Hindawi
Mobile Information Systems
Volume 2021, Article ID 1874584, 8 pages
https://doi.org/10.1155/2021/1874584

Research Article

Evaluation Model of College English Multimedia Teaching Effect Based on Deep Convolutional Neural Networks

Limei Geng

Department of Foreign Language, Jingdezhen University, Jingdezhen 333000, Jiangxi, China

Correspondence should be addressed to Limei Geng; limeigeng2021@126.com

Received 19 May 2021; Revised 18 June 2021; Accepted 1 July 2021; Published 28 July 2021

Academic Editor: Fazlullah Khan

With the acceleration of global integration, the demand for English instruction is increasingly rising. On the other hand, Chinese English learners struggle to learn spoken English due to the limited English learning environment and teaching conditions in China. The advancement of artificial intelligence technology and the advancement of language teaching and learning techniques have ushered in a new era of language learning and teaching. Deep learning technology makes it possible to solve this problem. Speech recognition and assessment technology are at the heart of language learning, and speech recognition technology is the foundation. Because of the complex changes in speech pronunciation, a large amount of speech signal data, the high dimension of speech characteristic parameters, and a large amount of speech recognition and evaluation computation, the large volume of speech signal processing requires higher requirements of hardware and software resources and algorithms. However, traditional speech recognition algorithms, such as dynamic time-warped algorithms, hidden Markov models, and artificial neural networks, have their advantages and disadvantages. They have encountered unprecedented bottlenecks, so it is difficult to improve their accuracy and speed. To solve these problems, this paper focuses on evaluating the multimedia teaching effect of college English. A multilevel residual convolutional neural network algorithm for oral English pronunciation recognition is proposed based on a deep convolutional neural network. The experiments show that our algorithm can assist learners in identifying inconsistencies between their pronunciation and standard pronunciation and correcting pronunciation errors, resulting in improved oral English learning performance.

1. Introduction

The demand for English learning [1] in China is increasingly rising due to global integration and China's increasing degree of internationalization. The tremendous Chinese pronunciation characteristics and the difference with English pronunciation with time and location constraints cause the lack of a domestic English learning environment. Also, good English teachers and standard classroom teaching [2] cannot fulfill the English learning needs for various reasons. These reasons caused comprehensive English teaching and learning to be a big problem for the people. English as a second language has become a study hotspot in the field of education. Artificial intelligence (AI)-enabled [3, 4] learning has solved this problem with the advancement of computer science [5] and technology [6] and improvements in language teaching [7, 8] and learning methods. This technology will disrupt the current language teaching and learning environment, allowing learners to learn independently and in any place. It will provide learners with reliable, objective, and timely pronunciation feedback and direct them; learners will also benefit from repeated listening to determine the difference between its pronunciation and standard pronunciation, correct their pronunciation errors, and improve their language learning performance [9–12].

The key to evaluating college English multimedia teaching [13, 14] is the recognition of spoken English. It refers to the technology of converting speech signals [15] into corresponding objects or texts by automatic recognition using AI and machine learning [16–19]. Speech recognition and assessment technologies have advanced rapidly in recent years, thanks to advances in deep learning, big data, and cloud computing technologies. As a result, it can model human neurons to interpret data through multilayer depth

transmission, which has been verified in speech recognition, and deep neural networks (DNNs) display outstanding advantages in solving complex problems. The rapid growth of graphing calculators and cloud computing technology has reduced the computational complexity of DNN [20, 21]. As a result, research on multimedia teaching and its assessment using technologies focused on deep learning will significantly increase English class teaching abilities.

On the one hand, most domestic English learners use portable devices to aid in oral English learning, such as language repeaters [22], MP3 players, and cell phones [23]. However, these tools do not conduct voice recognition and are restricted to inquiry and follow-up reading. The role, however, is unable to provide learners with a fair and objective pronunciation evaluation and feedback. On the other hand, in evaluating oral English, the current oral English test is still based on manual scoring with a strong subjective will, different standards, and slow speed. The problem is due to differences in the knowledge structure and experience of scoring experts and even differences in the decisions of the same expert [11, 24–27]. There are deviations in the evaluation of the same pronunciation. Subjective factors such as this lead to poor repeatability and stability of manual evaluation. Moreover, the use of manual scoring will consume a lot of manpower and material resources. Based on the above observations, this paper focuses on evaluating college English multimedia teaching effects as the primary research content. Therefore, in this paper, a multilevel residual convolutional neural network is proposed. The proposed scheme is used to recognize spoken English pronunciation based on the deep convolutional neural network [28–32]. The proposed algorithm has been tested, can help learners distinguish between their own and standard pronunciation, correct pronunciation errors, and improve the efficiency of oral English learning. The paper's key contributions are as follows:

(1) Based on a deep convolutional neural network, a multimedia-based English teaching [33, 34] impact evaluation model is proposed. It helps learners distinguish between their own and standard pronunciation, correct pronunciation errors, and increase the quality of oral English learning.

(2) This paper proposes a novel multilevel residual convolutional neural network, making up for the missing features to improve the recognition rate.

(3) Finally, the superiority of this method is proved through comparative experiments.

The rest of the paper is organized as follows. In Section 2, related work is studied. In Section 3, the methodology is given, whereas results and discussion are explained in Section 4. Finally, Section 5 concludes the paper.

2. Related Work

Different aspects of related work have been studied in this section, such as language recognition, speech processing, and feature extraction mechanisms.

2.1. Spoken Language Recognition Process. The general process of speech recognition [35, 36] is shown in Figure 1. First, the computer's sound card is used to digitize the voice analog signal and collect the voice signal. According to the Nyquist sampling theorem, in the process of analog/digital signal conversion, the sampling frequency $f_{s\_\max}$ is greater than 2 times the highest $F_{\max}$ in the signal, as shown in the following equation:

$$f_{s\_\max} \geq 2 * F_{\max}. \tag{1}$$

Then, the sampled digital signal can express the adequate information in the original speech signal more completely. Since the frequency of everyday speech is generally between 40 and 4000 Hz, the sampling frequency is set as 8 kHz in this paper. Then, the obtained speech signal is preprocessed, which includes preweighting, framing, winding, and endpoint detection. Then, the characteristic parameters of the preprocessed speech signal are extracted. Finally, the speech feature parameters can be selected for model training or pattern matching.

2.2. Spoken Speech Signal Preprocessing. Under the influence of oral and nasal radiation and glottis excitation, the high-frequency end of the average power spectrum of the speech signal is attenuated by 6 dB/OCT at about 800 Hz. Therefore, before the speech signal analysis, a 6 dB/OCT high-frequency lifting preweighted digital filter is generally adopted to enhance the high-frequency part of the speech signal so that the spectrum of the speech signal becomes flat. The spectrum of the whole frequency band from low frequency to high frequency can be obtained with the same signal-to-noise ratio. The calculation equation of the filter response function is as follows:

$$H(z) = 1 - \partial z^{-1}, \tag{2}$$

where ∂ is the pre-emphasis coefficient, which is taken as 0.9375 in this paper. In this way, the result $y(n)$ after pre-emphasis processing can be expressed by the input speech signal $x(n)$ as follows:

$$y(n) = x(n) - \partial x(n-1). \tag{3}$$

2.3. Feature Extraction. Voice signal function parameter extraction aims to remove redundant data that are not important to voice processing and analyze and process the voice signal. The original speech signal has a large amount of data. It has too much information that interferes with the semantics due to the difference of the speakers, the loudness, and the length of the sound. Hence, it is not suitable for direct use in speech processing. The quality of feature parameters directly impacts speech processing efficiency, and a suitable feature extraction method will yield better results. As a result, function parameters from the original voice signal must be extracted. The ideal voice function describes only semantic information, and the total amount of voice data is also tiny.

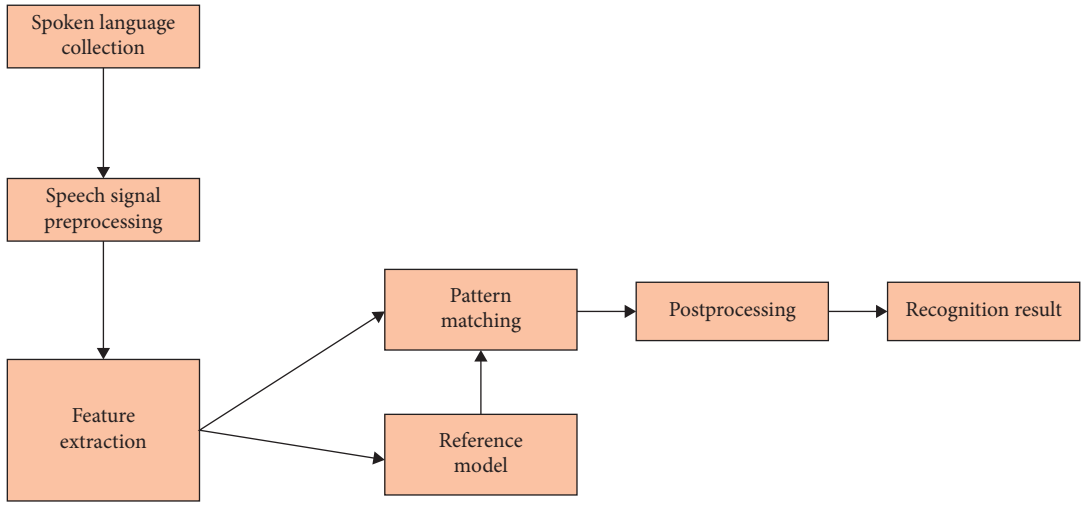

Figure 1: Spoken language recognition process.

3. Methodology

In this section, different stages of the methodology are discussed. The details about the convolutional neural network and the proposed multilevel residual convolutional neural network model are elaborated.

3.1. Convolution Neural Network. A convolution neural network (CNN) is put forward based on sparse interaction, parameters, share, and the critical thought of three identical mappings. It works through the set the convolution kernel size, and step length can be implemented in a small area in the image. In this way, the local characteristics of the image are extracted. Setting the pooling layer to dimension reduction of convolution image effectively reduces the number of network weights and improves the training efficiency. CNN will extract features at low, mid, and high levels in the image. Each convolution kernel in the convolution layer is equivalent to a feature extractor. The more the layers in a network there are, the more the features that can be deleted at different levels. The structure of CNN is shown in Figure 2.

A convolutional layer, an activation layer, a pooling layer, a completely connected layer, and an output layer are usually included in a CNN. CNN's central component is the convolution layer. Feature extraction at different levels can be achieved by setting the convolution kernel. The calculation equations between CNN convolution kernels are as follows:

$$y_l = w_l y_{l-1} + b_l, \tag{4}$$

where y_l represents the output of the l-th convolutional layer, w_l represents the tensor of the convolution kernel in the l-th convolutional layer, y_{l-1} represents the convolution output of the $l-1$-th convolutional layer, and b_l represents the bias of the l-th convolutional layer.

The down-sampling layer is also known as the pooling layer. The pooling layer will reduce the input function map's dimensionality and remove the key features, minimizing overfitting to some degree. The average pooling method adds up all the values in the window to average. It uses the average as the final sampling value. The calculation process is shown in Figure 3(a). The maximum pooling method takes the maximum value in each window as the final. The sampling value and the calculation process are shown in Figure 3(b). It is worth noting that the maximum pooling method is used.

The completely linked layer is the essential structure for classifying the high-level feature data obtained by the convolutional and pooling layers. The input feature map after convolution and pooling operations is compressed into a one-dimensional matrix, and then the one-dimensional matrix is input to the fully connected layer for training to realize the learning and memory of the target information. A large number of end-to-end neurons connect the fully connected layer. Each neuron is equivalent to a memory unit. Pattern recognition and classification can be realized through reasonable parameters such as weights and biases. The training process optimizes each neuron's parameters and bias values by forwarding propagation and back propagation algorithms and determining the optimal parameters.

3.2. Multilevel Residual Convolutional Neural Network Model. CNN can handle grid data such as images well. Each convolution kernel can extract information such as image texture and edge features from different levels and improve the recognition efficiency by increasing the number of convolution layers. However, with the deepening of the convolutional layer, the extracted feature information becomes less and less semantic. Problems such as loss of original feature information are prone to occur, resulting in slower training convergence speed and difficulty in improving the recognition rate. Based on the above analysis, this article improves the typical CNN structure and designs a multilevel residual convolutional neural network, as shown in Figure 4.

The proposed multilevel residual convolutional neural network contains multiple convolutional pooling layers and

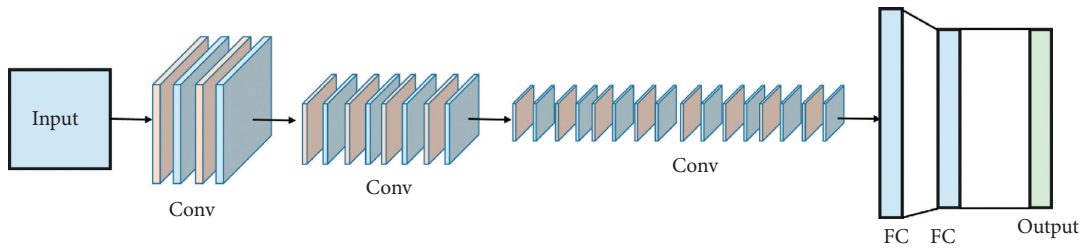

Figure 2: Schematic diagram of the structure of CNN.

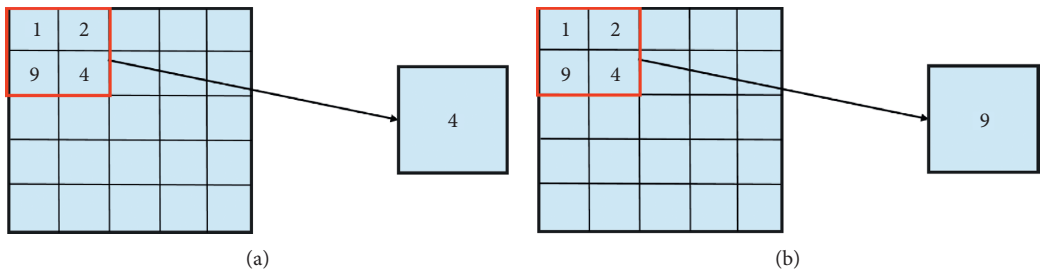

Figure 3: Pooling methods: (a) average pooling; (b) max pooling.

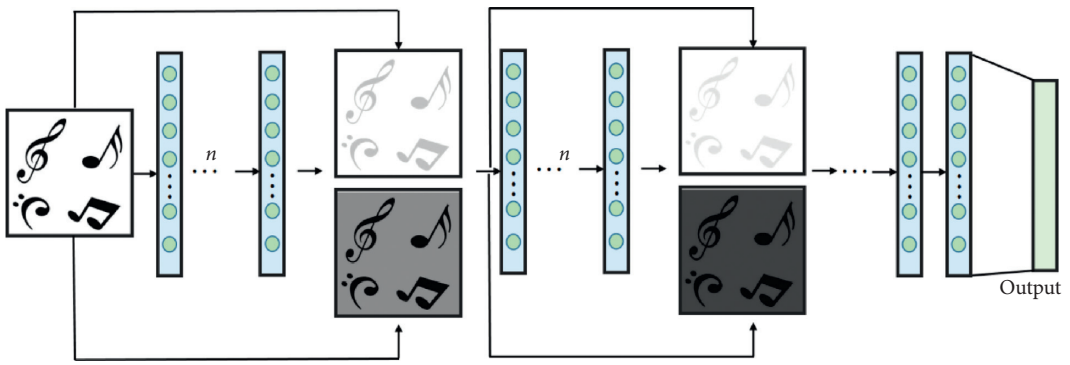

Figure 4: Multilevel residual CNN.

a multilevel residual structure. The multilevel residual structure can transmit original information across multiple convolutional layers to compensate for missing features. It represents the number of convolutional layers that the residual structure spans. The multilevel residual convolutional neural network designed in this paper further reduces the amount of calculation. It improves the recognition efficiency based on solving the shortcomings of the CNN structure.

The proposed model is improved using the residual structure by connecting the original information of the first n convolutional layers with the current layer. The structure of the multilevel residual retains the original information to the maximum extent. It adjusts the dimension of the original input features by adding control coefficients. Hence, it can effectively improve the recognition rate and accelerate the convergence rate. The principle is shown in Figure 5.

Assuming that the input when the residual structure is derived is x_i and the output after the residual is introduced is x_{i+n}, the output of the multilevel residual structure is as follows:

$$x_{i+1} = \sigma\left(w_{i+n}F\left(x_{i+n-1}\right) + b_{i+n} + \alpha x_i\right), \tag{5}$$

$$F\left(x_{i+n-1}\right) = \begin{cases} \sigma\left(w_{i+n}F\left(x_{i+n-1}\right) + b_{i+n}\right), & n \neq 1, \\ \beta x_i, & n = 1, \end{cases} \tag{6}$$

where α and β are control coefficients, which are used to limit the dimension of input features.

Assuming that the loss function is C, the weight update formula with backpropagation is as follows:

$$\frac{\partial C}{\partial x_i} = \frac{\partial C}{\partial x_{i+n}}\frac{\partial x_{i+n}}{\partial x_i} = \left[\left(\beta\prod_{k=1}^{n} w_{i+k} + \alpha\right)x_i + T\left(w, b\right)\right]\frac{\partial C}{\partial x_{i+n}}. \tag{7}$$

When the convolutional neural network is differentiated layer by layer, the weight w will gradually decrease or even approach zero. It will lead to the gradient update of back-propagation approaching zero, namely, the phenomenon of feature loss. The original feature information of the first n

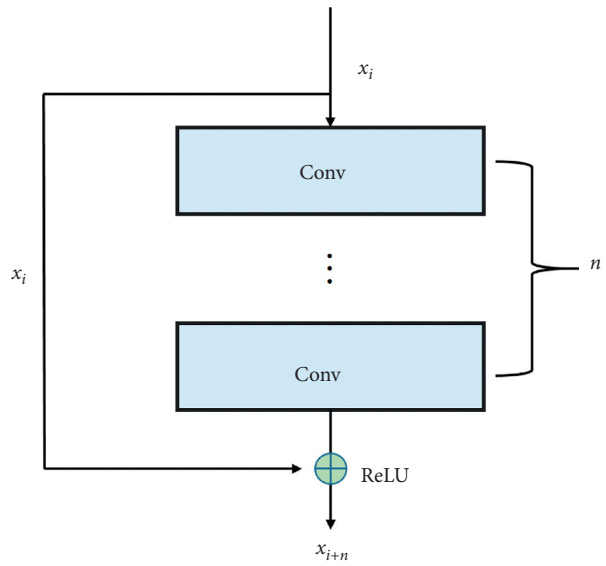

Figure 5: Multilevel residual structure.

convolutional layers can be introduced by adding a multilevel residual structure to supplement the features to the greatest extent. Meanwhile, the control parameters A and B can reduce the feature dimension, accelerate the training speed, and improve the training effect.

4. Experiments and Results

In this section, a detailed discussion on the experimental setup and result is performed. The following subsections, experimental setup, evaluation methods, datasets, experimental results, and model performance testing and analysis are performed.

4.1. Experimental Setup. This experiment uses the TensorFlow 2.0 toolkit and Matlab2018b to complete the construction of the network model, data preprocessing, and the realization of the training algorithm. The server platform configuration used for training is shown in Table 1.

4.2. Evaluation Methods. The main focus of the intonation test is to see if the material information in the pronunciation sentence is complete and correct. The MFCC coefficients based on the human ear hearing model are used as intonation assessment parameters in this paper, and the speech recognition model is developed using a deep belief network for speech recognition to assess if the content is complete and accurate. To judge the pronunciation, the correlation coefficient of the standard sentence and the MFCC function of the input sentence are both determined. Both intonation assessment and input are integrated on English pronunciation accuracy if it is consistent and fluent.

The term "speech speed" generally refers to the "pronunciation speed," which measures how quickly a speaker pronounces words. It can be expressed as the number of

Table 1: Server platform configuration.

Equipment	Model
GPU	GTX1080Ti 11Gx4
CPU	Intel Xeon E5-2665x2
CUDA	3584
OS	Ubuntu16.04 LTS

syllables N spoken in a unit of time T. It can be approximately calculated as the total speech length, including pauses. Since different speakers talk at different speeds, different people pronounce the same sentence differently depending on the length of the sentence. Furthermore, the speaker's emotional state affects speech tempo. For example, the pace of speech is generally slightly faster in anger and happiness than in a calm state. In contrast, the speed of speech is generally slower in the state of sadness. The length ratio A between the test sentence and the regular sentence is calculated in this paper using the speech rate evaluation based on the speech duration. The calculation equation is as follows:

$$\varphi = \frac{\text{Len}_{\text{std}}}{\text{Len}_{\text{test}}}, \tag{8}$$

where Len_{std} is the duration of the standard sentence and Len_{test} is the duration of the test sentence.

4.3. Data. The subjects of this paper are college students in our school, a total of 57 people, including 37 boys and 20 girls. Subjects were recorded by CoolEdit, a recording software, with a sampling rate of 16 kHz and 16-bit coding. The recording contains 10 sentences, all of which are commonly spoken English sentences.

4.4. Experimental Results. We compared the DHMM, CDHMM, TDA-GTS, and KASWT methods in the same experimental setting to somewhat check the superiority of the algorithm in this paper. The comparison results of their recognition rates are shown in Table 2.

The recognition rate of the proposed model is 97.17 percent, which is higher than the above models, as shown in Table 2 and Figure 6. As a result, the algorithm presented in this paper is both rational and accurate. It can be used to assess the impact of college English multimedia instruction.

4.5. Model Performance Testing and Analysis. Figures 7 and 8 show that the proposed model will reduce the loss function value below 0.2. At the same time, DHMM can only drop to about 0.5. The loss function values of the KASWT method drops to about 0.2 and stops converging. It demonstrates that our model performs better in terms of convergence. In general, this algorithm outperforms the compared methods in terms of convergence efficiency and speed in the experiment.

Table 2: Comparison of experimental results with different methods.

Methods	Accuracy
DHMM	0.9019
CDHMM	0.9415
TDA-GTS	0.9308
KASWT	0.9258
Ours	**0.9717**

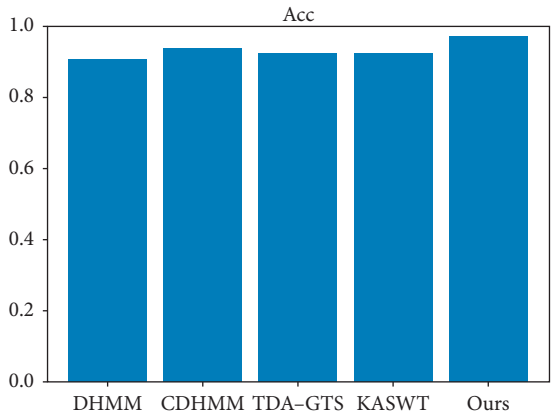

Figure 6: Histogram of comparison experiment.

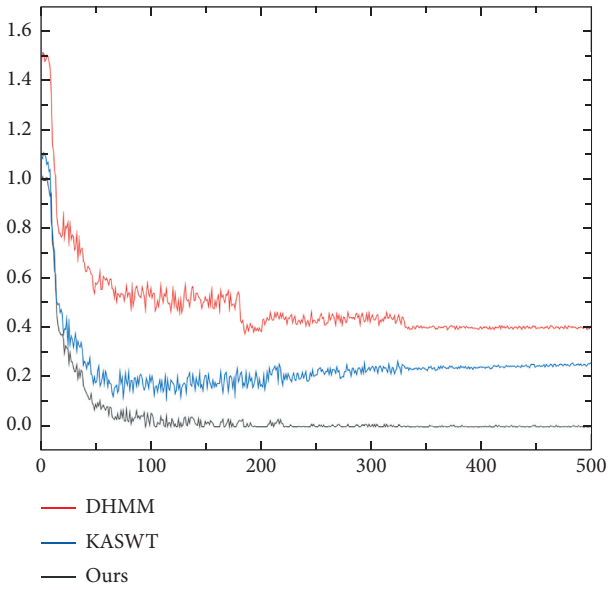

Figure 7: Comparison results of convergence performance.

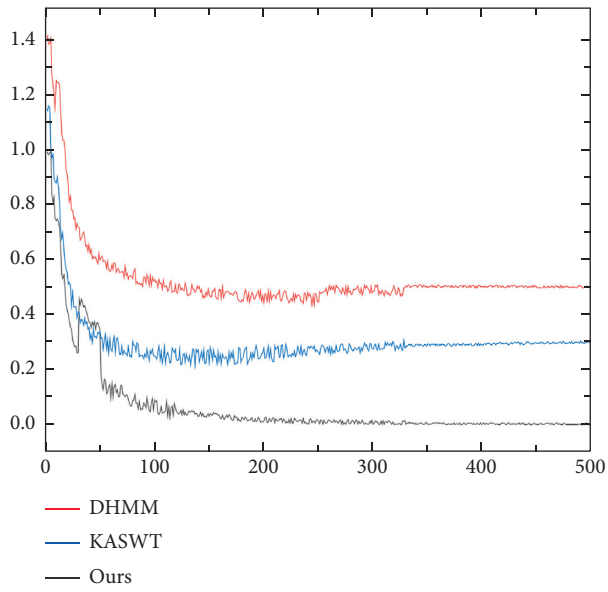

Figure 8: Comparison results of convergence speed.

5. Conclusion

Large-scale voice signal processing necessitates more challenging software and hardware resources and algorithms due to the complex changes in voice pronunciation, a large amount of data in voice signals, the high dimensionality of voice function parameters, and a large number of calculations for voice recognition and evaluation. Traditional speech recognition algorithms such as dynamic time warping, hidden Markov models, and artificial neural networks each have their own set of benefits and drawbacks. They have hit unheard-of bottlenecks, and it is impossible to boost their accuracy and pace anymore. In response to these issues, the emphasis of this article is on evaluating the effects of college English multimedia teaching. A multilevel residual convolutional neural network is proposed to assess spoken English pronunciation using. The proposed algorithm has been tested, which helps learners distinguish between their own and standard pronunciation, correct pronunciation errors, and increase the quality of oral English learning.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author declares there are no conflicts of interest.

Acknowledgments

This research was supported by the impact of online learning behavior on blended learning performance by the Leading Group of Jiangxi Provincial Educational Programming Research Topics (No. 17YB278).

References

[1] I. W. Suryasa, I. G. P. A. Prayoga, and I. W. A. Werdistira, "An analysis of students motivation toward English learning as second language among students in Pritchard English academy (PEACE)," *International Journal of Social Sciences and Humanities*, vol. 1, no. 2, pp. 43–50, 2017.

[2] M. T. Sathish, V. Sornaganesh, G. Sudha, and A. V. Chellama, "A study on shift of traditional classroom methods to online teaching methods in higher education scenario during lockdown," *International Journal of Multidisciplinary Research and Development*, vol. 7, no. 7, pp. 86–100, 2020.

[3] Y. Zhang, K. Cheng, F. Khan, R. Alturki, R. Khan, and A. U. Rehman, "A mutual authentication scheme for establishing secure device-to-device communication sessions in the edge-enabled smart cities," *Journal of Information Security and Applications*, vol. 58, 2021.

[4] F. Khan, A. U. Rehman, and M. A. Jan, "A secured and reliable communication scheme in cognitive hybrid ARQ-aided smart city," *Computers & Electrical Engineering*, vol. 81, 2020.

[5] Z. Huang, Y. Zhang, Q. Li et al., "Joint analysis and weighted synthesis sparsity priors for simultaneous denoising and destriping optical remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 6958–6982, 2020.

[6] F. Khan, A. U. Rehman, M. Usman, Z. Tan, and D. Puthal, "Performance of cognitive radio sensor networks using hybrid automatic repeat reQuest: stop-and-wait," *Journal of Mobile and Network Applications*, vol. 23, pp. 479–488, 2018.

[7] G. Kessler, "Technology and the future of language teaching," *Foreign Language Annals*, vol. 51, no. 2, pp. 205–218, 2018.

[8] K. Makhmudov, "Ways of forming intercultural communication in foreign language teaching," *Science and Education*, vol. 1, no. 4, 2020.

[9] N. Xu and W. H. Fan, "Research on interactive augmented reality teaching system for numerical optimization teaching," *Computer Simulation*, vol. 37, no. 11, pp. 209–212+304, 2020.

[10] M. W. Zhao, Z. M. Fan, and R. F. Wang, "Design of synchronous teaching system for hydraulic component theory

and experiment based on interconnection," *Machine Tool & Hydraulics*, vol. 48, no. 2, pp. 111–114, 2020.

[11] X. Yu, J. Yang, and Z. Xie, "Training SVMs on a bound vectors set based on Fisher projection," *Frontiers of Computer Science*, vol. 8, no. 5, pp. 793–806, 2014.

[12] L. J. Yao and L. Zhang, "Design and implementation of Chinese architecture history teaching system based on mixed reality technology," *Journal of Computer Applications*, vol. 39, no. 9, pp. 207–212, 2019.

[13] L. P. Jiang, X. P. Li, L. Zhang, J. Z. Chen, and Y. Y. Dong, "Design of virtual teaching experience space system based on wearable human body perception," *China Educational Technology*, vol. 407, no. 12, pp. 34–40+67, 2020.

[14] W. Gong, L. Tong, W. Huang, and S. Wang, "The optimization of intelligent long-distance multimedia sports teaching system for iot," *Cognitive Systems Research*, vol. 52, pp. 678–684, 2018.

[15] M. Sarma, P. Ghahremani, D. Povey, N. K. Goel, K. K. Sarma, and N. Dehak, "Emotion identification from raw speech signals using DNNs," in *Proceedings of the Interspeech 2018*, pp. 3097–3101, Hyderabad, India, September 2018.

[16] F. Khan, M. A. Jan, A. U. Rehman, S. Mastorakis, M. Alazab, and P. Watters, "A secured and intelligent communication scheme for IIoTenabled pervasive edge computing," *IEEE Transaction on Industrial Informatics*, vol. 17, no. 7, pp. 5128–5137, 2021.

[17] M. A. Jan, F. Khan, S. Mastorakis et al., "Lightweight and secure communication for energy-efficient IoT in health informatics," *IEEE Transactions on Green Communications and Networking*, 2021.

[18] J. Sun, F. Khan, J. Li, M. D. Alshehri, R. Alturki, and M. Wedyan, "Mutual authentication scheme for ensuring a secure device-to-server communication in the internet of medical things," *IEEE Internet of Things Journal*, 2021.

[19] W. Sun, P. Zhang, Z. Wang, and D. Li, "Prediction of cardiovascular diseases based on machine learning," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 30–35, 2021.

[20] Q. Chen, N. Li, X. Yang, R. Alturki, M. D. Alshehri, and F. Khan, "Impact of residual hardware impairment on the IoT secrecy performance of RIS-assisted NOMA networks," *IEEE Access*, vol. 9, pp. 42583–42592, 2021.

[21] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local-global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.

[22] K. Barkaoui, "Examining repeaters' performance on Second Language proficiency tests: a review and a call for research," *Language Assessment Quarterly*, vol. 14, no. 4, pp. 420–431, 2017.

[23] B. Klímová, "Mobile phones and/or smartphones and their apps for teaching English as a foreign language," *Education and Information Technologies*, vol. 23, no. 3, pp. 1091–1099, 2018.

[24] X. Yu, F. Jiang, J. Du, and D. Gong, "A cross-domain collaborative filtering algorithm with expanding user and item features via the latent factor space of auxiliary domains," *Pattern Recognition*, vol. 94, pp. 96–109, 2019.

[25] X. Yu, Y. Chu, F. Jiang, Y. Guo, and D. Gong, "SVMs classification based two-side cross domain collaborative filtering by inferring intrinsic user and item features," *Knowledge-Based Systems*, vol. 141, pp. 80–91, 2018.

[26] X. Yu, D. Zhan, L. Liu, H. Lv, L. Xu, and J. Du, "A privacy-preserving cross-domain healthcare wearables recommendation algorithm based on domain-dependent and domain-independent feature fusion," *IEEE Journal of Biomedical and Health Informatics*, p. 1, 2021.

[27] M. Yu, T. Quan, Q. Peng, X. Yu, and L. Liu, "A model-based collaborate filtering algorithm based on stacked AutoEncoder," *Neural Computing and Applications*, 2021.

[28] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020.

[29] W. Cai, Y. Song, and Z. Wei, "Multimodal data guided spatial feature fusion and grouping strategy for e-commerce commodity demand forecasting," *Mobile Information Systems*, vol. 2021, Article ID 5568208, 2021.

[30] J. Zhang, X. Jin, J. Sun, J. Wang, and A. K. Sangaiah, "Spatial and semantic convolutional features for robust visual object tracking," *Multimedia Tools and Applications*, vol. 79, no. 21, pp. 15095–15115, 2020.

[31] Y. Ding, X. Zhao, Z. Zhang, W. Cai, and N. Yang, "Graph sample and aggregate-attention network for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2021.

[32] Y. Gu, A. Chen, X. Zhang, C. Fan, K. Li, and J. Shen, "Deep learning based cell classification in imaging flow cytometer," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 2, pp. 18–27, 2021.

[33] L. Liang, Q. Yin, and C. Shi, "Exploring proper names online and its application in English teaching in university," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 24–29, 2021.

[34] J. Xiao, Y. Dai, and X. Shi, "Translation and influence of one two three... infinity in China," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 18–23, 2021.

[35] J. Park and S. Kim, "Noise cancellation based on voice activity detection using spectral variation for speech recognition in smart home devices," *Intelligent Automation & Soft Computing*, vol. 26, no. 1, pp. 149–159, 2020.

[36] J. Jo, H. Kim, I. Park, B. C. Jung, and H. Yoo, "Modified viterbi scoring for HMM-based speech recognition," *Intelligent Automation & Soft Computing*, vol. 25, no. 2, pp. 351–358, 2019.