International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 8375-8384

https,//doi.org/10.70517/ijhsa463716

A self-supervised learning-based multimodal action and identity recognition model for low bandwidth conditions

Ziyang Guo[1,*]
[1] Department of Information Science and Technology, Shanghai Ocean University, Shanghai, 201316, China
Corresponding authors: (e-mail: a1002655181@163.com).

Abstract Due to the limited data transmission under low bandwidth conditions, the performance of traditional multimodal motion and identity recognition cannot be fully released. In this paper, based on the data collected under four motion modes: standing, slow walking, running and walking up and down stairs, 20-dimensional eigenvalues including three-dimensional eigenvalues and combined vector eigenvalues are calculated and analyzed to complete the selection of eigenvalues for the four human motion modes. The acquired feature values are fused into a unified spatio-temporal graph convolutional network (ST-GCN) framework to extract the global spatio-temporal features of the action from both time and space dimensions, and carry out end-to-end training. Meanwhile, in terms of model structure, the feature recalibration structure based on the attention mechanism is selected to recalibrate the shared layer features, and a multimodal action and identity recognition model based on the ST-GCN algorithm is constructed. The accuracy of this model for action recognition can be as high as 99.76% under specific sample division conditions.

Index Terms spatio-temporal graph convolutional network, feature recalibration structure, attention mechanism, action and identity recognition

I. Introduction

The 53rd Statistical Report on Internet Development in China, released by China Internet Network Information Center (CNNIC) in March 2024, shows that as of December 2023, the number of Internet users in China exceeded 1 billion, the Internet penetration rate was 77.5%, and the number of cellular Internet of Things (IoT) end-users has exceeded 2 billion [1], [2]. Among many network devices, portable smart devices, due to their portability and practicality, make users participate in increasingly diverse network activities through such devices, covering instant messaging, financial management, electronic payment, healthcare and online entertainment [3]-[5]. In the context of such a large user group and evolving online services, data security and privacy protection in smart devices pose more serious challenges [6].

Currently, most of the widely adopted identification mechanisms in smart devices still rely on traditional knowledge-based methods, such as passwords and graphical locks, where users need to rely on their memory to enter the appropriate information for authentication [7]. In addition, static biometrics, such as fingerprint scanning and iris scanning, have begun to be integrated into certain devices, which are becoming increasingly popular in personal devices such as cell phones and laptops, and biometrics such as face recognition are being used in public places such as train stations and airports to match a person's identity [8]-[10]. Compared with traditional authentication methods that rely on memory, biometrics provide a higher level of security and a smoother user interaction experience [11]. However, these biometrics also face some inherent flaws and limitations [12]. For example, when using fingerprint recognition, the user's finger is required to be in contact with the device, which means that the reliability of the system is greatly reduced in case of finger injuries or unclean surfaces [13]. Similarly, iris recognition and face recognition technologies, although they can perform effectively at a certain distance, their performance is often affected by surrounding lighting conditions or line-of-sight obstructions [14]. Meanwhile, in low-bandwidth conditions such as subway stations and high-speed railways, the instability of the network can lead to problems such as delayed or interrupted data transmission, reduced recognition accuracy, poor user experience, and increased potential security risks during biometric recognition [15]-[18]. Therefore, there is an urgent need to study a multimodal action and identity recognition method for low-bandwidth environments.

In this paper, based on human wrist gait features, we collect motion data under four motion modes: standing, slow walking, running, and walking up and down stairs, and describe the arithmetic process of 20-dimensional eigenvalues such as peak mean, standard deviation, and covariance, as a method for selecting human wrist gait eigenvalues under different motion modes. Then the overall framework of spatio-temporal graphic convolutional network (ST-GCN) is elucidated, as well as the modeling steps of the joint points of the human skeleton sequence in two different latitudes, time and space, under the framework of this algorithm. The selected human action feature values are added to the spatio-temporal graph convolutional network (ST-GCN) framework, and the process of feature recalibration based on the feature recalibration structure of the attention mechanism is analyzed to recalibrate the features, so as to propose a multimodal action and identity recognition model based on the ST-GCN algorithm. Finally, the recognition performance and learning ability of the model are evaluated.

II. Multimodal action and identity modeling

II. A. Selection of eigenvalues for human movement patterns

In the field of classification and recognition for different research contents, selecting the most representative eigenvalues is one of the keys to classification and recognition research, and the selection of eigenvalues will have a decisive impact on the subsequent classification and recognition work. There are many methods and ideas in the selection of eigenvalues, for example, in the study of human movement pattern recognition based on wrist gait feature information, the data is a set of continuous fluctuation data with time sequence, and its data volume is very large, so before extracting the eigenvalues, it is necessary to first determine the eigenvalue selection method. The common methods are time domain analysis and frequency domain analysis.

Frequency domain analysis refers to the conversion of time domain signals into frequency domain signals through the Fourier transform to analyze and evaluate the signals from the frequency domain perspective, which is mostly applied to the research of the system with high demand for stability and dynamic characteristics. The motion gait is a quasi-periodic state, and there will be slight differences in the movements in the two adjacent cycles, and its stable characteristics are shown in a certain long time detection, while the dynamic characteristics of the gait need to be obtained through the analysis of the data of the test time is long. The time domain analysis method refers to the evaluation and analysis of the stability, transient characteristics and steady state characteristics of the system as a whole according to the time domain expression of the input variables. For the essential characteristics that gait has, the gait characteristic values can be better extracted when the wrist gait characteristic data are analyzed using time domain analysis. According to the feature analysis, the mean value, peak mean value, standard deviation, covariance and skewness of the data can be extracted as the eigenvalues in the classification and recognition of movement patterns.

(1) The mean value represents a measure of the trend in the test data set. In gait features, the mean value represents a constant level presented by the swing speed and movement amplitude of the wrist during human walking. By taking the mean value, we can get the convergence level of the wrist gait feature data under each motion mode, and use it to determine the pattern attribution of the feature data in the motion pattern classification and recognition experiment.

(2) The peak value indicates the maximum instantaneous value of the fluctuation data during the test time, while the peak mean value refers to the peak concentration trend in all the test data, which represents the size of the variation range of the wrist gait data. The larger the peak value, the larger the wrist acceleration, the larger the range of motion, and the larger the trend of wrist rotation.

(3) Standard deviation is a commonly used quantity in probability statistics for the distribution of statistical data, and the greater the dispersion of a set of data, the greater its standard deviation. From the motion characterization, it is known that the acceleration and angular velocity of the wrist in different modes have more obvious differences, so the standard deviation of the data has a strong characteristic representativeness. The standard deviation calculation formula is shown in equation (1):

$$\sigma(x)=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})^2} \tag{1}$$

where σ is the standard deviation, N is the sample size of the data, and $\bar{X}$ is the mean of the data sample.

(4) Covariance is often used to calculate the overall error between two sets of variables assessed, representing the correlation between two sets of data. When performing motion characterization, it can be seen that the distribution of acceleration of up and down stairs motion is more similar to that of the slow walking state, but there are some differences in the distribution of angular velocity, so the up and down stairs state can be identified by calculating the covariance value of acceleration and angular velocity. The covariance calculation formula is equation (2):

$$cov(a,\omega)=\frac{\sum_{i=1}^{N}(a_i-a_{avg})(\omega_i-\omega_{avg})}{N-1} \tag{2}$$

N is the number of samples of the data, and the combined values of acceleration a_i and angular velocity ω_i are a_{avg} and ω_{avg}, respectively.

(5) Kurtosis is used to measure the kurtosis of the probability distribution of a random variable in an experiment. For the study of motor gait, kurtosis represents the slope of the peak point of the data fluctuation, and the greater the acceleration kurtosis, the greater the degree of intensity of the motor state. The extraction of kurtosis value can effectively realize the accurate classification of running mode and other three motion modes, and the kurtosis calculation formula is equation (3):

$$K = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^4 f_i}{N\sigma^4} \tag{3}$$

where N is the number of data samples, $\bar{X}$ is the mean value of data samples, f_i is the sample interval of data, and σ is the standard deviation of data samples.

According to the motion characterization, the motion data of the standing mode basically did not fluctuate significantly, while the other three motion modes would have different wrist gait data due to different motion influencing factors. For the collected and pre-processed experimental data, three-dimensional eigenvalues including peak mean, standard deviation and covariance of acceleration and angular velocity as well as acceleration kurtosis and combined vector eigenvalues, totaling 20-dimensional eigenvalues, can be obtained according to the above calculation formula. The selected eigenvalues exhaustively characterize the acceleration and angular velocity between different motion patterns, which provides a basic guarantee for the efficiency and accuracy of the subsequent motion pattern recognition. Distribution calculations were performed for the distribution of the 20-dimensional eigenvalues.

II. B.Multimodal action and identity recognition model based on ST-GCN algorithm

II. B. 1) Spatio-Temporal Graph Convolution ST-GCN

In ST-GCN, the human skeleton map is captured by a pose estimation algorithm or kinectis camera and transformed into joint point coordinate information, which in turn constructs the corresponding connectivity relationships based on the skeletal data. The constructed sequence data of human joint point maps are inputted into the spatio-temporal graph convolution network ST-GCN, and graph convolution in both spatial and temporal dimensions is performed to extract more advanced feature maps. Finally, the final results are output by classifying them through fully connected layers, classifiers, etc. The overall flow of ST-GCN recognition is shown in Fig. 1.

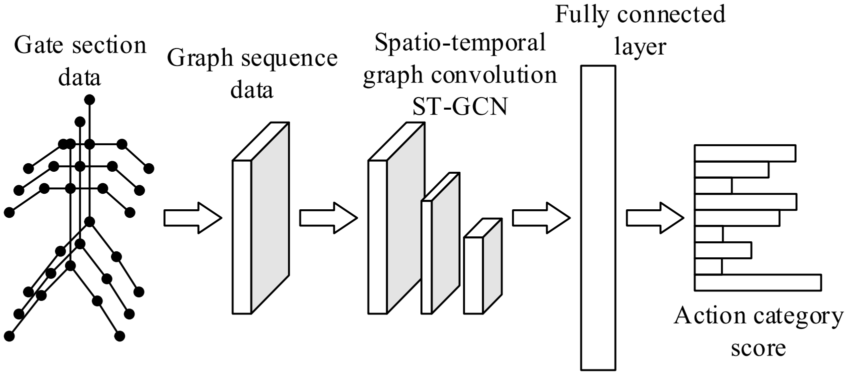


Figure 1: ST-GCN identification process

The joints in the skeleton sequence are modeled in two different dimensions, time and space, respectively, as shown in Fig. 2. The dark gray dots represent the human joints under the current frame, while the light gray dots represent the human joints under the adjacent frames (i.e., the frame before or the frame after the current frame), and the blue lines represent the realistic connections among the human joints, e.g., the connections between the head and the neck, the left hand and the left shoulder, the right hand and the right shoulder, and so on. The green lines represent the temporal connections between the same human body joints in neighboring frames, such as the connection between the position of the right hand node in the previous frame, the position at this time, and the position in the later frame. Translated in mathematical language, this can be expressed as inputting a sequence of human body joints (N,T), where N denotes N joints of the human body and T denotes the length of the input sequence. An undirected graph $G=(V,E)$ is constructed from this, with V representing the set of graph nodes, i.e., equation (4):

$$V=\left\{v_{ti}\mid t=1,2,\cdots,T,i=1,2,\cdots,N\right\} \tag{4}$$

E represents the set of edges, which consists of two parts, E_s and E_t, where E_s denotes the realistic connectivity of human joints on the current frame, and E_t denotes the temporal connectivity of the same joints under different frames.

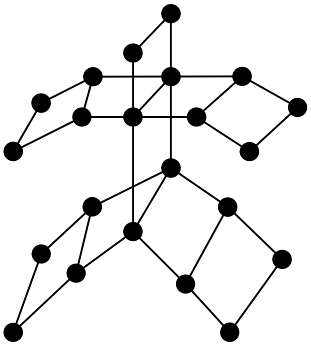


Figure 2: Time - Space diagram

From the well defined undirected graph G in the above paragraph, the graph convolution operation is then defined under the current frame (at the spatial dimension level). For the node $V_{\tau i}$ on the τ frame, it can be expressed as equation (5):

$$f_{out}(v_{\tau i})=\sum_{v_{\tau j}\in s_i}\frac{1}{|T_{ij}|}f_{in}(v_{\tau j})w\left(l(v_{\tau j})\right) \tag{5}$$

where: v represents the node of G, f_{in} represents the feature mapping, s_i is the sampling region of the convolution of the target node $v_{\tau i}$, the weight function w is used to provide the weight vector, the mapping function l assigns the weights to the feature vectors, and the subsets contained in s_i are different in size. O denotes the skeleton center of gravity, the red region is s_i, and s_i consists of three subsets: s_{i1} is the target node itself (red circle), s_{i2} is the set of centripetal nodes (yellow circle), and s_{i3} is the set of centrifugal nodes (blue circle), and each subset has its own label and mapping l, and $|T_{ij}|$ denotes the vertex $v_{\tau j}$ is in the base of the s_i subset.

A transformation of the formula yields the graph convolution realized in the spatial dimension as Eq. (6):

$$f_{out}=\sum_{k=1}^{K_t}w_k\left(f_{in}\left(\tilde{A}_k \square M_k\right)\right) \tag{6}$$

where K_t denotes the size of the convolution kernel, $\tilde{A}$ is the normalized form of the adjacency matrix A, M is a learnable weight matrix, and the □ notation denotes the dot product.

The graph convolution is shown in Figure 3.

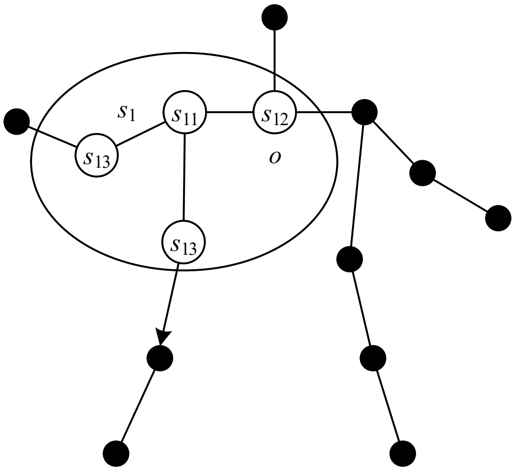


Figure 3: Graph conwolutional

Next define the graph convolution on neighboring frames (i.e., one frame before or one frame after the current frame). For each vertex v_{ti}, centered on it, looking one frame forward and one frame backward, there are and only those two corresponding identical joints, which means that in the whole sequence of human joints, each human joint has two fixed neighbor nodes from the time dimension. Therefore, only two-dimensional convolution of the feature graph output from the model is needed to complete the graph convolution operation in the time dimension.

In the overall structure of the original ST-GCN network model, the whole network contains 10 layers of ST-GCN modules, except for the first ST-GCN module, the last nine ST-GCN modules include not only graph convolution and time convolution modules, but also residual networks. After the feature map is processed by Polling layer and FC layer, it enters the Softmax classifier and finally outputs the result.

II. B. 2) Structure of feature recalibration based on attention mechanism

The essence of the attention mechanism is to allocate processing resources to more valuable input information and suppress useless information to improve the efficiency and accuracy of task processing. The compressive excitation (SE) module is an attention weighting operation based on feature channels, which can filter the feature information to obtain more effective shared feature representations to improve the multi-task recognition effect, and its structure is shown in Fig. 4, where X denotes the input feature map, and Y denotes the output feature map.The SE module extracts global information on the feature map output from the convolutional layer (Conv), and accordingly establishes the dependency relationship between each channel, and adaptively generates the attention weights in the learning process, so as to strengthen and suppress the information of each channel and realize the recalibration of features.

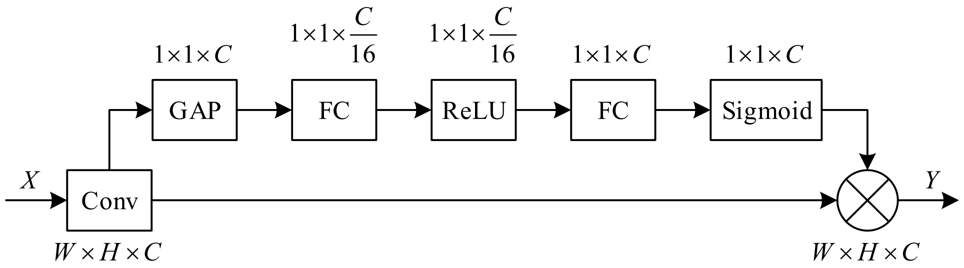


Figure 4: SE module structure

The input feature map X (dimension $W \times H \times C$) goes through Conv, and then sequentially goes through GAP (Global Pooling), which characterizes the 2D features of each channel of the feature map as real values, and compresses the feature map of each channel. Then the input feature dimension is reduced to 1/16 of the original by FC (Fully Connected Layer), ReLU (Revised Linear Unit) is activated, and the second FC upgrades the features to the original input dimension, and finally the sigmoid activation function generates the weights of each channel, which is multiplied and weighted with the convolved feature map to obtain the output feature map Y (with dimension $W \times H \times C$).

The SE module performs global pooling (GAP) on the output feature map of the convolutional layer with dimension $W \times H \times C$, i.e., the 2D features on each channel of the feature map are characterized as a real value by global averaging to compress the feature map for each channel. The dimensionality of the resulting output features corresponds to the number of channels C of the feature map, and the output value on each channel has a global receptive field that represents the global feature information of that channel.

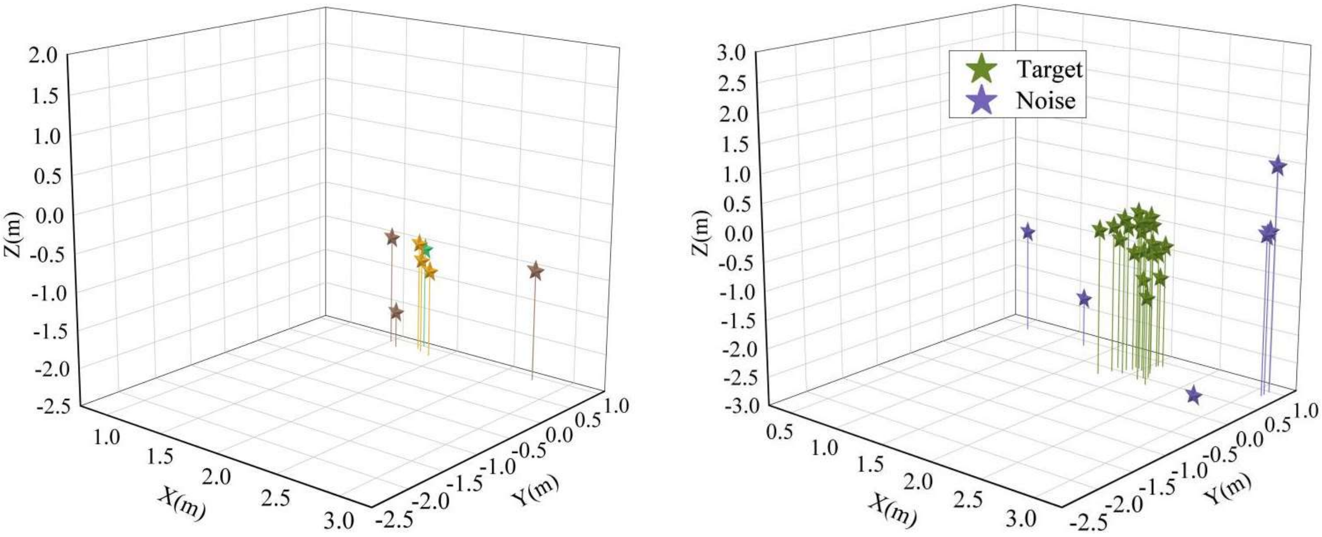
The structure also introduces a gate excitation mechanism derived from recurrent neural networks, which learns the non-mutually exclusive relationship between channels through a fully connected layer and uses the activation function as a gate control for obtaining the excitation on each channel.The SE module reduces the dimension of the input features to 1/16 of the original dimension using one fully connected layer, and after activating it using a modified linear unit (ReLU), it utilizes a second fully connected layer to dimension up to the original input dimension. The bottleneck structure formed by the two fully connected layers reduces the amount of model parameters in the feature weight generation process, reduces the computational complexity, and increases the nonlinearity of the model. A sigmoid activation function is used to generate the weights of each channel at the end of the network, and the value of this weight is related to the learnable model parameters, and the feature weight correction is done automatically by multiplicatively weighting the convolved feature map.

III. Recognition Performance Test and Learning Ability Evaluation of Models

III. A. Network Recognition Performance Analysis of ST-GCN Algorithm

III. A. 1) Multi-frame fusion and clustering

During the action data acquisition, the execution time of each action is 2 to 3 s. The point cloud data of Vayyar radar is 40 to 60 per frame, and the point cloud data of TI radar is 4 to 20 per frame. Compared with the Vay-yar millimeter wave radar, the point cloud of the TI millimeter wave radar is more sparse. In order to obtain better recognition, 6 frames (0.2s) of TI radar point cloud data were fused to enhance the quality of the point cloud data. Multi-frame fusion can improve the data quality of the point cloud, better access to the spatial geometric features of the human body movements, and provide richer and more intuitive feature information for subsequent human movement recognition. Figure 5 shows the effect of point cloud data fusion before and after clustering for TI millimeter wave radar. The green points in Fig. 5(b) represent the effective point cloud, and the purple points indicate the noisy points that have not been clustered. From the figure, it can be seen that through the multi-frame fusion and clustering process, the single-frame point cloud is changed from several points to dozens of points, and the interference points are filtered out, and the quality of the point cloud data is significantly improved, which is conducive to the later use of consecutive multi-frame data for human action recognition.



(a) Point cloud in single frame

(b) Point cloud after fusion clustering

Figure 5: Comparison of the TI radar point cloud after multi-frame fusion and clusterin

III. A. 2) Impact of different fusion frame numbers on network recognition performance

In order to evaluate the effect of different fusion frame numbers on the recognition performance of the proposed model in this paper, the data from TI radar are fused for 3, 6, 9 and 12 frames, corresponding to fusion frame periods of 0.1 s, 0.2 s, 0.3 s and 0.4 s. In order to keep the same fusion frame period as that of the TI data, the data from Vayyar radar are fused for 1, 2, 3, 4 frames, respectively, in this paper. The fused set of data is referred to as the post-fusion frame. All experiments use 10 consecutive post-fusion frames as a sample, and the results of the recognition accuracy experiments at different numbers of fusion frames are shown in Fig. 6. It can be observed from the figure that the data of TI radar has the highest recognition accuracy of 96.15% at the number of fusion frames 6. And for the data of Vayyar radar, the highest recognition accuracy of 99.23% is achieved when the number of fused frames is 2. When the number of fused frames is greater than 2, the recognition accuracy slightly decreases. It can be seen that the selection of the number of fusion frames has a significant effect on the recognition accuracy, and the optimal number of fusion frames varies for different datasets. Therefore, in the subsequent experiments, the number of fusion frames is set to 6 for the TI radar and 2 for the Vayyar radar, i.e., the period of fusion frames is 0.2s for both of them.

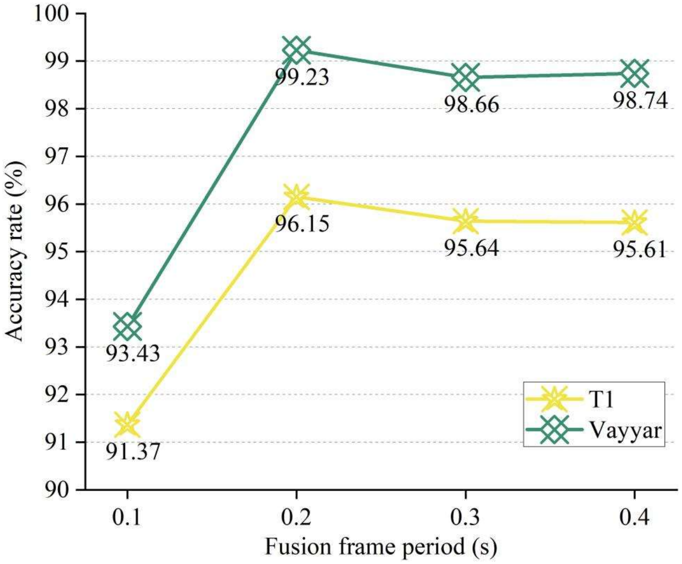


Figure 6: Accuracy with different fused frame

III. A. 3) Effect of different sample divisions on network recognition performance

In order to investigate the effect of different sample divisions on the recognition performance of the model network proposed in this paper, experiments were conducted with different numbers of fused frames within a single sample, and the experimental results of the recognition accuracy with different sample divisions are shown in Fig. 7. The experimental results show that, for the TI dataset, with the increment of the number of fused frames in the samples, the recognition accuracy is gradually increased and reaches a maximum of 99.59% when the sample contains 30 frames, and then decreases slightly. It then decreases slightly. For the Vayyar dataset, the recognition accuracy also increases with the number of fused frames and peaks at 20 frames. It should be noted that the recognition accuracy increases substantially when the sample contains 10 frames of data. This is due to the fact that the average duration of each action is about 2 seconds (about 10 fused frame cycles). If the number of sample frames is less than 10, the complete course of the action may not be fully captured, which in turn affects the recognition accuracy. However, the recognition accuracy reaches a maximum of 99.76% when the number of sample frames is increased to 20 fusion frames (approximately 3 seconds), which at this point contains approximately 2 full cycles of action execution. In practice, the selection of the number of fusion frames within a single sample needs to balance action completeness with computational efficiency. Therefore, in subsequent experiments, the number of post-fusion frames for a single sample was set to 20.

III. B. Attention Mechanism Learning Experiment Results and Analysis

III. B. 1) Data analysis and pre-processing

After filtering and normalizing the data, the angular changes of the hip and shoulder joints of Volunteer 1 during the same time period were selected for analysis. Among them, the hip and shoulder joints showed periodic patterns of change in all three degrees of freedom of motion. However, there were significant differences in the waveform trends of the joint angles in different parts of the body, especially in the abduction and adduction directions of the hip and shoulder joints. The visualization of feature dimensionality reduction using ST-GCN is shown in Fig. 8, where it can be seen that the orange and cyan points form two separated clusters. This shows that the ST-GCN algorithm is able to distinguish between the two different joint angle data and they have more distinctive features in the high dimensional space. There is a clear spatial separation between the two clusters, which indicates that the data of the hip and shoulder joints have more distinctive differences in the multidimensional space.

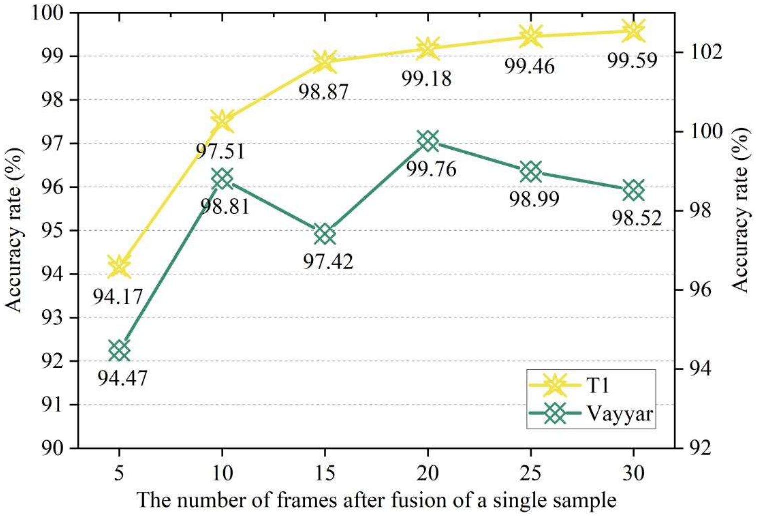
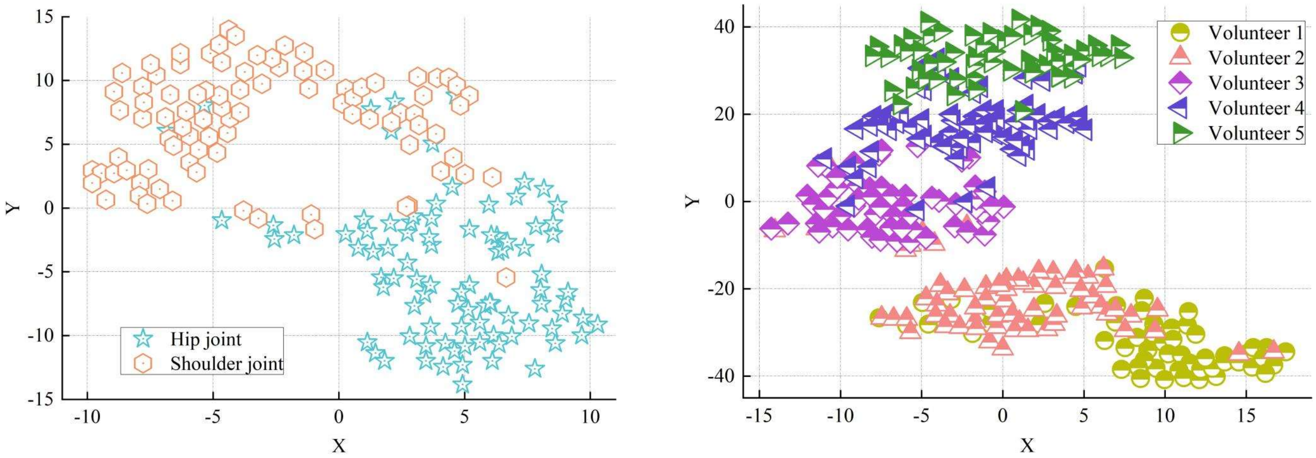


Figure 7: Accuracy with different sample divisions



(a) Hip joint and shoulder joint

(b) Different volunteer gaits

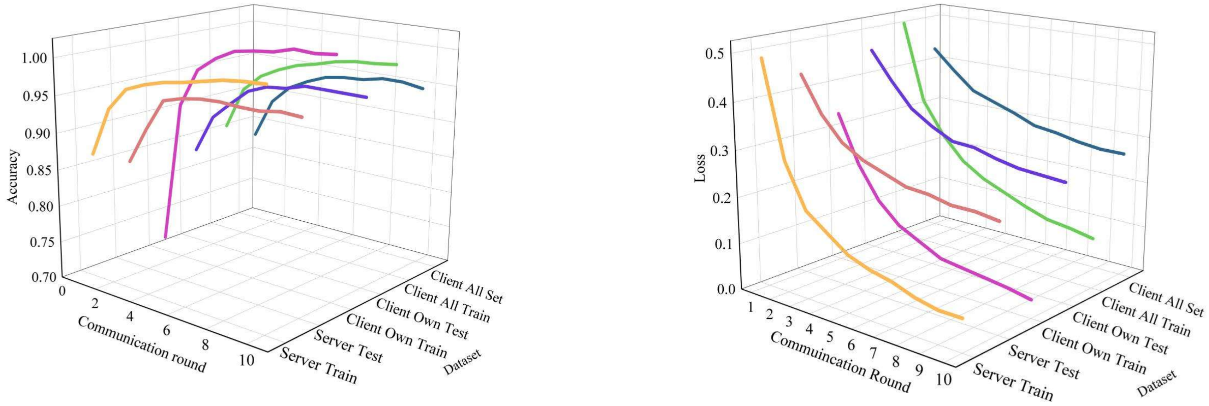
Figure 8: Dimensionality reduction visualization of the original data

Figure 8(b) shows the ST-GCN visualization results of the hip gait data of five volunteers. The different colored points in the figure represent the gait data of different volunteers, and the data of each volunteer formed independent clusters in the low-dimensional space. The high degree of aggregation of data points within each cluster indicates that the gait data of the same volunteer maintains a high degree of consistency in the high-dimensional feature space. The distinct spatial intervals between clusters of different colors revealed significant differences in gait characteristics between individual volunteers, effectively showing the differences between individuals in the multidimensional feature space in the low-dimensional space. However, it is also noted that there is a partial overlap between the data of Volunteer 1 and Volunteer 2, which implies that there is a similarity between the gait characteristics of these two volunteers in some dimensions. Due to a certain degree of randomness in the ST-GCN algorithm, it did not completely distinguish the gait feature differences between all individuals.

III. B. 2) Comparison of different parameter settings

Note that multiple parameters are involved in mechanism learning, and different parameter settings directly affect the training process and final performance of the model. Parameters such as local batch size, local learning rate and local training rounds are obtained for multiple experiments. For the number of local training rounds, the hipGait dataset is used, 5 clients are set up, each client has 2 participants, and the number of communication rounds is set to 10.

The training process for the hipGait dataset with local training rounds of 10 is shown in Fig. 9. In the results presented in Fig. 9(a), it is observed that the clients present high accuracy rates when training the model on their local dataset, all of them reaching 93.00% and above after 10 training rounds. This phenomenon is usually due to the relatively small amount of local data and the fact that the client's model is directly overfitted to these data, which allows the model to adapt well to its training data, but when faced with a wider range of or unseen data, the model's generalization ability may be insufficient, which leads to unsatisfactory accuracy rates on the independent test set. In contrast, in the loss value curve demonstrated in Fig. 9(b), the client shows good convergence performance when training the model on its local dataset, with the lowest loss value converging to 0.00 after 10 training sessions.



(a) Accuracy curve

(b) Loss rate curve

Figure 9: The local training round of the hipGait dataset is a 10-hour training process

III. B. 3) Comparison of different client number settings

Under the attention mechanism learning framework, the number of clients and their data volume distribution have a significant impact on the learning performance. However, due to various constraints and scenario limitations in practical applications, clients usually collect data independently, making this ideal data distribution difficult to realize. To address this practical challenge, experiments are conducted by setting up different numbers of clients and making each client hold a different amount of user identity data when conducting gait identity research. In the hipGait dataset, there were 24 participants whose data were randomly assigned to 2, 3, and 5 clients, where in the three-client setup, 8 users were assigned to each client. For the whuGait dataset, it contains 120 participants whose data are randomly assigned to 2, 7, 11, 18, 43, and 66 clients.

The statistical characteristics of the data from the 12 clients are shown in Figure 10, where it can be observed that there is a significant imbalance in the distribution of the amount of data among the clients between [2420,3469]. Such data distribution may have an impact on the training effect of the model in this paper.

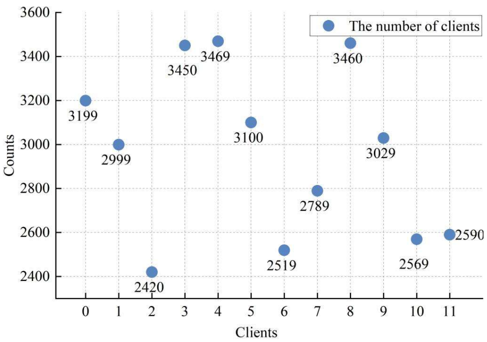


Figure 10: Data statistical characteristics of 12 clients

IV. Conclusion

In this paper, after extracting human wrist gait eigenvalues based on different motion pattern features, the eigenvalues are incorporated into a spatio-temporal convolutional network framework, and the attention mechanism is used to recalibrate the features of the task-sharing layer to establish a multimodal action and identity recognition model based on the ST-GCN algorithm. In the training and analysis of the network recognition performance, the model achieves a recognition accuracy of 96.15% when the number of fused frames of TI radar data is 6, 99.23% when the number of fused frames of Vayyar radar data is 2, 99.59% when the sample of TI dataset contains 30 frames, and 99.59% when the number of fused frames of Vayyar dataset sample is increased to 20. The recognition accuracy is as high as 99.76% when fused frames are included. In addition, the multimodal action and identity recognition model based on the ST-GCN algorithm shows excellent action recognition performance and convergence performance in the attention mechanism learning experiments when trained on the local dataset with the accuracy stabilized at 93.00% and above after 10 trainings, with the lowest loss value converging to zero.

Funding

This work was supported by Key R&D Program of Shandong Province: Action plan to boost scientific and technological innovation in rural revitalization (No: 2023TZXD051).

Disclaimer/Publisher's Note

References

[1] Hong, J. (2025). The Chinese Netizens in the New Chinese Society. In China's Internet in the 2000s: Challenges, Dilemmas, and Battles (pp. 139-150). Singapore: Springer Nature Singapore.

[2] Wang, J., Hu, Y., & Xiong, J. (2024). The internet use, social networks, and entrepreneurship: evidence from China. Technology Analysis & Strategic Management, 36(1), 122-136.

[3] Zhu, N., Zhao, W., Yan, X., Shen, Y., & Ma, Y. (2023). Intelligent wearable and portable security detection technologies integrated into protective equipment: A review. Journal of Industrial Textiles, 53, 15280837231189890.

[4] Sodhro, A. H., Awad, A. I., van de Beek, J., & Nikolakopoulos, G. (2022). Intelligent authentication of 5G healthcare devices: A survey. Internet of Things, 20, 100610.

[5] Lu, Y., Wang, D., Obaidat, M. S., & Vijayakumar, P. (2022). Edge-assisted intelligent device authentication in cyber–physical systems. IEEE Internet of Things Journal, 10(4), 3057-3070.

[6] Zheng, D., Alkawaz, M. H., & Johar, M. G. M. (2025). Privacy protection and data security in intelligent recommendation systems. Neural Computing and Applications, 1-18.

[7] Shen, S. S., Kang, T. H., Lin, S. H., & Chien, W. (2017, May). Random graphic user password authentication scheme in mobile devices. In 2017 International conference on applied system innovation (ICASI) (pp. 1251-1254). IEEE.

[8] Wechsler, H. (2012). Biometric security and privacy using smart identity management and interoperability: Validation and vulnerabilities of various techniques. Review of Policy Research, 29(1), 63-89.

[9] Liang, Y., Samtani, S., Guo, B., & Yu, Z. (2020). Behavioral biometrics for continuous authentication in the internet-of-things era: An artificial intelligence perspective. IEEE Internet of Things Journal, 7(9), 9128-9143.

[10] Rui, Z., & Yan, Z. (2018). A survey on biometric authentication: Toward secure and privacy-preserving identification. IEEE access, 7, 5994-6009.

[11] Bhuva, D. R., & Kumar, S. (2023). A novel continuous authentication method using biometrics for IOT devices. Internet of Things, 24, 100927.

[12] Shende, S. W., Tembhurne, J. V., & Ansari, N. A. (2024). Deep learning based authentication schemes for smart devices in different modalities: progress, challenges, performance, datasets and future directions. Multimedia Tools and Applications, 83(28), 71451-71493.

[13] Riaz, I., Ali, A. N., & Ibrahim, H. (2024). Loss of fingerprint features and recognition failure due to physiological factors-a literature survey. Multimedia Tools and Applications, 83(39), 87153-87178.

[14] Mudunuri, S. P., & Biswas, S. (2015). Low resolution face recognition across variations in pose and illumination. IEEE transactions on pattern analysis and machine intelligence, 38(5), 1034-1040.

[15] Forssell, H., Thobaben, R., Al-Zubaidy, H., & Gross, J. (2017, December). On the impact of feature-based physical layer authentication on network delay performance. In GLOBECOM 2017-2017 IEEE Global Communications Conference (pp. 1-6). IEEE.

[16] Wang, D., Cheng, H., He, D., & Wang, P. (2016). On the challenges in designing identity-based privacy-preserving authentication schemes for mobile devices. IEEE Systems Journal, 12(1), 916-925.

[17] Zukarnain, Z. A., Muneer, A., & Ab Aziz, M. K. (2022). Authentication securing methods for mobile identity: Issues, solutions and challenges. Symmetry, 14(4), 821.

[18] Xie, H., & Zhao, J. (2015). A lightweight identity authentication method by exploiting network covert channel. Peer-to-peer Networking and Applications, 8, 1038-1047.