A Guqin Notation Recognition System using Machine Learning Methods

Bowen Yang[1], Shun Kuremoto[2, 3], Mamiko Koshiba[2], Shingo Mabu[2], Hiroo Hieda[3], Takashi Kuremoto[1*]

[1]Nippon Institute of Technology University, [2]Yamaguchi University, [3]Institute for Future Engineering,
kuremoto.takashi@nit.ac.jp

Abstract: In the notation of the Chinese musical instrument Guqin, the symbols indicating the posture of the fingers and how to play are written in Chinese-like characters –Jianzu Pu. It is very difficult to understand and use the Guqin musical notation in the Jiazu Pu for the nowadays players. In this research, deep learning is proposed to recognize Guqin notation automatically. Firstly, a database of single Jianzu Pu was made with multiple handwriting images and augment data. Secondly, to increase the accuracy of classification, multiple machine learning models such as VGG16, SVM, and a combination of VGG16 and SVM proposed in our previous work were adopted. Thirdly, the comparison experiments were performed using 1,500 images which in 15 kinds of patterns of single Jianzu Pu. The accuracies of test data (unknown) were 63.1%, 85.07%, and 88.33% given by SVM, VGG16, and VGG16+SVM, respectively. VGG+SVM model reached 99.11% training accuracy after 40 training times.

Key-Words: *Deep Learning, Guqin's Notation, Pattern Recognition, Image Identification*

1. Introduction

Along with the flow of the times, there is a current situation that many histories and cultures have been lost. One of the threatened traditional cultures is a musical instrument called “Guqin" (Figure 1), though Guqin ranks the first among the four arts of traditional Chinese culture, i.e., “Guqin (Chinese harp), Go (Chinse chess), calligraphy, and Chinese painting”, and was invented before 2,500 years ago in China [1]-[3].

The Guqin music notations were written in the glyph and characters, recording the movement of fingers, string sequences, and timbre of the Guqin performance in the early time. Since the ancient notation was difficult to be understood and used by the normal players, it was simplified in the Tang dynasty (A.D. 618-907) as “Jianzi Pu”, which means “character reduced notation” (Figure 2). As a special notation for Guqin music, Jianzi Pu converts the letters of the character staff into symbols, combines the symbols with "reduced characters", and records fingering movements and string numbers to be plucked. Meanwhile, Jianzi Pu does not describe the tempo of the piece, although it describes how to use the fingers and the dynamics to play a melody. As a result, the performer needs to grasp the intention of the piece, imagine various scenes and emotions, and perform it with his/her own rhythm. Deciphering the reduced-letter notation and determining how to play each Guqin piece is technically called “Da Pu" [1]. However, although there are more than 600 musical scores for the Guqin produced in thousands of years, it is said that only about 100 of them are able to be typed and are often performed, due to the difficulty and hugeness of the reduced-letter notation understanding.

Received: 2023/01/19, Accepted: 2023/11/20
*Corresponding author: Takashi Kurmeoto
E-mail address: kuremoto.takashi@nit.ac.jp

Figure 1 Two Guqins [2]

Figure 2 A character of Jianzi Pu (Guqin notation)

Research on automatic transcription of reduced-letter notation (Jianzi Pu) began at the same time as handwritten character recognition [4]-[6]. In recent years, a research group at the Central Conservatory of Music (Beijing, China) has been conducting research on reduced-letter score identification using deep learning, but the results have yet to be announced as we know. On the other hand, there are several successful examples of handwritten character recognition using deep learning such as in [7]. However, different from the study of handwritten character recognition by artificial intelligence methods, the original data of Jianzi Pu are difficult to be collected and annotated without the contribution of the specialists of Guqin musicians.

In this study, we created a dataset of one Guqin music "Sen-O-So" written by Jianzi Pu at first, then we proposed multiple machine learning methods to recognize the notation automatically. The dataset was composed by the images of the original single characters of Jianzi Pu and the images of augmentation yielded by image processing such as rotation, size enlargement, size reduction, inversion and filtering. The machine learning methods used in this study were a well-known deep learning model VGG16, a classic classifier SVM (support vector machine), and a combination of VGG16 and SVM proposed in our previous work [10]-[12]. training accuracy

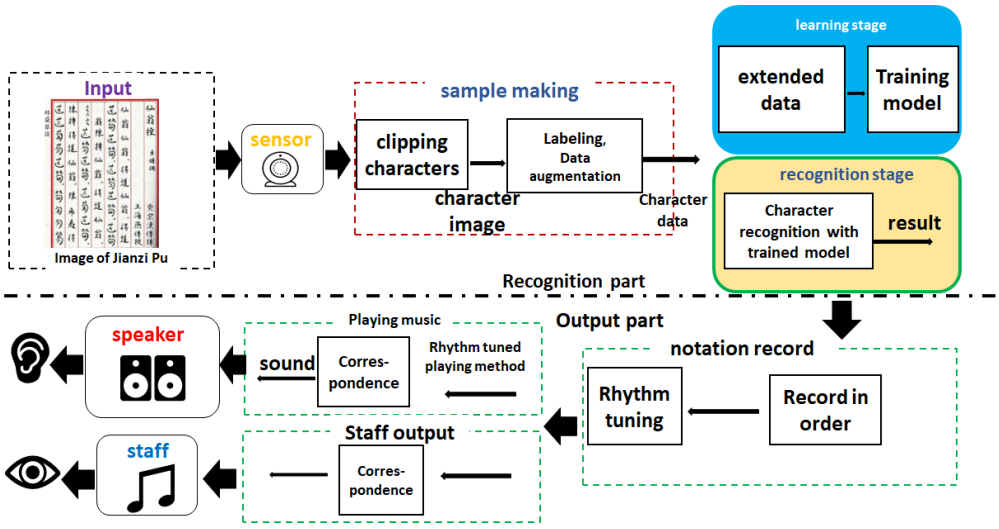


Figure 3 A recognition system for Guqin music proposed in this study

The comparison experiments were performed using 1,500 images which included 15 kinds of patterns of single Jianzu Pu appearing in the Guqin music of Sen-O-So. The average accuracies of test data (unknown data) were 63.1%, 85.07%, and 88.33% given by SVM, VGG16, and VGG16+SVM, respectively. Specially, to the training data, VGG+SVM model reached 99.11% training accuracy after 40 training times.

2. A Recognition System for Guqin Notation

In this research, we constructed a database of single characters of Guqin music Sen-O-So (details are described in Section 2.1 and Section 3.1) and proposed a recognition system for Guqin notation (see Figure 3).

2.1 Single character segmentation

The single characters in Jianzi Pu images are obtained by the processing of an automatic segmentation as shown in Figure 4. The original Guqin notations (handwritten or typographic style) captured by camera were binarized at first, then histogram equalization and horizontal/vertical projection filtering were performed to segment (extract) the single characters (Figure 5).

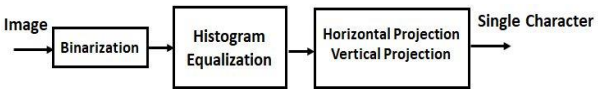
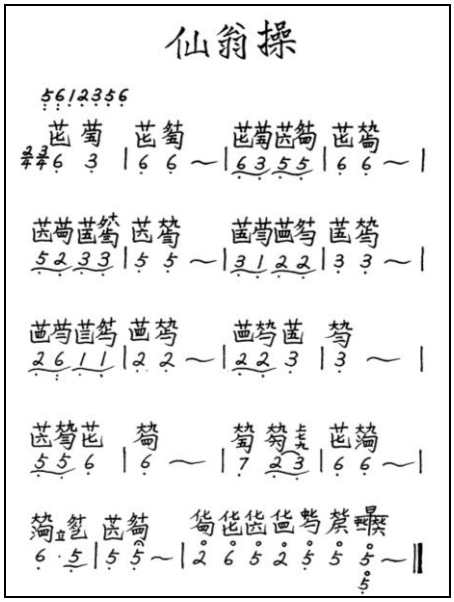


Figure 4 A processing of segmentation for single characters

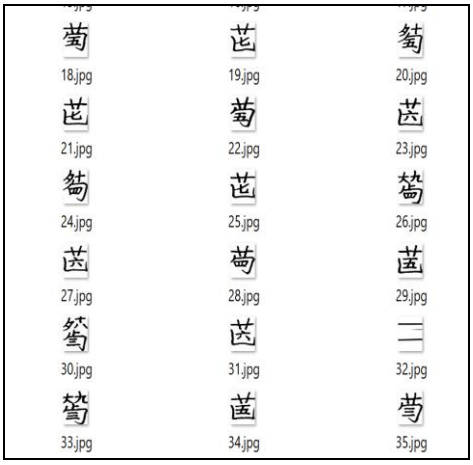
Augment of single-character image data was executed by image processing such as rotation, inversion, enlargement (zoom-in), and reduction (zoom-out). Some examples of the single character augment results are shown in Figure 6.

2.2 Machine learning models

Support vector machine (SVM) [8], deep convolutional neural network VGG16 [9], and a hybrid model VGG16 with SVM, which introduces SVM instead of fully connected layer of VGG16 [10] – [12], are adopted to the Guqin notation recognition system (Figure 3).



(a) An original image of Guqin notation "Sen-O-So"



(b) Segmentation results of single characters

Figure 5 Single character segmentation of Jianzi Pu

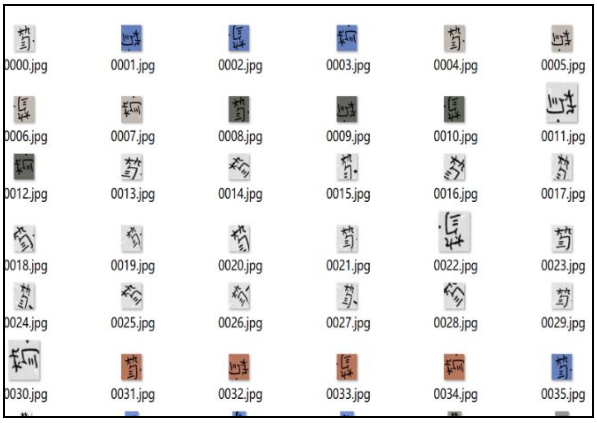


Figure 6 Single character augment of Jianzi Pu

SVM is a classic supervised learning method and powerful for the pattern recognition of high-dimensional data.VGG16 is a famous deep learning model in the 3rd AI boom. It consists of 16 layers of convolutional neural network (CNN) and three layers of fully connected layers. In our previous works, a hybrid deep learning model VGG16 with SVM was proposed by a fine-tuned VGG16 and a SVM which replaces the fully connected layers of the VGG16. In other words, it is a kind of SVM using the features of input data, i.e., the output of max-pooling layer of VGG16, as its input elements. After the training of the hybrid model, it is often performed better than a single SVM or the original VGG16 in the high dimensional classification problems.

Figure 7 shows the configuration of the hybrid VGG16 with SVM model used in this study. The construction of the model is given in two stages. In the first half, VGG16 is trained (fine-tune) to extract the features of the input image. In the second half, the part of the fully connected layers (FC layers) of VGG16 are replaced by SVM, and the whole model is trained again (fine-tuning again). Finally, the test data which are not used in the training process are utilized to verify the performance of the hybrid model VGG16 with SVM.

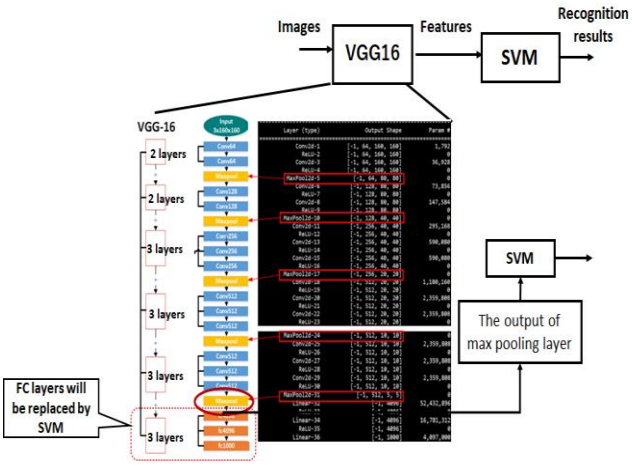


Figure 7 A hybrid deep learning model VGG16 with SVM

3. Experiment and Results

To verify the effectiveness of the proposed system, a simple Guqin notation "Sen-O-So" (see Figure 5 (a)) were used in the recognition experiment. As the first trial, we created a database with the first two lines of Sen-O-So, and the recognition accuracies of different machine learning models described in the Section 2 were compared by the experiment results.

3.1 Experiment preparation and setup

We collected 31 handwritten versions of "Sen-O-So" ("仙翁操" in Chinese and Japanese) Guqin notation from the Internet. Using the first two lines of Sen-O-So, single character of Jianzi Pu extraction processing was performed (Figures 4 & 5). As the result, 15 patterns of single character images (image size: 100×128) were obtained. By performing data augmentation on these single character images, i.e., rotation, reversion, size enlarge (zoom in), size reduction (zoom out), filtering to the original images of single characters, 1,500 sample images in the 15 classes were obtained as the training samples and test samples of the system as shown in Figure 6.

The original VGG16 model was trained by ImageNet, a database of 14 million sheets and 20,000 types of images [9]. In this study, fine-tuning of VGG16 was performed using the single character images of "Sen-O-So". SVM used here was a one-versus-the-rest model for multi-class classification. The number of input dimensions to the SVM was 12,800 according to the single character image size (100×128). In the case of VGG16 with SVM model, the number of the output of max pooling layer of VGG16 was 4,096, and it was the number of input dimensionality to the SVM part of the hybrid model.

In the single character recognition experiments, 1,200 training data and 300 test data were randomly selected from 1,500 sample images.

3.2 Experimental result

Table 1 shows the classification accuracies of the different models in training process (Acc (%)) and

validation process, i.e., the optimized model's performance to the unknown data (Val_acc (%)) after 40 training times (epochs). The classification accuracies of the hybrid model VGG16 with SVM in boldface type in Table, marked 99.11% to training data and 88.33% to validation (unknown) data, these were higher than the accuracies by other models. On the other hand, the training time of SVM was 8.53 seconds, which was the shortest among three models (Table 1).

Table 1 Accuracies and training time of different models

	SVM	**VGG16**	**VGG16+SVM**
Val_acc (%)	63.16	85.07	**88.33**
Acc (%)	-	94.33	**99.11**
Training time	**8.53 sec**	12.32 min	12.08 min

3.3 Analysis of learning performance

The learning performance of the original VGG16 and the hybrid model VGG16 with SVM was shown in Figure 8 and Figure 9, respectively. The comparison of them was depicted in Figure 10. It was confirmed that the convergence of VGG16 with SVM was faster than that of VGG16.

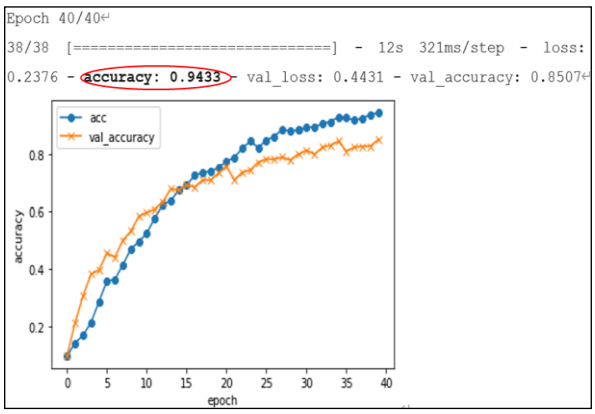


Figure 8 Learning performance of VGG16

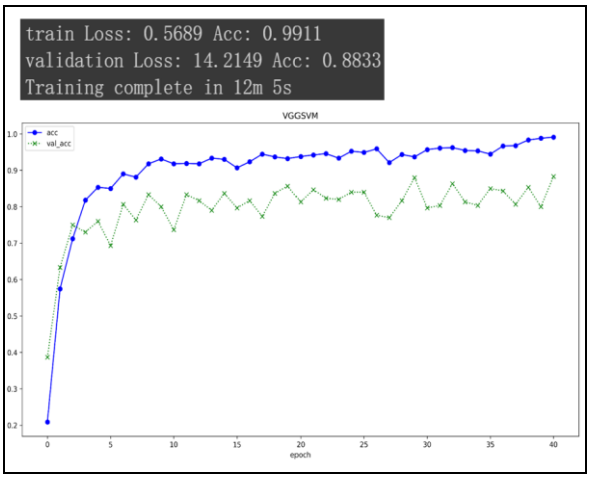


Figure 9 Learning performance of the hybrid model VGG16 with SVM

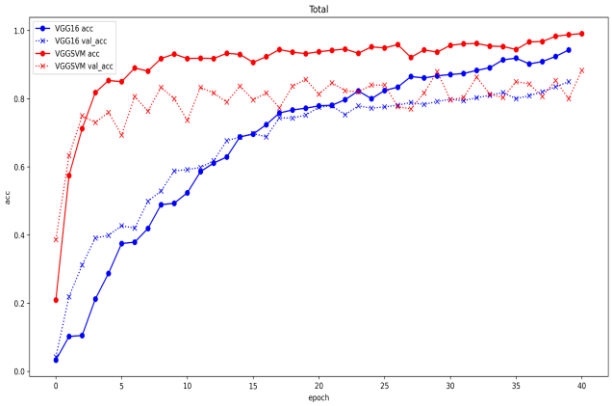


Figure 10 Comparison of learning performances between VGG16 and the hybrid model VGG16 with SVM

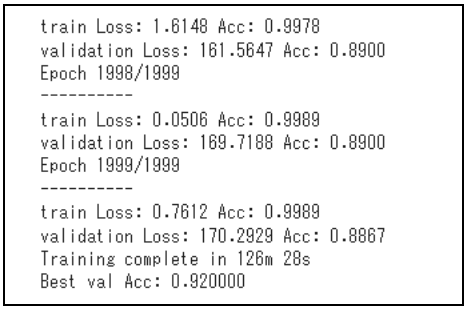


Figure 11 Experiment results of VGG16 with SVM model after 2,000 training times (epochs)

In fact, when the training time of VGG16 with SVM reached at 2000 times (epochs), the recognition accuracy of 15 classes single characters marked 99.89%, and the validation accuracy "Val_acc" also increased to 88.67%

(Figure 11). The training process consumed more than two hours (126m 28s) in the case of a computer of us (Intel Core i-5, 6 cores, 2.60GHz, GeForce GTX1650, 16GB memory).

4. Output of the Staff Notation

After the single characters of Jianzi Pu were recognized we tried to output the recognition result as a stave and audio form. Single characters of Jianzi Pu were matched to staff notations as same as shown in Figure 5 (a) and staff notations were transformed to an audio file by a free software the music21 [13]. An example of the output of the staff notation and audio form is shown in Figure 12.

Figure 12 The output of the staff notation system

5. Conclusions

To recognize the Guqin notation “Jianzi Pu”, a system with machine learning models including SVM, VGG16, and a hybrid model VGG16 + SVM was developed in this study. Recognition accuracies to the single character of a Jianzu Pu “Sen-O-So” showed the priority of the hybrid model VGG16 with SVM comparing to the former two classifiers. To realize the restoration of Guqin music, the system combined the recognition results of single characters of Jianzi Pu with staff notations and transformed them to audio data.

Acknowledgments

This research was supported by Grants-in-Aid for Scientific Research (No. 22H03709).

References

[1] 中国スクリプト,「琴と箏の歴史と構造の違い」：http://chugokugo-script.net/chugoku-bunka/kin.html (2016)

[2] Japan Society for Promotions of Guqin : https://www.Guqin.jp/about/

[3] 栂尾亮子：「打譜に関する調査報告」, お茶の水音楽論集, 第 10 号 (2008)

[4] Zhi-xiao Pan, Chang-le Zhou: “Text Segmentation and Extraction from Images of Guqin Jianzi Pu”, Mind and Computation, Vol.1, No.2, pp. 286-295 (2007) (in Chinese)

[5] Chen Shi: “Guqin Notation and Music Style Recognition”, Computer Science (2016)

[6] 王 利, 孫 洋, 罗 兆麟, 張輝:「AI 自动翻译“减字谱”——以《流水》和《不染》为例」,芸術教育, Vol.3 (2019)

[7] 竹内正広,早坂太一,大野亙,加藤弓枝,山本和明,石間衛, 石川徹也:「ディープラーニングによるくずし字認識組み込みシステムの開発」,人工知能学会全国大会 (2019)

[8] J. Platt: "Sequential minimal optimization: A fast algorithm for training support vector machines", www.microsoft.com/research/ (1998)

[9] K. Simonyan & A. Zisserman, "Very deep convolutional networks for large-scale image recognition", International Conference on Learning Representations (ICLR 2015)

[10] 呉本　尭，佐々木敬彬，間普真吾：“ニューラルネットワークとサポートベクトルマシンを併用した EEG 信号識別手法の検討”，平成 30 年電気学会電子・情報・システム部門大会論文集，TC16-13, pp.593-597 (2018)

[11] T. Kuremoto, T. Sasaki, S. Mabu: “Mental task recognition using EEG signal and deep learning methods”, Stress Brain and Behavior, Vol. 1, pp.18-23 (2019)

[12] Y Mori, T. Kuremoto, S. Mabu: “Facial Expression Recognition with DCNN and SVM”, The 7th International Conference on Innovative Application Research and Education (ICIARE 2019), pp. 64-68, (2019)

[13] MIT: “music21: a toolkit for computer-aided musicology”, http://web.mit.edu/music21