International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-14

https://doi.org/10.70917/ijcisim-2026-0119
Article

Construction of Multi-Channel Teaching Effect Evaluation System Based on Deep Learning and Enterprise Cooperation Operation Practice

Yang Wang [1] and Jing Ma [2,*]

[1] School of Continuing Education, Xinjiang Normal University, Urumqi, Xinjiang, 830000, China
[2] Urumqi Education Research Centre, Urumqi, Xinjiang, 830000, China; evhunter123@126.com

Abstract: The continuous development of deep learning technology has driven the intelligentization of classroom teaching. This paper proposes a real-time teaching evaluation system by combining CVAE-GAN image enhancement with the lightweight YOLOv5 object detection framework. To address the issue of blurred student expressions in classroom scenarios, a conditional variational adversarial generative network is employed for facial reconstruction. The SE attention mechanism and GSConv module are integrated to optimize the YOLOv5s network, enhancing detection performance while maintaining the number of parameters. Experiments show that the improved YOLOv5 model achieves an mAP of 79.78% and an F1 score of 0.82. The accuracy rates for the five facial expressions are 97.15%, 92.63%, 92.74%, 90.82%, and 91.44%, respectively. The system can precisely identify changes in students' facial expressions and classroom attention levels in one-minute intervals, providing teachers with references for instructional adjustments.

Keywords: CVAE-GAN; lightweight YOLOv5; SE attention mechanism; GSConv module; teaching evaluation

1. Introduction

Over the past few decades, most higher vocational and technical colleges have developed significantly, cultivating a large number of applied talents for enterprises and maintaining sustained, positive cooperative relationships with them. These partnerships have evolved from initial superficial collaborations to comprehensive, in-depth collaborations today. Schools and enterprises have become mutually interdependent. Enterprises leverage the schools' strengths in talent cultivation to continuously supply various types of talent, while schools utilize enterprises' financial and practical advantages to implement programs such as “order-based classes” and “apprenticeship systems,” thereby enhancing the quality of talent cultivation and promoting the growth and development of the schools [1-3]. Both parties benefit mutually and are inseparable. However, certain issues have emerged during the development process. Due to the lack of policy support, enterprises often prioritize short-term practical benefits in cooperation and lack the willingness to establish long-term school-enterprise cooperation mechanisms with schools [4-5]. Additionally, the limited research capabilities of higher vocational and technical colleges have, to some extent, constrained the development space for school-enterprise cooperation [6]. Therefore, deepening school-enterprise cooperation in teaching while conducting teaching quality assessments holds significant importance for cultivating core professional talent beneficial to industry development [7-8].

In current school-enterprise cooperation models, schools lack real-time monitoring data on students at off-campus practical training bases and lack corresponding quality analysis. Therefore, implementing teaching quality supervision can reduce information asymmetry between schools and enterprises, helping schools regularly track students' progress [9-10]. It can also strengthen schools' guidance and cultivation

of students' professional skills, enhance students' professional practical abilities, and effectively integrate them into actual job positions within enterprises [11-12]. Additionally, establishing a teaching effectiveness evaluation system can to some extent encourage professional students to organically integrate theoretical knowledge with practical knowledge, thereby enhancing their core competitiveness upon graduation, increasing their targeting and competitiveness in the talent market, and effectively achieving high-standard employment [13-16].

Numerous scholars have conducted related research on this topic. Literature [17] emphasizes that students' practical work capabilities should serve as the core of the school-enterprise cooperation teaching evaluation system. Based on this, a high-quality engineering teaching process evaluation system is constructed under conditions of quality cultivation, school-enterprise integration, and resource sharing, providing a new pathway for the collaborative talent cultivation management mechanism between universities and enterprises. Literature [18] proposes a comprehensive evaluation system for talent cultivation quality involving universities, enterprises, education experts, and students. This system has good applicability to local school-enterprise cooperation operations, effectively ensuring the quality of school-enterprise cooperation and collaborative education, and possesses strong promotional and application value. Literature [19] utilizes the analytic hierarchy process to construct an evaluation indicator system for vocational college students' graduation projects under the school-enterprise cooperation framework. The evaluation results to some extent reflect the development quality of school-enterprise cooperation practical education and hold significant implications for cultivating high-level talent. Literature [20] indicates that the establishment of a project performance evaluation system under the school-enterprise cooperation framework is the foundation for the collaborative operations between vocational colleges and enterprises. The evaluation results can accurately diagnose issues within the cooperative mechanisms, fully tap into cooperative potential, expand cooperative scope, and enhance cooperative outcomes. Literature [21] established a performance evaluation system for school-enterprise cooperation based on the decision tree algorithm. This system effectively addresses the inherent issues of low efficiency and unfair assessment in traditional evaluation systems, while significantly reducing manual labor and promoting the efficient development of school-enterprise cooperation operations. Literature [22] constructed a quality evaluation system applicable to the application-oriented talent cultivation model of school-enterprise cooperation based on the CIPP educational evaluation model. By conducting systematic supervision and effective control of the teaching process, it achieves the goal of cultivating high-quality talent. It is evident that establishing an effective teaching effectiveness evaluation system can intensify reforms and innovations in educational models, enhance the operational effectiveness of school-enterprise cooperation, and ultimately promote deep school-enterprise collaboration.

This paper designs a CVAE-GAN network that integrates conditional variational autoencoders (CVAE) with generative adversarial networks (GAN) to reconstruct high-resolution facial images. Through end-to-end adversarial training, the quality of generation is improved, and the accuracy of facial expression feature extraction in the classroom is enhanced. YOLOv5 is optimized with lightweight modifications, incorporating the SE attention mechanism and hybrid convolution modules to achieve channel weighting and optimize the bounding box regression loss function. While reducing computational load, the system enhances fine-grained feature extraction capabilities. Through collaborative operations with industry-academia partnerships, the system enables multi-dimensional evaluations such as quantifying attention levels during teaching processes and analyzing interactive states, driving the digital transformation of education.

2. Analysis of Deep Learning-Based Student Learning Performance Detection Technology

2.1. CVAE-GAN Network Model

2.1.1. GAN Network Model

Inspired by zero-sum games, GANs learn the probability distribution of training samples by having two networks (generator network and discriminator network) compete with each other to generate realistic images. This method has since been applied to various fields such as image restoration and text generation. The basic structure of a GAN network consists of a generator and a discriminator. where the generator G aims to generate as realistic images as possible to deceive the discriminator D, while the discriminator D aims to distinguish between generated images and real images. This creates a dynamic "game" process between the generator and discriminator. The original GAN network first randomly samples noise data z from the noise data distribution, inputs it into the generative model to obtain the generated sample $G(z)$, and then inputs it into the discriminator D. The discriminator D is a binary

classification network that outputs the probability that the data is real, where real is 1.0 and fake is 0.0. Initially, the discriminator's output is expected to be close to 0.0, but the generator G must deceive D as much as possible. As D gains a deeper understanding of real samples (by learning from an increasing number of samples), G must continuously improve its forgery techniques to deceive D. Figure 1 shows the GAN network structure.

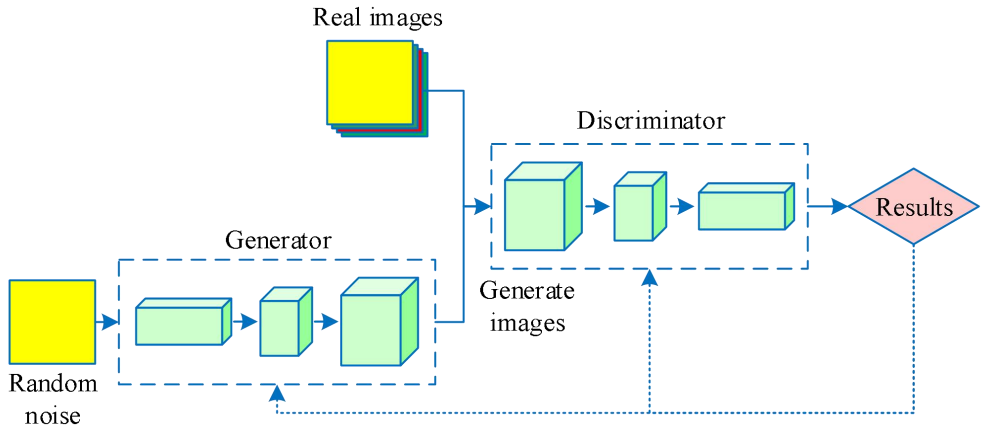

Figure 1. GAN network structure.

The training of generator and discriminator networks is an alternating evolutionary process, and its objective function can be expressed as two parts, one being the discriminator loss:

$$L_D = -E_{x\sim p_{data}}[\log D(x)] - E_{z\sim p_z}[\log(1 - D(G(z)))] \tag{1}$$

The discriminator's goal is to distinguish between real and fake samples as accurately as possible, so it uses the cross-entropy loss function between the predicted values and the actual values of the images to measure this. Another part of the GAN network's objective function is the generator loss:

$$L_{GD} = -E_{z\sim p_z}[\log D(G(z))] \tag{2}$$

For the generator, the goal is to deceive the discriminator as effectively as possible, with the discriminator $D(G(z))$ output approaching 1.0. The objective function of the GAN network is essentially two optimization problems: first optimize D, then optimize G. First, enhance the performance of the generator, then use the enhanced generator and real samples to enhance the performance of the discriminator. The two networks are trained alternately, and ultimately, an effective generator that can generate realistic images is obtained. However, in practice, GAN networks are very difficult to train, and it is challenging to achieve a dynamic balance between the generator and the discriminator. The generated results are often unstable, a phenomenon known as mode collapse.

2.1.2. Conditional Variational Autoencoder

Variational autoencoders (VAEs) are another commonly used generative model. Their key idea is to generate new images by randomly sampling from a random vector, based on variational and Bayesian theory. Generally speaking, the training process after sampling follows a standard distribution. A variational autoencoder consists of two parts: one performs variational inference on the input image, mapping it to a mean and variance that follow a normal distribution, referred to as the encoder; the other learns to approximate the probability distribution of the original data, compiling the encoder's output into an image, referred to as the decoder. The loss function of a VAE is divided into two parts: one uses the Kullback-Leibler (KL) divergence to measure the similarity between the features of the encoded data and the true posterior distribution.

$$L_{KL(p_1\|p_2)} = -1/2^*[2\log\sigma + 1 - \sigma^2 - \mu^2] \tag{3}$$

where p_1 and p_2 represent the prior distribution and the estimated posterior distribution, respectively, and σ and μ represent the variance matrix and mean matrix of the input image inferred by the encoder, respectively. VAE minimizes the KL divergence by constraining the parameters σ and μ. Another loss function is the cross-entropy loss function, which is used to reduce the distance between the original

data y and the generated data y'. The loss function of the variational autoencoder is the sum of the two. To generate images of a specific category, researchers proposed the conditional variational autoencoder (CVAE) model based on the variational autoencoder. The loss calculation for CVAE is the same as that for VAE, except that the labels are converted to one-hot representations and added to the input of the encoder and decoder to generate sample data of a specific category. Since CVAE only roughly calculates the "distance" between the original data and the generated data, this limits the quality of the images generated by CVAE.

2.1.3. CVAE-GAN Network

For CVAE, the generated images are relatively stable but of low quality. For GAN, the generated images are clear, but the generation process is unstable, leading to pattern collapse issues. Therefore, to address the shortcomings of both, this paper proposes combining CVAE and GAN to obtain the CVAE-GAN structure. Figure 2 shows the CVAE-GAN network structure.

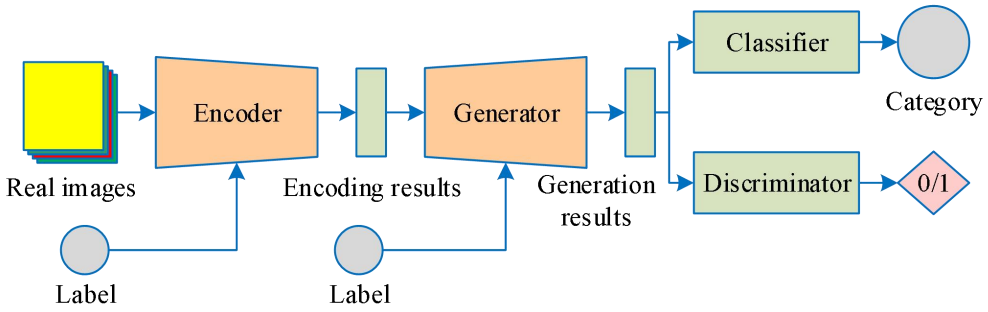

Figure 2. CVAE-GAN network structure.

As shown in the figure, CVAE-GAN consists of four parts: encoder, generator, classifier, and discriminator. The encoder encodes the input image into a latent vector z, and adding the category c can generate a higher-quality latent vector. The generator converts the latent vector into an image, and adding the category c enables the generation of images of a specific category. The classifier is used to determine the category of the generated image, while the discriminator is used to determine whether the image is a real image. These four networks are connected, and image label information is introduced for end-to-end training, enabling the generation of fine-grained images of a fixed category.

2.2. Lightweight Improvements Based on YOLOv5

2.2.1. Deep Learning-Based Object Detection Model (YOLOv5)

YOLOv5 builds on the strengths of its predecessors while introducing a series of innovations aimed at further enhancing the performance and efficiency of object detection. YOLOv5 divides the model into five distinct versions based on width and depth: n, s, m, x, and l, corresponding to FLOPs of 4.50×10^9, 16.50×10^9, 49.5×10^9, 109.5×10^9, and 205.5×10^9, respectively, with corresponding weight sizes of 3.90MB, 14.20MB, 40.80MB, 89.40MB, and 166.90MB. As the network structure becomes wider and deeper, the computational load and weight sizes increase significantly. Although detection accuracy improves slightly, the computational resource consumption increases exponentially, and real-time performance also decreases significantly. Therefore, this paper improves upon the YOLOv5s model, which has a smaller computational load and faster inference speed, to meet the requirements of real-time performance and accuracy. The YOLOv5s network structure consists of an input layer, a backbone network, a neck network, and a detection layer.

2.2.2. SE Attention Mechanism

In deep learning image processing, the attention mechanism can help models focus on important features in images while ignoring unimportant information. By introducing the attention mechanism, the performance and accuracy of models can be significantly improved. As a channel attention mechanism, the SE attention mechanism has achieved remarkable results in deep learning image processing. It adaptsively learns the weights of each channel, enabling the network to learn more important features during training and thereby enhancing the model's ability to extract semantic information from images.

Figure 3 shows the structure of the SE attention mechanism. The key steps of the SE attention

mechanism include compression and activation. During the compression process, the input feature map is operated on using global average pooling, and the average value is calculated for each channel so that the network can capture the global information of each channel. For a feature map of size $C \times H \times W$, after global average pooling, a feature tensor $z \in \mathbb{R}^C$ of size $C \times 1 \times 1$ is obtained, whose mathematical expression is as follows:

$$z_i = \frac{1}{H \times W} \sum_{j=1}^{H} \sum_{k=1}^{W} X_{i,j,k} \tag{4}$$

$$FC1(z, W_1) = \mathrm{Re}\,LU(W_1 \cdot z) \tag{5}$$

$$FC2(FC1(z, W_1), W_2) = \sigma(W_2 \cdot FC1(z, W_1)) \tag{6}$$

$$X'_{i,j,k} = X_{i,j,k} \cdot FC2(FC1(z, W_1), W_2) \tag{7}$$

where H and W represent the height and width of the feature map, respectively, j and k represent the horizontal and vertical pixel coordinates of the feature map, respectively, σ is the Sigmoid activation function, $FC1$ represents the first fully connected layer, $FC2$ represents the second fully connected layer.

Two fully connected layers are introduced during the activation process to learn the attention weights for each channel. First, the feature tensor output from global average pooling is fed into a small fully connected layer (FC1), then the ReLU activation function is applied, as shown in formula (5). The output is then fed into the second fully connected layer (FC2), and the resulting output is fed into the Sigmoid activation function for normalization, as shown in formula (6), at which point the attention weights for each channel are obtained. Finally, the original feature map X is multiplied by the learned channel attention weights to obtain the weighted feature map X' as shown in Formula (7). The SE attention mechanism focuses more on important channels, thereby enhancing the network's ability to perceive critical information.

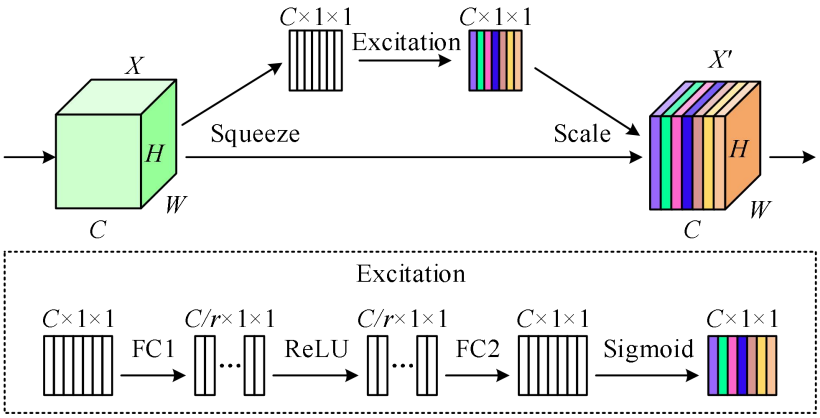

Figure 3. Structure of the SE attention mechanism.

2.3. Lightweight Convolution Module Settings

2.3.1. GSConv Principle

In CNNs, the number of model parameters and floating-point operations is currently reduced mainly by using deep separable convolution (DSC) operations instead of standard convolution SC. Figure 4 shows the convolution processes of SC and DSC.

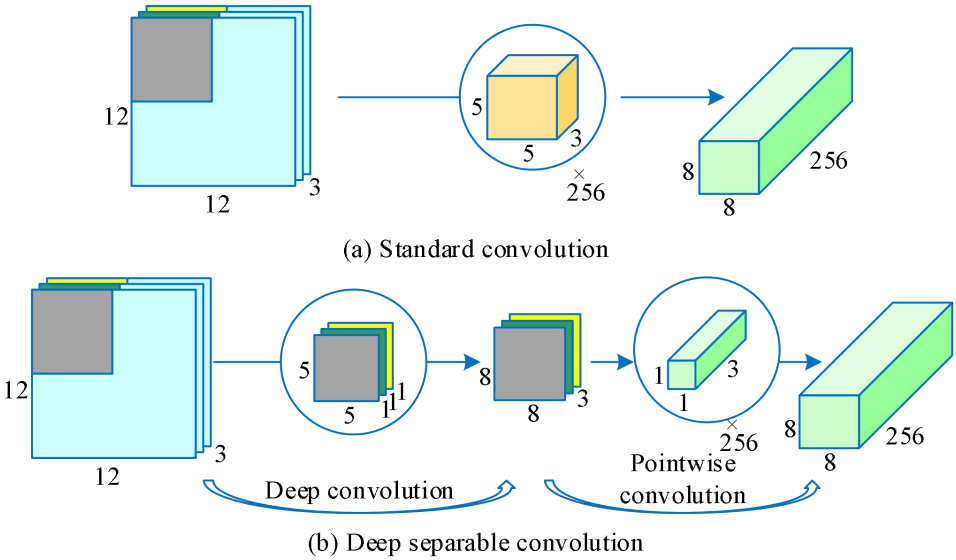

Figure 4. Calculation process of SC and DSC.

Standard convolution SC simultaneously performs feature extraction and channel fusion on the input feature map. The use of channel-dense convolution maximizes the retention of hidden connections between each channel, providing both spatial description and semantic depth.

Depth-Separable Convolution (DSC) decomposes standard convolution into depth convolution (DW) and pointwise convolution (PW). In DW convolution, the convolution kernel is split into single-channel form, and convolution operations are performed on each channel without altering the depth of the input feature image. This results in an output feature map with the same number of channels as the input feature map. Next, pointwise convolution performs a 1×1 standard convolution on the output feature map of the DW convolution, achieving channel-wise fusion to obtain a feature map of the same size and depth as the standard convolution, i.e., $9\times9\times360$. In this case, due to the limited number of channels in DSC, the feature map has insufficient dimensions, and the channel information of the input image is separated during the computation process, resulting in significantly lower feature extraction and fusion capabilities compared to SC.

Equation (8) is the parameter calculation formula for standard convolution (excluding bias), and Equation (9) is the parameter calculation formula for depth-separable convolution (excluding bias). Where K_h is the height of the convolution kernel, K_w is the width of the convolution kernel, C_{in} is the number of channels in each convolution kernel and also the number of channels in the input feature map, and C_{out} is the number of channels in the output feature map. Typically, the time complexity of convolution calculations is defined by the number of parameters. Therefore, by separating feature extraction and channel fusion, deep separable convolution significantly reduces the number of parameters.

$$param(SC) = K_h \times K_w \times C_{in} \times C_{out} \tag{8}$$

$$param(DSC) = K_h \times K_w \times C_{in} + C_{in} \times C_{out} \tag{9}$$

In order to make the output of lightweight convolutional DSC as close to SC as possible, GSConv uses a shuffle operation to combine DSC with SC, allowing the information generated by the SC channel dense convolution operation to permeate every part of the information generated by DSC. Figure 5 shows the structure of the GSConv module.

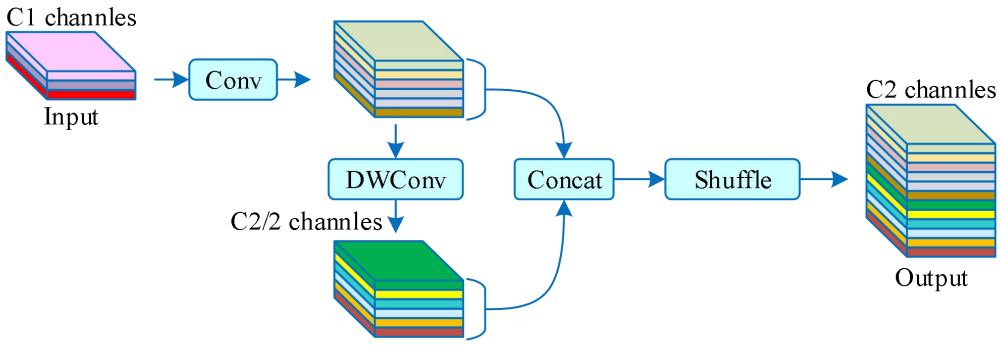

Figure 5. GSConv module structure.

The number of parameters for GSConv (without bias) is given by the formula shown in Equation (10). Comparing this with Equations (8) and (7), it can be seen that the number of parameters for GSConv lies between those of SC and DSC. In Formula (10), when comparing GSConv with SC, the number of parameters in GSConv is $\left[-2.0/\left(C_{in}+1\right)+2.0\right]$ times fewer than that in SC. The floating-point computation volume of these convolutions is multiplied by the height and width of the input image in the following formula, resulting in a significant reduction in floating-point operations, thereby greatly reducing the cost of convolution and lowering the model complexity.

$$param(asConv) = K_h \times K_w \times \frac{C_{out}}{2}(C_{in}+1) \tag{10}$$

$$\frac{param(SC)}{param(CSConv)} = -\frac{2}{C_{in}+1}+2 \tag{11}$$

2.3.2. Loss Function Optimization Design

To further improve the accuracy and detection rate of student classroom behavior detection, this paper designs an optimized default loss function for the benchmark network.

The IoU loss function is of great value to deep learning-based detectors, as it makes the predicted bounding box regression position more accurate. In the benchmark network, the predicted box regression loss function defaults to CIoU loss, but CIoU does not take into account the mismatch between the predicted box and the real box in terms of direction. This directional mismatch may result in slower convergence speed and lower convergence efficiency, causing the predicted bounding box to "wander aimlessly" during model training, ultimately degrading model performance. Based on the above reasons, this paper's algorithm replaces the CIoU loss function with the latest SIoU loss function. The SIoU loss function addresses issues related to the angle, distance, shape, and overlap area between the predicted and ground truth bounding boxes. Figure 6 shows the parameters of the SIoU loss.

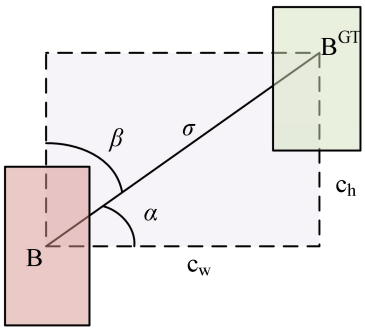

Figure 6. Each parameter in the SIoU loss.

The angle loss formula is shown in Formula (12):

$$\Lambda = 1 - 2\sin^2\left(\arcsin\left(\frac{c_h}{\sigma}\right) - \frac{\pi}{4}\right) \tag{12}$$

In the formula, c_h is the height difference between the center points of the predicted box and the true box, and σ is the distance between the center points of the predicted box and the true box. Based on the angle loss, the distance loss is redefined, and the distance loss formula is shown in Formula (13):

$$\Delta = \Sigma_{t=x,y}\left(1 - e^{-\gamma\rho_t}\right) \tag{13}$$

In the formula:

$$\rho_x = \left(\frac{b_{cx}^{gt} - b_{cx}}{c_w}\right)^2, \rho_x = \left(\frac{b_{cy}^{gt} - b_{cy}}{c_h}\right)^2, \gamma = 2 - \Lambda \tag{14}$$

When the angle α approaches 0°, the contribution of distance loss is greatly reduced. Conversely, when α approaches 50°, the contribution of distance loss becomes increasingly significant.

3. Experiment on Teaching Effectiveness Evaluation Based on Deep Learning

3.1. Comparative Experiments

3.1.1. Model Performance Comparison Experiment

The study compared three classic object detection algorithms with the improved YOLOv5 algorithm proposed in this paper, including YOLOv8, YOLOv5(l,s), and Faster-RCNN. Performance was evaluated by comparing mAP values, F1 scores, parameters, and FPS under the same experimental conditions. Table 1 shows the comparison results: Compared to YOLOv8l, YOLOv5l, YOLOv5s, and Faster-RCNN, the improved YOLOv5 model demonstrates significant performance improvements. It achieves 79.78% mAP and 0.82 F1, with fewer parameters (12.17M) and faster inference speed (70.63 FPS), making it more suitable for deployment on edge devices and capable of meeting the real-time requirements for detecting student classroom behavior.

Table 1. Results of the comparative experiment on student behavior recognition.

Model	mAP/%	F1	Param/M	FPS
YOLOv8l	75.05	0.78	35.41	73.18
YOLOv5l	73.52	0.71	38.24	70.81
YOLOv5s	74.16	0.72	28.49	75.32
Faster-RCNN	67.87	0.75	128.56	73.56
Improve YOLOv5	79.78	0.82	12.17	70.63

3.1.2. Comparison Experiment on Facial Expression Recognition in Classrooms

The effectiveness of improving YOLOv5 in the task of recognizing students' facial expressions in the classroom is evaluated using accuracy (Ac). Accuracy is the proportion of correctly classified samples out of the total number of samples in a classification problem. In the task of facial expression recognition, accuracy reflects the model's ability to correctly identify all facial expression categories.

The student faces recorded by the teacher's camera are generally blurry. First, image high-resolution reconstruction technology is used to restore facial expression images. Figure 7 shows the experimental results of the original facial expressions and those reconstructed and enhanced using CVAE-CAN. The recognition accuracy of facial expression images reconstructed and enhanced using CVAE-CAN ranges from 75% to 90%, with a maximum of 87.85%. In contrast, the recognition accuracy of the original, unprocessed images ranges from 70% to 85%, with a maximum of 84.79%. The results indicate that CVAE-CAN's image processing affects the model's accuracy in recognizing students' facial expressions.

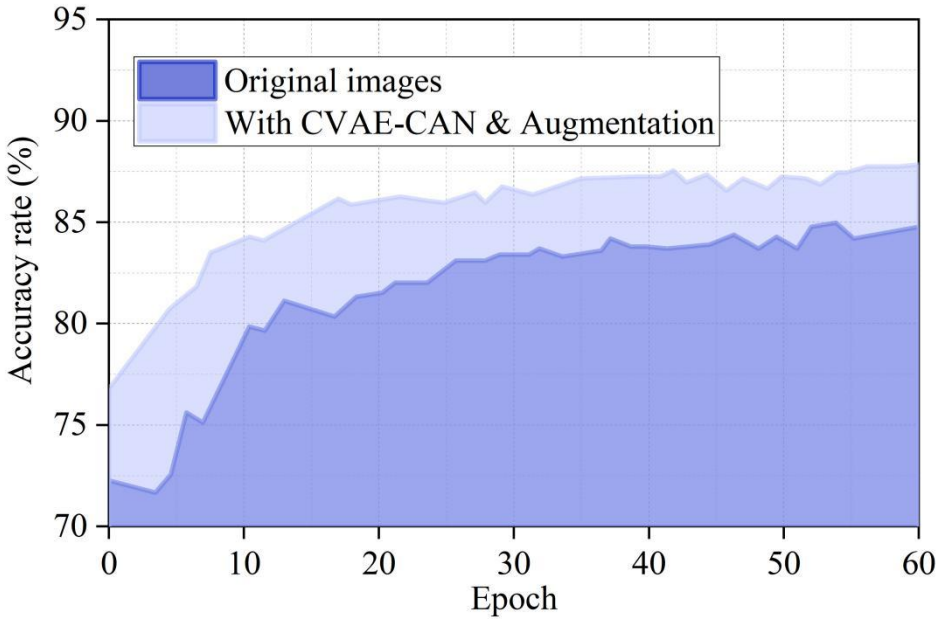

Figure 7. Comparison of original data and enhanced of CVAE-CAN.

The facial expression images reconstructed and enhanced using CVAE-CAN are the subject of this study. Figure 8 compares the recognition performance of the POSTERv2, DACL, EAC, and improved YOLOv5 facial expression recognition models. The improved YOLOv5 achieves the highest accuracy rate of 95.93%, effectively meeting the requirements for accurately identifying students' facial emotions in classroom settings. The maximum accuracy rates for DACL, EAC, and POSTERv2 are 93.07%, 89.52%, and 90.91%, respectively, which are lower than those of the improved YOLOv5, and their fitting processes are relatively slower.

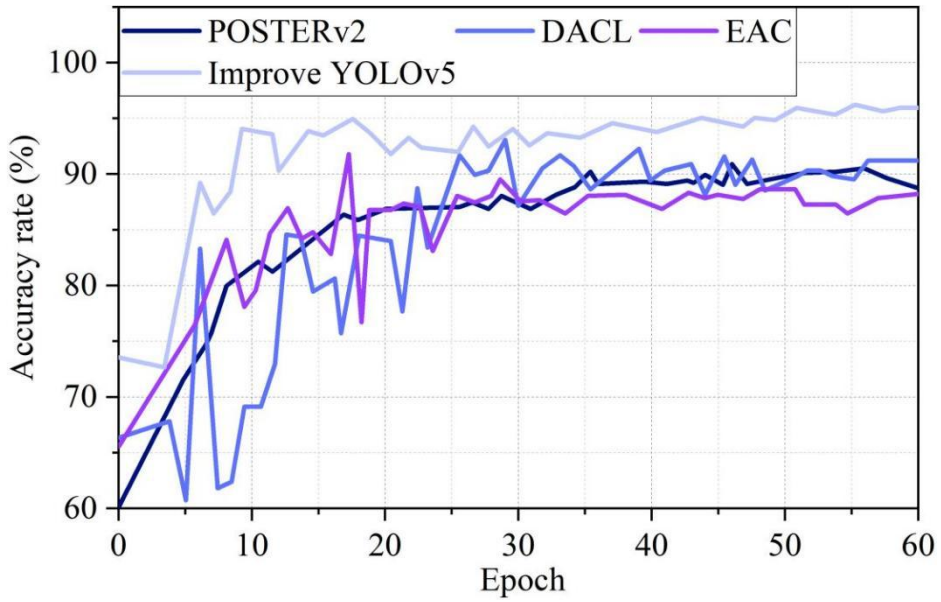

Figure 8. Comparison of recognition effects of expression recognition models.

3.2. Comparison of Discrete Expression Recognition Performance Before and After Improvement

Common classroom expressions among students include neutral, pleasant, tired, confused, and resistant, with each expression exhibiting diverse and subtle differences in expression. To assess the improved model's recognition performance, a comparative experiment was conducted to evaluate the specific classification results of classroom expressions before and after model improvements. Figure 9

shows the confusion matrix of the pre-improved YOLOv5 model on the self-built dataset. Figure 10 shows the confusion matrix of the improved YOLOv5 model on the self-built dataset. The recognition accuracy rates for the five discrete student expressions before improvement were 90.27%, 81.55%, 68.93%, 40.26%, and 40.65%, respectively. Except for neutral and pleasant expressions, the recognition accuracy rates for the remaining expressions were relatively low. After improvement, the recognition accuracy rates for the five discrete student expressions were 97.15%, 92.63%, 92.74%, 90.82%, and 91.44%, respectively, with improved recognition accuracy rates for all expression categories. Especially for the three expression categories of fatigue, confusion, and resistance, the recognition accuracy rates improved by 23.81%, 50.56%, and 50.79%, respectively. The model improvements are effective and can be applied to real-time assessment of teaching effectiveness.

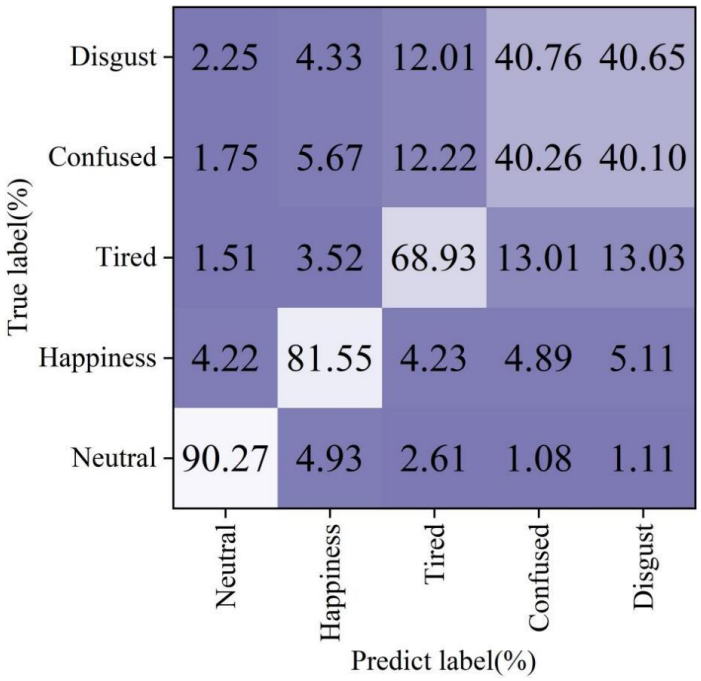

Figure 9. Discrete expression confusion matrix before improvement.

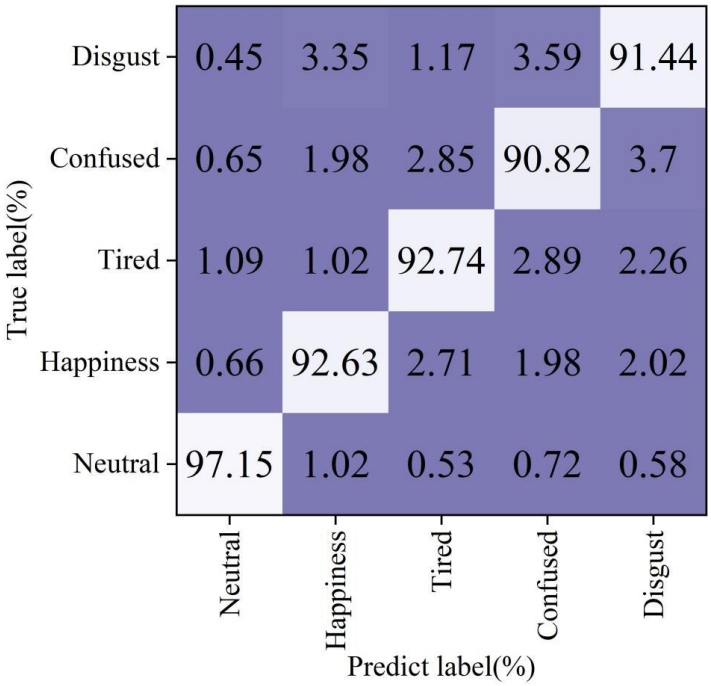

Figure 10. Discrete expression confusion matrix after improvement.

3.3. Evaluation and Analysis of Classroom Teaching Effectiveness

3.3.1. Student Focus Evaluation

Deploy the improved model in a real-time feedback system for real-time assessment of classroom teaching effectiveness in the Business Japanese program at a certain university. Attention span is used as the evaluation metric to measure students' classroom learning performance and teachers' teaching effectiveness. To facilitate teachers in observing students during lectures and adjusting teaching methods and approaches, 40 Business Japanese students were randomly selected as the system's primary research subjects, with two observers conducting remote observations via cameras. The course content “Business Japanese Etiquette” was taught to the research subjects to obtain their attention span data over an 80-minute period.

Figure 11 shows the changes in Student A's classroom focus based on the results of facial expression recognition. From the figure, it can be seen that Student A's focus reached a peak of 94.62% during the entire class, indicating that the student maintained a high level of focus throughout the class. However, from the 4th minute to the 6th minute, the student's focus decreased, possibly due to the fact that the course had just begun, and the student transitioned from conscious focus to unconscious focus. The observers noted that due to the system's reminder function, the student promptly adjusted their classroom state. From the 11th minute onwards, the student returned to a focused state and reengaged with the class. By the 34th minute, the student began to show signs of fatigue, and their focus started to decline. The teacher's timely interaction with Student A during the class helped the student return to a focused state. By the 70th minute, as the class neared its end, the focus level dropped sharply. Although the student's focus was affected by certain factors during the class, the overall evaluation indicates that the student remained in an active state. Students can use the focus curve to review the knowledge points covered by the teacher during the corresponding time period. Throughout the teaching process, the assessment system can quickly and accurately identify students' classroom expressions and calculate their focus levels, enabling teachers to promptly adjust individual students' focus states. Additionally, teachers did not exhibit any resistance to using the system.

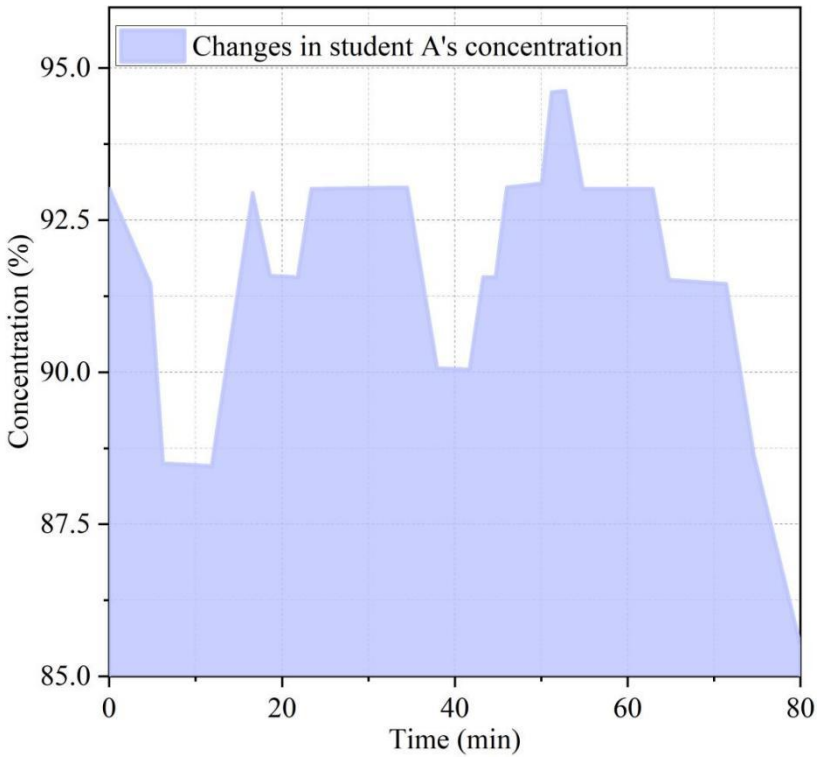

Figure 11. Changes in students' concentration in class.

3.3.2. Classroom Focus Evaluation

Figure 12 shows the average changes in attention levels among 40 students throughout the entire class. Within the first 5 minutes of the class, students' attention levels rapidly dropped from a high of 94.62% to around 87.05%. Subsequently, under the teacher's reminders and as the lesson progressed, students'

attention levels exhibited dynamic fluctuations during the learning process from 12 minutes to 70 minutes. As the class neared its end at 70 minutes, students became easily distracted by other matters, resulting in a significant decline in attention levels. Overall, students' attention levels remained above 85% throughout the class, indicating a high level of engagement. Based on the identification results and attention level calculations, teachers can identify key time points in the dynamic changes of attention levels during actual teaching, adjust the pace and content of instruction, and engage in timely interaction with students to keep their interest sustained throughout the class.

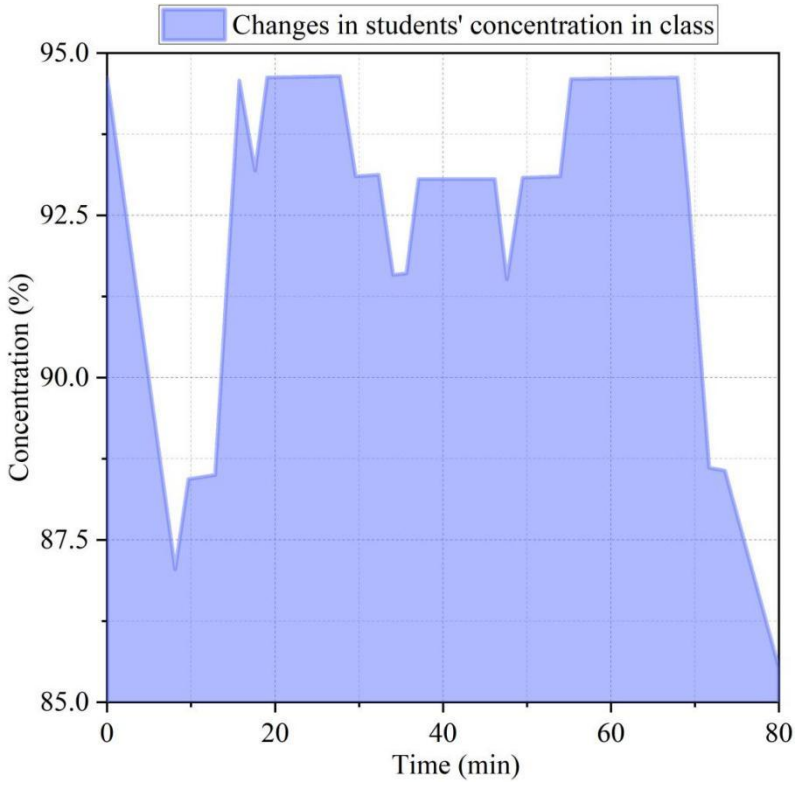

Figure 12. The changes in classroom concentration.

3.3.3. Changes in Classroom Interaction Status Over Time

Table 2 shows the changes in students' interactive states over time during the entire class. Starting from the beginning of the class, interactive states were analyzed at 10-minute intervals. At each time point, the average interactive score is a comprehensive indicator reflecting the overall interactive situation in the classroom. A higher average interactive score indicates that students in the classroom perform better in terms of interaction, have higher levels of attention, and that the teacher's instructional strategies are effective. Based on the average interactive scores, student interaction levels were lowest around the 10-minute mark and the 80-minute mark, at 0.755 and 0.700, respectively. Interaction levels peaked around the 30-minute and 60-minute marks, reaching 0.925 and 0.911, respectively. Overall, the changes in interaction levels align with the adjustments made by the teacher based on the identification results, further demonstrating that using an assessment system for real-time expression recognition and classification can effectively evaluate teaching effectiveness and facilitate adjustments.

Table 2. The changes of classroom interaction status over time.

Time	Average interaction value	Number of interlocutors	Non-interactive number of people
00:00	0.832	20	20
00:10	0.755	10	30
00:20	0.904	35	5

00:30	0.925	36	4
00:40	0.893	28	12
00:50	0.870	25	15
00:60	0.911	32	8
00:70	0.801	15	25
00:80	0.700	9	31

4. Deep learning-Based Corporate Cooperation and Operation Strategies

1) Adhere to the production-oriented training model under the school-enterprise cooperation operation mode of "promoting learning through production."

The training of students in colleges and universities is based on the standard of adapting to future work requirements. Therefore, the best solution in the educational process of schools is to "promote learning through production," so that students can fully understand what their future work will be and what they need to learn in school. Through the school-enterprise cooperation operation mode of "promoting learning through production," they can grasp market demand. Two primary approaches can be adopted. The first is the "on-campus training base integration model," where the school provides the facilities, while the enterprise supplies the technical expertise and training personnel. The second form is the "external on-campus training base," where the training base is established within the enterprise, allowing students to directly experience production and practical work in their future potential work environment. Relevant training personnel utilize a multi-channel teaching assessment system based on deep learning to receive real-time feedback on students' classroom expressions, thereby improving the quality of classroom content and teaching methods. This provides students with effective "industry-driven education" information, helping them understand occupational requirements in advance.

2) Adhere to a strategy of distinctive development in the construction of productive training bases.

The production-oriented teaching practices of university-enterprise cooperation should have Chinese characteristics, prioritizing students' education and practical experience to truly lay a solid foundation for their future careers. The management model of the bases should approach the modern enterprise environment of quasi-scientific management. Through "scientific teaching systems," "comprehensive teaching facilities," "strict organizational discipline," "meticulous scientific attitudes and team spirit," and "advanced teaching methods," students receive training and development in occupational ethics and comprehensive qualities within the authentic online production environment of on-campus production-oriented training bases. Introducing multi-channel teaching assessment systems based on deep learning into teaching facilities enhances the modernization and intelligence of teaching.

5. Conclusion

This paper constructs a teaching evaluation system based on deep learning and verifies its effectiveness, providing technical support for school-enterprise cooperation practices. The mAP and F1 scores of YOLOv5 are improved to 79.78% and 0.82%, respectively, with only 12.17 million parameters and an inference speed of around 70.63 FPS, featuring lightweight characteristics. The accuracy rates of the five types of facial expressions all exceed 90%, and the recognition accuracy rates of fatigue, confusion, resistance, with significant improvements of 23.81%, 50.56%, and 50.79%, respectively. This enables precise identification of students' classroom focus and interaction status. Future research can further explore cross-scenario detection capabilities and deepen privacy protection mechanisms in corporate collaborations to safeguard students' development.

References

1. Cheng, S., Chen, Z., & Wang, J. (2024, October). Research on the reform of school-enterprise cooperative teaching and education mode. In Proceedings of the 2024 International Conference on Artificial Intelligence and Teacher Education (pp. 131-135).
2. Zhenyan, Q., Pruettikul, S., & Howattanakul, S. (2023). School-Enterprise Cooperation Educaional Management Mode of University in Unhui Province. Journal of Roi Kaensarn Academi, 8(9), 354-364.
3. Peng, F., & Zhang, F. (2021). Research and Analysis on the Model of School-Enterprise Collaborative Education in Higher Vocational Colleges. Advances in Educational Technology and Psychology, 5(4), 66-72.

4. Qin, Q., & Lei, Y. (2024). Research on existing problems and countermeasures in school-enterprise cooperation in private higher vocational colleges. Journal of Education and Educational Research, 7(1), 222-226.
5. Li, C., Li, G., & Shi, Y. (2019). Analysis of the construction and implementation of the new mode of school-enterprise cooperation in higher vocational colleges from the perspective of educational reform. Open Journal of Social Sciences, 7(11), 246.
6. Fan, X., Liu, Y., Han, Y., Shi, Y., Liu, H., & Ma, L. (2019). Research on the Cooperative Education Mode of Schools and Enterprises in the New Era. DEStech Transactions on Social Science, Education and Human Science.
7. Yu, Y., Gu, H., Liang, B., Chen, X., Chen, Z., & Wang, L. (2024). Performance Evaluation of School Enterprise Collaborative Innovation and Practice of Innovation and Entrepreneurship Education Based on the Improved AHP Fuzzy Comprehensive Evaluation Method. Discrete Dynamics in Nature and Society, 2024(1), 5583728.
8. Indarti, S. (2021). The effects of education and training, management supervision on development of entrepreneurship attitude and growth of small and micro enterprise. International Journal of Organizational Analysis, 29(1), 16-34.
9. Bian, F., & Wang, X. (2021). School enterprise cooperation mechanism based on improved decision tree algorithm. Journal of Intelligent & Fuzzy Systems, 40(4), 5995-6005.
10. Zhengli, Y. A. N. G., Shixiang, C. A. I., Heng, L. U., Xia, X. I. A. N. G., & Guohong, L. I. (2021). Research of Cultivating Innovative Talents in the School-Enterprise Cooperation Model. Experiment Science and Technology, 19(2), 132-136.
11. Paľová, D., Czaja, A., & Vejačka, M. (2018). Innovative approach to education improvement via enterprise-education collaboration. Quality Innovation Prosperity, 22(3), 68-82.
12. Chen, K., Tham, J., & Khatibi, A. (2025). Research on Strategies for Enhancing the Employment Ability of Vocational College Graduates from the Perspective of School Enterprise Cooperation. Frontiers in Educational Research, 8(4).
13. Cadez, S., Dimovski, V., & Zaman Groff, M. (2017). Research, teaching and performance evaluation in academia: the salience of quality. Studies in Higher education, 42(8), 1455-1473.
14. Ratnay, G., Indriaswuri, R., Widayanthi, D. G. C., Atmaja, I. M. P. D., & Dalem, A. A. (2022). CIPP Evaluation Model for Vocational Education: A Critical Review. Education Quarterly Reviews, 5(3), 1-8.
15. Misbah, Z., Gulikers, J., Dharma, S., & Mulder, M. (2020). Evaluating competence-based vocational education in Indonesia. Journal of Vocational Education & Training, 72(4), 488-515.
16. VARZHAPETYAN, A. G., SEMENOVA, E. G., FOMINA, A. V., BALASHOV, V. M., & BALASHOVA, K. V. (2019). Assessing the quality and effectiveness of additional vocational education. Revista Espacios, 40(02).
17. Yuan, S., Shu, H., & Yang, Y. (2022, August). Exploration and Practice of High-Quality Engineering Teaching Process Evaluation System of College-Enterprise Cooperation in Colleges. In International Conference on Computer Science and Education (pp. 560-566). Singapore: Springer Nature Singapore.
18. Yan, X., Zhang, W., Wang, H., Jia, L., Hou, Y., Xu, T., ... & Wang, Y. (2019, March). Construction of the Quality Assurance System for Collaborative Education of Schools and Enterprises Cooperation in Local Universities. In 2018 8th International Conference on Education and Management (ICEM 2018) (pp. 144-147). Atlantis Press.
19. Jiang, W., Chai, P., Chen, Y., & Wang, S. (2015, July). University-enterprise Cooperation Practical Teaching Evaluation Model for polytechnic based on AHP and Fuzzy. In 2015 International Conference on Computational Science and Engineering (pp. 282-285). Atlantis Press.
20. Tan, S., & Zhang, C. (2019, May). Study on the Construction of Performance Evaluation Index System of School-enterprise Cooperation Project in Higher Vocational Colleges. In 1st International Conference on Business, Economics, Management Science (BEMS 2019) (pp. 616-621). Atlantis Press.
21. Yan, Y. (2022). Decision Tree Algorithm in the Performance Evaluation of School-Enterprise Cooperation for Higher Vocational Education. Mathematical Problems in Engineering, 2022(1), 4151168.
22. Xu, Z. (2022). Research on Application Evaluation Index System of University Enterprise Cooperation Informatization Based on CIPP. In The 2021 International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy: SPIoT-2021 Volume 1 (pp. 1065-1070). Springer International Publishing.