Recognition of Teaching Features and Behaviors in Online Open Courses Based on Image Processing

Sheliang Li, Huaqi Chai[*]

School of Management, Northwestern Polytechnical University, Xi'an 710072, China

Corresponding Author Email: chaifam@nwpu.edu.cn

ABSTRACT

High-quality online open courses have a wide audience. To further improve the quality of these courses, it is critical to analyze the teaching behaviors in class, which are the manifestation of the overall quality of the teacher. Considering the popularity of image processing-based behavior recognition in many disciplines, this paper explores deep into the teaching features and behaviors in online open courses based on image processing. Firstly, a coding scale was designed for teaching behaviors in online open courses. Next, the principle of optical flow solving was explained for teaching video images. Then, a teaching behavior feature extraction model was established based on dual-flow deep CNN, and used to extract the key points of teacher body and the behavior features of the teacher. After that, a teaching behavior recognition method was developed combining histogram of oriented gradients (HOG) and support vector machine (SVM) to accurately allocate the teaching features and behaviors to the corresponding teaching links. Finally, the proposed model was proved effective through experiments. Based on the recognized teaching behaviors, the frequency and duration of such behaviors were subject to comparative analysis, revealing the teaching features in high-quality online open courses.

1. INTRODUCTION

The rapid development of the Internet heralds the age of knowledge economy. Featured by openness and sharing, open educational resources greatly facilitate the learning and education around the world. Over the years, large public education systems have been formed and spread across the globe [1]. The Chinese Ministry of Education has released a series of policies on teaching quality and reform, calling for the construction of high-quality online open courses. Thanks to the efforts of domestic colleges, a total of 2,000 high-quality online open courses have gone live, all are freely accessible by the public [2-4].

High-quality online open courses have a wide audience. To further improve the quality of these courses, it is critical to analyze the teaching behaviors in class, which are the manifestation of the overall quality of the teacher [5]. Based on the selected cases of high-quality classroom teaching videos, the teaching features and behaviors can be quantified and qualified through video image analysis. Besides, this approach helps to explore the structure of teaching behaviors in online open courses, the knowledge framework lectured by the teacher, and the association between the two items.

In recent years, behavior recognition based on image processing has been applied to medical care, sports, smart home, and many other fields, attracting extensive attention from domestic and foreign scholars [6, 7]. To identify the abnormal behaviors of passengers, Fornaser et al. [8] implemented image processing in the safety monitoring of elevator car, introduced optical flow features into the reconstructed three-dimensional (3D) convolutional neural network (CNN), and analyzed the image features of elevator

car monitoring video by adjusting learning rate and optimizing transfer learning; In this way, the accuracy of the abnormal behavior recognition was increased to 95.1%.

The traditional feature extraction algorithms face several problems: the extracted features are not diverse enough; the similar and complex behaviors are not accurately recognized [9-12]. To solve these problems, Fuentes et al. [13] constructed an attack behavior recognition system supported by the big data of sensors, trying to recognize the attacks by detainees in smart prisons, built up a behavior recognition network that extracts spatiotemporal features and statistical features, and included the squeeze-stimulus module based on space and channels into the network, thereby elevating the recognition accuracy of network behaviors.

Compared with single-person behavior, the recognition of multi-person interactions is very difficult, requiring the extraction of high dimensional features [14-16]. Using Canny operator and local binary pattern (LBP), Lentzas and Vrakas [17] extracted body edge features and text features from video images, and described the background with complex dynamic features with the optical flow histogram, thereby preventing the recognition of multi-person interactions from being affected by light intensity and angle. Martinelli et al. [18] constructed a sparse coding space pyramid matching model for the coding and pooling of static fusion features and optical flow trajectory in video images, and realized the extraction of low-rank matrix and classified recognition of human behavior features, with the help of robust principal component analysis (PCA) and support vector machine (SVM). Lande and Gejji [19] established a VGG deep CNN with two pretrained branches: one branch maps the confidence at the key points in image feature region, and the other regresses the correlation

vectors between the key points; the abnormal behaviors, which are difficult to define or distinguish, the dynamic and static features of the target were combined through mathematical statistics before being processed.

In education and teaching, human behavior recognition is mostly adopted to analyze the student behaviors in the classroom environment [20-23]. Through optimization and clustering of skeletal node behaviors, Mabrouk and Zagrouba [24] proposed a group discovery algorithm of multi-person interactions, reduced the dimensionality of bone data features by the PCA, and classified the features with a multi-kernel SVM classifier; their method was proved valid and effective on measured datasets. To identify teacher-student interactions in a complex classroom environment, Jaouedi et al. [25] constructed a teacher-student interaction database containing five types of behaviors, trained the transfer learning model with you look only once (YOLO) V3 and the Xception model, and realized the visualized output of the detected and identified teacher-student interactions.

To sum up, the scholars engaging in behavior recognition rarely take the teachers of high-quality online open courses and their classroom teaching videos as the research carrier. Even fewer researchers have identified the teaching link or analyzed the teaching features and behaviors identified in the videos.

To make up for the gap, this paper explores deep into the teaching features and behaviors in online open courses based on image processing. Firstly, Section 2 designs the coding scale for teaching behaviors in online open courses. Next, Section 3 explains the principle of optical flow solving for teaching video images. Then, Section 4 sets up a teaching behavior feature extraction model based on dual-flow deep CNN, which consists of a channel for key point prediction and the other for feature prediction, and uses the model to extract the key points of teacher body and the behavior features of the teacher. After that, Section 5 presents a teaching behavior recognition method combining histogram of oriented gradients (HOG) and SVM, and accurately allocates the teaching features and behaviors to the corresponding teaching links. Finally, the proposed model was proved effective through experiments. Based on the recognized teaching behaviors, the frequency and duration of such behaviors were subject to comparative analysis, and the features of teaching behaviors in online open courses were fully summarized.

2. DESIGN OF CODING SCALE

This paper mainly discusses the teaching behaviors and their corresponding teaching links in the videos of high-quality online open courses. Drawing on the cases of information technology-based coding of teaching media on classroom behaviors, the authors optimized the coding system for teaching behaviors in online open courses, and summarized the codes in the dimension of teaching behaviors and the dimension of teaching links. As shown in Table 1, there are 20 codes in the dimension of teaching behaviors, including 8 self-control behaviors, 1 teacher-student simultaneous behavior, 9 response behaviors influenced by students, and 2 uncontrollable behaviors; there are three core pedagogies and four auxiliary teaching links in the dimension of teaching links.

The video analysis software ELAN was employed to acquire video images of online open courses. Under a fixed sampling interval, the teaching behaviors and teaching links in the images were obtained through repeated observations, and used to verify the subsequent behavior recognition model and teaching link classification model.

Table 1. Coding of teaching behaviors in online open courses

Dimensions	Code	Content	Type
Teaching behaviors	1	Asking questions	Self-control behaviors
	2	Praise or encouragement	
	3	Explanation	
	4	Coaching	
	5	Giving an instruction	
	6	Physical demonstration	
	7	Playing video	
	8	Emphasis	
	9	Chatting	Teacher-student simultaneous behavior
	10	Responding to passive statement by students	Response behaviors influenced by students
	11	Responding to active statement by students	
	12	Responding to student speeches	
	13	Responding to student presentation	
	14	Triggering student thinking through discussion	
	15	Correcting errors	
	16	Guiding student-student mutual evaluation	
	17	Responding to student summary	
	18	Responding to student evaluation	
	19	Silence	Uncontrollable behaviors
	20	Confusion	
Teaching links	TKT	Traditional knowledge teaching	Core pedagogies
	IT	Interactive teaching	
	HT	Heuristic teaching	
	SM	Self-evaluation and mutual evaluation	Auxiliary teaching links
	DIS	Discussion	
	SI	Summary and induction	
	ET	Experiment and training	

3. PRINCIPLE OF OPTICAL FLOW SOLVING

If the optical flow representing the brightness in the teaching video of online open course changes smoothly and continuously, the change rate of the optical flow can be considered as zero. Let P_i and P_{i+1} be the current frame and next frame in the video, respectively; $f=(f_1(a), f_2(a))^T$ be the two-dimensional (2D) displacement field of pixel a. Then, the optical flow solving method can be optimized by combining image grayscale with the 2D velocity field:

$$\min_f \left\{ \int_T \begin{array}{l} \left(|\nabla f_1|^2 + |\nabla f_2|^2 \right) da \\ + \gamma \int_T \left(P_{i+1}(a+f(a)) - P_i(a) \right)^2 da \end{array} \right\} \quad (1)$$

Formula (1) shows that the optical flow consists of an L_2 regular term that keeps its change smooth and a constraint of optical flow data. To make the algorithm more robust and suitable for boundary areas with a large gradient, the L_2 regular term was substituted with a more robust term: the brightness difference between pixels was characterized by the similarity score between the current and next frames:

$$\int_T \left\{ \begin{array}{l} \Psi\left((f, \nabla f, \ldots)\right) \\ + \gamma \Phi\left[P_i(a) - P_{i+1}(a+f(a)) \right] \end{array} \right\} da \quad (2)$$

Formula (2) shows that the optical flow now encompasses a priori regular term $\Psi(\nabla f)$ and image data fidelity $\Phi(a)$. The importance between the two parts is balanced by a weight coefficient γ. Suppose $\Phi(a)=|a|$ and $\Psi(\nabla f)=|\nabla f|$. Then, formula (2) can be converted into a function composed of a fully variable regular term and a data penalty:

$$\int_T \left\{ |\nabla f| + \gamma |P_i(a) - P_{i+1}(a+f(a))| \right\} da \quad (3)$$

The first-order Taylor expansion of $P_{i+1}(a+f(a))$ at $a+f_0$ can be expressed as:

$$\begin{aligned} P_{i+1}(a+f) &= P_{i+1}(a+f_0) \\ &+ f\nabla P_{i+1}(a+f_0) - f_0\nabla P_{i+1}(a+f_0) \end{aligned} \quad (4)$$

Then, the residual of the first-order video image can be described as:

$$\begin{aligned} e(f) &= P_{i+1}(a+f_0) + f\nabla P_{i+1}(a+f_0) \\ &- f_0\nabla P_{i+1}(a+f_0) - P_i(a) \end{aligned} \quad (5)$$

Let f_r be the r-th component of f. Then, the optical flow solving can be further simplified as:

$$\int_T \left\{ \gamma |e(f)| + |\nabla f_r| \right\} da \quad (6)$$

Introducing a small constant ε and an auxiliary variable q that approximates f:

$$\int_T \left\{ \sum_r |\nabla f_r| + \sum_r \frac{1}{2\varepsilon}(f_r - q_r)^2 + \gamma |e(f)| \right\} da \quad (7)$$

Fixing the value of q_r, the following formula can be solved:

$$\min_{f_r} \int_T \left\{ |\nabla f_r| + \frac{1}{2\varepsilon}(f_r - q_r)^2 \right\} da \quad (8)$$

Fixing the value of f_r, the following formula can be solved:

$$\min_q \sum_r \frac{1}{2\varepsilon}(f_r - q_r)^2 + \gamma |e(q)| \quad (9)$$

The obtained q value can be expressed as:

$$q = f + \begin{cases} \gamma\varepsilon\nabla P_{i+1} & \text{if } e(f) < -\gamma\varepsilon|\nabla P_{i+1}|^2 \\ -\gamma\varepsilon\nabla P_{i+1} & \text{if } e(f) > \gamma\varepsilon|\nabla P_{i+1}|^2 \\ -e(f)\nabla P_{i+1}/|\nabla P_{i+1}|^2 & \text{if } e(f) \leq \gamma\varepsilon|\nabla P_{i+1}|^2 \end{cases} \quad (10)$$

Our optical flow method was realized based on OpenCV, an open-source visual library, using Python. The video frames and optical flow images were extracted manually from high-quality online open courses.

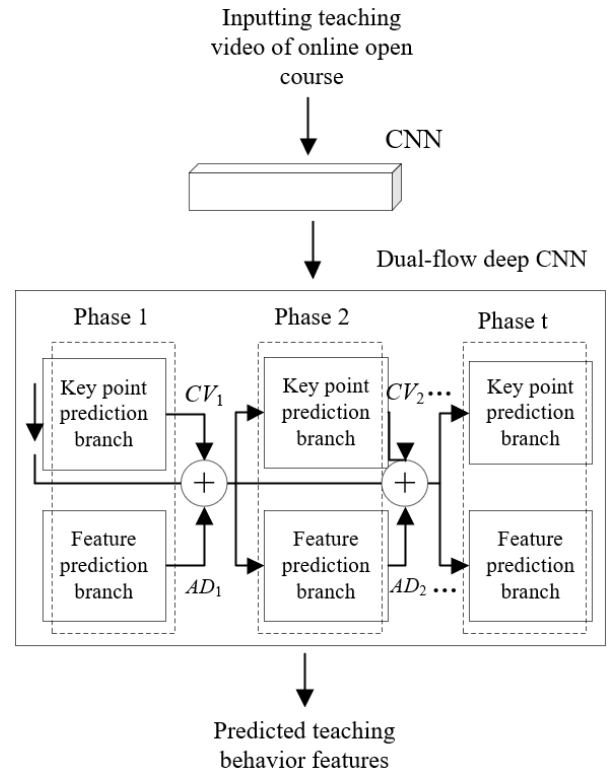## 4. TEACHING BEHAVIOR FEATURE EXTRACTION MODEL

Figure 1. Structure of teaching behavior feature extraction network

This section mainly elaborates the structure and principles of the teacher body target detection model. The traditional

dual-flow neural network extracts the dynamic and static features of the image via the time channel and space channel, respectively. In our dual-flow deep CNN, there are two channels, namely, key point prediction branch and feature prediction branch, responsible to extract the key points of teacher body and features of teaching behaviors. Figure 1 illustrates the structure of the proposed dual-flow deep CNN. The predicted results of the two branches were merged through multiple phases into the skeleton map of the instantaneous behaviors of the teacher. Figure 2 provides the structure of the CNN for preprocessing video images.
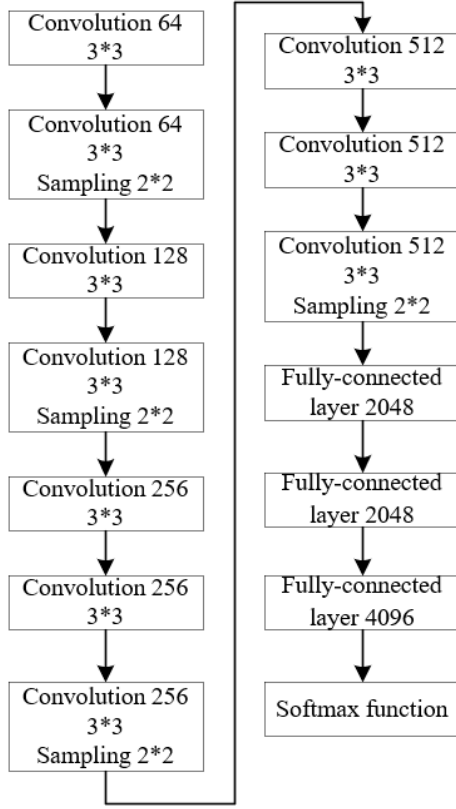

Figure 2. Structure of the proposed CNN

The key point prediction branch can forecast and track the key points in the teacher body, which appears in the teaching video of online open course. Let $CV_i=c$ be the confidence of the prediction of each key point; FA be the feature map set; $p_1^i(FA)$ be the probability of predicting the i-th key point at position c of the image in the initial phase. The c value is outputted by classifier ϕ_1^1:

$$\phi_1^1\left(FA_c\right) \rightarrow \left\{p_1^i\left(FA\right)\right\}_{i\in(0,1,\dots,N)} \quad (11)$$

where, N is the number of key points. The feature prediction branch can accurately forecast the association between key points. Let $\rho_2^r(FA)$ be the association score between key points d_1 and d_2 identified in the initial phase. Then, the predicted result is outputted by classifier ϕ_2^1:

$$\phi_2^1\left(FA_c\right) \rightarrow \left\{\rho_2^r\left(FA\right)\right\}_{r\in(0,1,\dots,R)} \quad (12)$$

Starting from phase 2, the confidence map, associated domain, and feature map set outputted by the two branches in the current cycle should be merged into the input of the next cycle. After t phases, the final confidence CV_t and associated domain AD_t can be outputted. Let $\Phi(\cdot)$ be the mapping of the context feature in phase t-1. Then, the output of the key point prediction branch in phase t can be expressed as:

$$\varphi_1^t\left(FA_c^1,\Phi_t\left(c,C_{t-1},AD_{t-1}\right)\right)$$
$$\rightarrow \left\{p_1^i\left(FA,C_{t-1},AD_{t-1}\right)\right\}_{i\in(0,1,\dots,N+1)}, \forall t\geq 2 \quad (13)$$

Similarly, the mapping of the context feature of ρ_2^{t-1} can be denoted as $\Gamma(\cdot)$. Then, the output of the feature prediction branch in phase t can be expressed as:

$$\varphi_2^t\left(FA_c^1,\Phi_t\left(c,C_{t-1},AD_{t-1}\right)\right)$$
$$\rightarrow \left\{\rho_2^r\left(FA,C_{t-1},AD_{t-1}\right)\right\}_{r\in(0,1,\dots,R+1)}, \forall t\geq 2 \quad (14)$$

Before being outputted, the confidence map of key point positions and its connection method were gradually refined through multiple phases. At the end of each phase, the loss function was calculated before outputting the predicted result, such as to prevent vanishing gradients induced by too many convolutional layers in the network. Next, a weight $W(i)$ was defined to reduce the probability of incorrect positioning of key points. Let $CV_t^k(i)$ be the confidence of the i-th key point in the k-th confidence map outputted by the key point prediction branch in phase t, and $CV_{GT}^k(i)$ be the labeled confidence. Then, the loss function can be calculated by:

$$Loss_1^t = \sum_{k=1}^{K}\sum_{N}W(i)\cdot\left\|CV_t^k(i)-CV_{GT}^k(i)\right\| \quad (15)$$

Let $AD_t^r(i)$ be the vector of the i-th key point in the associated domain of the r-th body part outputted by the feature prediction branch in phase t, and $AD_{GT}^r(i)$ be the labeled vector. Then, the loss function can be calculated by:

$$Loss_2^t = \sum_{r=1}^{R}\sum_{N}W(i)\cdot\left\|AD_t^k(i)-AD_{GT}^k(i)\right\| \quad (16)$$
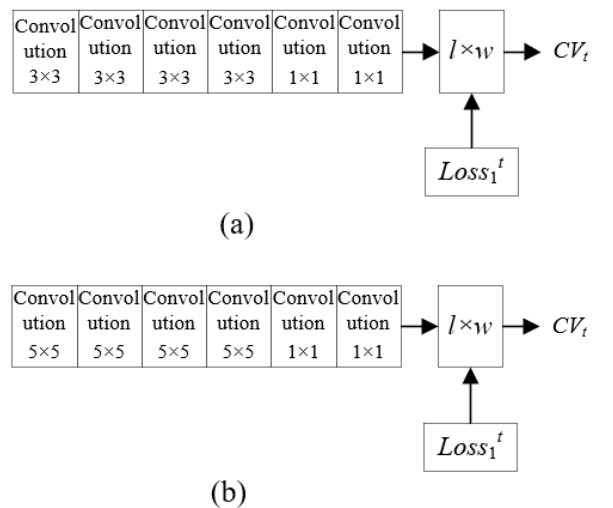

Figure 3. Structure of key point prediction branch
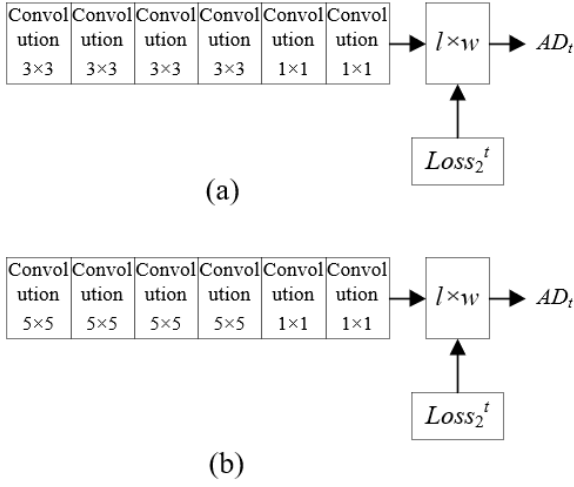
(a)

(b)

Figure 4. Structure of feature prediction branch

Figures 3 and 4 present the structures of key point prediction branch and feature prediction branch, respectively. Combining formulas (15) and (16), the overall loss function of the entire network can be obtained as:

$$Loss = \sum_{t=1}^{T}\left(Loss_1^t + Loss_2^t\right) \tag{17}$$

Let CV_t^i be the confidence at the i-th key point on teacher body corresponding to each pixel $D=(x,y)$ in video image; CV_t be the corresponding set of confidence maps; $l \times w$ be the length and width of the video image. Then, we have:

$$p_t^i[x,y] = p_t^i(FA) \tag{18}$$

If the position of CV_{GT}^i in video image obeys normal distribution, and if pixel D lies on the peak of the curve when it approaches the labeled position δ_i of the i-th key point, then the confidence CV_{GT}^i of i key points in the continuously sampled frames can be expressed as:

$$CV_{GT}^i(D) = e^{-\frac{\|D-\delta_i\|_2^2}{\sigma^2}} \tag{19}$$

$$CV_i^*(D) = \max_i CV_{GT}^i(D) \tag{20}$$

As for the connection between key points, the positions of key points g_1 and g_2 in teacher body TR are denoted as δ_{g1} and δ_{g2}, respectively. Then, the associated domain of the body, i.e., the vector at point i in the s-th limb vector field of the body, can be expressed as:

$$AD_s^*(i) = \begin{cases} \frac{\delta_{g2} - \delta_{g1}}{\|\delta_{g2} - \delta_{g1}\|_2}, \text{if i falls on TR} \\ 0, \text{otherwise} \end{cases} \tag{21}$$

If any pixel i between δ_{g1} and δ_{g2} does not fall on teacher body TR, $AD_s^*(i)$ equals zero; otherwise, $AD_s^*(i)$ equals the unit vector between δ_{g1} and δ_{g2}. Let b be the $AD_s^*(i)$ value when i falls on teacher body TR; $b_\perp$ be the unit vector perpendicular to b; τ be the width of teacher body TR. Then, whether i falls on TR can be judged by:

$$0 \le b \cdot \left(i - \delta_{g1}\right) \le \left\|\delta_{g2} - \delta_{g1}\right\|_2$$
$$and \left|b_\perp \cdot \left(i - \delta_{g1}\right)\right| \le \tau \tag{22}$$

The association between δ_{g1} and δ_{g2}, i.e., the weight of the segment between the two points, can be characterized by the linear integral values of the associated domain of each pixel on the segment between the two points:

$$LI = \int_{x=0}^{x=1} \rho_t^r\left(i(x)\right) \cdot \frac{\delta_{g2} - \delta_{g1}}{\left\|\delta_{g2} - \delta_{g1}\right\|} dx$$
$$= \int_{x=0}^{x=1} \rho_t^r\left((1-x)\delta_{g1} + x\delta_{g2}\right) \cdot \frac{\delta_{g2} - \delta_{g1}}{\left\|\delta_{g2} - \delta_{g1}\right\|} dx \tag{23}$$

5. HOG+SVM-BASED RECOGNITION OF TEACHING BEHAVIORS

In this paper, the teaching behavior features in video images are extracted and classified, using the HOG in local areas and the SVM classifier. The HOG+SVM algorithm is explained in Figure 5.
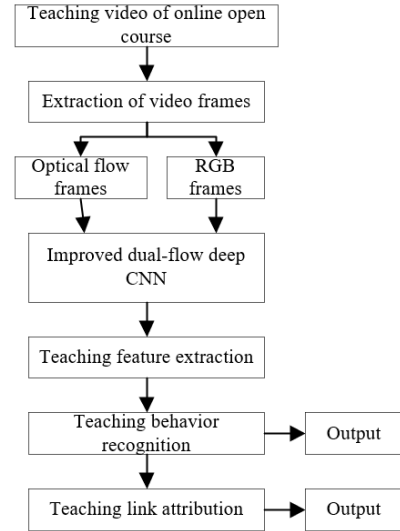

Figure 5. Workflow of teaching behavior recognition and teaching link attribution

The HOG-based feature extraction first normalizes the grayscale of the skeleton map. In the grayscale map, the pixel value can be expressed as:

$$P(x,y) = P(x,y)^{0.5} \tag{24}$$

The gradients of pixel (x, y) in horizontal x and vertical y directions can be described by:

$$\begin{cases} GA_x(x,y) = P(x+1,y) - P(x-1,y) \\ GA_y(x,y) = P(x,y+1) - P(x,y-1) \end{cases} \tag{25}$$

The gradient amplitude at (x, y) can be expressed as:

$$GA(x,y) = \sqrt{GA_x(x,y)^2 + GA_y(x,y)^2} \qquad (26)$$

The gradient direction at (x, y) can be expressed as:

$$\eta(x,y) = acrtan\frac{GA_y(x,y)}{GA_x(x,y)} \qquad (27)$$

The HOG feature descriptor was adopted to segment each video image into eight associated domains, whose gradient direction obeys the weighted distribution of gradient amplitude. Each domain corresponds to a histogram. Then, the gradient amplitudes of the interval composed of several domains were normalized, and the behavior feature vectors in all intervals were combined into the final HOG feature.
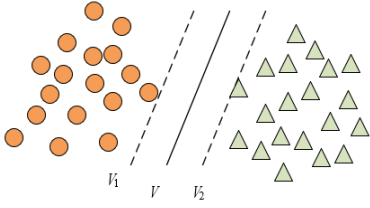

Figure 6. Data classification by SVM

Figure 6 explains the data classification by SVM. Facing the clusters of different types of samples, the cluster interval was defined as the distance between V_1 and V_2, and the optimal classification plane as V. Let I_i and E_i be a training sample and its expected class, respectively. The E_i value is either +1 or -1. Suppose $\alpha x + \beta = 0$ is the classification plane function for training samples. Then, the constraint on correct classification can be expressed as:

$$E_i(I_i\alpha + \beta) - 1 \geq 0 \qquad (28)$$

The classification interval can be expressed as:

$$\min_{\{I_i|E_i=1\}}\frac{I_i\alpha + \beta}{\|\alpha\|} - \max_{\{I_i|E_i=1\}}\frac{I_i\alpha + \beta}{\|\alpha\|} = \frac{2}{\|\alpha\|} \qquad (29)$$

The classification aims to maximize 2/||α||, i.e., minimize the following formula:

$$\Lambda(\alpha) = \frac{1}{2}\|\alpha\|^2 \qquad (30)$$

By Lagrangian multiplier method, the solving problem can be transformed into:

$$LA = \frac{1}{2}\|\alpha\|^2 - \sum_{i=1}^{M}\lambda_i E_i(I_i\alpha + \beta) + \sum_{i=1}^{M}\lambda_i \qquad (31)$$

To minimize the value of formula (31), the partial differentials of α and β were set to zero. Under the constraint of "making the gradients of LA relative to α and β zero while satisfying $\lambda_i \geq 0$", the maximum of the following formula was solved relative to i:

$$U(\lambda) = \sum_{i=1}^{M}\lambda_i - \frac{1}{2}\sum_{i=1}^{M}\lambda_i\lambda_j E_i E_j(x_i, y_j) \qquad (32)$$

Let $\lambda_i^{'}$ be the optimal solution. Then, the weight coefficient vector of the optimal classification plane can be expressed as:

$$\alpha' = \lambda_i' E_i I_i \qquad (33)$$

To solve the extreme values like maximum and minimum, the optimal solution must satisfy:

$$\lambda_i'\{[I_i\alpha + \beta]E_i - 1\} = 0 \qquad (34)$$

The teaching behaviors can be classified and attributed to teaching links by importing the optical flow data on video frames and the skeleton map of instantaneous teaching behaviors into the HOG+SVM classifier for training.

6. EXPERIMENTS AND RESULTS ANALYSIS

Table 2 displays the experimental results on feature extraction and behavior recognition of the proposed HOG+SVM classifier. The results of optical flow test and RGB test were merged, according to the test results on the image libraries of online open courses, which have different space and time attributes. As shown in Table 2, the proposed classifier achieved relatively high recognition rate, and strong resistance to disturbances, reflecting its applicability to teaching behavior recognition in videos on different scenes with different brightness levels. The recognition accuracy was even higher after feature fusion.

Table 3 compares the performance of our model with traditional dual-flow CNN, VGGNet, and ResNet in teaching behavior recognition. It can be seen that our model realized a mean recognition accuracy of 88.6%, which is higher than that of the contrastive models. Although the real-time performance was not ideal, our model consumed a shorter training time than the deep learning networks with a massive number of parameters. Despite its simplicity, our model can effectively learn the teaching behavior features.

Figure 7 compares the performance between different algorithms. The comparison further confirms the superiority of our algorithm in teaching behavior recognition over other network models. The traditional dual-flow CNN was not far behind our algorithm in terms of recognition effect. But our model was more efficient, less susceptible to scene changes, and less limited by illumination.

Table 2. Experimental results on feature extraction and behavior recognition

Image library	Branch input	Test accuracy	Training time	Feature fusion result
Image library of online open courses 1	RGB frames	85.15%	97.11s	89.3%
	Optimal flow frames	87.21%	342s	
Image library of online open courses 2	RGB frames	84.78%	125.53s	88.9%
	Optimal flow frames	86.29%	672s	

Table 3. Experimental results on teaching behavior recognition

Classification model	Mean recognition accuracy	Number of layers	Training duration
Our model	88.6%	16	5.6h
Traditional dual-flow CNN	81.7%	13	7.8h
VGGNet	80.3%	14	15.5h
ResNet	78.5%	16	21.7h

After accurately recognizing teaching behaviors, the HOG+SVM classifier was adopted to process the skeleton map corresponding to the optical flow data of video frames and the instantaneous teaching behaviors, and output the attribution results on the seven teaching links, namely, traditional knowledge teaching, interactive teaching, heuristic teaching, self-evaluation and mutual evaluation, discussion, summary and induction, and experiment and training.

Table 4 presents the experimental results on four teachers of different subjects for one class hour. During classroom teaching, "responding to student presentation", "correcting errors", and "responding to student evaluation" belong to different teaching behaviors, but are attributed to the same teaching link: "interactive teaching". The more obvious the features of the teaching behaviors in the same class, the more accurate the attribution.

Figure 7. Performance comparison between different algorithms

Based on the teaching behaviors recognized for the 4 teachers, the authors went on to compare the frequency and duration of different teaching behaviors. As shown in Figure 8, there was not a significant correlation between the frequency and duration of teaching behaviors; but the proportions of the frequency and duration of a behavior could reflect the behavior features of the corresponding teacher in online open course.

Table 4. Results of teaching link attribution

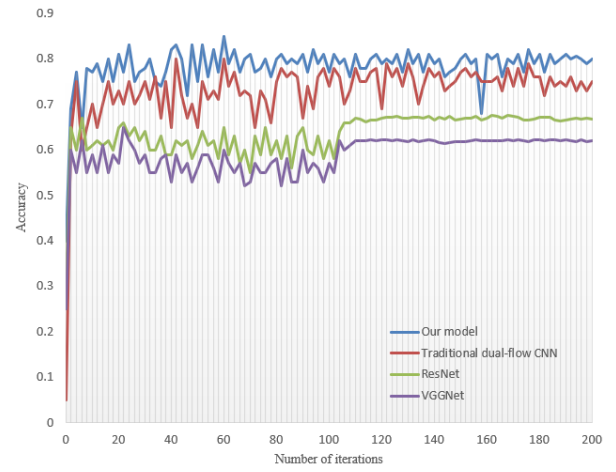|	Teacher 1	Teacher 2	Teacher 3	Teacher 4
Traditional knowledge teaching	87.14%	86.23%	82.78%	85.12%
Interactive teaching	88.27%	87.18%	90.84%	89.24%
Heuristic teaching	73.19%	72.55%	71.72%	77.38%
Self-evaluation and mutual evaluation	79.67%	77.31%	78.93%	79.43%
Discussion	91.93%	90.34%	89.87%	91.48%
Summary and induction	81.33%	80.05%	81.75%	78.82%
Experiment and training	94.49%	92.12%	91.87%	92.53%

Figure 8. Frequency and duration of different teaching behaviors

Comparing Figures 8 (a)-(d), "asking questions" occurred very frequently, while "responding to student answers", "giving an instruction", and "coaching" took large proportions. Hence, these behaviors are essential to online open course.

Heuristic and interactive teaching behaviors can stimulate student enthusiasm for teaching activities, provoking them to reflect on what they have learned. Giving an instruction with gesture helps the teacher to dominate the classroom, while bending to coach students could improve the teaching quality.

For the traditional explanation behavior, the frequency proportion was usually smaller than duration proportion, because knowledge explanation lasts from a couple of seconds to several minutes. In open course teaching, however, the explanation behavior did not take up the largest proportion. Centering on student demand, the teaching behaviors can live up to the student-oriented teaching philosophy of open course.

Figure 9 summarizes the probability of classifying each teaching behavior in online open course to each teaching link. An obvious law can be inferred from the classification probability of teaching behaviors to each link, concerning the online open courses given by four teachers on different subjects. The frequency of traditional knowledge teaching generally fell in 13-17%; the highest frequency belonged to experiment and training (26-38%); heuristic teaching and interactive teaching maintained a stable proportion of 12-32%; discussion, summary and induction, as well as self-evaluation and mutual evaluation occurred at the frequency of 22-46%. The arrangement of different teaching links reflects the teacher's understanding of teaching objectives, strategies, and processes, as well as classroom moderation. The above analysis enables teachers to reflect on their teaching strategy, and to avoid silence or frequent/persistent occurrence of a single teaching behavior.
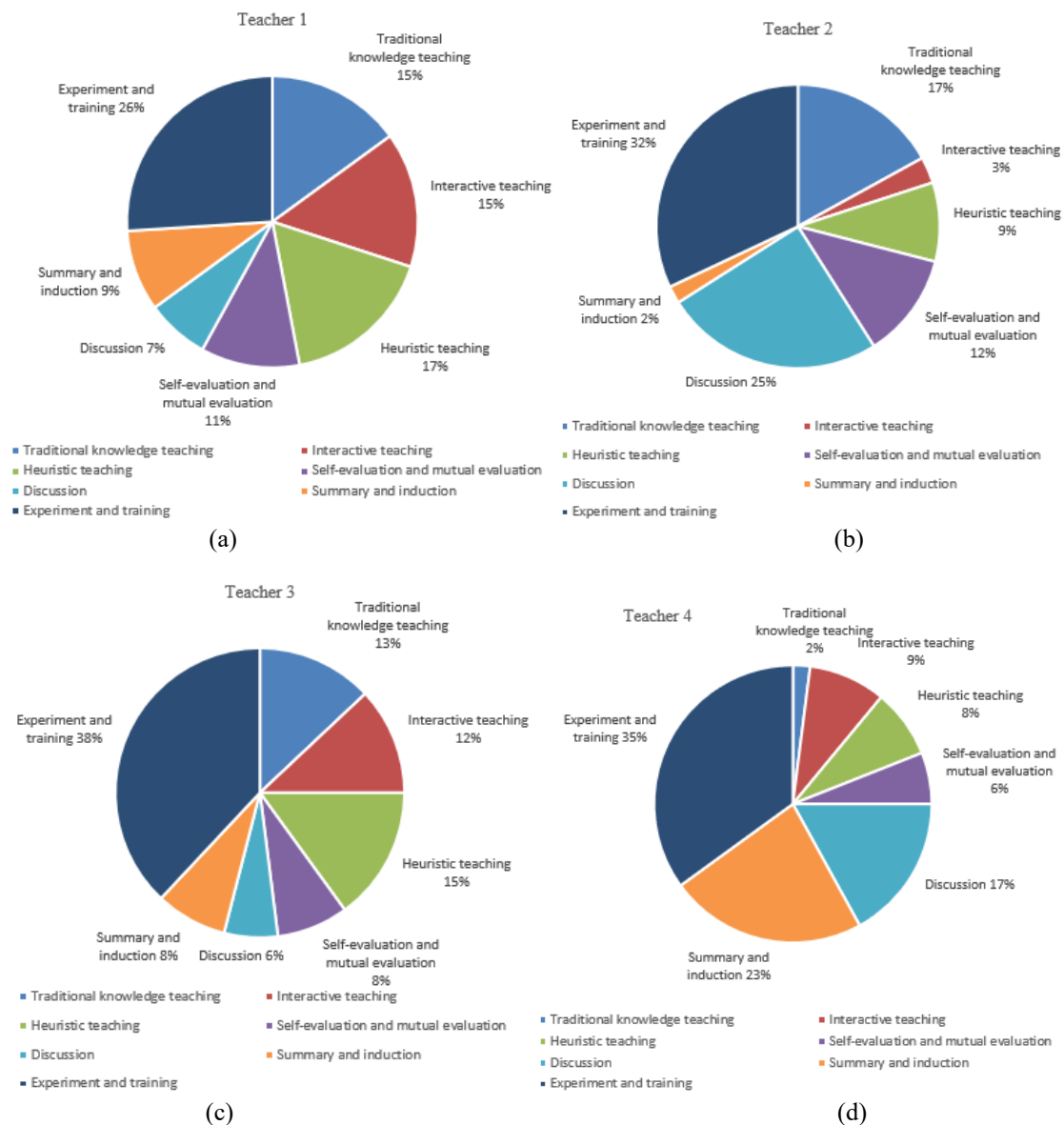

(a)

(b)

(c)

(d)

Figure 9. Classification probability of teaching behaviors to each link

7. CONCLUSIONS

Based on image processing, this paper investigates the teaching features and behaviors in online open courses. Firstly,

the authors designed the coding scale for teaching behaviors in online open courses, and expounded the philosophy of optical flow solving for teaching video images. Secondly, the authors created a teaching behavior feature extraction model

based on dual-flow deep CNN, and successfully extracted the key points of teacher body and the behavior features of the teacher. Thirdly, comparative experiments were conducted to verify the high recognition rate and strong anti-interference ability of our model after feature fusion. Fourthly, the HOG+SVM teaching behavior recognition method was introduced to accurately allocate the teaching features and behaviors to the corresponding teaching links. Fifthly, experimental results demonstrate that our model enables teachers to reflect on their teaching strategy, and to avoid silence or frequent/persistent occurrence of a single teaching behavior. Finally, the frequency and duration of recognized teaching behaviors were compared in a class hour, revealing the teaching behavior features in online open courses of different subjects.

REFERENCES

[1] Jadhav, N., Sugandhi, R. (2018). Survey on human behavior recognition using affective computing. In 2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN), 98-103. https://doi.org/10.1109/GCWCN.2018.8668632

[2] Lushan, M., Bhattacharjee, M., Ahmed, T., Rahman, M.A., Ahmed, S. (2018). Supervising vehicle using pattern recognition: Detecting unusual behavior using machine learning algorithms. In 2018 IEEE Region Ten Symposium (Tensymp), 277-281. https://doi.org/10.1109/TENCONSpring.2018.8692071

[3] De, P., Chatterjee, A., Rakshit, A. (2017). Recognition of human behavior for assisted living using dictionary learning approach. IEEE Sensors Journal, 18(6): 2434-2441. https://doi.org/10.1109/JSEN.2017.2787616

[4] Hägele, G., Sarkheyli-Hägele, A. (2020). Situational hazard recognition and risk assessment within safety-driven behavior management in the context of automated driving. In 2020 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA), 188-194. https://doi.org/10.1109/CogSIMA49017.2020.9216183

[5] Kim, J., Min, K., Jung, M., Chi, S. (2020). Occupant behavior monitoring and emergency event detection in single-person households using deep learning-based sound recognition. Building and Environment, 181: 107092. https://doi.org/10.1016/j.buildenv.2020.107092

[6] Uddin, M.Z., Hassan, M.M., Alsanad, A., Savaglio, C. (2020). A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare. Information Fusion, 55: 105-115. https://doi.org/10.1016/j.inffus.2019.08.004

[7] Chikhaoui, B., Ye, B., Mihailidis, A. (2018). Aggressive and agitated behavior recognition from accelerometer data using non-negative matrix factorization. Journal of Ambient Intelligence and Humanized Computing, 9(5): 1375-1389. https://doi.org/10.1007/s12652-017-0537-x

[8] Fornaser, A., Mizumoto, T., Suwa, H., Yasumoto, K., De Cecco, M. (2018). The influence of measurements and feature types in automatic micro-behavior recognition in meal preparation. IEEE Instrumentation & Measurement Magazine, 21(6): 10-14. https://doi.org/10.1109/MIM.2018.8573587

[9] Huang, C.W., Wu, P.W., Su, W.H., Zhu, C.Y., Kuo, S.W. (2016). Stimuli-responsive supramolecular materials: photo-tunable properties and molecular recognition behavior. Polymer Chemistry, 7(4): 795-806. https://doi.org/10.1039/C5PY01852H

[10] Batchuluun, G., Kim, Y.G., Kim, J.H., Hong, H.G., Park, K.R. (2016). Robust behavior recognition in intelligent surveillance environments. Sensors, 16(7): 1010. https://doi.org/10.3390/s16071010

[11] Eftekhari, H.R., Ghatee, M. (2018). Hybrid of discrete wavelet transform and adaptive neuro fuzzy inference system for overall driving behavior recognition. Transportation research part F: traffic psychology and behaviour, 58: 782-796. https://doi.org/10.1016/j.trf.2018.06.044

[12] Akine, S., Onuma, T., Nabeshima, T. (2018). A novel graphite-like stacking structure in a discrete molecule and its molecular recognition behavior. New Journal of Chemistry, 42(12): 9369-9372. https://doi.org/10.1039/C8NJ01315B

[13] Fuentes, A., Yoon, S., Park, J., Park, D.S. (2020). Deep learning-based hierarchical cattle behavior recognition with spatio-temporal information. Computers and Electronics in Agriculture, 177: 105627. https://doi.org/10.1016/j.compag.2020.105627

[14] Madokoro, H., Nakasho, K., Shimoi, N., Woo, H., Sato, K. (2020). Development of invisible sensors and a machine-learning-based recognition system used for early prediction of discontinuous bed-leaving behavior patterns. Sensors, 20(5): 1415. https://doi.org/10.3390/s20051415

[15] Duffy, A.G., Hughes, G.P., Ginzel, M.D., Richmond, D.S. (2018). Volatile and contact chemical cues associated with host and mate recognition behavior of Sphenophorus venatus and Sphenophorus parvulus (Coleoptera: Dryophthoridae). Journal of chemical ecology, 44(6): 556-564. https://doi.org/10.1007/s10886-018-0967-8

[16] Roviello, G.N., Oliviero, G., Di Napoli, A., Borbone, N., Piccialli, G. (2020). Synthesis, self-assembly-behavior and biomolecular recognition properties of thyminyl dipeptides. Arabian Journal of Chemistry, 13(1): 1966-1974. https://doi.org/10.1016/j.arabjc.2018.02.014

[17] Lentzas, A., Vrakas, D. (2019). Non-intrusive human activity recognition and abnormal behavior detection on elderly people: A review. Artificial Intelligence Review, 1-47. https://doi.org/10.1007%2Fs10462-019-09724-5

[18] Martinelli, F., Mercaldo, F., Orlando, A., Nardone, V., Santone, A., Sangaiah, A.K. (2020). Human behavior characterization for driving style recognition in vehicle system. Computers & Electrical Engineering, 83: 102504. https://doi.org/10.1016/j.compeleceng.2017.12.050

[19] Lande, D.N., Gejji, S.P. (2018). Molecular recognition, conformational behavior, and spectral characteristics of oxatub [4] arene macrocycle. The Journal of Physical Chemistry A, 122(2): 714-723. https://doi.org/10.1021/acs.jpca.7b12472

[20] Kwak, J., Gong, S., Sung, Y. (2016). Behavior network-based risk recognition method. In Proceedings of the 3rd International Conference on Intelligent Technologies and Engineering Systems (ICITES2014), 201-205. https://doi.org/10.1007/978-3-319-17314-6_26

[21] Lawanont, W., Inoue, M. (2018). An unsupervised learning method for perceived stress level recognition based on office working behavior. In 2018 International Conference on Electronics, Information, and

Communication (ICEIC), pp. 1-4. https://doi.org/10.23919/ELINFOCOM.2018.8330700

[22] Lindow, F., Kaiser, C., Kashevnik, A., Stocker, A. (2020). AI-based driving data analysis for behavior recognition in vehicle cabin. In 2020 27th Conference of Open Innovations Association (FRUCT), pp. 116-125. https://doi.org/10.23919/FRUCT49677.2020.9211020

[23] Reiß, S., Roitberg, A., Haurilet, M., Stiefelhagen, R. (2020). Activity-aware attributes for zero-shot driver behavior recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition

Workshops, pp. 902-903.

[24] Mabrouk, A.B., Zagrouba, E. (2018). Abnormal behavior recognition for intelligent video surveillance systems: A review. Expert Systems with Applications, 91: 480-491. https://doi.org/10.1016/j.eswa.2017.09.029

[25] Jaouedi, N., Boujnah, N., Htiwich, O., Bouhlel, M.S. (2016). Human action recognition to human behavior analysis. In 2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), pp. 263-266. https://doi.org/10.1109/SETIT.2016.7939877