ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

A Study of Paradigm Transfer and Competency Extension in Teaching Evaluation of Curriculum Civics in Meta-Universe Scenarios

Binbin Song[1,*], *Yuhan Huang*[2], *Zijin Wu*[2], *Yiwen Yuan*[2] *and Jiayi Yang*[2]

[1] Auditing School, Henan University of Economics and Low, Zhengzhou, Henan, 450046, China

[2] Accounting School, Henan University of Economics and Low, Zhengzhou, Henan, 450046, China

SUMMARY: *Smart classrooms have become a significant trend due to the high pace of development and extensive use of artificial intelligence and digital technologies, which opens up new opportunities to innovate the assessment of Civics and Political Science education. The current paper presents MASA theory as seen through the eyes of the meta universe and constructs a multimodal classroom information acquisition framework with the goal of facilitating paradigm shift and functional improvement in the process of course evaluation. Specifically, a better EfficientNet-driven model is created to recognize the facial expressions in class, then a more effective DenseNet-driven model is trained to recognize the learning behavior and action of students, which are both evaluated on a self-made dataset. Based on this, an intelligent evaluation method of Civics teaching is built based on the estimation of probability and the combination of learners expression with behavioral performance. The practical application outcomes show that out of ten students, Students 1, 2, and 3 have the best learning condition with each recording a score of over 0.400 and the other ones require further improvement. The offered approach will be able to track the status of the students during the learning process, and successfully complete the evaluation procedures of the Civics course. Such results enable offering more specific guidance and counseling, which in turn supports the overall, healthy, and sustainable development of students.*

KEYWORDS: *MASA theory; EfficientNet network; DenseNet network; probability value calculation; evaluation of course Civics teaching*

1 Introduction

Conventional ideological and political theory instruction has always been inclined to adopt a hard irrigation strategy towards theoretical training [1]. In contrast, ideological and political teaching nowadays emphasizes more on integrating both explicit teaching and implicit influence and advocates a soft irrigation model that provides a supportive learning setting where professional education and ideological-political education are advanced simultaneously and support one another [2, 3].The subject Curriculum Civics and Politics is very close to the pedagogical concept of combining the professional course with ideological and political education. It can thus be viewed as a successful method of enhancing educational synergy between disciplinary teaching and values-based cultivation and hence promote the overall development of students [4].Under these circumstances, it has been widely debated in

*hncybbs@126.com
https://doi.org/10.65102/is2026375
Published: 30 April, 2026

governments, societies, and universities how to develop the ideological and political aspect of curriculum teaching, and what kind of educational effects such integration may bring..

Curriculum Civics has reached a critical point in the new age with increased development, and further reforms have been undertaken. This method of teaching is now being introduced in a wide variety of universities and the issue of how it might be evaluated in terms of teaching quality in a scientific and efficient way has become a focal point as well as a significant challenge in order to make sure that the intended objectives are met [5]. Nevertheless, although most of the institutions have added the assessment of Curriculum Civics teaching quality to their regular teaching assessment procedures, there are still flaws in the high level planning and systematic evaluation model. Some of the existing issues are unclear indicator systems, lack of measurability of standards, insufficient evaluation content, overly narrow methods, and poor coordination between teaching and learning evaluation [6-8]. One of the main reasons is that the existing evaluation practices have not completely demonstrated the importance of guidance, diagnosis, motivation, feedback and improvement in assessing the teaching quality of the ideological and political aspects of courses. They have also not made proper alignment between the various aspects of construction, management, application, education, evaluation and reform. Consequently, the roles of professional-course teachers are not clearly specified, the ideological and political aspect of the courses does not play a major role, the integration of teaching and educating is not done effectively, and the practical application of evaluation results is not sufficient [9-12].To sum up, it is only by continuously refining and optimizing the evaluation system of Curriculum Civics teaching that the implementation effects of teachers instructional practice can be measured more precisely, feedback and adjustment processes traced, the overall quality of Curriculum Civics teaching enhanced and more reliable builders and successors raised to meet the long-term development needs of the country.

This paper builds a research framework of paradigm transfer and ability extension of course Civics teaching evaluation in meta-universe scenario with the support of MASA theory. Firstly, the classroom expressions and actions of course civics are defined and the classroom expressions and actions dataset is self-developed. Secondly, based on MnasNet, the ReLU activation function is introduced, and the homemade dataset is used for pre-training on the improved EfficientNet to make the network have good feature extraction ability, and then the network model obtained from the pre-training is fine-tuned to obtain the final expression recognition model, and the performance test is conducted on the dataset. Then the behavior recognition model based on DenseNet is studied, and the attention mechanism is introduced on its basis to improve it, and the teacher-student behavioral action recognition model is designed after modularizing the attention mechanism, and the experimental results are compared and analyzed. Finally, using probability value calculation and weight fusion, a multi-feature fusion assessment mechanism of students' Civics classroom status is established based on students' classroom expressions and behaviors, and the mapping relationship between specific expressions and behaviors and students' classroom status is completed, and it is applied to actual cases.

2 Migration of MASA Paradigm for Teaching Evaluation of Curriculum Civics in Meta-Universe

The digital transformation of the evaluation of higher vocational courses should comprehensively consider the evaluation situation, evaluation content, evaluation standard, evaluation subject and other important elements, organically combined with the students' professional characteristics, the industry's new technology and new techniques, the company's new trend of development to plan the evaluation of the reform path, pay attention to the integration and integration of the school and the school-based realities of the timely adjustments, towards the direction of diversification and digital intellectual development.

2.1 The implementation program of the evaluation of curricular civics in the meta-universe

2.1.1 Evaluation of Virtual Reality Integration

Applying the VR virtual reality technology and multimodal features of the meta-universe to innovate the teaching form, improve the “real”-based evaluation mode, implement the educational evaluation reform of “promoting the real with the virtual and integrating the real with the virtual”, and provide a feasible path to optimize the evaluation system of the Civic and Political Teaching of the curriculum. To build a “Belt and Road” cross-lingual virtual simulation base, with typical VR virtual reality scenarios and practical exercises, through virtual reality, situation simulation, role-playing, content simulation, confrontation mechanism and other multi-modal scenarios, to evaluate students' core qualities in human-computer interaction, and to implement the Civic-Constitutional assessment.

2.1.2 Implementing knowledge and action

Adhering to the concept of integrating knowledge with action, we have innovated the reform of the evaluation of works, emphasizing "internalizing in the heart and externalizing in action". We enable students to learn by doing and do while learning. We carry out the reform of the evaluation of works in accordance with the idea of "exploration - teaching - learning - practice - refinement - production - competition". According to the professional skills requirements of students, we design tasks for the production of works, enabling students to shift from "talking on paper" to "practical work". Collect outstanding ideological and political education works in courses, build a database of excellent course works, and enable the evaluation of works to feed back into classroom teaching.

2.1.3 Focus on non-quantifiable increases

In VR virtual reality, the blink evaluation system can effectively reflect students' emotional attitudes, dynamically record students' learning growth through intelligent evaluation, further explore value-added evaluation, and give more innovativeness to the ideological immersion education. Applying AI smart evaluation, big data and other digital technologies to optimize accurate evaluation effectively solves the difficult problem of fragmentation and inefficiency in student data management.

2.2 Multimodal Instructional Evaluation of Human-Computer Interaction in the MASA Paradigm

Individual differences and the dynamic factors of the communication environment emphasized by MASA theory are also supplemented by the multimodal expression of emotion. Figure 1 shows the human-machine cooperative path in the human-machine emotional interaction model based on MASA theory.

Specifically, the human-machine cooperative path contains the human path and the machine path. As shown in the solid line, the human subject initiates the emotional communication demand, and based on the basic emotion generation mechanism of the human subject, the human subject transforms the emotional communication demand into physiological or behavioral data signals, and enters the “human-machine” game. If the game succeeds, the human subject will continue to move along the original human task path towards an emotion-oriented or task-oriented emotional response (see Figure 1① route); if the game fails, the emotional communication dominance will be ceded to the machine subject, and the machine subject will dominate along the machine task path to continue to generate emotional responses (see Figure 1② route).

Similarly, as shown in the dotted line, the machine subject initiates the demand for emotional communication, and based on the basic emotion generation mechanism of the machine subject, the machine subject transforms the demand for emotional communication into a multimodal emotional expression signal, and enters the “machine-human” game. If the game succeeds, it continues to move along the original machine path to conscious anthropomorphization or unconscious anthropomorphization of the human subject's emotional cognition (see Figure 1 ③ route); if the game fails, it cedes the dominant right of emotional communication to the human subject, and the human subject dominates along the path of human beings to continue to generate emotional responses (see Figure 1 ④ route). It is important to point out that, since the benefit of human development is always the important red line of human-machine emotional communication, the “human-machine” game should be embedded in the human privacy protection risk warning mechanism, and the machine emotion generation threshold should also be placed in the “machine-machine” game. Therefore, the “human-machine” game should be embedded in the human privacy protection risk warning mechanism, and the machine emotion generation threshold should also be placed in the “machine-human” game, so as to form an adjustable and controllable human-machine symbiotic human-machine emotional relationship.

Therefore, under the theory of MASA paradigm, multimodal emotion expression based on human-machine interaction can well accomplish the evaluation of the teaching of Civics under the meta-universe scenario.

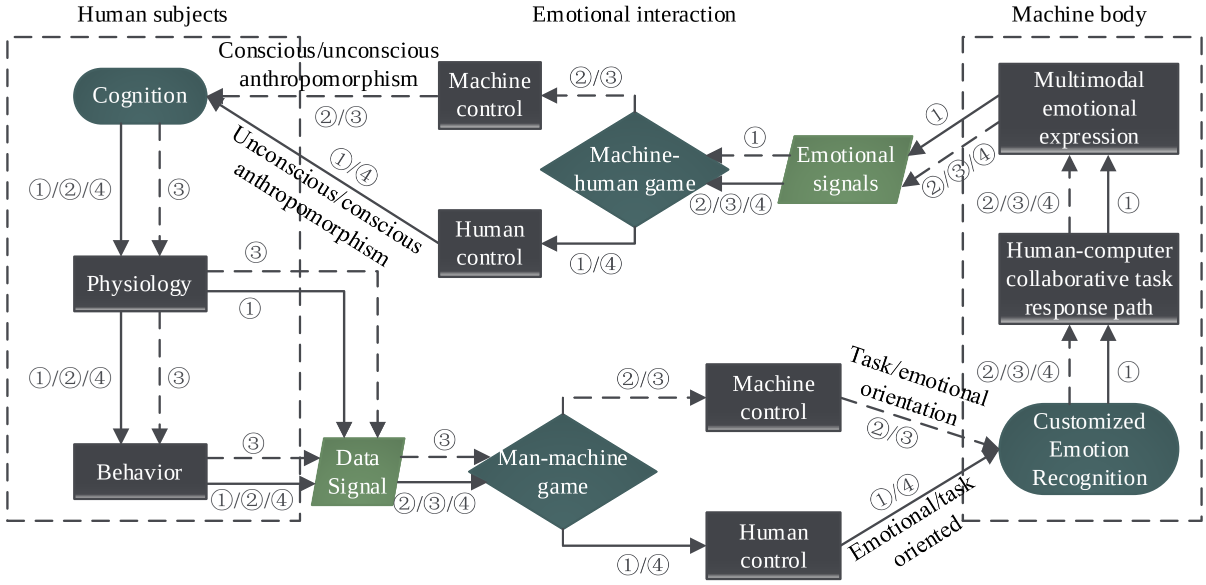


Figure 1: Interactive mode of interactive interaction based on masa theory

3 Methods of Analyzing Data for Classroom Curriculum Civics Evaluation in the MASA Paradigm

3.1 Definition of classroom expressions and gestures

3.1.1 Definition of classroom expressions

On this basis, this paper combines students' state law in class, synthesizes daily expressions and classroom expressions, and considers the ease of recognition and representativeness of expressions, and categorizes students' common facial expressions in class into five kinds, which are happy, concentrated, curious, bored and tired. Teacher expressions were categorized as happy, curious, focused, sad, angry and tired, and the degree of their positivity was classified. The results of classroom expression categorization of teachers and students are shown in Table 1.

Table 1: The teacher and the student classroom expression classification

Student		Teacher	
Expression	State	Expression	State
Revulsion	Negativity	Angry	Negativity
Afraid	Negativity	Revulsion	Negativity
Pleasure	Negativity	Afraid	Negativity
Heartbreak	Negativity	Pleasure	Positive
Surprise	Negativity	Heartbreak	Negativity
		Surprise	Positive

3.1.2 Definition of classroom actions

In this study, students' learning behaviors were specified as: looking up to listen, raising hands, reading and writing; and non-learning behaviors were specified as: looking left and right, sleeping. Meanwhile, after reviewing the relevant literature and summarizing the practice, the teacher's teaching behaviors were specified as: lecturing, asking questions, writing boards, and indicating key points, and the non-teaching behaviors were specified as: constantly looking down and sitting down. The results of classroom action classification are shown in Table 2.

Table 2: Class action classification results

Type	Classification
Student	Lecture, Wirting, Hand, Sleep, Get down, Standing
Teacher	Teach, Question, Board, Emphasis, Sit down, Bow down

3.2 Data Acquisition and Processing

3.2.1 Creation of the classroom expression dataset

In this study, the classroom environment of video resources and research experience, classroom teacher and student emotions are categorized into five types: happy, curious, focused, tired and bored.

In this paper, a total of 60 45-minute classroom teaching videos were selected to build a self-constructed classroom expression dataset. We first use the ffmpeg tool to split the original video data into frames and convert each frame into a separate image file for subsequent image processing and analysis. Next, face detection and cropping is implemented using Dlib technique to localize and detect faces in each image and crop them into single face images. By cutting and cropping the images, multiple face images are separated to facilitate subsequent data processing and analysis. To further improve the quality and usability of the dataset. Finally, we performed batch pixel adjustment and processing on each face image, saved it as a standard face expression image with a size of 48*48, and flipped, random brightness adjustment, and random contrast adjustment on each image in order to reduce the noise and interference, improve the image usability, and increase the diversity and robustness of the dataset. Eventually, the study constructed an enhanced classroom expression dataset containing a total of 5000 faces, which is suitable for expression recognition analysis research in the field of education, and the dataset is noted as CE-K12.

3.2.2 Production of the classroom action dataset

The true footage of classroom teaching surveillance was used as the data source and the ffmpeg tool was used to extract video segments, which produced 400 clips altogether. In order to ensure that the datasets were of good quality and were usable in practice, the length of the clips was restricted to 1-3 seconds, and the classroom teaching behaviors found in every segment were tagged manually based on the classifications given in Content 3.1. From this fact, the paper develops one-person video sets containing various classroom activities such as attentive listening with the head up, raising hands, writing, reading, looking around, sleeping on the desk, asking questions, board writing, and explaining the most important issues. The dataset is called CA-K12 and it should provide an excellent data support to the educational research as well as applications related to teaching.

3.3 Classroom expression recognition

3.3.1 Classroom expression recognition model design

(1) Network structure design

A simple convolutional neural network can be defined as in equation (1):

$$N = \bigodot_{i=1\cdots s} e_{i=1\cdots s} F_i^{L_i}\left(X_{\langle H_i, W_i, C_i \rangle}\right) \tag{1}$$

where F_i is the convolution operation of the i layer, L_i is F_i repeated L_i times in the i th stage, and $\left(H_i, W_i, C_i\right)$ is the dimensionality of inputs to the i th layer.

The EfficientNet network structure in this paper is based on MnasNet, which is searched using the AutoML method to derive coefficients that allow all convolutional layers to expand in the same proportion. The related mathematical expression is shown in equation (2):

$$\begin{aligned} &\max_{d,w,r} Accuracy(N(d,w,r)) \\ &\text{s.t. } N(d,w,r) = \bigodot_{i=1\cdots s} F_i^{d \cdot L_i\left(X\langle r \cdot H_i, r \cdot W_{i,w} \cdot c_i \rangle\right)} \end{aligned} \tag{2}$$

where d, w and r are the network depth, width and resolution, respectively. F_i, L_i, H_i, W_i, C_i are the predefined parameters of the baseline network, respectively. And the searched correlation coefficients are adjusted as in equation (3):

$$d = \alpha^{\Phi}, w = \beta^{\Phi}, r = \gamma^{\Phi}$$
$$\text{s.t. } \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \tag{3}$$
$$\alpha \geqslant 1, \beta \geqslant 1, \gamma \geqslant 1$$

where α, β, γ are constants searched out using the grid, representing the depth, width, and resolution of the adjustment grid; and Φ are customized correlation coefficients controlling the model augmentation.

EfficientNet uses the MBConv layer in MobileNet V2 as the backbone network of the model, while using the Squeeze and Ex-citation methods in SENet for optimization.

The defining equation of ReLU is shown in (4):

$$f(x) = \max(0, x) = \begin{cases} x, x > 0 \\ 0, \text{Other} \end{cases} \tag{4}$$

The modified network architecture is composed of an input layer, a 3×3 convolutional layer, an MBConv block containing seven 3×3 convolution kernels, another MBConv block with nine 5×5 convolution kernels, a 1×1 convolutional layer, a pooling layer, a ReLU activation layer, a dropout layer, and a fully connected layer.

The EfficientNet-B0 architecture is implemented in this paper as the base model but maintaining its initial advantages. The ReLU activation module is added at the same time in order to make the extracted features more expressive. Next, a dropout mechanism is added to decrease the possibility of overfitting when training.

(2) Loss function

The cross-entropy loss has been used in this research in order to calculate the output error of the network. Cross-entropy is the discrepancy between two probability distributions of the same random variable and it can be considered as the difference between true probability distribution and the predicted one in machine learning, as demonstrated in Equation (5):

$$H(p,q) = -\sum_{i=1}^{n} p(x_i)\log\left(q(x_i)\right) \tag{5}$$

where $p(x_i)$ denotes the ground-truth probability distribution of the sample,whereas $q(x_i)$ represents the probability distribution predicted for that sample.

(3) Optimizer

This paper has chosen the Adam optimizer because it can be used to adjust the learning rate during the training phase adaptively. Adam is an optimization algorithm of first order based on gradient descent that is applied to stochastic objectives. It is very applicable to high dimensional problems and the parameters can be optimized automatically through the learning procedure. The updating rule of the parameters is defined in Equation (6):

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{v_t}} m_t \tag{6}$$

where θ is the parameter to be updated, η is the learning rate, m_t is the mean of the first moment of the gradient, and $\sqrt{v_t}$ is the variance of the second moment.

3.3.2 Classroom expression recognition results and analysis

(1) Experimental environment and evaluation index

This research uses Windows 10 as its operating platform. Hardware configuration consists of Intel CPU i5-10400F and 8 GB NVIDIA GeForce GTX 1070Ti graphics card. Python 3.7 and PyTorch 1.7.1 are used as the programming language and deep learning framework respectively. Transfer learning and data augmentation are added to training to even further improve model performance. Training is performed on the network using 2000 epochs and 16 as the batch size. Weighted cross-entropy is considered as the loss function and Adam is chosen as the optimizer. The starting point of the learning rate is set at 0.01 and is subsequently updated dynamically by exponential decay.

The experiments are based on recognition accuracy, parameter quantity, confusion matrix, and F1-score as the measures of evaluation. Accuracy means the proportion of correctly classified samples to total sample set. Parameter quantity is the amount of trainable parameters in the model, and higher values usually indicate a more complicated network. The confusion matrix is a tabular form of true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN). F1-score is a harmonic mean of precision and recall, and the value can vary between 0 and 1.

(2) Experimental results and analysis

、In order to clarify the training process better, CE-K12 dataset, which has been created in the previous section, is taken as the experimental example. Figure 2 shows the loss curve and precision curve of the optimized network trained on CE-K12. It was found out that following the 200 epochs of training the loss on the training set and the validation set were decreased under the improved EfficientNet model whereas the recognition accuracy increased to around 85%.

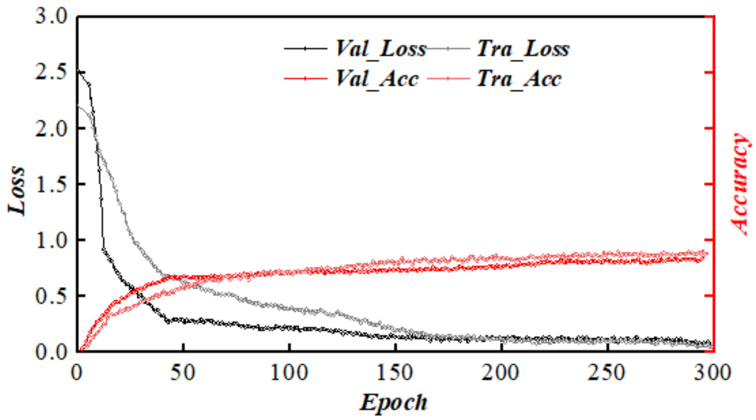


Figure 2: Training precision and loss curve

In the course of the experiments with the CE-K12 dataset, the last assessment of the enhanced model was conducted on the test set in order to prove its benefits more convincing, and the confusion matrix obtained is illustrated in Fig. 3. The matrix shows that the suggested model achieves recognition rates of 93 and 88 percent in relation to the two categories of happy and surprised respectively, whereas the three classes of anger, fear and sadness are relatively lower in terms of performance. It is also possible because happy and surprised expressions have more salient facial texture features and it is easier to distinguish them using the model. Surprise and fear are particularly prone to misclassification since both expressions have widened eyes and open mouths, but the mouth opening in fear is not as noticeable as those in surprise. Moreover, fear and sadness have similar visual indicators, like frowning and forehead wrinkling, making fear the hardest category to recognize. All three (anger, fear, and sadness) are negative emotions and they have a high degree of similarity among themselves, making recognition even more challenging.

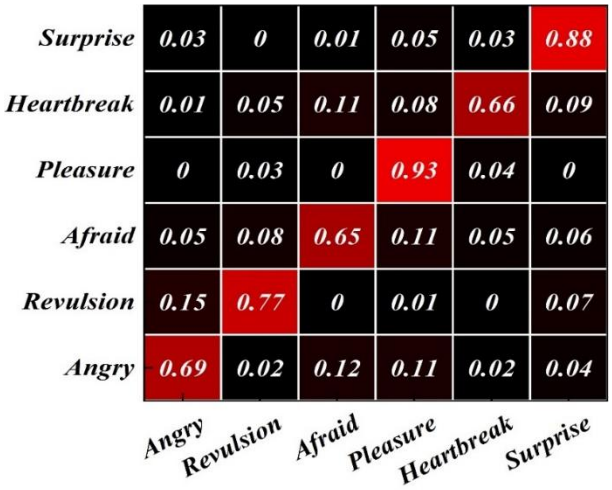


Figure 3: Identify the resulting confusion matrix

In this paper, the improved EfficientNet is compared and analyzed with other algorithms as shown in Table 3, and it is found that the recognition accuracy of the network model proposed in this paper reaches 85.51%, which achieves better results in terms of accuracy and F1-score value compared with other algorithms, and there is only a small increase in the number of parameters. It is verified that the method can recognize the classroom expressions of teachers and students well.

Table 3: Data set experiment comparison

Method	Accuracy/%	Parameter	F1-score
VGG16	69.85	135225516	0.71
ResNet50	71.65	25523222	0.77
Xception	66.89	16816555	0.69
Inceptionv3	70.66	22628179	0.73
MobileNetv2	72.62	3556528	0.75
EfficientNet	72.88	1685682	0.73
Our	85.51	1715684	0.81

3.4 Classroom Action Recognition

3.4.1 DenseNet network

The DenseNet model cleverly connects the features on the channel to achieve feature reuse, which greatly reduces the number of parameters to be computed and the computational cost.The network structure of the DenseNet model is shown in Fig. 4, which is mainly composed of convolutional layer, sense block, transition layer, pooling layer, and fully-connected layer, and can be formed into DenseNet of different layers by cascading with different number of sense blocks and transition layers. By using different numbers of sense blocks and transition layer to cascade, DenseNet with different layers can be composed.

Among them, the sense block is the core module of DenseNet, for an L-layer senseblock, it contains a total of $\frac{L(L+1)}{2}$ connections, which is a kind of dense connection compared with the residual connection of ResNet model and can realize the reuse of features.The output of the 1st layer of Denseblock can be expressed as:

$$x_l = H\left(\left[x_0, x_1, \cdots, x_{l-1}\right]\right) \tag{7}$$

where $H_l(\cdot)$ denotes a nonlinear transformation composed of a BN (Batch Normalization) layer, the ReLU activation function, and a 3x3 convolution. Even though DenseBlock design allows features to propagate all the way to later layers, too many channel-wise concatenations increase the number of channels in the feature map exponentially, drastically reducing the model computation rate. Hence, the $H_l(\cdot)$ operation is defined to reduce the dimensionality of features and hence reduce the computational burden of the network.

As illustrated in Figure 4, connecting the DenseBlock to the Transition layer is intended to shrink the size of the concatenated feature maps and further enhance model performance. The Transition layer contains a convolution kernel of size 1 together with an average pooling layer, and introduces the parameter θ, which specifies that the number of feature maps produced by the DenseBlock is reduced to θ times the original value; this parameter is commonly set to 0.5. In this study, DenseNet-169 is adopted as the baseline model.

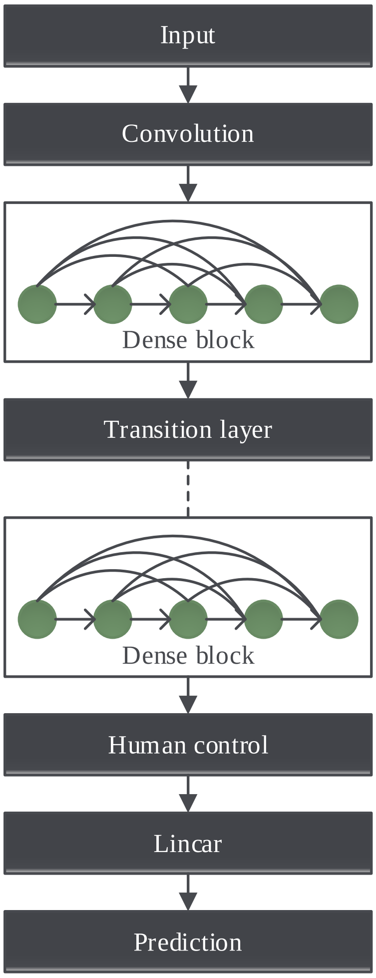


Figure 4: DenseNet model network structure

3.4.2 Attention mechanisms

A study was conducted, and a new modular unit, SE (Squzee & Excitation) block, was constructed in the paper to spontaneously calibrate the feature response in the channel dimension by establishing the relationship between different channels. Its structure is shown in Fig. 5.

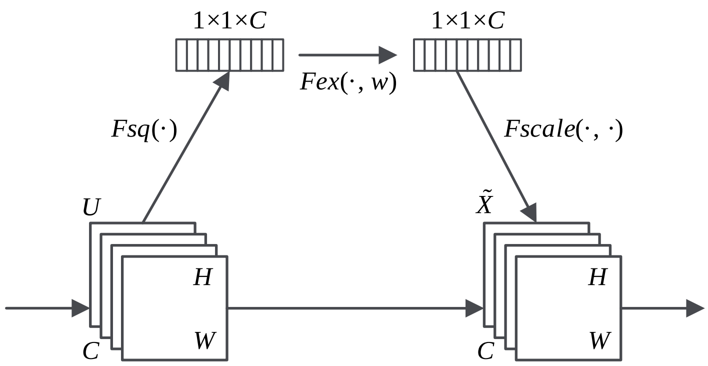


Figure 5: SE module structure

In Fig. 5, the SE block is given an input U with channel number C and then its features are recalibrated by three operations. First, the features are compressed in the spatial dimension to a real number that has a global receptive field in a way that represents the global distribution of feature responses on a channel. This operation can be represented by equation (8), where z denotes the statistic of U compressed to the spatial dimension, $z \in R^C, U \in R^{H\times W\times C}$. i.e:

$$z_c = F_{sq}\left(u_c\right) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j) \tag{8}$$

Then, the resulting C real numbers from compression are subjected to an excitation operation, which is similar to the gate mechanism in CNN networks, where the correlation between each feature channel is learned by training the w parameters to generate the corresponding weights for each feature channel. This operation can be expressed by equation (9), s is the excitation parameter after sigmoid activation. i.e:

$$s = F_{ex}(z,W) = \sigma(g(z,W)) = \sigma\left(W_2\delta\left(W_{1z}\right)\right) \tag{9}$$

Finally, the weights output after excitation are regarded as the importance of each channel feature, and are weighted to the original features by multiplication operation to realize the feature recalibration of the channel dimension. This operation can be expressed by Eq. (10), and $\tilde{x}_c$ is the output feature after recalibration. Then:

$$\tilde{x}_c = F_{scale}\left(u_c, s_c\right) = s_c u_c \tag{10}$$

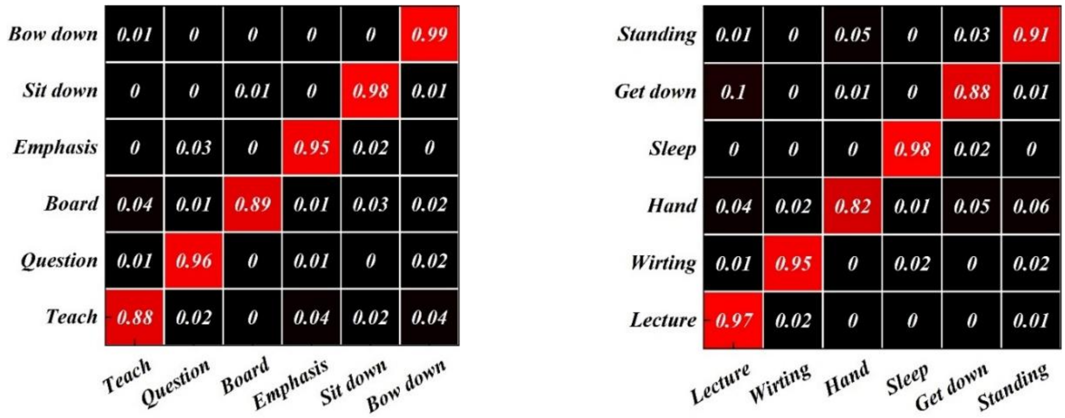
3.4.3 Classroom Action Recognition Results and Analysis

(1) Experimental setup

A deep learning system was used in the training phase of the given experiment. The batch size was selected to be 64 and the number of epochs to be 100 with an initial learning rate of 0.1. When the validation loss did not decrease after 10 straight steps, the learning rate was reduced by a factor of 0.1 to a minimum of 0.0001. On every experimental trial, the model with the highest recognition accuracy was kept as the best-performing model on the given training process.

(2) Experimental results and analysis

Confusion Matrix is an example of a specialized matrix form that can be used to provide a visual representation of the performance of a deep learning algorithm. Each column in the matrix refers to the predicted class, and each row refers to the true class in the matrix. The confusion matrices of recognitions of different teacher and student actions are given in Figure 6. The values on the diagonal of the displayed matrix represent the prediction accuracy of the model. It can be seen in Figure 6 that both the teacher-behavior model and the student-behavior model had quite similar recognition rates. In the teacher-side results (Fig. 6a), the lowest recognition accuracy was achieved by lecturing and it was because this behavior comprises temporal data and thus more difficult to recognize than the others, whereas bowing the head recorded the highest accuracy at 0.99. In the student-side (Fig. 6b), the lowest recognition accuracy, 0.82, could be found in the writing category, because the school behaviors like resting the cheek on the hand or scratching the head might be mistaken with raising their hands very easily, which in turn decreases recognition performance.



(a)Teacher (b) Student

Figure 6: Confusion matrix of various actions

Detection rate, recall and mean precision are commonly used as measures of evaluation in object detection tasks. The resulting values of the evaluation indicators of teacher and student classroom behavior recognition are shown in Table 4. The data in the table shows that the accuracy of classification and recall are higher than 88 percent, which is a good performance of recognition. To put it another way, the suggested behavior recognition framework works well when recognizing classroom behavior of both teachers and students.

Table 4: The student behavior identifies statistical indicators

Teacher			Student		
Full action	*Accuracy*	*Recall*	Full action	*Accuracy*	*Recall*
Teach	0.884	0.968	*Lecture*	0.927	0.953
Question	0.882	0.956	*Wirting*	0.969	0.883
Board	0.969	0.924	*Hand*	0.988	0.993
Emphasis	0.958	0.971	*Sleep*	0.957	0.913
Sit down	0.894	0.897	*Get down*	0.899	0.947
Bow down	0.908	0.915	*Standing*	0.917	0.942

Teacher Behavior Recognition A commonly used in the field of behavior recognition, dual-stream based fusion neural network is selected for comparison with the improved DenseNet network used in this paper for teacher behavior recognition. In the dual-stream based fusion network, the video input is divided into 2 parts, spatial (RGB image) and temporal (optical flow image), so as to carry out the extraction and processing of the information between consecutive frames in the video. The recognition network of this paper and the network used for comparison are trained and tested separately on the same dataset, and the comparison of the accuracy and loss rate convergence of behavior recognition with the same number of trainings is shown in Fig. 7. In addition, this paper also compares the accuracy of the improved DenseNet network with the better performing, dual-stream based fusion network for video behavior recognition. After running on this labeled dataset, the recognition accuracy of the dual-stream based fusion network is 88.52% and that of the improved DenseNet network is 94.21%. It is easy to see that in the practical application scenarios of this paper, the training convergence efficiency and recognition accuracy of the improved DenseNet network are both better than the dual-stream based fusion network to a certain extent, and it can provide better recognition efficiency for the behavior recognition work of teaching videos.

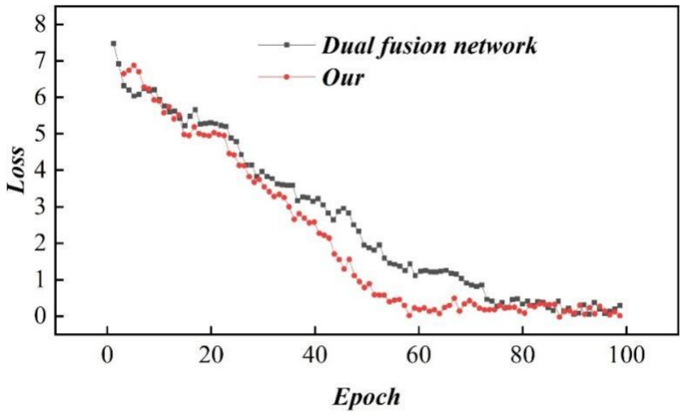


Figure 7: Contrast of loss rate curve

4 Methods of comprehensive assessment of the quality of teaching Civics in courses

4.1 Algorithm for evaluating the teaching of course civics by integrating expressions and behaviors

4.1.1 Assessment algorithm framework

Before carrying out the multimodal intelligent classroom teaching assessment, the students in the classroom video are firstly identified by their expressions and behaviors. After obtaining the results of expression recognition and behavior recognition, the intelligent classroom teaching assessment algorithm is used to obtain the results of students' classroom status assessment.

In the process of student behavior recognition, the test classroom video frame is input into the target detection model to get the student target frame in the video frame, and then the student target image is preprocessed, and then the preprocessed student target image is input into the behavior recognition model to extract the student behavioral features, and then the logistic regression algorithm is used to compute the probability of the student's behavioral states, and the maximum probability is taken to be the student's behavioral state at this moment. The maximum probability value is taken as the behavioral state of the student at this moment. The logistic regression algorithm formula is expressed as (11):

$$y = \frac{1}{1+e^{z}} \tag{11}$$

where y denotes the output probability vector and z denotes the extracted student behavioral feature vector.

4.1.2 Teaching Evaluation Algorithm Implementation

One of the most significant stages of the intelligent classroom teaching evaluation algorithm based on the combination of student facial expressions with behavioral data is the determination of the weight coefficients, since the selected weightings have an immediate impact on how expression recognition and behavior recognition influence the overall end-of-the-evaluation result. The cause behind this was that three groups of weighting schemes were created and the total evaluation outcomes as well as the related labels in various conditions were compared. Finally a combination of weights with relatively good performance was chosen, i.e., the weight of expression recognition as 0.65 and the weight of behaviour recognition as 0.35.

Since behavior recognition and facial expression recognition were performed simultaneously, no expression recognition was done during the period when the behaviour of the student was defined as head-down or head-turn to limit the interference with the final total assessment. The statistical analysis of frequencies and probabilities with the three categories of positive, negative and neutral state is based on the recognition outputs of students expressions and behaviors. Following this, the final comprehensive evaluation value is calculated by incorporating the weights assigned to expression and behavior, and the students listening status is analyzed and concluded based on the outcome.

To begin with, by $countpos\_expression$, $count_{neg\_expression}$, $count_{neu\_expression}$, let the number of students identified in classroom videos as having a positive, negative and neutral expression respectively and Sum be the overall number of students in the classroom video.

$$Sum = count_{pos\_expression} + count_{neg\_expression} + count_{neu\_expression} \tag{12}$$

Then, the probability values P of positive expressions and positive behaviors are calculated separately:

$$P_{pos\_expression/behavior} = \frac{count_{pos\_expression/behavior}}{Sum} \tag{13}$$

Finally, weights are assigned to the expression recognition results and behavioral recognition results, using μ_1 to denote the weight value assigned to the expression recognition results and μ_2 to denote the weight value assigned to the behavioral recognition results, and calculating the final composite evaluation value Com_Score:

$$Com\_Score = P_{pos\_emption} \times \mu_1 + P_{pos\_behavior} \times \mu_2 \tag{14}$$

For the comprehensive assessment value to analyze and discuss the students' overall classroom status assessment level, mining the students' overall classroom status assessment level mapping relationship, the classroom comprehensive assessment level determination method is shown in Figure 8.

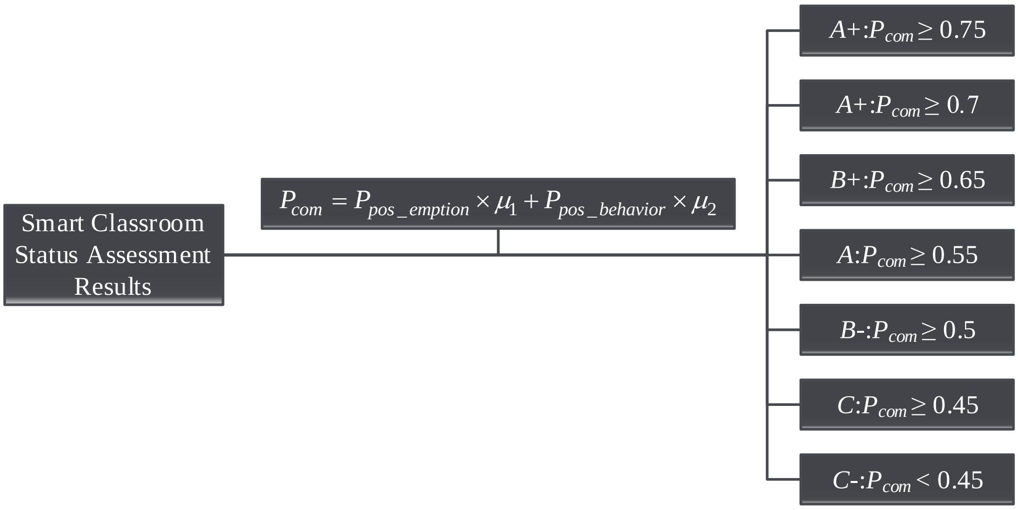


Figure 8: Class evaluation method

4.2 Acquisition and Processing of Multimodal Information in the Civics Classroom

Student behavior recognition is performed based on the classroom video images acquired with acquisition frequency τ , where the ratio of the number of students attending the classroom for the group of students in positive and neutral states can be represented by $\eta(t)$, where N is the total number of students in the classroom, and $A_i(t)$ ~ $F_i(t)$ are the student actions. Then:

$$\eta(t) = \sum_{i=1}^{N} (A_i(t) + B_i(t) + \cdots + F_i(t)) / N \tag{15}$$

Positive emotions, neutral emotions and negative emotions. Positive emotions include "surprise" and "happiness", while negative emotions include "disgust", "anger", "fear" and "sadness". In this case, positive and neutral emotions can reflect the proportion of the number of attentive listeners in the student population is used to represent. $G_i(t), H_i(t), M_i(t)$ represent positive, negative and neutral emotions, then:

$$\mu(t) = \sum_{i=1}^{N} (G_i(t) + H_i(t) + M_i(t)) / N \tag{16}$$

4.2.1 Classroom Behavior Detection

In this paper, we chose a video with a video length of 45 minutes of the students' class video, took a frame every 5 seconds to obtain a picture, and selected four students in the video to track and analyze respectively as shown in Figure 9, from which we can observe that the behavioral state of student one has been in a positive state, indicating that the student's attention has been kept at a high level during the learning process. Student two's $\eta(t)$ value has been in the neighborhood of 0.35, indicating that the student was not motivated throughout the class, showing a state of disinterest in the course. Student 3's line shows a clear downward trend in the second half of the lesson, indicating that the student may have lost interest in listening to the lesson because the problems that arose at this point in time were not solved in a timely manner. Student four's fold line shows an overall upward trend, and the value at the beginning of the lesson may be due to the fact that the student has not yet entered the learning state and is in a state of sleepiness at the beginning of the lesson.

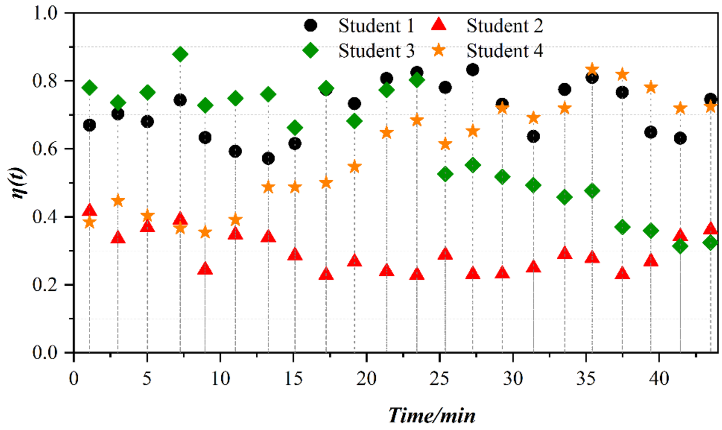


Figure 9: Four students' $\eta(t)$ lines

4.2.2 Learning Mood Detection

The four typical students to be tested in this paper, the detection of student emotional state, the test results are shown in Figure 10, the four students, compared with students 3, 4, student one's learning mood fluctuations are not large, indicating that the students for the classroom focus on the degree of influence by external factors, but the $\mu(t)$ value has always remained in the upper and lower fluctuations in the 0.68, as a whole, the student's motivation to learn is high. Student 2 is not focused enough in the course of learning, and his learning state is not good, so he needs to stimulate his enthusiasm for learning. Student three's value reached over 0.72 at the beginning of the lesson, and started to drop when the lesson reached 23 minutes, indicating that the student's attention was easily diverted and he gradually lost interest in the learning process. Student four's mood was not high enough at the beginning, but as the course progressed, the value of $\mu(t)$ became higher and higher, and the student's interest became stronger and stronger, which can be achieved by appropriately deepening the content of the course's Civics and providing more challenges and opportunities to promote the student's in-depth thinking and learning growth.

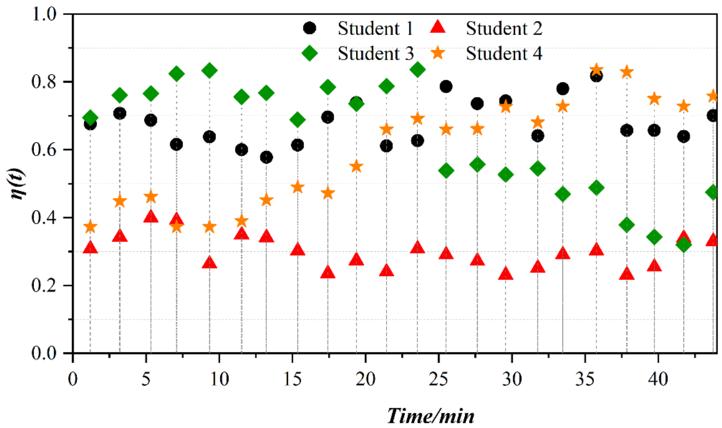


Figure 10: The emotional state contrast of four students

4.3 Practical application cases and analysis

In this section, the learning evaluation method based on multimodal information fusion proposed in 4.1 is applied in a real classroom scenario for research. The camera installed in the classroom is utilized to capture students in real time, and the classroom video with a duration of 45 minutes is used as the experimental data, and one frame is taken every 5 seconds to obtain the classroom pictures for processing. Recorded and assessed in the behavioral dimension and expression dimension respectively, thus obtaining students' behavioral state as well as expression information, and deriving the quantitative values of learning cognitive attention and learning emotion. Finally, the two dimensions of information are fused to calculate the students' final learning assessment result score.

4.3.1 Data acquisition

In this paper, 10 students were tracked for analysis, and Table 5 shows the behavioral recognition results of the 10 students, and Table 6 shows the expression recognition results of the students.

Table 5: Student behavior identification results

Student NO.	Sit up	Read a book	Raise one's hand or hands	Stand	Drop the table	Play mobile phone
Student 1	70	70	210	160	50	40
Student 2	140	105	158	130	36	20
Student 3	96	128	153	150	42	31
Student 4	102	86	147	175	76	29
Student 5	137	162	70	71	23	86
Student 6	172	217	42	96	30	44
Student 7	172	200	67	96	30	35
Student 8	189	166	72	65	45	57
Student 9	100	107	30	57	142	164
Student 10	98	82	20	10	190	200

Table 6: Student expression recognition results

Student NO.	Happy	Wonder	Neutral	Sad	Fear	Detest
Student 1	152	111	136	32	32	85
Student 2	131	174	115	40	42	58
Student 3	165	168	109	23	53	50
Student 4	153	121	111	30	19	13
Student 5	79	97	256	43	36	53
Student 6	99	107	272	44	32	33
Student 7	80	97	260	78	19	19
Student 8	76	87	243	90	27	22
Student 9	51	47	108	108	89	111
Student 10	60	62	133	95	87	91

In this paper, the positive, neutral, and negative behaviors of these 10 students who were continuously tracked were counted, in which the results of the behavioral state identification statistics are shown in Figure 11, and the results about the expression state statistics are shown in Figure 12. From the figure, we can observe that students' behaviors and expressions present an asynchronous pattern in the classroom. Student 1 has the best behavior in the classroom, but student 3 has the best emotional performance in the classroom, and student 10 performs poorly in both dimensions. Taken together, these 10 students have the same teacher and the same learning environment, but due to the variability among individual students, they have their own characteristics on different dimensions in the learning evaluation model.

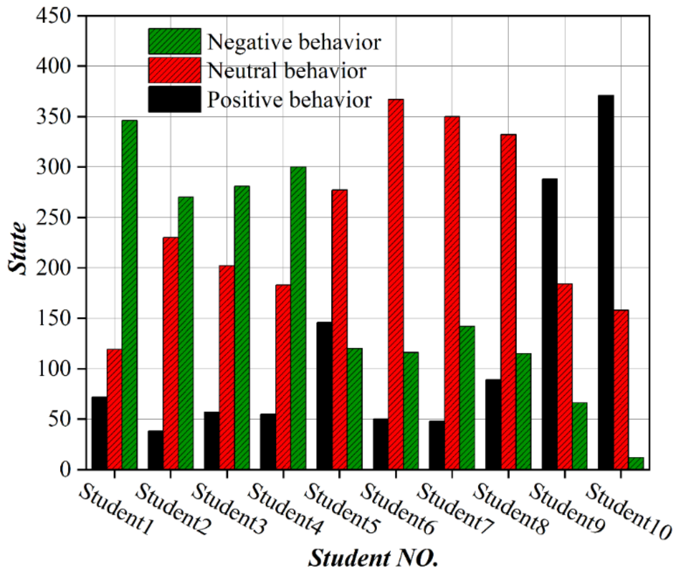


Figure 11: Behavior status recognition results

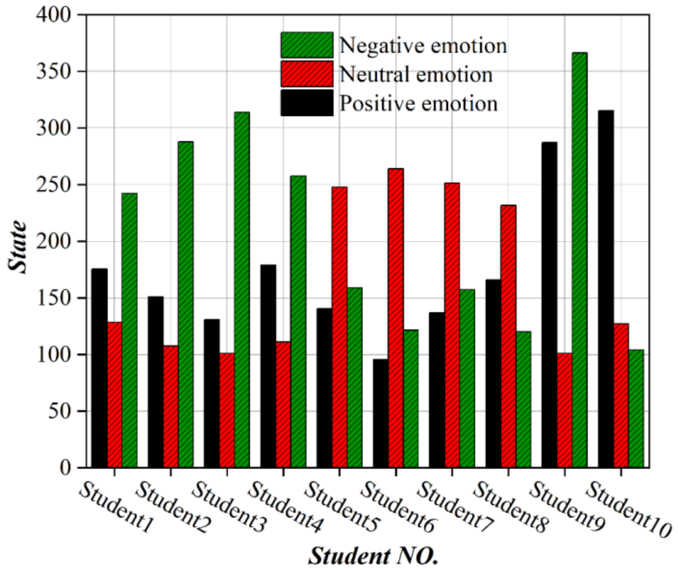


Figure 12: Emotional status recognition results

4.3.2 Calculation of learning assessment results

Wen processed the information detected from different dimensions and the data collected from the teacher-student interaction platform, combined them into a dataset, and used the method in Section 4.1 to perform hierarchical fusion, which fused the information from different data levels to obtain a comprehensive learning evaluation result as shown in Table 7. It can be seen that there are differences in the final evaluation scores of these 10 students, with Student 1, Student 2 and Student 3 having the best learning status with scores of more than 0.400, Student

9 and Student 10 having a clear need for the teacher to implement timely interventions, and the remaining students being at an intermediate level. The division of student status levels fully proves that the model in this paper can effectively screen out student learning status and provide teachers with valuable feedback results. These feedback results enable teachers to grasp the real state of students' learning in a timely manner, including information on students' interest in learning, learning progress, and learning behaviors.

Table 7: The course is the result of the evaluation of the course

	Cognitive attention	Learning emotion	Fusion result
Student 1	0.489	0.405	0.434
Student 2	0.497	0.359	0.407
Student 3	0.369	0.419	0.401
Student 4	0.291	0.377	0.347
Student 5	0.202	0.386	0.322
Student 6	0.299	0.319	0.312
Student 7	0.205	0.368	0.311
Student 8	0.367	0.338	0.348
Student 9	0.256	0.2082	0.225
Student 10	0.211	0.306	0.273

5 Conclusion

Traditional evaluation methods are not only laborious and difficult but also have the emergence of evaluation feedback untimely problems, with the development of information technology, the education industry is also widely used, gradually moving towards intelligence. In this paper, under the meta-universe perspective, MASA theory is introduced to innovate the evaluation paradigm of course Civics teaching and improve its evaluation ability. The main work is as follows:

First, in terms of the evaluation means of classroom Civics and Politics course teaching, the classroom expression and action recognition algorithms are optimized to realize the structuring of classroom data. In the homemade classroom dataset, the accuracy of the improved EfficientNet network in recognizing classroom teachers' and students' expressions reaches about 85%, and the accuracy of the improved DenseNet network in recognizing classroom teachers' and students' actions reaches 94.21%, which verifies the possibilities of the algorithms in practical applications.

Secondly, the evaluation method of intelligent classroom teaching integrating students' expressions and behaviors in teaching classroom videos is proposed to complete the mapping relationship between specific expressions and behaviors and students' classroom status.

Finally, in the real classroom situation, the comprehensive assessment of the teaching of the ideology and politics course is completed, and teachers can use these feedback results to adjust the teaching strategy, formulate a more personalized teaching plan, and improve the teaching effect. For example, for students with a low level of commitment, teachers can use more vivid and interesting teaching methods to stimulate students' interest in learning; for students with a slower learning progress, teachers can use step-by-step teaching methods to help students gradually master the knowledge points; for students with inappropriate learning behaviors, teachers can use incentives to guide students to form good learning habits. Therefore, the model proposed in this paper is of great significance for improving the teaching quality of online education.

Funding

This article is part of the research results of the 2026 Henan Province Academic Degree Postgraduate Core Course Project "Intermediate Microeconomics" (Project Number: YJS2026XSKC49).

This article is a phased research outcome of the undergraduate quality engineering project of Henan University of Economics and Law, "Research on the Reform and Practice of the Teaching Paradigm of Public Economics" (Project No. 300470).

This article is a phased research outcome of the teaching reform project of Henan University of Economics and Law, titled "Research on the Path and Method of Integrating Innovation and Entrepreneurship Education into Professional Education" (Project No. 330297).

This article is a partial research result of the 2024 Joint Research and Practice Project on Education and Teaching Reform of Henan University of Economics and Law, titled "Research on the Path and Method of Building a Comprehensive Evaluation System for Students under the Background of Academy System and Community

This article is a partial research result of the 2022 Henan Province blended online and offline first-class undergraduate course reform "Public Economics" (certificate number [2022] 38012).

This article presents some of the research findings from the 2022 undergraduate teaching project special research project on the paradigm reform and practical research of the "Public Economics" course.

About the Author

Binbin Song was born in Nanyang, Henan P.R, China, in 1980. He obtained a Doctor's degree from Beijing Normal University. He is currently working at the School of Auditing, Henan University of Economics and Law. His main research direction is innovative governance, ideological and political education.

Yuhan Huang was born in Anyang, Henan P.R, China, in 2000. She obtained a bachelor's degree from Henan University of Economics and Law. She is currently studying at the School of Accounting, Henan University of Economics and Law. Her main research direction is the theory and practice of financial accounting.

Zijin Wu was born in Pingdingshan, Henan P.R, China, in 2001. She obtained a bachelor's degree from Henan University of Economics and Law. She is currently studying at the School of Accounting, Henan University of Economics and Law. Her main research direction is the digital transformation of accounting.

Yiwen Yuan was born in Xuchang, Henan P.R, China, in 2002. She obtained a bachelor's degree from Henan Institute of Finance and Finance. She is currently studying at the School of Accounting, Henan University of Economics and Law. Her main research direction is equity incentives and corporate governance.

Jiayi Yang was born in Xuchang, Henan, P.R, China, in 2003. She obtained a bachelor's degree from Henan University of Economics and Law. She is currently studying at the School of Accounting, Henan University of Economics and Law. Her main research direction is the digital transformation and financial management.

References

[1] Lin, Y. (2015). Inter-subjective Ideological and Political Education in Recent Years-A Review. J. Pol. & L., 8, 223.

[2] Mei, Z. (2024, July). Integration of curriculum ideology and politics in higher education: a case study of English linguistics in China. In Frontiers in Education (Vol. 9, p. 1389469). Frontiers Media SA.

[3] Gong, Y., Lu, X., & Cui, D. (2021). Research on the construction strategies of “curriculum ideology and politics” for professional courses. International J. Frontiers in Sociology, 3(8), 118-122.

[4] Chen, B., & Pan, X. (2022). Practice and exploration of integrating curriculum ideology and politics into digital logic circuit course. Open Journal of Social Sciences, 10(4), 23-30.

[5] Chunxia, X. (2021, July). Curriculum ideological and political evaluation system for universities based on SWOT analysis. In 2021 International Conference on Education, Information Management and Service Science (EIMSS) (pp. 66-70). IEEE.

[6] Cheng, P. (2022). Evaluation method of ideological and political classroom teaching quality based on analytic hierarchy process. Scientific Programming, 2022(1), 6554084.

[7] Qiu, J. (2023). Research on the Establishment of Curriculum Ideology and Politics and College English Evaluation Model based on" Internet+". Advances in Education, Humanities and Social Science Research, 4(1), 8-8.

[8] Lv, X. (2021, January). A quality evaluation scheme for curriculum in ideological and political education based on data mining. In 2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) (pp. 649-652). IEEE.

[9] Sun, W., & Wang, H. (2023). Discussion on the Curriculum Organization and Teaching Method for Curriculum Ideological and Political Education. Frontiers in Educational Research, 6(23).

[10] Zhang, Y. (2025). Construction of Teaching Evaluation System for Curriculum Ideology and Politics in Colleges and Universities from the Perspective of" Three-All Education". Lecture Notes in Education, Arts, Management and Social Science, 3(6), 87-93.

[11] Liu, W., & He, C. (2022). Curriculum-Based Ideological and Political Education: Research Focuses and Evolution. International Education Studies, 15(5), 28-35.

[12] Wu, Z. (2018). An Improved Performance Evaluation Index System and Fuzzy Evaluation Model of College Students' Ideological and Political Education. Educational Sciences: Theory & Practice, 18(5).