Journal of Nonlinear and Convex Analysis
Volume 26, Number 6, 2025, 1861–1877

Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

COGNITIVE PATTERN KNOWLEDGE-MATCHING METHOD BASED ON DEEP SELF-ATTENTION TRANSFORMATION FOR PHYSICAL EDUCATION

GUOBIAO YANG, WANYING YANG, AND YONGXIANG WANG*

Abstract. Teaching students according to their aptitude is an ancient and effective educational principle that emphasizes tailoring the content and approach of education to each student's unique abilities and interests. Using artificial intelligence technology, sports majors students can be analyzed through big data to learn habits, athletic abilities, and interests, provide them with personalized learning resources and paths, and conduct a real-time and accurate evaluation of their learning effects as feedback data so that teachers can adjust teaching strategies in time to ensure that the teaching content matches the learning needs of students. However, artificial intelligence technology faces large technical barriers among professional sports teachers, and a complete and mature technical system needs to be formed before it can be used in sports teaching. Given the above needs, this paper researches the cognitive mode-knowledge pattern-matching method based on a deep self-attention transformation twin network. Firstly, the VARK learning style measurement theory was used to determine the cognitive mode of students' sports majors. Using transformers, the content of images, voices, and short videos on the public teaching platform was used to establish the reasoning process between multimodal learning resources and students' cognitive patterns. The deep self-attention transformation twin network was constructed to match students' cognitive and knowledge patterns accurately. Finally, taking a basic sports course as a typical case, the accuracy and practicability of the matching method proposed by the team were verified. The verification results show that the matching method can achieve high matching accuracy among college students.

1. Introduction

In teaching activities, students, as the main learning body, have significant individual differences, mainly reflected in cognitive styles, intellectual types, and learning methods. Numerous research findings indicate that students with different cognitive patterns exhibit different preferences and learning efficiency in learning [11, 12]. For example, field-independent learners prefer independent and loosely

2020 *Mathematics Subject Classification.* 62H30, 68T10.

Key words and phrases. Depth self-attention transformation, cognitive model, knowledge pattern, twinning network, matching method.

This study was supported by Regular Scientific Research Project of Shaanxi Provincial Sport Bureau (2023012), the "14th Five-Year Pla" education scientific planning project of 2023 annual topic of Shaanxi Province (SGH23Y2271) and the 2025 Graduate Education and Teaching Reform Research Project of Xidian University (JGYB2509).

Key Research and Development Program of Shaanxi (Program No. 2025SF-YBXM-146).

*Corresponding author.

structured teaching, while field-dependent learners prefer tightly structured teaching. Choosing teaching strategies based on students' cognitive patterns can significantly improve teaching effectiveness. For field-independent learners, problem-driven and self-exploration teaching methods can be adopted; For field-dependent learners, more structured guidance and feedback can be used [10]. On this basis, sports are closely interconnected and not isolated from disciplines such as science and technology. The participation of students' sports majors not only enriches the diversity of educational activities but also facilitates the integration of sports knowledge with that of science and engineering, promoting interdisciplinary crossover and fusion.

Combining science and engineering universities with students majoring in sports can bring diversity in professional backgrounds to educational activities. Science and engineering students typically possess solid theoretical foundations in science and logical thinking abilities. In contrast, students' sports majors boast practical skills, excellent physical fitness, and superb athletic abilities, contrasting the two. This diversity in professional backgrounds not only enriches the content of educational activities but also aids in exploring the different responses of various knowledge structures and ability systems to experimental tasks or intervention measures. Especially under certain experimental conditions, such as physical fitness tests and sports recovery, students' sports majors may exhibit distinct physiological and psychological reactions compared to other students, providing more extensive and in-depth data for experiments.

Students' sports major research findings often have high practical application value. They can directly apply experimental results to actual training and competitions, enhancing athletic performance and achievements, particularly in sports science, sports injury prevention and rehabilitation, and physical conditioning. This potential for practical application elevates the professional expertise of students' sports majors and offers science and engineering students opportunities to transform theoretical knowledge into practical applications.

Meanwhile, science and engineering universities usually possess relatively complete experimental facilities and research conditions, providing robust support for the smooth conduct of experiments. Due to the needs of their professional training, students' sports majors adhere well to experimental rules and operational procedures, ensuring the accuracy and reliability of experimental data. This feasibility of experimental control lays a solid foundation for the smooth progress of educational activities.

To achieve a match between cognitive patterns and teaching methods, especially in the field of physical education, educators can further collect and analyze a large amount of data from sports majors' students' learning, training, and competition processes, deeply understand their cognitive patterns and learning characteristics, match their cognitive patterns with knowledge patterns, construct teaching content targeted, adopt heuristic education methods suitable for their cognitive patterns, face the cognitive reality of learners, and deliver educational resources to learners in the form of graded teaching, thereby improving the timeliness, interactivity, and self-learning ability of teaching. However, this work is difficult for manpower to complete.

The rapid development of artificial intelligence technology has injected innovative concepts and models into modern teaching and provided new technologies and means for constructing students' cognitive and knowledge patterns [4, 16, 27]. In recent years, the attention mechanism has garnered extensive attention, and its parallel computing function and inter-feature dependency construction ability have achieved good results in natural language processing tasks. In the context of event detection tasks, Wu et al. [22] used the attention mechanism to dynamically determine the amount of information drawn from word or character-level embeddings, thereby enhancing the model's understanding of textual semantic features. Ding R et al. [5] addressed the challenge of identifying relationships between parameters and events within the text by proposing multiple attention layers to extract intra-sentence relationships, facilitating the extraction of deeper semantic information. Wang Xianxian et al. [19] proposed the AttIndRNNCapsNet model for the default of event elements in the Viennese language by introducing the intrinsic properties of events and event elements into an independent recurrent neural network combined with the attention mechanism to obtain higher-level semantic features of the text. Wang et al. [21] proposed using a syntactic dependency graph to construct a graph neural network to solve the ambiguity problem of monolingual words and fully extract the information between words, and at the same time, enhance the node information of the syntactic dependency graph by combining with the attention mechanism, to enhance the model's ability to recognize event triggers. Ahmad et al. [1] combined the advantages of GCN and attention and proposed a Graph Attention Transformer Encoder (GATE), which learns structured contextual representations using a self-attention mechanism, allowing it to capture long-distance dependencies and apply them to different types of languages.

Herzig et al. mapped visual features and GLoVe word vectors to a common semantic space through a semantic transformation module and concatenated the features of the two modalities. Then, they used a graph self-attention module to assign attention scores to the concatenated features to fuse and update node representations and finally classified them to obtain scene maps [8, 18]. Similarly, Mi et al. concatenated the word vectors corresponding to visual features and object categories and updated node features through object-level and relationship-level graph attention networks, respectively. Finally, the scene map was classified [17]. Kiros et al. processed object labels and relationship labels into word vectors using language embedding models and injected them as language information into the features of visual nodes. They used a module similar to a Transformer encoder as a message propagation method to integrate language information while propagating visual information [9]. However, these methods need the ability to mine multimodal contextual information between and within images and text. These methods either concatenate image and text features or cannot fully utilize cross-modal contextual information [15]. Either the need for more consideration for contextual information between word vectors within the text modality leads to suboptimal text features. Due to the need for the ability to perceive multimodal contextual information, these scene graph generation methods often obtain suboptimal representations of multimodal contextual features, which affects the effectiveness of scene graph generation. The features of different modalities often interact during the forward propagation process of the model, thereby learning the semantic correlation between entities in different modalities. Lee et al. used pre-trained object detectors to extract visual features of salient regions in images. They designed a cross-modal attention mechanism to achieve interaction between region and word features, thereby using local alignment of the two modalities to infer the similarity between images and texts [13]. Wang et al. encoded the geometric position of the region based on Lee et al. to obtain a graph text-matching model that is sensitive to the position of image content [20]. Jiuxiang et al. modeled salient regions in images and words in natural language descriptions as visual and textual images using cross-modal attention mechanisms and graph convolutional neural networks for intra-modal interaction [7]. Before calculating the degree of image text matching, this method utilizes various cross-modal information exchange mechanisms to interact with the features of two different modalities to discover and utilize the complementary information between modalities and better align the two modalities' local/global semantic information. Therefore, compared to the previous method, this method has better retrieval performance. However, due to the introduction of modal interaction processes, such methods' training time and matching score calculation time will be longer. To match student cognitive patterns with knowledge patterns, it is necessary to utilize the dynamic changes in short-term and long-term interests in student learning and life. The educational process must be based on the student's cognitive level, from low to high, and from conceptual ambiguity to clarity. A multimodal mapping between video, text, and phonetic values should be established to describe videos through cognitive pattern keywords, visualize knowledge pattern types, and achieve effective reasoning in the learning process [23, 26].

This article is based on personalized education and utilizes the Transformer deep self-attention transformation twin network in response to the above needs. By matching the cognitive model of students with the knowledge patterns of learning resources, a cognitive knowledge pattern matching method based on the deep self-attention transformation twin network is proposed to support the diverse needs of personalized cognition among students' sports majors.

2. Construction of student cognitive and knowledge patterns

2.1. Construction of Student Cognitive Patterns.

The construction of student cognitive patterns is based on the acquisition of multidimensional and multidimensional data from students, abstracting their multidimensional features, and constructing a student cognitive model by assigning "labels" to present student portraits [3, 6, 25]. This article constructs a cognitive model for students from the following three dimensions: basic information, learning ability analysis, and learning style model.

(1) Basic information about students: Explicit data can be obtained directly from the school's open data platform, including individual information such as student name, gender, age, ethnicity, class, sports specialty, physical data, interests, and other personal information; At the same time, social relationship characteristics between students can be extracted from the already obtained big data, and all students in the class form classmate relationships. In learning group activities, dynamic same-group relationships are also formed.

(2) Learning ability analysis: Students' learning ability is the most important information that teachers, students, schools, and society pay attention to, and it is also the key to determining whether they can successfully match their knowledge patterns. However, conventional scores and other information make it difficult to analyze the cognitive differences and learning process data between students. More scientific and modern detection methods are needed to conduct cognitive analysis on students and comprehensively obtain their knowledge and ability levels.

(3) Learning style model: Learning style refers to the personal learning habits and preferences gradually formed by students in the learning environment and sports process, with relative stability. It is an important characteristic indicator in personalized education. The currently recognized learning style models are the Felder Silverman model, the Kolb model, the VARK model, etc.

This paper uses the VARK Learning Style Scale to test and conduct a questionnaire survey and basic learning tests on 50 students' sports majors, measuring their learning foundations, learning styles, and learning preferences in school. The cognitive patterns of students, including visual, auditory, read-write, and kinesthetic, were obtained [14]. Among them, visual students have strong observation and imagination, are good at learning through visual charts, are good at using image symbols to connect concepts, concretize abstract things, have strong perception and understanding of colors, lines, and charts, have a strong memory of images, images, and scenes, and can quickly pay attention to details. For students with this type of cognition, using images to promote memory and convert boring textual data into graphical models for learning is suitable. Read-write students have strong reading and summarizing abilities and can better understand textual knowledge concepts; Auditory students have strong listening abilities and are adept at learning through auditory reception of information; Kinesthetic students have strong action and practical abilities and are skilled in learning through hands-on participation in practice.

2.2. **Construction of Knowledge Pattern Material Library.** This article uses a deep self-attention transformation to construct the correlation between visual features and knowledge patterns of multimodal data, as shown in Fig. 1.

Establishing knowledge pattern keywords can achieve the mapping relationship between semantic information of knowledge concepts within sports and deep transformation network models. Users input different descriptions of the course's knowledge concepts (such as easily understandable course content, more in-depth but difficult-to-understand course content, etc.) into the semantic reasoning module and then match them with the knowledge pattern keywords in the knowledge pattern material library [24].

In addition, a feedback mechanism is added to address the situation where the semantic information provided by users concerning certain sports terminology or specific training methods cannot find a corresponding knowledge pattern. This mechanism allows users to add the desired knowledge pattern to the list of keywords in a certain knowledge pattern in the database to associate the newly added keywords with a specific sports knowledge pattern.

As the output of this section, the classification of sports knowledge patterns as a strong constraint is crucial for matching subsequent student cognition with

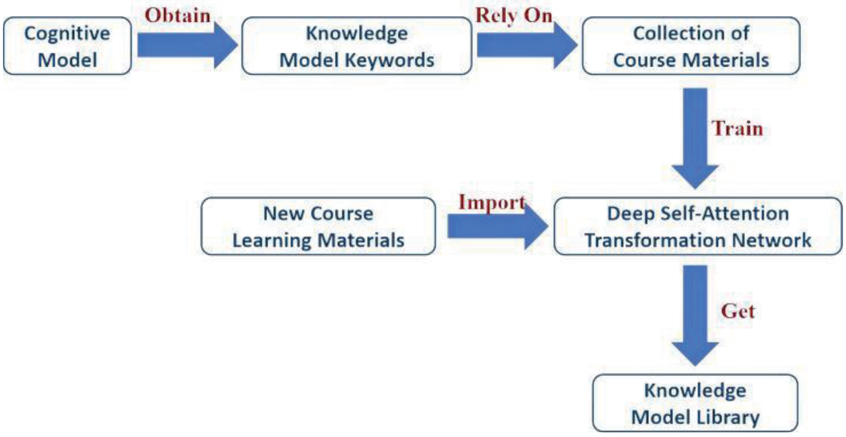

Cognitive Model
Obtain
Knowledge Model Keywords
Rely On
Collection of Course Materials
Train
New Course Learning Materials
Import
Deep Self-Attention Transformation Network
Get
Knowledge Model Library

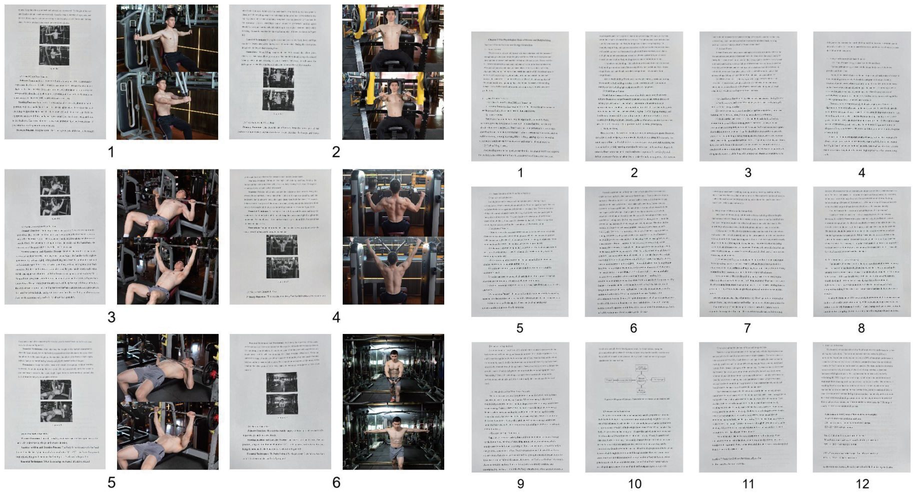
FIGURE 1. Construction process diagram of knowledge pattern material library

multimodal sports learning materials and the interpretability of problem tracing. Then, the VARK learning style measurement method is used to extract cognitive attribute features of students with specific sports needs, including information on their knowledge level, learning style, emotional state, etc., to reflect their cognitive state comprehensively. Subsequently, the extracted cognitive attribute features will be combined with historical cognitive classification features to reflect better the long-term cognitive state and changing trends of students. This can further optimize the modeling of cognitive states by utilizing students' past learning behaviors and performances, and the resulting student cognitive patterns will be used as keywords in the sports knowledge pattern material library. Next, many sports learning materials will be collected and input into the deep self-attention transformation twin network. Use many sports course content learning materials to train the neural network and obtain a knowledge pattern material library corresponding to each cognitive mode. When new sports course learning materials are generated or collected, they are input into the network, and the deep self-attention transformation twin network is used to match and classify them into the corresponding sports knowledge pattern material library.

In summary, in response to the needs of students' sports majors and combining many learning resources on open platforms, a knowledge pattern material library based on the basic sports course has been constructed. Two types of learning resources for students with different cognitive modes have been collected, namely visual and read-write knowledge pattern material libraries. This library will also serve as the library for experimental verification of the matching method in this article.

3. CONSTRUCTION OF DEEP SELF-ATTENTION TRANSFORM TWIN NEURAL NETWORKS

3.1. Network Structure of Cognitive and Knowledge Pattern Matching. Convolutional neural networks occupy a central position in computer vision tasks



(a) (b)

Figure 2. Local image of knowledge pattern material library: (a) visual type; (b) read-write type

such as image recognition by their powerful spatial feature extraction capability. Its convolutional and pooling layers are well-designed to efficiently capture spatial information and local features such as edges and textures in images. Meanwhile, CNN effectively reduces the number of model parameters, lowers computational cost, and improves training speed and model generalization ability through parameter sharing and sparse connectivity. Compared with traditional fully connected neural networks, CNNs are more efficient in processing images, especially on large-scale datasets. CNNs perform well in image classification and can be applied to various computer vision tasks, such as target detection, semantic segmentation, and video analysis. Therefore, CNNs have become the preferred solution in the field of image recognition and are widely used in various computer vision tasks.

Therefore, we build a knowledge pattern material library based on deep neural networks and convolutional neural networks. Firstly, the model utilizes deep neural networks and convolutional neural networks to collect a large amount of sports course knowledge materials and trains the network to correspond the collected sports course materials with the knowledge pattern keywords in the knowledge pattern keyword index table, obtaining a knowledge pattern material library about this cognitive pattern. Then, based on the constraints of student cognitive encoding and knowledge pattern classification encoding as input features, a good connection and matching relationship can be established between student cognitive state and knowledge patterns. Finally, a classification output is obtained using the Decoder module of the Transformer for decoding to determine whether the input course content materials match various knowledge patterns. This output can help teachers better create teaching materials, understand students' cognitive states and learning needs, and provide personalized teaching and guidance. The process is shown in Figure 3.

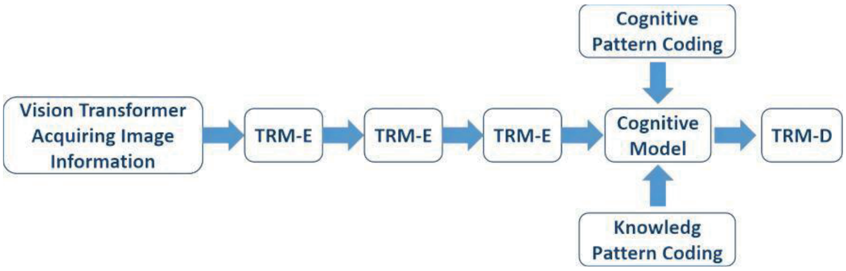


FIGURE 3. Network structure of cognitive pattern knowledge pattern matching

By utilizing deep self-attention transformation, the knowledge pattern features in the material library are matched with the knowledge pattern keywords obtained through student cognitive patterns to construct a knowledge pattern material library. The Transformer twin network model obtains a weighted feature vector through the "Self-attention" module, which focuses on the long-term learning, life, entertainment, and other information of students in school. It is then passed on to the network's selection mechanism and encoding encoder module, known as the Feed Forward Neural Network. The system integrates and maps information consistent with their cognitive patterns through multi-feature fusion and filtering, influenced by students' cognitive patterns. This process transforms the input into embedded Query, Key, and Value vectors. Using the Encoder-Decoder Attention module in the network, the matching content (Value) is obtained based on the similarity between Query and Key. The relationship between the current student's cognitive and encoded knowledge feature vectors is outputted to determine the knowledge pattern of the concept. Complete the mapping from student cognitive patterns to student knowledge pattern matching and use it as input and knowledge pattern constraint for subsequent intelligent, personalized retrieval.

3.2. Construction of Deep Self-Attention Transform Twin Neural Network Vision Transformer.

(1) Input images from the knowledge material library into the Transformer

Firstly, to input an image into the Transformer, each image pixel can be flattened into a sequence, which can then be input into the Transformer. However, this can cause excessive computational complexity, such as for a 224×224 size image, which requires input of 50176 embedding vectors. As the number of pixels increases, the computational complexity increases by a square level. Therefore, ViT splits the image into patches individually as input to the Transformer. For example, if the input image size is (224224,3) and each patch size is 16×16, it will result in $14 \times 14 = 196$ patches. Therefore, the number of this input is acceptable. Compared to CNN, which can only perform correlation analysis on adjacent elements, Transformer can calculate and consider the correlation of global elements, which is also the advantage of Transformer.

(2) The Structure Construction of Vision Transformer

Firstly, the divided patches are transformed into embedding vectors through a linear mapping layer. Then, a class token representing the class is added at the starting position of these embedding vectors, and position embedding is added to each embedding vector. These vectors are then input into the Transformer Encoder for calculation. Finally, the MLP multi-head attention block outputs the classification result, as shown in Figure 4. How does the size of the input vector change in the linear mapping layer. Firstly, the linear mapping layer is implemented through a convolutional layer consisting of 768 convolutional kernels with a size of 16 × 16 and a step size of 16. After the image with a size of (224224,3) is input into the network, it is transformed into a tensor with a size of (197768) through the convolutional layer. Then, after adding a category vector, the tensor size is (197768). Finally, position encoding and tensor size are added (197768). The structures of the Encoder Block and MLP Block for the Transformer are shown in Figure 4.

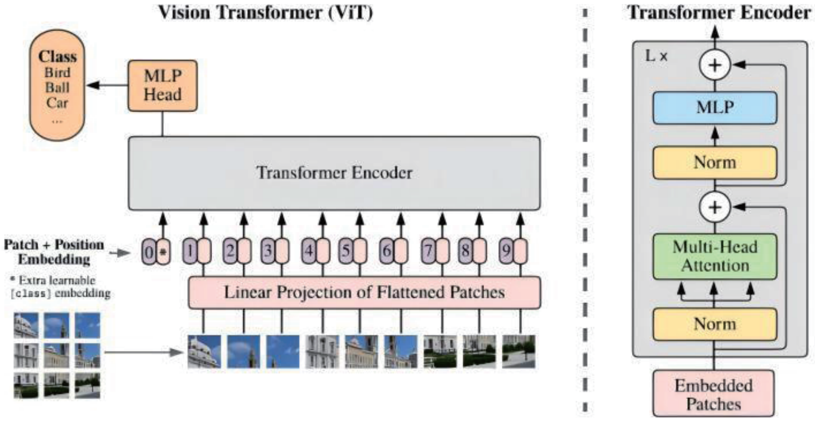


Figure 4. Structure of Vision Transformer

Firstly, the processed tensor is input into the Encoder Block, which undergoes a layer normalization module and then enters the multi-head attention module to calculate self-attention. Then, it is dropped out and connected through a residual before layer normalization. This processing method can optimize the accuracy of the input vector information. Finally, an MLP Block is used, and another Dropout is performed to obtain the output classification result. After inputting the tensor into the MLP Block, a linear, fully connected layer is first used to increase the number of channels by four times. Then, the GELU activation function is used, and after a Drop Out, a fully connected layer is used to restore the number of channels. Finally, the output result is obtained.

3.3. **Construction of Deep Self-attention Transform Twin Neural Networks.** Firstly, the Patch Embed module is established, in which initialization parameters are defined. The image size is (224224), and the size of each patch is (16,16). RGB images are passed in with three channels, and the ViT base model is used. The dimension of the embedded vector passed in is 768, and layer normalization is not performed by default. After defining the image size and patch size, continue to define the number of patches and the number and size of convolution kernels. Next is the forward propagation process. As the ViT model requires a fixed input image size, an error will be reported if the input image size does not match the required size. Next, the image will be flattened by flattening its height and width, swapping dimensions 1 and 2, and finally, normalizing it to convert it into an embedding vector and input it into the self-attention module. This module is a module for processing input images, which converts them into vector sequences and serves as input for subsequent modules.

Regarding the Multi-head Self attention module in Vision Transformer, first define the initialization parameters, where the dimension is the dimension of the embedding vector output by PatchEmbed, with an attention headcount of 8 and no bias of q, k, and v. Next, calculate the dimension of each headcount, and input q, k, and v into a linear fully connected layer, where the dimension will be tripled. Then, perform Dropout, pass through a linear, fully connected layer, and Dropout again. Then comes the forward propagation stage, where the parameters passed in are the number of processed images, the number of patches plus one, and the dimension of the embedding vector. Next, the q, k, and v vectors are obtained through matrix operations specific to each head. Each head has its own q, k, and v vectors and performs multi-head self-attention transformation. Next, each row of the obtained result is processed with softmax, and the result is weighted and summed with matrix V. After processing the matrix, the results of each head are concatenated to obtain a global attention result. Finally, the output is obtained through fully connected and dropout layers. In this module, the most core Multi-head Self-attention operation in the Transformer was mainly performed, calculating the correlation between different patches in the input image and obtaining the Q, K, and V matrices for different patches based on global considerations, as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3.1}$$

The MLP module first defines initialization parameters, the number of input nodes, the number of hidden layer nodes, and the activation function GELU. Then, through the first fully connected layer, the input is the number of input nodes, and the output is the number of hidden layer nodes. After processing the data with the activation function, the output is obtained through the fully connected layer. The dimension of this output is the same as the dimension of the input Multi-head Self-attention module, which restores the dimension of the vector. The main function of this module is to perform image classification processing, using a multi-layer perceptron to apply the parameters obtained from Multi-head self-attention to multiple fully connected layers to achieve a prediction of the final result.

The Transformer Encoder Block module first defines initialization parameters, including dimensions, number of headers, q, k, v vectors, activation function GELU, and dropout parameters. Then, it defines the first layer normalization function, uses the Multi-head Self-attention module, defines the second layer normalization function, calculates the vector dimension passed in the MLP module, and finally inputs the obtained parameters into the MLP module for the forward propagation stage.

Finally, there is the Vision Transformer module, which first defines initialization parameters and then calls the PatchEmbed module, Multihead Self-attention module, MLP module, and Transformer Encoder Block module to complete the classification of the input images.

4. Verification of cognitive pattern knowledge pattern matching methods

Based on the knowledge pattern material library and deep self-attention transformation twin neural network constructed in the previous text, this chapter will take the basic sports course as an example and use the knowledge images of concepts in the established knowledge pattern material library to test the established twin network, experimentally verifying the accuracy and feasibility of the proposed matching method.

4.1. **Selection of research subjects.**

Firstly, this article selects two types of students with significant differences in cognitive patterns. The first type is students with strong receptivity, learning ability, and a preference for exploring conceptual knowledge. Their cognitive patterns are defined as read-write. Due to their ability to quickly understand new knowledge, we hope to match more in-depth learning resources to this group of students. Therefore, we hope to match course materials with more conceptual textual content in the content of images to this group of students so that they can accurately understand the fundamental principles of learning this knowledge from the perspective of conceptual principles and also better and faster enable them to learn the content they want to learn; The second type is students who have weak acceptance ability and are not easy to understand obscure and difficult to understand knowledge points. Their cognitive mode is generally visual, and they need help to accept new knowledge during the learning process, making it difficult to understand the knowledge points of principle concepts. Therefore, we aim to match some materials with illustrations to help these students visualize and concretize abstract principle concepts to understand better how this knowledge point is derived or how the concept is applied in practical life. This can help them better and more deeply understand this knowledge point.

4.2. **Experimental Program Design.**

A test experiment was designed and implemented to investigate the accuracy and effectiveness of personalized learning resources designed for students with two cognitive modes: read-write and visual. The experiment was conducted with a sample of second-year students' sports majors from a university, focusing on two groups of students with significant differences in cognitive modes, and 30 students with stable grades and good mental health were selected as the test subjects to ensure the reliability of the results. Subsequently, we randomly subdivided these students into two groups: the test group and the control group. Throughout the experimental cycle, we strictly controlled all variables other than the systematic input information, such as video browsing, voice input, social input, etc., to ensure the accuracy of the experimental results, and conducted the test and recorded the results in the same period every other day for subsequent data analysis.

After four weeks of systematic testing and data collection, we used statistical methods to analyze the data in a comprehensive and in-depth manner. The results are shown in Table 1. Seven typical technical solutions were selected for comparative verification to verify the superiority of the technical solution proposed in this paper: SCAN, BFAN, DPRNN, CAAN, GSMN, SHAN, and SGRAF. Based on the student experimental samples provided in this paper, the core indicators of the read-write and visual cognitive patterns were calculated, respectively. Then, the data obtained using the technical solution proposed in this paper were compared with the results obtained by the seven typical technical solutions. As shown in Table 1, the technical solution DSAT proposed in this paper exhibits superior recognition and matching effects in terms of R@1, R@5, and R@10 for both cognitive patterns, verifying the technical advancement of the proposed DSAT scheme.

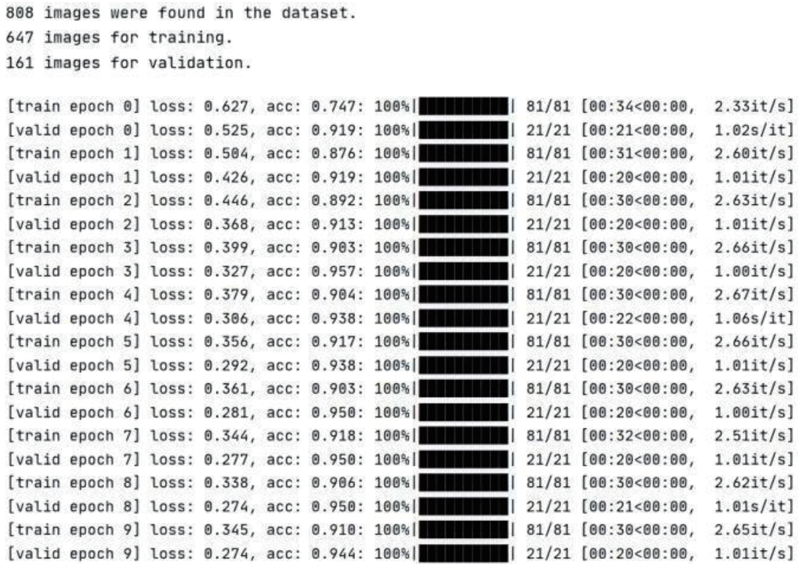
Table 1. Data processing results

Method	Read-Write Cognitive Pattern			Visual Cognitive Pattern			rsum
	R@1	R@5	R@10	R@1	R@5	R@10	
SCAN*	67.2	90.1	95.7	48.1	77.6	85.5	464.2
BFAN*	68.4	91.6	-	50.7	78.1	-	-
DPRNN	70.3	91.3	95.5	55.3	81.5	88.1	482.0
CAAN	70.6	91.5	97.0	52.2	79.3	87.7	478.3
GSMN*	76.0	94.2	97.2	57.9	82.0	89.3	496.6
SHAN*	74.4	93.8	96.8	55.6	81.7	88.2	490.5
SGRAF*	77.9	94.4	97.3	58.3	83.4	88.6	499.9
DSAT*	**77.3**	**94.6**	**97.5**	**60.4**	**84.6**	**90.1**	**504.5**

4.3. **Twin Network Training.** Firstly, input the content related to the course of the basic sports from the prepared knowledge pattern material library into the training network of the deep self-attention transformation twin network. The training process is shown in Figure 5. Eight hundred-eight photos were used to train the network, with 647 photos as the training set and 161 photos as the test set. Ten epochs were set to obtain the training parameters that needed to be learned.

After obtaining the trained parameters, the course material content can be matched with the corresponding cognitive students. As shown in Figure 6, first select images with more conceptual text to input into the network, hoping that the network can match them to students with read-write cognitive patterns. The following matching results are obtained after inputting the image into the prediction network. After inputting the image with more conceptual text into the system, it is recognized and classified by the Transformer's multi-head attention mechanism. The corresponding category displayed is the read-write type student cognitive pattern, with a probability above 83%, indicating good matching results.

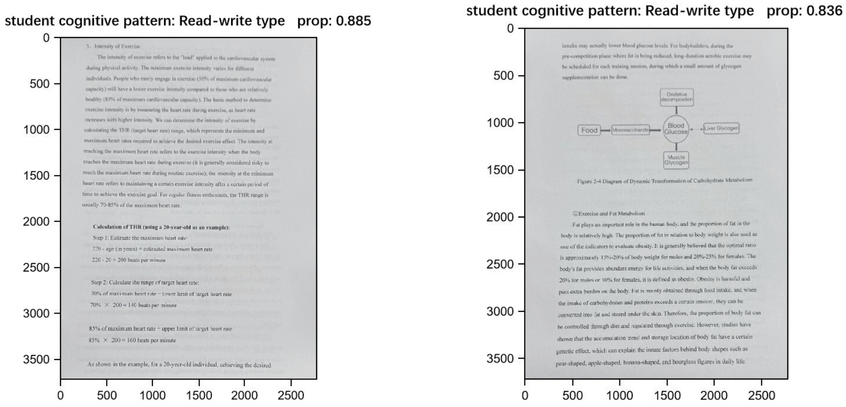
Next, we input a learning resource with a lot of image content, hoping that the twin network can match it to students with visual cognitive patterns. After inputting the image into the network, the result is shown in the following figure. The output result shows that the corresponding student's cognitive pattern category is a visual cognitive pattern, with a matching degree of over 80%. We have also successfully matched the course content to students with suitable cognitive pattern types, as shown in Figure 7.



```
808 images were found in the dataset.
647 images for training.
161 images for validation.

[train epoch 0] loss: 0.627, acc: 0.747: 100%|██████████| 81/81 [00:34<00:00,  2.33it/s]
[valid epoch 0] loss: 0.525, acc: 0.919: 100%|██████████| 21/21 [00:21<00:00,  1.02s/it]
[train epoch 1] loss: 0.504, acc: 0.876: 100%|██████████| 81/81 [00:31<00:00,  2.60it/s]
[valid epoch 1] loss: 0.426, acc: 0.919: 100%|██████████| 21/21 [00:20<00:00,  1.01it/s]
[train epoch 2] loss: 0.446, acc: 0.892: 100%|██████████| 81/81 [00:30<00:00,  2.63it/s]
[valid epoch 2] loss: 0.368, acc: 0.913: 100%|██████████| 21/21 [00:20<00:00,  1.01it/s]
[train epoch 3] loss: 0.399, acc: 0.903: 100%|██████████| 81/81 [00:30<00:00,  2.66it/s]
[valid epoch 3] loss: 0.327, acc: 0.957: 100%|██████████| 21/21 [00:20<00:00,  1.00it/s]
[train epoch 4] loss: 0.379, acc: 0.904: 100%|██████████| 81/81 [00:30<00:00,  2.67it/s]
[valid epoch 4] loss: 0.306, acc: 0.938: 100%|██████████| 21/21 [00:22<00:00,  1.06s/it]
[train epoch 5] loss: 0.356, acc: 0.917: 100%|██████████| 81/81 [00:30<00:00,  2.66it/s]
[valid epoch 5] loss: 0.292, acc: 0.938: 100%|██████████| 21/21 [00:20<00:00,  1.01it/s]
[train epoch 6] loss: 0.361, acc: 0.903: 100%|██████████| 81/81 [00:30<00:00,  2.63it/s]
[valid epoch 6] loss: 0.281, acc: 0.950: 100%|██████████| 21/21 [00:20<00:00,  1.00it/s]
[train epoch 7] loss: 0.344, acc: 0.918: 100%|██████████| 81/81 [00:32<00:00,  2.51it/s]
[valid epoch 7] loss: 0.277, acc: 0.950: 100%|██████████| 21/21 [00:20<00:00,  1.01it/s]
[train epoch 8] loss: 0.338, acc: 0.906: 100%|██████████| 81/81 [00:30<00:00,  2.62it/s]
[valid epoch 8] loss: 0.274, acc: 0.950: 100%|██████████| 21/21 [00:21<00:00,  1.01s/it]
[train epoch 9] loss: 0.345, acc: 0.910: 100%|██████████| 81/81 [00:30<00:00,  2.65it/s]
[valid epoch 9] loss: 0.274, acc: 0.944: 100%|██████████| 21/21 [00:20<00:00,  1.01it/s]
```

FIGURE 5. Twin network training process



(a)

(b)

FIGURE 6. Read-write cognitive pattern matching test results

Through the above experiments, we found that the deep self-attention transformation twin neural network excelled in completing the task of matching the learning content of different basic sports courses to students with corresponding cognitive patterns. The experimental results indicate that the cognitive knowledge pattern matching method based on deep self-attention transformation twin neural network constructed in this paper has certain feasibility and effectiveness.

4.4. **Model Performance Evaluation.** To validate the superiority of deep self-attention transformation twin neural network in mitigating the interference of noisy features in the matching process, the validation was carried out on the Flicker30k dataset, where DSAT is the abbreviation of the methodology in this chapter, R@1,

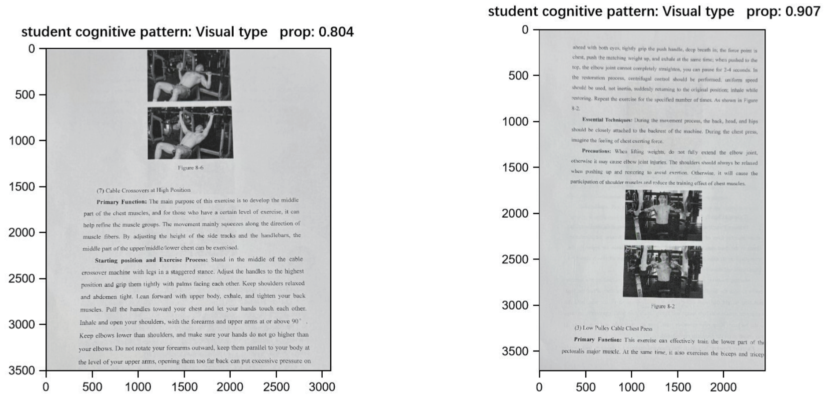

student cognitive pattern: Visual type prop: 0.804
0
500
1000
1500
2000
2500
3000
3500
0
500
1000
1500
2000
2500
3000
student cognitive pattern: Visual type prop: 0.907
0
500
1000
1500
2000
2500
3000
3500
0
500
1000
1500
2000

Figure 8-6

(7) Cable Crossovers at High Position

Primary Function: The main purpose of this exercise is to develop the middle part of the chest muscles, and for those who have a certain level of exercise, it can help refine the muscle groups. The movement mainly squeezes along the direction of muscle fibers. By adjusting the height of the side tracks and the handlebars, the middle part of the upper/middle/lower chest can be exercised.

Starting position and Exercise Process: Stand in the middle of the cable crossover machine with legs in a staggered stance. Adjust the handles to the highest position and grip them tightly with palms facing each other. Keep shoulders relaxed and abdomen tight. Lean forward with upper body, exhale, and tighten your back muscles. Pull the handles toward your chest and let your hands touch each other. Inhale and open your shoulders, with the forearms and upper arms at or above 90° . Keep elbows lower than shoulders, and make sure your hands do not go higher than your elbows. Do not rotate your forearms outward, keep them parallel to your body at the level of your upper arms, opening them too far back can put excessive pressure on

(a)

ahead with both eyes, tightly grip the push handle, deep breath in; the force point is chest, push the matching weight up, and exhale at the same time; when pushed to the top, the elbow joint cannot completely straighten, you can pause for 2-4 seconds. In the restoration process, centrifugal control should be performed, uniform speed should be used, not inertia, suddenly returning to the original position; inhale while restoring. Repeat the exercise for the specified number of times. As shown in Figure 8-2.

Essential Techniques: During the movement process, the back, head, and hips should be closely attached to the backrest of the machine. During the chest press, imagine the feeling of chest exerting force.

Precautions: When lifting weights, do not fully extend the elbow joint, otherwise it may cause elbow joint injuries. The shoulders should always be relaxed when pushing up and restoring to avoid exertion. Otherwise, it will cause the participation of shoulder muscles and reduce the training effect of chest muscles.

Figure 8-2

(3) Low Pulley Cable Chest Press

Primary Function: This exercise can effectively train the lower part of the pectoralis major muscle. At the same time, it also exercises the biceps and tricep

(b)

Figure 7. Visual cognitive pattern matching test results

R@5 and R@10 in the table are Recall@k, which is subdivided into two task settings of read-write cognitive pattern and visual cognitive pattern, and contains a total of six metrics data. The *rsum* in the last column of the table is the sum of the recall of the first six columns, which gives a more comprehensive picture of the performance of the matching model.

Table 2. Results of data matching

Method	Read-Write Cognitive Pattern			Visual Cognitive Pattern			*rsum*
	R@1	R@5	R@10	R@1	R@5	R@10	
MSCOCO 1k test							
SCAN*	72.3	94.1	98.8	58.4	88.2	94.4	506.2
BFAN*	74.7	95.5	-	59.6	88.5	-	-
DPRNN	75.6	95.8	98.5	62.1	89.3	95.3	516.6
CAAN	75.4	95.3	98.2	61.2	89.1	95.6	514.8
GSMN*	78.0	96.2	98.6	63.6	90.0	95.2	521.6
SHAN*	76.2	96.7	98.7	62.5	89.5	95.8	519.4
SGRAF*	79.8	96.4	98.1	63.9	90.3	96.1	524.6
DSAT*	**79.9**	**96.6**	**98.7**	**64.7**	**91.4**	**95.6**	**526.9**
MSCOCO 5k test							
SCAN*	50.3	82.2	90.2	38.7	69.0	80.9	411.3
CAAN	52.7	83.6	90.4	41.4	70.2	82.3	420.6
SGRAF*	58.1	84.9	92.1	41.6	70.7	81.6	429.0
DSAT*	**60.4**	**86.5**	**92.3**	**41.8**	**72.4**	**81.9**	**435.3**

The results of the comparison experiments are shown in Table 2. Compared with the previous methods, the DSAT proposed in this paper can get some improvement in each metric. In the task setting of read-write cognitive pattern, compared with the baseline model SGRAF, the DSAT model can achieve a good improvement in R@1, R@5, and R@10 by 2.0%, 0.2%, and 0.3%, respectively, with the most important metric, R@1, achieving a great improvement. In the task setting of visual cognitive pattern, compared with the baseline model SGRAF, the DSAT model was able to achieve a great improvement in each metric, with the most important metric R@1 improving by 1.6%, whereas R@5 improved by 1.5% and R@10 by 1.4%.

Finally, the composite metric rsum shows that DSAT significantly improves by 7% compared to SGRAF. DSAT also achieves a more significant performance improvement on Flicker30k, which fully proves the importance of the DSAT model in mitigating noise feature interference in the matching process.

5. Conclusion

Starting from the concept of "teaching according to individual needs", this article utilizes artificial intelligence and information technology to empower personalized education deeply. Firstly, the knowledge patterns of sports majors' student cognitive patterns and learning resources were constructed, and the process inference between multi-model learning resources and student cognitive patterns was completed. Then, the Transformer deep self-attention transformation twin network matched the student's cognitive patterns and knowledge patterns, obtaining learning resources suitable for the student's cognitive patterns. Finally, the feasibility and accuracy of the matching method proposed in this article were experimentally verified, with a matching accuracy of up to 80%.

References

[1] W. U. Ahmad, N. Peng and K. W. Chang, *GATE: Graph attention transformer encoder for cross-iingual relation and event extraction*, in: AAAI Conf. Artif. Intell., AAAI, Association for the Advancement of Artificial Intelligence, May. 2021, pp. 12462–12470.

[2] K. Arulkumaran, M. P. Deisenroth, M. Brundage and A. A. Bharath, *Deep reinforcement learning: A brief survey*, IEEE Signal Processing Magazine **34** (2017), 26–38.

[3] J. Bernard, T. Chang, E. Popescu and S. Graf, *Learning style Identifier: Improving the precision of learning style identification through computational intelligence algorithms*, Expert Systems with Applications **75** (2017), 94–108.

[4] L. Deng, *Artificial intelligence in the rising wave of deep learning: The historical path and future outlook [perspectives]*, IEEE Signal Processing Magazine **35** (2018), 180–177.

[5] R. Ding and Z. Li, *Event extraction with deep contextualized word representation and multi-attention layer*, in: Lect. Notes Comput. Sci., Springer-Verlag, Dec. 2018, pp. 189–201.

[6] P. Grifoni, A. D'Ulizia and F. Ferri, *When language evolution meets multimodality: Current status and challenges toward multimodal computational models*, IEEE Access **9** (2021), 35196–35206.

[7] J. Gu, J. Cai, S. Joty, L. Niu and G. Wang, *Look, imagine and match: improving textual-visual cross-modal retrieval with generative models*, in: Proc IEEE Comput Soc Conf Comput Vision Pattern Recognit, IEEE Computer Society, Nov. 2017, pp. 7181–7189.

[8] R. Herzig, M. Raboh, G. Chechik, J. Berant and A. Globerson, *Mapping images to scene graphs with permutation-invariant structured prediction*, in: Adv. neural inf. proces. syst., Neural information processing systems foundation, Dec. 2018, pp. 7211–7221.

[9] R. Kiros, R. Salakhutdinov and R. S. Zemel, *Unifying visual-semantic embeddings with multimodal neural language models*, ArXiv, **abs/1411.2539** (2014): 13.

[10] B. M. Lake, T. D. Ullman, J. B. Tenenbaum and S. J. Gershman, *Building machines that learn and think like people*, Behavioral and Brain Sciences, **40** (2017): e253.

[11] M. Lang, *Learning styles and on-line learning analytics: An analysis of student behaviour based on the honey and mumford model*, in: Lect. Notes Comput. Sci., Springer Science and Business Media Deutschland GmbH, Nov. 2023, pp. 154–166.

[12] H. Y. Lee, Y. P. Cheng, W. S. Wang, C. J. Lin and Y. M. Huang, *Exploring the learning process and effectiveness of STEM education via learning behavior analysis and the interactive-constructive-active-passive framework*, Journal of Educational Computing Research **61** (2023), 951–976.

[13] K. H. Lee, X. Chen, G. Hua, H. Hu and X. He, *Stacked cross attention for image-text matching*, in: Lect. Notes Comput. Sci., Springer Verlag, Mar. 2018, pp. 212–228.

[14] W. L. Leite, M. D. Svinicki and Y. Shi, *Attempted validation of the scores of the VARK: Learning styles inventory with multitrait-multimethod confirmatory factor analysis models*, Educational and Psychological Measurement **70** (2010), 323–339.

[15] K. Li, Z. Wu, K. C. Peng, J. Ernst and Y. Fu, *Guided attention inference network*, IEEE Transactions on Pattern Analysis and Machine Intelligence **42** (2020), 2996–3010.

[16] W. M. Lim, A. Gunasekara, J. L. Pallant, J. I. Pallant and E. Pechenkina, *Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators*, The International Journal of Management Education **21** (2023): 100790.

[17] L. Mi and Z. Chen, *Hierarchical graph attention network for visual relationship detection*, in: Proc IEEE Comput Soc Conf Comput Vision Pattern Recognit, IEEE Computer Society, Jun. 2020, pp. 13883–13892.

[18] J. Pennington, R. Socher and C. D. Manning, *GloVe: Global vectors for word representation*, in: EMNLP - Conf. Empir. Methods Nat. Lang. Process., Proc. Conf., Association for Computational Linguistics (ACL), Oct. 2014, pp. 1532–1543.

[19] X. Wang, L. Yu, S. Tian and R. Wang, *Missing argument filling of Uyghur event based on independent recurrent neural network and capsule network*, Acta Automatica Sinica **47** (2021), 903–912.

[20] Y. Wang, H. Yang, X. Qian, L. Ma, J. Lu and X. Fan, *Position focused attention network for image-text matching*, in: IJCAI Int. Joint Conf. Artif. Intell., International Joint Conferences on Artificial Intelligence, Aug. 2019, pp. 3792–3798.

[21] Z. Wang, B. Li and Y. Wang, *Event detection based on multilingual information enhanced syntactic dependency GCN*, in: Lect. Notes Comput. Sci., Springer Science and Business Media Deutschland GmbH, Aug. 2022, pp. 378–390.

[22] Y. Wu and J. Zhang, *Chinese event extraction based on attention and semantic features: A bidirectional circular neural network*, Future Internet **10** (2018): 95.

[23] C. Zhang, Z. Yang, X. He and L. Deng, *Multimodal intelligence: representation learning*, Information Fusion, and Applications, IEEE Journal on Selected Topics in Signal Processing **14** (2020), 478–493.

[24] S. Zhang, N. Hui, P. Zhai, J. Xu, L. Cao and Q. Wang, *A fine-grained and multi-context-aware learning path recommendation model over knowledge graphs for online learning communities*, Information Processing and Management **60** (2023): 103464.

[25] Y. Zhang, Y. Zhao, Y. Dong and B. Du, *Self-supervised pretraining via multimodality images with transformer for change detection*, IEEE Transactions on Geoscience and Remote Sensing **61** (2023), 1–11.

[26] Z. Zhao, P. Xu, C. Scheidegger and L. Ren, *Human-in-the-loop extraction of interpretable concepts in deep learning models*, IEEE Transactions on Visualization and Computer Graphics **28** (2022), 780–790.

[27] J. Zheng, Q. Zhang, S. Xu, H. Peng and Q. Wu, *Cognition-based context-aware cloud computing for intelligent robotic systems in mobile education*, IEEE Access **6** (2018), 49103–49111.

Manuscript received July 1, 2024
revised January 3, 2025

G. Yang
School of Marxism, Xi'an Jiaotong University, Xi'an, China;
Department of Physical Education, Xidian University, Xi'an, China
E-mail address: gbyang@xidian.edu.cn

W. Yang
Department of Physical Education, Xidian University, Xi'an, China
E-mail address: 23221215234@stu.xidian.edu.cn

Y. Wang
School of Marxism, Xi'an Jiaotong University, Xi'an, China
E-mail address: yxwang2015@xjtu.edu.cn