Journal of Combinatorial Mathematics and Combinatorial Computing, 118: 143–154
DOI:10.61091/jcmcc118-11
http://www.combinatorialpress.com/jcmcc
Received 10 September 2023, Accepted 20 December 2023, Published 29 December 2023

Combinatorial Press

*Article*

# Human Behavior Recognition Based on CNN-LSTM Hybrid and Multi-Sensing Feature Information Fusion

**Chaoyu Fan**[1,*]

[1] Department of Electrical Engineering, Columbia University, New York,10027, USA.

* **Correspondence:** cf2859@columbia.edu, jakefan179@163.com.

**Abstract:** To address the human activity recognition problem and its application in practical situations, a CNN-LSTM hybrid neural network model capable of automatically extracting sensor data features and memorizing temporal activity data is designed and improved by integrating CNN and gated recurrent units as a variant of RNN. A multi-channel spatiotemporal fusion network-based two-person interaction behavior recognition method is proposed for two-person skeletal sequential behavior recognition. Firstly, a viewpoint invariant feature extraction method is used to extract two-player skeleton features, then a two-layer cascaded spatiotemporal fusion network model is designed, and finally, a multi-channel spatiotemporal fusion network is used to learn multiple sets of two-player skeleton features separately to obtain multi-channel fusion features, and the fusion features are used to recognize the interaction behavior, and the weights are shared among the channels. Applying the algorithm in the paper to the UCF101 dataset for experiments, the accuracy of the two-person cross-object experiment can reach 96.42% and the accuracy of the cross-view experiment can reach 97.46%. The method in the paper shows better performance in two-player interaction behavior recognition compared to typical methods in this field.

**Keywords:** Two-player interaction behavior, CNN, LSTM, Spatiotemporal fusion network,
Multi-channel

## 1. Introduction

Human activity recognition refers to the classification of activities into known predefined human activity categories based on the temporal data obtained from sensors. With the current rapid development and massive application of wearable devices, the collected data such as human activity and vital signs are becoming more and more accurate, thus making the quality of capturing human activity information and the accuracy of HAR increasingly high [1]. The improvement of HAR accuracy is of great significance for health surveillance systems, remote healthcare, human-computer interaction, rehabilitation medicine and other fields [2]. In the field of intelligent transportation, the behavioral recognition technology can automatically identify traffic violations such as pedestrians/vehicles running red lights and unsafe driving by drivers to ensure people's travel safety; in the field of medical monitoring, the technology can achieve real-time monitoring of patients and accidental fall detection to ensure that patients can receive timely treatment and assistance; in the field of safety production, real-time monitoring of the whole process of production operations can be achieved, and the detection of accidents occurring in the process of operations and production can be achieved. In the field of

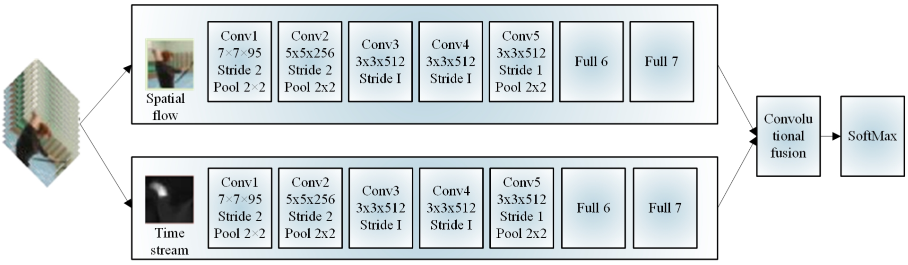
Spatial flow
Conv1 7×7×95 Stride 2 Pool 2×2
Conv2 5x5x256 Stride 2 Pool 2x2
Conv3 3x3x512 Stride I
Conv4 3x3x512 Stride I
Conv5 3x3x512 Stride 1 Pool 2x2
Full 6
Full 7
Time stream
Conv1 7×7×95 Stride 2 Pool 2×2
Conv2 5x5x256 Stride 2 Pool 2x2
Conv3 3x3x512 Stride I
Conv4 3x3x512 Stride I
Conv5 3x3x512 Stride 1 Pool 2x2
Full 6
Full 7
Convolu tional fusion
SoftMax

**Figure 1.** Dual-Stream Convolutional Neural Network Structure

safety production, real-time monitoring of the whole process of production operations can be realized, and timely alarms can be provided for actions that may lead to safety hazards in the process of operations and production, ensuring that operations and production are carried out within a safe and controllable range, and safeguarding the personal safety of personnel and property [3].

Although human activity recognition can also be performed through video capture, issues such as privacy protection, capturing blind spots and ethics make this form of information collection flawed and limited in its application. Ke et al. [4] mentioned that wearable health monitoring systems have the advantage of being non-invasive to the human body when performing real-time monitoring, and that HAR ensures the accuracy of the information collected and the safety of the subject while circumventing issues such as privacy protection. Qayyum et al. [5] summarized the application of wearable device data for chronic disease prevention and management, noting that effective data feedback can increase human activity, enhance patient health, improve disease prognosis, reduce healthcare costs and help clinical users make healthcare decisions; Acharya et al. [6] proposed an integrated sensor network, the Care Net and used for remote health care and healthcare; Tsai and Chen [7] summarized the application of wirelessly transmitted sensor networks in rehabilitation medicine, noting that sensor feedback can help in the immediate monitoring of patients undergoing rehabilitation training, and that this information is also of research value at the level of helping to correct rehabilitation postures.

Behavior recognition tasks based on video analysis require the creation of a library of action and pose samples and the training of the designed model to achieve the classification of behavior in video [8]. Traditional methods rely on manual extraction of features, and due to the small amount of data in the early sample library, simple scenes and single actions, traditional methods can meet certain needs [9]. However, with the popularity of video surveillance technology, the application scenes become more and more complex, and the video features extracted using traditional methods can no longer meet the actual needs in terms of recognition accuracy, making it difficult to make full use of the actual value of video surveillance [10].

In this paper, we propose a two-person interaction behavior recognition method based on multi-channel spatiotemporal fusion network for two-person skeleton sequence behavior. Neural networks that have undergone continuous development have unique advantages in solving HAR tasks [11]. The advantages of CNN and RNN in neural networks are considered, and a hybrid CNN-LSTM model is designed and improved in order to automatically extract features from sensor data and incorporate consideration of temporal dependency issues to achieve recognition of basic human states and common rehabilitation and health care activities [12].

## 2. Preliminary

### *2.1. Two-Stream CNN*

Dual-stream CNN has good recognition effect in the field of human body recognition [13]. Its network architecture is shown in Figure 1.

The dual-stream architecture is one of the current benchmarks in the field of human behavior recognition, and scholars at home and abroad have further explored the dual-stream architecture on

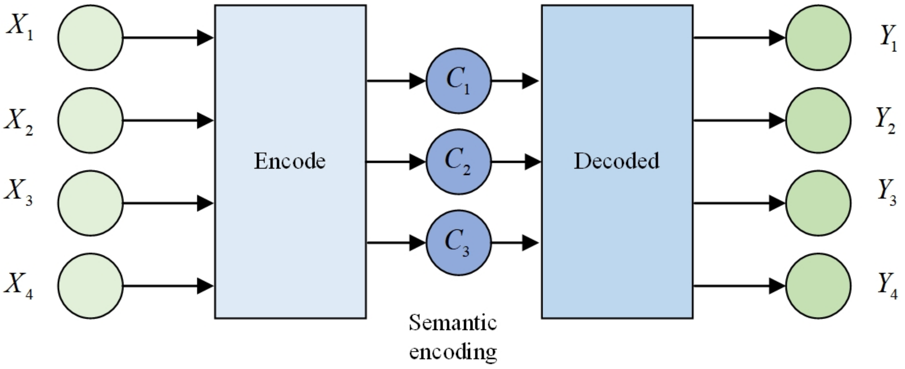

**Figure 2.** Encoder-Decoder Framework

its basis [14]. The earliest dual-stream convolutional model is based on the VGG-16 convolutional network, and adds residual blocks to the network for the temporal and spatial channels respectively to enhance the network's ability to extract temporal and spatial features, and finally fuses the feature information of the two channels [15].

On the other hand, the deep network model is optimized by adding an attention module to the residual network, and most attention models today are based on the Encoder-Decoder framework, which can be understood as the process of first transforming a given sequence X into a fixed-length vector by encoding it and then decoding it into a target output sequence *Y*. The Encoder-Decoder framework is shown in Figure 2.

The Encoder-Decoder framework was proposed to lay the foundation for building network models that can selectively extract feature information.

Through the efforts of many scholars, research based on dual-stream CNN models has become more and more mature, and many achievements have been made in the field of human behavior recognition [16]. Although the dual-stream network can well combine the static and dynamic feature information of human behavior and has the characteristics of strong stability and high recognition accuracy, it is undeniable that its high performance is based on the training of a large number of data samples, and in practical applications, many scenarios are due to the inability to collect enough sample information for training, which will make the dual-stream CNN appear in the training process overfitting. This also leads to the problem that in practice it cannot achieve the theoretical recognition accuracy [17].

### 2.2. *Spatiotemporal Fusion Feature Learning*

The spatiotemporal feature fusion network uses a two-layer cascade structure, with the first layer learning spatial features, the second layer learning temporal features, and finally outputting spatiotemporal fusion features.

The spatial feature learning network layer is shown in Figure 3, which learns the spatial relationship features of the skeleton at $t, t \in (1, ..., T)$ time. To maintain the temporality of the sequence, a $M$ one-dimensional CNN filter $\omega$ is used to filter the sequence $F$ of length $T$and dimension $N$, and a one-dimensional maximum pooling layer extracts the maximum features in the neigh bour hood to obtain a $M$ feature map $f_m$ of scale $(T, P)$, see Eq. (1):

$$f_m = \sigma\left(\omega_{(1,l)} * F + b\right), m = (1, 2, ..., M). \tag{1}$$

The bi-directional long and short term memory network (Bi-LSTM) is used to learn the correlation of the nodes in the feature map $f_m$ space. The LSTM network is defined as shown in (2) and contains input gate $i_t$, output gate $o_t$, forgetting gate $f_t$ and memory gate $c_t$, which can avoid the gradient disappearance caused by RNN [18]. The Bi-LSTM network consists of a combination of a forward LSTM and a backward LSTM. Point $t, t \in (1, ..., P)$, the values on feature map $(1, ..., M)$ form a

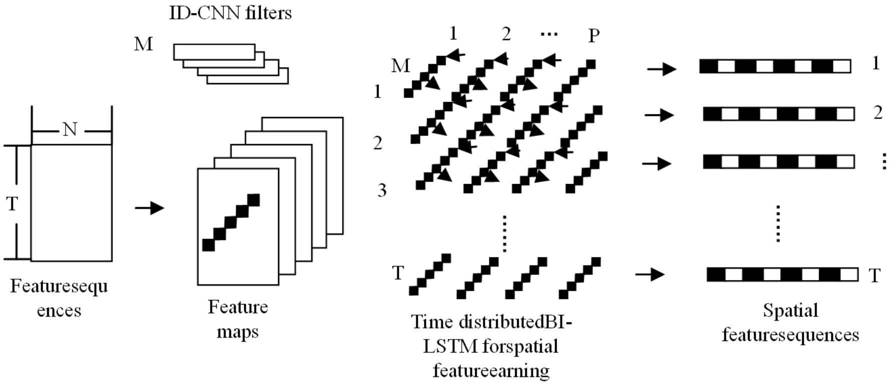

**Figure 3.** Spatial Feature Learning

$M$dimensional feature vector $\overrightarrow{fs}(i,t) = [f_1(i,t), f_2(i,t), ..., f_M(i,t)]$. Inputting $\overrightarrow{fs}(i,t), i \in (1, ..., P)$ into the Bi-LSTM network, the output is a spatial feature representation $fsr(t)$ of node correlations at the moment $t$, see (3):

$$\begin{cases} i_t = \sigma\left(W_{iy}y_t + W_{ih}h_{t-1} + b_i\right), \\ f_t = \sigma\left(W_{fy}y_t + W_{fh}h_{t-1} + b_f\right), \\ o_t = \sigma\left(W_{oy}y_t + W_{oh}h_{t-1} + b_o\right), \\ c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh\left(W_{cy}y_t + W_{ch}h_{t-1} + b_c\right), \\ h_t = o_t \cdot \tanh(c_t). \end{cases} \tag{2}$$

$$\begin{cases} hs_i^{\text{forward}}(t) = LSTM^{\text{forward}}(h_{i-1}, fs(i,t), c_{i-1}), \\ hs_i^{\text{backward}}(t) = LSTM^{\text{backward}}(h_{i-1}, fs(i,t), c_{i-1}), \\ fsr(t) = \left\{hs_i^{\text{forward}}(t), hs_i^{\text{backward}}(t)\right\}. \end{cases} \tag{3}$$

The time-domain LSTM network layer learns the temporal features of the sequence, inputting the spatial features $fsr(t)$ at time $t$ into the time-domain LSTM network, and learning the temporal correlation of the sequence at time $1 \sim T$, to obtain vector $fsr$ , which represents the temporal fusion features of the behavioral sequence, see Eq. (4):

$$fsr = LSTM(fsr(t)). \tag{4}$$

## 3. Multichannel Weight-Sharing Spatiotemporal Feature Fusion Network

A multi-channel feature fusion network based on distance features was designed to learn single and double association features respectively, as shown in Figure 4. The features input to each channel of this network belong to the distance features of the skeleton and have the same physical meaning. Therefore, a multi-channel weight-sharing spatiotemporal feature fusion network structure is used. Multiple convolutional kernels are used to extract features in a single channel, and the remaining channels use the same structure with shared convolutional kernel weights [19]. The multi-channel weight sharing spatiotemporal fusion network structure has two advantages, one is to reduce the network parameters, and the other is to avoid the gradient dispersion during the training process of the multi-channel structure.

The four sequences of $D_1, D_2, D_{c1}, D_{c2}$ are fed into the structure 1DCNN-LSTM LSTM spatiotemporal feature fusion model, which is set to implement the function $F_{SPT}$, with the same structure of

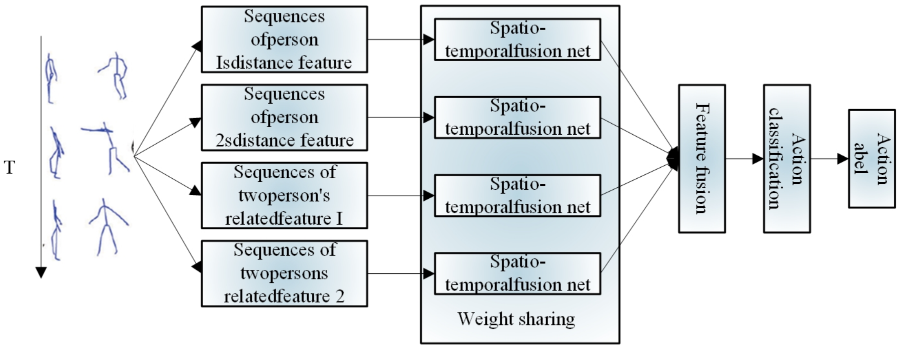
T
Sequences ofperson 1sdistance feature
Sequences ofperson 2sdistance feature
Sequences of twoperson's relatedfeature I
Sequences of twopersons relatedfeature 2
Spatio-temporalfusion net
Spatio-temporalfusion net
Spatio-temporalfusion net
Spatio-temporalfusion net
Weight sharing
Feature fusion
Action classification
Action abel

**Figure 4.** Multichannel Spatiotemporal Fusion Network

the four branches and shared weights, as shown in (5):

$$\begin{cases} fsr_{d1} = F_{SPT}\left(D_1\right), \\ fsr_{d2} = F_{SPT}\left(D_2\right), \\ fsr_{dc1} = F_{SPT}\left(D_{c1}\right), \\ fsr_{dc2} = F_{SPT}\left(D_{c2}\right). \end{cases} \tag{5}$$

The outputs are fused together to form a multi-branch fusion feature $f_{fusion}$, see Eq. (6):

$$f_{fusion} = \left[fsr_{d1}, fsr_{d2}, fsr_{dc1}, fsr_{dc2}\right]. \tag{6}$$

The interaction behavior labels were obtained by learning fusion features using the fully connected network, see Eq. (7):

$$L = soft\max\left(W * f_{fusion}\right). \tag{7}$$

## 4. Experimental Results and Analysis

Experiments were conducted on an Ubuntu 16.04 system with an RTX2070 graphics card using TensorFlow as the back-end and Keras as framework. The hyperparameters of the entire network were: maximum number of iterations 16, batch size 12, input image dimensions (112, 112, 16, 3) for height, width, number of consecutive frames and channels respectively, initial learning rate 0.005, learning rate reduced to one-tenth of the original rate after every 4 rounds, optimizer is stochastic gradient descent, using Nesterov momentum with a momentum value of 0.9. In the test, the input video was time-domain segmented into 3 segments, 16 consecutive RGB images were taken from each segment as network input, and for video frames less than the required number, they were replicated in a loop, and the average of the scores of the 3 video segments was used as the final classification result [20].

Table 1 shows the accuracy of the Inflated VGGNet-16 network designed in this paper compared with the C3D network and other mainstream networks by center-cutting the data as input. Figure 5 shows the data curves of the corresponding experiments in Table 1, where IVGG-16, 19 and IRes-18, 34, 50 represent the corresponding Inflated versions of VGG-16, 19 and Res-18, 34, 50 respectively.

It can be seen from Table 1 and Figure 5 that the small dataset could not satisfy the training of the deep network model of VGGNet-19, while for the relatively shallow residual networks ResNet-18 and ResNet-34 could not model the behavioral actions well, among the VGGNet-16, VGGNet-19, Res-Net-18, ResNet-34 and ResNet-50 of each extended 3D network, this paper models the network with the best recognition performance and the highest accuracy rate [21]. It can also be seen that simply extending the VGGNet-16 network from 2D to 3D is less accurate than the C3D network and converges slowly when trained from scratch, partly because of the insufficient amount of data and partly because the network is initialized from scratch. However, after extending the pre-trained model

| Method | Accuracy/% |
|---|---|
| C3D | 63.2 |
| VGGNET-6(3D) | 49.8 |
| Inflated Resnet -18 | 43.3 |
| Inflated Res Net-34 | 57.7 |
| Inflated Resnet-50 | 61.9 |
| Inflated VGGNET-16 | 72.4 |
| Inflated VGGNET-19 | 71.0 |

**Table 1.** Comparison of this Paper's Algorithm with the C3D Algorithm and Other Mainstream Networks on the UCF101 Dataset Split1

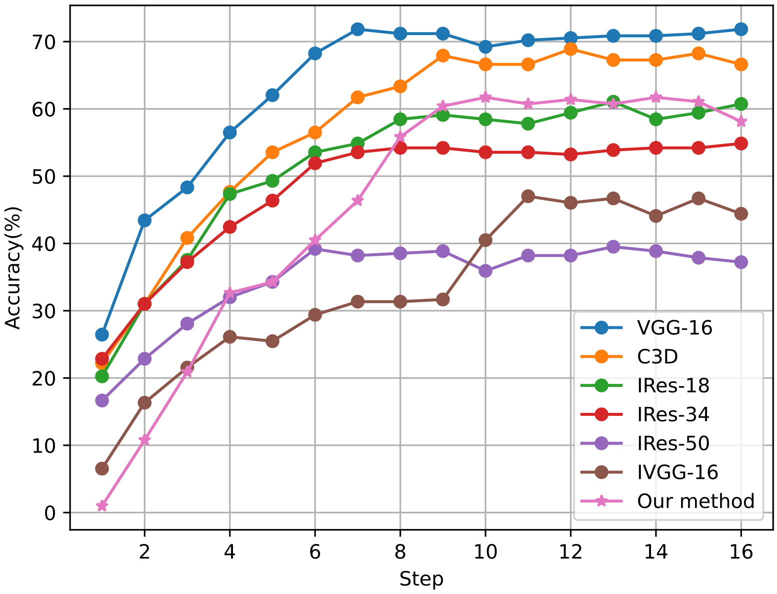

**Figure 5.** Accuracy Curve of UCF101 Split1 Verification Set

| Data set | Accuracy/% | | | |
|---|---|---|---|---|
| | Split1 | Split2 | Split3 | Average |
| UCF101 | 88.9 | 90.8 | 89.4 | 89.7 |
| HMDB-51 | 62.4 | 61.0 | 62.0 | 61.8 |

**Table 2.** Average Accuracy of the Algorithms in this Paper on the two Datasets

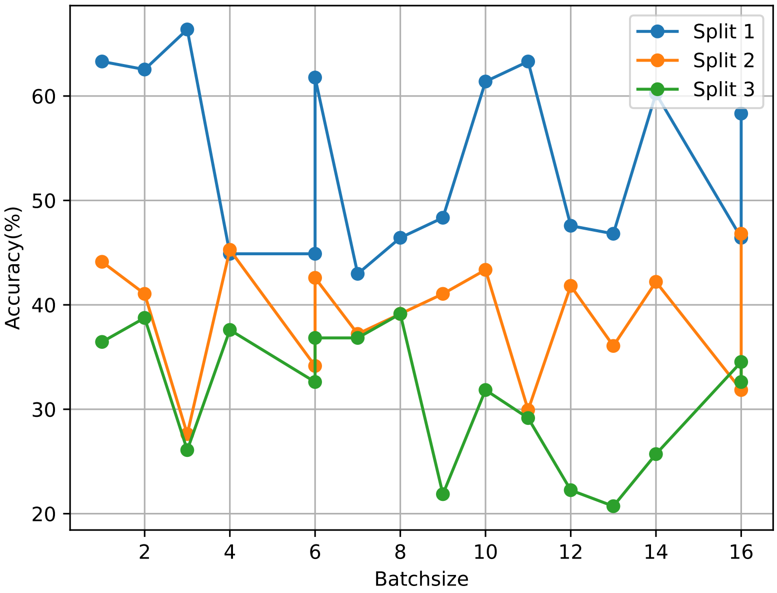

**Figure 6.** UCF101 Test Set Accuracy Curve

on the ImageNet dataset to 3D and initializing it, the network converged with improved speed and 72.3% accuracy [22].

Table 2 shows the accuracy of the algorithm in this paper on three standard training-test set grouping schemes (Split1, Split2, Split3) for two datasets with 10-fold enhancement of the input data and a batch normalization layer added to the network, and the final accuracy of the algorithm is averaged over the three components.

Figures 6 and 7 show the corresponding test accuracy curves for the 3 scenarios on UCF101 and HMDB-51 respectively. The batch size was 32 at the time of testing, and 16 batches were randomly selected for testing on both datasets and the average was taken as the final accuracy.

Table 3 shows the accuracy comparison with several classical behavior recognition algorithms as well as algorithms with different data forms as input, it can be seen that the model in this paper is able to achieve a higher recognition accuracy than the classical i DT, two-stream and C3D networks in both cases. The use of the Kinetics dataset to pretrain the network and the combination of different network structures such as two-stream CNN with 3D convolution can all improve the recognition accuracy on small datasets such as UCF101 [23]. It can also be seen that the improvement of the designed network is greater on the HMDB-51 dataset compared to the two-stream algorithm, indicating that the 3D network structure is more favorable for modelling in the temporal dimension and for extracting motion information compared to the 2D network.

Figure 8 illustrates real-time video prediction utilizing the planned network. The projected action categories and accompanying prediction probabilities are displayed in the two lines of text in the upper left corner. The constructed network has a very high recognition accuracy for actions with modest backdrop changes or obvious moving targets, as shown in the figure, and it continues to have a high recognition accuracy for actions with big background changes and moving targets that are more or

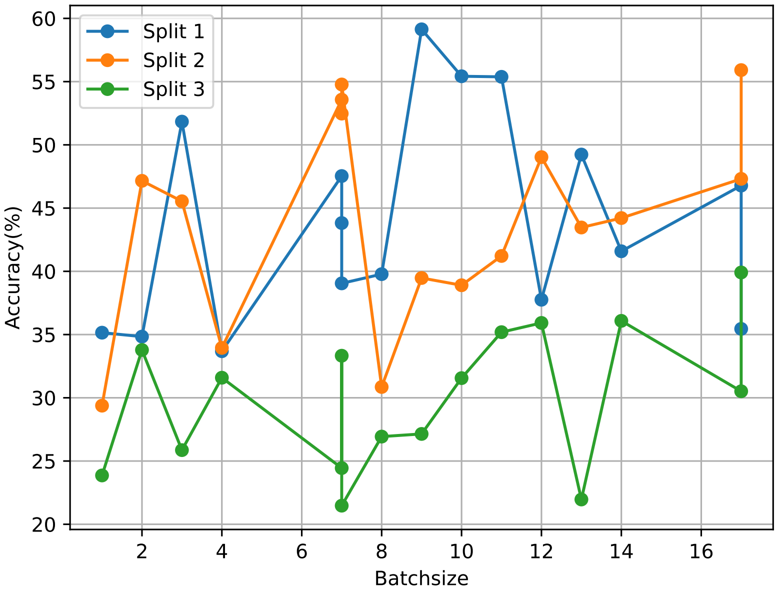

**Figure 7.** HMDB-51 Test Set Accuracy Curve

| Method | Accuracy/% | |
|---|---|---|
| | UCF101 | HMDB-51 |
| I DT +FV | 86.0 | 57.3 |
| Two- stream | 88.1 | 59.5 |
| Motion Vector +FV | 78.6 | 46.8 |
| RGB +Enhanced Motion vector | 86.5 | |
| RGB+ R GB Diff | 87.4 | |
| two-stream 13 D (Kinetics) | 98.1 | 81.0 |
| C3D + liner SVM | 83.4 | |
| R (2 + 1) D-RGB (Kinetics) | 96.9 | 74.6 |
| T-c3d (Kinetics) | 92.6 | 62.5 |
| MARS + RGB Flow (Kinetics) | 95.9 | |
| our method | 89.7 | 61.8 |

**Table 3.** Comparison Between the Algorithm in this Article and Current Advanced Algorithms

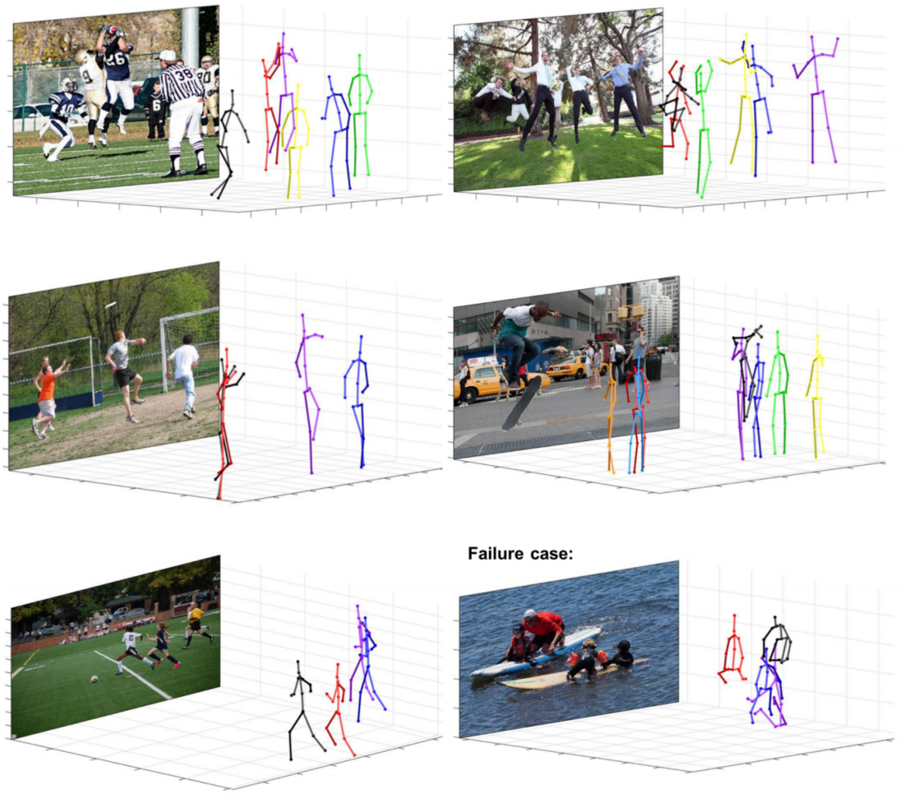

**Figure 8.** Algorithm Recognition Results on the UCF101 Dataset

less evident [24, 25].

Using a segmented network to separate salient parts of motion and combining attentional mechanisms for the recognition of actions occurring in videos, it is possible to further increase the accuracy of human action recognition by solving the intra-class variability and inter-class similarity problems [26].

The method presented in this paper, which successfully handles the issue of converting 2D portraits to the corresponding 3D models, has the great advantage that it does not require manual feature annotation to achieve good conversion results (although manual gender annotation can produce better results). This is evident from Figure 9. The situation in column 2 also demonstrates that the method in this paper is susceptible to errors due to the complexity of the actual scene: overlapping of different limbs in height, overlapping of different characters in the absence of depth information, and inability to accurately distinguish the orientation of the characters.

## 5. Conclusion

A multi-channel spatiotemporal fusion network based on deep learning for two-person interaction behavior recognition method, the experimental results verify that the CNN-LSTM structure can extract the spatiotemporal fusion features of behavior r sequences. A novel two-person behavior representation is designed to obtain a two-person perspective invariant feature representation using four sets of distance features to represent the original skeleton, which improves the perspective invariance of behavior features. A multi-channel weight-sharing spatiotemporal feature fusion network model is designed, using multiple channels to process each group of features separately and sharing the weights among multiple channels to extract multiple groups of spatiotemporal fusion features without increasing the network parameters. The method has a high accuracy rate in two-person behavior recognition, and experimental results comparing with typical algorithms in this field show that the proposed method has obvious advantages in two-person interaction behavior recognition. Multimodal behavior features will be introduced later to achieve behavior analysis of more complex scenes.

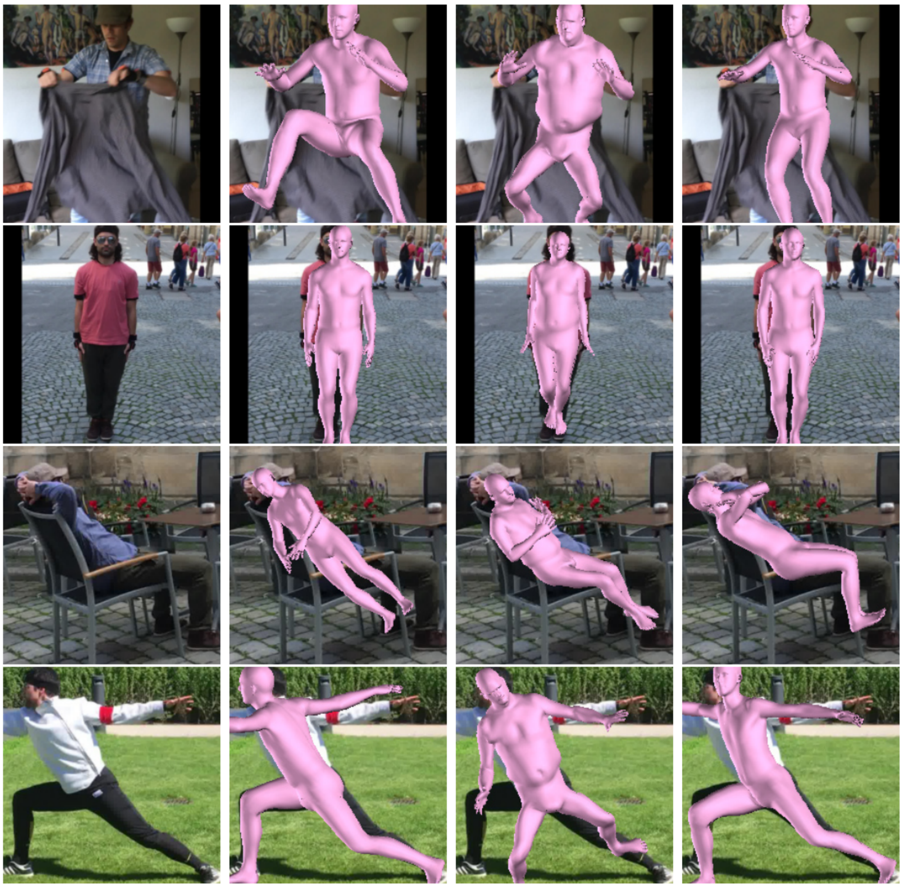

**Figure 9.** 3D Human Fusion Effect

## Funding

No funding is available for this research.

## Conflict of interest

The author declares no conflict of interests.

## References

1. Ye, Q., Li, R., Yang, H. and Guo, X., 2022. Human interactive behaviour recognition method based on multi-feature fusion. *International Journal of Computational Science and Engineering, 25*(3), pp.262-271.
2. Ye, Q., Li, R. and Guo, X., 2021, February. Human interaction behavior recognition based on local spatial-temporal and global feature. In *Journal of Physics: Conference Series* (Vol. 1754, No. 1, p. 012188). IOP Publishing.
3. Le, T., Singh, R. and Miller, T., 2021. Goal Recognition for Deceptive Human Agents through Planning and Gaze. *Journal of Artificial Intelligence Research, 71*, pp.697-732.
4. Ke, G., Chen, R.S., Chen, Y.C., Hu, Y.X. and Wu, T.Y., 2022. Simple multi-scale human abnormal behaviour detection based on video. *International Journal of Information and Computer Security, 17*(3-4), pp.310-320.
5. Qayyum, A., Razzak, I., Moustafa, N. and Mazher, M., 2022. Progressive ShallowNet for large scale dynamic and spontaneous facial behaviour analysis in children. *Image and Vision Computing, 119*, p.104375.

6. Acharya, D., Varshney, N., Vedant, A., Saxena, Y., Tomar, P., Goel, S. and Bhardwaj, A., 2021. An enhanced fitness function to recognize unbalanced human emotions data. *Expert Systems with Applications, 166*, p.114011.

7. Tsai, M.F. and Chen, C.H., 2021. Spatial temporal variation graph convolutional networks (STV-GCN) for skeleton-based emotional action recognition. *IEEE Access, 9*, pp.13870-13877.

8. Yahaya, S.W., Lotfi, A. and Mahmud, M., 2021. Towards a data-driven adaptive anomaly detection system for human activity. *Pattern Recognition Letters, 145*, pp.200-207.

9. Ameen, H.A., Mahamad, A.K., Saon, S., Ahmadon, M.A. and Yamaguchi, S., 2021. Driving behaviour identification based on OBD speed and GPS data analysis. *Advances in Science Technology and Engineering Systems Journal, 6*(1), pp.550-569.

10. Huang, S., Liu, X., Chen, W., Song, G., Zhang, Z., Yang, L. and Zhang, B., 2022. A detection method of individual fare evasion behaviours on metros based on skeleton sequence and time series. *Information Sciences, 589*, pp.62-79.

11. Li, Y., 2021. Dance motion capture based on data fusion algorithm and wearable sensor network. *Complexity, 2021*, Article ID 2656275, 11 pages.

12. Rahm, J. and Johansson, M., 2021. Assessment of outdoor lighting: Methods for capturing the pedestrian experience in the field. *Energies, 14*(13), p.4005.

13. Cárdenas, J., Blanca, M.J., Carvajal, F., Rubio, S. and Pedraza, C., 2021. Emotional processing in healthy ageing, mild cognitive impairment, and Alzheimer's disease. *International Journal of Environmental Research and Public Health, 18*(5), p.2770.

14. Effendi, J., Tjandra, A., Sakti, S. and Nakamura, S., 2021. Multimodal Chain: Cross-Modal Collaboration Through Listening, Speaking, and Visualizing. *IEEE Access, 9*, pp.70286-70299.

15. Zhang, C., Li, M. and Wu, D., 2022. Federated Multidomain Learning With Graph Ensemble Autoencoder GMM for Emotion Recognition. *IEEE Transactions on Intelligent Transportation Systems, 24*(7), pp.7631-7641

16. Tien, P.W., Wei, S., Calautit, J.K., Darkwa, J. and Wood, C., 2021. Vision-based human activity recognition for reducing building energy demand. *Building Services Engineering Research and Technology, 42*(6), pp.691-713.

17. Aza-Conde, J., Reyes, C., Suárez, C.F., Patarroyo, M.A. and Patarroyo, M.E., 2021. The molecular basis for peptide-based antimalarial vaccine development targeting erythrocyte invasion by P. falciparum. *Biochemical and Biophysical Research Communications, 534*, pp.86-93.

18. Guerra, S., Chung, R., Yerbury, J. and Karl, T., 2021. Behavioural effects of cage systems on the G93A Superoxide Dismutase 1 transgenic mouse model for amyotrophic lateral sclerosis. *Genes, Brain and Behavior, 20*(5), p.e12735.

19. Diraco, G., Rescio, G., Caroppo, A., Manni, A. and Leone, A., 2023. Human action recognition in smart living services and applications: context awareness, data availability, personalization, and privacy. *Sensors, 23*(13), p.6040.

20. Ibáñez, M.L., Miranda, M., Alvarez, N. and Peinado, F., 2021. Using gestural emotions recognised through a neural network as input for an adaptive music system in virtual reality. *Entertainment Computing, 38*, p.100404.

21. Hernandez, J., Valarezo, G., Cobos, R., Kim, J.W., Palacios, R. and Abad, A.G., 2021. Hierarchical Human Action Recognition to Measure the Performance of Manual Labor. *IEEE Access, 9*, pp.103110-103119.

22. Abdulazeem, Y., Balaha, H.M., Bahgat, W.M. and Badawy, M., 2021. Human action recognition based on transfer learning approach. *IEEE Access, 9*, pp.82058-82069.

23. Masuda, Y., 2022. Human recognition-behavioral adaptation system. *Human Arenas, 5*(1), pp.57-66.

24. Fang, L. and Sun, M., 2021. Motion recognition technology of badminton players in sports video images. *Future Generation Computer Systems, 124*, pp.381-389.

25. Guo, H., Wan, J., Wang, H., Wu, H., Xu, C., Miao, L., Han, M. and Zhang, H., 2021. Self-powered intelligent human-machine interaction for handwriting recognition. *Research, 2021*, p.4689869

26. Yamini, G. and Ganapathy, G., 2021. Enhanced sensing and activity recognition system using IoT for healthcare. *International Journal of Information Communication Technologies and Human Development, 13*(2), pp.42-49.