Research Article

Application of Artificial Intelligence and Big Data Technology in Basketball Sports Training

Si Zhong

College of Humanities, Zhaoqing Medical College, Zhaoqing, 526000 Guangdong, China

Correspondence should be addressed to Si Zhong; 201701390135@lzpcc.edu.cn

Received 5 June 2022; Revised 6 July 2022; Accepted 22 July 2022; Published 23 August 2022

Academic Editor: Kuruva Lakshmanana

Copyright © 2022 Si Zhong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Basketball is a sport with a wide range of complicated human movements, and the ability to accurately identify these movements is critical in both competition and training. Training athletes is largely based on the subjective observations and experiences of the coaches. Artificial intelligence and big data technologies can be used to monitor athlete training. It can also assist coaches in making decisions which significantly increase the athletic ability by detecting their movements. This research work proposed a mechanism called BSTARNet which is an action recognition method for basketball sports training. The method is based on artificial neural network (ANN), and network is trained through basketball sports big data. First, this work uses the Convolution Long Short-Term Memory (ConvLSTM) unit to extract the spatiotemporal information features of basketball sports training from videos. Second, this work establishes the Attention Long Short-Term Memory (AttLSTM) unit that combines the attention mechanism with the LSTM. The unit selectively scans each location, giving more attention to the area where the action takes place. Finally, the network framework is built by improving the ordinary encoder-decoder model. After that, the spatiotemporal information contained in the video is encoded based on the Darknet network model. In the decoding stage, the AttLSTM structure is used to replace the ordinary LSTM. These units are combined to form the BSTARNet architecture. Experiments are conducted to verify the effectiveness of the proposed method applied on action recognition in basketball sports training and achieved 89.5% mAP and 95.4% accuracy.

1. Introduction

In today’s relatively rich material life, sports are more and more popular among people and have become a vital part of daily life. Physical exercise helps to improve the body’s immunity, improves the function of the nervous system, can keep the athlete in good shape, and can keep their mood happy. Persistence in sports helps to promote the all-round development of the athlete. In addition, team sports can help people establish and maintain harmonious interpersonal relationships, improve their ability to communicate and coordinate, and cultivate participants’ awareness of teamwork, especially for teenagers. Beautiful sports movements have high viewing value and can make people’s lives more colorful. Watching sports events has become a common preference for sports enthusiasts around the world. In recent years, the popularity of sports such as basketball games on television (TV) screens and the Internet has fully reflected the popularity of sports [1–4].

If want to achieve your goals faster in sports, a professional training plan is indispensable. Since each athlete’s physique and starting level are different, the training plan needs to be specially customized for each athlete’s actual situation. As the hobby of sports continues to rise, the demand for professional training guidance is also rising, and scientific training plans need to be formulated by professional coaches. In the traditional situation, the coaches need to watch the performance of the athletes at the sports site and then formulate an appropriate training plan based on their rich experience. This solution has the following shortcomings. First, this situation consumes a lot of coaches’ time for watching the sports scene and squeezes the time for coaches to formulate training plans. Due to the limited number of professional coaches, not all sports enthusiasts can get...
training guidance from excellent coaches. Secondly, only through visual observation, coaches cannot obtain quantitative data information like acceleration and angular velocity; it is difficult to accurately grasp the deep-level information that is crucial for athletes [5–8].

A ball is tossed into an opponent’s basket in order to score or prevent the other team from collecting the ball and scoring according to particular rules in the game of basketball. Other ball games, such as soccer or volleyball, are less technical and tactical, but basketball is characterized by a high level of individual battle and synchronization. When it comes to a basketball game, each player’s degree of skill has a noticeable effect on the entire squad. Basketball teams suffer when their players lack the necessary basketball abilities, as their shortcomings are exposed and their levels of defense and attack suffer as a result. Athletes should receive rigorous, evidence-based basketball training as a result. Coaches traditionally develop training regimens for their teams based on the performance of their players in practice and competition. Coaches’ training theories and personal experiences inform this approach, which introduces an element of subjectivity. Athletes’ training quality is evaluated by comparing their performance to a variety of test standards, and coaches must manually calculate their training performance. There are however some disadvantages to using this approach. As a result, it is critical for the development of athletes’ competitive abilities and coaches’ decision-making abilities that sports data and motions of players can be precisely captured in real time [9–12].

AI has appeared in frontier fields. The development for artificial intelligence requires support of big data. Artificial intelligence is a method to guide the direction and collect laws in the world of big data. AI can process large datasets beyond traditional human methods. Machines can perform deep iterative learning through powerful logical computing capabilities to make predictions and decisions. At present, some researchers have applied artificial intelligence technology to the field of sports to identify movements in sports. Based on the recognition results of each movement obtained by using these research results, coaches can quickly understand the performance of multiple athletes even when they are not at the sports site. This improves the efficiency of coaches in formulating training plans, so that more sports enthusiasts can get training guidance from excellent coaches [13–15].

This research work develops an action recognition mechanism for basketball sports training named BSTARNet. This mechanism is based on artificial neural network where training of network is carried out through basketball sports big data. Spatiotemporal information features are extracted from basketball sports training videos as input data for the proposed system. The ConvLSTM structure is used to extract the spatiotemporal information from video which is then encoded using Darknet network model. In the decoding stage, the AttLSTM structure is used to replace the ordinary LSTM. The above units are combined into the BSTARNet architecture to accurately recognize the actions in basketball sports training.

2. Related Work

Wei et al. [16] propose an action recognition system that uses deterministic finite automata to recognize sports actions, which can be used to recognize the actions of athletes and referees in videos. Aoun et al. [17] and Bilen et al. [18] successively tried to use 2-dimensional (2D) convolution neural network (CNN) and 3D CNN to recognize actions in football game videos. The experimental result is that the recognition accuracy rate when using 3D CNN is higher than that when using 2D CNN. This shows that the features extracted with 3D CNN can better reflect the real situation of the original data than the features extracted with 2D CNN. Ullah et al. [19] use multiple cameras to cooperatively track players. This solution can solve the problem that players are occluded in partial views, which affects the recognition effect. Target linkages are captured in a state-space model [20], which makes use of model field particles that integrate appearance and motion models. Occupied players in football games can be tracked with this technology, which also works well for multitarget monitoring in other team sports characterized by similar physical characteristics and unexpected movements. Fani et al. [21] examine football movies’ structure, which includes the use of parallel feature-fusion networks to merge local and panoramic
characters to identify perspectives, and then provide an advanced Markov model to detect the playback status of shots. It is possible to identify transition effects using a threshold-based technique and Gaussian mixture models for event candidates [22]. Using an extreme learning machine classifier, you may identify event categories, mark critical events, and identify video playbacks in sports footage. In sports videos, event detection and video summary have been the center of user interest. According to Suzuki et al. [23], a deep extreme learning machine is used to assess age. In sports videos, event detection and video summary

3. Method

This work designs BSTARNet, an action recognition method for basketball sports training. The method is based on
artificial intelligence neural network and carries out network training through basketball sports big data, so as to realize high-performance basketball sports training action recognition.

3.1. Convolution Neural Network and Long Short-Term Memory. Convolution neural network (CNN) is a network that simulates the hierarchical cognition of the visual cortex and consists of convolutional layers, pooling layers, fully connected layers, and classifiers. When CNN receives the input information data, lower-level convolutional layer will first identify and extract the primary graphic features, and the abstract fusion of several underlying features will form the features of the high-level convolutional layer. Convolutional layer is important constituent unit for entire network. When convolution kernel performs the convolution operation on the image, it only convolves with a part of the area pixels, that is, the local connection. The advantage of this connection method over full connection is that it reduces feature redundancy and thus reduces parameters. However, each convolutional layer has several convolution kernels, and the weight parameters of each convolution kernel are the same, that is, the weights are shared. The pooling layer is to downsample the feature map obtained by the convolution layer, which is to make the feature map smaller according to the specified rules but the depth does not change. Commonly used pooling operations include maximum pooling and mean pooling, which are used to reduce the amount of training parameters and improved the training and learning speed. The fully connected layer is to convert the feature map into the input required by the classifier and then perform classification and recognition to obtain the output result.

Long Short-Term Memory (LSTM) successfully solves the defects of traditional recurrent neural network. LSTM improves the internal structure of traditional recurrent neural network (RNN) and uses a series of gate structures to control the state of the unit, including input gate and forget gate as well as output gate. Input gate is utilized to control how much information of the previous unit needs to be written into the unit state, the forget gate is to selectively forget or discard some information, and the output gate is to decide how much information needs to be saved to the next hidden state. To put it another way, this is like having three control switches, one for long-term preservation, one for immediate state input, and one for whether to use long-term state as the output of current LSTM. Figure 1 is a diagram of the internal structure of LSTM.

3.2. ConvLSTM Unit. Because of its unique internal structure, LSTM has the functions of association and memory, and it is good at dealing with timing problems and spatial information. However, the internal connection method of LSTM is fully connected. For 3D graphics data, spatial information is rich and each pixel has a strong correlation with

| Layer   | Number | Size   |
|---------|--------|--------|
| Conv    | 32     | 3×3    |
| Maxpool | -      | 2×2    |
| Conv    | 64     | 3×3    |
| Maxpool | -      | 2×2    |
| Conv    | 128    | 3×3    |
| Conv    | 64     | 1×1    |
| Conv    | 128    | 3×3    |
| Maxpool | -      | 2×2    |
| Conv    | 256    | 3×3    |
| Conv    | 128    | 1×1    |
| Conv    | 128    | 3×3    |
| Maxpool | -      | 2×2    |
| Conv    | 512    | 3×3    |
| Conv    | 256    | 1×1    |
| Conv    | 512    | 3×3    |
| Maxpool | -      | 2×2    |
| Conv    | 1024   | 3×3    |
| Conv    | 512    | 1×1    |
| Conv    | 1024   | 3×3    |
| Conv    | 512    | 1×1    |
| Conv    | 1024   | 3×3    |
| FC      | 1000   | 1×1    |
| Avgpool | -      | Global |

Figure 4: Proposed system structure diagram.
surrounding pixels. If you continue to extract features with FC-LSTM, it will cause unnecessary redundancy. Borrowing the local connection idea of the convolutional neural network, a new computing unit ConvLSTM combining convolution and LSTM is proposed. The specific improvement is to replace the point multiplication operation in FC-LSTM with a convolution operation.

\[
 f_t = \sigma \left( w_{xf} x_t + w_{hf} h_{t-1} + w_{cf} c_{t-1} + b_f \right),
\]

\[
 i_t = \sigma \left( w_{xi} x_t + w_{hi} h_{t-1} + w_{ci} c_{t-1} + b_i \right),
\]

\[
 o_t = \sigma \left( w_{xo} x_t + w_{ho} h_{t-1} + w_{co} c_{t-1} + b_o \right),
\]

\[
 \tilde{c}_t = \tanh \left( w_{xc} x_t + w_{hc} h_{t-1} + b_c \right),
\]

\[
 c_t = f_t c_{t-1} + i_t \tilde{c}_t,
\]

\[
 h_t = o_t \tanh \left( c_t \right).
\]

Figure 2 is a diagram of the internal structure of ConvLSTM, which is similar to LSTM unit. The difference between the two structures can be clearly found. The internal structure of ConvLSTM only adds a convolution operation between input to state and state to state, and other structures remain unchanged. When receiving an externally provided image sequence input, ConvLSTM performs feature extraction on image information in both spatial and temporal dimensions. The spatial dimension is to extract the spatial position information of each frame of the sequence, and the time dimension is to extract the time series information between the frames before and after. This double extraction method ensures the integrity of the spatiotemporal features of the video sequence, and the features that are not separated in time as well as space can improve recognition performance of the system.

3.3. AttLSTM Unit. The implementation of the deep learning attention mechanism is to retain the intermediate output results of encoder for input sequence and then assign different attention weights to different locations at the current moment. These weights are then weighted and summed with the intermediate output results, and the obtained result is used as the input of the LSTM at the next moment. This process is called a cycle. The number of loops depends on frame length of video sequence. The weight vector obtained in the last loop is weighted and summed with the output results obtained at the previous moment, and the obtained results are directly input into the classifier. In this paper, the unit composed of the attention mechanism and LSTM is called AttLSTM, which is composed of several attention blocks. Attention block is a new module that improves its internal structure operation based on LSTM network, that is, adding attention calculation.

\[
 e_t = \text{attention} \left( h_{t-1}, d_t \right),
\]

\[
 \alpha_t = \text{softmax} \left( e_t \right),
\]

\[
 \tilde{d}_t = \sum_{i=1}^{K} \alpha_t^i \tilde{d}_t^i.
\]

Figure 3 demonstrates the specific implementation process of the deep learning attention mechanism. The dotted box represents the calculation flow chart of the attention block. As can be seen from the figure, the context vector obtained after the encoder and the hidden state of the previous moment are used as input. Where attention represents a calculation method, and the result of the operation is normalized by softmax to obtain the attention weight vector.

3.4. BSTARNNet. Most algorithms use the common encoder-decoder framework. An input sequence is learned using LSTM units, encoded into a vector representation and decoded using an LSTM unit, in the typical LSTM model. The performance of this model is not sufficient when the input sequence is long. For the most part, this is because, during the encoding process, the structure can only represent the input as a fixed-length vector. Once the sequence is too long, the previous information will be forgotten and lost, so in this paper, when the action recognition of basketball sports training video is performed, the input sequence of the system is long or short. Therefore, by improving the traditional encoder-decoder structure, the AttLSTM unit is added in the decoding stage. When processing the vector obtained in the encoding stage, this unit will deemphasize the heavy to avoid discarding important information in the decoding stage. In fact, this unit can also be added in the encoding stage at the same time. The function is to assign different weights to different location regions when extracting features. However, considering the calculation time and algorithm complexity, this unit is only added in the decoding stage.

| Name               | Setting               |
|--------------------|-----------------------|
| CPU                | Intel Core i7         |
| GPU                | GTX 1080Ti            |
| Operating system   | Ubuntu16.04/cudnn/cuda |
| Development environment | PyTorch/Python    |
In this work, the ConvLSTM unit is selected to replace the ordinary LSTM unit in the encoder part, which can extract spatial information features and time series features at the same time. In the decoder part, the AttLSTM unit is used to replace the ordinary LSTM unit, which can assign attention to the content vector obtained in the encoding stage and assign different weights to different positions. Then, weighted summation with the original encoding vector, after several learning updates, the feature vector used for classification decision is obtained, and the final classification decision is made by the softmax classifier. Figure 4 is the system structure diagram of this paper.

This work uses Darknet as the encoder network; it combines the advantages of VGGNet and NIN structure. When this structure was proposed, it was applied in the field of target recognition. Compared with several other classical target recognition algorithms, it showed a very superior recognition performance. Therefore, this paper attempts to apply this network to video human action recognition. This network uses a convolutional kernel of size $3 \times 3$, and a $1 \times 1$ convolutional layer is added between the two convolutional layers. After this is a maximum pooling layer, global pooling is used to compress feature map into a $1 \times 1$ pixel value, which is used as the input of the classifier for decision classification. Table 1 is the specific structure of the Darknet network model.

The ConvLSTM structure has dual depth in time and space, which can effectively learn abstract representations of spatiotemporal information simultaneously. It is the combination of CNN and LSTM, that is, the convolution operation is integrated into LSTM. This unit is used in the same way as CNN, by stacking ConvLSTM units to build a deep neural network to extract video spatiotemporal features. In this paper, the ordinary convolution layer in the Darknet model is replaced by ConvLSTM, and size of convolution kernel is still $3 \times 3$ and $1 \times 1$. The $1 \times 1$ ConvLSTM layer is used for feature cross-channel fusion to reduce feature redundancy. Global mean pooling is used to compress the feature map size to $1 \times 1$, which is convenient for the network to make the final decision classification. The original Darknet model adds a fully connected layer after global pooling. This paper only uses this network model in the encoder, so this layer is not needed. The number of filters is appropriately adjusted according to the actual research in this paper.

The network model is used in the encoder only. On this basis, this work combines the AttLSTM unit to build a behavior recognition network BSTARNet for basketball sports training. The pipeline of BSTARNet is illustrated in Figure 5.

The network input is a continuous image sequence extracted at equal intervals, and the ConvLSTM layer and maxpooling layer stacked behind are connected according to the improved Darknet network model. The last layer of AttLSTM is used for the decoder part, and it also replaces the fully connected layer of the original Darknet, because the vector calculated by the decoding of AttLSTM can be directly applied to the softmax classifier.

4. Experiment and Analysis

4.1. Dataset. This work crawls basketball sports training videos from websites with big data technology, with six different training movements: shooting, catching, passing, dribbling, jumping, and running. Each type of action is completed by 100 different basketball training personnel in different scenes and contains a total of 1086 videos. This work selects 677 videos as the training set, and the remaining 409 videos as the test set. The evaluation metrics used are accuracy and mean Average Precision (mAP). The experimental environment is illustrated in Table 2.

4.2. Comparison Results with Other Methods. To confirm the validity of the proposed method, quantitative comparison is
performed with state of the art methods. The compared methods include LRCN [26], ALSTM [33], VideoLSTM [34], and CHAM [35]; the experimental results are illustrated in Figure 6.

AI has extensively been used for different problems and its evaluation in different filed, such as [36–38].

BSTARNet proposed in this work can achieve 89.5% mAP and 95.4% accuracy, which can be improved to varying degrees compared with other methods, which proves the superiority of this method.

4.3. ConvLSTM Experiment Result. This work combines convolution and LSTM, thereby designing the ConvLSTM unit. To verify effectiveness of this unit, performances without ConvLSTM and with ConvLSTM are compared, respectively, and the results are illustrated in Figure 7.

After using the ConvLSTM unit, the two performance indicators can be improved by 1.3% and 1.1%, respectively, which proves the feasibility of the ConvLSTM unit.

4.4. AttLSTM Experiment Result. This work embeds the attention mechanism into the LSTM to construct the AttLSTM unit. To verify effectiveness of this unit, performances without the AttLSTM unit and when the AttLSTM unit is used are compared, respectively, and the experimental results are illustrated in Figure 8.
After using the AttLSTM unit, the two performance indicators can be improved by 1.7% and 1.5%, respectively, which proves the feasibility of the AttLSTM unit.

4.5. Darknet Experiment Result. This work uses Darknet as the base network of the encoder, and in order to evaluate the feasibility of this network, it is compared with the VGG network and the INI network. The experimental results are illustrated in Figure 9.

As can be seen from the data in the figure, using the Darknet network as the basic network of the encoder can obtain the highest mAP and accuracy, which verifies the correctness of the choice in this work.

4.6. Video Frame Experiment Result. This work takes basketball sports training videos as input. To verify the impact of different video frames on performance, mAP and accuracy under different frame numbers are compared, respectively. The experimental results are illustrated in Figure 10.

As the number of frames increases, mAP and accuracy first rise to the maximum value and then decrease. When the frames is set to 10, the best performance is obtained.
The method is based on artificial intelligence neural network computing and carries out network training through basketball sports training performance simulation based on improved machine learning, "International Journal of Human Movement and Sports Sciences", vol. 14, no. 15, pp. 1–24, 2021.

In basketball training, the analysis of the players can visually display the performance of the players for the coach, formulate corresponding training plans, and improve the team’s game performance. With the application of video analysis technology, the movements of basketball players in training can be analyzed, and the human action recognition technology can be introduced to provide players with accurate action analysis results. This work designs BSTARNet, an action recognition method for basketball sports training. The method is based on artificial intelligence neural network and carries out network training through basketball sports big data, so as to realize high-performance basketball sports training action recognition. Considering that the research object is basketball training videos, it is proposed that the extraction of spatial location features and time series information features are inseparable. The ConvLSTM unit combined with convolution loops is used to extract spatiotemporal features at the same time in the encoding stage. ConvLSTM can not only extract spatial location features but also retain the associative memory function of LSTM in time series problems. In the decoding stage, the AttLSTM unit combining the attention mechanism and LSTM is used to redistribute the weight of the feature vector according to certain rules, which enhances the relevance of time series information. The action recognition in this paper adopts the Darknet network model, and then considering that the research object of this paper is video sequences, the structure is improved, and the BSTARNet network model is proposed by using ConvLSTM instead of Conv to extract spatiotemporal features. Comprehensive experiments verify the effectiveness of this method in action recognition for basketball sports training.

### Data Availability

The datasets used during the current study are available from the corresponding author on reasonable request.

### Conflicts of Interest

The author declares that he has no conflict of interest.

### References

1. J. Luo, W. Gao, and Z. L. Wang, “The triboelectric nanogenerator as an innovative technology toward intelligent sports,” *Advances in Materials*, vol. 33, no. 17, article 2004178, 2021.
2. R. Summerley, “The development of sports: a comparative analysis of the early institutionalization of traditional sports and E-sports,” *Games and Culture*, vol. 15, no. 1, pp. 51–72, 2020.
3. Z. Gao, H. Xue, and S. Wan, “Multiple discrimination and pairwise CNN for view-based 3D object retrieval,” *Neural Networks*, vol. 125, pp. 290–302, 2020.
4. S. Ding, S. Qu, Y. Xi, and S. Wan, “Stimulus-driven and concept-driven analysis for image caption generation,” *Neurocomputing*, vol. 398, pp. 520–530, 2020.
5. S. W. West, J. Clubb, L. Torres-Ronda et al., “More than a metric: how training load is used in elite sport for athlete management,” *International Journal of Sports Medicine*, vol. 42, no. 4, pp. 300–306, 2021.
6. D. A. Dobrosielski, L. Sweeney, and P. J. Lismann, “The association between poor sleep and the incidence of sport and physical training-related injuries in adult athletic populations: a systematic review,” *Sports Medicine*, vol. 51, no. 4, pp. 777–793, 2021.
7. T. H. Liu, W. H. Chen, Y. Shih, Y. C. Lin, C. Yu, and T. Y. Shiang, “Better position for the wearable sensor to monitor badminton sport training loads,” *Sports Biomechanics*, vol. 1, pp. 1–13, 2021.
8. N. Brown, C. J. Knight, and L. J. Forrest, “Elite female athletes’ experiences and perceptions of the menstrual cycle on training and sport performance,” *Scandinavian Journal of Medicine & Science in Sports*, vol. 31, no. 1, pp. 52–69, 2021.
9. P. Kamarasaks and D. Conte, “Changes in salivary markers during basketball long-term and short-term training periods: a systematic review,” *Biology of Sport*, vol. 39, no. 3, pp. 673–693, 2022.
10. F. D’Elia, D. D’Andrea, G. Esposito, G. Altavilla, and G. Raisia, “Increase the performance level of young basketball players through the use of high intensity interval training,” *International Journal of Human Movement and Sports Sciences*, vol. 9, no. 3, pp. 445–450, 2021.
11. R. Ji, “Retracted article: air pollution detection in plain area based on web server and visualization of basketball training,” *Arabian Journal of Geosciences*, vol. 14, no. 15, pp. 1–14, 2021.
12. W. Pengyu and G. Wanza, “Image detection and basketball training performance simulation based on improved machine learning,” *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 2493–2504, 2021.
13. P. Pareek and A. Thakkar, “A survey on video-based human action recognition: recent updates, datasets, challenges, and applications,” *Artificial Intelligence Review*, vol. 54, no. 3, pp. 2259–2322, 2021.
14. M. Ullah, M. Mudassar Yamin, A. Mohammed, S. Daud Khan, H. Ullah, and F. Alaya Cheikh, “Attention-based LSTM network for action recognition in sports,” *Electronic Imaging*, vol. 33, no. 6, pp. 302–1–302–6, 2021.
[15] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, and S. W. Baik, "Action recognition in video sequences using deep bidirectional LSTM with CNN features," *IEEE Access*, vol. 6, pp. 1155–1166, 2018.

[16] S. Baysal and P. Duygulu, "Sentioscope: a soccer player tracking system using model field particles," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 7, pp. 1350–1362, 2016.

[17] M. Fani, M. Yazdi, D. A. Clausi, and A. Wong, "Soccer video structure analysis by parallel feature fusion network and hidden-to-observable transferring Markov model," *IEEE Access*, vol. 5, pp. 27322–27336, 2017.

[18] A. Javed, A. Irtaza, Y. Khaliq, H. Malik, and M. T. Mahmood, "Replay and key-events detection for sports video summarization using confined elliptical local ternary patterns and extreme learning machine," *Applied Intelligence*, vol. 49, no. 8, pp. 2899–2917, 2019.

[19] G. Suzuki, S. Takahashi, T. Ogawa, and M. Haseyama, "Team tactics estimation in soccer videos based on a deep extreme learning machine and characteristics of the tactics," *IEEE Access*, vol. 7, pp. 153238–153248, 2019.

[20] J. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in *IEEE International Conference on Computer Vision*, pp. 4489–4497, 2015.

[21] J. Carreira and A. Zisserman, "Quo vadis, action recognition? A new model and the kinetics dataset," in *International Conference on Computer Vision and Pattern Recognition*, pp. 6299–6308, 2017.

[22] S. Branson, G. Van Horn, S. Belongie, and P. Perona, "Bird species categorization using pose normalized deep convolutional nets," 2014, http://arxiv.org/abs/1406.2952.

[23] J. Fu, H. Zheng, and T. Mei, "Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4438–4446, Hawaii, 2017.