Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 7501765, 11 pages
https://doi.org/10.1155/2022/7501765

Research Article

Optimal Design of Language Teaching Strategy System Based on Smart Video Mode

Jia Liu and Jianling Guo

Department of Foundational Disciplines, Shijiazhuang People's Medical College, Shijiazhuang, Hebei, China

Correspondence should be addressed to Jianling Guo; 2007026@muc.edu.cn

Received 4 May 2022; Revised 6 June 2022; Accepted 16 June 2022; Published 28 June 2022

Academic Editor: Jun Ye

The accuracy of video and goal enables students to learn and strengthen their ability constantly. Strengthening our country's study initiative degree can promote student study more effectively. As a new teaching method, students can not only obtain the basic knowledge, learning priorities, and difficulties needed for learning through video, but also understand the content of the text, the content of the article, and even cultivate students' interest in many related languages, such as writing, text, sound, image, color, and video, which can be displayed, clarified, and displayed intuitively, creating a free and relaxed learning environment, an interesting background teaching process, encouraging students to experience emotion, including physical experience, and being open and open. Establishing a complete and comprehensive ideological channel to further improve students' acceptance of information is helpful for students' analysis and training, understanding, and evaluation. Therefore, this paper first identifies video and excavates the intrinsic value of video application. This can provide technical and methodological support for the design of video teaching system.

1. Introduction

The accuracy of video and goals enables students to continuously learn and strengthen their abilities [1]. The concept of video learning was first proposed in the mid-1990s, and Carnegie Mellon University took the lead in introducing the concept of video learning into a research project, which was reorganized, understood and promoted, and approved [2]. At the same time, more and more educators begin to pay attention to video learning and vigorously promote the application of the concept of video learning to practical education and reform traditional teaching methods, which leads to modern video learning models, such as distance learning and university education [3]. Because video learning is very useful and convenient, many scholars and educators have taken a very positive attitude towards video learning. At the same time, our country has kept pace with the times and has begun to vigorously promote the use of video learning in the national education model [4]. The Ministry of Education has taken educational initiatives, such as the theory and practice of video teaching and video teaching. In addition, the state also attaches importance to video teaching in colleges and universities. Many of the best educational and training enterprises in China, such as the new Oriental educational institutions and elite online education, have developed their own video learning platforms to enhance the competitiveness of educational and training institutions in the new information age [5]. Reform education, such as Chinese teaching, is usually carried out in the form of text materials, with abstract and complex characteristics; the application of traditional teaching methods can easily lead to the low interest, motivation, and interest. For Chinese learning, as a new form of teaching, students can not only acquire the basic knowledge needed for learning through videos, learning priorities, and challenges, but also understand the content of the text, the content of the article, including stimulating people's interest in writing, writing abstracts, through words, sounds, images, colors, and videos, clarify and vividly display the teaching content, create a free and relaxed learning environment, a learning process background, vision, and fun, encourage students to carry out emotional experiments, open up comprehensive and comprehensive channels of thought, and further improve students' acceptance of information. It is helpful for students to carry out analysis training and improve students' understanding and evaluation of teaching [6]. In this paper, starting from the visual recognition of video, the intrinsic value of video use is excavated to provide technical and methodological support for the design of video education system.

2. Related Works

About video teaching, the literature suggests that the University of Nottingham (UK) has done an excellent job in information exchange, and user teachers can access the My Art Space platform. And their understanding of the subject or their views on difficult issues are recorded to the My Art Space platform, so that students in need can access it. Meanwhile, students, as the main body of the My Art Space platform, can also bring their learning experience to the platform, and students can communicate on the My Art Space platform on specific subjects. The literature suggests that Harvard University (USA) attaches great importance to the information obtained from students about school education programs. Therefore, the Department of Educational Science at Harvard University specializes in studying the experience and views of each student on the professional courses offered and proposing Hudl projects. Through systematic analysis data on the 8.344 million, more course-compliant staff training programs collected information on the learning of different professional students in different school curricula. The literature of South Dakota University in the United States has invested in the design of a student self-study service platform to facilitate the preservation of learning materials and the search for learning resources, which is a good example of the application of video learning technology abroad. According to the literature, the typical practice of video teaching technology university is adopted in vocational training. In practice, medical school students often face particularly difficult situations. In order to provide relevant expertise or to respond appropriately to these difficult situations, the University of Oslo, Norway, has designed a manual called Know mobile to effectively address the learning and work problems of medical school students. Documentation presents the current situation in Australia, where the Development Group has expanded the concept of video learning, not just to train specialized technicians through online video learning platforms, but also creatively incorporated the concept of evaluation and evaluation into video learning as a means of monitoring the effectiveness of students' learning practices and outcomes. Document Birmingham University (UK) designed a video learning project called Handler to solve some practical problems, such as the concept of lifelong learning, which is deeply rooted in today's society, but students in different age groups often have different expectations for lifelong learning. The above examples of video learning applications show that institutions of higher learning, teaching research, and teaching institutions are the pillars of applying and promoting video learning technology, which is also proved by the application of video learning technology. It helps solve some defects in the traditional education model and meet the learning needs of the whole society.

For the design of video teaching strategy system, literature video is regarded as an integral part of "video teaching," which is the alias of video teaching syllabus, focusing on classroom teaching video, and other supplementary resources, such as teaching design, textbooks, exercises, and tests, and teaching reflection [7]. Create a theme-centric, semistructured application environment for modular resources. Based on individualized learning content, teaching system element theory, and learning condition theory, a model of influencing factors determined in learning based on relevant national and international research results is developed [8]. The international model shows that students, environment, resources, and teachers are four factors that affect the effect of individualized learning. The explanation of the learning resource factor states that "designers or teachers must be able to integrate and select learning resources in advance according to the needs and interests of students," so the presentation of the organization and resources should be in line with the learning styles and learning expectations of the organizations of the United Nations system. It ensures that students have the opportunity to acquire knowledge through online learning resources and promote the development of their learning ability. Literature provides learning choices for students in the network by studying and analyzing the behavior of students accessing platform resources and making suggestions on appropriate learning resources [9]. Video is defined as "a view resource consisting of a complete unit or knowledge point of multiple knowledge atoms, completed within 2 to 20 minutes" [10]. In order to study how the production of teaching video affects the teaching of program knowledge, four introductory videos are designed, namely, "graphic sketch," "homework recording," "graphic illustration + teacher," and "homework recording and recording teacher," which are summarized into a broadcast classroom mode. Some people point out that students' preference for teaching videos is inconsistent and needs to take into account students' cognitive habits [11, 12].

According to previous studies, it can be seen that the individuation of students' cognitive style in the process of learning is reflected in their understanding of information and the process of compiling information [13]. Although the design of curriculum resources takes into account the different preferences of students in the process of learning, the individuation of students' cognitive style in the process of learning is reflected in their understanding of information [14]. As for the problem of accepting resources in self-study on the Internet, there is no special design research on the satellite resources that accord with the individualized learning mode, and the video resources for personalized learning are less.

3. Video Scene Recognition Based on Artificial Intelligence

3.1. Multimodal Complementary Video Scene Retrieval Method. In the past few years, video has been widely spread in social networks as a new way of communication. Unlike traditional videos, video adds more social attributes, such as comments and topics. The text information contained in these social attributes is very useful for analysis and understanding. The characteristics of multimedia analysis and understanding show that semantic information may be better understood in some multimedia. In some cases, such as in a single mode, and in the tasks of social media (such as referral systems), social characteristics are often common feature information with other data. Considering the semantic richness of text information, text information can be described as a multimodule video scene feature, plus audio-visual information contained in the video itself.

However, space learning faces some challenges and problems when applied to video in real situations. The correlation between visual, sound, and text models of video is weak, but there is complementarity between different parts. Table 1 provides the interrelationship between two modes: visual mode and the interrelationship between sound and text in size text between sound and text. Tambin analyzes the interrelationship between these three units and case categories, as given in Table 2; it can be seen from the table that the correlation between visual mode and category is greater. It can be seen from Table 3 that both unsupervised and supervised learning methods cannot fully reflect semantic information. These problems are called low correlation between different models. The video models with different relationships are weak and complementary. The integrated features of multimodal transport can be represented using supplementary information of multimodal transport as given in Tables 1–3.

Table 1: Correlation between three modes.

Modality	Correlation coefficient
Visual and audio	0.5036
Visual and text	0.5069
Audio and text	0.1217

Table 2: Correlation between individual modes and scene categories.

Modality	Correlation coefficient
Visual	0.6135
Audio	0.1755
Text	0.2827

Table 3: Correlation between representation of three modes in common subspaces and categories.

Modality	Corrcoef_CCA	Corrcoef_MVDA
Visual	0.2502	0.2235
Audio	0.2074	0.0234
Text	0.2849	0.0964

3.1.1. Method Description. Given the training set sample set D, each sample contains three modes of vision, sound, and text. During training, the dataset is divided into N batch inputs. The amount of data per batch is set to b_n, the final loss function representation as shown in

$$E_n(W, B) = -\frac{1}{\text{bn}}\sum_{i=1}^{bn} yi \ln \hat{y}_i + (1 - y_i)\ln(1 - \hat{y}_i). \tag{1}$$

This method uses the minimum batch gradient descent algorithm to optimize the network weight W and bias term B. Parameter updates are shown in formulas (3) and (4), respectively.

$$W = W - \eta_{n1}\frac{\partial E_n}{\partial W}, \tag{2}$$

$$B = B - \eta_{n2}\frac{\partial E_n}{\partial B}. \tag{3}$$

The loss of each batch is continuously adjusted by the backpropagation algorithm until the algorithm converges. Given the test set sample, it is input into the trained network structure, and the prediction results are obtained.

The objective function of learning is shown in

$$\min_{W,B} E_{\text{h}} = |\|Y - W^T \text{sgn}(B)\|^2 + |\left\|\text{S} - \frac{1}{t}\text{sgn}(B)^T \text{sgn}(B)\right\|^2 + \lambda\|W\|^2 \tag{4}$$

$$s.t. B = A^T \widehat{X}$$

$$\text{sgn}(B)\text{sgn}(B)^T = I.$$

Construction of the similarity matrix S is as follows:

$$s_{ij} = \left\{\begin{array}{cc} 1 & \text{if}y_i\text{issimilarto}y_i \\ -1 & \text{else.} \end{array}\right\} \tag{5}$$

For ease of solution, this section relaxes it as

$$\min_{W,B} E_h = ||Y - W^T B||^2 + ||S - \frac{1}{t}B^T B||^2 + \lambda||W||^2$$
$$s.t. \; BB^T = I. \tag{6}$$

Since the closed solution of variable B can not be obtained, the gradient descent method is used to optimize the solution:

The first step: initialization of W and B; the second step: partial derivation of W and B:

$$\frac{\partial E_h}{\partial W} = -B\left(Y - W^T B\right)^T + \lambda W, \tag{7}$$

$$\frac{\partial E_h}{\partial W} = -WY + \left(WW^T + (2 + g)I\right)B - \left(\frac{2}{m}\right)BS. \tag{8}$$

Step 3: update W and B

$$W = W - \eta_{h1}\frac{\partial E_h}{\partial W}, \tag{9}$$

$$B = B - \eta_{h2}\frac{\partial E_h}{\partial B}. \tag{10}$$

Step 4: bring the updated W and B into formula (6) and calculate the value. Continue iteratively performing step 1–step 4 until convergence.

3.1.2. Experimental Setup. First, the most suitable K value is selected by experiment. As given in Table 4, the network achieves the best results when $K = 5$. Therefore, this experiment sets the number of network layers to 5.

Experimental performance measures are mean average accuracy (mAP). This performance metric is also widely used in other literature and is a general standard for retrieval tasks. Given a query sample, the average accuracy (AP) is calculated by

$$AP = \frac{1}{M}\sum_{\mathrm{r}=1}^{R}\mathrm{pre}(r)\mathrm{rel}(r). \tag{11}$$

3.1.3. Results Analysis. As given in the table, the multimodal transport characteristics obtained by cascade direct method and the multimodal transport characteristics obtained by learning subsystem are poor in dimensional image search. The additional multimodal transport approach proposed in this section is more effective than other approaches. The combination of ciphers shows that the correlation and complementarity between different video data models with the best results in dimensional image search are very low. Results are given in Tables 5–7.

By nonlinearly transforming the properties of multimodule combinations into multimodal intermodal transport sensors, the method learns to express higher and more selective properties and automatically learns the relevance of each dimension to the semantics of outer space. The search work has increased due to the conversion of the learning method of characteristic changing hash monitoring into hash code, which keeps the similarity and difference within this category unchanged. Multimodal transport and hash learning methods are very effective.

3.2. Video Scene Classification Based on Consistency Semantic Learning. Different from the traditional video, the content of the video produced by different users is often very different, which is due to the subjectivity and randomness of the user image, resulting in the same image and the intention of expression.

Table 4: Comparison K network performance (mAP) at different values.

Values of K	mAP (@50)	mAP (@100)
$K = 3$	0.4255	0.4240
$K = 4$	0.4246	0.4200
$K = 5$	0.4504	0.4468
$K = 6$	0.4293	0.4298
$K = 7$	0.4253	0.4164

Table 5: Comparison of mAP performance of this section with traditional multimodal fusion methods.

Method	mAP (@50)	mAP (@100)
Concatenating	0.398	0.358
LDA	0.411	0.393
CCA	0.258	0.234
MvDA	0.282	0.250
Multilayer neural network	0.450	0.445
Proposed method	0.469	0.477

3.2.1. Method Description. Neural analytic network is an important method to learn various characteristics in computer vision. Excellent learning and performance in visual recognition of computer images and other visual missions ensure that the spatial characteristics of each image exit after passing the LSTM, the result of which is the concealment of each step; over time, it remains in the structure of the first layer of LSTM and is input as a LSTM unit at that time. The next procedure for calculating the hidden state can be found in the following formulas:

$$X_{ih}^{t} = O_l * \tanh\left(X_{ie}^{t}\right), \tag{12}$$

$$\begin{aligned} x_{ic}^{\mathrm{t}} &= f_t * x_{ic}^{t-1} + i_t * \widetilde{x}_{ic}^{t} \\ \mathrm{f_t} &= \sigma\left(W_{\mathrm{xf}}x_i^{\mathrm{t}} + W_{\mathrm{hf}}x_{ih}^{t-1} + b_f\right) \\ \mathrm{i_t} &= \sigma\left(W_{\mathrm{xi}}x_i^{t} + W_{\mathrm{hi}}x_{ih}^{t-1} + b_{\mathrm{i}}\right) \\ \mathrm{o_t} &= \sigma\left(W_{\mathrm{xo}}\mathrm{x}_i^{t} + W_{\mathrm{ho}}x_{ih}^{t-1} + b_{\mathrm{o}}\right) \\ \widetilde{x}_{ic}^{t} &= \tanh\left(W_{xc}x_i^{t} + W_{hc}x_{ih}^{t-1} + b_c\right), \end{aligned} \tag{13}$$

where α_t is the weight of each frame without attention module that is automatically weighted by the inner product. The calculation process of weight and inner product is as follows:

$$\alpha_t = \frac{\exp\left(x_{ih}^{t}\right)}{\sum_{t=1}^{n}\exp\left(x_{ih}^{t}\right)}, \tag{14}$$

$$Z_{\mathrm{i}}^{t} = \langle\alpha_{\mathrm{t}}, \mathrm{x}_{\mathrm{ih}}^{\mathrm{t}}\rangle. \tag{15}$$

After the attention module is weighted, the features of all frames pass through the LSTM layer again. LSTM layer is still a single-layer structure, and the output features of the layer are as follows:

Table 6: Comparison of mAP@50 performance of methods in this section with individual hash learning methods.

Method	8 bits	16 bits	32 bits	64 bits
LFH	0.388	0.365	0.406	0.359
KSH	0.338	0.393	0.452	0.439
SDH	0.239	0.223	0.223	0.292
COSDISH	0.330	0.369	0.400	0.375
Proposed method	0.469	0.452	0.455	0.454

Table 7: Comparison of mAP@100 performance of this section method with individual hash learning methods.

Method	8 bits	16 bits	32 bits	64 bits
LFH	0.395	0.366	0.406	0.358
KSH	0.341	0.377	0.431	0.411
SDH	0.247	0.219	0.220	0.266
COSDISH	0.330	0.369	0.403	0.378
Proposed method	0.477	0.453	0.455	0.453

$$z_{ih}^{n} = o_n * \tanh\left(z_{ic}^{n}\right), \tag{16}$$

$$\begin{aligned} z_{ic}^{n} &= f_n * z_{ih}^{n-1} + i_n * \tilde{z}_{ic}^{n} \\ \tilde{z}_{ic}^{n} &= \tanh\left(W_{zc} z_i^n + W_{hc} z_{ih}^{n-1} + b_c\right) \\ \mathrm{o}_n &= \sigma\left(W_{zo} z_i^n + W_{ho} z_{ih}^{n-1} + b_o\right) \\ \mathrm{o}_n &= \sigma\left(W_{zo} z_i^n + W_{ho} z_{ih}^{n-1} + b_o\right) \\ f_n &= \sigma\left(W_{zf} z_i^n + W_{hf} z_{ih}^{n-1} + b_n\right). \end{aligned} \tag{17}$$

Each bit of the predicted category is shown in the following equation:

$$\hat{\mathrm{y}}_{\mathrm{i}} = \mathrm{softmax}\left(W_{\mathrm{fc}} \mathrm{z}_{\mathrm{ih}}^{\mathrm{n}} + \mathrm{b}_{\mathrm{fc}}\right). \tag{18}$$

To improve the consistency of microvideos in the same scenario, due to inconsistent content, this section adopts a two-sector framework and a supervised learning mechanism, while maintaining the similarity of scene categories between two-branch network parameters and samples. The forum where you want to post is a stage. The results L3 the stage cross-$L1$, $L2$ and relative loss functions are as follows:

$$L_1 = -\sum\nolimits_{i=1}^{Pl} \sum_{j=1}^{k} y_{ij} \ln \hat{y}_{ij1} + \left(1 - y_{ij1}\right) \ln\left(1 - \hat{y}_{ij1}\right), \tag{19}$$

$$L_2 = -\sum\nolimits_{i=1}^{Pl} \sum_{j=1}^{k} y_{ij2} \ln \hat{y}_{ij2} + \left(1 - y_{ij2}\right) \ln\left(1 - \hat{y}_{ij2}\right), \tag{20}$$

$$L_3 = \frac{1}{2Pl} \sum\nolimits_{m=1}^{Pl} y_m d_m^2 + \left(1 - y_m\right) \max\left(\mathrm{marg}in - d_m, 0\right)^2. \tag{21}$$

The final objective function of the ACSL method is to minimize the weighted fusion of three losses. The calculation process of the objective function L is as follows:

$$\min_{W,B} L = \alpha L_1 + \alpha L_2 + \beta L_3, \tag{22}$$

where α and β are the equilibrium parameters, and these two parameters are hyperparameters. W and B are the network learning parameters. The optimization method used in this section is the stochastic gradient descent method.

3.2.2. Experimental Setup. In this section, the method network structure consists of two branches; each branch consists of two LSTM layers and one Attention layer, where the output dimension of the first LSTM layer is Q24, 1. The return_sequences "property value is True" indicates that all frames correspond to a 1024-dimensional output. And the output dimension of the second LSTM layer is 128, "The return_equences" property value is False, and it represents the final 128-dimensional output as given in Tables 8–10.

3.2.3. Experimental Results. The comparison of performance of dual-branch and single-branch networks is given in Table 11. Validation of LSTM layers is given in Table 12.

This section introduces the classification of semantic learning Scopic models. This method utilizes the spatial characteristics of the video scene through a central pre-training image recognition network. Ensure the semantic expression ability of these features, combine the spatial characteristics with the time series characteristics adopted by the LSTM, and introduce the attention mechanism when extracting the time series characteristics. The semantic related content in the video box is automatically removed. Some table features are automatically weighted to obtain a more neutral table of spatial and temporal features. Through dual network and supervised learning mechanism, the consistency of video content in the same scene and the difference of time characteristics are maintained. Temporary

Table 8: Comparison of the performance of methods in this section with existing methods on Maryland datasets.

Class	HOF + GIST	SFA	C3D	ACSL
Avalanche	0.200	0.600	1.000	1.000
Boiling water	0.500	0.700	0.900	1.000
Chaotic traffic	0.300	0.800	0.900	1.000
Forest fire	0.500	0.100	0.800	1.000
Fountain	0.200	0.500	0.900	1.000
Iceberg collapse	0.200	0.600	1.000	0.800
Landslide	0.200	0.600	0.800	0.800
Smooth traffic	0.300	0.500	0.800	0.800
Tornado	0.400	0.700	0.800	0.800
Volcanic eruption	0.200	0.800	0.900	0.800
Waterfall	0.200	0.500	0.700	0.400
Waves	0.800	0.600	1.000	0.600
Whirlpool	0.300	0.800	0.900	1.000
Average	0.330	0.600	0.860	0.850

Table 9: Performance comparison of methods in this section with existing methods on Yupenn datasets.

Class	HOF + GIST	SFA	C3D	ACSL
Beach	0.870	0.930	0.970	1.000
Elevator	0.870	0.970	1.000	1.000
Fire	0.630	0.700	1.000	1.000
Fountain	0.430	0.570	0.830	1.000
Highway	0.470	0.930	0.970	0.890
Lightning	0.630	0.870	0.930	1.000
Ocean	0.970	1.000	1.000	1.000
Railway	0.830	0.930	0.970	1.000
Rfiver	0.770	0.870	1.000	0.890
Sky	0.870	0.930	0.970	1.000
Snowing	0.470	0.700	0.930	0.560
Street	0.770	0.970	1.000	0.890
Waterfall	0.470	0.730	0.970	0.890
Windmill	0.530	0.870	1.000	0.890
Average	0.680	0.850	0.970	0.930

experimental results confirm that this method is successful in classifying video.

3.3. Video Scene Classification Based on Multimodal Semantic Enhancement. Video, as a new form of media, is widely spread in social platforms. In addition to the visual information and sound information in the traditional video, the video has a good auxiliary effect on the scene understanding of the video. Therefore, the feature fusion of visual, sound, and text modes can learn rich scene feature representation.

Regarding the MESL method, the semantic enhancement of strong semantic modes to weak semantic modes is realized by minimizing the semantic distance between weak semantic modes and strong semantic modes and the discriminant loss of single modes. So, the semantic distance minimization objective function is shown in the following formula:

$$\min_{W,B} \text{dis} = \alpha \left\| y_i^{v\_\text{out}} -_i^{a\_\text{out}} \right\|_2 + \beta \left\| y_i^{v\_\text{out}} - y_i^{t\_\text{out}} \right\|_2. \qquad (23)$$

Table 10: Performance comparison of methods in this section with existing methods on videoSceneData_10 datasets.

Class	HOF + GIST	SFA	C3D	ACSL
Museum	0.250	0.080	0.100	0.797
Pier	0.130	0.070	0.050	0.594
Garden	0.500	0.040	0.100	0.815
Office	0.030	0.020	0.050	0.594
Bridge	0.190	0.120	0.120	0.768
Racetrack	0.230	0.070	0.120	0.774
Landmark	0.210	0.060	0.050	0.788
Aquarium	0.300	0.470	0.050	0.818
Lake	0.060	0.070	0.120	0.683
Bowling alley	0.290	0.030	0.090	0.895
Average	0.220	0.100	0.100	0.753

Table 11: Comparison of performance of dual-branch and single-branch networks.

Class	Single-branch	Two-branch
Museum	0.774	0.797
Pier	0.622	0.594
Garden	0.877	0.815
Office	0.438	0.594
Bridge	0.752	0.768
Racetrack	0.744	0.774
Landmark	0.741	0.788
Aquarium	0.796	0.818
Lake	0.590	0.683
Bowling alley	0.860	0.895
Average	0.719	0.753

After the semantic enhancement, the cross-entropy loss of the three modes is calculated as follows:

$$\text{Loss}\_v = -\sum_{i=1}^{m} y_i^{v\_\text{out}} \ln\, y_i + \left(1 - y_i^{v\_\text{out}}\right) \ln\left(1 - y_i\right), \qquad (24)$$

$$\text{Loss}\_a = -\sum_{i=1}^{m} y_i^{a\_\text{out}} \ln\, y_i + \left(1 - y_i^{a\_\text{out}}\right) \ln\left(1 - y_i\right), \qquad (25)$$

$$\text{Loss}\_t = -\sum_{i=1}^{m} it\_\text{out} \ln\, y_i + \left(1 - y_i^{t\_\text{out}}\right) \ln\left(1 - y_i\right). \qquad (26)$$

Four trials are included in this section to validate the method of this section, that is, MESL, in the dimensional scenario classification. The first test was selected as a parameter to determine the equilibrium coefficient of the elements of the PE function. The second experiment is performance comparison to verify the comparison between the method and the method. The third trial is a combined trial to verify the effectiveness of weak semantic effectiveness and integration. The fourth test of multimodal transport is a confluence test to verify the consistency of the methods in this section and to evaluate the performance.

The integration method of multimodal transport is superior to that of single mode. The results of the combined tests are given in Tables 13 and 14. As given in Table 13, the

Table 12: Validation of LSTM layers.

Class	W/O-LSTM	W/LSTM
Museum	0.589	0.797
Pier	0.139	0.594
Garden	0.834	0.815
Office	0.775	0.594
Bridge	0.721	0.768
Racetrack	0.719	0.774
Laminuurk	0.668	0.788
Aquarium	0.776	0.818
Lake	0.500	0.683
Bowling alley	0.819	0.895
Average	0.684	0.753

Table 13: Modal semantic enhancement experiment.

Modal	Acc_before_enhancement	Acc_after_enhancement
Audio	0.3286	0.3427
Text	0.4153	0.4210
Visual	0.9816	0.9697

Table 14: Ablation experiment.

Modal	Accuracy
Audio	0.3427
Text	0.4210
Visual	0.9697
Visual + audio + text	0.9826

characteristics of sound and text molds are superior to the previous version. The performance of the visual model decreases slightly, mainly because the semantic speech and text models are weak, but the visual model has not been improved, and the characteristics of multimodal transport are better than the single mode in the classification of microvideo images. Details are given in Tables 13 and 14.

In order to make full use of the semantic pattern contained in the weak semantic pattern, the method of this section adopts the method of strengthening the semantic pattern to deal with the weak semantic situation and enhances the expression ability of the semantic model characteristics. For microvideo teaching, in some cases, it is impossible to determine the scene type directly according to the way. However, the type of scene can be judged by sound or part of the text, which is very complementary to the lack of semantic model vision. Therefore, it is suggested that the enhanced multimodal transport microvideo images be classified by minimizing the semantic distance between the strong and weak modules and the interpretation errors of different molds. And we further integrate the characteristics before and after, in order to better integrate the strong model and the weak model semantically. Through the weight of self-study method and the automatic learning fusion between different weight modes, the experimental results suitable for the weight supplement of multimodal transport prove the effectiveness of the algorithm.

4. Requirements for Language Video Design

4.1. Video Should Focus on Students' Original Cognition. According to the students' learning foundation and ability, we must fully consider, study, and understand the confusion points and obstacles in practice, as well as the obstacles encountered in practice. Coverage should be as broad as possible, taking into account as many students as possible, for example, the use of video that links new learning materials to past knowledge, or new learning materials to knowledge acquired from student courses, and guides students to reflect in-depth dialogue and gradually improve their cognitive and ability. Whether in writing or onstage, or in determining the key and difficult points in the textbook, students must be able to understand and learn knowledge.

4.2. Video to Support Classroom Teaching Effectively. For students with strong motivation, video is very obvious. Teachers should be familiar with the teaching content of a subject or chapter and should be familiar with the overall planning and difficulties, problems, and requirements of teaching materials: accordingly, for example, regarding the teaching of ancient poetry, students may have difficulties in understanding words, customs, and the ancient system of rights. Before that, video can be systematically produced for students to use, covering classical Chinese, special phrases, vocabulary, and ancient words in textbooks. It can display the priorities, difficulties, and doubts of ancient poetry in the eyes of students, help them familiarize themselves with ancient philosophy and framework, understand the differences of ancient Chinese, and understand the rich connotation of Chinese culture.

4.3. The Practice of Chinese Subject Should be Emphasized by Video. Language learning is very practical, and teachers must guide students to deepen their understanding of PRA language learning methods. Science and technology teachers must combine teaching needs with students' learning conditions before preparing materials on video content, integrating practical content, and enriching and expanding video content. For example, once visual expression education is provided in the classroom, teachers can combine this education with writing education and carry out training activities that contribute to improvement of students' ability to write and write, create video design, download materials from the website, and connect different learning elements through various learning lines, in a way of presentation and narration, including a wide range of historical background, and landscape characteristics. On the one hand, it enables students to observe and understand in a visual and detailed way; on the other hand, it improves students' rationality and systematization. Video enables students to obtain vision, which has a positive impact on expanding students' thinking and enriching written materials.

5. Optimization of Teaching Strategy System Design of Video Language

5.1. Cognitive Theoretical Basis of Multimedia Learning. Mayer cognitive theory of multimedia learning holds that meaningful learning must organize multimedia information according to different multimedia cognitive methods. The model shows that external information stimuli first enter sensory memory and then deliberately choose. From sensory memory, the results of multimedia cognitive experiments show that the use of multimedia has a positive impact on the retention and transfer of learning, which is superior to single learning. Multimedia display enables students to learn more deeply and understand learning content better, rather than putting learning content in a holistic form. Learning outcomes are better only when students focus enough working memory resources on useful information processing, as shown in Figure 1.

Through extensive practical research, the effectiveness of learning multimedia is proved. Video is a form of resources, combined with text and image, which is very important to image and sound and consistent with the meaning of multimedia learning. These characteristics are the basic concepts of this virtual view design research. Teaching videos spread knowledge through mixed music works and jointly display content. Educational video can be classified according to the prominence of sound and image and can also be used in the form of image, audio-assisted interpretation of the content, or in the form of illustrations. Mayer seven principles put forward media design, namely, multimedia knowledge, space proximity, time proximity, consistency, personal difference, and so on are helpful for understanding. They are the basis of current research on microinformation presentation design.

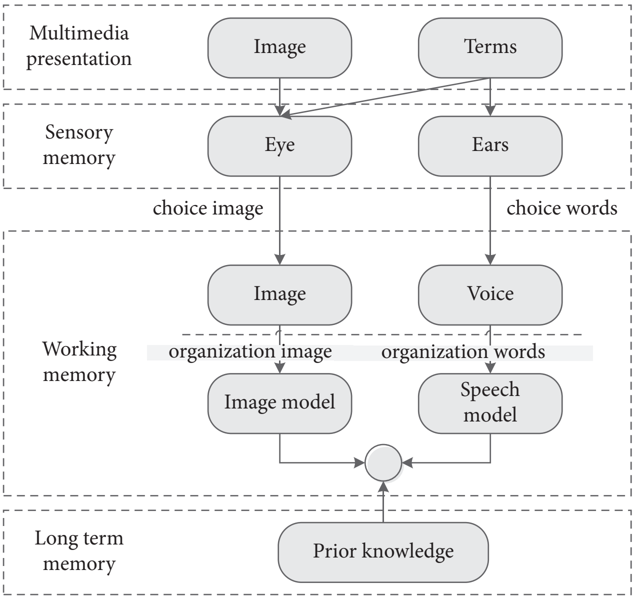

Figure 1: Cognitive model of multimedia learning.

5.2. Elements of a Video for VARK Learning Styles. The video consists of two parts: image and sound: video learning (i.e., interaction of visual and auditory channels), information stimulation, and information processing, plus features of four types of Vark learning styles, in order to enable interaction between body organizations to better understand learning content. As shown in Figure 2, images and sounds reflect the uniqueness of each style. Video oriented to Vark learning styles should contain different contents, as shown in Figure 2.

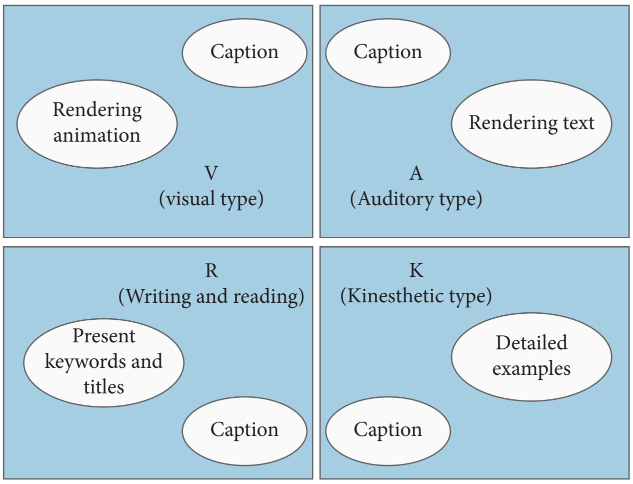

Figure 2: Classification of video elements for VARK learning style learners.

5.3. Facilitating Information Processing for Learners. Reducing the redundancy of information independent of learning content in video teaching design is helpful for students to process information and cognitive ability of redundancy effect according to multimedia learning theory. The principle of redundancy means that students can better learn visual materials composed of “animation” and “narration” rather than “animation,” “narration,” and “narration” of visual materials. “Screen text,” better knowledge in terms of migration, has too many animation effects, text, or subtitles and may be redundant, interfering with students’ information processing: on the one hand, animation, text, and subtitle are received through the eyes, which results in visual overload of information; on the other hand, when the same information appears in visual and auditory form, the brain processing ability is limited, and processing two kinds of the same information reduces the processing of other useful information and increases cognitive ability. The internal Figure 3 shows an information processing model.

Based on the above findings, the study concluded that video should be designed to reduce excessive information and avoid interference and that these images should no longer appear in the same subtitle and recording or music background; on the basis of different characteristics of learning styles, visual video is presented in animation form, emphasizing visualization of abstract content, while the remaining acoustic, text, and kinetic view frequencies take the form of PPT video screens combined with

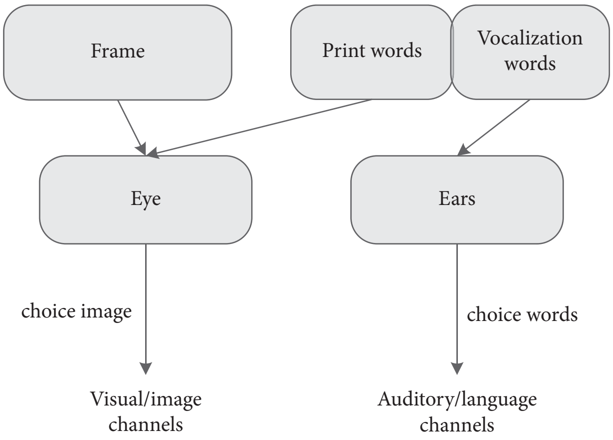

FIGURE 3: Information processing process model with redundant information.

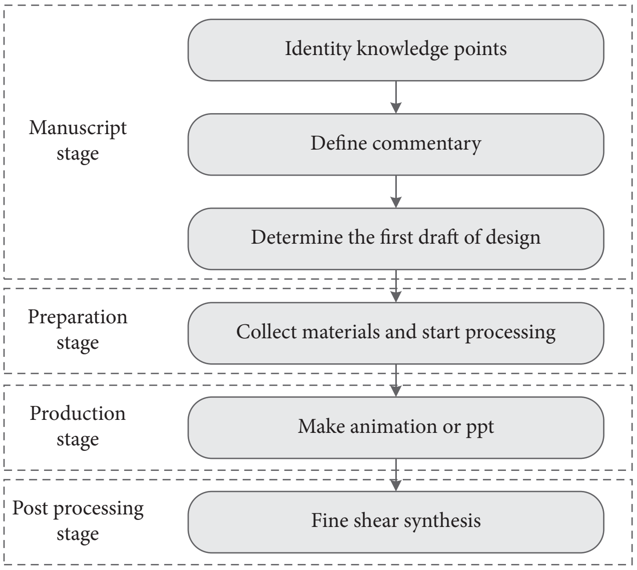

FIGURE 4: Production process for learning style microvideo.

demonstrations. PPT production designed many animation effects in order to be dynamic, but too many animation effects distracted the students' attention. As the basis of the principle of avoiding interference information redundancy, the complex influence of animation should be reduced as far as possible in design, only to gradually introduce some text and image characteristics, to reduce the simple influence of animation, without subtitles, and not to increase music.

5.4. Optimization of Teaching Strategy System Design for Microvideo Language. The microvideo development process for this study includes the drafting, preparation, fabrication, and subsequent processing phases, using the evaluation phase as a teaching practice. A project phase includes the identification of knowledge points and the definition of explanatory or explanatory text and the identification of design manuscripts; the preparatory phase includes the collection and processing of materials required for video production; the production phase of animation or PPT production materials; and the video synthesis phase, which simplifies video and audio production, processes the details of microvideo and improves the overall quality of microvideo. Figure 4 shows the specific process. Details are shown in Figures 4 and 5.

Students will study through the network learning platform, before class to complete the microvideo learning, and in the after-class stage, teachers will randomly ask questions to check academic performance, to ensure the integrity of academic performance. Teachers randomly select one or two students from each learning style to ask questions and determine the satisfaction of all students with the use of microvideo. The results were tested by students' feedback to the class and the survey of school design. Use microvideo, and use satisfaction questionnaire and interview students when designing microvideo.

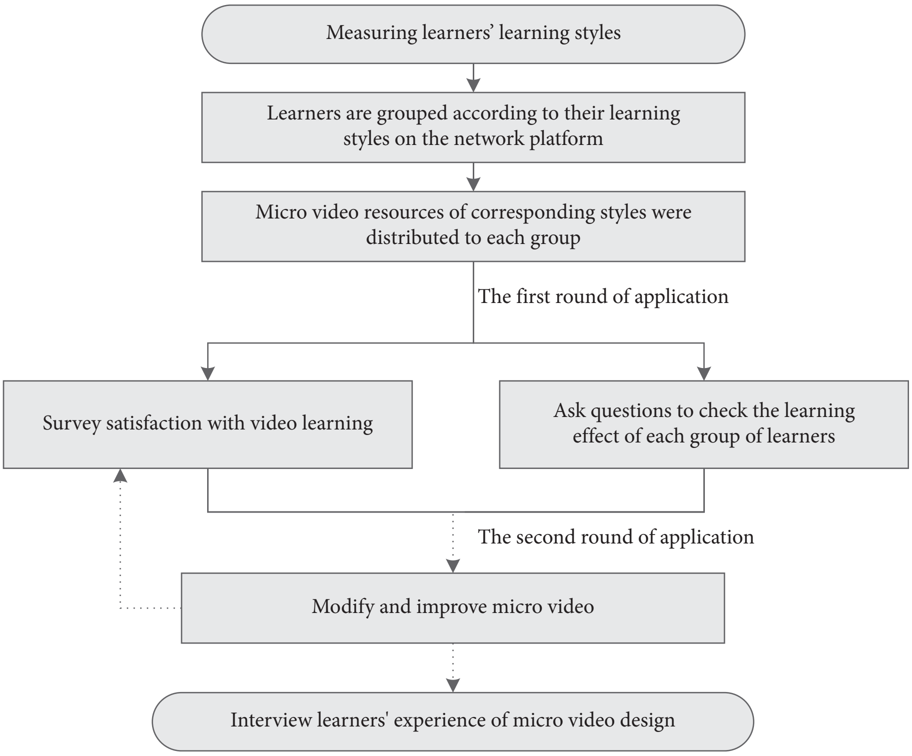
Measuring learners' learning styles
Learners are grouped according to their learning styles on the network platform
Micro video resources of corresponding styles were distributed to each group
The first round of application
Survey satisfaction with video learning
Ask questions to check the learning effect of each group of learners
The second round of application
Modify and improve micro video
Interview learners' experience of micro video design

Figure 5: Microvideo resource application process diagram.

6. Conclusion

During the era of Internet + education, the emergence of mobile terminals such as smart phones and mobile blackboards provides strong teaching support for traditional language classes, supporting language teaching, with emphasis on changing the use of cumbersome and ineffective traditional teaching methods. Using interactive online learning platform to support students' active learning and cooperative learning, we can strengthen the ability of problem finding and questioning by implementing reasonable video teaching. Effective video teaching has injected the contemporary water of life into the development of language subject. If we further improve the quality of video teaching and production, provide students with a platform for interactive thinking and demonstration, and further enrich the means of dissemination of video, such as making video in language, in front-line language classes, language video will be more widely used.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] M. B. Brown and J. K. Lippincott, "Learning spaces: more than meets the eye," *Educause Quarterly*, vol. 45, no. 1, pp. 14–16, 2012.

[2] D. Oblinger, "Leading the transition from classrooms to learning spaces," *Educause Quarterly*, vol. 28, no. 1, pp. 14–18, 2005.

[3] L. Sheremetov and A. G. Arenas, "EVA: an interactive Web-based collaborative learning environment," *Computers & Education*, vol. 39, no. 2, pp. 161–182, 2002.

[4] I. Blau, "Teachers for smart classrooms: the extent of implementation of an interactive whiteboard-based professional development program on elementary teachers instructional practices," *Interdisciplinary Journal of e-Skills and Lifelong Learning*, vol. 7, pp. 275–289, 2011.

[5] D. Kim, Y. Lee, W. L. Leite, and A. C. Huggins-Manley, "Exploring student and teacher usage patterns associated with student attrition in an open educational resource-supported online learning platform," *Computers & Education*, vol. 156, no. 10, Article ID 103961, 2020.

[6] N. S. Ali, K. Hodson-Carlton, and M. Ryan, "Students' perceptions of online learning: implications for teaching," *Nurse Educator*, vol. 29, no. 3, pp. 111–115, 2004.

[7] C. Bouras, G. Hornig, V. Triantafillou, and T. Tsiatsos, "Architectures supporting E-learning through collaborative virtual environments: the case of invite," in *Proceedings of the IEEE International Conference on Advanced Learning Technologies*, pp. 13–16, Madison, WI, USA, August 2001.

[8] A. A. Drozdova and A. I. Guseva, "Modern technologies of E-learning and its evaluation of efficiency," *Procedia - Social and Behavioral Sciences*, vol. 237, pp. 1032–1038, 2017, Feb.

[9] V. Demertzi and K. Demertzis, "A Hybrid Adaptive Educational Elearning Project Based on Ontologies Matching and Recommendation System," 2021, http://arxiv.org/abs/2007.14771.

[10] S. Adamu and J. Awwalu, ""The Role of Artificial Intelligence (AI) in Adaptive Elearning System (AES) Content FORMATION: Risks and Opportunities Involved," 2021, http://arxiv.org/abs/1903.00934.

[11] J. Cruz-Benito, J. C. Sánchez-Prieto, R. Therón, and F. J. García-Peñalvo, *Measuring students' acceptance to AI-driven assessment in eLearning: proposing a first TAM-based research model,"Learning and Collaboration Technologies. Designing Learning Experiences*, vol. 11590, pp. 15–25, Springer International Publishing, Cham, Switzerland, 2019.

[12] J. L. Mcbrien, R. Cheng, and P. Jones, "Virtual spaces: employing a synchronous online classroom to facilitate student engagement in online learning," *International Review of Research in Open and Distance Learning*, vol. 10, no. 3, 2009.

[13] A. Y. Kolb and D. A. Kolb, "Learning styles and learning spaces: enhancing experiential learning in higher education," *The Academy of Management Learning and Education*, vol. 4, no. 2, pp. 193–212, 2005.

[14] A. Ibrahim, M. W. Aulls, and B. M. Shore, "Development, validation, and factorial comparison of the Mc Gill Self-Efficacy of Learners for Inquiry Engagement (Mc SELFIE) survey in natural science disciplines," *International Journal of Science Education*, vol. 38, no. 16, pp. 2450–2476, 2016.