Research on Application of Convolutional Neural Network Algorithm

Ling Ye 1, Peng Yin 2,*

1 School of Foreign Languages, Hubei Engineering University, Xiaogan, 432000, China
2 Information Technology Center, Hubei Engineering University, Xiaogan, 432000, China.

*Corresponding author e-mail: PengYin@cnooc.com.cn

Abstract. This paper proposes a method based on deep learning to improve the efficiency of children's English learning. Firstly, the migration method is used to improve the accuracy of Resnet18 model. Then the accuracy of the model was further improved by adding channel attention mechanism. Finally, the model fine-tuning method is used to reduce the computational load of the model, so as to facilitate the transplantation to embedded devices and improve the efficiency of the classification model. By comparing with the classical classification model, it is found that the classification model proposed in this paper has higher accuracy and fewer parameters, so it is suitable for transplantation into the English teaching of children.

Keywords: Transfer learning; attention mechanism; model fine-tuning; parameter number.

1. Preface
With the continuous development of society and the impact of the social environment caused by the current epidemic, early childhood education is very important, but it cannot meet its needs. In view of the contemporary social environment and social factors, this article proposes a smart, lightweight English alphabet Identify the classification method. In the traditional sense, English teaching is limited to the teacher's teaching, and the students follow to learn, and its efficiency is not high [1].

2. Neural network model
This article refers to the residual network model Resnet18 [2], and applies the classification model to the teaching of English alphabets for children, so that 52 English alphabets can be used for kindergarten children to learn English directly, thereby improving the quality and efficiency of English teaching and liberating The teacher talked about English teaching methods.

3. Algorithm structure
In this paper, under the network model of ResNet18, the methods of transfer learning [3] and channel attention mechanism [4] are used to improve the accuracy of the model; the method of fine-tuning the
model is used to further improve the practicability of the model; finally, this paper proposes one A new network model (TR-ResNet), which can be applied to children's English teaching methods.

### 3.1. Transfer learning

Based on the training of a letter recognition classification model, this article needs to initialize the weights of the classification model, and the parameters are random. The data migration learning method used in this article is fine-tuned on the basis of training all the weight parameters of the pre-training. Quickly realize the convergence of the training network model, thereby improving the accuracy of classification and recognition. The pre-training weight comes from the weight of the large data set image [5]. The algorithm flow chart of adding transfer learning to English letter classification is shown in Figure 1.

![Fig. 1 Flow chart of letter classification and recognition algorithm combined with transfer learning](image)

### 3.2. Attention mechanism

This article uses the channel attention mechanism. The channel attention mechanism is added after the 3×3 convolution of BasicBlock to improve the accuracy of the model's classification of English letters, thereby improving the efficiency of English teaching and increasing the learning of children and children's enthusiasm and learning efficiency.

Among them, as shown in Figure 2, the left a is the BasicBlock of resnet18, and the right b is the T-BasicBlock model with the channel attention mechanism added in this article.

![Fig. 2 Comparison of model structure between adding channel attention mechanism and without adding channel attention mechanism](image)

### 3.3. Model fine-tuning

Model fine-tuning is a relatively classic optimization method, through continuous attempts to achieve the goal of optimizing the network model. In this paper, through constant debugging of the model, in order to prevent some data sets, pictures in English words from instantaneous explosions in the data set,
we replaced the activation function relu number in T-BasicBlock with relu6 to make the activation function non-linear. There will be no data explosion in the process; in addition, under the premise of ensuring that the number of layers of the network model does not affect the accuracy of the model or has minimal impact on the accuracy, we will reduce the number of residuals [2,2,2,2] becomes [2,1,1,1].

Finally, through three methods of transfer learning, attention mechanism and model fine-tuning, the network model structure of this paper can be obtained as shown in Table 1. Among them, R+T+A represents the residual fast after using transfer learning, channel attention mechanism and model fine-tuning on the basis of ResNet18’s BasicBlock. Among them, the R+T+A residual is 4 layers faster, so our proposed child English classification model ResNet14 has a total of 14 layers.

Table 1. Structure of network model in this paper

| LayerName | Output Size | Layer | Params(170,751,359) |
|-----------|-------------|-------|---------------------|
| ConV_1    | 112×112     | 7×7, 64, stride 2 | 130,146,591 |
| ConV_2    | 56×56       | 3×3, 64, stride 2 | 11,852,265 |
| ConV_3    | 28×28       | R+T+A, 1 | 12,563,445 |
| ConV_4    | 14×14       | R+T+A, 1 | 9,758,112 |
| ConV_5    | 7×7         | R+T+A, 1 | 6,426,934 |
| ConV_6    | 1×1         | pool, FC, SoftMax | 4,012 |

4. Experimental results

The experimental environment of this article is carried out on the ubuntu 18.04 operating system, and the computer is configured as i9 RTX3090 NVIDIA Studio graphics card with 24G single card memory. The data set was taken by the School of Foreign Languages, Hubei Institute of Technology. Among them, the resolution of each letter picture is 295×413, the horizontal and vertical resolution is 96dpi, the bit depth is 24, there are 62400 English letter pictures, 57200 training set, 5200 test machine, 52 types of English letters. The average distribution of the number of sheets is the same, and the actual training test set is one. The following time is the effect of one picture for testing. The evaluation indicators used in this article are the accuracy (accuracy, ACC) and loss function (Loss) of top1 and top3.

In order to determine that the TA-ResNet14 model in this article is suitable for the study of children’s English teaching, we use ResNet50, mobileNetV1[7], GoogleNet[8], VGG-16[9], VGG-19 and DenseNet-121[10] and TA-ResNet14 for comparison. Through the comparison of experimental data in Table 2, we can find that ResNet50 exceeds mobileNetV1, GoogleNet, GoogleNet, VGG-16, VGG-19 in the accuracy of top1 and top3, but the same reason is lower than TA-ResNet14, although DenseNet-121 has a higher accuracy rate than TA-ResNet14, but DenseNet-121 runs significantly longer than TA-ResNet14, so DenseNet-121 is not suitable for transplantation in early childhood education, with low efficiency and high resource consumption. It is best to conclude that our model TA-ResNet14 is suitable for application to the task of teaching English to young children.

Table 2. Comparison of data between this model and other classification models

| Model       | ACC top1(%) | ACC top3(%) | Time(s) |
|-------------|-------------|-------------|---------|
| ResNet50    | 89.01       | 90.69       | 2.0     |
| mobileNetV1 | 79.87       | 82.01       | 0.9     |
| GoogleNet   | 78.20       | 80.20       | 2.2     |
| VGG-16      | 80.46       | 82.10       | 2.3     |
| VGG-19      | 85.23       | 88.67       | 3.2     |
| DenseNet-121| 94.66       | 97.15       | 5.1     |
| TA-ResNet14 | 92.01       | 95.67       | 1.3     |

Figure 3 is the accuracy curve of TA-ResNet14 where A is the learning rate curve and B is the accuracy rate of top1 and top3. Through the experimental graph, we can get the same results as Table 2, which further proves the validity of the experimental data.
5. Conclusion

With the emergence of the epidemic and the development of society, English learning anytime and anywhere is essential. This paper proposes an automatic classification model of English letters combined with the deep learning method. First, transfer learning is used to improve the accuracy of the classification model; then the channel attention mechanism method is used to further improve the accuracy; finally the model fine-tuning method further reduces the amount of parameters and calculations the amount. Experiments and comparisons on the letter data set found that the effect is better than the traditional classification model, so it is concluded that the classification model proposed in this article is suitable for automatic English vocabulary teaching.

Acknowledgments

Fund Project: The Philosophy and Social Science Youth Project of Hubei Provincial Department of Education in 2019 "Research on the Development of College English Teachers Based on Blended Learning Model" (Project Number: 19Q171).

References

[1] Li Ya, Huang Jing. The application of Chinese traditional culture in college English teaching [J]. Journal of Hubei Open Vocational College, 2020, 33(24): 188-189.
[2] Lécun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
[3] Gu, Jiuxiang & Wang, Zhenhua & Kuen, et al. (2015). Recent Advances in Convolutional Neural Networks. Pattern Recognition. 77. 10.1016/j.patcog.2017.10.013.
[4] Krizhevsky, Alex & Sutskever, et al.(2012). Imagenet classification with deep convolutional neural networks. 1097-1105.
[5] Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409.1556.
[6] Szegedy, Christian & Liu, Wei & et al. (2015). Going deeper with convolutions. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 1-9. 10.1109/CVPR.2015.7298594.
[7] He, Kaiming & Zhang, Xiangyu et al.(2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90. resnet
[8] Pan, S, Tsang, I., Kwok, et al,Yang, Q. Domain Adaptation via Transfer Component Analysis[J]. IEEE Transactions on Neural Networks, 2011, 22(2):199-210.
[9] Hu, Dichao. (2020). An Introductory Survey on Attention Mechanisms in NLP Problems. 10.1007/978-3-030-29513-4_31.
[10] Woo S, Park J, Lee J Y, et al. Cham: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.