Application of Sanda-Assisted Teaching System Integrating VR Technology from a 5G Perspective

Zhaoquan Zhang*, Yong Ding
Police Command and Tactical College, Guangdong Police College, Guangzhou 510232, China

***Abstract*—Since the focus of Sanda teaching is to allow students to master martial arts techniques through confrontational exercises, Sanda teaching in colleges and universities commonly adopts contextualized teaching methods. However, this Sanda teaching method suffers from deficiencies such as poor teaching effectiveness and difficulty in reflecting the effectiveness of Sanda martial arts. In order to solve these problems and make up for the shortcomings of offline Sanda teaching, the study adopts the virtual reality technology and the fifth generation mobile communication technology to construct a Sanda-assisted teaching system for college students. In order to ascertain whether students complete the Sanda movement practice, the study has designed two models: a self-supervised model based on acceleration and angular velocity contrast learning and a multi-task semi-supervised model based on time-frequency contrast learning. These models aim to improve the analytical function of the Sanda-assisted teaching system and address the analytical deficiencies of the existing human movement identification algorithms. The results indicated that the maximum accuracy of the research-designed self-supervised model was 95.76% and 95.89% on the training and test sets, respectively. The multi-task semi-supervised model designed in the study plateaued after nearly 22 and 24 iterations on the training and test sets, respectively. The average response time of the research-designed system was 59ms, and the throughput could reach a maximum of 77651bit/s. The model and the research-designed system both worked well, and they can lower the risk of student injuries while offering technological support for Sanda-assisted teaching and learning in higher education institutions.**

Keywords*—VR technology; Sanda; teaching system; motion recognition; feature extraction

I. INTRODUCTION

Sanda, as an excellent traditional sport of the Chinese nation, contains valuable national spirit, reflects strong national traditional colors, and has strong vitality and epochal character. University Sanda teaching (UST) can promote the development of Sanda sport in the new era and make it glow with more beautiful colors [1-2]. Most colleges and institutions have embraced the contextual teaching style since it aligns with Sanda teaching's goal. However, this Sanda teaching method has shortcomings such as poor teaching effect and difficulty in reflecting the effectiveness of Sanda martial arts [3]. In order to solve this problem, constructing Sanda-assisted teaching system (SATS) becomes particularly important. Virtual reality (VR) has been progressively incorporated into the domains of education, healthcare, gaming, and entertainment with the advancement of computer technology [4]. Feng and other specialists created a model based on the combination of VR technology and Web application design to increase students' enthusiasm to learn. They then used the model's data to teach physical education (PET). The findings demonstrated how well VR and PET work together to increase students' excitement for athletics and interest in studying [5]. Gao and other researchers designed a medical system based on remote VR technology to address the problem of medical informatization. The system contained two major functional sections, namely, consultation management and system management. Among them, the consultation management section included all remote business consultation processes, and the system management version involved managing medical resources and user data. The findings demonstrated significant time and cost savings in the areas of remote VR technology adoption, safety, and rehabilitation-improvements of 85%, 92%, and 96%, respectively [6]. To help architecture students learn more effectively, researchers like Elgewely created a VR platform for architectural details based on VR technology and building information modeling. Additionally, the platform was validated in three main areas, namely pilot, system usability and immersion and learning gains. The results showed that students' learning progress increased by 30% after experiencing the VR environment [7]. Experts such as Almousa designed a VR-based virtual clinical simulation system with Oculus Quest headset in order to promote global academic collaboration. In addition, the system was able to connect and communicate in real time with an instructor control panel application. The results showed that this virtual clinical simulation system was able to provide realistic clinical training in a virtual space simulating a hospital environment and promote academic collaboration [8].

In SATS, the most important thing is to judge whether the student has completed the exercise or not, and this requires the use of human motion recognition (HMR) algorithm model. Common HMR methods include traditional machine learning, such as K-nearest neighbor classification algorithms, self supervised learning (SSL) methods, semi supervised learning, and wearable device based methods [9-10]. To solve the problem that the existing HMR-based algorithms cannot fully explore the spatio-temporal properties of motion, Yangzhi and other researchers designed a HMR algorithm based on a spatio-temporal attention graph convolutional network model that integrates spatial and temporal attention mechanisms. The comparison results show that the algorithm designed in the study has improved the accuracy of Top-1 and Top-5 by 5.0% and 4.5%, respectively [11]. To increase the precision of video HMR and the computational effectiveness of large-scale datasets, Gao and other specialists created a motion capture

multidimensional data model and deep learning framework based video image motion recognition. In addition, the study also used Gaussian mixture model and gradient histogram. The outcomes indicated that the average value (AV) of the classification accuracy of the method was 85.79% and the maximum running speed was 20 frames per second [12]. To improve the accuracy of the rehabilitation robot on HMR and to reduce the recognition time, Chen and other researchers designed an improved sparrow search algorithm based on multiple strategies. According to the results, this upgraded algorithm's recognition accuracy was 2.835% higher than that of the original classifier, which should make it easier for the rehabilitation robot to understand the intention behind a person's movements [13]. Ye and colleagues developed an enhanced time-slotted video down sampling technique based on Gaussian model and proposed a human interaction recognition algorithm based on parallel multi-feature fusion network to identify human actions. Convolutional kernels with various scales were employed in the study to extract features. The findings demonstrate that the approach can identify six interaction acts with an accuracy of 88.9% [14].

In summary, the current studies on VR-based systems and HMR are relatively rich and use a variety of methods. However, these studies also have certain shortcomings, such as not fully considering the complexity and specificity of human motion sensing signals, and still facing the problems of underutilization of human motion sensing data and difficulty in feature extraction. Therefore, in an attempt to compensate for the shortcomings of Offline Sanda teaching (OST), the study designed a SATS based on VR and 5th generation mobile communication technology (5G), constructed self-supervised model with contrastive learning of acceleration and angular velocity (SSMCLAA) and semi-supervised multi-task model with time-frequency (TF) contrastive learning (SSMTFCL). The study aims to improve the speed and accuracy of HMR, accelerate the judgment of whether students complete Sanda exercises, reduce the probability of student injuries, and improve the effectiveness of UST. The innovativeness of the study is that it combines 5G technology and VR technology to alleviate the problem of insufficient data labeling, and realize the consistency of TF characteristics of data and the improvement of HMR speed and accuracy.

There are five sections to the study overall. Section II is the design of the research methodology, which includes the construction of SATS based on 5G and VR technologies, the design of SSMCLAA model and SSMTFCL model. Section III is the performance validation of SATS, SSMCLAA model and SSMTFCL model. Discussion is given in Section IV. Section V is the conclusions, shortcomings and future prospects of the study.

II. Methods and Materials

To address the problems of OST, the study designed SATS based on 5G and VR technologies, and designed the structure and functions of teaching system (TS). In order to realize the judgment of whether students complete the exercises in TS, the research adopts the HMR algorithm, and designs the corresponding recognition model for the problems existing in the current HMR, respectively.

A. SATS Design Based on 5G and VR Technologies

UST suffers from deficiencies such as poor teaching effectiveness, difficulty in reflecting the effectiveness of Sanda martial arts, and the frequency of accidental injuries or even serious permanent injuries to students. To compensate for the shortcomings of OST, the study adopted 5G and VR technologies. Since VR technology can simulate real-world situations in three dimensions, it has been included into college and university curricula. The majority of these institutions have started to create VR teaching laboratories [15]. 5G technology has ultra-high speed rate, which can quickly enjoy 360° panoramic VR and display more high-definition VR images [16]. The structure of SATS is shown in Fig. 1.

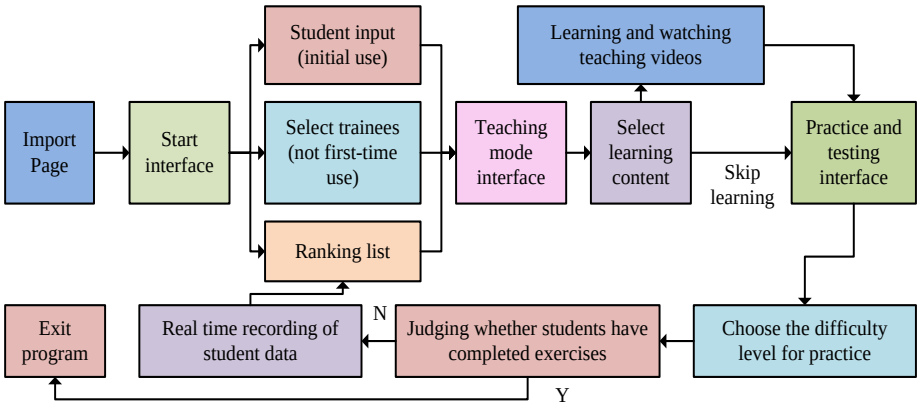


Fig. 1. The structure of sanda-assisted teaching system.

In Fig. 1, the structure of SATS involves import page, start interface, ranking list, student input (initial use), select trainees (not first-time use), teaching mode interface, select learning content, learning and watching teaching videos, skip learning, practice and testing interface, choose the difficulty level for practice, judging whether the difficulty level is too high or too low level for practice, judging whether students have completed exercises, real time recording of student data, and exit program. SATS involves four main functions, which are action demonstration function, practice and test function, interaction function, and data analysis function. Among these, the action demonstration function requires 5G technology to

guarantee that students can view the video with clarity and fluidity without experiencing any latency. Both the practice and test function and the interaction function need to allow students to follow the instructions on the VR interface to practice and get feedback from the system. The composition of the interaction function is shown in Fig. 2.

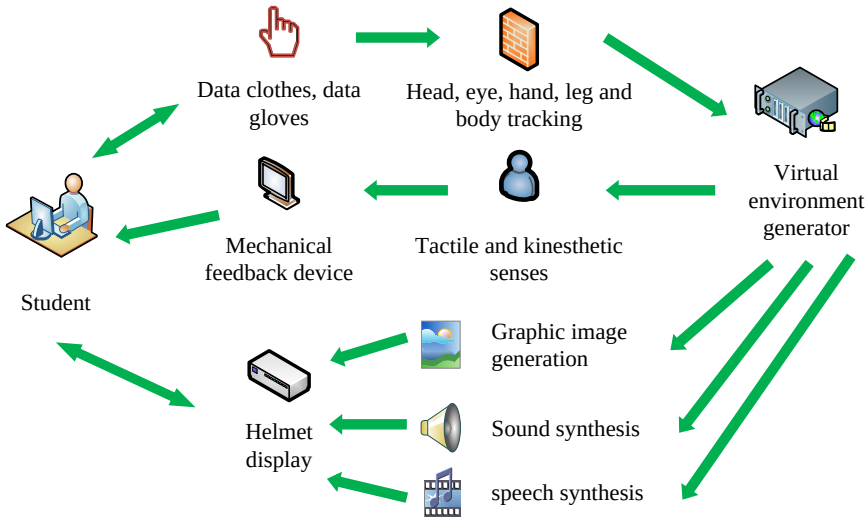


Fig. 2. Composition of interactive functions.

In Fig. 2, the interaction function requires the use of different wearable devices, a data suit, data gloves and a helmet display, as well as a mechanical feedback device. The camera in the helmet display is able to recognize the movement of the student's hand. The data suit and data gloves are able to track the student's head, eyes, hands, legs and body, and the mechanical feedback device mainly involves haptic and kinesthetic senses. The data analysis function is to collect the data changed in the VR scene through the data collection device in the background, and analyze and diagnose these data to determine whether the students have completed the exercise or not. In the parts that follow, the study's design will be thoroughly examined for the purpose of diagnosing these results.

B. Construction of the SSMCLAA Model

To make a judgment on whether a student has completed the exercise or not, the study uses the HMR algorithm. The human movement time series data collected using wearable sensing devices has a complex multidimensional spatial structure. To address the problems of complexity and variability of the collected data and the difficulty of feature extraction, the study designed the SSMCLAA model. SSL utilizes the input data itself as a supervisory signal and is beneficial for almost all types of downstream tasks [17]. Wearable sensing devices integrate various types of miniature sensing elements, such as accelerometers, gyroscopes, etc., which are capable of monitoring and tracking human activities in real time and continuously [18]. Therefore, the multidimensional time series data samples used in the study are composed of 3D acceleration data and 3D angular velocity data. The SSMCLAA model adopts the SSL framework, and its specific structure is shown in Fig. 3.

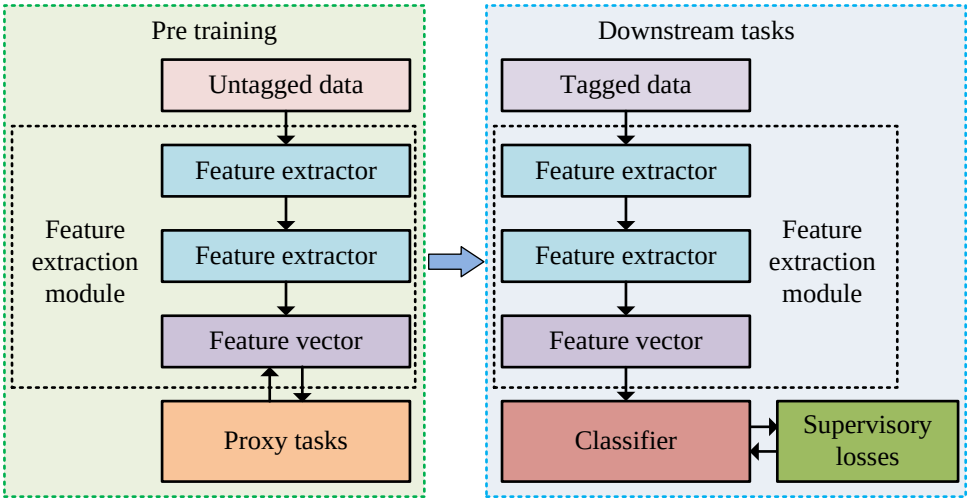


Fig. 3. The specific structure of the SSMCLAA model.

In Fig. 3, the SSL framework contains two main parts, which are pre-training and downstream tasks. Among them, the pre-training part contains unlabeled data (ULD), feature extraction module (FEM), and agent task, where FEM is divided into feature extractor (FE) + eigenvector. In order to carry out the agent task in the pre-training part, the study adopts the contrast learning technique. This technique is a common way to improve the model representation. The downstream task part contains labeled data (LD), FE trained in the pre-training part, eigenvector, classifier and supervised loss. Since the FE in the pre-training part only learns the feature representation of the data, the downstream task part needs to connect the FE trained in the pre-training part with the classifier, and then supervise the training with a small amount of LD to achieve action classification. In order to fully extract the acceleration and angular velocity features of the students when learning Sanda, the study constructs two independent FEs at the feature extraction module, and both of them use the deep residual network structure. In Fig. 4, the FEs' structure is displayed.

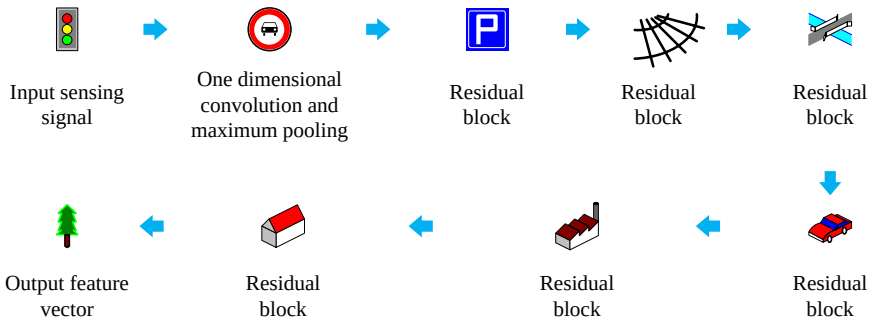


Fig. 4. The structure of a feature extractor.

In Fig. 4, the first step of the FE is the input sensing signal, i.e., the acceleration or angular velocity of the student while performing Sanda learning. The second step is a one-dimensional convolution and maximum pooling operation; after which it goes through six residual blocks. The combination of these six residual blocks together forms a deep residual network. The residual block is mainly composed of three consecutive one-dimensional convolutional layers and an activation function ReLU, and then the output and input of the convolution operation are directly added to solve the problem of gradient vanishing during the training process of deep neural networks, and to enable deep neural networks to learn deeper feature representations. The third step is the output eigenvector. The residual block mainly contains three 1D convolutional layers and ReLU activation function. The purpose of contrast learning pre-training is to enhance the FE representation. The commonly used formulaic definition of contrast learning is shown in Eq. (1) [19].

$$\max_{f_1,f_2}\sum_{i=1}^{N}\left\{s\left(z_1\left(w_{i1}\right)\right),\left(z_2\left(w_{i2}\right)\right)-\sum_{j=1,j\neq i}^{N}\left[s\left(z_1\left(w_{i1}\right)\right),\left(z_1\left(w_{j1}\right)\right)+s\left(z_1\left(w_{i1}\right)\right),\left(z_2\left(w_{j2}\right)\right)\right]\right\} \quad (1)$$

In Eq. (1), f_1 and f_2 represent different FEs, respectively. i and j are both ordinal numbers, and N is the total features. w_i and w_j represent different random samples, z_1 and z_2 are the projected heads of the upper and lower branches (ULB), respectively. $z_1\left(w_{i1}\right)$ and $z_2\left(w_{i2}\right)$ represent the features of the ULB of w_i, respectively, and both are set as positive samples. $z_1\left(w_{j1}\right)$ and $z_2\left(w_{j2}\right)$ are the features of the ULB of w_j, respectively, and are set as negative samples. s is the metric. The structure of the contrast learning pre-training is shown in Fig. 5.

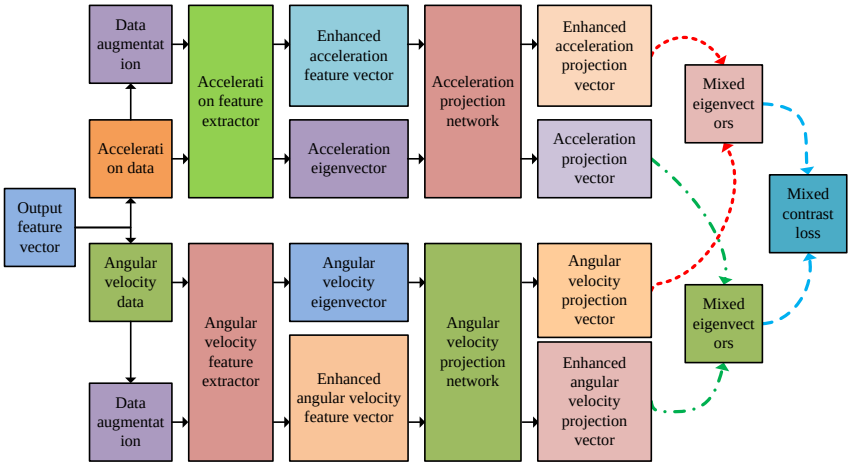


Fig. 5. Comparative learning pre training structure.

In Fig. 5, the contrast learning pre-training structure is mainly divided into the upper and lower halves. Among them, the upper half is the acceleration contrast learning process and the lower half is the angular velocity contrast learning process. In order to facilitate the subsequent unsupervised contrast learning, the study enhances both raw data. The study adds a projection network and obtains a new projection eigenvector in an effort to enhance the impact of FE even more. The acceleration contrast learning loss L_b is shown in Eq. (2).

$$L_b = -\sum_{i \in I} \log \frac{\exp\left(sim\left(z_i^b, z'^b_i\right)/\chi\right)}{\sum_{h \in H(i)} \exp\left(sim\left(z_i^b, z_h^b\right)/\chi\right)} \tag{2}$$

In Eq. (2), I represents the set of acceleration data after data augmentation, and $H(i)$ denotes the subscripts of the rest of the sample data except for the serial number i. χ is the parameter, z_i^b and z'^b_i are the corresponding projected eigenvector before and after the acceleration data enhancement, respectively. z_h^b is the projected eigenvector corresponding to the acceleration data other than the serial number i. $sim\left(z_i^b, z'^b_i\right)$ is the cosine similarity, which is solved as shown in Eq. (3) [20].

$$sim\left(z_i^b, z'^b_i\right) = \frac{\left(z_i^b\right)^T \cdot z'^b_i}{\left\|z_i^b\right\| \left\|z'^b_i\right\|} \tag{3}$$

The expression for $H(i)$ is shown in Eq. (4).

$$H(i) = I \setminus \{i\} \tag{4}$$

The angular velocity comparison learning loss L_g is shown in Eq. (5).

$$L_g = -\sum_{i \in I} \log \frac{\exp\left(sim\left(z_i^g, z'^g_i\right)/\chi\right)}{\sum_{h \in H(i)} \exp\left(sim\left(z_i^g, z_h^g\right)/\chi\right)} \tag{5}$$

In Eq. (5), z_i^g and z'^g_i are the projected eigenvectors corresponding to the angular velocity data before and after enhancement, respectively. z_h^g is the projected eigenvector corresponding to the angular velocity data other than the ordinal number i. The hybrid contrast loss function (LF) L_{bg} is shown in Eq. (6).

$$L_{bg} = -\sum_{i \in I} \log \frac{\exp\left(sim\left(z_i^{bg}, z'^{bg}_i\right)/\chi\right)}{\sum_{h \in H(i)} \exp\left(sim\left(z_i^{bg}, z_h^{bg}\right)/\chi\right)} \tag{6}$$

In Eq. (6), z_i^{bg} and z'^{bg}_i represent the hybrid eigenvector corresponding to the data before and after data enhancement, respectively. z_h^{bg} is the hybrid eigenvector corresponding to the data other than the ordinal number i. The solution of z_i^{bg} and z'^{bg}_i is shown in Eq. (7).

$$\begin{cases} z_i^{bg} = z_i'^b \oplus z_i^g \\ z'^{bg}_i = z_i^b \oplus z'^g_i \end{cases} \tag{7}$$

The total LF L_{sum} for the pre-training phase of contrast learning is shown in Eq. (8).

$$L_{sum} = L_b + L_g + \beta L_{bg} \tag{8}$$

In Eq. (8), β represents the parameter. The downstream task of the SSMCLAA model is to classify students' Sanda practice movements. In the downstream task, a simple two-layer linear mapping network and activation function are used for the classifier structure, and the cross-entropy loss function (CELF) L_{out} is used for the LF, as shown in Eq. (9) [21].

$$L_{out} = \frac{1}{m} \sum_{i, w_i \in G^d} \sum_{t=1}^{T} k_{it} \log\left(\hat{k}_{it}\right) \tag{9}$$

In Eq. (9), k_{it} and $\hat{k}_{it}$ represent the actual and predicted distribution probabilities, respectively. m is the size of the randomly selected data subset G^d. t and T are the action category ordinal number and total number.

C. Construction of the SSMTFCL Model

To make a judgment on whether a student has completed the Sanda exercise, research designed the HMR model of SSMCLAA to solve the problems of complex and variable collected data and the difficulty of feature extraction. However, the SSMCLAA model is unable to learn the common feature space structure of LD and ULD at the same time. In addition, human motion sensing signals are not only fluctuating and periodic, but also have multimodal characteristics such as time and frequency-domains (FDs) [22]. However, the current HMR only considers unilateral information in the time or frequency domain during feature extraction, ignoring the consistent relationship between time and frequency of human action sensing data, as well as the study of multi-task framework [23-24]. Based on these problems, in order to better judge whether students complete the Sanda exercise, the study also designed the SSMTFCL model and arranged it together with the SSMCLAA model in SATS to judge whether students complete the Sanda exercise. Fig. 6 depicts the SSMTFCL model's structure.

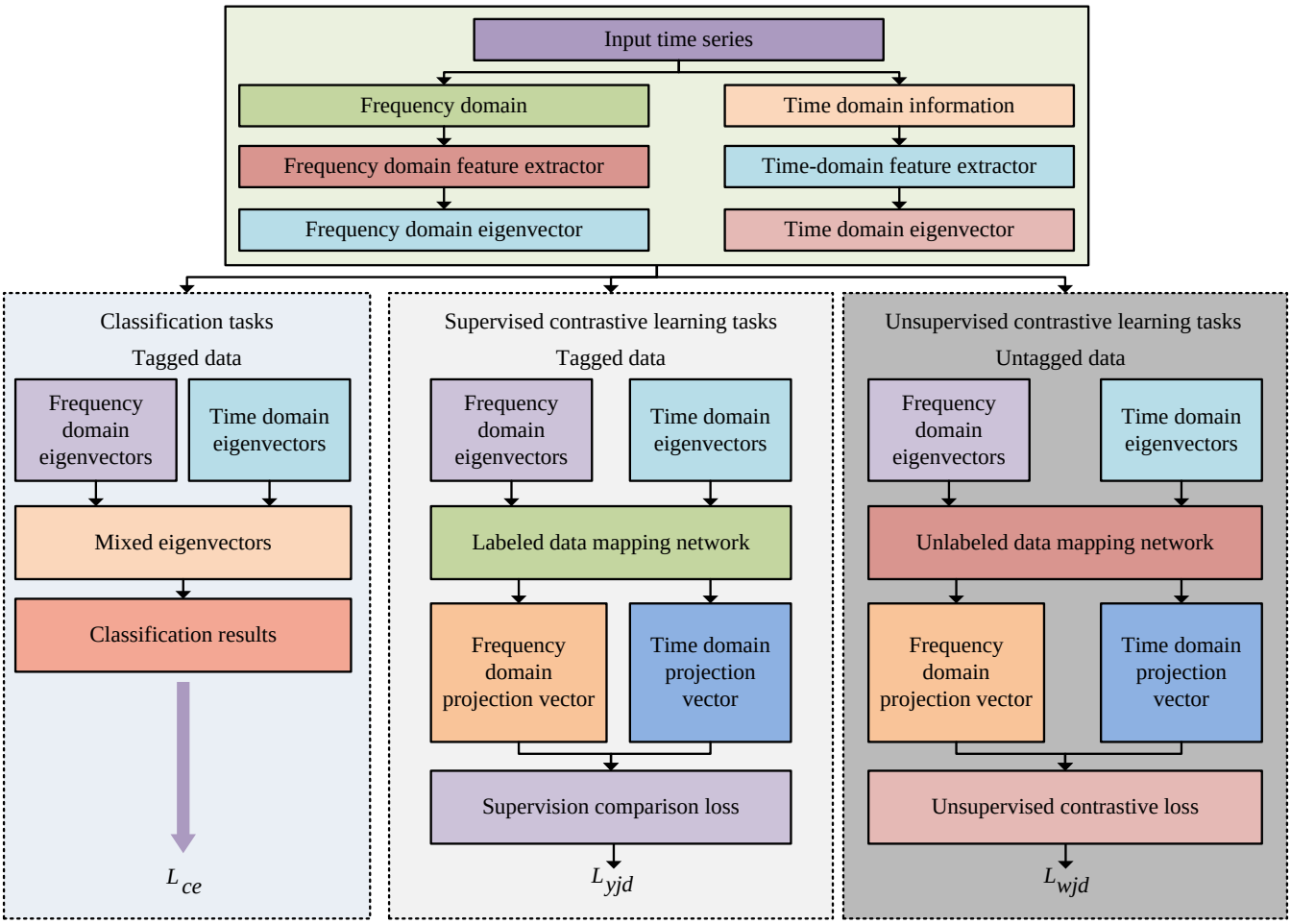


Fig. 6. The structure of SSMTFCL model.

In Fig. 6, the first step of the SSMTFCL model is to input the time series, and the second step is to obtain the FD information x_i' and the time-domain (TD) information x_i. The third step is to input x_i' and x_i into the FD FE V_F and the TD FE V_Q respectively. The fourth step is to output the corresponding FD eigenvector y_i' and the TD eigenvector y_i. The fifth step is to carry out the three tasks in the respective modules, i.e., classification task (CT), supervised contrast learning task and unsupervised contrast learning task, and output the CELF L_{ce}, supervised contrast learning LF L_{yjd} and unsupervised contrast learning LF L_{wjd}. In the CT, in order to maximize the consistency between the frequency domain and the TD of the samples, the study directly compares the TD projection vectors and the FD projection vectors of the data. To obtain the FD information, the study used Fourier transform. The Fourier transform has the advantage of good frequency localization and clearly shows the frequency components contained in the signal [25-26]. The process of Fourier transform is shown in Eq. (10) [27].

$$F(\omega)=\frac{1}{\sqrt{2\pi}}\int f(t)e^{i\omega t}dt \tag{10}$$

In Eq. (10), $f(t)$ represents a non-periodic function and $F(\omega)$ is the representation of the $f(t)$ function in the FD. $e^{i\omega t}$ is a complex exponential function and ω represents the angular frequency. The CELF L_{ce} for the CT is the same as the CELF L_{out} used in the downstream task of the SSMCLAA model, with only a slight difference in the values taken. The expression of L_{ce} is shown in Eq. (11).

$$L_{out}=\frac{1}{m}\sum_{i,x_i\in G^d}\sum_{t=1}^{T}k_{it}\log\left(\hat{k}_{it}\right) \tag{11}$$

To capture deeper TF information, the study constructed two independent FEs to extract the human action eigenvector from TD information and FD information, respectively. The TF feature extraction module is shown in Fig. 7.

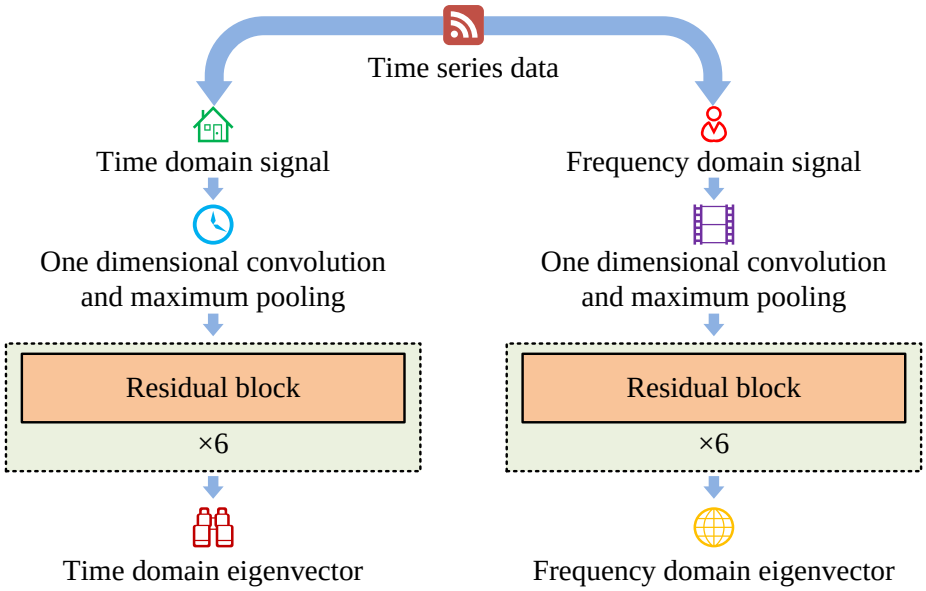


Fig. 7. Time-frequency feature extraction module.

In Fig. 7, the TF feature extraction module mainly consists of a TD signal FE and a FD signal FE. Both signal FEs involve one-dimensional convolution, maximum pooling, six residual block structure and eigenvector output. Both labeled and ULD are subjected to the same feature extraction method, and their respective obtained TF eigenvectors are used in subsequent multi-task learning modules. The supervised comparison learning task and the unsupervised comparison learning task in the SSMTFCL model are mainly used to mine the internal features of ULD and LD, and the supervised comparison learning task takes into account the data labeling information when calculating the loss. The unsupervised comparison LF L_{wjd} is shown in Eq. (12).

$$L_{wjd} = -\sum_{i \in R} \log \frac{\exp\left(sim\left(u_i^v, u''^v_i\right)/\chi\right)}{\sum_{h \in H(i)} \exp\left(sim\left(u_i^v, u_h^v\right)/\chi\right)} \tag{12}$$

In Eq. (12), R is the set of TD eigenvector and FD eigenvector, and u_i^v and u''^v_i denote the TD projection vectors and FD projection vectors of ULD, respectively. u_h^v is the TD projection vector corresponding to the ULD other than the ordinal number i. Minimization of unsupervised contrast loss can improve the model's representational ability. The study uses an unsupervised contrast learning task to learn the features of ULD, and also uses supervised contrast learning to perform deep mining of the deep structural features of LD. The study trains two contrast learning tasks in parallel to learn the feature space on the full data. The LF L_{yjd} for supervised contrast is shown in Eq. (13).

$$L_{yjd} = -\sum_{i \in R} \frac{1}{\left|K(i)\right|} \sum_{k \in K(i)} \log \frac{\exp\left(sim\left(u_i, u'_i\right)/\chi\right)}{\sum_{h \in H(i)} \exp\left(sim\left(u_i, u_h\right)/\chi\right)} \tag{13}$$

In Eq. (13), $K(i)$ represents the set of sample data of the same category as sample x_i. u_i and u'_i represent the TD projection vector and FD projection vector of the LD, respectively. u_h is the TD projection vector corresponding to the LD other than the serial number i. u_i and u'_i are solved as shown in Eq. (14).

$$\begin{cases} u_i = G_c\left(y_i\right) \\ u'_i = G_c\left(y_i'\right) \end{cases} \tag{14}$$

In Eq. (14), G_c represents the LD projection network. In order to allow the feature encoder to learn the intrinsic structure of the LD more deeply, the study minimizes the supervised comparison loss as well. In order to achieve overall consistency of data features, the study adopts a multi-task learning framework. The study co-trained the three tasks of the model so that they jointly participate in the optimization of the TF feature encoder. Eq. (15) displays the SSMTFCL model's total LF.

$$L_{\text{SSMTFCL}} = L_{ce} + \theta L_{yjd} + \varphi L_{wjd} \tag{15}$$

In Eq. (15), both θ and φ are scaling parameters.

III. RESULTS

To validate the performance of the research design SATS and the corresponding student action recognition classification model, the study sets up the experimental environment, experimental parameters and experimental dataset. In addition, the study also describes the comparison model and comparison system. The model comparison mainly involves accuracy, error and time consumption, while the system comparison mainly involves memory occupation, throughput and response time.

A. SSMCLAA Model Performance Validation

To validate the performance of the SSMCLAA model, the study uses the UCI-HAR dataset [28]. Since the length of the time series data is not uniform, the study uses a sliding window mechanism to segment the raw data, and the number of samples after processing is 6400. The study divides the dataset into a training set and a test set with a division ratio of 7: 3. The ratio of LD to ULD in the training set is 1: 9. The β parameter of the SSMCLAA model is set to 1, and the learning rates for the pre-training and downstream task phases are 1e-5 and 1e-3, respectively. Other SSL methods selected for comparison in the study are semi-supervised time model (SemiTime), self-supervised of human activity recognition (SelfHAR) and simple framework for contrastive learning of representations (SimCLR). The operating system used for the experiments is Windows 11 (64-bit), and the processor is Intel Core i5-12600 K with a maximum RWI of 4.9 GHz, a maximum accelerated power consumption of 130 W, and a maximum RAM of 128 GB. Fig. 8 compares the accuracy of the various models in the training and test sets.

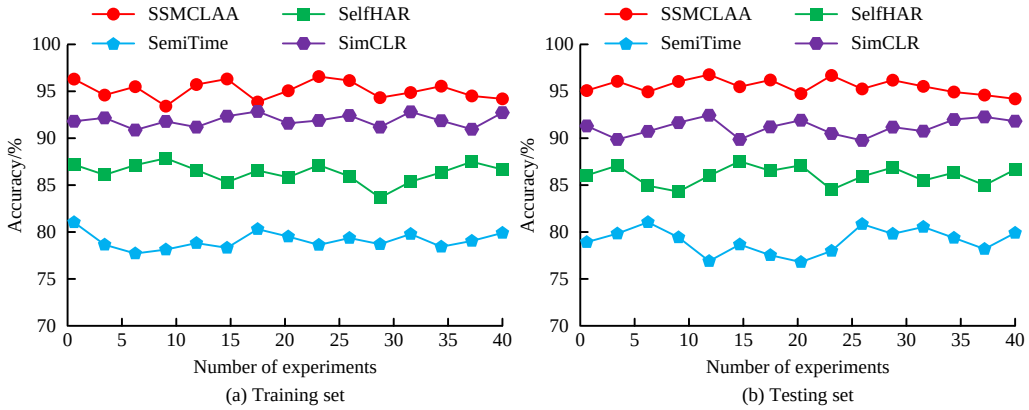


Fig. 8. Comparison of accuracy between different models on training and testing sets.

In Fig. 8(a), the maximum value (MaxV) of the accuracy of the research design SSMCLAA model is 95.76% and the minimum value (MinV) is 93.81% on the training set. The MaxVs of accuracy of SemiTime model, SelfHAR model and SimCLR model are 80.50%, 87.42% and 93.03%, respectively, and the MainVs are 77.64%, 83.89% and 90.71%, respectively. The accuracy of SSMCLAA model is significantly higher than the comparison models. In Fig. 8(b), on the test set, the MaxV of accuracy is appeared on SSMCLAA model with a value of 95.89%. The MainV of accuracy is appeared on SemiTime model with a value of 76.75%. The mean values of accuracy for SelfHAR model and SimCLR model are 86.13% and 91.24% respectively. The SelfHAR and SimCLR models outperformed the SemiTime model. In summary, the research design SSMCLAA model has a higher action recognition accuracy and is able to better recognize and classify students' Sanda actions in order to determine whether the students have completed Sanda training or not. Comparison of mean absolute error (MAE) and mean squared error (MSE) of different models are shown in Table I.

TABLE I. COMPARISON OF MAE AND MSE OF DIFFERENT MODELS

Model	MAE					MSE				
	Number of experiments					*Number of experiments*				
	1	*2*	*3*	*4*	*5*	*1*	*2*	*3*	*4*	*5*
SemiTime	1.21	1.37	1.19	1.32	1.28	1.31	1.29	1.38	1.42	1.46
SelfHAR	1.03	0.97	1.13	0.95	1.05	1.16	1.21	1.25	1.09	1.18
SimCLR	0.95	1.06	0.98	1.01	0.94	1.07	1.12	1.05	0.99	1.14
SSMCLAA	0.74	0.62	0.67	0.58	0.69	0.64	0.51	0.68	0.73	0.66

In Table I, the maximum and MainVs of MAE for the SSMCLAA model of the research design are 0.74 and 0.58, respectively. As a whole, the MAE values of the SSMCLAA model are significantly lower than those of the comparison models. The MaxVs of MAE for SemiTime model, SelfHAR model and SimCLR model are 1.37, 1.13 and 1.06, respectively, and the MainVs are 1.19, 0.97 and 0.94, respectively. In addition, the MainV of MSE occurs on the SSMCLAA model with a value of 0.51. The MaxV occurs on the SemiTime model with a value of 1.29. The mean values of MSE for the SemiTime model, SelfHAR model, SimCLR model and SSMCLAA model are 1.372, 1.178, 1.074 and 0.644, respectively. In summary, the values of MAE and MSE for the research-designed SSMCLAA model are significantly smaller than those of the comparison models, which suggests that the SSMCLAA model has a smaller classification error. This suggests that the SSMCLAA model can determine whether or not students have finished the Sanda training with a lower classification error. The time-consuming comparison of different models is shown in Fig. 9.

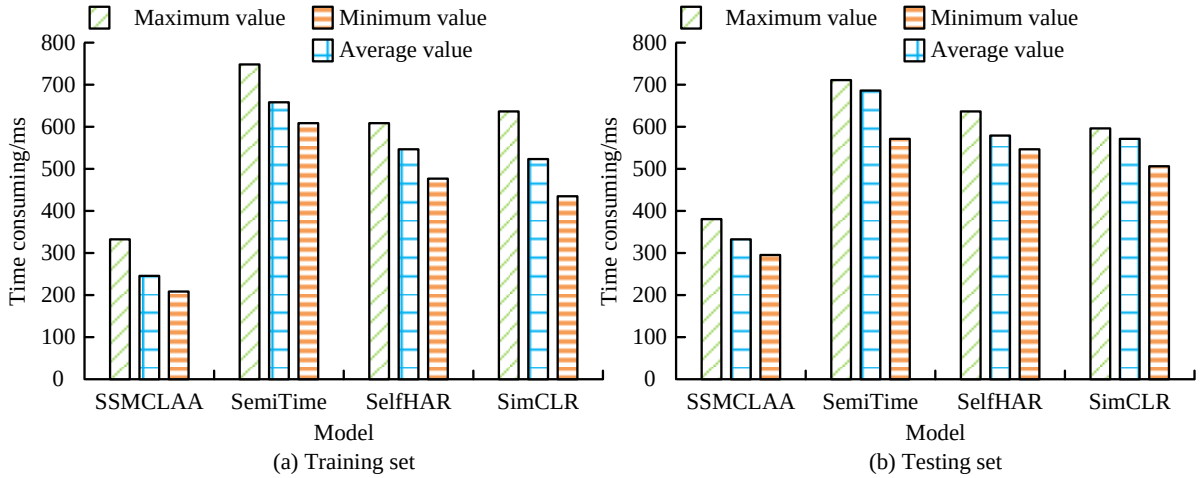


Fig. 9. Comparison of time consumption of different models.

In Fig. 9(a), on the training set, the time-consuming mean of the research design SSMCLAA model is 242 ms, and the time-consuming mean of the comparison models SemiTime, SelfHAR, and SimCLR are 670 ms, 543 ms, and 521 ms, respectively. The time-consuming mean of the SSMCLAA model is less than that of the comparison models, which are 328 ms, 201 ms, and 179 ms. The SSMCLAA model has a greater advantage in time. In Fig. 9(b), on the test set, the time-consuming AV of the SSMCLAA model is 337 ms, which is significantly lower than that of the comparison model. In conclusion, the research designed SSMCLAA model has less time consuming and can make a judgment on whether the students complete the Sanda training in a shorter period of time and reduce the waiting time of the students.

B. SSMTFCL Model Performance Validation

The dataset used in the study, the method the dataset is partitioned, and the experimental setup are all consistent with the SSMCLAA model performance validation, which is necessary to validate the SSMTFCL model's performance. The values of θ and φ in the LF of the SSMTFCL model are both 1.0, and the iterations of the model is 300. In addition, four classical semi-supervised algorithm models are selected for comparative validation of the research models, the double-Πmodel (Π-model), the MeanTeacher model, the multi task learning model (MTL), and the MixMatch model that combines Mixup and Fixmatch. Comparison of LF curves for different models is shown in Fig. 10.

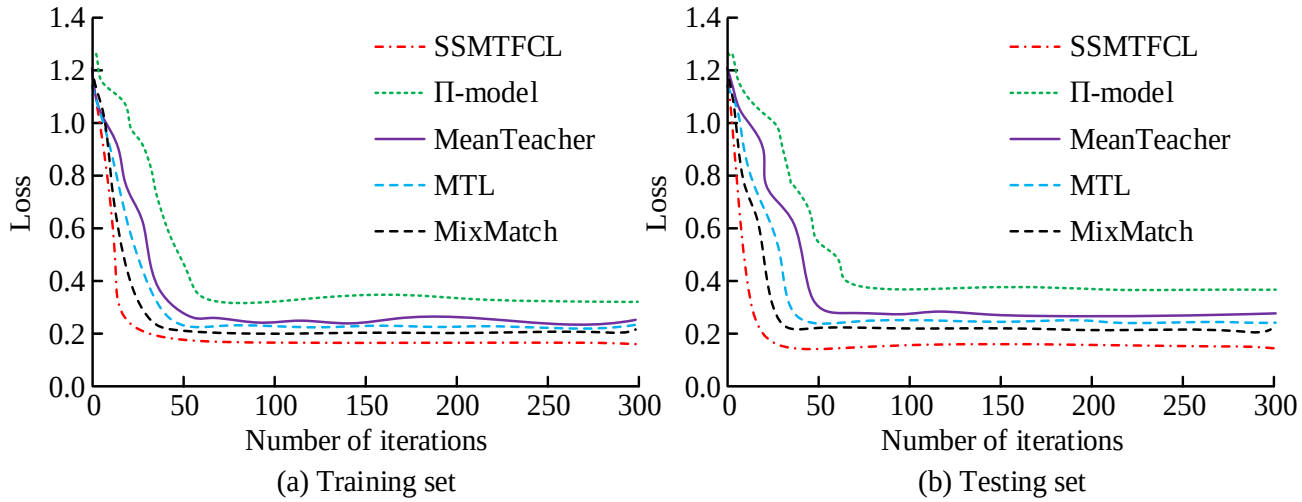
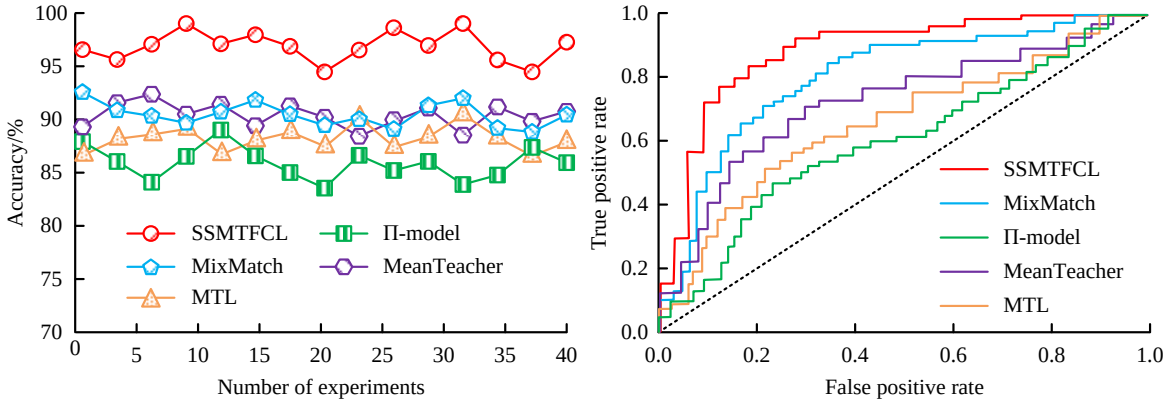


Fig. 10. Comparison of LF curves for different models.

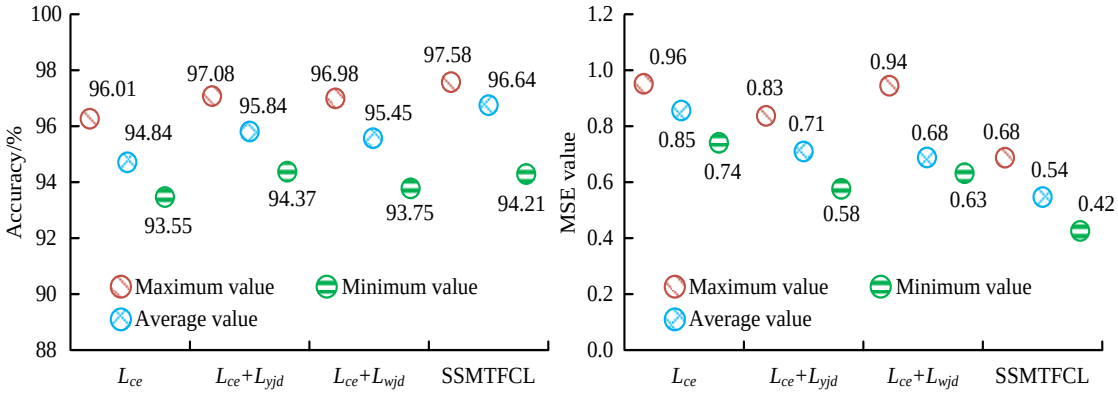
In Fig. 10(a), on the training set, the research design SSMTFCL model plateaus after almost 22 iterations, while the Π-model, MeanTeacher, MTL, and MixMatch models plateau after 57, 52, 48, and 43 iterations, respectively. The MainVs of loss values for the five models are 0.17, 0.33, 0.25, 0.23, and 0.21, respectively. In Fig. 10(b), on the test set, the MainV of loss values for the SSMTFCL model is 0.15, and the MainVs of loss values for the comparison models Π-model, MeanTeacher, MTL, and MixMatch are 0.36, 0.27, 0.24, and 0.22, respectively. The five models leveled off after nearly 24, 59, 54, 47, and 45 iterations, respectively. In summary, the research designed SSMTFCL model converges faster and has better performance with smaller loss values. A comparison of the accuracy and receiver operation characteristic, (ROC) curves for the different models is shown in Fig. 11.



(a) Comparison of accuracy of different models (b) Comparison of ROC curves of different models

Fig. 11. Comparison of accuracy and ROC curves of different models.

In Fig. 11(a), the MaxV of accuracy of the research design SSMTFCL model is 97.58%, the MainV is 94.21%, and the AV is 96.64%. The mean values of accuracy for the comparison models Π-model, MeanTeacher, MTL and MixMatch are 86.17%, 89.42%, 87.55% and 90.73% respectively. The accuracy of SSMTFCL model is significantly higher than the comparison model. In Fig. 11(b), the area under the ROC curve of the research design SSMTFCL model is the largest with a value of 0.945. It is followed by the MixMatch and MeanTeacher models with values of 0.899 and 0.857, respectively. The smallest are the MTL and Π-model models with values of 0.821 and 0.798, respectively. In summary, the research design SSMTFCL model has higher accuracy and better performance. To verify the effectiveness of the three tasks in the SSMTFCL model, the study conducted ablation experiments. L_{ce} is the LF for the CT, and L_{yjd} and L_{wjd} are the LFs for the supervised comparison learning task and the unsupervised comparison learning task, respectively. Fig. 12 displays the ablation experiment's outcomes.



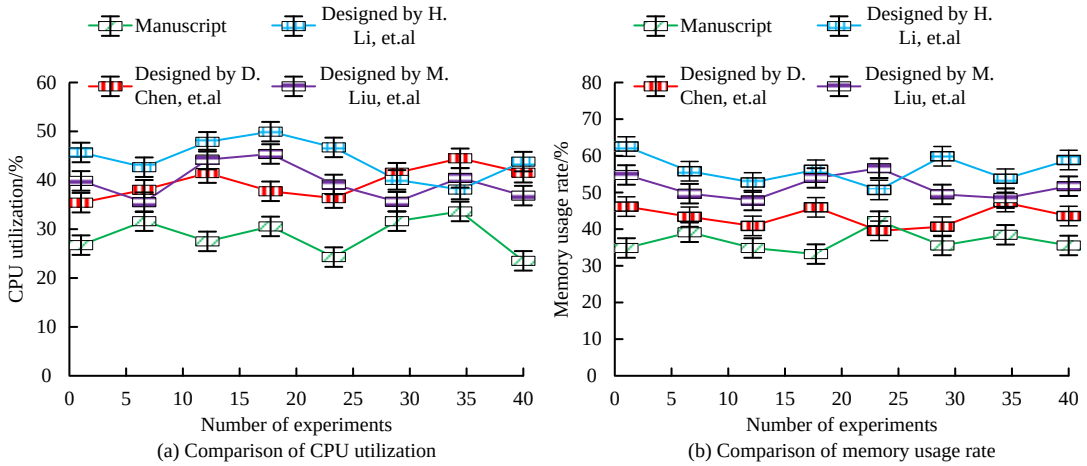
(a) Comparison of accuracy (b) Comparison of MSE value

Fig. 12. Results of SSMTFCL model ablation experiment.

In Fig. 12(a), the accuracy mean of the model using only the CT is 94.84%, which is 1.8% lower than the accuracy mean of the full SSMTFCL model of 96.64%. The average accuracy of the model using the CT and the supervised comparative learning task is 95.84%, while the average accuracy of the model using the CT and the unsupervised comparative learning task is 95.45%. In Fig. 12(b), the MSE mean values of the model using only the CT, the model using the CT and the supervised contrast learning task, and the model using the CT and the unsupervised contrast learning task are 0.85, 0.71, and 0.68, respectively. The MSE mean value of the full SSMTFCL model is 0.54. In conclusion, the multi-task learning framework used in the study is effective.

C. SATS Performance Validation

To validate the performance of the research design SATS based on 5G and VR technologies, the study selected other TSs for comparison. Among them, there are psychological virtual simulation experiment TS designed by experts such as D. Chen, PET aid network system designed by scholars such as Li, and piano playing TS designed by researchers such as Liu [29-31]. In addition, the experimental environment configurations used for system performance validation are consistent with those used for SSMCLAA model performance validation. A comparison of the central processing unit (CPU) utilization and memory occupancy of the different systems is shown in Fig. 13.



(a) Comparison of CPU utilization (b) Comparison of memory usage rate

Fig. 13. Comparison of CPU utilization and memory usage rate in different systems.

In Fig. 13(a), the average CPU utilization of the research design SATS is 27.8%, and the average CPU utilization of the psychological virtual simulation experiment TS, the PET aid network system, and the piano playing TS are 37.8%, 42.8%, and 39.7%, respectively. The CPU utilization of the research design SATS is significantly smaller than the comparison system. In Fig. 13(b), in terms of memory occupancy, the AV of memory occupancy of the research design system is 38%, and the AVs of memory occupancy of the other three comparison models are 43%, 57%, and 52%, which are 5%, 19%, and 14% higher than those of the research design system, respectively. In conclusion, the research design SATS has better performance. The response time and throughput comparisons of the different systems are shown in Table II.

TABLE II. COMPARISON OF RESPONSE TIME AND THROUGHPUT OF DIFFERENT SYSTEMS

System	Response Time/ms					Throughput/(bit/s)				
	Number of experiments					*Number of experiments*				
	1	*2*	*3*	*4*	*5*	*1*	*2*	*3*	*4*	*5*
Designed by D. Chen, et.al	136	143	132	158	132	63142	64498	65506	68504	63187
Designed by H. Li, et.al	145	162	157	140	142	60988	60543	61279	63885	64771
Designed by M. Liu, et.al	121	137	128	136	134	64312	67329	68634	66032	61062
Manuscript	63	57	61	55	59	74521	76282	75113	76895	77651

In Table II, in terms of response time, the AV of the research design system is 59 ms, and the AVs of the psychological virtual simulation experiment TS, the PET aid network system, and the piano playing TS are 140.2 ms, 149.2 ms, and 131.2 ms, respectively. The response time of the research design system is significantly lower than the comparison system. In addition, the throughput of the research design system can reach a maximum of 77651 bit/s and a minimum of 74521 bit/s. The MaxVs of the throughput of the psychological virtual simulation experiment TS, the PET aid network system, and the piano playing TS are 68504 bit/s, 64771 bit/s, and 68634 bit/s, respectively, and the MainVs are 63142 bit/s, 60543 bit/s and 61062 bit/s. In conclusion, the research design SATS has shorter response time, larger system throughput and better performance.

IV. DISCUSSION

In response to the shortcomings of offline Sanda teaching, this paper designs the SSMCLAA model and SSMTFCL model, and based on this, constructs a Sanda auxiliary teaching system that integrates 5G and VR technology. The results showed that the average memory usage of the Sanda assisted teaching system designed in the paper was 38%, which was 5%, 19%, and 14% lower than the average memory usage of the other three compared systems, respectively. The average response time is 59ms, which is 81.2ms, 90.2ms, and 72.2ms lower than the average response time of the other three compared systems, respectively. The paper design of a Sanda auxiliary teaching system can improve the effectiveness of Sanda teaching. Liu X et al. designed a decision support system to evaluate the functionality of 5G networks and artificial intelligence in higher education context teaching research in order to achieve teaching objectives. The results show that 5G networks and artificial intelligence algorithms can enhance the effectiveness of situational teaching in higher education [32]. This result is similar to the research findings.

V. CONCLUSION

To make up for the shortcomings of OST, the study constructs SATS based on 5G and VR technologies, and constructs SSMCLAA model and SSMTFCL model. The results revealed that the maximum accuracy of SSMCLAA model was 95.76% on the training set. The MaxVs of accuracy of the comparison models SemiTime, SelfHAR and SimCLR were 80.50%, 87.42% and 93.03%, which were 15.26%, 8.34% and 2.73% lower than the MaxVs of accuracy

of the SSMCLAA model, respectively. The MainVs of MAE and MSE for the SSMCLAA model were 0.58 and 0.51, respectively, and the mean values of time consumed for the training and test sets were 242ms and 337ms, respectively. The SSMCLAA model performed relatively well. On the training set, the SSMTFCL, Π-model, MeanTeacher, MTL, and MixMatch models plateaued after almost 22, 57, 52, 48, and 43 iterations, respectively, with minimum loss values of 0.17, 0.33, 0.25, 0.23, and 0.21, respectively. The AVs of accuracy for the five models were 96.64%, 86.17%, 89.42%, 87.55% and 90.73%. The SSMTFCL model performs better. The AV of CPU utilization for the research design SATS was 27.8%, the AV of memory occupancy was 38%, the AV of response time was 59 ms, and the maximum and MainVs of throughput were 77,651 bit/s and 74,521 bit/s, respectively.

The research design SATS has good performance. The study also has some shortcomings. One, the test of SATS on students in terms of Sanda is limited to the technical level, and future research can include a theoretical question-answering session to enhance students' theoretical knowledge base and further avoid students' injuries. Secondly, the helmet used in the study and other types of sensing devices can cause discomfort to some students after prolonged use, and future research can do in-depth exploration on the comfort of wearable devices. Thirdly, the Sanda auxiliary teaching system still lacks human-machine training, game mode, and competition mode. Future research can compensate for this module to further enrich the Sanda auxiliary teaching system. Fourthly, for the data collected by different data collection devices, future research can simply annotate their location information to improve the recognition efficiency of students' Sanda movements. Fifthly, some students are prone to multiple compound movements due to non-standard movements during Sanda practice, which increases the difficulty of judging Sanda movements. Future research can further improve and optimize the model for recognizing composite actions.

REFERENCES

[1] S. Feng, “Experimental study on attention blink of male sanda athletes of different sports levels under fatigue state,” J. Educ. Educ. Res., vol. 7, pp. 305-309, March 2024.

[2] H. Zhao, “Design of sanda action reconstruction model based on 3D images,” Wirel. Commun. Mob. Com., vol. 2021, pp. 1-6, January 2021.

[3] J. Xu and S. Ariyasajsiskul, “Sports injury of sanda wushu class of students in Xi'an city,” Int. J. Sociol. Anthropol. Sci. Rev., vol. 4, pp. 465-476, January 2024.

[4] Y. Asham, M. H. Bakr, and A. Emadi, “Applications of Augmented and Virtual Reality in Electrical Engineering Education: A Review,” IEEE. ACCESS., vol. 11, pp. 134717-134738, November 2023.

[5] Y. Feng, C. You, Y. Li, Y. Zhang, and Q. Wang, “Integration of computer virtual reality technology to college physical education,” J. Web Eng., vol. 21, pp. 2049-2071, December 2022.

[6] J. Gao, L. Chong, X. Qiao, and F. Tian, “Telemedicine virtual reality based skin image in children's dermatology medical system,” Comput. Intell., vol. 38, pp. 229-248, June 2021.

[7] M. H. Elgewely, W. Nadim, A. Elkassed, M. Yehiah, and S. Abdennadher, “Immersive construction detailing education: building information modeling (BIM)-based virtual reality (VR),” Open. House. Int., vol. 46, pp. 359-375, July 2021.

[8] O. Almousa, R. Zhang, M. Dimma, J. Yao, A. Allen, L. Chen, P. Heidari, and K. Qayumi, “Virtual reality technology and remote digital application for tele-simulation and global medical education: an innovative hybrid system for clinical training,” Simulat. Gaming., vol. 52, pp. 614-634, May 2021.

[9] P. P. Groumpos, “A critical historic overview of artificial intelligence: issues, challenges, opportunities, and threats,” AIA., vol. 1, pp. 197-213, July 2023.

[10] M. He, G. Song, and Z. Wei, “Human behavior feature representation and recognition based on depth video,” J. Web. Eng., vol. 19, pp. 883-902, December 2020.

[11] L. I. Yangzhi, J. Yuan, and H. Liu, “Human skeleton-based action recognition algorithm based on spatio-temporal attention graph convolutional network model,” J. Comput. Appl., vol. 41, pp. 1915-1921, July 2021.

[12] P. Gao, D. Zhao, and X. Chen, “Multi-dimensional data modelling of video image action recognition and motion capture in deep learning framework,” IET. Image. Process., vol. 14, pp. 1257-1264, April 2020.

[13] P. Chen, H. Wang, H. Yan, J. Du, Y. Ning, and J. Wei, “sEMG-based upper limb motion recognition using improved sparrow search algorithm,” Appl. Intell., vol. 53, pp. 7677-7696, July 2023.

[14] Q. Ye, H. Zhong, C. Qu, and Y. Zhang, “Human interaction recognition method based on parallel multi-feature fusion network,” Intell. Data. Anal., vol. 25, pp. 809-823, July 2021.

[15] M. Xu, “Application of human-computer interaction virtual reality technology to the design of ice and snow landscapes,” Int. J. Hum. Robot., vol. 19, pp. 74-88, April 2022.

[16] S. Makhsuci, S. M. Navidi, M. Sanduleanu, and M. Ismail. “A review of doherty power amplifier and load modulated balanced amplifier for 5G technology,” Int. J. Circ. Theor. App., vol. 51, pp. 2422-2445, July 2023.

[17] A. D. Desai, B. M. Ozturkler, C. M. Sandino, R. Boutin, M. Willis, S. Vasanawala, B. A. Hargreaves, C. Ré, J. M. Pauly, and A. S. Chaudhari, “Noise2Recon: enabling SNR-robust MRI reconstruction with semi-supervised and self-supervised learning,” Magn. Reson. Med., vol. 90, pp. 2052-2070, July 2023.

[18] I. Okpala, C. Nnaji, and I. Awolusi, “Wearable sensing devices acceptance behavior in construction safety and health: assessing existing models and developing a hybrid conceptual model,” Constr. Innov.: Inf. Process. Manage., vol. 22, pp. 57-75, January 2022.

[19] A. Tokhmetova and A. Y. Albagachiev, “Comparative analysis of a numerical method and machine learning methods of temperature determination of a doped lubricating layer with experimental data,” J. Mach. Manuf. Reliab., vol. 52, pp. 509-515, September 2023.

[20] P. Rathnasabapathy and D. Palanisami, “A theoretical development of improved cosine similarity measure for interval valued intuitionistic fuzzy sets and its applications,” J. Amb. Intel. Hum. Comp., vol. 14, pp. 16575-16587, June 2023.

[21] K. Li, B. Wang, Y. Tian, and Z. Qi, “Fast and accurate road crack detection based on adaptive cost-sensitive loss function,” IEEE. T. Cybernetics., vol. 53, pp. 1051-1062, February 2023.

[22] A. Natarajan, V. Krishnasamy, and M. Singh, “Device-free human motion detection using single link wiFi channel measurements for building energy management,” IEEE. Embed. Syst. Lett., vol. 15, pp. 153-156, September 2023.

[23] Q. Wu, Q. Huang, and X. Li, “Multimodal human action recognition based on spatio-temporal action representation recognition model,” Multimed. Tools. Appl., vol. 82, pp. 16409-16430, November 2023.

[24] D. Noh, H. Yoon, and D. Lee, “A Decade of Progress in Human Motion Recognition: A Comprehensive Survey From 2010 to 2020,” IEEE. ACCESS., vol. 12, pp. 5684-5707, January 2024.

[25] M. J. Grotevent, S. Yakunin, D. Bachmann, C. Romero, J. R. V. D. Aldana, and M. Madi, et al. “Integrated photodetectors for compact Fourier-transform waveguide spectrometers,” Nat. Photonics., vol. 17, pp. 59-64, October 2023.

[26] A. Khalili Golmankhaneh, K. K. Ali, R. Yilmazer, and M. K. A. Kaabar, “Local fractal fourier transform and applications,” Comput. Methods. Diffe., vol. 10, pp. 595-607, July 2022.

[27] K. Skrai, J. Petrovi, and P. Pale, “Classification of low-and high-entropy file fragments using randomness measures and discrete fourier transform coefficients,” Vietnam. J. Comput. Sci., vol. 10, pp. 433-462, July 2023.

[28] Y. Kaya and E. K. Topuz, "Human activity recognition from multiple sensors data using deep CNNs," MULTIMED. TOOLS. APPL., vol. 83, pp. 10815-10838, June 2024.

[29] D. Chen, X. Kong, and Q. Wei, "Design and development of psychological virtual simulation experiment teaching system," Comput. Appl. Eng. Educ., vol. 29, pp. 481-490, July 2021.

[30] H. Li, H. Zhang, and Y. Zhao, "Design of computer-aided teaching network management system for college physical education," Comput. Aided. Des. Appl., vol. 18, pp. 152-162, February 2021.

[31] M. Liu and J. Huang, "Piano playing teaching system based on artificial intelligence-design and research," J. Intell. Fuzzy Syst.: Appl. Eng. Technol., vol. 40, pp. 3525-3533, February 2021.

[32] X. Liu, M. Faisal, and A. Alharbi, "A decision support system for assessing the role of the 5G network and AI in situational teaching research in higher education," SOFT. COMPUT., vol. 26, pp. 10741-10752, October 2022.