Back to the Origin: An Intelligent System for Learning Chinese Characters

Jinglei Yu, Jiachen Song, Yu Lu(✉), and Shengquan Yu

Advanced Innovation Center for Future Education, Faculty of Education, Beijing Normal University, Beijing, China
luyu@bnu.edu.cn

Abstract. Learning Chinese characters is a challenging task for both native and foreign beginners. One major reason is that most Chinese characters in writing are distinct from each other and lack of directly phonetic clues. Fortunately, many Chinese characters' original forms have iconicity that indicates their meanings. By leveraging on these characteristics and the latest computer vision (CV) techniques, we design and build an intelligent system that could automatically retrieve the iconic and original forms of Chinese characters. Furthermore, the system could provide learners with different styles of the character in a chronological order to bridge the original form and the most commonly used one. Specifically, we adopt the SE-Resnet-50 classification model for both character recognition and style recognition tasks, and design a dedicated retrieval mechanism to properly select the representative characters in different styles for learners. A specific user interface is designed for beginners to upload, recognize, remember, and understand the Chinese characters.

Keywords: Language learning · Character recognition · Computer vision

1 Introduction

Chinese character recognition is difficult for non-native and even native beginners due to several reasons. First, the structure of individual character normally depicts an object or represents some abstract notions rather than based on alphabet. Hence, different Chinese characters in writing are distinct from each other and the number of commonly used characters exceeds 3,000. Second, each Chinese character can be written in five major chronologically formed scripts that are still utilized today, namely seal script, clerical script, cursive script, running script, and regular script. Many characters' current commonly used forms in regular script cannot reflect any visual meaning that can be easily perceived by beginners, while their original forms in seal script can. In addition, Chinese characters are not directly related to their pronunciations, and thus the learners have to recognize the characters apart from the pronunciations.

I. Roll et al. (Eds.): AIED 2021, LNAI 12749, pp. 457–461, 2021.
https://doi.org/10.1007/978-3-030-78270-2_81

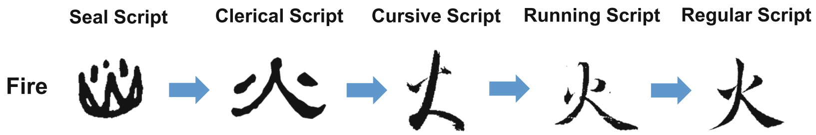

Fig. 1. The chronological evolving process of the character "Fire" in five scripts.

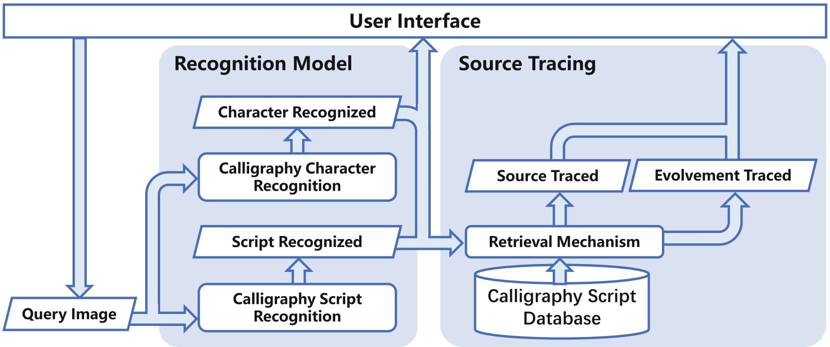

Fig. 2. The simplified block diagram of the system.

The previous studies have shown that iconicity, i.e., structural similarity between the character and its referent, is effective for children in reading [5]. Fortunately, many Chinese characters' original forms are iconic that can be easily recognized. Taking character "Fire" as example, its original form in seal script is a simplified picture of fire. Figure 1 illustrates the chronological evolving process of it in the five scripts. Hence, we design and implement an intelligent system that could automatically recognize and trace back to the original form of the given Chinese character. Beside the original form, the system could also show the evolvement in the five scripts of the given character. The system specifically utilizes the calligraphy collections with high art value, which demonstrates the elegant forms of different scripts and easily arouses the learners' interests.

2 System Design

The system consists of two modules, namely *recognition model* module and*source tracing* module. Both modules connect to the upper user interface and support the interaction with the learners. Figure 2 illustrates the block diagram of the built system. Briefly speaking, the system first receives the input image of an individual character through the user interface. Since structural similarities of the same character among five scripts are generally higher than them among different characters, we treat character recognition and script recognition tasks separately for more accurate classification. In the recognition model module, two convolutional neural network (CNN) [3,4] models are utilized to accomplish the two tasks. The recognition results would be delivered to the source tracing

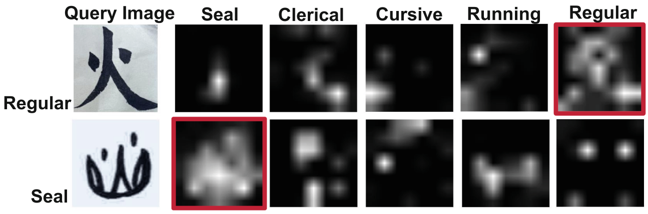

Fig. 3. Important region visualization of two query images on five classes. The brighter the regions are, the more attention the model pays to for the particular class. The correct recognition results are marked in red and illustrated on the left.

module. The source tracing module receives the recognized character and then retrieves different scripts of it in calligraphy script database. The database stores 35,563 images of the Chinese characters in the five scripts, and the representative images would be selected based on the input image. The selection criteria are based on the latent features extracted by the CNN model for calligraphy script recognition. Finally, the system displays the selected images in the chronological order to demonstrate the evolving process of the character, particularly highlighting the seal script as its original form. We will elaborate the two modules in the following parts.

2.1 Recognition Model Module

In recognition model module, both the character recognition and script recognition tasks can be regarded as the image classification task in CV. Specifically, we adopt SE-ResNet-50 [2] to recognize both the character and its script. ResNet [1,8] is a widely used CNN structure, which utilizes residual learning to train deep neural network. We adopt 50 layers ResNet as a based backbone with 49 convolutional layers and one fully connected layer. Meanwhile, squeeze-and-excitation (SE) blocks work as the self-attention function [7] on channels to improve classification accuracy.

For the character recognition model, we set each character as a class and first select 150 characters having five scripts. In total, 35,563 images of 150 characters are used for training and validation. Using 5-fold cross validation, the F1 score of the built model achieves 0.90. For the script recognition model, we treat each script as a class and five in total. In this way, the model is trained to learn the latent feature of each calligraphy script. We utilize gradient-weighted class activation mapping (Grad-CAM) [6] method to visualize the localization of the important regions, where the model attempts to discriminate different classes separately. As shown in Fig. 3, the model correctly responses to the ground-truth class, meanwhile responses to several local regions that partially reflect the script features in other classes. A total of 70,111 images of 2,177 characters are collected and, using 5-fold cross validation, the F1 score of the built model achieves 0.88.

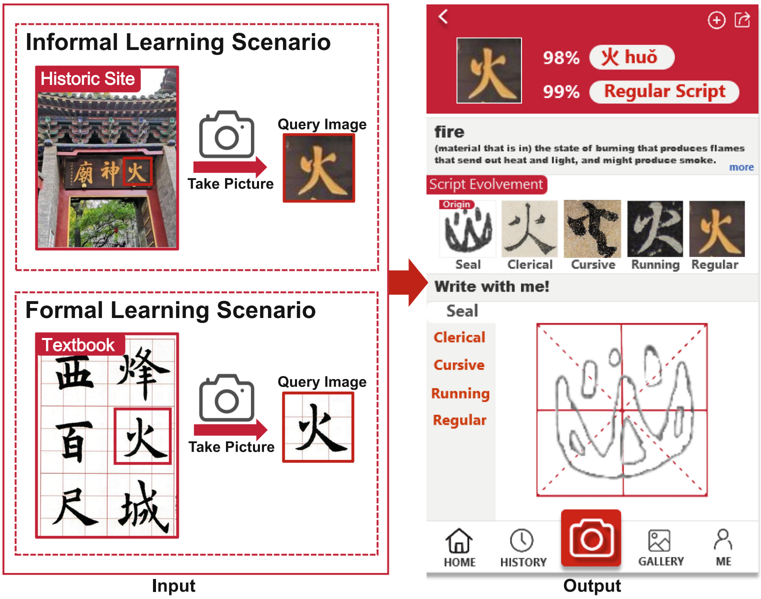

Fig. 4. The system usage scenarios and its user interface.

2.2 Source Tracing Module

Source tracing module selects different scripts of the recognized character in calligraphy script database using the retrieval mechanism. Specifically, the mechanism takes advantage of the features learned by the script recognition model. We extract features before last fully connected layer in float array with a size of 512, then measure the Euclidean distance of features between the query and script dataset images. The selections are the five representative characters in each script that are most similar to the input image.

2.3 User Interface

Learners could easily use the built system to learn the individual Chinese character in either formal or informal learning environment, such as taking pictures and uploading one photo from a historic site or a textbook. As shown in Fig. 4, the recognized character and its script would be shown at the top of the user interface, including the pronunciation and the models' confidences. The user interface also provides translation and explanation in English. Besides, the evolvement of the recognized character is displayed from the seal script to the regular script in the chronological order. More importantly, the original form of the recognized character is shown in the center of the user interface, learners could also counterdraw the character in different scripts to consolidate memory during writing.

3 Conclusion

By leveraging on the latest CV techniques in artificial intelligence, we design and implement an intelligent system to trace back the origin and evolvement of the Chinese characters. Empowered by the system, learning Chinese characters could be more intriguing, meaningful, and effective. We are deploying the system in multiple local schools and meanwhile improving the system to cover more basic Chinese characters.

Acknowledgement. This research is supported by the National Natural Science Foundation of China (No. 62077006), the Fundamental Research Funds for the Central Universities.

References

1. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
2. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
3. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Adv. Neural. Inf. Process. Syst. **25**, 1097–1105 (2012)
4. LeCun, Y., et al.: Backpropagation applied to handwritten zip code recognition. Neural Comput. **1**(4), 541–551 (1989)
5. Luk, G., Bialystok, E.: How iconic are Chinese characters? Bilingualism **8**(1), 79 (2005)
6. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)
7. Vaswani, A., et al.: Attention is all you need. arXiv preprint arXiv:1706.03762 (2017)
8. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1492–1500 (2017)