigma\left(A_{GL}\right) \tag{12}$$

Let $SF_g$ be the fusion flow network; $JG_E$ be the recognition results of the fusion flow network. Then, the fusion flow network can be expressed as:

$$JG_E = SF_g\left(CH_R^1, CH_R^2, CH_R^3, CH_\sigma^1, CH_\sigma^2, CH_\sigma^3\right) \tag{13}$$

The action recognition result $SSR$ of the entire network can be expressed as:

$$SSR\left(A_{OR}, A_{GL}\right) = \frac{\omega \cdot JG_R + \lambda \cdot JG_\sigma + \varepsilon \cdot JG_E}{\omega + \lambda + \varepsilon} \tag{14}$$

where, $\omega$, $\lambda$, and $\varepsilon$ are the weights of the space flow network, time flow network, and fusion flow network, respectively. The default values of $\omega$, $\lambda$, and $\varepsilon$ are 1, 1, and 10, respectively.

By the time axis, the image data of match video frames are divided equally into three segments. From each segment, one image frame and five continuous optical flow images are selected randomly. The selected images are processed by the space flow network, time flow network, and fusion flow network. The outputs are the standard sports actions recognized in these segments. Figure 5 shows the structure of the standard sports action recognition network based on time-space feature information fusion.

Finally, all recognition results are averaged and fused to obtain the final recognition result $SSRT$ on standard sports actions. Let $A_{OR}^\tau$ and $A_{GL}^\tau$ be the frame of the $\tau$-th segment of match video data $VD$, and the corresponding optical flow image, respectively, $\tau$=1, 2, 3; $JG_R^\tau$, $JG_\sigma^\tau$, and $JG_E^\tau$ be the recognition result of the $\tau$-th segment of match video data $VD$ obtained by the space flow network, time flow network, and fusion flow network, respectively. Then, the $SSRT$ can be calculated by:

$$\begin{aligned} SSRT\left(VD\right) &= \frac{\sum_{\tau=1}^{3} SSR\left(A_{OR}^\tau, A_{GL}^\tau\right)}{3} \\ &= \frac{\omega \cdot \sum_{\tau=1}^{3} JG_R^\tau + \lambda \cdot \sum_{\tau=1}^{3} P_\sigma^\tau + \varepsilon \cdot \sum_{\tau=1}^{3} JG_E^\tau}{3 \cdot \left(\omega + \lambda + \varepsilon\right)} \end{aligned} \tag{15}$$

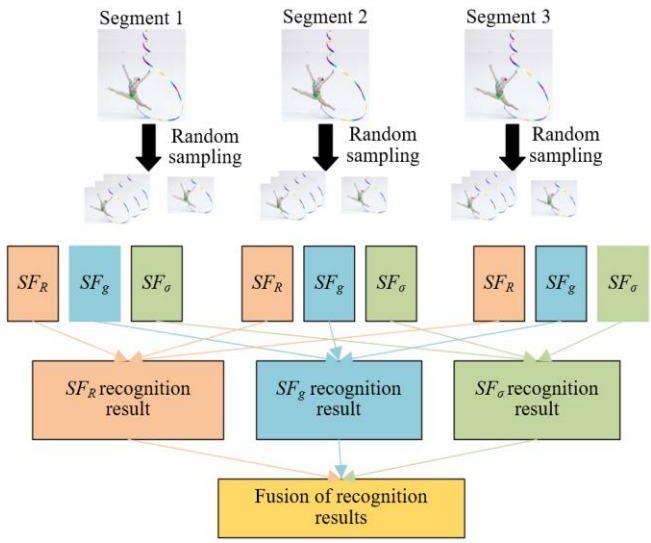


**Figure 5.** Structure of the standard sports action recognition network based on time-space feature information fusion

## 4. EXPERIMENTS AND RESULTS ANALYSIS

Using OpenCV library of Python programming software, this paper extracts individual frames from the video clips on sports matches. On the CUDA parallel computing frame, TVL1 optical flow algorithm was employed to extract the

optical flow images from the extracted sports action frames. Taking cross entropy function as the loss function, the model was optimized through stochastic gradient descent (SDG), with the momentum parameter being 0.9. Table 1 presents the experimental settings of the space flow network, time flow network, and fusion flow network. After reaching a certain number of training cycles, the learning rate would automatically drop to 1/10 of the current value. The number of the last iteration is the maximum number of training cycles. In this case, the learning rate dropped to zero, and the network training terminated.

Based on our algorithm, the fusion strategy of local features can be adjusted by changing the threshold setting. Here, the threshold is mainly used for feature classification: whether the features should be allocated to all cluster centers or only to the nearest center. The threshold values depend on the highest attentions core of all local features for all the actions in the video frame. The highest score multiplied with the threshold parameter yields the threshold. The threshold should not be too large or too small. Otherwise, the local action features of the frame would be filtered out or allocated to all cluster centers, making it impossible to extract representative local features.

To find the ideal threshold, comparative experiments were carried out with different number of cluster centers: 20, 40, 80, and 160. As shown in Figure 6, compared with the traditional method without any limitation on threshold, the mAP of our algorithm tended to be stable, after the threshold surpassed 0.85. To stabilize the experimental effect, the threshold parameter was fixed at 0.88 in the subsequent experiments.

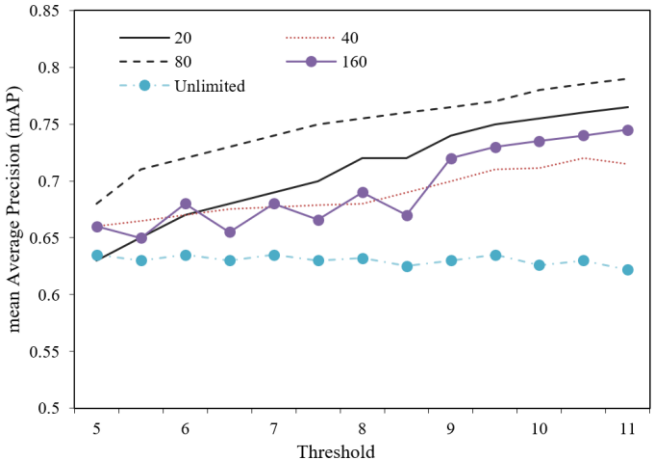


**Figure 6.** mAP of our algorithm at different thresholds

**Table 1.** Experimental settings of different networks

| Phase | | $SF_R$ | $SF_\sigma$ | $SF_g$ |
|---|---|---|---|---|
| Batch size | | 45 | 42 | 17 |
| *Dropout* | | 0.7 | 0.8 | 0.5 |
| Training cycles | Sample set 1 | 50,90,130,180 | 160,250,360,500 | 60,120,180 |
| | Sample set 2 | 60,100,150,220 | 170,270,470,600 | 50,100,150 |

**Table 2.** mAP of dimensionality reduction before fusion

| **Parameter setting** | | **80,1000** | **80,500** | **80,160** | **80,80** | **80,40** |
|---|---|---|---|---|---|---|
| Sample set 1 | Before fusion | 0.748 | 0.958 | 0.926 | 0.954 | 0.932 |
| | After fusion | 0.925 | 0.911 | 0.953 | 0.972 | 0.948 |
| Sample set 2 | Before fusion | 0.827 | 0.922 | 0.926 | 0.948 | 0.886 |
| | After fusion | 0.852 | 0.754 | 0.928 | 0.975 | 0.964 |
| **Parameter setting** | | **80,20** | **40,160** | **40,80** | **40,40** | **40,20** |
| Sample set 1 | Before fusion | 0.851 | 0.915 | 0.962 | 0.985 | 0.937 |
| | After fusion | 0.885 | 0.914 | 0.958 | 0.925 | 0.941 |
| Sample set 2 | Before fusion | 0.874 | 0.829 | 0.915 | 0.938 | 0.977 |
| | After fusion | 0.816 | 0.822 | 0.948 | 0.937 | 0.936 |

**Table 3.** mAP of fusion before dimensionality reduction

| **Fusion scale** | | **2000** | **1000** | **500** | **250** |
|---|---|---|---|---|---|
| Sample set 1 | Before fusion | 0.925 | 0.853 | 0.911 | 0.928 |
| | After fusion | 0.887 | 0.952 | 0.941 | 0.842 |
| Sample set 2 | Before fusion | 0.937 | 0.928 | 0.846 | 0.882 |
| | After fusion | 0.857 | 0.826 | 0.837 | 0.819 |
| **Fusion scale** | | **160** | **80** | **40** | **20** |
| Sample set 1 | Before fusion | 0.948 | 0.916 | 0.852 | 0.942 |
| | After fusion | 0.916 | 0.815 | 0.937 | 0.981 |
| Sample set 2 | Before fusion | 0.916 | 0.937 | 0.849 | 0.918 |
| | After fusion | 0.846 | 0.915 | 0.982 | 0.938 |

Apart from threshold, the dimensionality of the local eigenvector of sports actions is another important parameter of the proposed algorithm. Similarly, comparative experiments were designed to reveal the influence of local eigenvector dimensionality on recognition accuracy. Specifically, the dimensionality of the local eigenvector was reduced from 1,000 to 500, 160, 80, 40, and 20. The standard actions of rhythmic gymnastics were recognized on the sports match video database repeatedly. The experimental results are displayed in Tables 2 and 3.

During the dimensionality reduction of local features through principal component analysis (PCA), it was found that the dimensionally reduced local features are more recognizable than the original local features. Comparing

Tables 2 and 3, it can be learned that the sports actions were recognized more accurately when the local features went through dimensionality reduction before fusion, than when they went through fusion before dimensionality reduction. The mAP of sports action recognition was 0.974 when dimensionality reduction precedes fusion, and 0.931 when fusion precedes dimensionality reduction. The former approach realizes the better accuracy, mainly because the PCA filters out the non-salient local features. These features cannot be eliminated through PCA dimensionality reduction, when the local features have already been fused.

Further experiments found that a low-dimensional feature descriptor led to a better action recognition effect than a high-dimensional feature descriptor. To improve model recognition in real applications, it is not necessary to blindly increase the dimensionality of the feature descriptor.

Our experiments also show that local features greater than 1,000 dimensions contain much redundant information. The results of Tables 2 and 3 suggest that our model can better recognize sports actions, using a low-dimensional feature descriptor. Based on the proposed attention-based local feature fusion method, the standard sports action recognition model increased the recognition efficiency on sports match video datasets by 13.1% and 8.6%, respectively. In cooperation with global features, the proposed feature fusion method further improved the recognition accuracy on sports match video datasets by 1.45% and 0.84%, respectively.

The effectiveness of the proposed feature fusion algorithm was further demonstrated through more comparison experiments. Figure 7 presents the precision-recall (P-R) curves of our method and single-feature extraction algorithms. Our method, which combines local feature fusion and global features, achieved superior recognition performance than the local feature extraction algorithm and the global feature extraction algorithm, on sports match video datasets. In addition, our algorithm can effectively enhance the recognition accuracy.

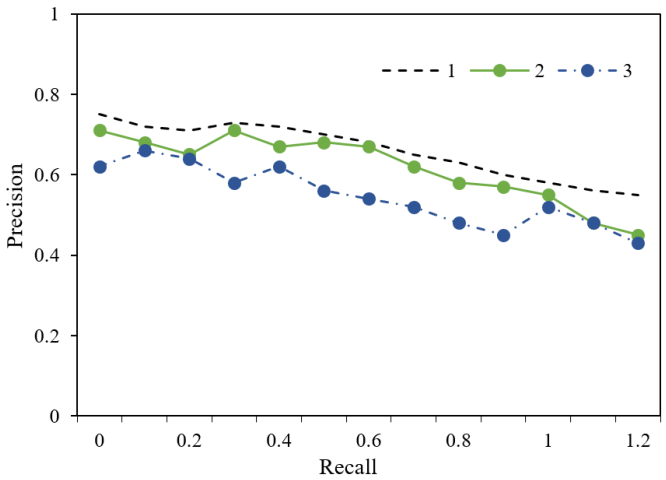


**Figure 7.** P-R curves of different models

## 5. CONCLUSIONS

This paper proposes an image recognition method of standard actions in sports videos, which merges local and global features. After summing up the functions and performance required for the recognition of standard actions of sports, the authors proposed an attention-based local feature extraction algorithm for the frames of sports match videos. On this basis, a sampling algorithm was developed based on time-space compression, and a standard sports action recognition algorithm was designed based on time-space feature fusion, trying to fuse the time-space features of the standard actions in sports match videos, and to overcome the underfitting problem of direct fusion of time-space features extracted by the attention mechanism. The mAP of our algorithm was tested at different thresholds, providing a suitable threshold parameter for subsequent experiments. Then, comparative experiments were carried out to measure the influence of the threshold parameter on recognition accuracy. The results show that the sports actions can be recognized more accurately, when dimensionality reduction of local features precedes fusion, than when fusion precedes dimensionality reduction. Further, the P-R curves of our method and single-feature extraction algorithms were obtained, indicating that our algorithm can effectively enhance the recognition accuracy.

During the experiments and writing, the authors noticed several key issues worthy of further exploration and reflection: (1) The proposed standard sports action recognition network only works when all the target sports actions have been fully executed by the network. Hence, the proposed model may not apply to the unlearned sports actions. (2) The feature extraction and feature fusion steps could be combined to improve the computing efficiency of our model.

## REFERENCES


[1] Xiang, L. (2013). An approach to evaluating the physical fitness of Chinese national women's football players with linguistic information. Journal of Computational Information Systems, 9(8): 2965-2971.

[2] Fang, W. (2020). Based on J2EE research and realization of national fitness sports convenience service platform. Journal of Physics: Conference Series, 1693(1): 012003. 10.1088/1742-6596/1693/1/012003

[3] Dong, Y., Ji, Q. (2014). Research on badminton sports in national fitness activities. Frontier and Future Development of Information Technology in Medicine and Education, pp. 1941-1945. https://doi.org/10.1007/978-94-007-7618-0_229

[4] Zhang, J.H. (2014). Research on the status quo and the character of the development of the national fitness program. Advanced Materials Research, 998: 1769-1772. https://doi.org/10.4028/www.scientific.net/AMR.998-999.1769

[5] Pan, J. (2017). An advanced mode analysis for spread and development of national traditional sports based on internet information technology. Boletin Tecnico/Technical Bulletin, 55(18): 420-425.

[6] Wang, Z., Zhu, D. (2021). Sports monitoring method of national sports events based on wireless sensor network. Wireless Communications and Mobile Computing, 2021: 5739049. https://doi.org/10.1155/2021/5739049

[7] Pan, W., Liu, B., Song, Z. (2021). Edge Computing-induced caching strategy for national traditional sports video resources by considering unusual items. International Journal of Distributed Systems and Technologies (IJDST), 12(2): 1-12. https://doi.org/10.4018/IJDST.2021040101

[8] Guo, C.Y., Xu, S.J., Luo, G. (2020). Research on the integration development of sports intangible cultural

heritage and national fitness. IOP Conference Series: Earth and Environmental Science, 510(3): 032002. https://doi.org/10.1088/1755-1315/510/3/032002
[9] Jiang, J. (2017). Study on the inheritance and development of national traditional sports from the perspective of culture. Agro Food Industry Hi-Tech, 28(1): 862-865.
[10] Wang, P. (2021). Research on sports training action recognition based on deep learning. Scientific Programming, 2021: 3396878. https://doi.org/10.1155/2021/3396878
[11] Liu, H., Li, J. (2017). Sports biomechanics characteristics of breakthrough technical action of step holding and cross step holding of basketball player. Agro Food Industry Hi-Tech, 28(3): 780-784.
[12] Zhou, Q. (2017). An international comparative study of children's learning goals in sports and health from the perspective of action development. Agro Food Industry Hi-Tech, 28(3): 733-737.
[13] Sun, C., Ma, D. (2021). SVM-based global vision system of sports competition and action recognition. Journal of Intelligent & Fuzzy Systems, 40(2): 2265-2276. https://doi.org/10.3233/JIFS-189224
[14] Tejero-de-Pablos, A., Nakashima, Y., Sato, T., Yokoya, N., Linna, M., Rahtu, E. (2018). Summarization of user-generated sports video by using deep action recognition features. IEEE Transactions on Multimedia, 20(8): 2000-2011. https://doi.org/10.1109/TMM.2018.2794265
[15] Wang, W. (2017). Research on sports action recognition based on somatosensory technology. Agro Food Industry Hi-Tech, 28(3): 2588-2592.
[16] Hollett, T. (2019). Symbiotic learning partnerships in youth action sports: Vibing, rhythm, and analytic cycles. Convergence, 25(4): 753-766. https://doi.org/10.1177%2F1354856517735840
[17] Su, N. (2017). Research on sports action recognition based on wireless sensor. Agro Food Industry Hi-Tech, 28(3): 2637-2640.
[18] Abdellaoui, M., Douik, A. (2020). Human action recognition in video sequences using deep belief networks. Traitement du Signal, 37(1): 37-44. https://doi.org/10.18280/ts.370105
[19] Wang, Y.N., Yang, Y.M., Li, Y. (2020). Recognition and difference analysis of human walking gaits based on intelligent processing of video images. Traitement du Signal, 37(6): 1085-1091. https://doi.org/10.18280/ts.370621
[20] Liu, G., Zhang, D., Li, H. (2014). Research on action recognition of player in broadcast sports video. International Journal of Multimedia and Ubiquitous Engineering, 9(10): 297-306. http://dx.doi.org/10.14257/ijmue.2014.9.10.29
[21] Li, H., Tang, J., Wu, S., Zhang, Y., Lin, S. (2009). Automatic detection and analysis of player action in moving background sports video sequences. IEEE Transactions on Circuits and Systems for Video Technology, 20(3): 351-364. https://doi.org/10.1109/TCSVT.2009.2035833
[22] Ni, S., Yao, D. (2021). Sports dance action recognition system oriented to human motion monitoring and sensing. Wireless Communications and Mobile Computing, 2021: 5515352. https://doi.org/10.1155/2021/5515352
[23] Jiang, H., Tsai, S.B. (2021). An empirical study on sports combination training action recognition based on SMO algorithm optimization model and artificial intelligence. Mathematical Problems in Engineering, 2021: 7217383. https://doi.org/10.1155/2021/7217383
[24] Zhai, S. (2021). Research on 3D modeling and detection methods of wrong actions in sports. 2021 International Conference on Control Science and Electric Power Systems (CSEPS), Shanghai, China, pp. 107-111. https://doi.org/10.1109/CSEPS53726.2021.00029