*Research Article*

# Text Reconstruction Method of College English Textbooks from the Perspective of Language Images

**Xiaojuan Peng** (ID)

*Guilin University of Electronic Technology, Guilin 541004, China*

Correspondence should be addressed to Xiaojuan Peng; deliapeng0826@guet.edu.cn

With the continuous development of social economy, English learning plays an increasingly important role in daily communication. However, the update speed of English textbooks is far lower than the development speed of English. How to reconstruct the text of English textbooks in colleges and universities to improve the learning effect of college English has become an urgent problem to be solved. In order to solve this problem more effectively, this paper proposes a text reconstruction method for college English textbooks from the perspective of language images. First, this paper proposes a text reconstruction network model for college English textbooks, which includes two self-network models, namely, a text feature extraction network and an image feature extraction network. Second, text designs an English text feature reconstruction network to fuse image features and text features to guide the generation of new English texts according to the generated emotions. Finally, through a large number of experiments, it is proved that the text reconstruction method of college English textbooks from the perspective of language images can effectively generate new texts of college English textbooks, enhance the emotional color of college textbooks, and improve the effect of English learning.

## 1. Introduction

English plays an important role in the daily affairs of internationalization, but learning English [1, 2] is often a difficult problem. Usually, English learning requires teachers to reconcept and design the content of the text [3, 4], improve students' interest in learning in a reasonable way, turn boring learning into active learning, and deepen the understanding of the text content [5, 6]. The text reconstruction method of college English textbooks [7, 8] has gradually become an effective method to improve college students' English learning. Through the reconstruction of a variety of English knowledge in English textbooks and the use of diversified text forms, college students can improve the English learning methods and skills [9, 10] mastered by college students and deepen their understanding of the text content. The method of text reconstruction of English textbooks is widely used in various learning tasks such as English learning in primary and secondary schools, professional English training, and English reading teaching.

The teaching method of English textbooks is relatively simple. Traditional learning methods [11] often fail to stimulate students' desire to learn, limit students' imagination and comprehension ability, and cannot improve students' learning effect [12, 13]. Usually, teachers in colleges and universities will give a certain degree of guidance to study [14, 15] and reorganize [16, 17] the structure and content of college English textbooks, reconstruct the old-fashioned content in the text into lively cases, stimulate students' imagination, and improve students' learning effect. However, many teachers find some problems in the teaching process. First of all, each country has a certain degree of cultural difference [18, 19]. In order to improve students' ability to master English and improve their learning effect, students need to constantly feel the emotions [20, 21] and foreign cultures reflected in the text. Secondly, each teacher's teaching style [22, 23] is very different, and the understanding and reconstruction of the text content are very different. How to make students have a strong interest in the learning of English textbooks often requires teachers to pay a

lot of money, time, and energy [24, 25]. The difficulties in learning traditional college English textbooks are that the content of the text is relatively old-fashioned and lacks emotional color. College students will have a dull learning mood during learning, which will reduce the effect of learning.

With the continuous progress of science and technology, artificial intelligence methods represented by deep learning technology [26, 27] continue to solve the problems existing in real teaching. Deep learning technology is mainly divided into two aspects. One is the neural network method based on image processing [28, 29], which is mainly used to process image tasks and is widely used in vision-related tasks such as image recognition, image classification, and image segmentation. The second is the neural network method based on text processing [30, 31], which mainly deals with text tasks and is widely used in natural language processing tasks such as text emotion recognition, text emotion classification, and text prediction. In the task of text reconstruction of college English textbooks, it is mainly related to the content of the text and the images in the book. The text features and image features are extracted through neural networks, and new textbook texts with specific emotions can be generated by adding emotional factors and continuously improve students' learning interest and learning effect.

In order to allow college students to better master English and improve the effect of English learning, this paper proposes a text reconstruction method for college English textbooks from the perspective of language images. Firstly, this paper proposes a text feature extraction network based on college English textbooks. It mainly includes two subnetworks; the two subnetworks are the text feature extraction network and the image feature extraction network [32, 33]. Secondly, this paper designs an English text feature reconstruction network [34], which is mainly used to fuse image features and text features, reconstruct college English textbook texts according to the generated emotions, and generate texts with characteristic emotions. Finally, through a large number of experiments, it is proved that the text reconstruction method of college English textbooks based on the perspective of language images can effectively reconstruct the texts of college English textbooks, enhance the emotional color of college textbooks, and improve the English learning ability of college students' teaching burden. The main purpose of this paper is to solve the difficulty of learning English textbooks in colleges and universities, improve the learning effect of college students, and reduce the workload of teachers in schools.

## 2. Related Work

### 2.1. Image Feature Extraction.
Image features exist widely in people's daily life and usually contain human emotional information, which is an effective way of emotional communication and communication between people. Human emotions are usually expressed through images, and this emotion representation method is widely used in intelligent monitoring, online learning, autonomous driving, and other fields.

Image emotional feature extraction is firstly to preprocess the input image, then use the convolutional neural network to extract the emotional information in the image, and then classify according to the extracted emotional information. Image emotion feature extraction is an indispensable link in neural network, and the effect of emotion feature extraction ultimately affects the effect of emotion classification. Convolutional neural networks can obtain higher-level and more abstract feature representations of images by directly extracting the features of images for learning, thus obtaining more essential features in images, and thus making deep learning features more accurate and general.

### 2.2. Text Feature Extraction.
In recent years, text data has grown rapidly through the Internet, and a large amount of text data has been continuously accumulated. These massive data contain a lot of valuable information. Natural language processing is the main method to solve these pieces of text feature information. It uses the text to train a classifier model and then uses the trained network model to divide and supplement the new text.

Sentiment recognition of text content is also called sentiment analysis. It mainly divides the text into a variety of emotion types according to the meaning and emotional information expressed by the text, which is a multisentiment classification problem. By analyzing and researching these text data, important feature information is extracted, which can be used to analyze the public's attention to hot topics and emotional tendencies, which provides important research ideas for correctly guiding the direction of social public opinion.

With the continuous advancement and development of technology, multimodal data is still a very challenging task at present. In image and text multimodal data, the information contained in text and image is generally complementary. Compared with single-modal data of text or images, multimodal data contains more comprehensive information and can better display and explain the emotional characteristics used. First, the amount of information contained in data of different modalities is often different. Sentiment analysis of multimodal data requires effective understanding and extraction of emotional features of multimodal data. Compared with the traditional single-modal sentiment method, the multimodal sentiment analysis task needs to combine the effective information of multiple modalities to extract the feature interaction between the modal information and the multimodal information in a reasonable way.

### 2.3. English Text Reconstruction.
It is often difficult to improve the English ability of college students through traditional English teaching methods, which are not conducive to students' good grasp of English syntax knowledge and enhance the effect of English learning. Therefore, on the basis of traditional college English textbook learning combined with English practical teaching activities, many teachers have proposed methods of text reconstruction of college English textbooks from the levels of syntactic

analysis, text sorting, title prediction, and textual analysis to improve English learning. The learning effect has a certain reference value.

English text prediction is a high-level idea of text reconstruction in English textbook learning. It requires readers to predict the unknown content of English textbooks based on existing information and personal understanding of textbook content. Mind map refers to presenting the text content in a more specific, vivid, and hierarchical organization based on the thematic events in English textbooks, using keywords, images, color changes, and other main branches. Based on the mind map of text content, it can help students have a deeper understanding of the content of English textbooks, splicing fragmentary knowledge, grasp the key points in learning, and understand English textbooks from a macrolevel. Reconstruction of text content is a common way of learning, which means that learners, from the perspective of the author, combine the knowledge and content they have learned and rewrite the text content, trends, trends, and so on on the basis of understanding and learning English textbooks. Reconstruction can deepen the understanding of the text content of the textbook and realize the process of learning, accumulating knowledge, and transforming the text content of college English textbooks. Text reconstruction can rearrange and understand the content of English textbooks, improve students' interest in learning, and deepen their understanding of the text in textbooks.

## 3. Methods of Text Reconstruction in College English Textbooks

In order to more effectively obtain the feature information of each modality and the interaction of feature information between modalities, this paper adopts a feature fusion model based on standard transformer structure to fuse the extracted image and text features and use it to generate emotional new text message in Figure 1. First, this paper proposes a text feature extraction network, which uses the LSTM network structure to encode temporally continuous text feature information, then uses the attention mechanism to obtain the text feature information that this paper pays more attention to, and finally generates text feature vectors for input into college English textbooks for text reconstruction in the network. Second, this paper proposes an image feature extraction network to input images from textbooks into the network structure. Its main function is to extract human emotional information through images, input the generated image feature vector into the text reconstruction network of college English textbooks, and add more emotional features to the newly generated text information. With two subnetworks and a backbone network, the text reconstruction method of college English textbooks proposed in this paper can generate new content with specific emotions through a generative method. Deep learning methods can generate various types of text reconstruction content, increase students' interest in learning, and greatly reduce the workload of school teachers.

*3.1. Text Feature Extraction Network.* The text feature extraction network proposed in this paper is improved on the basis of Att-CNN-BiGRU, which mainly consists of three aspects: text vectorization, text feature extraction, and feature vector generation, as shown in Figure 2. The first part is the vectorization of text features. The main function is to map each word in the input text to a vector representation, and the text vector is represented by the word vector.

Text vectorization is to map the text into a vector that can be recognized and processed by a computer. It is mainly to avoid the problems of one-hot encoding vector sparseness and dimension disaster. Generally speaking, text vectors are mainly composed of word vectors and position vectors. The word vector is to convert the words in the text into vectors that can express semantic information. Considering the meaning of the words and the influence between words, the semantic similarity of words is calculated. The position vector is to convert the position features of the words in the text into vectors and make it clear that the word in the sentence is the trigger word. The position feature is defined as the relative distance between the current word and the candidate word, which represents the relative position of the current word in the sentence.

$$S = \{w_1, w_2, \ldots, w_n,\},$$
$$X = \{x_1, x_2, \ldots, x_n,\},$$
$$x_i = [e_i, pe_i], \quad d = d_w + d_p, \tag{1}$$

where S represents the sentence, $w_i$ represents the $i$-th value in the sentence, and $n$ represents the length of the sentence. A similar method is used in this paper to map the word $w_i$ to a real-valued vector $x_i$; usually, a sentence consists of multiple values.

The text feature extraction network model proposed in this paper is used to extract text features. Text features are divided into lexical-level features and sentence-level features. The lexical-level feature is to extract the background knowledge of the word, including the word's part of speech, semantic information, emotional information, and classification information. Sentence-level features are the contextual information of the entire sentence in the text, such as grammatical features, degree of association, and emotional expression. Convolution operations are used to extract lexical features in text, and BiLSTM and attention mechanism are used to extract sentence-level features.

Convolutional neural networks can only extract local feature information within the convolution window and cannot perform feature extraction and association on textual context information. In order to fully consider the lexical-level features and sentence-level features of each word and enhance the degree of association between words, this paper introduces a self-attention mechanism to obtain lexical-level features more comprehensively while avoiding lexical position information in the pooling operation of loss.
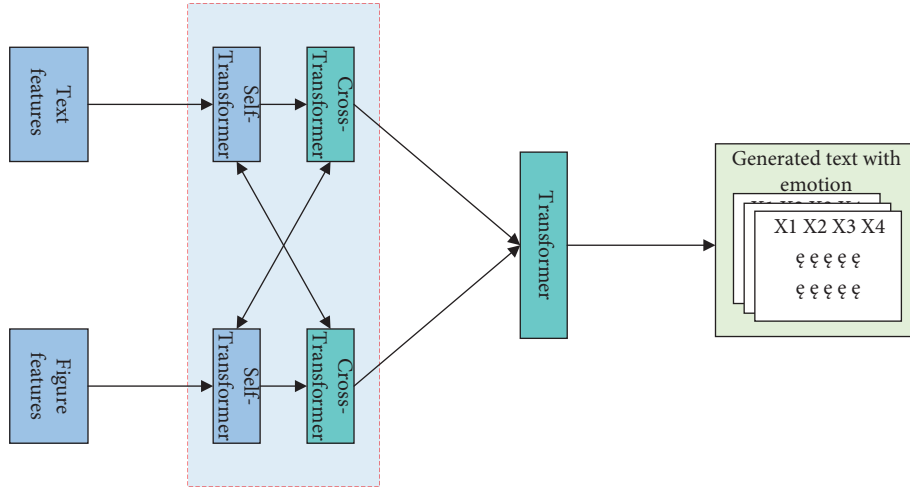
Figure 1: Structure diagram of the text reconstruction network of college English textbooks.
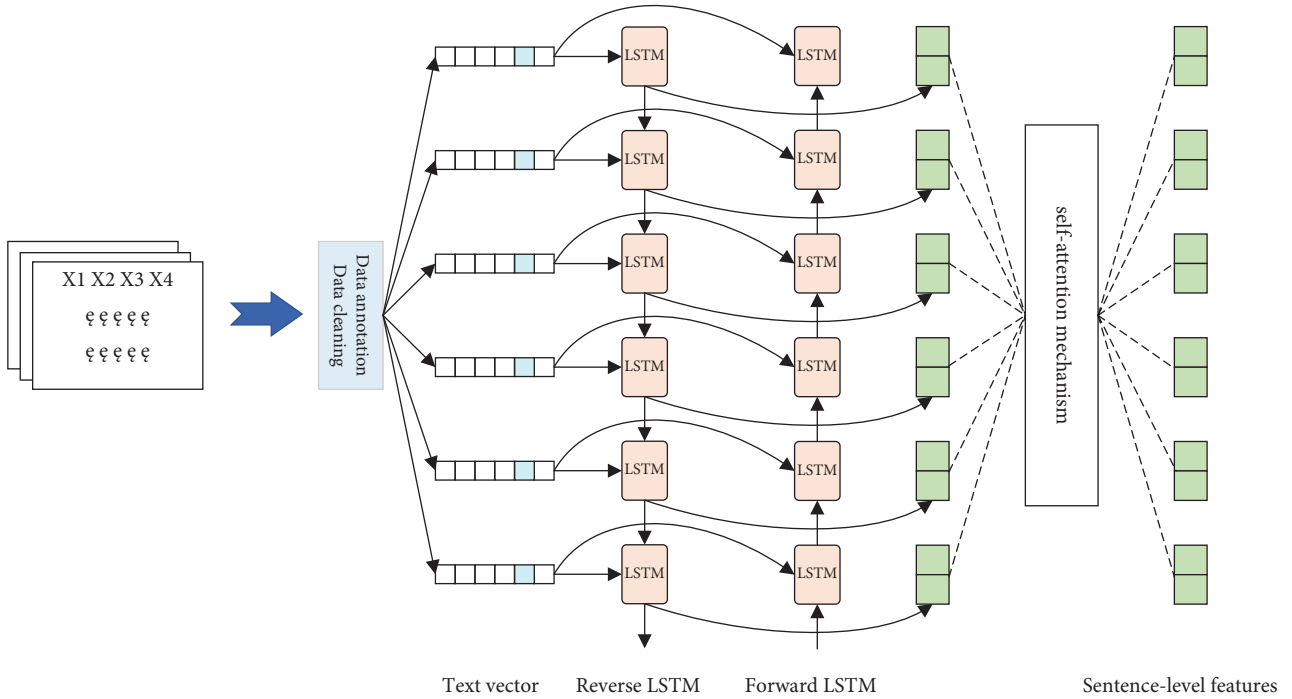


Figure 2: Network structure diagram of text feature extraction.

$$f_{att}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V,$$

$$Q \in R^{n \times d_k}, \quad K \in R^{n \times d_k}, V \in R^{n \times d_v}, \qquad (2)$$

where $T$ represents the transpose operation of the matrix and dividing by $\sqrt{d_k}$ is to prevent the inner product result from being too large in the experiment. The self-attention mechanism essentially reencodes the input matrix into a new matrix after considering the global feature information through the convolution operation of the matrix. The self-attention mechanism assigns different weights to different vocabulary-level features and considers the global information of the vocabulary and the relationship between related associations so as to obtain the corresponding vocabulary features or weights.

### 3.2. Image Feature Extraction Network.

The AlexNet neural network is an 8-layer convolutional neural network with three hidden layers in Figure 3. Its basic structure includes an input layer, 5 convolutional layers, 3 pooling layers, 2 fully connected layers, and an output layer. The hidden layer consists of convolutional layers and pooling layers. Convolutional neural networks can obtain multidimensional features of images through convolution layers, and different
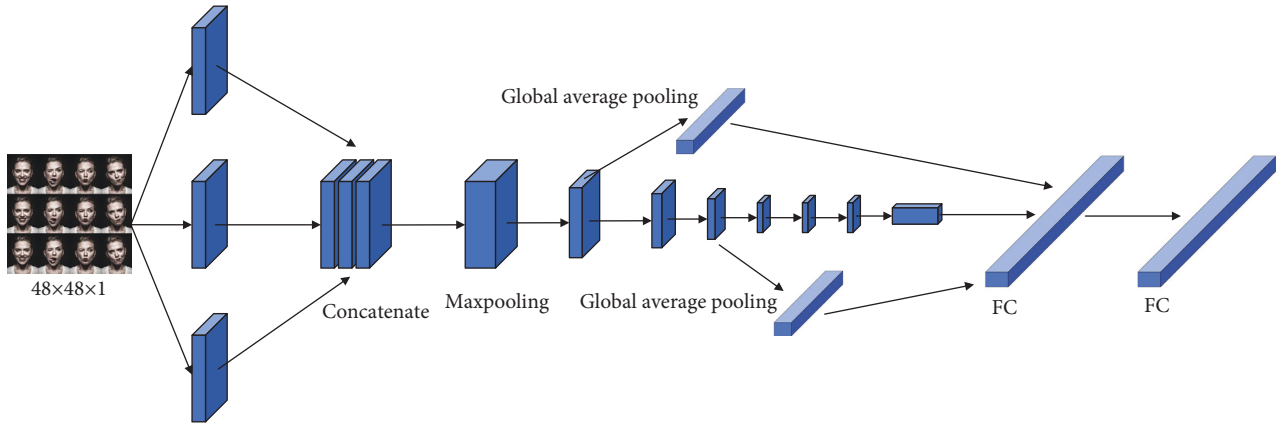
FIGURE 3: Network structure diagram of image feature extraction.

convolution kernels can extract different image features. For emotional images of faces, convolution kernels of different scales can be used for feature extraction, and feature information of different scales can be obtained. In this paper, various types of convolution kernels are used to obtain richer image features so that the feature information of the input data can be more accurately expressed.

The sizes of the multiscale convolution kernels in this paper are $1 \times 1$, $3 \times 3$, and $5 \times 5$, which extract features from the input image data, respectively, and add a BN layer at the end of the convolution layer to improve the multiscale convolution network. Finally, the obtained multidimensional expression features are connected and fused. This feature connection method ensures the richness of features. Among them, the $1 \times 1$ convolution kernel can organize information across channels, improve the expressive ability of the network model, and greatly enhance the nonlinear features while keeping the feature scale unchanged. At the same time, the channel can also be adjusted, the pixels on different channels are linearly combined, and then the nonlinear operation is performed to reduce the dimension and the number of parameters.

After cross-layer information fusion, the edge information is also enhanced while retaining the rich high-dimensional features of the feature map. In this paper, the first and second pooling layers of the network are connected to the fully connected layer, and the low-level feature information and high-level feature information are fused as the input of the softmax layer. It can effectively express the feature information of input data, alleviate the problem of gradient disappearance and explosion, enhance the propagation of features in different dimensions, and effectively utilize multidimensional feature information. Cross-layer connection fuses low-level features and high-level features to more accurately describe the feature information of the input data. However, too many parameters are generated in this way, which will cause parameter explosion and overfitting. There are many redundant features, which will reduce the training speed of the network and affect the accuracy of recognition. Therefore, a global average pooling layer is added at the end of the network model to average each

feature map, and finally, the result is input to the softmax layer.

### 3.3. LSTM Neural Network.
Because neural networks such as RNN retain previous data information when processing time-series-based data, the previous data will have less and less effect on the model with the subsequent data input; that is, there is a problem of long-distance dependence. In addition, some unimportant data will also be retained by neural network models such as RNN, resulting in data redundancy. In order to solve the above problems, this paper introduces the LSTM network structure. It has the characteristics of maintaining long-term memory and has good performance in processing time-series-based data. The network model structure is shown in Figure 4.

Equation (3) represents the forget gate of LSTM calculated from the input data and the hidden state. Equation (4) indicates that the input data and the weight of the forgetting gate are subjected to linear operation, and the output of the forgetting gate is obtained to represent the memory level of the long-term memory state. Equations (5) and (6) represent the input gate part of the LSTM.

$$C = \text{sigmoid}\left(W_1\left(x_{i-1}, h_{i-1}\right) + b_1\right), \tag{3}$$

$$C' = \text{sigmoid}\left(W_1\left(x_{i-1}, h_{i-1}\right) + b_1\right) \times c_i, \tag{4}$$

$$C_1'' = \text{sigmoid}\left(W_2\left(x_{i-1}, h_{i-1}\right) + b_2\right), \tag{5}$$

$$C_2'' = \text{sigmoid}\left(W_3\left(x_{i-1}, h_{i-1}\right) + b_3\right). \tag{6}$$

Through the calculation of the above formula, the LSTM network model can complete the update of the existing long-term memory elements, as shown in (7) and (8). Finally, the output gate, LSTM network model is improved. The LSTM network model comprehensively considers the influence of two aspects of the current long-term memory and the current input data elements in the output, as shown in formulas (9) and (10). Finally, the result of the long-term memory is activated by the tanh function, which finally represents the size and positive or negative of the actual output of the LSTM.
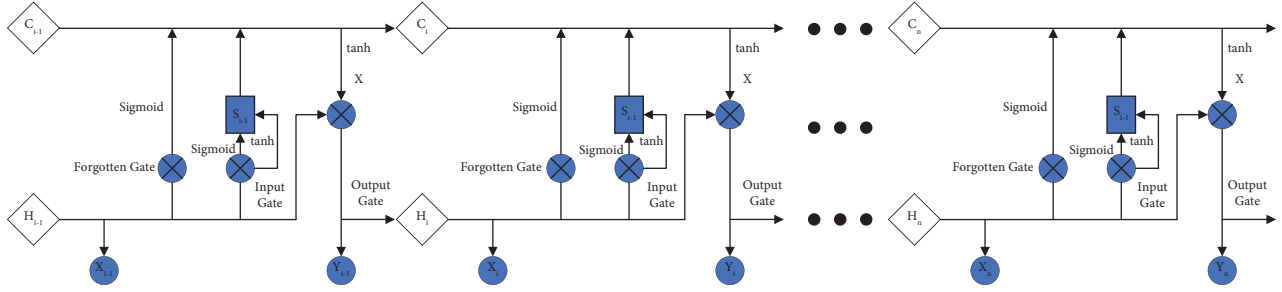
FIGURE 4: LSTM neural network structure diagram.

$$C'' = C_1'' \times C_2'', \tag{7}$$

$$c_i = C' \times C'', \tag{8}$$

$$y' = \text{sigmoid}\left(W_3\left(x_{i-1}, h_{i-1}\right) + b_3\right), \tag{9}$$

$$y_{i-1} = y' \times \tanh\left(c_i\right). \tag{10}$$

## 4. Experimental Results and Analysis

*4.1. Experimental Setup.* In terms of text feature extraction and processing, this paper uses the ACE2005 English prediction library as training data. It mainly annotates information such as event mentions, event trigger words, and event elements and is widely used in feature extraction tasks in English. It divides the dataset into training, validation, and test sets in a roughly $8:1:1$ ratio. At the same time, the precision rate, recall rate, and $F1$ value are selected as important indicators for model performance evaluation.

In terms of image datasets, this paper uses the COCO large-scale dataset to train and test the network model. The main purpose of image training is to add emotional information to the newly generated text.

The text uses Adam optimizer to optimize the parameters of the neural network model. The initial learning rate is $10^{-3}$, which is gradually reduced to $10^{-4}$ and $10^{-5}$. By adjusting the learning rate, the loss function value of the network model is continuously reduced, and the prediction accuracy is continuously improved until a stable equilibrium state is reached.

There are many types of human emotions, and different datasets have different types. In order to prove the effectiveness of the proposed text reconstruction method for college English textbooks, this paper screened these emotion types. Both LSTM and GRU units are capable of processing sequences of events. In general, the LSTM prediction effect is better, and the GRU unit is actually more efficient. In this paper, more attention is paid to the generation effect of the text content, so the LSTM unit is adopted.

*4.2. Evaluation Indicators.* The loss function value is an important indicator for evaluating the actual performance of the model. When training a neural network, in order to improve the prediction effect of the network model, the loss

function value is often reduced to a very low level in order to obtain more ideal network model parameters. The loss function is an important reference index to measure the performance of the neural network. Usually, the smaller the loss function value in the test set, the better the function model. For the text reconstruction task of college English textbooks, we use the MSE mean square loss function as the loss function in this paper to fit the regression problem of the predicted value.

$$\text{MSE} = \frac{1}{n}\sum_{i=0}^{n}\left(\hat{y}_i - y_i\right)^2. \tag{11}$$

Text or images based on time series usually contain hidden relationships. Through the training of neural network models, latent features between continuous data can be discovered to realize data prediction. In this paper, the MSE loss function is used to continuously adjust the parameters of the network model in reverse to improve the prediction effect of the network model.

In addition, in order to evaluate the prediction effect of the network model, this paper also adopts the MAE mean absolute error as the evaluation index of the model. MAE represents the mean absolute value of the error between the predicted value and the true value. Compared with MSE, MAE can directly reflect the difference with the original data. The smaller the MAE of the network model test data, the better the prediction result of the model.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|. \tag{12}$$

*4.3. Comparison and Analysis of Experimental Results.* As shown in Figure 5, A–E represent the number of emotion categories in the image dataset in the subnetwork, which are angry, natural, fearful, happy, and sad, respectively. Correspondingly, the percentages in the figure represent the proportion of each type in the image dataset. From the figure, we can clearly see that the number of types B is the largest and the number of types E is the least. Natural and normal images make up the largest proportion of our dataset, and sad images make up the least. Image emotional feature extraction network is mainly trained on this dataset to extract emotional features in images.

Figure 6 shows the distribution of sentiment features in the text dataset. This paper mainly divides text sentiment features into two categories: one is positive sentiment, and
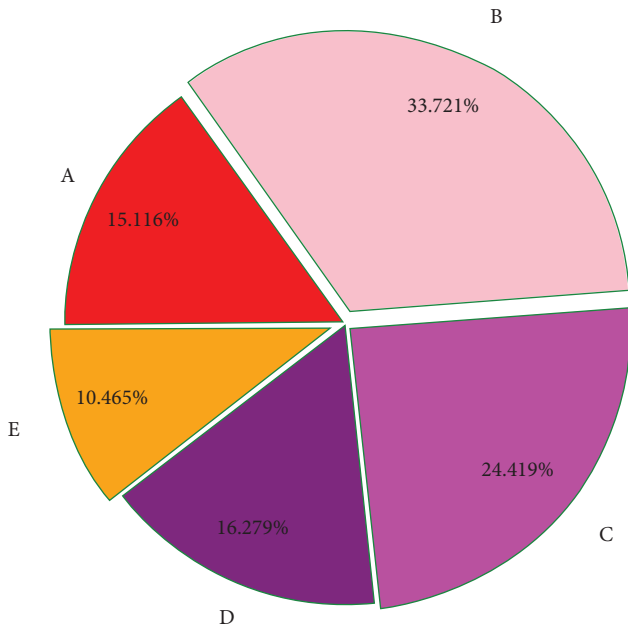
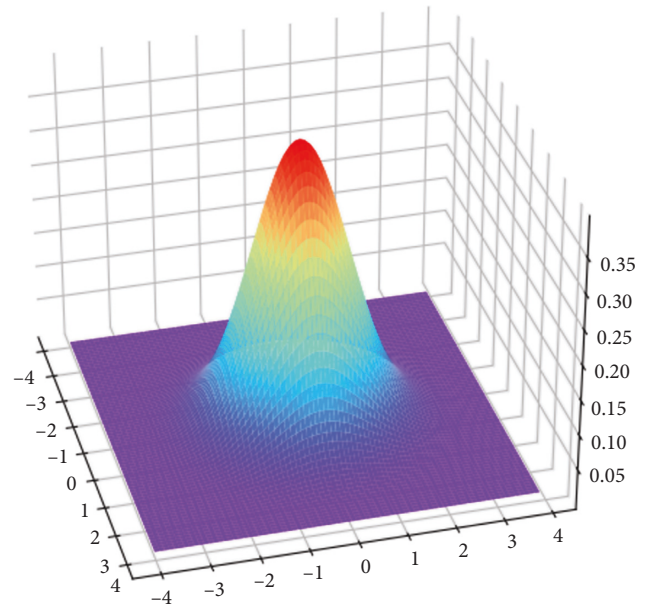Figure 5: Proportion of emotion categories in image datasets.



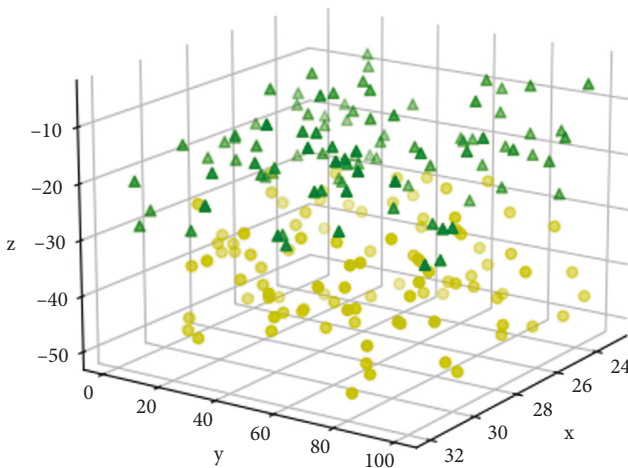Figure 7: Network model optimization learning.



Figure 6: Sentiment dispersion in text datasets.



Figure 8: Confusion matrix for image network model learning.

the other is negative sentiment. Positive emotions include joy, enthusiasm, and confidence, and negative emotions include anger, fear, and sadness. In the figure, we use green to represent positive sentiment in the text and yellow to represent negative sentiment in the text dataset. On the whole, the demarcation line between negative emotion and positive emotion is relatively clear, which is suitable for model training and testing.

Figure 7 shows the results of network model optimization learning. The two axes in the plane represent the model parameter settings of the image feature extraction and text feature extraction subnetworks. From the figure, we can see that the parameters of the network model can be optimized by setting some parameters in the image and text feature extraction subnetworks.

Figure 8 shows the learning effect of the image network model. The lighter the color value, the better the prediction effect of the model. From the above figure, we can clearly see
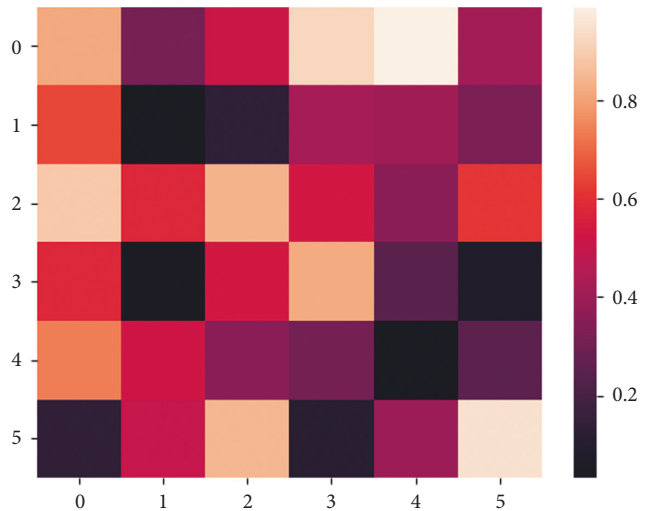
that the network model in this paper has a certain accuracy in distinguishing emotion categories, which meets the requirements of image feature extraction.

Figure 9 shows the learning effect of the text feature extraction network. The deeper the red value of the color value, the worse the prediction effect, and the darker the blue value, the better the prediction effect. The text feature extraction network in this paper expands the category of emotion and has been able to classify 10 emotion types more accurately.

Figure 10 shows the learning effect of the backbone network model. The $x$-axis and $y$-axis, respectively, represent the number of iterations of the subnetwork model, which is the number of times of training in different data sets. The $z$-axis represents the effect of learning the backbone network model together with the two subnetwork models. From the
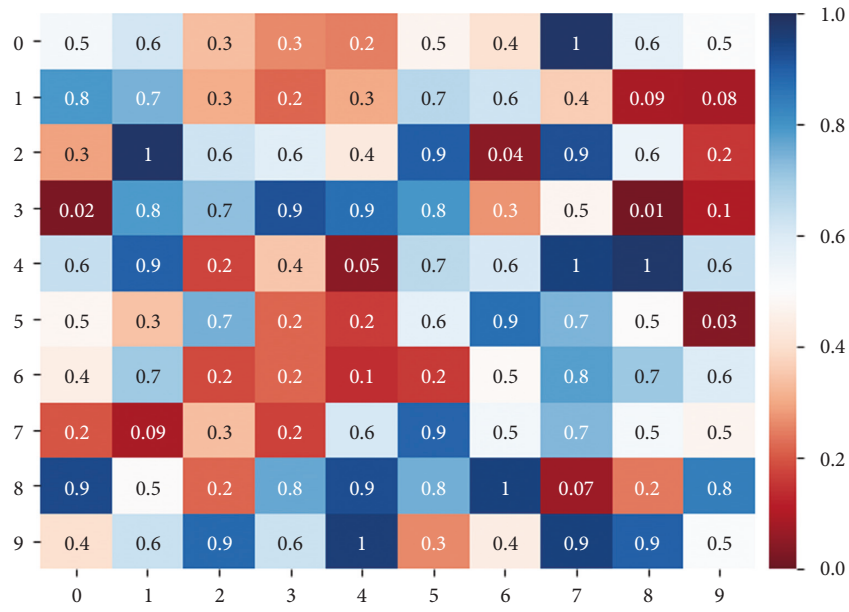
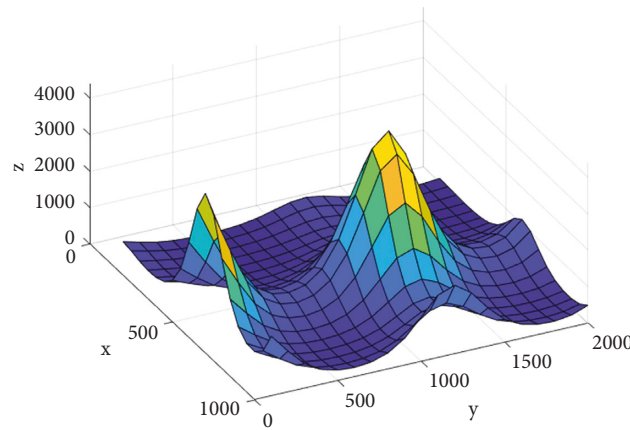Figure 9: Prediction effect of text feature network model.



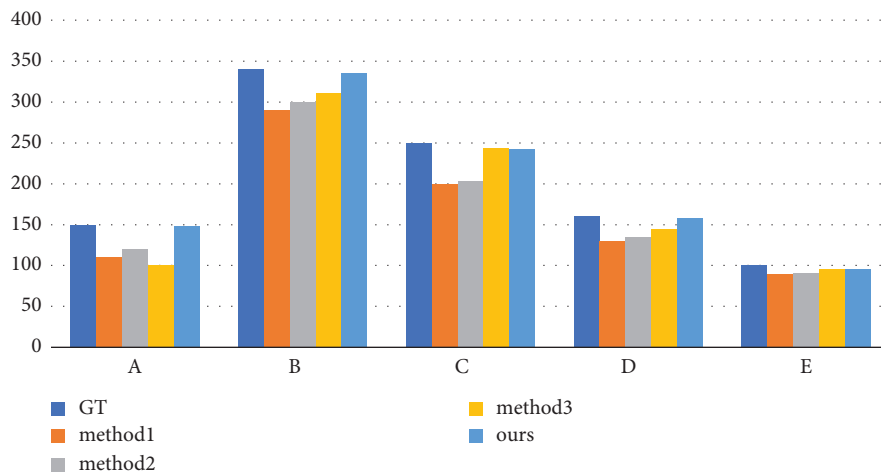Figure 10: The learning effect of the backbone network model.



Figure 11: Comparison of various neural network models.

figure, we can see that multiple experiments and settings through multiple network models can effectively improve the generation effect of the entire network model.

This paper compares the methods of various neural network models, and the experimental results are shown in Figure 11. It should be noted that all experiments are trained and tested on the same test platform and only the structure of the network model is different. All experiments are trained from scratch with the same learning rate and optimizer. GT represents the real value of the data predicted by the network model, and other colors represent the experimental results obtained by CNN, RNN, transformer, and the method in this paper, respectively. From the data, we can clearly see that the method in this paper has a better prediction effect on the data and is closer to the true value.

## 5. Summary

In order to improve the learning effect of college English, this paper proposes a text reconstruction method for college English textbooks from the perspective of language images. In this paper, the text feature extraction network and image feature extraction network are used to obtain text features and image features, respectively. Then, this paper designs a multimodal feature fusion network and a college English text generation network, which mainly guides the reconstruction of new college English texts according to the emotional features in image texts, generates new texts of college English textbooks with emotions, and improves college English learning effect. Through a large number of experiments, it is proved that the reconstruction method of college English textbooks based on the perspective of language and images can improve the learning ability of college students and reduce the teaching burden of college teachers. The neural network model adopted by the text can handle multimodal data and can be extended to more data types in future work, such as audio data and video data.

## Data Availability

The dataset can be obtained from the corresponding author upon request.

## Conflicts of Interest

The author declares that there are no conflicts of interest.

## References

[1] H. Dehghanzadeh, H. Fardanesh, J. Hatami, and O. TalaeeNoroozi, "Using gamification to support learning English as a second language: a systematic review," *Computer Assisted Language Learning*, vol. 34, no. 7, pp. 934–957, 2021.

[2] A. Hashemi and S. I. Kew, "The effects of using blended learning in teaching and learning English: a review of literature[J]," *The Eurasia Proceedings of Educational and Social Sciences*, vol. 18, pp. 173–179, 2020.

[3] F. Nappi and C. Spadaccio, "Keep fumbling around in the dark when it comes to infective endocarditis, or produce new, reliable data to redesign the guidelines?" *The Journal of Thoracic and Cardiovascular Surgery*, vol. 155, no. 1, pp. 75-76, 2018.

[4] D. J. Fremont, J. Chiu, D. D. Margineantu, D. Osipychev, and S. A. Seshia, *Formal Analysis and Redesign of a Neural Network-Based Aircraft Taxiing System with VerifAI[C]//International Conference on Computer Aided Verification*, Springer, Cham, pp. 122–134, 2020.

[5] K. Wijekumar, B. J. F. Meyer, P. Lei, and D. L. HernandezAugust, "Improving content area reading comprehension of Spanish speaking English learners in Grades 4 and 5 using web-based text structure instruction," *Reading and Writing*, vol. 31, no. 9, pp. 1969–1996, 2018.

[6] J. M. Vargas Vásquez and R. Zuñiga Coudin, "Graphic organizers as a teaching strategy for improved comprehension of argumentative texts in English[J]," *Actualidades Investigativas en Educación*, vol. 18, no. 2, pp. 32–54, 2018.

[7] B. Setyono and H. P. Widodo, "The representation of multicultural values in the Indonesian Ministry of Education and Culture-Endorsed EFL textbook: a critical discourse analysis," *Intercultural Education*, vol. 30, no. 4, pp. 383–397, 2019.

[8] H. P. Widodo, "A critical micro-semiotic analysis of values depicted in the Indonesian ministry of national education-endorsed secondary school English textbook," *Situating Moral and Cultural Values in ELT Materials*, Springer, Cham, pp. 131–152, 2018.

[9] M. Delphine, N. Sylvestre, N. Gabriel, and N. Wenceslas, "A psychometric analysis of the study skills questionnaire for university of Rwanda undergraduate students at national police college," *Creative Education*, vol. 13, no. 03, pp. 862–885, 2022.

[10] T. Bonsaksen, M. C. Småstuen, M. M. Thørrisen, and T. FongLimBrown, "Factor analysis of the Approaches and Study Skills Inventory for Students in a cross-cultural occupational therapy undergraduate student sample," *Australian Occupational Therapy Journal*, vol. 66, no. 1, pp. 33–43, 2019.

[11] I. A. Krishnan, G. D. Mello, S. A. Kok, and V. N. SabapathyMunianChingKandasamy RamalingamBaskaranKanan, "Challenges faced by hearing impairment students during COVID-19," *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, vol. 5, no. 8, pp. 106–116, 2020.

[12] A. Syakur, L. Musyarofah, S. Sulistiyaningsih, and W. Wike, "The effect of project based learning (PjBL) continuing learning innovation on learning outcomes of English in higher education," *Budapest International Research and Critics in Linguistics and Education (BirLE) Journal*, vol. 3, no. 1, pp. 625–630, 2020.

[13] X. Wang, Y. Liu, and B. Ying, "The effect of learning adaptability on Chinese middle school students' English academic engagement: the chain mediating roles of foreign language anxiety and English learning self-efficacy[J]," *Current Psychology*, vol. 6, pp. 1–11, 2021.

[14] T. Xiong and Z. m Yuan, ""It was because I could speak English that I got the job": neoliberal discourse in a Chinese English textbook series," *Journal of Language, Identity and Education*, vol. 17, no. 2, pp. 103–117, 2018.

[15] A. Syakur, E. Junining, and M. K. Mubarok, "Developing English for specific purposes (ESP) textbook for pharmacy students using on-line teaching in higher education[J]," *Britain International of Linguistics Arts and Education (BIoLAE) Journal*, vol. 2, no. 1, pp. 467–474, 2020.

[16] H. Gong, Y. Sun, and X. Feng, "Tablegpt: few-shot table-to-text generation with table structure reconstruction and

content matching[C]," in *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 1978–1988, Spain, 2020.

[17] T. M. Paixo, R. F. Berriel, M. C. S. Boeres et al., "Fast (er) reconstruction of shredded text documents via self-supervised deep asymmetric metric learning[C]," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14343–14351, Seattle, WA, USA, 2020.

[18] M. Brunton and C. Cook, "Dis/Integrating cultural difference in practice and communication: a qualitative study of host and migrant Registered Nurse perspectives from New Zealand," *International Journal of Nursing Studies*, vol. 83, pp. 18–24, 2018.

[19] H. Forbes-Mewett, K. Hegarty, and R. Wickes, "Regional migration and the local multicultural imaginary: the uneasy governance of cultural difference in regional Australia," *Journal of Ethnic and Migration Studies*, vol. 6, pp. 1–18, 2021.

[20] I. W. Suryasa, I. N. Sudipa, I. A. M. Puspani, and I. M. Netra, "Translation procedure of happy emotion of English into Indonesian in kṛṣṇa text," *Journal of Language Teaching and Research*, vol. 10, no. 4, p. 738, 2019.

[21] W. Suryasa, I. N. Sudipa, and I. A. M. Puspani, "Towards a change of emotion in translation of kṛṣṇa text[J]," *Journal of Advanced Research in Dynamical and Control Systems. Special Issue on Social Sciences*, vol. 11, no. 2, pp. 1221–1231, 2019.

[22] B. Vermote, N. Aelterman, W. Beyers, and M. AperBuysschaertVansteenkiste, "The role of teachers' motivation and mindsets in predicting a (de)motivating teaching style in higher education: a circumplex approach," *Motivation and Emotion*, vol. 44, no. 2, pp. 270–294, 2020.

[23] K. J. Bartholomew, N. Ntoumanis, A. Mouratidis, and S. KatartziThøgersen-NtoumaniVlachopoulos, "Beware of your teaching style: a school-year long investigation of controlling teaching and student motivational experiences," *Learning and Instruction*, vol. 53, pp. 50–63, 2018.

[24] Y. Ding, D. Zeng, M. Li et al., "Towards efficient human-machine collaboration: real-time correction effort prediction for ultrasound data acquisition," *Medical Image Computing and Computer Assisted Intervention - MICCAI 2021*, Springer, in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 461–470, 2021.

[25] D. Goldmann, S. Schramm, and M. Galek, "Virtual double pulse tests to reduce measuring time and effort in semiconductor loss modeling[C]," in *Proceedings of the PCIM Europe 2019; International Exhibition and Conference for Power Electronics, Intelligent Motion*, pp. 1–7, Renewable Energy and Energy Management. VDE, 2019.

[26] S. Dargan, M. Kumar, M. R. Ayyagari, and G Kumar, "A survey of deep learning and its applications: a new paradigm to machine learning," *Archives of Computational Methods in Engineering*, vol. 27, no. 4, pp. 1071–1092, 2020.

[27] M. Veres and M. Moussa, "Deep learning for intelligent transportation systems: a survey of emerging trends," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 8, pp. 3152–3168, 2020.

[28] A. A. Abdulrahman, M. Rasheed, and S. Shihab, "The analytic of image processing smoothing spaces using wavelet," *Journal of Physics: Conference Series*, vol. 1879, no. 2, 2021.

[29] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, "Comparison of full-reference image quality models for optimization of image processing systems," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1258–1281, 2021.

[30] P. Pandiarajaa, "A survey on machine learning and text processing for pesticides and fertilizer prediction[J]," *Turkish Journal of Computer and Mathematics Education (TURCO-MAT)*, vol. 12, no. 2, pp. 2295–2302, 2021.

[31] T. O'Reilly, D. G. Feng, D. J. Sabatini, and D. J. WangGorin, "How do people read the passages during a reading comprehension test? The effect of reading purpose on text processing behavior," *Educational Assessment*, vol. 23, no. 4, pp. 277–295, 2018.

[32] L. Stappen, A. Baird, and L. Christ, "The MuSe 2021 multimodal sentiment analysis challenge: sentiment, emotion, physiological-emotion, and stress[M]," in *Proceedings of the 2nd on Multimodal Sentiment Analysis Challenge*, pp. 5–14, 2021.

[33] J. Tang, K. Li, and X. Jin, "CTFN: hierarchical learning for multimodal sentiment analysis using coupled-translation fusion network[C]," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, vol. 1, pp. 5301–5311Long Papers, Thailand, 2021.

[34] B. Zhang, F. Yu, and Y. Gao, "Joint learning for relationship and interaction analysis in video with multimodal feature fusion[C]," in *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 4848–4852, China, 2021.