Publication Date: 31 July 2024
Archs Sci. (2024) Volume 74, Issue 4 Pages 142-152, Paper ID 2024419.
https://doi.org/10.62227/as/74419

Research on the Coordinated Development of Higher Vocational English Education and Student Management Based on Deep Learning

Yaoli Wang[1,*]
[1]Wuxi City College of Vocational Technology, Wuxi, Jiangsu, 214073, China.
Corresponding authors: (e-mail: sunshinewyl@126.com).

Abstract As a key link in the development of higher vocational English education, student management lays the foundation for talent cultivation and affects the speed and quality of vocational skills training. This paper proposes a student behavior recognition algorithm based on YOLOv5, and introduces the EIoU loss function and Varifocal Loss function to optimize the YOLOv5 network, in order to take into account the real-time detection and higher accuracy. At the same time, the classroom behavioral states are classified, and the fuzzy comprehensive evaluation method is combined to quantitatively assess the degree of concentration, and empirical evidence of higher vocational English education is carried out with X university as the research object. The results show that the average accuracy rate of the student behavior recognition algorithm proposed in this paper is 0.86875, which can more accurately identify student learning behaviors in the English classroom and improve the efficiency of student management. In addition, the concentration scores of individual students and all students in the whole class are 78.073 and 78.274, respectively, which are both higher than 75, indicating that students tend to be more concentrated in the whole class. Based on the results of the evaluation and analysis of student behavior and classroom concentration, this paper puts forward a three-point strategy for the coordinated development of higher vocational English education and student management, which provides a reference basis for promoting the science of college management construction and the development of student training and education.

Index Terms yolov5, EIoU, fuzzy comprehensive evaluation, English language education, student management

I. Introduction

With the development and application of deep learning theory, teaching methods in the field of education are undergoing a revolution. For how to integrate deep learning theory into teaching design and apply it in practice, it has become a hot topic in educational research [1]–[3]. Among them, university English course teaching and student management have become an important direction of educational reform, which aims to effectively integrate and coordinate English education and student management to promote the overall development of students [4]–[6]. However, how to design a teaching model for higher vocational English courses that can promote students' deep learning and effectively achieve English learning goals is still an issue that needs to be thoroughly studied and explored.

Deep learning, as a concept and idea of learning, contrasts sharply with surface learning and strategic learning; deep learning is not simply memorizing and understanding, but actively exploring, analyzing, critiquing, and reflecting on what is being learned, to develop intrinsic understanding and deep insight [7]–[9]. Deep learning is aimed at the independent construction and application of knowledge, building a systematic knowledge system under students' independent inquiry and being able to apply it flexibly. It requires students to be able to self-monitor and regulate the learning process, knowing when they need more understanding and when they need to adjust their learning strategies to achieve the most effective learning [10]–[12]. English is a language output-based activity, which requires students to master systematic knowledge of English and internalize it, and apply it flexibly after active internalization. Teachers need to give students full autonomy when organizing higher vocational English teaching activities, in the creation of an immersion classroom atmosphere, guiding students to independently accumulate a wealth of materials; in the use of network resources, so that students independently find problems, solve problems, learn to think in English, and encourage students to reflect and think critically, not only to accept and understand the knowledge, but also to evaluate and question the knowledge, to develop their own understanding and opinions [13]–[15].

Literature [16] introduces a deep learning feature extraction method that applies the idea of deep learning to multimodal feature extraction. It is used to convert features of different modalities into features of the same modality. A hybrid network English participle processing method was proposed. Literature [17] combined local knowledge with global knowledge and deep learning methods to propose a memory neural network method that combines local knowledge with global knowledge. The method significantly improves the quality of English translation, resulting in more effective and richer context vectors that more accurately represent the contextual situation. Literature [18] designed a grammar analysis method combining attention mechanism, word embedding and CNN seq2seq using deep learning algorithms on the basis of seq2seq model, and the experimental results show that the method designed in this study is effective in grammar analysis, and it can be applied and popularized in practical English teaching. Literature [19] presents a new system for practicing lexical stress in second language English learning with Amazon Alexa home assistant. The main scientific contribution of this work is a deep learning model for automatically assessing lexical stress in non-native English speakers. The results show that the system is able to interactively create vocabulary for a specific speaker. The system proposed in the literature [20] is an intelligent writing scoring system for teaching English at university level. It uses popular big data analytics and deep learning to differentiate training algorithms. Compared with traditional manual scoring, the technique is more convenient, fast, concise and effective. It is of great significance to improve the efficiency of college English writing teaching. Literature [21] proposes a new English education model based on artificial intelligence and evaluates students' comprehensive English ability through deep learning. Experiments show that the evaluation model constructed in this paper is effective and verifies the feasibility of the artificial intelligence method applied to college English education. Literature [22] focuses on constructing a neural network model, explaining the concept of restricted Boltzmann machine, integrating BP algorithm, and generating the differentiation of multi-parameter evaluation indexes of college students' spoken English. Finally, the article establishes a reference framework for the evaluation index system mainly from the needs of college students' English speaking ability.

In this paper, the classical target detection model YOLOv5 is first selected on the basis of deep learning algorithms for recognizing student behaviors, and combined with EIoU and Varifocal Loss, an optimized loss function strategy is used to make the model more suitable for the task of student behavior recognition in complex classrooms and to obtain higher accuracy. On the basis of the identification of students' behavioral states, they are divided according to the degree of positivity, and the fuzzy comprehensive evaluation method is used to quantify the degree of concentration, and the students' classroom concentration scores and their concentration grades are counted. Then we selected the video of English education classroom recording of the first unit of the first book of the eighth grade in X college teaching platform for empirical analysis, firstly, we verified the validity and feasibility of the student behavior identification model proposed in this paper from the perspectives of analysis of the model training effect and the overall analysis of the classroom students' behavior, and then we analyzed the classroom concentration degree of the students in terms of individuals and the whole classroom. Finally, the results of the empirical analysis are combined to explore the coordinated development strategy of higher vocational English education and student management.

II. Deep Learning-Based Student Behavior Recognition Model

A. Student behavior recognition algorithm based on YOLOv5

For further research on how to balance real-time detection with high accuracy for the task of student behavior recognition, the classical target detection model YOLOv5 is chosen for the experiments, which possesses high accuracy.

1) YOLOv5 modeling

The YOLOv5 model is a more classic algorithm in the YOLO family, with innovative improvements that provide better performance than networks of the same class, while having a smaller number of parameters. The combination of different network depths and network widths gives flexibility when used in the YOLOv5 network, and the network structure is shown in Figure 1. The input side uses adaptive anchor frame calculation to set the anchor frame size automatically, and uses adaptive padding to prevent the large number of black edges that appear when the input image is scaled, which affects the detection results. The standard of experimental equipment is reduced, and a smaller Mini-BatchSize can be used to train better results, reducing GPU memory usage.Focus slicing technique is an innovative approach of YOLOv5 in the Backbone backbone network part, which can be used to increase the speed of computation.Cross-stage localized structure (CSP) in the CSPNet can solve the computation bottlenecks. The Spatial Pyramid Pooling (SPP) structure can be used to process images of different sizes, and is used in the network to take into account the fusion of features at different levels, enriching the expression of the feature map to obtain, and at the same time, the multi-scale features can help the model converge quickly and improve the accuracy of the model.Neck part of the use of the feature pyramid network combined with the path aggregation network, firstly, after the up-sampling operation, the high-level feature information and low-level features are fused, and the computation of the feature information and low-level features is performed. low-level features are fused to compute the predicted feature map and compute the network's understanding of the semantic features.The output of each stage of the FPN connects the feature pyramid PAN, which conveys the strong semantic features through the top-down FPN layer, while conveys the strong localization features through the bottom-up feature pyramid, and the parameter

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

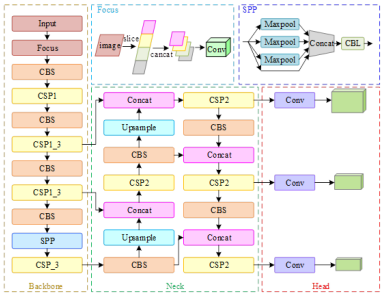

Figure 1: YOLOv5 Network structure

aggregation is performed from the different backbone layers, for the different detection layers.

2) Optimizing the YOLOv5 network

YOLOv5 utilizes the loss function to measure the gap between the true value and the predicted value during the training process, so the selection of the loss function plays a very important role in the training of the model. In this paper, we optimize the loss function of YOLOv5 network according to the needs of student behavior recognition task.

1) Introduction of EIoU loss

YOLOv5s internally embedded loss function IoU and CIoU loss function, the choice of appropriate localization loss function will make the convergence faster and can achieve better results. This paper uses two types of loss functions for experiments, and for the shortcomings of the built-in loss, this paper optimizes the design of the loss function on the basis of the original model and introduces the EIoU loss function. The calculation methods of different loss functions are described below:

IoU is the degree of intersection of the true and predicted frames and is calculated L_{IoU} by the formula:

$$L_{IoU} = 1 - \frac{X \cap Y}{X \cup Y} = \frac{W}{S}, \tag{1}$$

where X and Y are the real frame and the predicted frame respectively, when frame X and Y intersect, W is the overlapped part of the two frames, and S is the overall area of the two frames, when there is a situation where there is no overlapped part of the two frames, the value of the IoU is 0. Therefore, the distance between the two frames can not be judged by the overlapped area of the two frames only, and the IoU still has a big drawback in the prediction.

GIoU is a non-overlapping area penalty term added to IoU, which solves the situation that the gradient can't be backtracked when the real box and the predicted box don't intersect. As shown in Figure 1, assuming that rose-pink A is the real box, yellow B is the predicted bounding box, the orange-filled part is the penalty term, and the pink box C in which the predicted box is the smallest outside rectangular box with the real box, L_{GIoU} is given by:

$$L_{GIoU} = 1 - IoU + \frac{|C - (X \bigcup Y)|}{|C|}. \tag{2}$$

In student behavior recognition, when it appears that the real frame is included with the predicted frame, the second half of the penalty term of the GIoU loss function fails, and the GIoU degenerates into the IoU, and the model convergence slows down.

In order to compensate for the defects of GIoU, the CIoU loss is proposed in the latest version of YOLOv5, and the formula for L_{CIoU} is:

$$\begin{aligned} L_{CloU} &= 1 - CIoU \\ &= 1 - \left(IoU - \frac{\rho^2\left(b, b^{gt}\right)}{c^2} - \alpha v \right), \end{aligned} \tag{3}$$

$$\alpha = \frac{v}{(1 - IoU) + v}, \tag{4}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2, \tag{5}$$

where v is used to measure the aspect ratio of the detection frame to the real frame, $\rho^2\left(b, b^{gt}\right)$ denotes the Euclidean distance between the centroids of the prediction rectangle frame and the labeling rectangle frame, and C denotes the diagonal distance of the smallest closure region that can contain both the prediction rectangle frame and the labeling rectangle frame.

CloU is used in the aspect ratio consistent parameter v is not clearly defined, and there is still a gap with the confidence level of the real degree, which will prevent the model to carry out the effectiveness of the optimization of the similarity of the problem. Therefore, the EIoU loss function is introduced to accurately describe the problem of the difference between the real frame and the predicted frame width and length while calculating the Euclidean distance of the center point.EIoU is based on CIoU splitting the loss term of the aspect ratio into the difference between the predicted width and height and the width and height of the smallest external frame, and the EIoU is calculated as shown in Figure 2. Where A is the real frame, B is the predicted frame, c and c^{ct} are the centers of A and B respectively, d is the diagonal distance of the minimum outer rectangle of A and B, w and h are the width and length of A respectively, w^{ct} and h^{ct} are the width and length of B respectively, D_w and D_h are the width and length of the minimum outer rectangle formed by A and B, and the EIoU formula is:

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp}. \tag{6}$$

$$L_{EIoU} = L_{IoU} + \frac{\rho^2\left(c, c^{ct}\right)}{d^2} + \frac{\rho^2\left(w, w^{ct}\right)}{D_w^2} + \frac{\rho^2\left(h, h^{ct}\right)}{D_h^2}. \tag{7}$$

In the formula L_{dis} is the normalized Euclidean distance of the center point, assigning closer prediction boxes to

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

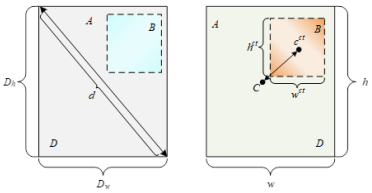

Figure 2: EIoU calculate

different target students to improve the positioning accuracy of the center point. L_{asp} is the edge length loss, using ρ is to calculate the Euclidean distance operation, separate the actual error of the encircling box width and length, which enhances the sensitivity of the model to the length and width of the prediction box, and further accelerates the convergence speed of the model, so in this paper, the localization loss is replaced by EIoU.

2) Introduction of Varifocal Loss Function

YOLOv5 network training process, the program will be each round of sub-model training output and the true value of the loss function, confidence and classification loss using binary cross-entropy loss function, the formula is:

$$\begin{aligned}&\text{BCE With Logits Loss}\,(p,y)\\&=-y\log(p)-(1-y)\log(1-p)\,.\end{aligned}\tag{8}$$

The target detection network generates dense anchor frames and matches the true frames to some of the prior frames; prior frames that match are positive samples, and those that do not match are negative samples. In most dense target detector training, the task appears to have too large a difference between the target category and the background category, Focal Loss adds a larger weight to harder-to-learn samples a and a smaller weight to easy-to-distinguish samples, increasing the weight of the hard-to-distinguish samples and decreasing the weight of the easy-to-distinguish samples, Focal Loss is defined as:

$$\begin{aligned}&\text{Focal Loss}\,(p,y,\alpha,\gamma)\\&=\begin{cases}-\alpha(1-p)^{\gamma}\log(p)\,, y=1\\-(1-\alpha)\,p^{\gamma}\log(1-p)\,, y=0\end{cases}.\end{aligned}\tag{9}$$

However, the positive samples are very few compared to the negative samples, it is more important to keep the positive samples and make full use of the supervisory signals of the positive samples, which is more conducive to the training of the model. When calculating the loss, the positive samples tend to master the direction of model training and contribute more in terms of LOSS, so Varifocal Loss proposes an asymmetric weighting operation with the formula:

$$\begin{aligned}&\text{VF Loss}\,(p_s,q)\\&=\begin{cases}-q\,(q\log(p_s)+(1-q)\log(1-p_s))\,, q>0\\-\alpha_t p_s^{\gamma}\log(1-p_s)\,, q=0\end{cases},\end{aligned}\tag{10}$$

where p_s is the predicted IOU perceptual classification score, q represents the target score, and for the foreground class, $q > 0$ represents the IOU score between the predicted and real frames generated by the positive examples, and for the background class, q is all zeros for all classes. γ is the attenuation factor controlling p_s, by which the attenuation contribution of the negative examples can be reduced by $p_s\gamma$, and the weight of the positive examples is kept unchanged, preserving the learning information of the positive examples and allowing the model to spend a small amount of money on centralized training higher quality samples and learn more useful information. Eiou and varifocal loss loss function are more suitable for model training than existing loss functions.

B. Evaluation of classroom concentration

The above section identifies students' learning behavioral states so as to analyze and study classroom concentration. In this section, the classification results of classroom behavioral states are mainly divided into categories according to the degree of positivity, based on which concentration is quantified by combining the fuzzy comprehensive evaluation method, and concentration scores and their concentration grades are finally derived.

Teachers in the teaching process teach content that is difficult or easy, and the teaching methods are also different. Therefore, in the course of class, students will have different emotional changes at different times. Based on the identification of classroom learning behaviors, seven basic expressions of students appearing in the classroom can be obtained. These seven expressions are categorized according to their degree of positivity, i.e., positive (surprised, happy), neutral (neutral), and negative (sad, disgusted, scared, angry).

For the more vague concept of classroom concentration, this paper will introduce a fuzzy comprehensive evaluation algorithm to evaluate students' facial expressions and behavioral outcomes from two aspects. The function model of fuzzy comprehensive evaluation usually contains three (U, V, R) or four elements (U, V, R, w), of which $U = \{u_1, u_2 \cdots\cdots u_n\}$ is the factor set, which refers to the factors that directly affect the evaluation object as the elements of the aggregate, where u_i represents the ith influence factor, n is the number of factors. For each of these evaluation factors can be regarded as a single evaluation factor, i.e., a level 1 indicator, under which a second level of evaluation factors, i.e., level 2 indicators, can be set. $V = \{v_1, v_2 \cdots\cdots v_n\}$ is the set consisting of the various evaluation results that the evaluator may make on the evaluation object. w is the weight, which is used to indicate

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

the importance of each evaluation factor. R is the fuzzy relationship matrix, which indicates the degree of affiliation of V that can be determined independently from a single factor starting from the evaluation of the evaluation object, also called single-factor fuzzy evaluation. After determining the hierarchical fuzzy subset, it is necessary to quantify the evaluated object on each evaluation factor u_i one by one, so as to determine the degree of affiliation of the evaluated object to each hierarchical fuzzy subset, and thus derive the fuzzy relationship matrix:

$$R = [r_{ij}], \tag{11}$$

where r_{ij} denotes the degree of affiliation of a certain evaluated object to the hierarchical fuzzy subset v_j from a single factor u_i.

In this paper the design process of fuzzy comprehensive assessment of students' classroom attentiveness from their behavioral state is as follows:

1) Identifying evaluation factors
In this paper, in the study of classroom concentration, the main purpose is to identify the students' expression and behavioral status, for three types of emotions (positive, neutral, negative) and two types of behaviors (positive, negative), there are six combinations of situations, in different combinations of situations the expression and behavior on the concentration of the degree of influence is different.
2) Determining the evaluation level
In this paper, the students' concentration situation in the course is categorized into concentration, more concentration and less concentration, and thus the evaluation set is set to these three kinds, which is identified by three letters V_1, V_2 and V_3, and the three concentration levels are assigned as S_1, S_2 and S_3.
3) Establishing the main factor subset
In this paper, we mainly analyze the classroom concentration from facial expression and behavioral state, and the main factor subsets are different in different combination cases, for example, in the first combination case, positive expression and positive behavior are set as the factor set, and defined as F_1, F_2, respectively, and weight W is set as w_1 and w_2, respectively, and the sum of their weights is 1, which is expressed as:

$$\sum_{i=1}^{n} w_i = 1 \, (w_i \geq 0) \, . \tag{12}$$

The rest of the combination cases build the main factor subset in the same way as above.
4) Establishing sub-factor sets
For the combination of different categories of expressions and behaviors, the establishment of the sub-factor set is not the same. If there are m secondary evaluation factors under the first level one indicator and n secondary evaluation factors under the second level one indicator, the sub-factor set of the first level one indicator is set to $F_{11}, F_{12}, \cdots, F_{1m}$, its weight $W1$ is set to $w_{11}, w_{12}, \cdots, w_{1m}$ respectively, and the sum of its weights is 1, which is expressed as:

$$\sum_{i=1}^{n} w_{is} = 1 \, (w_{is} \geq 0) \, . \tag{13}$$

The sub-factor set for the second level 1 indicator is set as $F_{21}, F_{22}, \cdots, F_{2n}$, and its weight $W2$ is set as $w_{21}, w_{22}, \cdots, w_{2n}$. In the first combination case, for example, the two sub-factor sets for positive emotions are determined as Happy (F_{11}) and Surprised (F_{12}), and their weights $W1$ are set as w_{11} and w_{12}, respectively, and similarly, the sub-factor set for positive behaviors is determined as Heads Up Listening (F_{21}), and its weight w_2 is set as w_{21}. The method of establishing the sub-factor sets for the rest of the combination cases is the same as above. Where n is the number of secondary indicators and i represents the ith primary indicator. The final classroom concentration evaluation indexes for the seven combinations are shown in Table 1, in which the seventh combination contains both negative and positive behaviors.

Combination situation	Primary indicator	Symbol	Weighting	Secondary in
1	Active emotion	F_1	w_1	Joyfuln
				Surpri
	Active behavior	F_2	w_2	Look up an
2	Active emotion	F_1	w_1	Joyfuln
				Surpri
	Negative behavior	F_2	w_2	Bow the
				Look in all d
3	Neutral emotion	F_1	w_1	Neutra
	Active behavior	F_2	w_2	Look up an
4	Neutral emotion	F_1	w_1	Neutra
	Negative behavior	F_2	w_2	Bow the
				Look in all d
5	Negative emotion	F_1	w_1	Revulsi
				Sad
				Fear
				Ange
	Active behavior	F_2	w_2	Look up an
6	Negative emotion	F_1	w_1	Revulsi
				Sad
				Fear
				Ange
	Negative behavior	F_2	w_2	Bow the
				Look in all d
7	Negative behavior	F_1	w_1	Bow the
				Look in all d
	Active behavior	F_1	w_1	Look up an

Table 1: The evaluation index of classroom concentration in seven combinations

5) Establishment of a single-factor evaluation matrix
The single-factor evaluation matrix established in different combinations is not the same. If there are m second-level evaluation factors affecting the first first-level evaluation factors and n second-level evaluation factors affecting the second first-level evaluation factors, the corresponding single-factor evaluation matrix is:

$$R11 = (r_{ij})_{m\times 3} \, (i = 1, 2, \cdots, m; j = 1, 2, 3) \, . \tag{14}$$

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

$$R12 = (r_{pq})_{n\times 3}\,(p = 1, 2, \cdots, n; q = 1, 2, 3)\,. \quad (15)$$

Taking the first combination case as an example, according to Table 1 on the determination of the factors affecting positive emotions and positive behaviors as well as the weights, which leads to the formation of a single-factor level evaluation matrix for positive emotions and positive behaviors, denoted as:

$$R11 = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{bmatrix} = (r_{ij})_{2\times 3}\,(i = 1, 2; j = 1, 2, 3)\,. \quad (16)$$

$$R12 = \begin{bmatrix} r_{11} & r_{12} & r_{13} \end{bmatrix} = (r_{pq})_{1\times 3}\,(p = 1; q = 1, 2, 3)\,. \quad (17)$$

6) Conducting fuzzy comprehensive evaluation

The evaluation model for fuzzy comprehensive evaluation is:

$$B = W * R = (b_1, b_2, \cdots\cdots, b_n)\,. \quad (18)$$

By analogy, the evaluation model for the two first level factors in this paper is expressed as:

$$B11 = W1 * R11 = [w_{11}, w_{12}, \cdots, w_m] * [r_j]_{m\times 3} \quad (i = 1, 2, \cdots, m; j = 1, 2, 3). \quad (19)$$

$$B12 = W2 * R12 = [w_{21}, w_{22}, \cdots, w_{2n}] * [r_{pq}]_{n\times 3} \quad (p = 1, 2, \cdots, nq = 1, 2, 3). \quad (20)$$

where $W1$ and $W2$ denote the weight matrices of the second level factors, respectively.

Taking the first combination case as an example, the evaluation model of positive expression and positive behavior is obtained as:

$$B11 = W1 * R11 = [w_{11}, w_{12}] * \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{bmatrix}. \quad (21)$$

$$B12 = W2 * R12 = [w_{21}] * [r_{11}, r_{12}, r_{13}]\,. \quad (22)$$

The evaluation models of the two Level W factors were fused to obtain the final evaluation model of classroom attentiveness, with 1 being the weight matrix of the Level 1 factors, denoted as:

$$D = W * B = \begin{bmatrix} w_1 & w_2 \end{bmatrix} * \begin{bmatrix} B11 \\ B12 \end{bmatrix}. \quad (23)$$

The student's concentration score was calculated based on the given thresholds S_1, S_2, and S_3:

$$Y = D * S = D * \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix}. \quad (24)$$

The overall concentration score for the class was calculated as:

$$Score = \frac{\sum_{i=1}^{n} Y_i}{n}, \quad (25)$$

where n represents the number of students in the class.

III. Empirical Analysis of English Language Education in Higher Education

A. Analysis of student behavior

1) Analysis of model training effects

In order to verify the performance of the student behavior recognition algorithm based on YOLOv5 proposed in this paper, it is used to identify the behavior of individual learners. The study obtained a total of 800 classroom videos of English education classes in the year of 2023 from the public service platform through screen recording and crawler, and 138 qualified classroom video screens were screened to obtain 1,645 clips in the front viewpoint, 4,684 clips in the diagonal upper viewpoint, 4,315 clips in the rear viewpoint, and 2,682 clips in the teacher's viewpoint through lens slicing. In this study, the clips in the front viewpoint are selected as the data source for later data labeling and model training. In this section, based on the observation of English education classroom recorded video and scholars' research related to classroom student behavior, a set of student behavior classification indexes based on English education classroom is designed to classify student behavior into nine kinds of behaviors, including individual behavior and team behavior. Among them, individual behavior is the behavior that students do alone, and there are 7 kinds of behaviors: writing with the head down, reading with the head down, listening with the head up, raising the hand to speak, answering while standing up, turning the head to look at others, and abnormal behavior (yawning, sleeping, and walking around). Team behaviors are behaviors that students need to perform in cooperation with others, and there are two types: group discussion and teacher guidance.

Considering the balance of student behavior recognition accuracy and detection time, 400 rounds of training using YOLOv5s model were chosen. The dataset was randomly selected 75% as the training set and 25% as the test set. The main hardware equipment for this study includes i5-10400 (CPU), 32G RAM, and GT1030 (GPU). In terms of software, the programming language used in this experiment is Python 3.8 and Pytorch deep learning framework is used. The confusion matrix is a specific matrix used to present a visualization of the performance of deep learning algorithms. Each of its columns represents the predicted values and each row represents the actual categories, and the confusion matrix for student behavior recognition in the classroom is shown in Table 2. The bolded values on the diagonal represent the accuracy of the model predictions, and the average recognition accuracy is 0.86875.The student behavior recognition models all achieved relatively accurate results, and the accuracy of the two behaviors of head turning (0.70) and hand raising (0.79) recognition is relatively low. The possible reason is that in the front view, the students' eyes will follow the teacher's movement, and for behavior recognition, it is impossible to distinguish whether they are turning their heads or looking up to listen to the lesson, and similar movements such as the students' resting their cheeks and scratching their heads are misrecognized as raising their hands. The student behavior recognition algorithm based on YOLOv5 proposed in this

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

Dimension	Bend down	Look down	Look up	Turning head	Hand up	Standing	Panel discuss-ion	Teacher guidance	Back-ground FP
Bend down	**0.93**	0.06		0.02	1	0.02	0.545 0.01	0.234	0.21 0.005
Look down	0.04	**0.88**	0.03	0.03	0.04	0.01	0.433 0.01	0.183	0.30 0.001
Look up		0.01	**0.91**	0.16	0.07	0.01	0.37 0.01	0.104	0.35 0.006
Turning head		0.01	0.03	**0.70**	0.04		0.35	**0.442**	0.05 0.002
Hand up				0.03	**0.79**	0.02	0.425	0.251	0.02 0.001
Standing					0.02	**0.91**	0.568	0.201	0.03 0
Panel discuss-ion					7		0.618 **0.91**	0.152	0.02 0.001
Teacher guidance					8		0.542	0.392	0.002
Back-ground FP	0.01	0.02	0.01	0.04	0.03	0.01	0.615 0.04	0.162	0.001
					10		0.599	0.075	0.009
					11		0.624	0.104	0.001
					12		0.505	0.134	0.006
					13		0.644	0.058	0.005
					14		0.645	0.091	0.006
					15		0.47	0.193	0
					16		**0.738**	0.012	0
					17		0.726	0.065	0
					18		0.508	0.104	0.001
					19		0.591	0.005	0.012
					20		0.617	0.014	0.002

Table 2: The confusion matrix of the behavior recognition of students in the classroom

Table 3: The distribution of the behavior of 20 students in class

paper has an accuracy of 0.98 for student recognition, which is good and can analyze student learning behaviors in the English classroom more accurately in order to improve the efficiency of student management.

2) Overall analysis of student behavior in the classroom

In this subsection, the video of the English education classroom recording of the first unit of the eighth grade book in the X College Teaching Platform is selected for analysis, and the number of realistic students in the video is 20. The size of the video is 780 pixels $\times$ 460 pixels, and one frame of image is extracted every 2 seconds for the video screen, and a total of 851 effective images are obtained. Students have the main status in the classroom, not only need to analyze the behavioral status of different classrooms as a whole, but also need to monitor and analyze the learning status of each learner, so that teachers can choose the appropriate teaching methods according to the behavioral status of the students and realize targeted teaching. The overall analysis of classroom student behavior is to analyze the behavioral distribution status of 20 students in a classroom, and Table 3 shows the distribution of the behavioral status of 20 students in the classroom.

To analyze team behaviors as a whole, no teacher-directed behaviors occurred in this class, only four students (serial numbers 2, 3, 4, and 7) participated in the group discussion, indicating that it was a private discussion and exchange among the four students, which can be seen that the teacher did not organize a formal group discussion activity in this class. From the overall analysis of individual behavior, the two behaviors of looking down (0.557) and looking up (0.256) accounted for most of the time in the classroom, of which 18 people looked down at the books with a ratio of more than 0.4, and 8 people with a ratio of more than 0.6. 16 people looked up to listen to the class with a ratio of more than 0.2, and 5 people with a ratio of more than 0.3. From the two behaviors of raising their hands to speak up and standing to answer, which best reflect the degree of participation in the classroom, 16 people raised their hands to speak up and 16 people stood to answer the questions. Answer two behaviors, 16 students participated in the hand-raising behavior in this classroom, which accounted for four-fifths of the total number of students in the classroom. Due to the fact that students' standing up behavior at the end of the class accounted for about 0.005 of the distribution of behaviors, 18 people stood up to answer the teacher's questions after the correction.

From the distribution of the overall behavioral state of students in the classroom, this classroom is a typical mixed classroom teaching, classroom head down reading and head up listening to two kinds of behavior accounted for most of the classroom teaching time, there is no formal group discussion and the teacher teaching activities occur, the classroom teacher-student interaction is mainly carried out through the teacher question student answers. From the students' individual level analysis of personal behavior serial number 16 and 17 students, the proportion of the two behaviors of head-up listening and head-down reading are 0.974 and 0.908 respectively, indicating that students 16 and 17 do not have a high degree of classroom participation and are not highly engaged in learning. As for the eight students No. 1-5, 8, 12 and 15, the proportion of the three behaviors of reading with head down, writing with head down and listening with head up are relatively evenly distributed, among which the proportion of the behavior of writing with head down of the students No. 1, 4 and 5 is more than 0.2, which indicates that these eight students have a high degree of learning engagement. In summary, the behavior recognition algorithm based on YOLOv5 proposed in this paper can well identify the eight behaviors of students' head-up listening, head-down writing, head-down reading, hand-raising, standing, group discussion, and teacher's guidance, which provides a way of thinking about how to monitor the learners' learning status in the English education classroom and student management.

B. Analysis of classroom concentration evaluation results

This section focuses on analyzing and evaluating the concentration status of individual students, all students during the time with the whole class. By displaying the students' concentration scores line graphs to analyze the students' concentration in the classroom, and to judge the teacher's

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

teaching situation, and to make comments and suggestions on the teacher's teaching and students' learning.

1) Individual student concentration evaluation

By analyzing the concentration situation of individual students, it is possible to investigate the development of individual students' concentration relative to the reasons for their existence, for this reason, this paper collects the learning video of the students of the eighth grade C class in the teaching platform of the X university as experimental data, and selects the learning situation of one of the students R to draw a line graph of his concentration for specific example analysis. The length of the video acquisition is 15 minutes, and 29846 pictures can be obtained by decomposing the video, and 229892 valid pictures are obtained by classifying them through image recognition. In order to ensure that the line graph formed by the acquired data information is more obvious, and also to ensure that the distribution of its concentration is more evenly distributed, this paper carries out the division of numerical thresholds for the four levels of classroom concentration, and the score thresholds corresponding to the four concentration evaluation sets are 100, 75, 50, and 25. For the weighting of the indexes all the time, the paper finds that classroom action recognition is more important than classroom action recognition based on the interviews with teachers and actual observation of classroom teaching. Based on the interviews with teachers and the actual observation of classroom teaching, it can be found that the classroom action recognition can better show the learners' concentration in the classroom. According to the fuzzy synthesized matrix for relevant calculations can be obtained every 15 seconds of the student's concentration score, the video will be 15s as a time interval, drawing the students' classroom concentration curve as shown in Figure 3.

According to the classroom concentration score of student R, we can see that the concentration score of the students in the whole class is $78.073 > 75$, which means that the students' state in the whole class is close to concentration. In Figure 3, we can see that the students' concentration declined rapidly in the first 120s of the learning period, which shows that the students may have shifted their attention or looked around during this period, which led to the poor concentration without listening carefully to the lesson. In this situation of inattention in the last 2 minutes after the attention gradually rise and in the next period of time in a more stable learning state, in the following 8 minutes, although there are some ups and downs, but the ups and downs state is not very big, only in the 7 minutes and 15 seconds there is a sudden drop, and then in the 7 minutes and 30 seconds when the minutes to climb rapidly, so it can be seen that the student in the following ten minutes of listening to the lesson is still The student is still able to return to the class in a timely manner, even though his attention is sometimes shifted by other factors. This is also relatively consistent with our concentration curve, in which there is a lack of concentration in the early stages of learning, but with the development of time, their concentration will improve. According to the line graph of students R's concentration, it can be judged that the teacher's teaching content and teaching methods in the classroom are in a better state, more suitable for students R's learning needs, able to capture students R's attention, and promote students R's learning well, but the teacher's introduction at the beginning is not ideal. At the same time, students can correct their problems of not being able to concentrate in time at the beginning of the class, and try to adjust themselves as quickly as possible in the first few minutes of learning, so that they can enter the learning state in time.

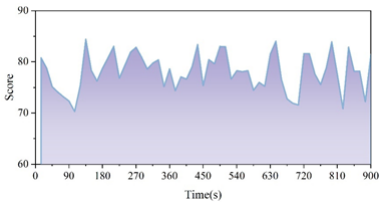

Figure 3: Students' class concentration curve

2) Evaluation of concentration of all students

Similarly, through the corresponding analysis of the concentration situation of all students, we can judge the change of the concentration situation of all students, so as to analyze the reasons for the change of students' concentration, and at the same time, through the relevant judgment of the data, we can evaluate the teaching situation and teaching methods of teachers. Therefore, by taking the students' classroom learning video as the experimental data, we can obtain the concentration level of all students and draw the concentration level line graph for the analysis of teachers and students. The length of the video acquisition is about 20 minutes, through the video decomposition to obtain each student's learning behavior is about 25646 pictures, so as to carry out the analysis of learning behavior, concentration analysis to 15 seconds for the unit of concentration achievement calculation, and finally the whole class of all the students' concentration situation for calculation, drawing concentration curve as shown in Figure 4, with 30s as the time interval. The concentration score of the whole class for all students is 78.274, which is higher than 75, so all students can basically maintain a state of concentrated listening in this time interval. At the beginning of the first minute, there is a rising trend in students' concentration, followed by a slow decline, it can be seen that at the beginning of the students' learning attention is not completely focused on the classroom, it is more likely to be interfered by other external factors and lead to memory distraction, but after 2 minutes, students' concentration then rises and stays stable, which shows that students in the middle of the learning period, they are able to focus on the classroom content, but after 10 minutes, there is a rise in the concentration of all students. However, after 10 minutes, there is a trend of sudden rise and fall of concentration, which is very obvious. It can be seen that students' concentration will inevitably decline after concentrating for a period of time, while at the end of the lesson,

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

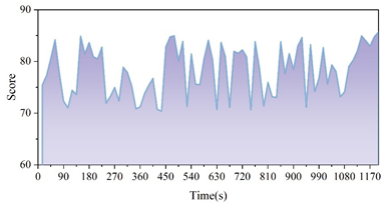

Figure 4: Concentration of all students

students' concentration has a rapid climb and stabilized, which can be seen that students are more excited at the end of the lesson so that they can concentrate on the lesson. According to the concentration curve of all students, it can be seen that the teacher's teaching method and teaching level are good, which can attract students' attention and make them basically in a state of concentration. However, due to the teacher's introduction is not good, so that the students in the beginning of the attention can not enter the class in a timely manner, this point is the teacher needs to improve, but also timely remind the students as soon as possible to enter the learning state, will focus on the class. Because the middle section of classroom teaching belongs to the area where students' concentration is highly concentrated, the more important knowledge points can be introduced in this period of classroom teaching to let students learn and explore better, and students can study or review their weak knowledge points independently in this period. In the latter part of the classroom, teachers need to slow down the attenuation of students' concentration through some specific reminders and warnings, so that students' concentration can be maintained in a more focused situation.

IV. Strategies for Coordinated Development of English Language Education and Student Management in Higher Education

The coordination of higher vocational English education and student management should start from insisting on openness and guidance, cultivating the correct concepts and literacy of teachers and students from the perspective of comprehensive service, accurately grasping the connotation and characteristics of the innovation and development of higher vocational student management, striving for the expansion of the new space of English education, and constantly enhancing the vitality and adaptability of the work of students, and endeavoring to build a new era of integration of education, management and service. The pattern of "education management". In this section, based on the analysis of students' classroom behavior and evaluation of concentration above, the following strategies for the coordinated development of higher vocational English education and student management are proposed:

1) Focus on improving the talent cultivation function of student management, and create a new model of "balanced + interactive" English education. The foundation of higher education lies in the establishment of morality and humanism, which is carried out throughout the whole process of talent cultivation in colleges and universities, and requires a balanced and interactive student management and English education model to promote the scientific development of talent cultivation, and to provide a good field to support the construction of an educational ecology of equality, respect, harmony and sharing under the premise of emphasizing the return of inter-subjectivity. The school can function based on yolov5's student behavior recognition algorithm in the classroom multimedia equipment, in the course of English teaching, can give the student the state of the student class in real time through the teaching system feedback to the teachers, the teacher and the student state is targeted management.
2) Strengthen the implementation of school management. In order to better promote the coordinated development of student management and English education in colleges and universities, colleges and universities must strengthen the implementation of the management system, to do in strict accordance with the system to manage the school, so as to ensure that all aspects of the school in a reasonable and orderly manner, but also to provide a scientific model for the practice of English language education, which will enable students to be strict from the reality, and continue to improve their overall quality. The school will also provide a model of scientific practice for English education, so that students can strictly demand themselves from reality and continuously improve their comprehensive quality.
3) Promote the scientific nature of the establishment of student management system in colleges and universities. Colleges and universities in the construction of the management system, must be from the practical point of view, the construction of the system should be derived from the actual, more to be applied to the actual. If the system established is not in line with the actual development, then it is difficult to ensure the implementation of the system, which will make some of the imperfections of the system utilized, and it is easy to cause adverse effects on students. Therefore, promoting the scientific nature of the establishment of student management system in colleges and universities can advance the development of English education.

V. Conclusion

Realizing the coordinated development of English education and student management is an important initiative to promote the healthy and sustainable development of higher education. This paper constructs a student behavior recognition model based on deep learning, and takes the actual video of the English education classroom in X university as an example for student behavior analysis and classroom concentration evaluation, and the results show that:

1) The YOLOv5-based student behavior recognition model has a recognition accuracy of 0.98 for student behavior, which can analyze student behavior in English classroom more precisely. It is also found that two

individual behaviors of students looking down (0.557) and looking up (0.256) in the classroom account for most of the time in the classroom, which indicates a low level of classroom participation and a low level of learning engagement. Teachers can improve their English education methods according to the students' learning detection status so as to improve the efficiency of students' management.

2) The concentration scores of individual students and all students in the whole class are 78.073 and 78.274 respectively, which are both greater than 75, so individual students and all students can basically maintain a state of concentration on listening to lectures in this period of time. According to the results of classroom concentration evaluation, we can judge the teaching situation of teachers and make comments and suggestions on teachers' teaching and students' learning.

In conclusion, based on the results of the identification and analysis of deep learning on the teaching behavior of higher vocational English, the optimization strategy of student management is proposed to improve English education, and research ideas are provided to achieve the coordinated development of higher vocational English education and student management.

Funding

1) This article is a part research finding of the 2022 "Jiangsu Province Social Science Application Research Excellent Project in Foreign Language": Study on the Value-added Evaluation Mechanism of English Core Literacy for Vocational College Students from the Perspective of Curriculum Ideology and Politics (22SWC-58).
2) This article is a part research finding of the 2023 "Jiangsu Province Social Science Application Research Excellent Project in Foreign Language Project": Research on the Path of Digital Literacy Cultivation for Students in Higher Vocational English Teaching (23SWC –06).

References

[1] Xu, Y., Yu, J., & Buehrer, R. M. (2020). The application of deep reinforcement learning to distributed spectrum access in dynamic heterogeneous environments with partial observations. *IEEE Transactions on Wireless Communications, 19*(7), 4494-4506.
[2] Cho, Y., & Kim, J. (2021). Production of mobile english language teaching application based on text interface using deep learning. *Electronics, 10*(15), 1809.
[3] Ramachandram, D., & Taylor, G. W. (2017). Deep multimodal learning: a survey on recent advances and trends. *IEEE Signal Processing Magazine, 34*(6), 96-108.
[4] Guo, W., Tian, W., Ye, Y., Xu, L., & Wu, K. (2020). Cloud resource scheduling with deep reinforcement learning and imitation learning. *IEEE Internet of Things Journal, 8*(5), 3576-3586.
[5] Alwasiti, H., Yusoff, M. Z., & Raza, K. (2020). Motor imagery classification for brain computer interface using deep metric learning. *IEEE Access, 8,* 109949-109963.
[6] Xiangmin, L. (2019). Characteristics and rules of college english education based on cognitive process simulation. *Cognitive Systems Research, 57*(OCT.), 11-19.
[7] David, Stevens, Karen, & Lowing. (2018). Observer, observed and observations: initial teacher education english tutors' feedback on lessons taught by student teachers of english. *English in Education, 42*(2), 182-198.
[8] Yujie, N. (2017). Study on talent training mode of exhibition english courses in higher vocational colleges. Journal of Higher Education.
[9] Li, P. The Application of Multimodal Teaching Model Based on VAR Model in English Teaching in Colleges and Universities. *Applied Mathematics and Nonlinear Sciences,* 9(1).
[10] Taub, G. E., Sivo, S. A., & Puyana, O. E. (2017, July). Group differences between English and Spanish speakers' reading fluency growth in bilingual immersion education. In *School Psychology Forum, Research in Practice* (Vol. 11, No. 2, pp. 45-51). National Association of School Psychologists.
[11] Iqbal, A. (2021). Innovation speed and quality in higher education institutions: the role of knowledge management enablers and knowledge sharing process. *Journal of Knowledge Management, 25*(9), 2334-2360.
[12] Gundsambuu, S. (2019). Internationalization of Higher Education and English Medium Instruction in Mongolia: Initiatives and Trends. *Educational Studies,* (1 (eng)), 215-243.
[13] Zhang, Y. (2019). Exploration of student management mechanism in higher vocational colleges under the coupling model of institutionalization and humanization. *Basic & clinical pharmacology & toxicology,* (S2), 125.
[14] Zhang, F., & She, M. (2021). Design of english reading and learning management system in college education based on artificial intelligence. *Journal of Intelligent and Fuzzy Systems, 5*(2), 1-10.
[15] Zhang, X., & Lin, D. (2019). Exploration on the innovation of education management in colleges and universities in the big data era. *Basic & clinical pharmacology & toxicology,*(S9), 125.
[16] Wang, D., Su, J., & Yu, H. (2020). Feature extraction and analysis of natural language processing for deep learning english language. *IEEE Access, 8*, 46335-46345.
[17] Zhang, T. (2022). Deep learning classification model for English translation styles introducing attention mechanism. *Mathematical Problems in Engineering, 2022*(1), 6798505.
[18] Zhang, G. (2020). A study of grammar analysis in English teaching with deep learning algorithm. *International Journal of Emerging Technologies in Learning (iJET), 15*(18), 20-30.
[19] Korzekwa, D., & Kostek, B. (2019). Deep learning model for automated assessment of lexical stress of non-native english speakers. *The Journal of the Acoustical Society of America, 146*(4), 2956-2957.
[20] Qin, F. (2022). College english intelligent writing score system based on big data analysis and deep learning algorithm. Journal of Database Management.
[21] Wu, F., Chen, Y., & Han, D. (2022). Development countermeasures of college english education based on deep learning and artificial intelligence. *Mobile Information Systems, 2022*(1), 8389800.
[22] Wang, X. (2017). The framework of the multi-parameter evaluation index system for college spoken english based on deep learning theory. *Revista de la Facultad de Ingenieria, 32*(15), 583-590.