Automatic Chinese National Pentatonic Modes Recognition Using Convolutional Neural Network

Zhaowen Wang[1] Mingjin Che[2] Yue Yang[1] Wenwu Meng[2]
Qinyu Li[2] Fan Xia[2] Wei Li[3]

[1] Central Conservatory of Music [2] Sichuan Conservatory of Music [3] Fudan University

Introduction

Automatic music key/mode recognition is a fundamental task of MIR. The existing works are mainly for western major and minor keys, including approaches using template matching, hidden Markov models, or neural networks. But the automatic recognition of Chinese national pentatonic modes from audio signals has been less studied due to various reasons, for example complicated variants and lack of annotated datasets. In this paper, we design residual convolutional neural network models for recognition and then train and test them on our self-built CNPM dataset. This work will contribute to the development of multicultural MIR, computational ethnomusicology and five-tone music therapy.

The modern Chinese national pentatonic modes theory was created after the early twentieth century. It is based on the five tones of Gong, Shang, Jue, Zhi and Yu and has combined western and Chinese music theories. A complete mode name here can be divided into three parts: the pitch of tonic, the mode pattern, and the mode type. Besides, there is a concept called TongGong system, which can assist us in classifying.

Some examples of the mode scales.

Classification basis	Names	Amount	Labels
TongGong System	C; ♯C (♭D); D; ♯D (♭E); E; F; ♯F (♭G); G; ♯G (♭A); A; ♯A (♭B); B	12	0-11
Pitch of Tonic	C; ♯C (♭D); D; ♯D (♭E); E; F; ♯F (♭G); G; ♯G (♭A); A; ♯A (♭B); B	12	0-11
Mode Pattern	Gong; Shang; Jue; Zhi; Yu	5	0-4
Mode Type	Pentatonic; Hexatonic (Qingjue); Hexatonic (Biangong); Heptatonic Yayue; Heptatonic Qingyue; Heptatonic Yanyue	6	0-5

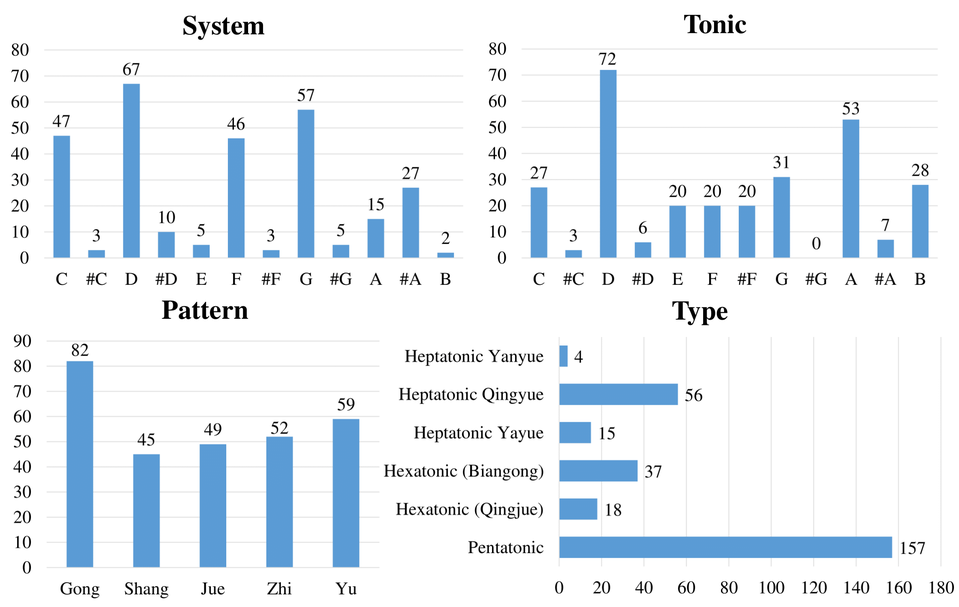
Definition of the mode category.

Type	Step names
Pentatonic	Gong / C - Shang / D - Jue / E - Zhi / G - Yu / A
Hexatonic (Qingjue)	Gong / C - Shang / D - Jue / E - Qingjue / F - Zhi / G - Yu / A
Hexatonic (Biangong)	Gong / C - Shang / D - Jue / E - Zhi / G - Yu / A - Biangong / B
Heptatonic Yayue	Gong / C - Shang / D - Jue / E - Bianzhi / ♯F - Zhi / G - Yu / A - Biangong / B
Heptatonic Qingyue	Gong / C - Shang / D - Jue / E - Qingjue / F - Zhi / G - Yu / A - Biangong / B
Heptatonic Yanyue	Gong / C - Shang / D - Jue / E - Qingjue / F - Zhi / G - Yu / A - Run / ♭B

Six mode types. The pitch after each step name is just one of the possible matches, listed for better understanding.

Dataset

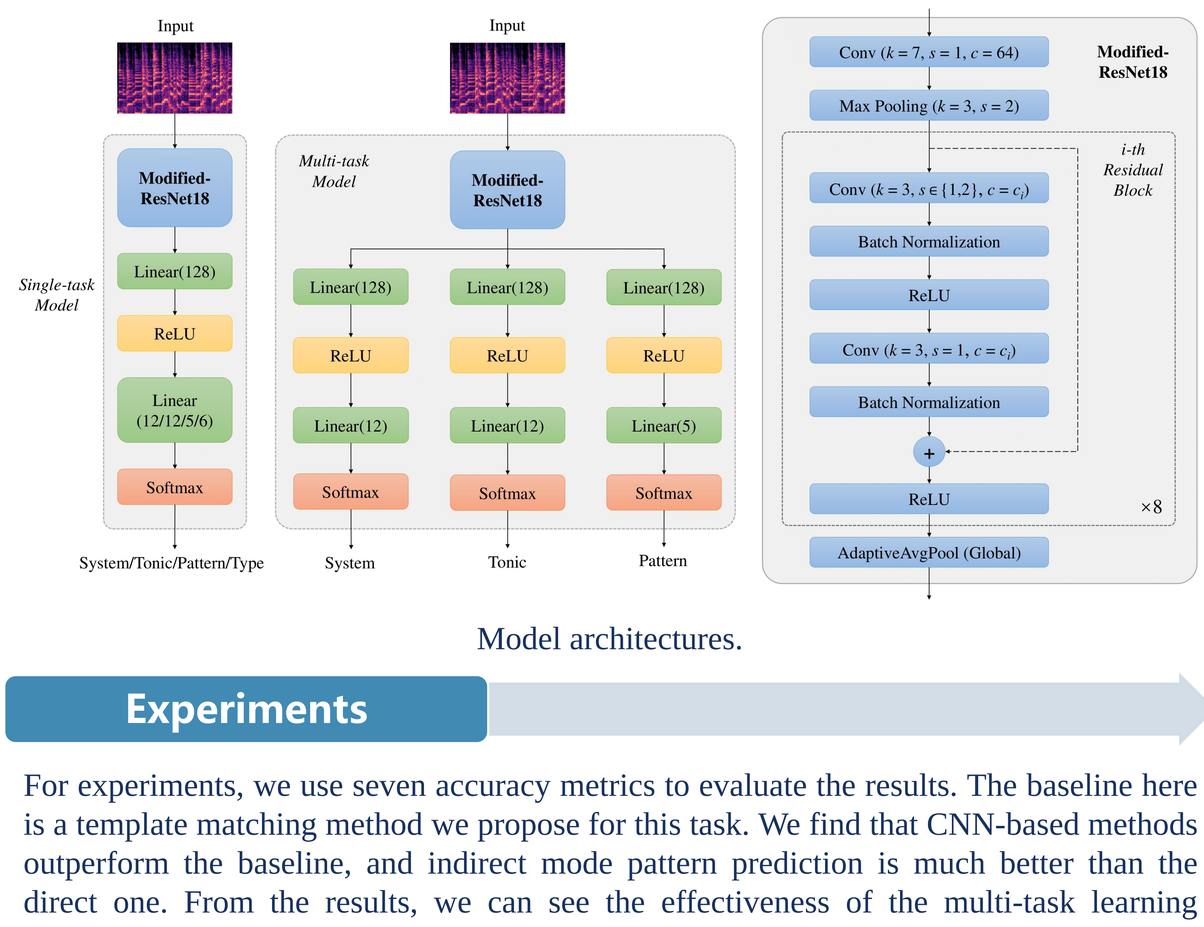
We use and expand our self-built CNPM Dataset. After expanding, the dataset contains 287 music recordings, including instrumental and vocal music, with instruments mainly being traditional Chinese instruments such as Guzheng, Guqin, Pipa, etc. We select music pieces with clearer modes, more stable pitches, and mostly performed by contemporary musicians to avoid controversy in judging the modes. Labels are annotated by faculty and students in related disciplines. Due to data imbalance, data augmentation is adopted.



Data distribution of the CNPM Dataset.

Models

Since there are 360 (12×5×6) kinds of modes theoretically, it is not reasonable to classify them directly. So we design CNN models to predict which TongGong system the mode belongs, the pitch of tonic, the mode pattern and the mode type separately. The backbone of our models is a modified ResNet18. We use both single-task and multi-task models with different strategies for identification. The mode pattern can be obtained either directly from the model or indirectly through the predicted system and tonic.



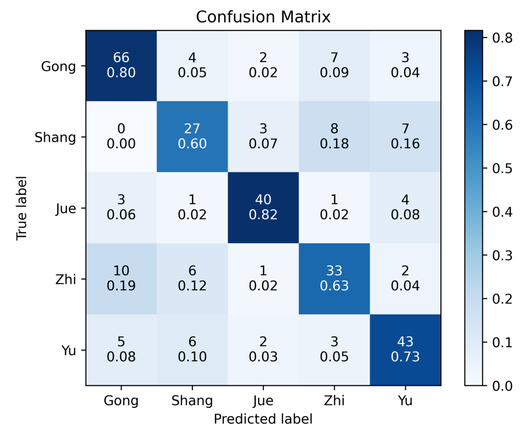
Model architectures.

Experiments

For experiments, we use seven accuracy metrics to evaluate the results. The baseline here is a template matching method we propose for this task. We find that CNN-based methods outperform the baseline, and indirect mode pattern prediction is much better than the direct one. From the results, we can see the effectiveness of the multi-task learning strategy.

Metric	Object
ACC1	System
ACC2	Tonic
$ACC3_D$	Pattern (Directly)
$ACC3_I$	Pattern (Indirectly)
ACC4	Tonic and Pattern
ACC5	Type
ACC6	Tonic, Pattern and Type

Accuracy metrics.



Confusion matrix for the pattern prediction.

ACC (%)	Baseline	Res-ST	Res-MT
ACC1	85.71	**88.50±1.80**	85.36±1.44
ACC2	71.78	74.85±9.28	**79.07±3.25**
$ACC3_D$	-	58.58±6.24	**63.79±7.74**
$ACC3_I$	66.55	71.01±10.30	**71.74±5.64**
ACC4	64.46	69.27±9.70	**69.65±5.21**
ACC5	48.08	**57.87±4.81**	-
ACC6	31.01	**42.15±3.65**	-

Accuracy results under 5-fold cross-validation.

Conclusion

In this paper, we explore the automatic mode recognition for Chinese national music composed after the early 20th century with the characteristics of Chinese national pentatonic modes. We use ResNet-based CNN models and compare them with template matching methods. The training and testing are conducted on our self-built CNPM dataset. The results for system, tonic and pattern classification are generally satisfactory, but can still be improved. We have combined the musical domain knowledge when designing the methodology used in this paper. Considering the relationship between system, tonic and pattern, we adopt the indirect prediction method and the multi-task learning strategy, both of which lead to better results.

Acknowledgement

This work was supported by National Key R&D Program of China (2019YFC1711800), NSFC (62171138). We also thank the Student Author Grant provided by ISMIR 2022.

Contact: wzw@mail.ccom.edu.cn, weili-fudan@fudan.edu.cn