ATMLP: Attention and Time Series MLP for Fall Detection

Huan Zhang, Wenhua Cui, Tianwei Shi, Ye Tao, Jianfeng Zhang

Abstract—**Addressing the problem that the human body is difficult to be correctly identified in the process of falling, this study proposes an Attention Mechanism and Time Series MLP (AT-MLP) model. To realize the capture of body features during the state and process of the fall of the human body, the time series is combined in the submodule. Thus, the subspace Angle Time Series algorithm (ATS) and the subspace Angle Difference-value Time Series algorithm (ADTS) are designed. The AT-MLP stage composed of Convolutional Block Attention Module (CBAM) and MLP blocks is constructed in the model to strengthen the relation between the time and space dimensions. Comparing with other methods on the Le2i Fall detection Dataset (LFDD) to verify the effectiveness of the AT-MLP model, the experimental results show that the AT-MLP model effectively improves the detection efficiency of human falls and has strong generalization ability.**

Index Terms—**AT-MLP, Fall Detection, Subspace Node Angle, Time Series**

I. INTRODUCTION

WITH the continuous development of computer technology, video-based behavior recognition has achieved rapid progress. Fall detection has always been a major topic in the study of abnormal human behavior. The application of computer technology on human fall detection presents a significant scientific research value and reality significance [1]. The methods of human fall detection are roughly divided into two categories: The first is fall detection based on spatial information collected by external sensors. The second is to use the human keypoints detection to determine whether the human body is in a fall state [2]. Because the information collected by the sensor is quite different from the external situation, the detection accuracy is low [3-5]. Most of the mainstream methods use human key point detection.

Recently, the majority of studies are based on the human pose, geometric shape information [6]-[8], and the background of the moving object for pose extraction to achieve the purpose of fall detection. For example, Hua et al. used keypoints vectorization method was exploited to eliminate irrelevant information in the initial coordinate representation. Then, the observed keypoint sequence of each person was input to the pose prediction module adapted from sequence-to-sequence (seq2seq) architecture. By predicting the sequence of key points in the future, it is possible to judge whether a person falls or not [9]. Chua et al. [10] proposed to divide the human body into three parts: head, torso, and feet, and detect the falls by analyzing the shape changes of the human silhouettes. Fan et al., proposed a difference score method (DSM) based on adjacent dynamic images in the temporal sequence [11], which converts the four stages of falls (standing, falling, fallen and not moving) into four categories. Rougier et al. [12] suggested fitting the shape of the human body in the video sequence by fitting an ellipse and subsequently tracking whether the human silhouette falls. Agrawal et al. [13] proposed an improved Gaussian mixture model background subtraction method, which calculates the aspect ratio and midpoint distance of the human body to determine whether it falls. Sun et al. [14] suggested a hierarchical feature detection algorithm based on an improved GMM background update. According to the sensitivity judgment of each feature to different states of the human body, the features of the height ratio of the human body centroid and the rectangular aspect ratio were used to analyze the distinguished human behavior. Chamle et al. [17] used the Adaboost classifier to classify the data and judge the falling behavior of the human body. Poonsri et al. [18] employed a mixture of the Gaussian model (MoG) and background subtraction with an average filter model, combine the Principal Component Analysis (PCA) of human silhouettes for feature extraction. Alaoui et al. [19] introduced the optical flow algorithm to estimate the speed and direction of human movement. The implementation of these methods is based on the Le2i Fall detection Dataset (LFDD), but the accuracy and generalization ability still need improvement. Although a two-class SVM classifier and a Long Short-Term Memory (LSTM) were trained using the angle information between the vectors formed by the frames of the video sequence [15] to discriminate between fallen and unfallen states, these methods lack the ability to capture the characteristics of the human body in the process of falling, and still, show the problem of high false detection rate.

To solve the above-mentioned issues, this study proposes

Huan Zhang is a postgraduate student of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China. (e-mail: hzzhbest@163.com).

Wenhua Cui is a Professor of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China. (corresponding author to provide phone: 130-1963-5901; e-mail: cwh@systemteq.net).

Tianwei Shi is an associate Professor of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China. (e-mail: tianweiabbcc@163.com).

Ye Tao is a lecturer of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China. (e-mail: taibeijack@163.com).

Jianfeng Zhang is a postgraduate student of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China. (e-mail: jianfengzhang177@163.com).
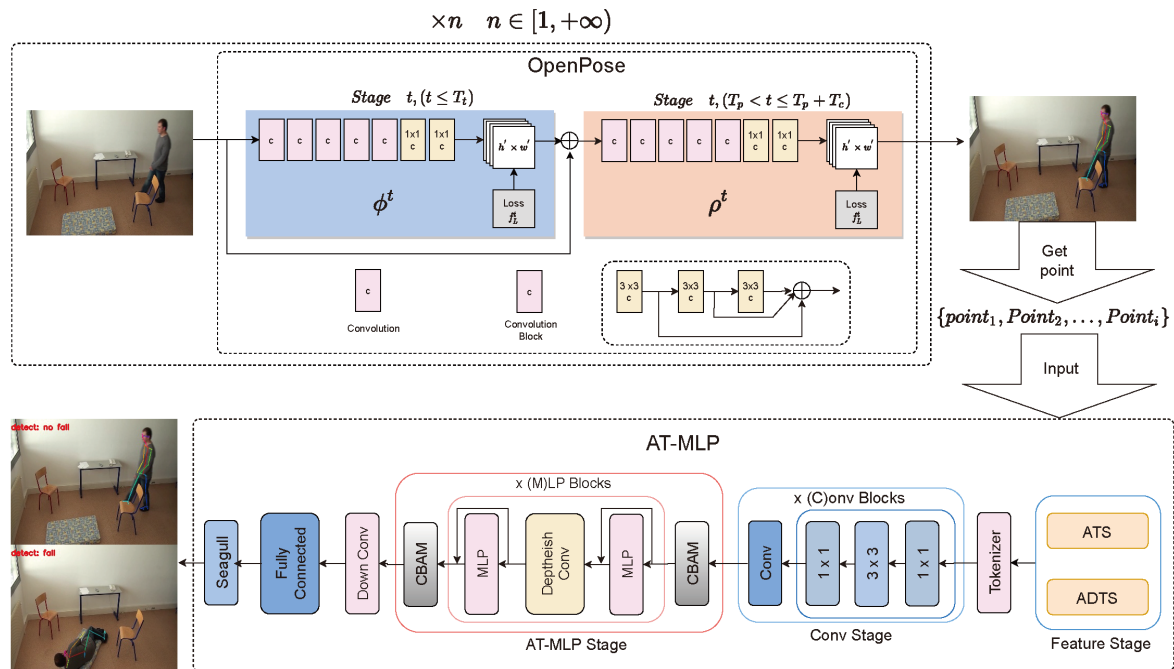
Fig. 1. Overall architecture of the model

a new detection model (AT-MLP), designs ATS and ADTS algorithms, and constructs an AT-MLP stage with CBAM blocks in the model. The ATS and ADTS algorithms use the abscissa and ordinate of the subspace nodes extracted by the OpenPose network, and calculate the node angle to construct a multi-dimensional feature matrix with time series, so that the spatiotemporal feature capture of the detected object is clearer during the falling process. Simultaneously, the AT-MLP stage module effectively assigns channel weights, extracts key information from different channels, and focuses on capturing temporal and spatial feature values. The experimental results show that the AT-MLP model has a significantly superior performance and generalization ability, and the accuracy rate reaches 96.21%. Compared with various methods [15], [17]-[19] under the same dataset, the prediction ability and detection accuracy of the model have been effectively improved.

II. METHOD

A. Architecture

This study is based on the AT-MLP model, using the OpenPose network to extract the joint points of the skeleton, and performing fall recognition on the extracted joint point information. This model mainly includes three parts: a feature stage, Conv Stage and AT-MLP stage, which can extract temporal feature information and effectively interact with spatial information. The recognition effect of low-level features such as the edge or texture background is obvious, and the architecture is shown in Figure 1.

In the process of human fall detection, it is necessary to perform frame extraction processing on the fall video, input the processed image into the OpenPose network for extraction of the skeleton joint points, select subspace nodes, and construct the subspace node angle matrix and transmit it to the AT-MLP model. In this setting, the subspace node angle matrix was firstly input into the feature module

composed of ATS and ADTS, which focuses on capturing the features of the human fall and its process, forming a fall-prediction mechanism and inputting it to the Tokenizer module of the model in the form of a feature matrix. Note that the Tokenizer module contains three convolution blocks formed by 3×3 convolution, batch normalization, and Rectified Linear Units (ReLU) activation functions, as well as a maximum pooling layer. In the Conv Stage, multiple convolution blocks were present. After the input features were obtained, the residual path was used to add the input and output features to facilitate the interaction of spatial information. In the AT-MLP stage, the attention weights were sequentially inferred along the two dimensions of space and channel and then multiplied with the original feature map. Adaptive adjustment of the obtained features, and the 1×1 convolution was used to stack the convolutions. Local information interaction was carried out in a layer-by-layer manner.

B. ATS

The ATS algorithm is based on the calculation of the human subspace node angle and time series. Firstly, the video frame was extracted and identified by the OpenPose network and detected the same joint point at multiple consecutive time points. The coordinate $Point_i$ is shown in Formula (1):

$$Point_i = \{T_1 = (x_1, y_1, t_1), T_2 = (x_2, y_2, t_2), \cdots, T_n = (x_n, y_n, t_n)\} \quad (1)$$

where i represents the selected bone joint point. Regarding time series, that is, Univariate Time Series (UTS), as shown in Formula (2):

$$UTS = (uts_1, uts_2, \cdots, uts_t, \cdots, uts_T) \quad (2)$$

where t represents the time of the time interval, T represents the end time, ust_t represents the observation value at time t, and the range of t is $1 \le t \le T$.

Integrate $Point_i$ with time series to $Point_i'$, is as shown in Formula (3):

$$Point'_i = \begin{bmatrix} [T_1, T_2, ..., T_m]_1, [T_{m+1}, T_{m+2}, ..., T_{2m}]_2, \cdots \\ , [T_{(m-i)+1}, T_{(m-i)+2}, ..., T_k]_n \end{bmatrix} \quad (3)$$

where m represents the number of picture frames contained in a single time series, and n represents the number of time series.

Divide the whole data into n blocks, calculate the angle for each time series block, and build the subspace node $angle_{k,t}$, as shown in Formula (4):

$$angle_{k,t} = \begin{cases} 0, & T_{i,t}(2) = T_{j,t}(2) \\ \dfrac{T_{i,t}(2) - T_{j,t}(2)}{T_{i,t}(1) - T_{j,t}(1)}, & \begin{bmatrix} i \in [0, 24] \\ j \in [0, 24] \end{bmatrix} \\ \infty, & T_{i,t}(1) = T_{i,t}(1) \\ None, & T_{i,t}(2) = T_{j,t}(2), T_{i,t}(1) = T_{j,t}(1) \end{cases} \quad (4)$$

Among them, the subspace node angle is defined as an array $angle_{k,t}$, which contains all the angle information of the k subspace nodes selected at time t. i, j are the serial numbers corresponding to the selected different joint points, $T_{i,t}(1)$ and $T_{i,t}(2)$ are the abscissa and ordinate of the selected associative subspace nodes, and the subspace angle information in the video is obtained based on the $angle_{k,t}$.

Integrate $angle_{k,t}$ with time series to construct a single block feature matrix S that reflects the changes of $angle_{k,t}$ in time order, as shown in Formula (5):

$$S_{i,j,t} = \begin{bmatrix} angle_{i,t} \\ angle_{i+1,t} \\ \vdots \\ angle_{j,t} \end{bmatrix} \quad (5)$$

where i and j are the corresponding start and end coordinates of the time series block $Point'$, respectively. All the obtained single block feature matrices are spliced to form an n-dimensional feature matrix L, as shown in Formula (6):

$$L = (S_{1,i,1}, S_{i+1,2i,2}, \cdots, S_{(k-i)+1,k,t}) \quad (6)$$

C. ADTS

Considering the difference between the human body's falling state and process, the ADTS algorithm has a clear recognition effect on the falling process, realizes the prediction of the falling action, and makes the feature change smoother. First, take the angle difference between the current sequence moment t and the previous moment $t-1$ as $\Delta angle_{k,t}$:

$$\Delta angle_{k,t} = \begin{bmatrix} \Delta angle_{0,t} = angle_{0,t} - angle_{0,t-1} \\ \Delta angle_{1,t} = angle_{1,t} - angle_{1,t-1} \\ \vdots \\ \Delta angle_{k,t} = angle_{k,t} - angle_{k,t-1} \end{bmatrix} \quad (7)$$

where $\Delta angle_{k,t}$ is the angle change difference of the selected *k-th* subspace node at time t, $\Delta angle_{k,t}$ and $\Delta angle_{k,t-1}$ represent the angle values at time t and $t-1$ of the *k-th* subspace node, respectively.

Since $Point'_t$ contains k frame images, the calculated

$\Delta angle_{k,t}$ contains $t-1$ data. To solve the problem of the difference between the two dimensions, add the $\Delta angle_{k,1}$ to the sequence in front of $\Delta angle_{k,2}$ in the form of padding, and assign it to 0, so that the two dimensions are the same. Get the feature matrix M with the time dimension:

$$M = [\Delta angle_{k,1} \quad \Delta angle_{k,2} \quad \Delta angle_{k,3} \quad \ldots \quad \Delta angle_{k,t}] \quad (8)$$

M is divided into blocks by column, the dimensions are consistent with the ATS feature matrix, and the divided matrix $M_{i,j}$ is obtained:

$$M_{i,j} = \begin{bmatrix} \Delta angle_{0,i} & \Delta angle_{0,i+1} & \ldots & \Delta angle_{0,j} \\ \Delta angle_{1,i} & \Delta angle_{1,i+1} & \ldots & \Delta angle_{1,j} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta angle_{k,i} & \Delta angle_{k,i+1} & \ldots & \Delta angle_{k,j} \end{bmatrix} \quad (9)$$

where i and j are the corresponding start and end coordinates of the time series block $Point'$, respectively.

The segmented $M_{i,j}$ and ATS feature matrix are concatenated to form a new feature matrix F with the features of the falling process, which is expressed as:

$$F = ATS(angle_{ij}) \oplus ADTS(angle_{ij}) \quad (10)$$

Input the calculation results of ATS and ADTS into the Tokenizer module of the AT-MLP model to capture the posture features during the falling state and process. The expansion of the feature matrix F is shown in Formula (11):

$$F = \begin{bmatrix} \begin{bmatrix} angle_{0,1} & \Delta angle_{0,1} \\ angle_{1,1} & \Delta angle_{2,1} \\ \vdots & \vdots \\ angle_{k,1} & \Delta angle_{k,1} \end{bmatrix} \cdots \begin{bmatrix} angle_{0,m} & \Delta angle_{0,m} \\ angle_{1,m} & \Delta angle_{2,m} \\ \vdots & \vdots \\ angle_{k,m} & \Delta angle_{k,m} \end{bmatrix} \cdots \\ \begin{bmatrix} angle_{0,t-m} & \Delta angle_{0,t-m} \\ angle_{1,t-m} & \Delta angle_{2,t-m} \\ \vdots & \vdots \\ angle_{k,t-m} & \Delta angle_{k,t-m} \end{bmatrix} \cdots \begin{bmatrix} angle_{0,t} & \Delta angle_{0,t} \\ angle_{1,t} & \Delta angle_{2,t} \\ \vdots & \vdots \\ angle_{k,t} & \Delta angle_{k,t} \end{bmatrix} \end{bmatrix} \quad (11)$$

The sequence features are obtained according to the trend features of the time series, and the feature points of the time series variable dimension are extracted in combination with the feature stage. The feature extraction is shown in Figure 2.
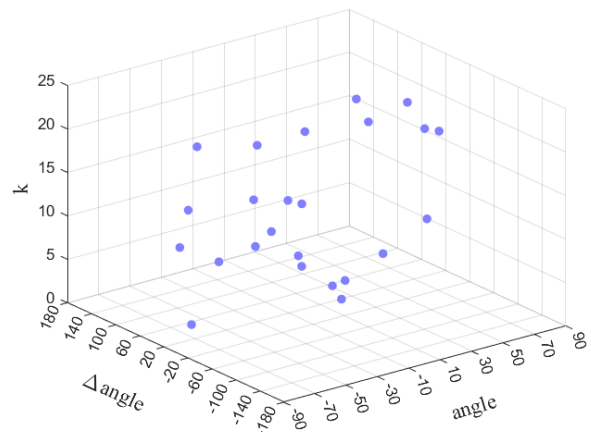

Fig. 2. Schematic diagram of feature points extraction

D. AT-MLP Stage

The AT-MLP stage consists of the CBAM and MLP blocks. Due to the lack of consideration of the feature space dimension and the correlation between channels in the

model, the extracted feature relation was insensitive. To better adapt to the subspace node information processed by ATS and ADTS, CBAM was introduced to enhance the feature association between space and time. CBAM includes two blocks, the Channel Attention Module (CAM) and Spartial Attention Module (SAM), which can focus local information on the two parts of the spatial and channel dimensions, respectively. The CBAM network structure is shown in Figure 3.
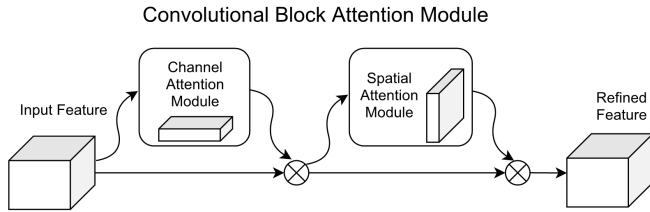

Fig. 3. CBAM network structure diagram

The CAM performed Global Maximum Pooling and Global Average Pooling on the input matrix, allowing the network to pay more attention to major feature maps. After the feature maps were obtained, they were passed to the shared fully connected layer. Before the Sigmoid activation operation, the output of feature values by the shared fully connected layer was added based on element-wise, and finally, the multiplication based on element-wise was performed to generate the input features required by the block. The calculation Formula is:

$$M_c(F) = \sigma(W_1(W_0(F_{avg}^C)) + (W_1(W_0(F_{max}^C)))) \qquad (12)$$

where σ represents the Sigmoid operation. The channel attention mechanism assigns different weights to each channel, learning the importance of different channels.

The input of the SAM is the output of the CAM. After Global Maximum Pooling and Global Average Pooling operations, concatenate channel features and 7×7 convolutions were performed for dimensionality reduction. Finally, after processing by the Sigmoid, the multiplication operation was performed directly. The calculation Formula is:

$$M_s(F) = \sigma(f^{7 \times 7}([F_{avg}^S; F_{max}^S])) \qquad (13)$$

where $f^{7 \times 7}$ represents the 7×7 convolution operation. The spatial attention mechanism performs corresponding spatial transformation in the image spatial domain information, so that key information is extracted.

In the AT-MLP model, CBAM blocks were added to the input and output terminals of MLP blocks, respectively. In the MLP blocks, the input features were normalized (LayerNorm) and then passed through the fully connected layer and the Gaussian Error Linear Unit (GELU) activation, and the obtained feature map was added to the input features. The 1×1 convolution was adopted to reduce the constraints on the input dimension. The middle convolutional layer used a 3×3 depthwise convolution for spatial information interaction. MLP blocks used a patch merging layer to down-sample feature maps. In the case of a small number of parameters, the AT-MLP effectively improves the accuracy of knowing whether the human body is in a falling state. The network structure of MLP blocks is shown in Figure 4.
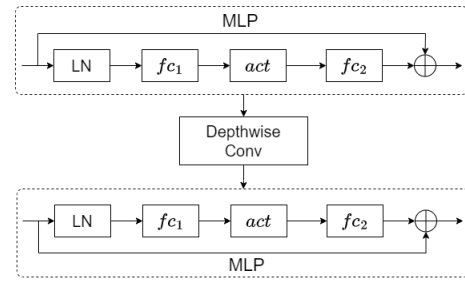

Fig. 4. MLP Blocks network structure diagram

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. Datasets

The LFDD [16] includes four scenes: Home, Coffee room, Office, and Lecture room, which are composed of 191 videos. Among them, 143 videos contain falls, 48 contain several normal activities, movements, and body transfer, including from chair to sofa. The pixel resolution is 320×240, and the frame rate is 25 frames per second. Falling postures include front fall, side fall, rear side fall and other postures. The shooting camera is at the top of the indoor environment room, and the video contains only one person. The video sequences of the dataset include various complex conditions, such as lighting factors and occlusions. The LFDD dataset contains both bright and low light conditions, as well as textured backgrounds. Furthermore, there are frames with an abnormal falling behavior and no people shown in the video.

B. Experimental Setting

This experiment was performed on a server with Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz 2.20 GHz (2 processors) and NVIDIA TITAN X(Pascal) with 12 GB of graphics memory. The operating system was Window10 64-bit, using Python 3.8 and Pytorch 1.9. During the experiments, all learning rates were initialized to 0.01.

C. Evaluation Metrics

To verify the effectiveness of the proposed model and algorithm, this study used four evaluation indexes: accuracy, precision, recall, and F-score. The model's performance was evaluated and compared with other human fall detection algorithms.

Accuracy represents the proportion of falling and non-falling samples to all samples when the model is correctly classified, as shown in Formula (14):

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad (14)$$

Precision represents the proportion of fall samples to all fall samples, as shown in Formula (15):

$$Precision = \frac{TP}{TP + FP} \qquad (15)$$

Recall represents the proportion of fall samples in the actual fall samples, as shown in Formula (16):

$$Recall = \frac{TP}{TP + FN} \qquad (16)$$

As a comprehensive evaluation index that balances precision and recall, F-score is a commonly used comprehensive evaluation index for classification models, as

shown in Formula (17):

$$F - score = \left(1 + \beta^2\right) \times \frac{Precision \times Recall}{\beta^2 \times Precision + Recall} \quad (17)$$

where

TP is a positive sample predicted by the model as a positive class, which represents that the human body is falling in both the test result and real result;

TN is a negative sample predicted by the model to be a negative class, representing that the human body is non-falling in both the test result and real result;

FP is a negative sample predicted by the model to be a positive class, which represents the model test result when the human body is falling, but the real result is a falling state;

FN is predicted by the model as a positive sample of the negative class, representing that the human body is in a falling state in the model test result, but the real result is a falling state.

In Formula (16), when $\beta = 1$, it is called F1-score, and the weights of precision and recall are identical.

D. ATS assessment

Firstly, this study evaluates the performance of the original model with a time dimension of 1. Secondly, the ATS algorithm was introduced into the AT-MLP model, and no other processing was performed. The training epochs were 5000, and the performance was evaluated for the time dimensions of 3, 5, 7, 9, 11, 13, and 15. Compared with the model without the ATS, the performance of time dimensions 3 and 5 is improved by about 8.43% and 7.71%, respectively. According to the analysis of the experimental results, the time series in the ATS is shorter, which is more conducive to the capture of the fall process, as shown in Figure 5.
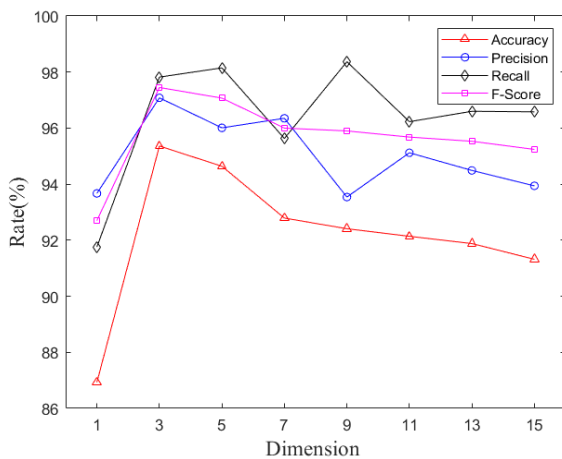

Fig. 5. Comparison of model performance in different dimensions

Although the performance of time dimensions 7, 9, 11, 13, and 15 has improved compared with time dimension 1, it has decreased compared with time dimension 3. This is because the time series is extremely long, resulting in overwhelming information, data confusion, and the network is not able to extract useful information. Figure 6(a) represents the train information of time dimension 3, and Figure 6(b) represents the test information. In the following experiments, the time dimension was set to 3.
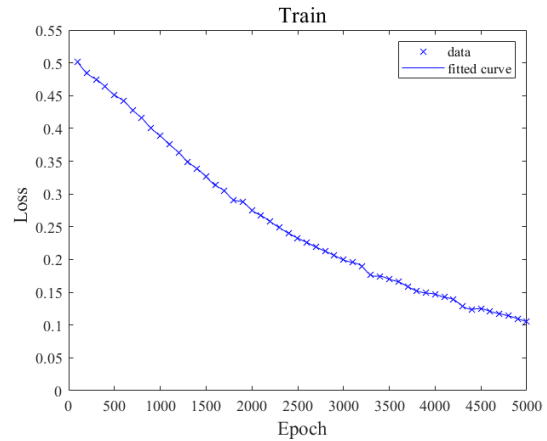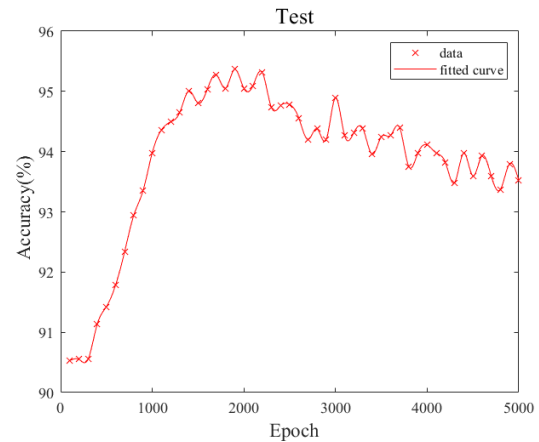

Fig. 6(a). Train loss

Fig. 6(b). Test accuracy

E. ADTS assessment

Through the comparison of features, it is found that the information capture between channels was not enough. In this study, the constructed ADTS algorithm was experimentally analyzed to enhance the ability of feature capture of the human body in the process of falling. Based on the fact that there was only an ATS algorithm in the model, the time dimension was selected as 3 to construct the ADTS algorithm. The experimental results show that after using the ADTS algorithm, the model performance was improved by about 1.29%. The experimental results are illustrated in Figure 7(a/b).
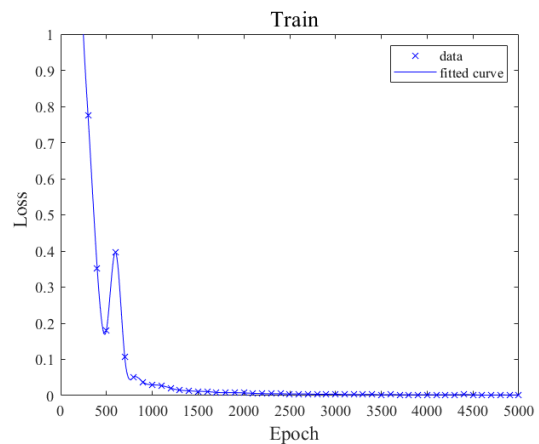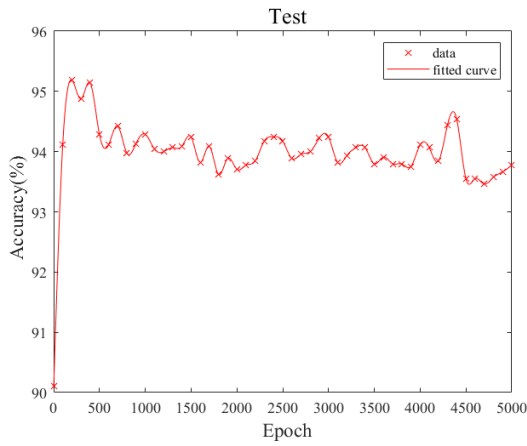

Fig. 7(a). Train loss
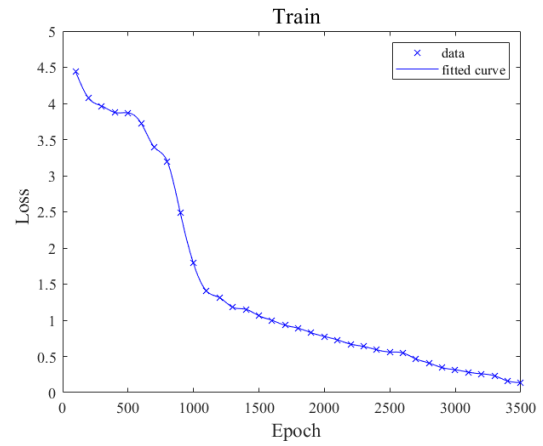
Fig. 7(b). Test accuracy

Analysis by the model loss value: When the training epoch reached 3500, the model tended to converge. To reduce the training time, the training epochs were adjusted to 3500 in subsequent experiments. Loss information is illustrated in Figure 8(a/b) when the training epoch is 3500.
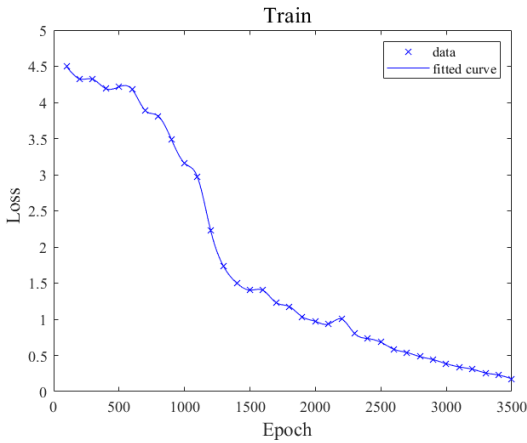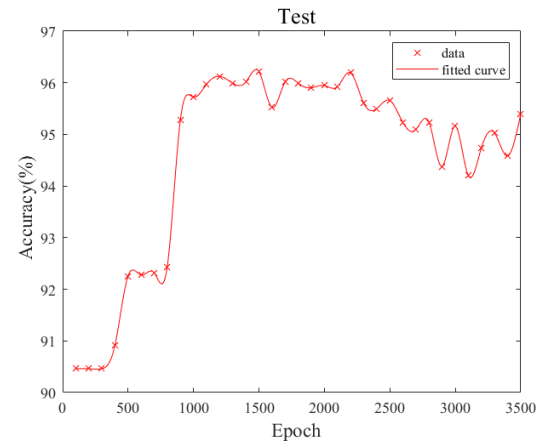

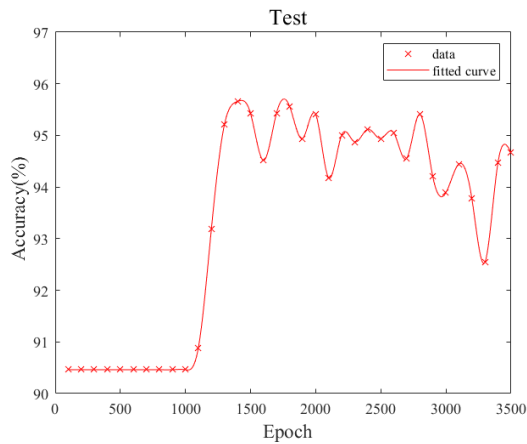Fig. 8(a). Train loss

Fig. 8(b). Test accuracy

F. AT-MLP stage assessment

To better capture the time-varying relationship on the channel, the corresponding information of *angle* and $\Delta angle$ were made clearer. CBAM blocks were added to the model to build the AT-MLP stage. The experimental results show that the model's performance was improved by about 0.56% compared to before the construction, as shown in Figure 9(a/b).

Fig. 9(a). Train loss

Fig. 9(b). Test accuracy

G. Experimental results

The effectiveness of the ATS algorithm, ADTS algorithm and AT-MLP Stage in the model were verified by four evaluation indicators, and the results are shown in Table 1. For the unimproved model, the subspace node angle was added to OpenPose, the time dimension 1 was selected for evaluation, while the accuracy rate was 86.93%. By adding the ATS algorithm, the accuracy increased to 95.36%. The impact of the ADTS algorithm on the network is further evaluated. Adding the ADTS algorithm helps to capture the posture of the human body during the falling process, and the accuracy rate increased to reach 95.65%. Finally, adding the AT-MLP stage to the model increases the accuracy to 96.21%. The performance of the proposed model is improved by 9.28% compared with the original model.

TABLE I
EXPERIMENTAL RESULTS IN THE LE2I FALL DETECTION DATASET

Approaches	Accuracy (%)	Precision (%)	Recall (%)	F-score
OpenPose+angle	86.93	93.66	91.76	0.93
OpenPose+angle+ATS	95.36	97.08	97.82	0.97
OpenPose+angle+ ATS+ADTS	95.65	93.94	96.58	0.95
OpenPose+angle+ AT-MLP	96.21	97.38	98.47	0.98

H. Compare with other methods

To verify the performance of the proposed model, this study compares it with other methods using four evaluation indexes, accuracy, precision, recall, and F-score on the same dataset LFDD. The results are shown in Table 2. The accuracy of the suggested AT-MLP model is 96.21% on the LFDD dataset, which is 11.61% superior to that of the angle combined with SVM, and the angle combined with LSTM in the comparison method. Moreover, compared with the three indexes of precision, recall, and F-score, the proposed model improved significantly. Compared with the literature [16]-[18], the accuracy rates are increased by 16.9%, 10%, and 26.98%. Simultaneously, the proposed models show a better performance in terms of precision, recall, and F-score.

TABLE II
COMPARISON OF PERFORMANCE WITH OTHER ON THE LE2I
FALL DETECTION DATASET

Approaches	Accuracy (%)	Precision (%)	Recall (%)	F-score
Adaboost classifier [17]	79.31	79.41	83.47	0.81
GMM +PCA[18]	86.21	89.13	93.00	0.91
Opticalflow + von Mises distribution [19]	69.23	69.84	94.56	0.79
Angle + Distance + Resnet50+ SVM[15]	84.60	90.00	90.00	0.9
Angle + Distance + LSTM[15]	84.60	89.00	89.00	0.89
Ours:OpenPose+Angle + AT-MLP	96.21	97.38	98.47	0.98

IV. CONCLUSION

In this study, an AT-MLP model is proposed for human fall detection, and the ATS and ADTS algorithms are designed in combination with time-series to capture the falling state and body features during the fall. Firstly, the video frame is identified based on the Open-Pose network, the node angle information of the human subspace is constructed, and the multidimensional matrix calculated by the ATS and the ADTS algorithms are concatenated. Among them, the ADTS algorithm is used to calculate the angle difference between the current and previous moments of the video frame. In the AT-MLP model, an AT-MLP stage composed of CBAM and MLP blocks is constructed, which realizes the enhanced capture of the correlation between different dimensions. In this study, experiments were conducted on the dataset LFDD to evaluate the performance of the ATS algorithm and the ADTS algorithm in different time dimensions. An experimental analysis of the overall AT-MLP model was carried out. The experimental results show that the proposed AT-MLP model has superior recognition accuracy and generalization ability. Compared with the comparison method, the AT-MLP model has significantly improved in terms of accuracy and so on.

REFERENCES

[1] Usiholo Iruansi, Jules R. Tapamo, and Innocent E. Davidson, "Classification of Power-line Insulator Condition using Local Binary Patterns with Support Vector Machines," *IAENG International Journal of Computer Science*, vol. 46, no.2, pp300-310, 2019

[2] W.Y. Lin, C.H. Chen, Y.J. Tseng, Y.T. Tsai, C.-Y. Chang, H.-Y. Wang, and C.-K. Chen, "Predicting post-stroke activities of daily living through a machine learning-based approach on initiating rehabilitation," *International Journal of Medical Informatics*, vol. 111, pp. 159-164, 2018.

[3] T. N. Gia, V. K. Sarker, I. Tcarenko, A. M. Rahmani, T. Westerlund, P. Liljeberg, and H. Tenhunen, "Energy efficient wearable sensor node for IoT-based fall detection systems," *Microprocessors and Microsystems*, vol. 56, pp. 34-46, 2018.

[4] L. Abdi, and S. Hashemi, "To combat multi-class imbalanced problems by means of over-sampling techniques," *IEEE transactions on Knowledge and Data Engineering*, vol. 28, no. 1, pp. 238-251, 2015.

[5] H. Xiao, X. Wang, Q. Li, and Z. Wang, "Gaussian mixture model for background based automatic fall detection." *International Conference on Cyberspace Technology IET*, pp. 234-237.

[6] G. Sun, and Z. Wang, "Fall detection algorithm for the elderly based on human posture estimation." *2020 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, pp. 172-176.

[7] Z. Huang, Y. Liu, Y. Fang, and B. K. Horn, "Video-based fall detection for seniors with human pose estimation." *2018 4th International Conference on Universal Village (UV)*, pp. 1-4,2018.

[8] B.H. Wang, J. Yu, K. Wang, X.Y. Bao, and K.M. Mao, "Fall detection based on dual-channel feature integration," *IEEE Access*, vol. 8, pp. 103443-103453, 2020.

[9] M. Hua, Y. Nan, and S. Lian, "Falls prediction based on body keypoints and seq2seq architecture." *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW) IEEE*, pp. 0-0, 2019.

[10] J.L. Chua, Y. C. Chang, and W. K. Lim, "A simple vision-based fall detection technique for indoor video surveillance," *Signal, Image and Video Processing*, vol. 9, no. 3, pp. 623-633, 2015.

[11] Y. Fan, M. D. Levine, G. Wen, and S. Qiu, "A deep neural network for real-time detection of falling humans in naturally occurring scenes," *Neurocomputing*, vol. 260, pp. 43-58, 2017.

[12] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Robust video surveillance for fall detection based on human shape deformation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 5, pp. 611-622, 2011.

[13] S. C. Agrawal, R. K. Tripathi, and A. S. Jalal, "Human-fall detection from an indoor video surveillance." *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT) IEEE Computer Society*, pp. 1-5. 2017.

[14] P. Sun, F. Xia, H. Zhang, D. Peng, M. A. Xi, and Z. Luo, "Research of human fall detection algorithm based on improved Gaussian mixture mode," *Computer Engineering and Applications*, 2017.

[15] B. D. Romaissa, O. Mourad, N. Brahim, and B. Yazid, "Fall detection using body geometry in video sequences." *2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1-5,2020.

[16] I. Charfi, J. Miteran, J. Dubois, M. Atri, and R. Tourki, "Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and adaboost-based classification," *Journal of Electronic Imaging*, vol. 22, no. 4, pp. 041106, 2013.

[17] M. Chamle, K. Gunale, and K. Warhade, "Automated unusual event detection in video surveillance." in 2016 *International Conference on Inventive Computation Technologies (ICICT)*, vol. 2, pp. 1–4, 2016.

[18] A.Poonsri and W. Chiracharit, "Fall detection using gaussian mixture model and principle component analysis," in 2017 9th *International Conference on Information Technology and Electrical Engineering (ICI-TEE). IEEE*, pp. 1–4, 2017.

[19] A. Y. Alaoui, A. E. Hassouny, R. Th   ami, and H. Tairi, "Video based human fall detection using von Mises distribution of motion vectors." in 2017 *Intelligent Systems and Computer Vision (ISCV). IEEE*, pp. 1–5, 2017.

HUAN ZHANG was born in Liaoning Province, P. R. China, received the B. Sc degree in Software Engineering from University of Science and Technology Liaoning, Anshan, P. R. China, in 2020.

She is currently pursuing the M. Sc degree in Software Engineering with University of Science and Technology Liaoning, Anshan, P. R. China. Her research interest is computer vision.

WENHUA CUI was born in Liaoning Province, P. R. China, received the M. Sc degree in Electromechanical control and automation from Dalian University of Technology, Dalian, P. R. China, in 1998, received the Ph.D. degree in Control Theory and Control Engineering from Dalian University of Technology, Dalian, P. R. China, in 2014,

She is currently a professor in the School of Computer and Software Engineering, University of Science and Technology Liaoning. She has published more than 20 academic papers and established more than 20 scientific research projects. The main research directions are control theory, sensor measurement and control, intelligent IoT, information security, computer network, machine vision, etc.

TIANWEI SHI was born in Liaoning Province, P. R. China, received the M. Sc degree in Control Theory and Control Engineering from University of Science and Technology Liaoning, Anshan, P. R. China, in 2010, received the Ph.D. degree in Mechatronic Engineering from Northeastern University, Shenyang, P. R. China, in 2016

He is currently an associate professor in the School of Computer and Software Engineering, University of Science and Technology Liaoning, Anshan, P. R. China. The main research directions are brain-computer interface physiological electrical signal analysis and processing, machine vision, etc.

YE TAO was born in Liaoning Province, P. R, received the B. Sc degree in Computer Science and Technology from Anshan Normal University, received the M. Sc degree in Computer Science and Technology from Northwest Normal University.

He is currently a lecturer in the School of Computer and Software Engineering, University of Science and Technology Liaoning, Anshan, P. R. The main research direction is image encryption

JIANFENG ZHANG was born in Heilongjiang Province, P. R. China, received the B. Sc degree in Automation from University of Science and Technology Liaoning, Anshan, P. R. China, in 2020.

He is currently pursuing the M. Sc degree in Control Science and Engineering with University of Science and Technology Liaoning, Anshan, P. R. China. His research interest is computer vision.