Hindawi
Discrete Dynamics in Nature and Society
Volume 2024, Article ID 9786837, 1 page
https://doi.org/10.1155/2024/9786837

Retraction

Retracted: Deep Learning-Based Detection and Identification Method for Sports Health Video Dissemination

Discrete Dynamics in Nature and Society

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Pang, "Deep Learning-Based Detection and Identification Method for Sports Health Video Dissemination," *Discrete Dynamics in Nature and Society*, vol. 2022, Article ID 1628165, 9 pages, 2022.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2022, Article ID 1628165, 9 pages
https://doi.org/10.1155/2022/1628165

Research Article

Deep Learning-Based Detection and Identification Method for Sports Health Video Dissemination

Yajun Pang [1,2]

[1]*College of Physical Education, Luoyang Institute of Science and Technology, Luoyang, Henan 471023, China*
[2]*Henan Province Engineering Research Center of Industrial Intelligent Vision, Luoyang, Henan 471023, China*

Correspondence should be addressed to Yajun Pang; yajun.pang@lit.edu.cn

Received 17 May 2022; Revised 12 June 2022; Accepted 16 June 2022; Published 5 July 2022

Academic Editor: Zaoli Yang

Sports health is gradually attracting attention, and computer vision technology is integrated into sports health to improve the quality of sports and increase the motivation of athletes. A deep learning sports health video propagation detection and recognition system is built through the mode of video propagation to provide real-time training information for sports and scientific body index parameters and exercise data for sports health programs. An athletics action estimation network (AAEN) is promoted, which initially obtains the correlation features and depth features between human skeleton key points through partial perception units. Then, all the joint point features are classified and correlated based on the affinity field range through the confidence map of the human skeletal node region. All video frames are then fused with similar joint features at the temporal level to extract motion key points in the time scale, and human posture prediction is achieved by fitting between the motion features and the dynamic database. To show the high efficiency of our method, we select three main databases for validation, and the results prove that AAEN outperforms by 13.96%, 16.90%, and 15.10% in precision, *F*1 score, and recall compared to the SOTA in sports health video detection and recognition. Our method also performs better overall in the same type of algorithms.

1. Introduction

With the improvement of living standards, physical health is gradually gaining attention. Most people value physical health because they want to keep their body in balance on the dietary and exercise levels. Proper physical activity promotes body metabolism, while physical health indicators can be used as a guide for assessing a person's physical health status. Based on the different parameters of the indicators, the specialist can determine the cause of the patient's illness and suggest a detailed set of sports rehabilitation tables based on rehabilitation science. For professional athletes, sports health can test the physical limits of the athlete and prevent the athlete from developing organismic injuries during the training process [1–3]. Sports health contains a variety of physical assessment parameters, such as real-time heart rate, respiratory rate, blood pressure, muscle stretch, and bone tolerance, based on which the athlete can receive more scientific training recommendations [4, 5].

Physical fitness is the most fundamental code in the field of sports. The effectiveness of training in sports will be directly proportional to the results of the assessment of training intensity and sports health. Physical training programs can have a huge impact on the physiological performance of athletes. According to our preliminary research, regarding athletes with more thorough physical fitness programs, the actual performance of athletes with better physical fitness programs is better than that of other athletes. Targeted strength training is also essential, depending on the sport. The effects of sports health planning are reflected in athlete field technique, mental fitness, physical strength, and tactical planning. In addition to prioritizing a prudent training process, good coaches will consider adding supplementary training facilities, such as computer-aided training systems and computer vision systems, to the training process [6, 7]. Only by incorporating real-time training data into the training program can a sustainable physical fitness program be scientifically developed.

The key performance of the sport health assessment is reflected in the initial training period, special period, competition period, peak period, and rest period. In the initial training period, the physical fitness of the athletes is the main concern. Through the indicators of training ability, training intensity, and cycle analysis, the basic fitness of the athletes will be understood, and scientific training plans will be formulated for the subsequent phases. In the special period, the training capacity and training intensity will be appropriately increased based on the data from the previous phase to complete the transition from adaptive training to competition-related intensity training. The most important part of this period is the issue of improving physical fitness in sports health planning of athletes. The main purpose of the competition period is to improve technical difficulty and learn competition tactics [8, 9]. The competition period usually lasts about 15 weeks and is carried out in phases. This phase of training can be effective in improving competition performance. The peak period usually lasts about one month and is the precompetition training phase. The intensity of training in this phase is gradually reduced to reserve strength and energy for the actual competition. This phase usually employs the taper rule, using a combination of 8/25/45 taper and program retention features, using fitness fatigue training to maximize recovery and maintain peak. Rest periods are postcompetition rest periods where training intensity and volume are reduced to one-third of the peak. The aim is to recover sufficiently to prepare the body for the next phase of training.

With the growth of people's demand for sports health, the combination of sports health industry and video has become the best model. It is difficult for the sports health planning model to meet the motivation and participation of people in sports. It makes athletes unable to grasp the essentials of sports quickly, resulting in low sports results. With the upgrade of computer vision technology, posture estimation techniques can be applied to track and field training [10]. The development of these technologies is often accompanied by video communication techniques that directly contribute to the athletic effectiveness of sports health and improve the understanding and interaction of the public with sports. The introduction of the sports health management system under the video communication system can achieve a win-win situation in sports quality and sports popularity [11]. Understanding the relationship between different sports and health is a prerequisite for health planning. There are many movements in sports health assessment, including athletes who need to understand the focus of technical movements. Therefore, deep learning methods under video dissemination system can learn sports action features and provide people with detailed action guidance and suggestions in sports health system.

The rest of the paper is organized as follows. Section 2 presents work related to different human pose recognition methods. Section 3 introduces the implementation process of detection and recognition of sports health video propagation paths based on deep learning network. Section 4 presents the experimental dataset and the analysis of the experimental data, and Section 5 summarizes the full paper, analyzing the shortcomings of the study and indicating future research directions.

2. Related Work

The most commonly used deep learning model for sports health video dissemination is the human pose prediction model. The most important aspect of human posture research is the estimation of spatial coordinates of joint points, yet this aspect is greatly influenced by the appearance of the human body [12]. Sports health video detection and recognition technology can provide real-time visual display and complete exercise data for sports. With this as a reference, doctors can tailor a scientific exercise plan and health cycle arrangement for each athlete. The core algorithm of sports health video detection technology is human action recognition algorithm, which is divided into single-player action recognition and multiplayer action recognition according to the number of people faced. After a large number of researchers, experimental verification can be seen; the more people, the worse the video action detection effect, where the single person action recognition is the preferred algorithm for most action recognition industry because of its good model stability and high robustness. For multiperson action recognition, video action detection results are affected by unstructured factors [13]. Some researchers in multiperson action experiments have found that crowd occlusion, overlapping light streams, and dynamic dark scenes can cause poor recognition results. It generally occurs that some human bones cannot be captured, resulting in spatial features and temporal features that cannot be connate, and human behavioral features cannot be matched with the action database. Although the efficiency of multiperson action recognition is not high, but considering that our research faces a large number of people for video detection, therefore, we choose to optimize on the basis of multiperson action recognition to improve the efficiency of multiperson video detection.

Sports health video detection has different effects when faced with multiplayer detection using different detection methods. Researchers in the literature [14] have focused their research on multitask action detection around top-down approaches. The authors used a convolutional neural network as the base network to capture the outer contours of the human body through center-of-mass localization. In the case of multiple people, the outer contours share a feature extraction layer and different people correspond to different center-of-mass contours, then the number of people is determined based on the number of centers of mass, and different numbers of people are divided into different pose estimation units, each of which contains a human skeleton segmentation algorithm that automatically assigns human skeletons to the people whose centers of mass are determined. The temporal relationship between the human skeletal nodes of the premise of the action recognition algorithm, the authors determined by this method to deal with the problem of skeletal segmentation of multiple people, and the experiments proved that the overall efficiency of the method is better, but it is affected by the human detector.

Researchers in the literature [15] proposed the heat map gradient method to detect multiperson gestures, and the authors found in their experiments that different gesture actions have different ways of heat map representation, and threshold range restrictions can segment the heat maps of different actions into broad categories of actions, and different network classes output different heat maps, and different categories of actions can be obtained according to the mapping between heat maps and human actions. For control of the number of people, the authors used the convolutional bit-pose algorithm to estimate the pixel weights of different people in the video, and the matching between people and poses is accomplished by the weights and the size of the heat map area.

Other researchers used stacked pyramid networks to improve the perceptual domain of human skeletal joint points and fuse the characteristics of different body parts by cascading them to improve the extension of action categories [16, 17]. In the literature [18], to solve the problem of inconsistent scales of multiperson action characteristics, the authors proposed a feature Atlas preprocessing method to effectively resolve the differences between action characteristics. The researcher in the literature [19] proposed a combined skeletal key point planning method that can compensate for the difference between features and improve the accuracy of skeletal recognition. Some other researchers have tried to use a linear regression neural network algorithm to localize skeletal points before moving to a dynamic joint algorithm of skeletal points to predict action classes [20, 21]. Researchers in the literature [22] found in their experiments that a bottom-up video action detection method is more effective, and the authors started with joint coordinate points of multiple individuals, whose joint point vectors are not oriented in the same direction for different individuals, and used this as a criterion to categorize each individual's joint points as a way to complete skeletal point segmentation, followed by video action recognition. The researchers in the literature [23] incorporated a residual network [24] into the video action recognition network and used adaptive image constraints to perform linear regression on skeletal points, but this method requires high hardware conditions, which leads to high experimental costs. Considering the experimental cost and computational complexity, researchers in the literature [25] performed clustering analysis of human skeletal joint points by video single-frame pixel embedding, and the clustering results at different levels mapped different video action classes.

In addressing the efficiency of video multiplayer action recognition, some researchers have tried to start with tracking algorithms that mimic the pixel tracking principle to achieve dynamic tracking of skeletal points as a way to analyze their behavioral action categories. Researchers in the literature [26], inspired by the flower pollination algorithm, set adaptive search windows by using the human center of mass of each frame of the video as a tracking point. This method improves the video detection accuracy of the human skeleton and achieves the target tracking of actions at the temporal level. The researchers in the literature [27] proposed an energy optimization strategy, and to reduce the influence of the experimental environment on the experimental results, the authors built a linear invariant system, and the experimental results showed that the method achieved 87% of the video action recognition accuracy. The researchers in the literature [28] found in a study of badminton escort robots that the derivative evolutionary algorithm can migrate learning to video action recognition and maintain dynamic skeletal point feature sharing at the temporal level, which facilitates the differentiation of differences between multiplayer actions.

3. Methods

3.1. Partial Perception. For video input, the human body is divided into skeletal nodes with each image frame of size $w \times h$. The deep neural network divides the body into different parts, each corresponding to a different joint confidence map H, and a body part affinity field (PAF) L, where L represents the number of joints. The number of frame rates of joint skeletal confidence maps $H = (H_1, \ldots, H_J)$ in video motion capture is J, where $H_j \in \mathbb{R}^{w \times h \times 2}$, $j \in \{1, \ldots, J\}$, H_j^{GT} represents the position of joint skeletal points in each frame of the video. Ifmmc1 is the video motion capture that starts from the overall level of human joints, and the labels of each joint part can be automatically generated by linear functions. For PAFs $L = (L_1, \ldots, L_c)$, C represents the total number of vectors of joint skeletal points, and different joints map different vector domains, where $L_c \in \mathbb{R}^{w \times h \times 2}, c \in \{1, \ldots, C\}$, L_c^{GT} represents the true unit vector of independent joints in the composition of skeletal points, each joint skeleton maps an independent unit vector j_1, j_2, the range of joints in each frame of the video consists of a combination of rectangular boxes, and the direction of vector j_1 within each rectangular box is consistent as j_2. Assuming a label of $y_{GT} = (H^{GT}, L^{GT})$, a model function of $P = (H, L)$, and an error parameter of $E_{L2}(P, y_{GT})$, the mathematical equations are expressed as follows:

$$E_{L2}(P, y_{GT}) = \sum_{j=1}^{J} \sum_{p} W(p) \| H_j(p) - H_j^{GT}(p) \|_2^2 + \sum_{c=1}^{C} \sum_{p} W(p) \| L_c(p) - L_c^{GT}(p) \|_2^2, \quad (1)$$

where P represents the pixel coordinates of joint skeletal points in each frame of the video, and W represents the dynamic mask of joint points, which is used to call function $W(p) = 0$ for optimizing video motion capture with

nonstructural factors in the case of abnormal experimental environment and multiple overlapping people. During the training process, the real features are easily deleted by mistake due to occlusion and ambient lighting. To solve this problem, we set an independent supervised self-loop f in each feature extraction layer to adaptively compensate gradient errors and prevent gradient explosion [29], which is mathematically expressed as follows:

$$f = \sum_{t=1}^{T_p} f_L^t + \sum_{t=T_p+1}^{T_p+T_c} f_S^t. \tag{2}$$

In order to reasonably match the skeletal joints with the corresponding parts of the body, a set of linear regression functions was used. In the vector domain generated by the skeletal joint points, the group of joint points is filtered according to the vector direction. The linear combination of the skeletal vectors with the corresponding body part domains was evaluated using confidence maps as the criteria. After the joints are matched with the body parts, the weight scores are then calculated, and the confidence weight values depend on the mapping rules between the skeletal points and the joint groups. The effect of the action of the joint affinity field is shown in Figure 1. When associating the dancer's left arm at the same time, different people will have different directions of affinity vectors for labeling.

3.2. Confidence Correlation. To explain the equation f mentioned in the previous section at a mathematical level, we label each skeletal point in the pixel coordinate system of the video frame and generate the corresponding joint confidence map S^*. The confidence map is generated from the joint group, so each confidence contains a direction vector information and a pixel coordinate information. The position changes of different skeletal points on the pixel coordinates can be converted into video motion capture information. In human motion detection experiments, each limb consists of a joint group and a skeletal point, and different combinations correspond to different peak ranges. Each person has a different peak range, so the peak range can be used to divide most people into independent individuals j. Each independent individual k is defined in the peak range or generates k confidence atlases $S_{j,k}^*$, where $x_{j,k} \in \mathbb{R}^2$ represents the real information of body part j of individual k in each frame of the video. Assuming $p \in \mathbb{R}^2$, then $S_{j,k}^*$ has the following mathematical expression.

$$S_{j,k}^*(p) = \exp\left(-\frac{\left\|p - x_{j,k}\right\|_2^2}{\sigma^2}\right), \tag{3}$$

where σ represents the joint combination peak, the video action recognition network starts with a single confidence map to resolve the action type of a single individual and then migrates to learn from multiple individuals, and the joint group confidence map range can filter the action category weights between different individuals.

$$S_j^*(p) = \max_k S_{j,k}^*(p). \tag{4}$$

The joint group confidence peaks in the video multiplayer hands-on recognition experiments limit the action trade-offs between individuals, and to maintain accuracy, we chose the mean value as the limiting condition. The confidence correlation process is shown in Figure 2, taking the athlete's hand confidence correlation as an example, and each hand appears with an infinite number of confidence correlation baselines. The association between the hands of the same person is the internal correlation, and the association with others is the external correlation. In the internal correlation, the correlation of the same hand is an internal positive correlation (shown as red dashed line), and the correlation between different hands is an internal negative correlation (shown as yellow dashed line).

3.3. Mapping and Association. In the first layer of the video multiplayer action recognition network, each frame of the video is characterized on-demand based on joint groups and skeletal points. Referring to the Google VGG network, we set VGG-19 as the base feature extraction network and select the joint group PAF $L^1 = \phi^1(F)$ in the skeletal feature initialization stage, where ϕ^1 denotes the joint features are denoted after the first stage of initialization. Each round of feature F is iteratively updated until the PAF logic criterion is satisfied before it can be output to the next action feature capture stage. The mathematical expressions are shown as follows:

$$L^t = \phi^t\left(F, L^{t-1}\right), \quad \forall 2 \le t \le T_p, \tag{5}$$

where ϕ^t represents the prediction result of the video action recognition network at stage t, and T_p indicates the number of iterations of the PAF. According to the internal logic of the PAF, the confidence map generated by each local joint iteration will be passed as a set of skeletal points, and the form of the pass is converted once after each T_p iterations.

$$\begin{aligned} S^{T_p} &= \rho^t\left(F, L^{T_p}\right), \quad \forall t = T_p, \\ S^t &= \rho^t\left(F, L^{T_p}, S^{t-1}\right), \quad \forall T_p \le t \le T_p + T_c, \end{aligned} \tag{6}$$

where ρ^t represents the action prediction result in the confidence iteration phase t , and T_c represents the number of skeletal confidence parameters for the joint group. Researchers in the literature [14] also refined the association of confidence maps with joint groups in their study of PAF, but the refinement led to halving of parameters and a significant reduction in the prediction of affinity fields. Therefore, in our network design, we used cluster analysis to adopt linear clustering for each joint group and bone, and different individual joint groups can be discriminated according to PAF channel logic. The experimental results demonstrate that our method is less effective in obtaining the connection between the confidence map and the action of the body as a whole, and for the extraction of fragmented features of body parts.

Figure 3 shows the effect of different iterative levels of PAFs between joint groups and bone points. As in the case of the athlete's leg node, the first stage is the refinement of the

Figure 1: Principle of joint affinity field action.

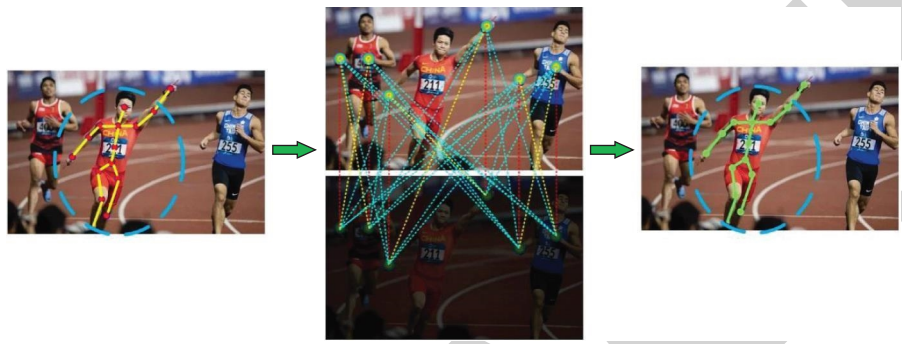

Figure 2: Confidence correlation process.

Figure 3: The effect of part affinity fields at different stages.

affinity field between the starting point and the endpoint. The second stage will connect the start point with the endpoint and refine the intermediate points. The third stage is the overall feature node refinement. The joint group PAF predictions of different individuals all share the same input of the network layer. To facilitate feature traversal of the joint group confidence map in the same individual, all network layers use the same parameter settings to prevent video frame pixel bias during traversal to properly guide the convolutional neural network to assign different features to different parts of the body during the iterative process. To prevent the confidence bias of the PAFs of the branch network and the backbone network, an additional L_2 loss function is added at the end of each video action recognition network. When matching real action features with labels, we added spatial loss functions in 3D for spatial location localization of individual mass centers. In addition, for the additional skeletal point features, we used data optimization and data padding to filter out the features with higher weights and then padding them into the real feature set.

3.4. Athletics Action Estimation Network. In order to meet the development needs of the sports and health industry, we propose a human action video recognition network, named Athletics Action Estimation Network (AAEN), with the network structure shown in Figure 4. The whole network of AAEN is divided into two stages, and the first stage of network iteration is to obtain the confidence features of the joint group confidence features ρ^t and match them with the individual skeletal affinity field ϕ^t. The network structure of the second stage is an optimization of the first stage, and the number of network iterations is adaptively selected according to time t. Each iteration updates the joint group confidence and skeletal affinity field, and when the parameter weights reach the specified values, then the output values are optimized in a continuous layer for feature optimization, where $t \in \{1, \ldots, T\}$ represents the supervised self-loop in the network layer.

We optimize based on the human pose estimation network proposed in the literature [30]. We design the network with more strata, and the number of strata is related to the time t, which is determined by the confidence of the joint group and the matching of the skeletal affinity field. When the weights reach the specified range, then the update iterations are stopped to output the action classification results. In the optimization of the network layer, we refer to the optimization strategy of Google VGG network and use conv 3×3 instead of conv 7×7 to increase the network width and improve the feature extraction limit of the

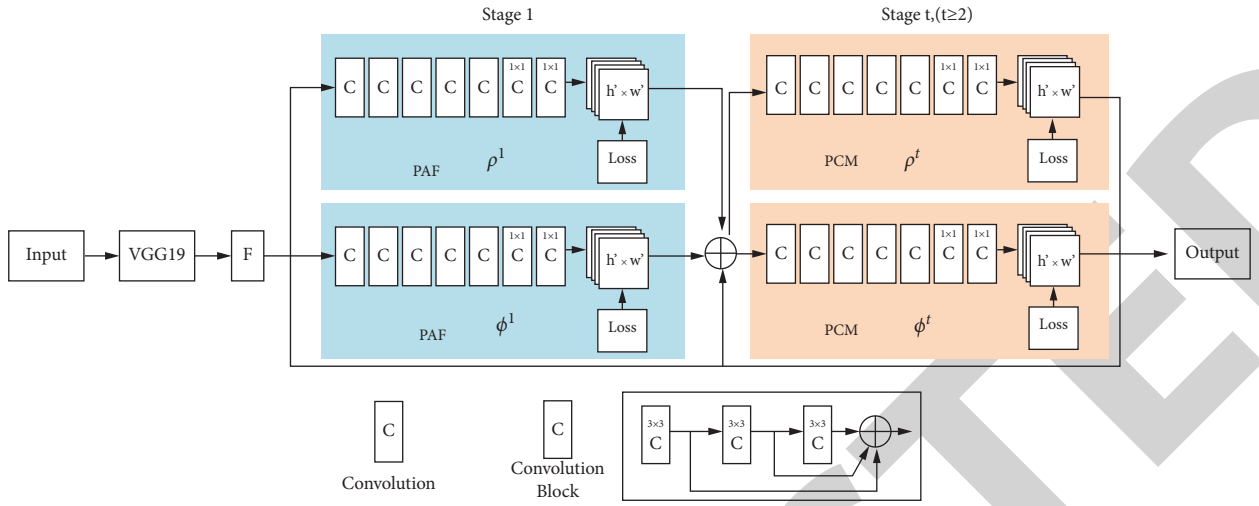

Figure 4: Athletics action estimation network.

network. We design the whole network as a sparse structure, and the initial and final layers of the network use residual connections for feature sharing, which solves the problem of parameter redundancy and overfitting. In addition, in the design of the network layer of the skeletal point affinity field, we add nonlinear units to extract high-level action features without affecting the extraction of low-level features by this network layer. Such a network design reduces the computational cost and improves the robustness of the video action recognition model.

4. Experiment

4.1. Dataset. To validate the performance of AAEN in sports video propagation detection, we performed experimental validation in the public datasets KTH [31], OSD [32], and UCF-Sports [33]. The KTH dataset was collected from 25 volunteers according to the actions specified. Each sample segment has a duration of 4 seconds and a pixel resolution of 160×120. The jogging and running datasets were chosen for the experiments to fit the research topic of sports health assessment. The OSD dataset is an Olympic sport-based dataset that contains all Olympic sports. The UCF-Sports dataset is a combination of sports videos from video websites and national channels, and this dataset is mainly oriented towards action recognition and human localization studies. We selected some data related to track and field sports for our experiments. The details of the three public datasets mentioned above are shown in Table 1.

4.2. Analysis of Results. We compared three methods, SVM [34], CNN [35], and OpenPose [36]. To ensure that the action recognition methods perform optimally independently in the experiments, we take a reset system approach for each method. In the method evaluation, we chose precision (P), $F1$ score, and recall (R) as the evaluation criteria. Each sports health detection metric is fed to each metric in the dataset, and the association between the metrics can indicate the performance of the sports health video propagation detection system and then the sports gesture prediction output with the help of motion joint feature classifier. Data labels can also be compared to react to the accuracy of human pose recognition. The performance of the sports health video propagation detection system in different public datasets is shown in Table 2.

The results in Table 2 indicate that the machine learning algorithms have poor performance in video recognition of human poses. Among the deep learning class of methods, CNN is the most widely used method, but video human pose recognition is not as accurate as the OpenPose algorithm. OpenPose's hierarchically interconnected network structure obtains better efficiency of action recognition, which allows for local perception and maximum fusion of memory information. Our approach AAEN adds an articulation group channel and a skeletal feature channel to the base model, starting from local fragmented information from the body to obtain bidirectional feature information. Therefore, the AAEN outperforms by 13.96%, 16.90%, and 15.10% in precision (P), $F1$ score, and recall (R) compared to the OpenPose algorithm.

Due to the low accuracy of SVM method in video human action recognition, it seriously affects the video human pose estimation in the second stage. We only keep the deep learning method as a comparative method for video human pose estimation. Before performing the video human action recognition work, according to the targeted experimental objectives of the three datasets, we will develop the optimization of the three datasets in order to be more adapted to the action recognition of sports health. Due to the large amount of data, we evaluated the datasets in terms of skeletal feature matching rate (SFMR), joint point matching rate (JPMR), and time node distribution rate (TNDR), and the results are shown in Table 3.

In Table 3, UCF-S maintains above 80% in both the skeletal feature matching rate and the node matching rate, and the recognition performance of the KTH is lower than that of the other two datasets in terms of the distribution rate

Table 1: Number of training sets and test sets.

	Datasets		
	KTH	OSD	UCF-S
Train	1332	1647	753
Test	232	315	106
Total	1564	1962	859

Table 2: Comparison of the accuracy of different methods of identification.

	KTH			OSD			UCF-S		
	P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
SVM	0.53	0.53	0.57	0.63	0.62	0.59	0.55	0.53	0.57
CNN	0.65	0.71	0.69	0.70	0.59	0.60	0.70	0.63	0.60
OpenPose	0.72	0.80	0.78	0.73	0.70	0.70	0.80	0.75	074
Ours	**0.86**	**0.91**	**0.81**	**0.83**	**0.80**	**0.82**	**0.87**	**0.92**	**0.92**

Table 3: Experimental results for different motion datasets.

	KTH	OSD	UCF-S
SFMR	0.72	0.75	0.84
JPMR	0.67	0.82	0.89
TNDR	0.43	0.66	0.57

Table 4: Comparison of the accuracy of human pose estimation in different sports events.

	1000 m race	Triple jump	Pole vault	Marathon
CNN	0.68	0.55	0.53	0.61
OpenPose	0.78	0.68	0.73	0.76
Ours	**0.95**	**0.86**	**0.91**	**0.94**

Figure 5: Results of human pose estimation based on AAEN for track and field events in the racing category.

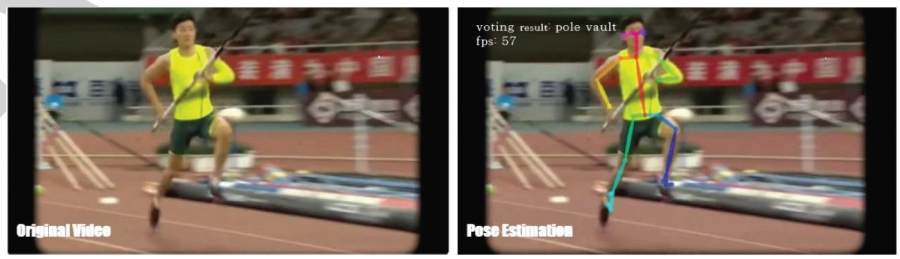

Figure 6: Results of human pose estimation based on AAEN for pole vault category track and field events.

of the temporal nodes. In order to maintain the balance between temporal and spatial features, we finally choose OSD as the validation dataset for video human pose estimation. We prioritized four representative sports, 1000-meter running, triple jump, pole vault, and marathon and compared the human action recognition video performance

of various methods. The results are shown in Table 4. And the results of our video body pose estimation method adopted AAEN in competition sports and pole vaulting-type sports are shown in Figures 5 and 6.

Table 4 shows that the video human pose estimation by CNN is not as effective as the OpenPose method, and our method is the most efficient, with an average accuracy of over 20 percentage points higher than the OpenPose algorithm. Therefore, human pose estimation is more suitable for running sports health video detection projects, such as 1000 m race and marathon. This is because running sports are simpler in terms of action characteristic dimension. In contrast, for triple jump events, the overall movement is more complex, the movement characteristics are more difficult to capture, and the prediction of pose at the temporal level is more difficult, so it is more effective than running sports.

5. Conclusion

In this paper, we investigate the details of the evaluation of sports health and find that the sports model is inefficient. To further improve the quality of sports and the motivation of athletes, we integrate deep learning human pose estimation algorithms into sports and build an integrated sports health video detection and recognition system that incorporates computer vision techniques and deep learning algorithms. An athletics action estimation network (AAEN) is promoted, which initially acquires position features and orientation features between key points of the human skeleton through partial perception units. Then, all nodal features are classified and correlated based on the affinity field range through the confidence map of the human skeletal node region. All video frames are then fused with similar joint features at the temporal level to extract motion features in the time scale, and video human action recognition is achieved by fitting between motion features and the dynamic database. We screened three main datasets for validation, and the results prove that our method is more efficient than machine learning methods. Our method also performs better overall in the same type of algorithms. To further validate the adaptability of our method to specific athletic events, we selected four sports, and the results prove that our method performs better in running-type sports.

Experiments from video propagation tests of sports events show that our method performs poorly in complex sports events. To solve this problem, in future research, we will consider starting from the mapping relationship between part and whole, borrowing the bidirectional loop structure of LSTM algorithm to highlight the fuzzy action features and weaken the noise features to achieve the effect of feature balance.

Data Availability

The dataset used to support the findings of the study can be accessed by contacting the author.

Conflicts of Interest

The author declares that they have no conflicts of interest.

Acknowledgments

This work was supported by the project aource: Science and Technology Projects of Henan Science and Technology Department. Name: Assessment of Tai-chi Action based on Vision Transformer (no. 222102320016).

References

[1] D. S. Lorenz, M. P. Reiman, and J. C. Walker, "Periodization," *Sport Health: A Multidisciplinary Approach*, vol. 2, no. 6, pp. 509–518, 2010.

[2] B. Macdonald, S. McAleer, S. Kelly, R. Chakraverty, M. Johnston, and N. Pollock, "Hamstring rehabilitation in elite track and field athletes: applying the British Athletics Muscle Injury Classification in clinical practice," *British Journal of Sports Medicine*, vol. 53, no. 23, pp. 1464–1473, 2019.

[3] A. Ek, J. Kowalski, and J. Jacobsson, "Training in spikes and number of training hours correlate to injury incidence in youth athletics (track and field): a prospective 52-week study," *Journal of Science and Medicine in Sport*, vol. 25, no. 2, pp. 122–128, 2022.

[4] P. Edouard, L. Navarro, J. Pruvost, P. Branco, and A. Junge, "In-competition injuries and performance success in combined events during major international athletics championships," *Journal of Science and Medicine in Sport*, vol. 24, no. 2, pp. 152–158, 2021.

[5] B H. DeWeese, G. Hornsby, M. Stone, and M H. Stone, "The training process: planning for strength-power training in track and field. Part 1: theoretical aspects," *Journal of sport and health science*, vol. 4, no. 4, pp. 308–317, 2015.

[6] B H. DeWeese, G. Hornsby, and M. Stone, M H Stone, "The training process: planning for strength-power training in track and field. Part 2: practical and applied aspects," *Journal of sport and health science*, vol. 4, no. 4, pp. 318–324, 2015.

[7] H. Woellik, A. Mueller, and J. Herriger, "Permanent RFID timing system in a track and field athletic stadium for training and analysing purposes," *Procedia Engineering*, vol. 72, pp. 202–207, 2014.

[8] P. Edouard, K. Hollander, L. Navarro et al., "Lower limb muscle injury location shift from posterior lower leg to hamstring muscles with increasing discipline-related running velocity in international athletics championships," *Journal of Science and Medicine in Sport*, vol. 24, no. 7, pp. 653–659, 2021.

[9] M. Liliana and P. Alina, "Optimization strategies theoretical training in competitive athletics," *Procedia - Social and Behavioral Sciences*, vol. 76, pp. 497–502, 2013.

[10] T. Timpka, J D. Périard, A. Spreco et al., "Health complaints and heat stress prevention strategies during taper as predictors of peaked athletic performance at the 2015 World Athletics Championship in hot conditions," *Journal of Science and Medicine in Sport*, vol. 23, no. 4, pp. 336–341, 2020.

[11] S. Li, B. Zhang, P. Fei, P M. Shakeel, and R D J. Samuel, "Computational efficient wearable sensor network health monitoring system for sports athletics using IoT," *Aggression and Violent Behavior*, Article ID 101541, 2020 in Press.

[12] S. Johnson and M. Everingham, "Clustered pose and nonlinear appearance models for human pose estimation," in *Proceedings of the BMVC*, vol. 2, no. 4, p. 5, September 2010.
[13] B. Sapp and B. Taskar, "Modec: multimodal decomposable models for human pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3674–3681, IEEE, Portland, OR, USA, June 2013.
[14] Z. Cao, T. Simon, S. E. Wei, and S. Yaser, "Realtime multiperson 2d pose estimation using part affinity fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291–7299, IEEE, Honolulu, HI, USA, July 2017.
[15] J. He, C. Zhang, X. He, and R. Dong, "Visual Recognition of traffic police gestures with convolutional pose machine and handcrafted features," *Neurocomputing*, vol. 390, pp. 248–259, 2020.
[16] G. Hua, L. Li, and S. Liu, "Multipath affinage stacked—hourglass networks for human pose estimation," *Frontiers of Computer Science*, vol. 14, no. 4, pp. 1–12, 2020.
[17] L. Zhu, F. Lee, J. Cai, H. Yu, and Q. Chen, "An improved feature pyramid network for object detection," *Neurocomputing*, vol. 483, pp. 127–139, 2022.
[18] M. Li, Z. Zhou, and X. Liu, "Multi-person pose estimation using bounding box constraint and LSTM," *IEEE Transactions on Multimedia*, vol. 21, no. 10, pp. 2653–2663, 2019.
[19] G. Papandreou, T. Zhu, and N. Kanazawa, "Towards Accurate Multi-Person Pose Estimation in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4903–4911, IEEE, Honolulu, HI, USA, July 2017.
[20] J. Xu, K. Tasaka, and M. Yamaguchi, "[Invited paper] fast and accurate whole-body pose estimation in the wild and its applications," *ITE Transactions on Media Technology and Applications*, vol. 9, no. 1, pp. 63–70, 2021.
[21] S. Ren, K. He, R. Girshick et al., "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, pp. 91–99, 2015.
[22] C. Wang, F. Zhang, and S. Ge, "A comprehensive survey on 2D multi-person pose estimation methods," *Engineering Applications of Artificial Intelligence*, vol. 102, Article ID 104260, 2021.
[23] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "DeeperCut: a deeper, stronger, and faster multi-person pose estimation model," in *Proceedings of the European Conference on Computer Vision*, Springer, Cham, Germany, pp. 34–50, September 2016.
[24] A. Martínez-González, M. Villamizar, and O. Canévet, "Efficient convolutional neural networks for depth-based multi-person pose estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 4207–4221, 2019.
[25] S. Liang, G. Chu, C. Xie, and J. Wang, "Joint relation based human pose estimation," *The Visual Computer*, vol. 38, no. 4, pp. 1369–1381, 2022.
[26] P. Ong, T. K. Chong, and K. M. Ong, "Tracking of moving athlete from video sequences using flower pollination algorithm," *The Visual Computer*, vol. 38, pp. 1–24, 2021.
[27] L. Van den Broeck, M. Diehl, and J. Swevers, "A model predictive control approach for time optimal point-to-point motion control," *Mechatronics*, vol. 21, no. 7, pp. 1203–1212, 2011.
[28] K. Rutten, J. De Baerdemaeker, J. Stoev, M. Witters, and B. De Ketelaere, "Constrained online optimization using evolutionary operation: a case study about energy-optimal robot control," *Quality and Reliability Engineering International*, vol. 31, no. 6, pp. 1079–1088, 2015.
[29] Q. Dang, J. Yin, B. Wang, and W. Zheng, "Deep learning based 2D human pose estimation: a survey," *Tsinghua Science and Technology*, vol. 24, no. 6, pp. 663–676, 2019.
[30] S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, "Imaging through glass diffusers using densely connected convolutional networks," *Optica*, vol. 5, no. 7, pp. 803–813, 2018.
[31] J. Ng and S. Gong, "Composite support vector machines for detection of faces across views and pose estimation [J]," *Image and Vision Computing*, vol. 20, no. 5-6, pp. 359–368, 2002.
[32] L. Kong, D. Huang, J. Qin, and W. Yunhong, "A joint framework for athlete tracking and action recognition in sports videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 2, pp. 532–548, 2019.
[33] J. Chen, R. D. J. Samuel, and P. Poovendran, "LSTM with bio inspired algorithm for action recognition in sports videos," *Image and Vision Computing*, vol. 112, Article ID 104214, 2021.
[34] G. Yao, T. Lei, and J. Zhong, "A review of Convolutional-Neural-Network-based action recognition," *Pattern Recognition Letters*, vol. 118, pp. 14–22, 2019.
[35] H. C. Shin, R. Roth H, M. Gao et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
[36] F. Angelini, Z. Fu, Y. Long, and S. Ling, "2D pose-based real-time human action recognition with occlusion-handling," *IEEE Transactions on Multimedia*, vol. 22, no. 6, pp. 1433–1446, 2019.