IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 40, No. 1, February, 2023, pp. 249-256
Journal homepage: http://iieta.org/journals/ts

Student Behavior Identification During Practice and Training Based on Video Image

Wei Chen, Xinqiao Fan, Fengwei Dai*, Tingting Chen

College of Commerce and Tourism, Hangzhou Vocational and Technical College, Hangzhou 310018, China

Corresponding Author Email: 2000010004@hzvtc.edu.cn

https://doi.org/10.18280/ts.400124

Received: 15 October 2022
Accepted: 12 January 2023

Keywords:
video image, student behavior, practice and training, convolution network model, behavior identification

ABSTRACT

Enriching and developing the connotation and value of labor education theories can help students in higher vocational colleges form correct viewpoint and attitude towards labor. Higher vocational colleges should put more efforts to education through practice based on the features of each discipline. Accurately identifying students' behavior in complex practice and training scenarios is very important for teachers to know about their status during practice and training, however, existing research results are not applicable to complex practice and training scenarios since they have neither considered how to improve the accuracy of static image identification while ensuring the model is lightweight structured, nor considered the time series information of students' behavior during practice and training in the collected video images. For this reason, this paper took the property management major as the subject to study the identification of student behavior during practice and training based on video image. In the paper, the students' practice and training content was divided into three aspects, a task of asking students to cooperate with each other to deal with an equipment failure emergency was adopted for the research, and a research idea of helping teachers figure out students' status during practice and training via identifying their actions and intentions during the said activities was determined. Then, a few pre-processing operations were performed on the captured video images of student behavior during practice and training, including removing abnormal image frames, filtering, and aligning, etc. After that, based on the collected video image data, the dynamic convolution kernel was improved and optimized, and a lightweight convolution network model was built for identifying student behavior during practice and training. At last, experimental results verified the validity of the proposed identification model.

1. INTRODUCTION

Labor is a foundation for the existence and development of human society, so it's a necessary task in pedagogy to teach students about labor. Enriching and developing the connotation and value of labor education theories can help students in higher vocational colleges form correct viewpoint and attitude towards labor [1, 2]. Cultivating students to form a good labor habit is an important content for vocational colleges to fulfill their mission of enabling students to achieve all-round development in terms of morality, intelligence, physical fitness, aesthetic perception, and labor [3-5]. Higher vocational colleges need to put more efforts to education through practice based on the features of each discipline. Some of them carry out labor-themed on-campus activities such as labor skill festival, labor culture festival, or labor week, hoping to fill the gap between labor education in class and the real world situations, thereby preparing students for the challenges of this new age [6-10].

Accurately identifying students' behavior in complex practice and training scenarios is very important for teachers to know about their status during practice and training, and it is also a prerequisite for the top-level design of labor education [11-18]. The maturing deep learning technology and image processing technique have laid a theoretical foundation for developing algorithms for identifying student behavior during practice and training based on video image.

Chen et al. [19] pointed out in their paper that intelligent identification of students' classroom behavior is becoming increasingly important in a smart education background. To improve the accuracy of intelligent identification of student behavior, the authors collected 7 typical classroom behavior images of 300 students, preprocessed the data, then a classic deep network model VGG16, which was trained on ImageNet dataset, was applied to the student classroom behavior identification task. Through experimental comparison with other deep learning models, they verified the high accuracy of this VGG16 network model in identifying students' classroom behavior. Xiao et al. [20] constructed a classroom teaching video library and a student classroom behavior library. In their work, classroom monitoring videos were extracted regularly in combination with the cv2 library to get a real-time picture stream of each student; then the spatial-temporal features of each student behavior were learned using convolutional neural networks so as to achieve real-time behavior identification in classroom teaching scenarios oriented to multi-student targets. Moreover, an intelligent teaching assessment model was constructed and an intelligent teaching assessment system based on student classroom behavior identification was designed and implemented, and the latter got good operation results on the classroom teaching video dataset. Scholar Pang [21] argues that student classroom behavior identification has important guiding significance for the development of distance education strategies. The author improved the traditional

clustering analysis algorithm using random forest, and combined a human skeleton model to identify students' classroom behavior in real time. Then, a network topology model was constructed based on the needs of behavior identification to build a network topology model. The error rate of feature reconstruction using spatio-temporal features was lower than that of a single feature, and the effectiveness of the extracted spatial angle features was verified through experiments based on the human skeleton model. Scholar Liu et al. [22] noticed the complex and slow-speed problem with conventional behavior identification process and proposed a method of student abnormal behavior identification in classroom video based on deep learning. For the poor effect of small target identification of the original network, the proposed method introduced a cascading improved RFB module by adding a branch to the RFB to increase the reference to peripheral visual field, which can enhance the feature extraction capability of original network and make full use of the shallow information to improve identification effect of small targets, and border regression calculation was performed by changing the border loss function to DIoU_Loss. The experimental results show that the improved network SE-Res2Net-DIoU achieved 80.1% in the accuracy of student abnormal behavior identification.

After reviewing relevant studies, it's found that existing research results are not applicable to complex practice and training scenarios since they have neither considered how to improve the accuracy of static image identification while ensuring the model is lightweight structured, nor considered the time series information of student behavior during practice and training in the collected video images. For this reason, this paper took the property management major as the subject to study the identification of student behavior during practice and training based on video image. In the second chapter, students' practice and training content was divided into three aspects, a task of asking students to cooperate with each other to deal with an equipment failure emergency was adopted for the research, and a research idea of helping teachers figure out students' status during practice and training via identifying their actions and intentions during the said activities was determined. In the third chapter, a few pre-processing operations were performed on the captured video images of student behavior during practice and training, including removing abnormal image frames, filtering, and aligning, etc. In the fourth chapter, based on the collected video image data, the dynamic convolution kernel was improved and optimized, and a lightweight convolution network model was built for identifying student behavior during practice and training. At last, experimental results verified the validity of the proposed identification model.

2. CONTENT OF STUDENTS' PRACTICE AND TRAINING AND STEPS OF BEHAVIOR IDENTIFICATION

After the impact of COVID-19 epidemic, the higher vocational education in China has posed some new requirements for talent cultivation, and an objective of "setting curriculum under ideological and political guidance" has been proposed for the purpose of training students to become virtuous and skilled talents with the ability to cope with the epidemic, to serve our people, and be good at management and operation, so that in the future, they could grow into honourable laborers who can care for people, have new ideas of management and operations, and fight the epidemic with courage. Practice and training is a major means of labor education, in order to create new content and approach for labor education in higher vocational colleges, labor education must be pragmatic and engaging. The content setting of students' practice and training is very important, so this paper took the property management major as the subject and divided the content of practice and training into three aspects:

The first content is owner management. One basic task of owner management is the management of property owners, specific content of this task includes: managing owner information such as the building unit number, room area, house orientation, the name, birthday, personality, contact number, and work unit of property owners and their family members. All these information should be registered and documented properly for easy inquiry in the future.

The second content is security management. The focus of security management is to prevent the occurrence of various accidents. Modern security management combines artificial prevention measures with technical prevention measures, wherein artificial prevention is carried out by personnel of special job posts, for example, typical security-related posts of a building include the VIP security posts, lobby security posts, mobile posts, parking lot security posts, and fire prevention posts; technical prevention adopts modern technologies to support artificial measures and carry out security management tasks, examples include the monitoring system, smoke sensor, displacement sensor, smoke sensor, temperature sensor, and infrared sensor, etc.

The third content is equipment maintenance and maintenance management. The repair and maintenance of equipment is an important task starting from the acceptance check of the equipment. Once an equipment is taken over by the property management department, its warranty scope and time should be figured out, and the maintenance work of the equipment should be checked clearly. At the same time, equipment operation procedures and equipment failure emergency response procedures should be formulated; monthly, quarterly, and annual maintenance plans should be made; equipment management system and equipment room management system should be established; a machine account should be set for each equipment, the parameters of each equipment, as well as each maintenance, should be properly recorded, so that every operator could have a clear understanding of the equipment at a glance. Moreover, a label should be made for each equipment and all equipment should be numbered, in this way, the records could be accurate enough. One last thing, the energy consumption budget should be made before the running of the equipment to achieve economic operation.

Research content of this paper is to help teachers know about students' status of practice and training by identifying their actions and intentions during practice and training activities. Taking the property management major as an example, this paper can assist teachers to check whether the students have completed tasks of owner management, security management, and equipment maintenance management in various practice and training scenarios. In our study, a task of asking students to cooperate with each other to deal with an equipment failure emergency was set to simulate a scenario of student practice and training, during which the actions between students when they pass equipment maintenance tools and components to each other were captured. During this

task, students were asked to work in pairs to complete jobs such as deliver and fetch the equipment maintenance tools and components, before the task, they were required to watch a video showing the specific steps, then they were asked to repeat the equipment failure emergency response task for multiple times until they reached the required practice and training level of professional property management.

3. PREPROCESSING OF VIDEO IMAGES OF STUDENT BEHAVIOR DURING PRACTICE AND TRAINING

When a student performs complex actions during task execution, the positions of his/her body joints would show obvious changes, and these data are important for accurate identification of student behavior during practice and training. To improve the efficiency and accuracy of identification, some non-critical joints were excluded. Besides, the angle between the position at which the student performs the task and the shooting position of the camera can also affect the identification effect, so this paper transformed the coordinates of joint positions to eliminate the influencing of camera tilt, the following formula gives the expression of the *i*-th joint in the *g*-th frame after motion trajectory angle transformation:

$$\bar{T}_i^g = \begin{bmatrix} a_i^g \\ u_i^g \\ c_i^g \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ \cos\omega & 0 & 0 \\ 0 & 0 & \sin\omega \end{bmatrix}, g \in M \tag{1}$$

Abnormal image frame can cause deviation in the prediction of behavior actions, and this deviation in the understanding of students' behavior intentions will lead to the result that a maintenance tool or component delivered to them may not meet their requirements, so this paper adopted the box plot method to analyze the images of student behavior and delete abnormal image frames in time.

Assuming: W_1-1.5SU and W_3+1.5SU respectively represent the lower limit and upper limit of the data of student behavior; W_1 represents the lower quartile, W_2 represents the median, W_3 represents the upper quartile, SU represents the difference between W_3 and W_1, that is SU=W_3-W_1.

$$\begin{cases} outlinear > W_3 + mSU \\ outlinear < W_1 + mSU \end{cases}, m = 1.5 or 3 \tag{2}$$

The value of m took 1.5 or 3, if m is equal to 1.5, then the behavior image data between W_1-1.5SU and W_3+1.5SU is within the normal value range, and data outside this range is considered as mild abnormal data; similarly, if m is equal to 3, then the behavior image data between W_1-3SU and W_3-3SU is within the normal value range, and data outside this range is considered as extreme abnormal data.

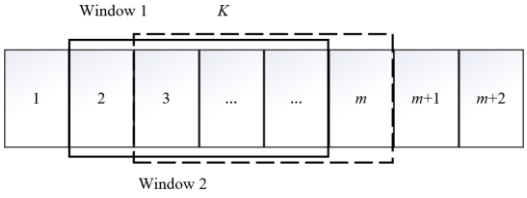


Figure 1. A schematic diagram of the filtering method

When a student performs an action during practice and training, tiny but quick movements such as trembling would produce noise in the collected image data, so in this study, the moving-average filter was adopted to perform filtering, as shown in Figure 1. Implementation process of this moving-average filtering method is detailed below, for student behavior image data segments A_1,...,A_K within a certain length K, the formula for calculating the average value is:

$$\bar{A} = \frac{1}{K} \sum_{i=0}^{K} A_{K-i} \tag{3}$$

Then, the student behavior image data at position A_1 was replaced by the average value; starting from A_2, again, K data points were taken and averaged, and the calculation result was used to replace A_2; after repeating this operation for m times, the m-times image filtering smoothing result was attained.

The camera collected real-time images of students' behavior actions, since there are differences in the duration of each action in the practice and training scenario, there will be differences in the length of each group of action image data if a same sampling frequency has been adopted, therefore, to facilitate the effective learning and prediction of student behavior during practice and training, each group of action image data should be align and this paper chose to use discrete Fourier transform to perform the processing. Assuming: $a(p)$ represents the time signal of non-periodic continuous motion images, then the Fourier transform of $a(p)$ is:

$$A(\theta) = \int_{-\infty}^{+\infty} a(p) o^{-j\theta p} dp \tag{4}$$

Let θ^{-l}_m=$e^{2lfj/m}$, then the above formula was discretized further to get:

$$A(l) = \sum_{0}^{m-1} a(m) \theta_m^{-l}, m = 0,1,...,m-1 \tag{5}$$

Students vary greatly in height, body shape, and daily behavior habit, when they were carried out the task, their positions relative to the camera were different as well. To avoid errors in the prediction of student behavior during practice and training caused by individual differences, the action image data needs to be normalized. Assuming: $\dot{T}^g_i$ represents the position of the n-th joint in the g-th frame after normalization; a^g_{SP}, b^g_{SP}, c^g_{SP} represent the *spinebase* coordinates in the g-th frame; after the local coordinate origin of the image was set, the coordinates of each joint in each frame could be attained based on the following formula:

$$\dot{T}_i^g = \begin{bmatrix} a_i^g - a_{SP}^g \\ b_i^g - b_{SP}^g \\ c_i^g - c_{SP}^g \end{bmatrix}, g \in M \tag{6}$$

4. STUDENT BEHAVIOR IDENTIFICATION MODEL

In the research of intention identification of student behavior during practice and training, only identifying static images or only considering the spatial features of actions can result in loss of temporal features of the actions, for instance, when a student tightens or loosens a screw, if we only observe

the behavior in a single frame, we can not judge whether he/she is tightening the screw or loosening it, and whether the behavior is an effective action of equipment failure emergency response or not is pending for judgement. Therefore, behavior identification based on video image plays a very important role in identifying student behavior during practice and training. In our study, the dynamic convolution kernel was optimized based on the collected video image data, and a lightweight convolution network model was built for behavior identification by terminal devices with less parameters used in specific practical scenarios.

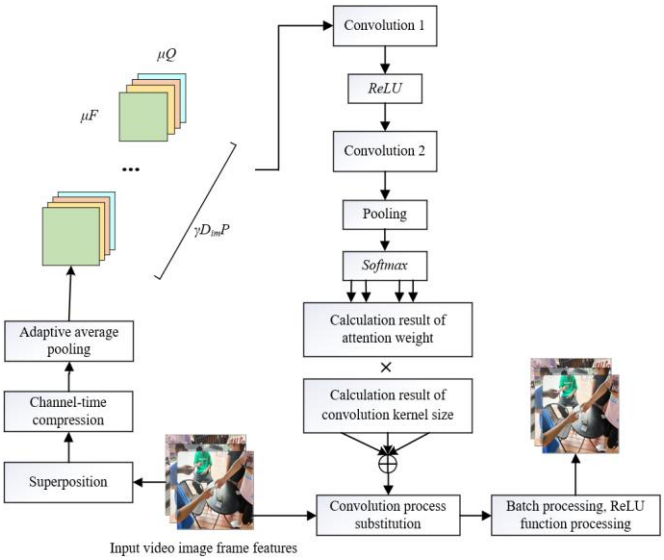


Figure 2. Calculation flow of 3D dynamic convolution

Conventional 3D dynamic convolution can superimpose all channel dimensions and time dimensions of the collected video images to form one dimension for processing, however, inevitably, this will introduce some redundant information that can affect behavior identification, so this paper proposed a kind of optimized 3D dynamic convolution kernel based on efficient attention to solve this problem, and its calculation flow is given in Figure 2.

Assuming: the size of input video is $D_{im}\times F\times Q\times P$, by superimposing time dimension and channel dimension, the input feature of video images can be superimposed into $D_{im}\times P\times F\times Q$, after that, the superimposed dimension can be compressed at the scale of $\gamma D_{im}\times P\times F\times Q$, wherein the value range of β is [0,1]. In order to reduce redundant information and retain useful information to the great extent, in this paper, down-sampling processing was performed on the spatial feature of video images after subjected to dimension compression at the scale of $\gamma D_{im}P\times\mu F\times\mu Q$, the processed video image features were then superimposed at the scale of $\gamma D_{im}P\times\mu^2 FQ$. Furthermore, the output result was processed by the fully-connected layer and the adaptive pooling layer, and finally the dimension of video image features had been reduced to L, namely the number of dynamic convolution kernels, at last, the attention weight was calculated by the *Softmax* function.

Assuming: $U(.)$ represents the superimposing operation of channel dimension and time dimension, $R(.)$ represents the dimension compression process of input features, $H(.)$ represents the process of adaptive average pooling which is used for downsampling the feature vector; G_1 and G_2 represent two layers of fully-connection operations in the module, $\varepsilon(.)$ represents the *ReLu* nonlinear function, then, for an input video of students' practice and training $C\in R^{D_{im}*F*Q*P}$, its dynamic weight can be calculated based on the following formula:

$$\beta = Soft\max\left(G_2\varepsilon\left(G_1\left(H\left(R\left(U\left(C\right)\right)\right)\right)\right)\right) \tag{7}$$

The random channel-time dimension was adopted to perform redundancy-removal dimension compression on 3D dynamic kernels to reduce the superimposed time and channel dimensions, specifically, $\gamma D_{im}P$ dimensions were randomly picked from $D_{im}P$ dimensions of video images:

$$\hat{E} = R\left(U\left(C\right)\right) \tag{8}$$

$$\hat{C} = \left\{\hat{C}_0, \hat{C}_1, ..., \hat{C}_{\gamma D_{im}P}\right\} \tag{9}$$

$$\hat{C} \subseteq C^{D_{im}P} \tag{10}$$

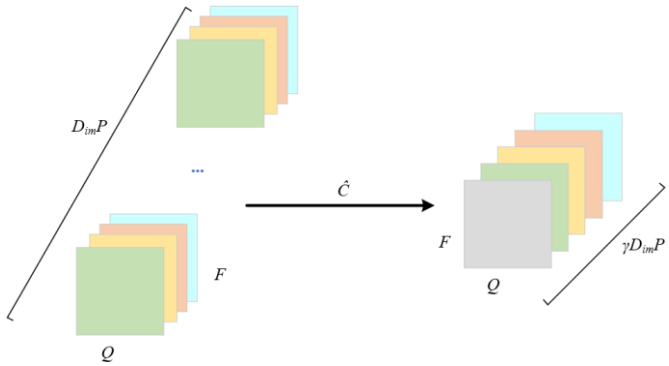


Figure 3. Flow of random dimension compression

Although the above method can significantly reduce the redundant information of video images in terms of time and channel, an improper compression scale can easily cause information loss of student behavior during practice and training in the time dimension, so this paper directly performed random dimension compression on the channel dimension to solve this problem, and the flow of this random dimension compression is shown in Figure 3. Assuming: $\hat{C}$ represents the random dimension compression operation, then the formulas for calculating dynamic weight after compression strategy optimization are given below:

$$\beta = Soft\max\left(G_2\varepsilon\left(G_1\left(H\left(U\left(R\left(C\right)\right)\right)\right)\right)\right) \tag{11}$$

$$\hat{C} = R\left(C\right) \tag{12}$$

$$\hat{C} = \left\{\hat{C}_0, \hat{C}_1, ..., \hat{C}_{\gamma D_{im}}\right\} \tag{13}$$

$$\hat{C} \subseteq C^{D_{im}} \tag{14}$$

Therefore, in case that bias is not considered and output dimension is $\dot{D}_{out}$, the parameter size of the fully connected layer is $\mu^2\gamma D_{im}PFQ\dot{D}_{out}$, which can be turned into $\mu^2\gamma D_{im}P\dot{D}_{out}$ by adopting 2D convolution with a l-sized kernel, such

operation can reduce parameter size and make the fully connected layer lighter, the following formula calculates the dynamic attention weight after improvement:

$$\beta = Soft\max\left(G_2\varepsilon\left(G_1\left(H\left(R\left(U\left(C\right)\right)\right)\right)\right)\right) \tag{15}$$

If the adaptive convolution operation of convolution kernel size is adopted, then the convolution kernel size at this time can be calculated by the following formula:

$$l = \phi\left(\dot{D}_{im}\right) = \left|\frac{\log_2\left(\dot{D}_{im}\right)}{\alpha} + \frac{y}{\alpha}\right|_{odd} \tag{16}$$

When l value is solved in the attention mechanism, the $\dot{D}_{im}$ of l value in $G_1(.)$ and $G_2(.)$ is $\gamma\dot{D}_{im}P$.

The improved dynamic convolution kernel is plug-and-play and can be applied to lightweight convolution neural networks. The depth-separable convolution with a kernel size of 3×3×3 can be replaced by a dynamic convolution kernel. Assuming: β_i^* represents the weight of the i-th attention weight, L represents the number of dynamic convolution kernels, then the convolution process is given by the following formula:

$$\begin{aligned} g\left(a_p\right) &= \begin{bmatrix} \bar{B}_1^{'} \\ \bar{B}_2^{'} \\ \cdots \\ \bar{B}_{D_{im}}^{'} \end{bmatrix} = \left[\beta_1^* \begin{bmatrix} \bar{q}_{11}^1 & 0 & \cdots & 0 \\ 0 & \bar{q}_{22}^1 & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & \bar{q}_{D_{im}D_{im}}^1 \end{bmatrix} \right] \\ &+\beta_2^* \begin{bmatrix} \bar{q}_{11}^2 & 0 & \cdots & 0 \\ 0 & \bar{q}_{22}^2 & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & \bar{q}_{D_{im}D_{im}}^2 \end{bmatrix} \\ &+\ldots+\beta_l^* \left[\begin{bmatrix} \bar{q}_{11}^l & 0 & \cdots & 0 \\ 0 & \bar{q}_{22}^l & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & \bar{q}_{D_{im}D_{im}}^l \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ \cdots \\ A_{D_{im}} \end{bmatrix} \right] \end{aligned} \tag{17}$$

When 3D point convolution of structure blocks in the network is replaced by dynamic convolution kernel, the convolution process can be expressed as:

$$\begin{aligned} g\left(a_p\right) &= \begin{bmatrix} \bar{B}_1^{'} \\ \bar{B}_2^{'} \\ \cdots \\ \bar{B}_{D_{im}}^{'} \end{bmatrix} = \left[\beta_1^* \begin{bmatrix} \bar{q}_{11}^1 & \bar{q}_{12}^1 & \cdots & \bar{q}_{1D_{im}}^1 \\ \bar{q}_{21}^1 & \bar{q}_{22}^1 & \cdots & \bar{q}_{2D_{im}}^1 \\ \cdots & \cdots & \ddots & \cdots \\ \bar{q}_{D_{out}1}^1 & 0 & \cdots & \bar{q}_{D_{out}D_{im}}^1 \end{bmatrix} \right] \\ &+\beta_2^* \begin{bmatrix} \bar{q}_{11}^2 & \bar{q}_{12}^2 & \cdots & \bar{q}_{1D_{im}}^2 \\ \bar{q}_{21}^2 & \bar{q}_{22}^2 & \cdots & \bar{q}_{2D_{im}}^2 \\ \cdots & \cdots & \ddots & \cdots \\ \bar{q}_{D_{out}1}^2 & \bar{q}_{D_{out}2}^2 & \cdots & \bar{q}_{D_{iout}D_{im}}^2 \end{bmatrix} \\ &+\ldots+\beta_l^* \left[\begin{bmatrix} \bar{q}_{11}^L & \bar{q}_{12}^L & \cdots & \bar{q}_{1D_{im}}^L \\ \bar{q}_{21}^L & \bar{q}_{22}^L & \cdots & \bar{q}_{2D_{im}}^L \\ \cdots & \cdots & \ddots & \cdots \\ \bar{q}_{D_{out}1}^L & \bar{q}_{D_{out}2}^L & \cdots & \bar{q}_{D_{iout}D_{im}}^L \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ \cdots \\ A_{D_{im}} \end{bmatrix} \right] \end{aligned} \tag{18}$$

According to above formula, compared with the convolution kernel before improvement, the improved convolution kernel only corrects the weight size when performing depth-separable convolution operations, when performing point-wise convolution, it can be considered as ordinary convolution with a kernel size of 1.

5. EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, for different fault equipment in each step of the equipment fault emergency response task, we designed several sets of actions correspond to different emergency response steps (Table 1). Action 1 refers to the state in which a student stands still with both arms hanging naturally at his/her sides. Actions 2-6 are the operations of turning the screw of a fault equipment, they are dynamic actions of students 1 and 2 extracted from steps 1-6. Overall speaking, by identifying the actions of students 1 and 2 in the current step, we can evaluate whether the behavior sequence and operations of students can meet requirements or not.

Table 1. Action design for the equipment fault emergency response task

Action No.	Corresponding step	Type	Action description
Action 1	-	Static	Students 1 and 2 stand relaxed with arms hanging naturally at their sides
Action 2	Step 1	Dynamic	Student 1 bends down near the screw position
Action 3	Step 2	Dynamic	Student 2 hands a screw to Student 1, Student 1 takes the screw and faces to the right
Action 4	Step 3	Dynamic	Student 1 presses the screw with left hand
Action 5	Step 4	Dynamic	Student 2 hands a wrench to Student 1, Student 1 takes the hexagon wrench with right hand and extends forward
Action 6	Step 5	Dynamic	Student 1 turns the screw
Action 7	Step 6	Dynamic	Student 1 rises to the left and returns the wrench to Student 2

At first, the collected video images of 6 steps were pre-processed, an example of angle change of figure in the image is given in Figure 4. Then, after ensuring all figures in images were in a normal horizontal state, abnormal data frames in the video image sequence were identified and processed. Taking the action image data of the head area of a student during step 2 as an example, the detection and removal results of abnormal image frames are shown in Figure 5. As can be seen from the figure, in the sample image sequence, three abnormal image frames had been successfully removed by the method proposed in this paper.

To enhance model validity, this paper built a 4-layer network model and performed random dropout experiment to prevent model structure from being too complicated. Specifically, in the experiment, the weight of 15% neurons in convolution layer 2 of the model was set to 0, that is, these

neurons didn't work during model training. The performance of the trained model on test set and validation set is given in Figure 6. Obviously, random dropout can shorten training time and effectively prevent network over-fitting. But after random dropout, the model showed difficulty in convergence, so the identification model constructed in this paper is not applicable to random dropout.

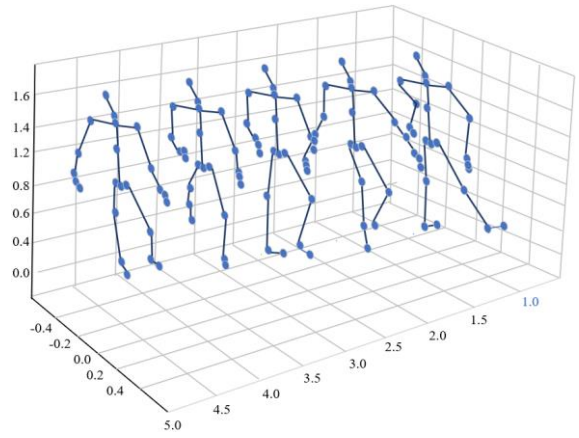
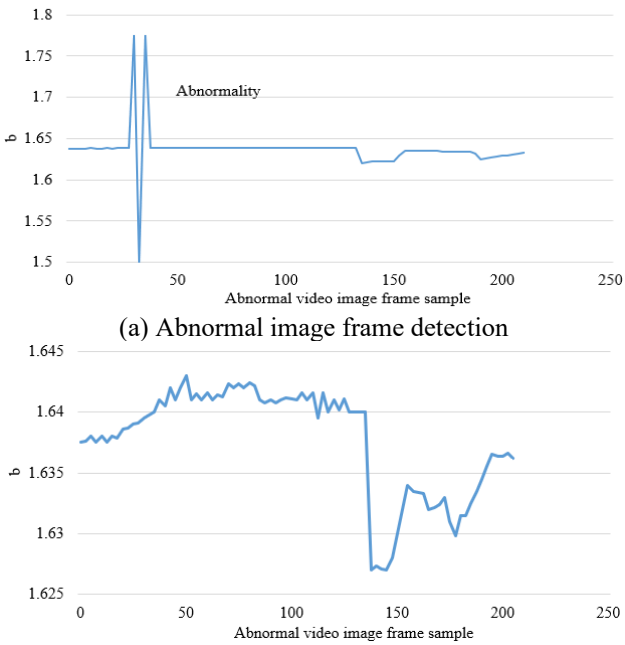


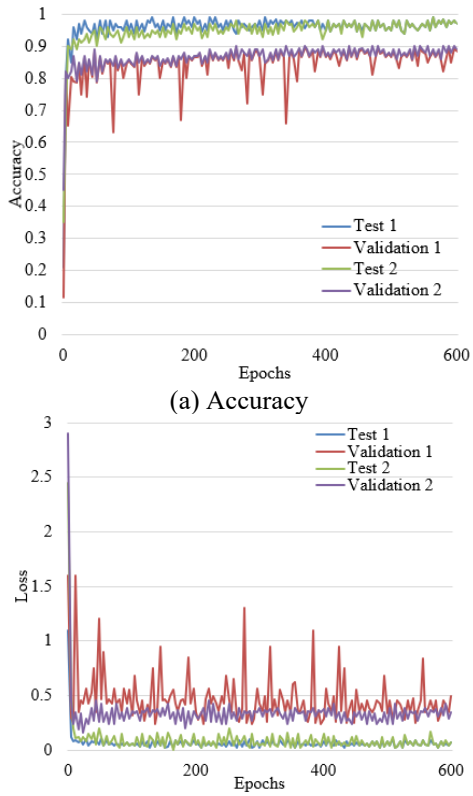
Figure 4. Angle change of figure in image



(a) Abnormal image frame detection

(b) Abnormal image frame removal

Figure 5. Abnormal image frame detection and removal



(a) Accuracy

(b) Loss value

Figure 6. Random dropout results of the model

After parameter debugging and setting, in order to verify the validity of the model, comparative experiment was performed on the proposed model, conventional CNN (convolution neural network), and conventional LSTM (long and short term memory network), and the model performance comparison results at different alignment scales are summarized in Table 2. According to the table, identification accuracy of the proposed model is better than that of CNN and LSTM, which has verified the validity of the proposed model, the model can effectively identify students' actions during practice and training. When image alignment length is 400, identification accuracy of the proposed model is the highest, reaching 98.69%. Compared with CNN and LSTM, the stable training time of the proposed model is longer, this is because the proposed model has introduced the optimized 3D dynamic convolution kernel of efficient attention, but in terms of the equipment failure emergency response task performed by students, the influence of a slightly longer training process is not much.

Table 2. Model performance comparison results at different alignment scales

Alignment scale	Model	Identification accuracy (%)	Training time (min)
200	The proposed model	97.78	17.85
	CNN	91.49	8.25
	LSTM	90.58	12.59
400	The proposed model	98.69	36.24
	CNN	93.43	22.23
	LSTM	91.92	28.46
600	The proposed model	97.87	78.55
	CNN	90.76	52.05
	LSTM	90.33	63.46
800	The proposed model	96.12	168.59
	CNN	89.59	107.15
	LSTM	86.21	128.34

Table 3. Model identification accuracy for 7 actions

Model	Action No.	Identification accuracy (%)	Action No.	Identification accuracy (%)
The proposed model	Action 1	99.85	Action 5	98.58
	Action 2	96.86	Action 6	99.25
	Action 3	95.99	Action 7	92.34
	Action 4	99.03		
CNN	Action 1	99.41	Action 5	93.34
	Action 2	88.46	Action 6	94.25
	Action 3	90.76	Action 7	90.68
	Action 4	91.35		
LSTM	Action 1	98.35	Action 5	90.13
	Action 2	89.56	Action 6	91.79
	Action 3	89.03	Action 7	89.08
	Action 4	90.89		

Table 3 summarizes model identification accuracy of 7 actions. In actual practice and training scenario, the models built for different actions differ in identification accuracy. Obviously, for action 1 (static state), all three models show a high accuracy; but for actions 3-5, since the models need to identify actions of two students, the identification accuracy is lower. However, on the whole, for the 7 actions, the identification accuracy of the proposed model is higher than that of CNN and LSTM.

6. CONCLUSION

This paper took the property management major as the subject and studied the identification of student behavior during practice and training based on video image. At first, students' practice and training content was divided into three aspects, a task of asking students to cooperate with each other to deal with an equipment failure emergency was adopted for the research, and a research idea of helping teachers figure out students' status during practice and training via identifying their actions and intentions during the said activities was determined. Then, a few pre-processing operations were performed on the captured video images of student behavior during practice and training, including removing abnormal image frames, filtering, and aligning, etc. After that, based on the collected video image data, the dynamic convolution kernel was improved and optimized, and a lightweight convolution network model was built for identifying student behavior during practice and training. In the experiment, this paper designed specific actions for the equipment fault emergency response task, pre-processed the collected video images of 6 steps, and verified the validity of the pre-processing steps; then a 4-layer network model was built and subjected to random dropout experiment. At last, performance of the proposed model was compared with conventional CNN and LSTM in experiment, and the results proved that the identification accuracy of the proposed model for 7 actions is higher than that of CNN and LSTM.

ACKNOWLEDGEMENT

This paper was supported by Hangzhou Research Base of Philosophy and Social Sciences (Hangzhou Institute for the Integration of Industry and Education) Pre-establishment Project (2022), General scientific research projects of Zhejiang Provincial Department of Education (Grant No.: Y202250697), Higher Education Research Project of Zhejiang Higher Education Association (Grant No.: KT2022279), Topic of the Research Conference on Party Construction and Ideological and Political Work of Hangzhou Vocational and Technical College (Grant No.: Dy2023010), Research project of the Communist Youth League in schools in Hangzhou (Grant No.: hzxx22048), Education and Teaching Reform Research Project of Hangzhou Vocational and Technical College (2022), and Hangzhou Philosophy and Social Science Research Base (Hangzhou Industry-education Integration Research Institute) Project (2021).

REFERENCES

[1] Yu, S. (2022). Research on the effectiveness of labor education for college students based on big data technology. Advances in Multimedia, 2022: 5328851. https://doi.org/10.1155/2022/5328851

[2] Wang, Y. (2022). Research on the labor education practice project of normal students under the background of artificial intelligence. In Artificial Intelligence in China: Proceedings of the 3rd International Conference on Artificial Intelligence in China, pp. 261-267. https://doi.org/10.1007/978-981-16-9423-3_33

[3] Aguelo, A., Coma-Roselló, T., Vicente-Sánchez, E., Baldassarri, S. (2022). Creating a network for employability: the creation of useful networks to help university students enter the labor market. IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, 17(3): 215-222. https://doi.org/10.1109/RITA.2022.3191258

[4] Zuo, B., Gao, J. (2022). Construction and coordination mechanism of college students' employment and labor relations in the Internet+ environment. International Journal of Emerging Technologies in Learning, 17(19): 135-149. http://dx.doi.org/10.3991/ijet.v17i19.34517

[5] Azan, W., Valiorgue, P., Peyrol, E., Hadid, P.H.B., Li, Y., Miranda, L.F.M. (2022). Proposal for an integrative performance framework based on Distributed Ledger Technology dedicated to higher education students entering the labor market. In 2022 IEEE 6th International Conference on Logistics Operations Management (GOL), Strasbourg, France, pp. 1-6. https://doi.org/10.1109/GOL53975.2022.9820106

[6] Yang, R., Zhang, Y., Liu, Y., Yang, J. (2022). Effects of Practice on software training video for college students. In 2022 4th International Conference on Computer Science and Technologies in Education (CSTE), Xi'an, China, pp. 300-304. https://doi.org/10.1109/CSTE55932.2022.00062

[7] Mazhitova, L.K., Syzdykova, R.N., Imanbayeva, A.K. (2021). Practice-oriented model of training students in physics at a technical university. Journal of Physics: Conference Series, 1929: 012030. https://doi.org/10.1088/1742-6596/1929/1/012030

[8] Nadrljanski, M., Nemetschek, V., Sanader, A. (2020). Training of student practical training managers. In Smart Education and e-Learning 2020, pp. 575-583. https://doi.org/10.1007/978-981-15-5584-8_48

[9] Nadrljanski, M., Pavlinović, M., Šimundić, S. (2020). Professional preparation of teachers for new models of student practical training. In Smart Education and e-Learning 2020, pp. 585-593. https://doi.org/10.1007/978-981-15-5584-8_49

[10] Rosch, D.M., Imoukhuede, P.I. (2016). Improving bioengineering student leadership identity via training and practice within the core-course. Annals of Biomedical Engineering, 44: 3606-3618. https://doi.org/10.1007/s10439-016-1684-5

[11] Nadrljanski, D., Vidović, K. (2020). Student practical training as an education factor. In Smart Education and e-Learning 2020, pp. 565-573. https://doi.org/10.1007/978-981-15-5584-8_47

[12] Uchida, H., Yuasa, K. (2019). A practice report on the active learning using business game for the teacher training students. In General Conference on Emerging Arts of Research on Management and Administration, pp. 42-48. https://doi.org/10.1007/978-981-13-6936-0_5

[13] Shi, J., Qi, E.S. (2013). Analysis of professional practice ability training of the student based on the extracurricular activities. In International Asia Conference on Industrial Engineering and Management Innovation (IEMI2012) Proceedings: Core Areas of Industrial Engineering, pp. 1691-1698. https://doi.org/10.1007/978-3-642-38445-5_178

[14] Lei, B., Liu, W., Shi, J., Yao, T., Wang, W., Hu, H. (2017). Exploration and practice of the cultivation of optoelectronic innovative talents based on the students innovation training program. In ETOP 2017 Proceedings, p. 1045242.

[15] Marqués-Molías, L., Esteve-González, V., Holgado-Garcia, J., Cela-Ranilla, J., Sánchez-Caballé, A. (2016). Student perceptions of ePortfolio as competence assessment during the practical training period for early childhood and primary school teaching. In European Conference on e-Learning, pp. 777-781.

[16] Li, H., Tang, Y., Huang, G. (2022). STEAM curriculum design and practical research: training students' scientific advanced thinking ability. In 2022 10th International Conference on Information and Education Technology (ICIET), Matsue, Japan, pp. 244-249. https://doi.org/10.1109/ICIET55102.2022.9779031

[17] Tarjányi, N., Tarjányiová, G. (2022). Video analysis of physical phenomena as a training of students for solving practical technical problems. In 2022 ELEKTRO (ELEKTRO), Krakow, Poland, pp. 1-4. https://doi.org/10.1109/ELEKTRO53996.2022.9803358

[18] Dumitru, D., Minciu, M. (2023). Better teacher-Better critical thinker. good practices for pre-service teacher training students in economics in synchronous online classes. In Technology and Innovation in Learning, Teaching and Education: Third International Conference, TECH-EDU 2022, Lisbon, Portugal, pp. 283-293. https://doi.org/10.1007/978-3-031-22918-3_22

[19] Chen, G., Ji, J., Huang, C. (2022). Student classroom behavior recognition based on openpose and deep learning. In 2022 7th International Conference on Intelligent Computing and Signal Processing, Xi'an, China, pp. 576-579. https://doi.org/10.1109/ICSP54964.2022.9778501

[20] Xiao, T., He, X., Wu, J. (2022). Student classroom behavior recognition and evaluation system based on YOLOX. In 2nd International Conference on Signal Image Processing and Communication, Qingdao, China, pp. 581-586. https://doi.org/10.1117/12.2644211

[21] Pang, C. (2021). Simulation of student classroom behavior recognition based on cluster analysis and random forest algorithm. Journal of Intelligent & Fuzzy Systems, 40(2): 2421-2431. https://doi.org/10.3233/JIFS-189237

[22] Liu, H., Ao, W., Hong, J. (2021). Student abnormal behavior recognition in classroom video based on deep learning. In Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering, Xiamen China, pp. 664-671. https://doi.org/10.1145/3501409.3501529