IMPROVED CHORD RECOGNITION USING SYNTHETIC DATA GENERATION BY ROBOT

Gerelmaa Byambatsogt and Gou Koutaki

Kumamoto University, Japan

Lodoiravsal Choimaa

National University of Mongolia, Mongolia

ABSTRACT

In this paper, we propose a new framework for generating big sized dataset using synthetic data generation by robotics. In learning-based recognition, for example, using convolutional neural networks (CNNs), it is critical for the performance, to collect high-quality and large amounts of training data. Previously, to increase the training data set, a data augmentation technique based on digital signal processing were applied to the original sound data. However, the data augmentation based on digital signal processing data is a limited method, because it depends on some previous knowledge of the data and cannot perform for all domains. On the other hand, we propose a new dataset collection technique using a robot that automatically plays instruments, by which it becomes possible to add high-quality data to training samples. Experimental results for guitar chord recognition show that the proposed method using CNNs and a guitar robot can outperform the CNN systems with the traditional data augmentation.

Index Terms— Chord Recognition, Data Augmentation, Convolutional Neural Network, Robot

1. INTRODUCTION

Chord, the mid-level musical feature, is an important element for understanding the music for the player. Currently, the analysis of computer music is a common method of analyzing musical sounds. Along with the computer music analysis, musical robotics was developed [1]. The robot has the advantage of being able to deal with any design. Therefore, the robot-musician has a benefit of playability, associated with dynamic variation or playing speed that are difficult for a human to perform. Data augmentation is a powerful technique, but it has some drawbacks. Traditional data augmentation is of poor quality and contains special unnatural artifact sounds and it does not completely reflect the richness of the acoustic sound. To solve these, we propose a synthetic data generation method (Fig. 1). We aimed to use a real chord to train a more accurate chord recognition system. Our contributions are: (1) We propose a new framework for generating training samples using robotics and prove its effectiveness through experimental results, (2) We investigated the best CNN architecture for our augmented training data by an exhaustive experimental setup. (3) We are constructing and will open a large dataset for guitar chord recognition on GitHub [1]. The dataset includes 47 types of a total of 4,660 chord samples with a size of 1.8 GB.

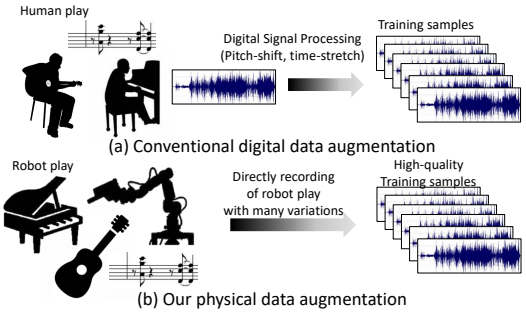


Fig. 1. In the data augmentation method, training samples are generated from a small number of training samples of a human play using digital signal processing. On the other hand, the proposed method can generate many high-quality training samples from the music play by the controlled robot.

2. PROPOSED METHOD

To enrich the guitar chord dataset, we created a robot that plays an acoustic guitar. The robot system is divided into three units, including a string pushing unit, a picking unit, and a linear mechanical unit (Fig. 2). A total of 24 solenoids were installed in pushing string unit because of the chord forms and six servomotors and solenoids in the picking unit. In a linear mechanical unit, a set of solenoids moves along with the guitar neck to play a variety of chords with a stepper motor.

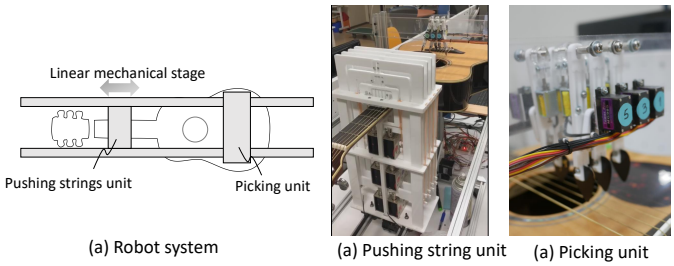
The proposed robot can automatically play many variations of chords and in many chord forms and strokes (fast, up/down, slowly). Chords was recorded in WAV format and separately in noisy environments.

3. EXPERIMENTAL RESULTS

3.1. Dataset

The dataset includes 48 types of the chords with major, minor, major 7th, and minor 7th of the twelve roots A, A#, B, C, C#, D, D#, E, F, F#, G and G#. In the experimental setup, we used 4 types of datasets. Table 1 represents the size of each dataset.

[1] https://github.com/gerelmaa/guitar_chord_from_robot



(a) Robot system (a) Pushing string unit (a) Picking unit

Fig. 2. Our developed guitar robot. The robot consists of a linear mechanical stage, a string pushing unit, a picking unit, and an actual guitar.

Table 1. The number of samples for each dataset.

Dataset	#samples
(A) Human (Original)	2100
(B) Human + aug. (Prev.)	7932
(C) Human aug. + robot (Prop.)	13030
(D) Test dataset	200

Human dataset consists of the guitar chord dataset used in [2]. In (B) dataset, we use time-stretching and pitch shift augmentations. (C) dataset is used as the proposed method. Test dataset is compiled from an online source that represents plays of different musicians in different environments.

3.2. Implementation and result

We used the Mel-spectrogram for the CNNs input feature. The proposed method is evaluated using 4 types of CNN architectures including (a) simple CNN, (b) deep CNN architecture [3], (c) CNN without fully connected layer [4], (d) Convolutional Recurrent neural network (CRNN) architecture [5]. The first architecture is a simple CNN network, while others are those used in the classification of musical instruments and chord recognition tasks. Table 2 shows the results of network architectures trained on each dataset. From the Table 2, the best performance found in the method that relies on human augmentation with a robot-played dataset [4], which is proposed method. For all networks, the accuracy of the test increases when the network is trained with the proposed method, except for the method proposed in [3], where the accuracy is the same as for the human augmentation dataset.

Table 2. Accuracy test [%] of CNN architectures.

Methods	(A) Original	(B) Prev.	(C) Prop.
(a) Simple CNN	36.50	43.29	48.95
(b) Deep CNN [4]	29.91	45.31	**49.47**
(c) Fully CNN [3]	35.41	48.43	48.43
(d) CRNN [5]	33.85	36.90	44.27

4. DISCUSSION AND LIMITATION

Using synthetic data by robot is effective in the task of recognizing guitar chords. The proposed method can be used for any kind of musical instrument. However, there are some limitations to this method. This method requires a hardware implementation of the robot The creation and playback of sound is directly dependent on the robot design. Creating a sound similar to the sound of a human musician requires a more sophisticated robotic design.

5. CONCLUSION

We proposed a synthetic data generation using a robot player to recognize musical chords in the MIR system. As an example of such a system, we have developed a robot playing guitar that could play many types of chords using various chord forms and strokes. We have recorded 4,660 sound files using the robot, and these were used for the machine learning system of CNNs. We evaluated our system using four kinds of deep learning methods, and our system provided better performance than the conventional data augmentation method, such as time stretching and pitch shift.

6. ACKNOWLEDGEMENTS

The authors would like to thank the MJEED research grant (Research profile code:J14C16) Higher Engineering Education Development Project.

7. REFERENCES

[1] Ajay Kapur, "A history of robotic musical instruments," in *ICMC*, 2005.

[2] Jean Jacques Droogenbroeck Marc Piérard Sébastien. Osmalskyj, Julien Embrechts, "Neural networks for musical chords recognition.," in *Actes des Journées d'Informatique Musicale (JIM)*, 2012.

[3] Y. Han, J. Kim, and K. Lee, "Deep convolutional neural networks for predominant instrument recognition in polyphonic music," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 1, pp. 208–221, Jan 2017.

[4] F. Korzeniowski and G. Widmer, "A fully convolutional deep auditory model for musical chord recognition," in *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, Sept 2016, pp. 1–6.

[5] S. Panwar, A. Das, M. Roopaei, and P. Rad, "A deep learning approach for mapping music genres," in *2017 12th System of Systems Engineering Conference (SoSE)*, June 2017, pp. 1–5.