



Visual Image Recognition of Basketball Turning and Dribbling Based on Feature Extraction

Bowen Zhang, Tianqi Wang*

Northeastern University Physical Education Department, Northeastern University Archives, Liaoning 110819, China

Corresponding Author Email: zhangbw@mail.neu.edu.cn

<https://doi.org/10.18280/ts.390624>

Received: 20 August 2022

Accepted: 5 November 2022

Keywords:

feature extraction, feature fusion, basketball, turning and dribbling, image recognition

ABSTRACT

The processing of basketball videos with complex contents faces several challenges in terms of global motion features, group motion features, and individual pose features. The current research cannot solve problems, such as the diverse spatiotemporal features of actions, the utilization of correspondence between spatiotemporal features, the increase of data volume, and the complexity of the network. To solve these problems, this paper studies the visual image recognition of basketball turning and dribbling based on feature extraction. Specifically, the optical flow image was introduced to establish the relationship between the velocity field of the basketball turning and dribbling and the grayscale of the image frame, such as to effectively depict the time variation of pixels. In addition, a convolutional neural network was established based on multi-feature learning to process the sports video image frames, and to extract more spatiotemporal features of basketball turning and dribbling. To improve the feature utilization of the action recognition model, this paper strengthens the extraction of dynamic and static features for the recognition of the player's basketball turning and dribbling in the same scene, and improves the existing convolutional neural network. Furthermore, the multi-feature learning of motion excitation and temporal aggregation of actions were completed. The proposed model was proved effective through experiments.

1. INTRODUCTION

In recent years, the number of sports videos has exploded thanks to the development of Internet technology and multimedia technology [1-4]. To acquire valuable information about players, actions, events from sports videos, it is necessary to analyze sports videos and summarize the tactics [5-9]. Compared with that of other sports, the video analysis of basketball is rather complicated, because basketball requires the cooperation between multiple players. Since the players move very fast, their motions are difficult to capture, making it easy to produce mutual occlusion between players [10-15]. The processing of basketball videos with complex contents faces several challenges in terms of global motion features, group motion features, and individual pose features. The current research cannot solve problems, such as the diverse spatiotemporal features of actions, the utilization of correspondence between spatiotemporal features, the increase of data volume, and the complexity of the network [16-23]. In addition, it is more difficult to analyze events during the movement.

Many scholars have investigated the intelligent training of professional competitive basketball. Their research ranges from video-based body tracking to intelligent motion prediction. Among the current human detection methods, the window-based human representation model is one of the most commonly used models. The model represents the core human body as a rectangular area or a combination of multiple areas with fixed relative positions. However, it does not provide the details of human limbs and torso. Wang and Xue [24] introduced the modeling steps of the human representation model for intelligent training of professional basketball

athletes, conducted intelligent data analysis, and applied the model to video information analysis. Eventually, the designed model was proved robust. Long and Long [25] attempted to predict the player behavior with the improved radial basis function (RBF) network, discussed the basic concepts of the RBF, and proposed the application scheme for the improved RBF-based prediction of player behaviors. Taking basketball game videos as an example, Long and Long the behavior of athletes in the game were predicted, and compared with the results. The comparison verifies the effectiveness of the proposed method. The ball motions of a player can be easily identified using the field-programmable gate array (FPGA). Yin and He [26] detailed the current state of FPGA application in image processing, and designed custom hardware and lightweight processors to overcome the limitations of specific current-level synthesis tools, thereby realizing parallel processing. The experimental results show that the designed framework can accurately distinguish the image edges, and meet the requirements of the image edge recognition of the current videos. Azhar et al. [27] aimed to analyze the parabolic motion when a basketball is thrown from a situational perspective, using descriptive experiment. The videos of the basketball being thrown from 2, 3, and 5m from the foot of the thrower were recorded, and analyzed by the tracker software. The parameters measured were the time the ball in the air, the furthest distance, the maximum altitude, and the total velocity.

Scholars at home and abroad have achieved a lot of valuable results in the identification of key postures or action classes in sports. However, there are still some problems to be solved, such as the diverse spatiotemporal features of actions, the utilization of correspondence between spatiotemporal features, the increase of data volume, and the complexity of the network.

To solve these problems, this paper studies the visual image recognition of basketball turning and dribbling based on feature extraction. Section 2 introduces the optical flow image to establish the relationship between the velocity field of the basketball turning and dribbling and the grayscale of the image frame, such as to effectively depict the time variation of pixels. Section 3 establishes a convolutional neural network based on multi-feature learning to process the sports video image frames, and to extract more spatiotemporal features of basketball turning and dribbling. To improve the feature utilization of the action recognition model, this paper strengthens the extraction of dynamic and static features for the recognition of the player's basketball turning and dribbling in the same scene, and improves the existing convolutional neural network. Furthermore, the multi-feature learning of motion excitation and temporal aggregation of actions were completed. The proposed model was proved effective through experiments.

2. PREPROCESSING OF SPORTS VIDEO DATA

Event analysis in the process of sports aims to identify the start and end points of specific events in sports videos. Figure 1 gives an example of a complete image frame series of a basketball turning and dribbling event. The recognition of the athlete and each part of the action in the image frames lay the basis for the detection of basketball turning and dribbling.



Figure 1. Detection of basketball turning and dribbling in sports videos

Spatial and temporal information are very important for the recognition of basketball turning and dribbling of athletes. Therefore, it is necessary to select appropriate sports video image frames to extract motion features. Otherwise, it would be impossible to extract the valuable spatiotemporal features of basketball turning and dribbling. To obtain the time information of sports video, this paper introduces the optical flow image, establishes the relationship between the velocity field of basketball turning and dribbling and the gray level of the image frames, and effectively illustrates the time variation of pixels.

For an image frame in a sports video, a pixel is denoted by (a,b) , the brightness of the pixel at time $e+\Delta e$ is denoted by $LI(a+\Delta a,b+\Delta b,e+\Delta e)$, the variation of (a,b) on axes a and b is

Δa and Δb , respectively; e is the time interval of pixel changes; v and u are the displacement components in the horizontal and vertical directions of the optical flow of (a, b) . Then, we have:

$$v = \frac{da}{de} \quad (1)$$

$$u = \frac{db}{de} \quad (2)$$

Suppose pixel (a, b) does not have brightness change as Δe approaches 0. Then, we have:

$$LI(a, b, e) = LI(a + \Delta a, b + \Delta b, e + \Delta e) \quad (3)$$

Suppose pixel (a,b) has brightness change when Δe is nonzero. Then, we have:

$$\begin{aligned} & LI(a + \Delta a, b + \Delta b, e + \Delta e) \\ &= LI(a, b, e) + \frac{\partial LI}{\partial e} \Delta e + \frac{\partial LI}{\partial a} \Delta a + \frac{\partial LI}{\partial b} \Delta b + \frac{\partial LI}{\partial e} \Delta e + \rho \end{aligned} \quad (4)$$

Ignoring the second-order infinitesimal term in formula (4), it is possible to obtain the velocity and direction change ∇LIh_q of each pixel when Δe approaches 0 at a certain moment:

$$-\frac{\partial LI}{\partial e} = \frac{\partial LI}{\partial a} \frac{da}{de} + \frac{\partial LI}{\partial b} \frac{db}{de} = \nabla LIh_q \quad (5)$$

Based on the above processing procedure, this paper obtains the optical flow image corresponding to the RGB image of the video frame for the player's basketball turning and dribbling.

3. SPATIOTEMPORAL FEATURE EXTRACTION

This section constructs a convolutional neural network based on multi-feature learning to process the sports video image frames, and to extract more spatiotemporal features of basketball turning and dribbling. Specifically, ResNet 101 and the shallow network BNInception were chosen to extract spatial features, which effectively reduces the computational complexity of the algorithm. Because the value range of RGB image pixels is $[0, 255]$, this paper also introduces cross-modal pre-training to discretize the optical flow field into this range, and sets the weight of the first convolutional layer of the pre-trained model, which is more suitable for processing optical flow images. Figure 2 shows the framework of the action recognition model.

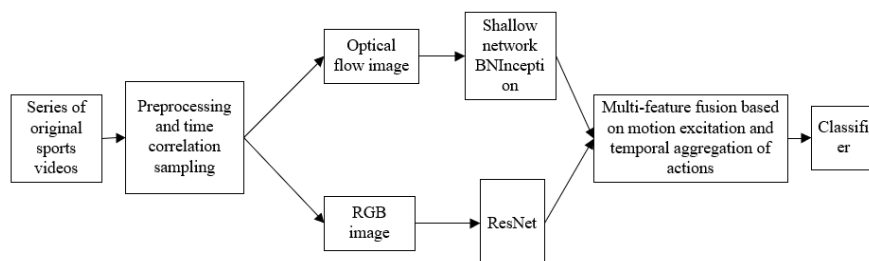


Figure 2. Framework of action recognition model

The sampling method of sports video image frames affects the training and results of the subsequent action recognition model. It is necessary to ensure the consistency of the length and size of the sports video image frame series, as well as the reasonability of the position being selected. This paper chooses to sample the image frames at equal intervals from the original video. Figure 3 illustrates the sampling process. Let m be the finally selected image frames, M be the total number of image frames in the original sports video; E be the selection interval. Then, m , M and E satisfy the following formula:

$$E = \frac{M}{m-1} \quad (6)$$

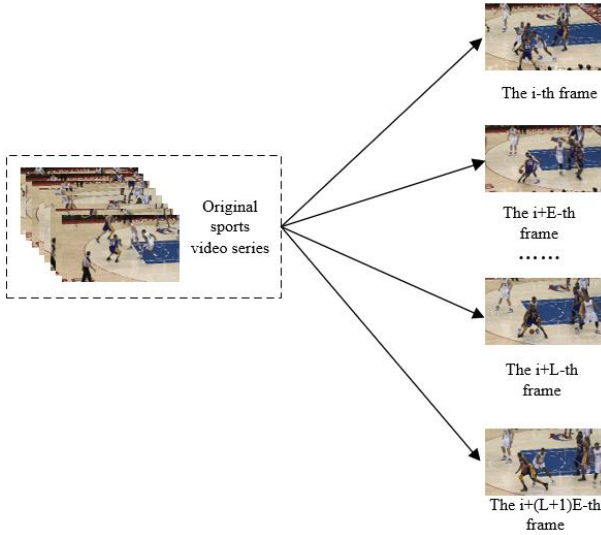


Figure 3. Sampling of sports video image frames

Based on the residual learning unit, ResNet101 has a major advantage: the performance does not change with the deepening of the network layers. Let $G(A)$ be the network output; A be the current input of block; $F(A)$ be the expected output. To fit an ideal $F(A)$, it is necessary to modify the hyperparameters in the network. The residual network of ResNet 101 realizes the approximation of $G(A)$ to $F(A)-A$ based on a short-circuit structure. Let a_k and b_k be the input and output of the k -th residual unit, respectively; G be the residual function; $f(a_k)=a_k$ be the rectified linear unit (ReLU). Then, the residual network can be expressed as:

$$b_i = f(a_i) + G(a_i, Q_i), (Q_i = \{Q_{i,l} | 1 \leq l \leq L\}) \quad (7)$$

$$a_{k+1} = g(b_k) \quad (8)$$

From the shallow layer k to the deep layer K , the learning features of the basketball athlete's turning and dribbling can be expressed as:

$$a_K = a_i + \sum_{i=1}^{K-1} G(a_i, Q_i) \quad (9)$$

It can be seen from the above formula that the output of ResNet101 is only determined by the sum of the output residuals of the previous layer. In this way, exploding or

vanishing gradients will not occur with the deepening of network layers, and the training effect of the network will not be significantly affected.

4. MODEL OPTIMIZATION

In the recognition of athlete's basketball turning and dribbling, it is wrong to emphasize spatial features over spatiotemporal features. The full utilization of the spatiotemporal correlation of continuous action features helps to improve the feature utilization by the action recognition model, making it easier to identify actions with inconsistent timing. Therefore, this paper strengthens the feature extraction for the recognition of basketball turning and dribbling in the same scene, improves the existing convolutional neural network, and conducts multi-feature learning based on the multi-feature learning of motion excitation and temporal aggregation of actions. Figure 4 shows the framework of the optimized model for turning and dribbling recognition.

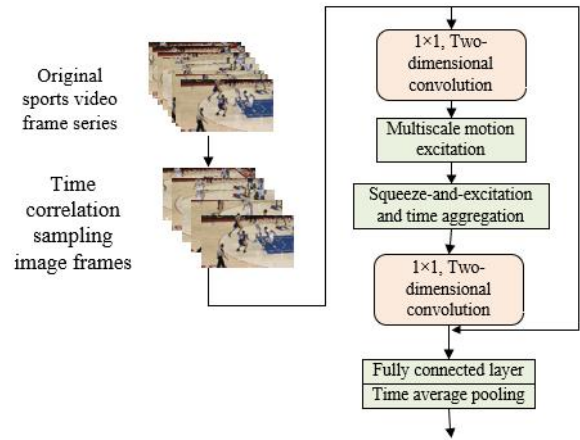


Figure 4. Framework of the optimized recognition model

The proposed action recognition model contains a multiscale motion excitation module, and supports the following functions: the modeling of fixed actions, expansion from pixel level to feature level, and learning of spatiotemporal features.

The shape of the spatiotemporal features input to the model is represented by M , E , D , F , and Q . The size of batch processing BS is denoted by M . The time dimension and feature channel are denoted by E and D , respectively. F and Q stands for the length and width of the corresponding space shape, respectively. In addition, $*$ and s stand for the convolutional operation and reduction scale. Then, the feature A^s after the channel size reduction satisfies:

$$A^s = CO_{RE} * A, A^s \in R^{M \times E \times D / s \times F \times Q} \quad (10)$$

This paper defines the difference between two adjacent frames $A^s(e)$ and $A^s(e-1)$ of the sports video image in the time dimension E as the feature-level motion $N(e-1)$. Let CO_{TR} be a two-dimensional convolution for channel changes. Then, we have:

$$\begin{aligned} N(e-1) &= CO_{TR} * A^s(e) - A^s(e-1), \\ 2 \leq e \leq E, N(e-1) &\in R^{M \times D / s \times F \times Q} \end{aligned} \quad (11)$$

Suppose $N(E)=0$ at the end of the time dimension. By traversing the time dimension, the motion matrix N^{e-1} of the player's basketball turning and dribbling can be constructed based on $[N(1), \dots, N(E)]$. The size of the matrix is $[M, E, D/s, F, Q]$.

To better illustrate the feature-level motions of the athlete's turning and dribbling, this paper extends the adjacent image frames $A^s(e)$ and $A^s(e-1)$ of the motion video to the three adjacent image frames $A^s(e+1)$, $A^s(e)$, and $A^s(e-1)$. Similarly, $A^s(e)$, which has completed one channel transform, is subtracted from $A^s(e+1)$, which has completed two channel transforms, producing the difference between the adjacent image frames $A^s(e)$ and $A^s(e-1)$ in dimension E. That is, the feature-level motion description $N(e)$ of the athlete's turning and dribbling at time e can be obtained by:

$$\begin{aligned} N(e) &= CO_{TA} * [CO_{TR} * A^s(e+1)] \\ &- CO_{TR} * A^s(e), 1 \leq e \leq E-1 \\ N(e) &\in R^{M \times D/s \times F \times Q} \end{aligned} \quad (12)$$

Similarly, we can obtain the motion matrix N^e of the size $[M, E, D/s, F, Q]$. Further, N^{e-1} is superimposed on N^e to obtain the feature-level motion description N between the three adjacent image frames $A^s(e+1)$, $A^s(e)$, and $A^s(e-1)$. To excite the motion sensitive channel of the said action, this paper summarizes the spatial information of all image frames, using the global average pooling layer:

$$N^r = Pool(N), N^r \in R^{M \times E \times D/s \times 1} \quad (13)$$

Then, a 1×1 two-dimensional convolution is adopted to restore the number of channels N^* , which are about to complete pooling, to $[M, E, D, 1, 1]$. After the activation of sigmoid function, the weight X of the dynamic features of the athlete's basketball turning and dribbling is obtained. Let ξ be the sigmoid function; CO_{EX} be the convolutional operation that restores the number of channels. Then, we have:

$$X = 2\xi(CO_{EX} * N^r) - 1, X \in R^{M \times E \times D \times 1 \times 1} \quad (14)$$

The feature activation of the turning and dribbling features is completed by multiplying the action with X . Meanwhile, the static features of the action, which plays a role in motion recognition, are preserved based on the residual connection. Let A^p be the output of the multiscale motion excitation module; \otimes be the channel-based multiplication. Then, we have:

$$A^p = A + A \otimes X, A^p \in R^{M \times E \times D \times F \times Q} \quad (15)$$

The basketball player turns and dribbles rather quickly. It is very difficult to establish an effective correspondence between temporal and spatial features. The problem cannot be solved by setting many local temporal convolution structures in the recognition model. Therefore, this paper constructs a squeeze-and-excitation time aggregation module, i.e., builds up a sub-convolution hierarchy. The transmission of valuable action features is highlighted by squeezing and exciting the shared features between convolutional layers. In this way, the receptive field of the time dimension is expanded, while the transfer and sharing of invalid action features is suppressed between convolutional layers.

The given features of the player's basketball turning and dribbling are divided along the channel dimension to form 4 image frame subseries of the shape $[M, E, D/4, F, Q]$. The last 3 image feature subseries are processed based on temporal and spatial sub-convolutions. Suppose the A_i^p is the output of the i -th image feature subseries; CO_{ZJ} is the spatial sub-convolution; CO_{SJ} is the temporal sub-convolution; G_{rt} be the squeeze-and-excitation method. Then, we have:

$$\begin{aligned} A_i^p &= A_i, & i &= 1 \\ A_i^p &= CO_{ZJ} * (CO_{SJ} * A_i), & i &= 2 \\ A_i^p &= CO_{ZJ} * [CO_{SJ} * (A_i + G_{rt}(A_{i-1}^p))], & i &= 3, 4 \end{aligned} \quad (16)$$

After the spatial and temporal sub-convolutions, the three image frame sub-series are parallel convolution structures. To expand the receptive field of the time dimension, it is necessary to short-circuit them to form a hierarchical convolution structure. The squeeze-and-excitation operation G_{rt} on the subseries is arranged before the residual connection is passed into the next subseries. Specifically, squeezing compresses the spatial features of each subseries into $[M, E, D/4, 1, 1]$. Then, adaptive learning is performed on the features based on two fully connected layers, producing new feature maps.

Let R be the feature map containing the importance information; G_{ta} be the weighted excitation function; V be the feature map passing through two fully-connected layers; Q_1 and Q_2 be the weights of the first and second fully connected layers, respectively; ε and ξ be sigmoid and ReLU activation functions, respectively. Both Q_1 and Q_2 are collectively referred to as activation function. Then, the above process can be characterized by:

$$R = G_{ta}(V, Q) = \varepsilon(Q_2 \xi(Q_1 V)) \quad (17)$$

The resulting new feature map is of the size $[M, E, D/4, 1, 1]$. In the squeeze-and-excitation time aggregation module, there are differences in the receptive fields of the subseries obtained via division. Before outputting the final results, this paper cascades the multiple outputs:

$$A^p = [A_1^p; A_2^p; A_3^p; A_4^p], A^p \in R^{M \times E \times D \times F \times Q} \quad (18)$$

With the increase of the receptive fields of the image frames, the model output will contain effective spatiotemporal feature correlations. This optimization strategy improves the recognition effect of the model.

5. EXPERIMENTS AND RESULTS ANALYSIS

This paper selects different segments of the same basketball game video, and further calculates the correlations between the image frames of the athlete' basketball turning and dribbling in these segments. Despite the short overlaps between the frames, there is an obvious correlation between the frames in the image frame series of the dribbling action. To disclose the influence of learning the spatiotemporal features of dribbling over the motion recognition of the model, the authors analyzed the spatiotemporal feature correlations between frames in the dribbling image frame series. According to the analysis results

in Table 1, there was an obvious difference in the spatiotemporal correlation between the frames in the extracted image frame series. The most significant correlation was observed between frames 5 and 6 in segment 1, and between

frames 2 and 3 in segment 2. Thus, the extracted spatiotemporal features of the image frames are effectively, and can be imported to the action recognition model for further feature learning.

Table 1. Spatiotemporal feature correlations between different image frames

	Correlation (%)	Frame 1	Frame 2	Frame 3	Frame 4	Frame 5	Frame 6
Video segment 1	Frame 1	61.28	39.21	41.16	47.38	46.29	43.61
	Frame 2	31.05	71.24	38.49	40.62	31.17	38.29
	Frame 3	47.15	39.62	55.27	41.57	49.53	44.62
	Frame 4	49.68	48.57	42.51	52.69	42.91	49.85
	Frame 5	40.31	30.16	46.39	41.32	55.16	44.37
	Frame 6	49.52	38.59	40.17	45.81	72.37	66.29
Video segment 2	Frame 1	52.69	28.37	32.58	46.29	48.15	41.27
	Frame 2	21.30	52.49	70.16	40.81	22.61	26.59
	Frame 3	33.27	21.01	58.31	43.68	49.28	49.81
	Frame 4	46.19	49.27	41.26	62.53	41.05	38.57
	Frame 5	40.58	23.61	49.38	47.51	55.36	40.13
	Frame 6	41.74	28.95	47.52	39.28	49.81	69.27

Table 2. Recognition accuracies before and after model optimization

Model	Type of backbone network	Pretraining method	Accuracy (%)	Fusion accuracy (%)
Traditional convolutional neural network	A	1	62.3%	64.1%
	B	1	65.8%	
Pre-optimization	A	1	69.6%	73.4%
	B	2	77.2%	
Post-optimization	A	1	71.7%	80.6%
	B	2	89.5%	

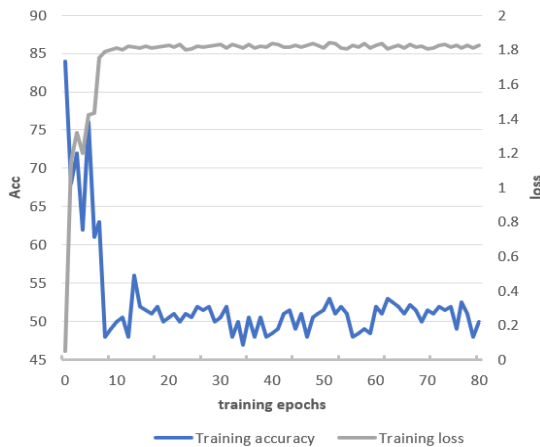


Figure 5. Training results of the RGB channels of the model

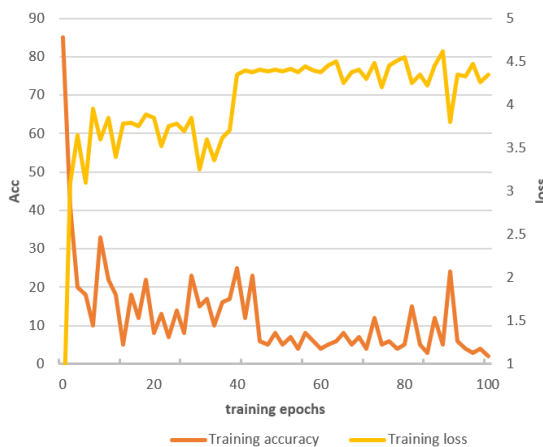


Figure 6. Training results of the optical flow channel of the model

Figures 5 and 6 show the training results of the RGB channels and optical flow channel of the model, respectively. The results include the loss function and recognition accuracy. It can be seen that the motion recognition accuracy of the RGB channels and the optical flow channel tended to be stable at 20 and 40 epochs, respectively. The comparison shows that, although the shallow network BNInception was adopted as the backbone of the optical flow channel, the long convergence time slows down the training of the entire model.

To verify the effectiveness of the additional motion excitation and time aggregation module, this paper compares the recognition accuracies before and after model optimization. The results are displayed in Table 2, where A and B stand for the backbone network ResNet101 and the shallow network BNInception, respectively; pretraining methods 1 and 2 refer to ImageNet and cross-modal pretraining, respectively. It can be seen that, the motion recognition accuracy of the optimized motion recognition model was significantly higher than that of the traditional convolutional neural network, and that of the model without the motion excitation and time aggregation module. The main reason is that the optimized model fuses the dynamic features and static features extracted by the RGB channels and the optical flow channel, such that the spatial accuracy of motion recognition is marked improved.

To further verify its effectiveness, our model was compared with the other video image-based action recognition models, including *HRBNN*, *DeepLSTM*, *ST-LSTM*, *Ind-RNN*, *TCN*, *ST-GCN*, and *SGN6*. The comparison results in Table 3 and Figure 7 reveal that our model achieved superior recognition performance than the other models. The recognition accuracy of our model was 7~8% higher than that of the other models on image sets 1 and 2. Among the turning and dribbling actions, reverse lay-up and cross-leg dribbling are the most easily confusing actions. The confusion rate of turning and dribbling and lay-up shoot was 1.45%, while that of reverse

lay-up and turning and dribbling was 2.11%. The experimental results show that our model can extract consistent static and dynamic features of basketball turning and dribbling, and recognize the actions ideally.

Table 3. Recognition accuracies of different models

Model	Video set 1	Video set 2
<i>HRBNN</i>	52.6	69.5
<i>DeepLSTM</i>	67.4	64.8
<i>ST-LSTM</i>	65.8	71.6
<i>Ind-RNN</i>	85.9	82.4
<i>TCN</i>	72.9	88.5
<i>ST-GCN</i>	85.2	81.6
<i>SGN</i>	88.5	94.5
Our model	91.02	98.9

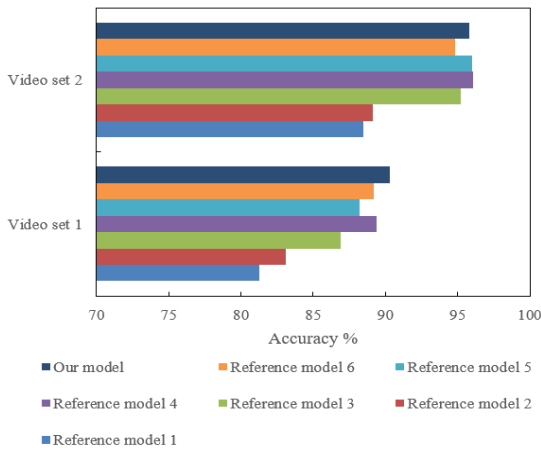


Figure 7. Recognition accuracies of different models

6. CONCLUSIONS

This paper studies the visual image recognition of basketball turning and dribbling based on feature extraction. Firstly, the optical flow image was introduced to establish the relationship between the velocity field of the basketball turning and dribbling and the grayscale of the image frame, such as to effectively depict the time variation of pixels. Next, a convolutional neural network was established based on multi-feature learning to process the sports video image frames, and to extract more spatiotemporal features of basketball turning and dribbling. To better use the features of the action recognition model, this paper strengthens the extraction of dynamic and static features for the recognition of the player's basketball turning and dribbling in the same scene, and improves the existing convolutional neural network. Furthermore, the multi-feature learning of motion excitation and temporal aggregation of actions were completed. Through experiments, the authors analyzed the spatiotemporal feature correlations between the frames in the image frame series of the dribbling action, and verifies the effectiveness of the extracted spatiotemporal features of the image frames, which can be imported to the action recognition model for further multi-feature learning. In addition, the authors obtained the training results of the RGB channels and optical flow channel of the model, compared the recognition accuracies before and after model optimization, and contrasted the performance of our model with other video image-based action recognition models. The experimental results show that our model can

extract consistent static and dynamic features of basketball turning and dribbling, and recognize the actions ideally.

REFERENCES

- [1] Xu, L., Choi, D., Yang, Z. (2022). Deep Neural network-based sports marketing video detection research. *Scientific Programming*, Article ID: 8148972. <https://doi.org/10.1155/2022/8148972>
- [2] Wu, S.J. (2021). Image recognition of standard actions in sports videos based on feature fusion. *Traitement du Signal*, 38(6): 1801-1807. <https://doi.org/10.18280/ts.380624>
- [3] Xu, C., Li, Y. (2022). Sports video moving object detection and tracking technology based on hybrid algorithm. In *Innovative Computing*, Singapore, 1799-1803. https://doi.org/10.1007/978-981-16-4258-6_234
- [4] Zhang, Y., Tang, H., Zereg, F., Xu, D. (2022). Application of deep convolution network algorithm in sports video hot spot detection. *Frontiers in Neurorobotics*, 16: 829445-829445. <https://doi.org/10.3389/fnbot.2022.829445>
- [5] Zhao, X. (2022). Research on athlete behavior recognition technology in sports teaching video based on deep neural network. *Computational Intelligence and Neuroscience*, Article ID: 7260894. <https://doi.org/10.1155/2022/7260894>
- [6] Liu, X., Liu, J., Fan, J., Li, X. (2022). Research on wushu sports feedback system based on real-time target intelligent tracking video processing. *Security and Communication Networks*, Article ID: 4599005. <https://doi.org/10.1155/2022/4599005>
- [7] Park, G.M., Hyun, H.I., Kwon, H.Y. (2022). Multimodal learning model based on video-audio-chat feature fusion for detecting e-sports highlights. *Applied Soft Computing*, 126: 109285. <https://doi.org/10.1016/j.asoc.2022.109285>
- [8] Liu, Y.G., Wu, Y. (2021). A multi-feature motion posture recognition model based on genetic algorithm. *Traitement du Signal*, 38(3): 599-605. <https://doi.org/10.18280/ts.380307>
- [9] Yang, X., Wang, H. (2022). Design and implementation of intelligent analysis technology in sports video target and trajectory tracking algorithm. *Wireless Communications and Mobile Computing*, Article ID: 5633066. <https://doi.org/10.1155/2022/5633066>
- [10] Yan, Z., Yu, Y., Shabaz, M. (2021). Optimization research on deep learning and temporal segmentation algorithm of video shot in basketball games. *Computational Intelligence and Neuroscience*, Article ID: 4674140. <https://doi.org/10.1155/2021/4674140>
- [11] Kakuta, M., Kinoshita, Y., Watanabe, I. (2021). An Imaging system for teleteaching support using multi-view video shot by basketball players. In *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, Kyoto, Japan, pp. 841-842. <https://doi.org/10.1109/GCCE53005.2021.9621794>
- [12] Lin, J., Sun, L., Song, J. (2020). Research on the application of artificial intelligence video feedback system in college basketball shooting teaching. In *2020 International Conference on Computer Science and Management Technology (ICCSMT)*, Shanghai, China,

- pp. 144-148.
<https://doi.org/10.1109/ICCSMT51754.2020.00035>
- [13] Zou, W., Jin, Z. (2020). Algorithm for motion video based on basketball image. In *The International Conference on Cyber Security Intelligence and Analytics*, pp. 792-799. https://doi.org/10.1007/978-3-030-43306-2_111
- [14] Chen, L., Wang, W. (2020). Analysis of technical features in basketball video based on deep learning algorithm. *Signal Processing: Image Communication*, 83: 115786. <https://doi.org/10.1016/j.image.2020.115786>
- [15] Quiroga, J., Carrillo, H., Maldonado, E., Ruiz, J., Zapata, L.M. (2020). As seen on tv: Automatic basketball video production using gaussian-based actionness and game states recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Seattle, WA, USA, pp. 894-895. <https://doi.org/10.1109/CVPRW50498.2020.00455>
- [16] Teket, O.M., Yetik, I.S. (2020). A fast deep learning based approach for basketball video analysis. In *Proceedings of the 2020 4th International Conference on Vision, Image and Signal Processing (24)*: 1-6. <https://doi.org/10.1145/3448823.3448882>
- [17] Nopiana, R., Akbar, M.F., Tamim, M.H. (2020). The Influence of using video media on basic movement skills lay-up shoot basketball. In *Journal of Physics: Conference Series*, 1539(1): 012020. <https://doi.org/10.1088/1742-6596/1539/1/012020>
- [18] Feng, H. (2020). Introduction of video analysis technology in basketball teaching under the background of "Internet+". In *Innovative Computing*, Singapore, pp. 1037-1042. https://doi.org/10.1007/978-981-15-5959-4_127
- [19] Zhou, J., Fu, W. (2020). Quality evaluation of degraded basketball video image restoration based on classification learning. *International Journal of Performability Engineering*, 16(3): 392-400.
- [20] Monezi, L.A., Calderani Junior, A., Mercadante, L.A., Duarte, L.T., Misuta, M.S. (2020). A video-based framework for automatic 3D localization of multiple basketball players: A combinatorial optimization approach. *Frontiers in Bioengineering and Biotechnology*, 8: 286. <https://doi.org/10.3389/fbioe.2020.00286>
- [21] Liang, Q., Mei, L., Wu, W., Sun, W., Wang, Y., Zhang, D. (2019). Automatic basketball detection in sport video based on R-FCN and Soft-NMS. In *Proceedings of the 2019 4th International Conference on Automation, Control and Robotics Engineering*, New York, NY, United States, pp. 1-6. <https://doi.org/10.1145/3351917.3351970>
- [22] Feng, S., Sheng, S. (2018). Construction of multifunctional video conversion-based multimedia teaching system for college basketball. *International Journal of Emerging Technologies in Learning (Online)*, 13(6): 176-176. <https://doi.org/10.3991/ijet.v13i06.8587>
- [23] Chen, L.H., Chang, H.W., Hsiao, H.A. (2017). Player trajectory reconstruction from broadcast basketball video. In *Proceedings of the 2nd International Conference on Biomedical Signal and Image Processing*, New York, NY, United States, pp. 72-76. <https://doi.org/10.1145/3133793.3133801>
- [24] Wang, L., Xue, Q. (2022). Intelligent professional competitive basketball training (IPCBT): From video based body tracking to smart motion prediction. In *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, Erode, India, pp. 1398-1401. <https://doi.org/10.1109/ICSCDS53736.2022.9761028>
- [25] Long, Y., Long, J. (2022). Basketball players' behavior prediction method based on video. In *International Conference on Cognitive based Information Processing and Applications (CIPA 2021)*, Singapore, pp. 269-275. https://doi.org/10.1007/978-981-16-5854-9_34
- [26] Yin, L., He, R. (2021). Target state recognition of basketball players based on video image detection and FPGA. *Microprocessors and Microsystems*, 80: 103340. <https://doi.org/10.1016/j.micpro.2020.103340>
- [27] Azhar, T.A.N., Mulyaningsih, N.N., Saraswati, D.L. et al. (2021). Video analysis of basketball throws for parabolic motion learning materials. In *Journal of Physics: Conference Series*, 1816(1): 012077. <https://doi.org/10.1088/1742-6596/1816/1/012077>