ChMusic: A Traditional Chinese Music Dataset for Evaluation of Instrument Recognition

Xia Gong
School of Music
Shandong University of Technology
Zibo, Chia
sdlggx@sdut.edu.cn

Yuxiang Zhu
No.2 High School (Baoshan) of
East China Normal University
Shanghai, China
3447535189@qq.com

Haidi Zhu
Shanghai Institute of Microsystem and Information Technology
Chinese Academy of Sciences
Shanghai, China
hdzhu@mail.sim.ac.cn

Haoran Wei
Department of Electrical and Computer Engineering
University of Texas at Dallas
Richardson, USA
haoran.wei@utdallas.edu

Abstract—**Musical instruments recognition is a widely used application for music information retrieval. As most of previous musical instruments recognition dataset focus on western musical instruments, it is difficult for researcher to study and evaluate the area of traditional Chinese musical instrument recognition. This paper propose a traditional Chinese music dataset for training model and performance evaluation, named ChMusic. This dataset is free and publicly available, 11 traditional Chinese musical instruments and 55 traditional Chinese music excerpts are recorded in this dataset. Then an evaluation standard is proposed based on ChMusic dataset. With this standard, researchers can compare their results following the same rule, and results from different researchers will become comparable.**

Index Terms—**Chinese instruments, music dataset, Chinese music, machine learning, evaluation of musical instrument recognition**

I. Introduction

Musical instruments recognition is an important and fascinating application for music information retrieval(MIR). Deep learning is great tool for image processing [1–3], video processing [4–6], natural language processing [7–9], speech processing [10–12] and music processing [13–15] applications and has demonstrated better performance compared with previous approaches. While deep learning based approaches rely on proper datasets to train model and evaluate performance.

Though there are already several musical instruments recognition dataset, most of these datasets only covers western musical instruments. For examples, Good-sounds dataset [16] has 12 musical instruments consisting of flute, cello, clarinet, trumpet, violin, sax_alto, sax_tenor, sax_baritone, sax_soprano, oboe, piccolo and bass. Another musical instruments recognition dataset called [17] has 20 musical instruments consisting of accordion, banjo, bass, cello, clarinet, cymbals, drums, flute, guitar, mallet_percussion, mandolin, organ, piano, saxophone, synthesizer, trombone, trumpet, ukulele, violin, and voice. Very few attention has been made to other culture's musical instruments recognition. Even for the very limited researches

base on Chinese musical instruments recognition [18, 19], dataset used for these research are not publicly available. Without an open access Chinese musical instruments dataset, Another critical problem is that researchers can not evaluate their model performance by a same standard, so results reported from their papers are not comparable.

To deal with the problems mentioned above, a traditional Chinese music dataset, named ChMusic, is proposed to help training Chinese musical instruments recognition models and then conducting performance evaluation. So the major contributions of this paper can be concluded as:

1) Propose a traditional Chinese music dataset for Chinese traditional instrument recognition, named ChMusic.

2) Come up with an evaluation standard to conduct Chinese traditional instrument recognition on ChMusic dataset.

3) Propose baseline approach for Chinese traditional instrument recognition on ChMusic dataset, code for this baseline is released through github https://github.com/HaoranWeiUTD/ChMusic

The rest of the paper is organized as follows: Section 2 covers details of ChMusic dataset. Then, the corresponding evaluation standard to conduct Chinese traditional instrument recognition on ChMusic dataset are described in Section 3. Baseline for Musical Instrument Recognition on ChMusic Dataset is proposed in Section 4. Finally, the paper is concluded in Section 5.

II. CHMUSIC DATASET

ChMusic is a traditional Chinese music dataset for training model and performance evaluation of musical instrument recognition. This dataset cover 11 musical instruments, consisting of Erhu, Pipa, Sanxian, Dizi, Suona, Zhuiqin, Zhongruan, Liuqin, Guzheng, Yangqin and Sheng, the indicating numbers and images are shown on Figure 1 respectively.

Each musical instrument has 5 traditional Chinese music excerpts, so there are 55 traditional Chinese music excerpts

Number	1	2	3	4
Name	Erhu	Pipa	Sanxian	Dizi
Image				
Number	5	6	7	8
Name	Suona	Zhuiqin	Zhongruan	Liuqin
Image				
Number	9	10		11
Name	Guzheng	Yangqin		Sheng
Image				

Fig. 1. Example of a traditional Chinese instrument.

in this dataset. The name of each music excerpt and the corresponding musical instrument is shown on Table I. Each music excerpt only played by one musical instrument in this dataset.

Each music excerpt is saved as a .wav file. The name of these files follows the format of "x.y.wav", "x" indicates the instrument number, ranging from 1 to 11, and "y" indicates the music number for each instrument, ranging from 1 to 5. These files are recorded by dual channel, and has sampling rate of 44100 Hz. The duration of these music excerpts are between 25 to 280 seconds.

This ChMusic dataset can be down-load from Baidu Wangpan by link pan.baidu.com/s/13e-6GnVJmC3tcwJtxed3-g and password xk23, or from google drive drive.google.com/file/d/1rfbXpkYEUGw5h_CZJ tC7eayYemeFMzij/view?usp=sharing

III. Evaluation of Musical Instrument Recognition on ChMusic Dataset

The diagram of musical instrument recognition can be simplified as Figure 2. Figure 2 consisting of two stages, corresponding to training stage and testing stage respectively. Training data is used for training stage to train a musical instrument recognition model. Then this model is tested by testing data during testing stage.

TABLE I
Musics played by traditional Chinese instruments

| Instrument Name	Music Number	Music Name
Erhu	1	"Ao Bao Xiang Hui"
	2	"Mu Yang Gu Niang"
	3	"Xiao He Tang Shui"
	4	"Er Xing Qian Li"
	5	"Yu Jia Gu Niang"
Pipa	1	"Shi Mian Mai Fu" excerpts
	2	"Su" excerpts
	3	"Zhu Fu" excerpts
	4	"Xian Zi Yun" excerpts
	5	"Ying Hua"
Sanxian	1	"Kang Ding Qing Ge"
	2	"Ao Bao Xiang Hui"
	3	"Gan Niu Shan"
	4	"Mao Zhu Xi De Hua Er Ji Xin Shang"
	5	"Nan Ni Wan"
Dizi	1	"Hong Mei Zan"
	2	"Shan Hu Song"
	3	"Wei Shan Hu"
	4	"Xiu Hong Qi"
	5	"Xiao Bai Yang"
Suona	1	"Hao Han Ge"
	2	"She Yuan Dou Shi Xiang Yang Hua"
	3	"Yi Zhi Hua"
	4	"Huang Tu Qing" excerpts
	5	"Liang Zhu" excerpts
Zhuiqin	1	"Jie Nian" excerpt 1
	2	"Jie Nian" excerpt 2
	3	"Hong Sao" excerpt
	4	"Jie Nian" excerpt 3
	5	"Zi Mei Yi Jia" excerpt
Zhongruan	1	"Yun Nan Hui Yi" excerpt 1
	2	"Yun Nan Hui Yi" excerpt 2
	3	"Yun Nan Hui Yi" excerpt 3
	4	"Yun Nan Hui Yi" excerpt 4
	5	"Yun Nan Hui Yi" excerpt 5
Liuqin	1	"Chun Dao Yi He" excerpt 1
	2	"Yu Ge" excerpt
	3	"Chun Dao Yi He" excerpt 2
	4	"Yu Hou Ting Yuan" excerpt
	5	"Jian Qi" excerpt
Guzheng	1	"Lin Chong Ye Ben" excerpt 1
	2	"Zhan Tai Feng" excerpt
	3	"Yu Zhou Chang Wan"
	4	"Lin Chong Ye Ben" excerpt 2
	5	"Han Jiang Yun"
Yangqin	1	"Zuo Zhu Fa Lian Xi Qu"
	2	"Fen Jie He Xian Lian Xi Qu"
	3	"Dan Yin Yu He Yin Jiao Ti Lian Xi Qu"
	4	"Lun Yin Lian Xi Qu"
	5	"Da Qi Luo Gu Qing Feng Shou"
Sheng	1	"Gua Hong Deng" excerpt
	2	"Qin Wang Po Zhen Yue" excerpt
	3	"Shui Ku Yin Lai Jin Feng Huang"
	4	"Shan Xiang Xi Kai Feng Shou Lian"
	5	"Xi Ban Ya Dou Niu Wu Qu"
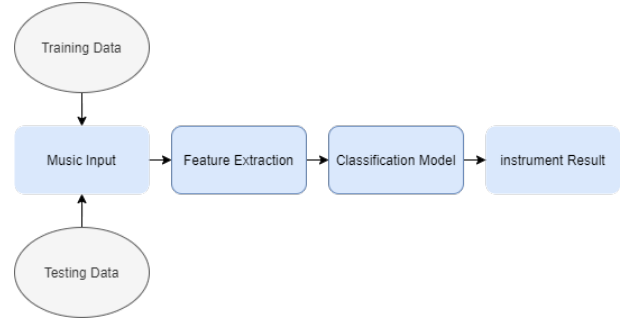
Fig. 2. Diagram of musical instrument recognition.

The widely used features for musical instrument recognition include linear predictive cepstral coefficient (LPCC) [20] and mel-scale frequency cepstral coefficients (MFCC) [21] . The commonly used classification models include k-nearest neighbors (KNN) model[18] , Gaussian mixture model(GMM)[22] , support vector machine (SVM) model[23], hidden markov model (HMM) [24] and deep learning based models [25] .

In ChMusic Dataset, musics with number 1 to 4 of each instrument are used as training dataset, musics with number 5 of each instrument are used as testing dataset. Every music is cut into 5 seconds clips without overlap, the ending part of each music which is shorter than 5 seconds is discarded. Each 5 seconds clips need to make a prediction for musical instrument classification result.

$$Accuracy = \frac{CorrectClassifiedClips}{AllClipsFromTestingData} \quad (1)$$

Classification accuracy will be the evaluation metrics for musical instrument recognition on ChMusic Dataset. And confusion matrix is also recommended to present the result.

IV. Baseline for Musical Instrument Recognition on ChMusic Dataset

Baseline for musical instrument recognition on ChMusic dataset is described in this section. Twenty dimension MFCCs are extracted from each frame, and KNN is used for frame-level data training and testing. As each test clip have 5 minutes in duration, a majority voting is conducted to get final classification result from frame-level resuls within the 5 minutes test clip.

TABLE II
Accuracy for KNN Models with Various K Value

Model	Accuracy(%)
KNN (K=3)	93.57
KNN (K=5)	93.57
KNN (K=9)	93.57
KNN (K=15)	94.15

KNN models with various K value are compared. Accuracy for KNN models with each K value are presented in Table II then confusion matrix for K=3, K=5, K=9 and K=15 are presented in Figure 3, Figure 4, Figure 5 and Figure 6 respectively.
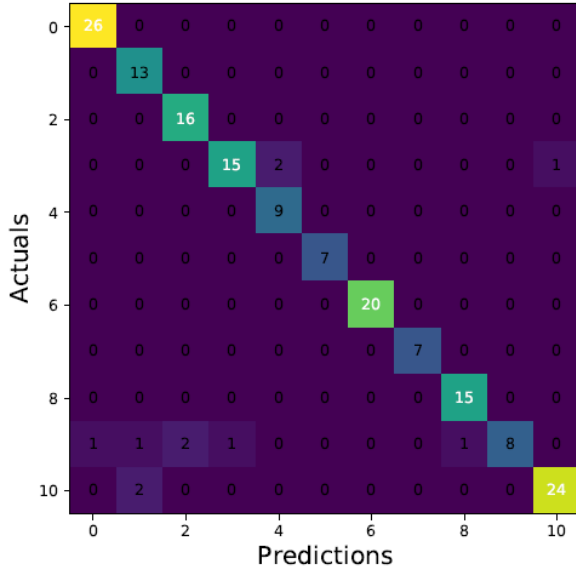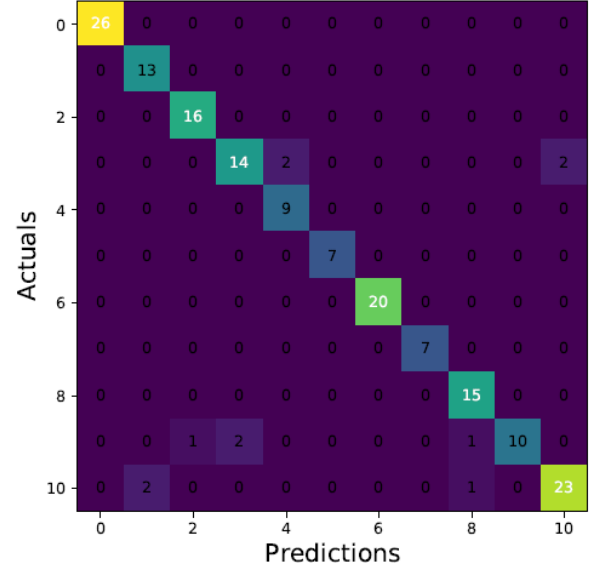
Fig. 3. Confusion matrix of KNN model with K=3.

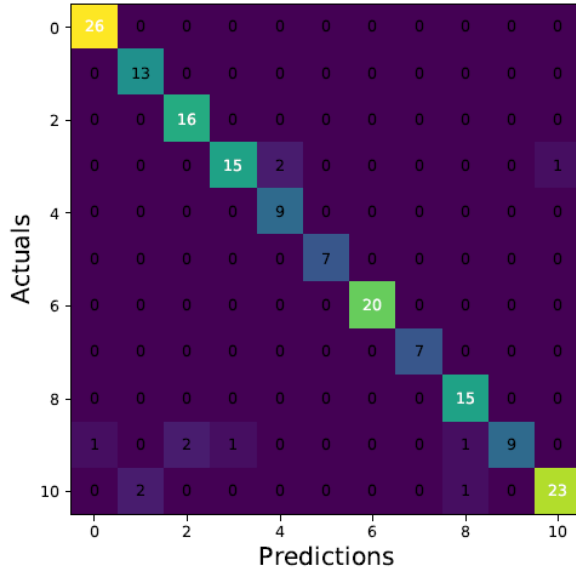Fig. 5. Confusion matrix of KNN model with K=9.
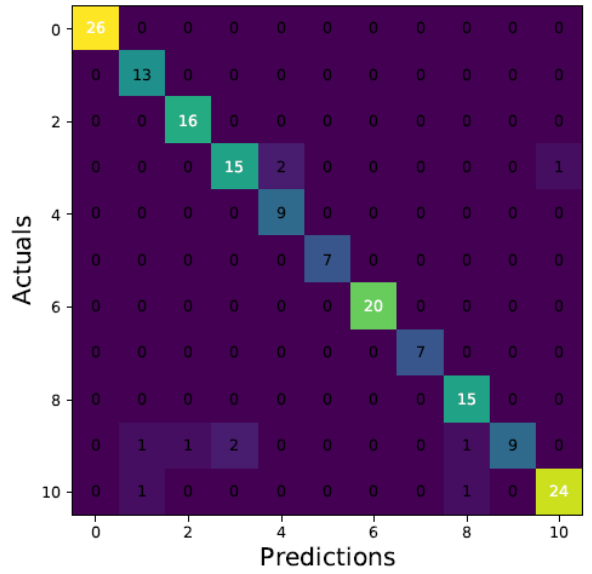

Fig. 4. Confusion matrix of KNN model with K=5.

Fig. 6. Confusion matrix of KNN model with K=15.

V. Conclusion

Musical instruments recognition is an interesting application for music information retrieval. While most of previous musical instruments recognition dataset focus on western musical instruments, very few works have been conducted on the area of traditional Chinese musical instrument recognition. This paper propose a traditional Chinese music dataset for training model and performance evaluation, named ChMusic. With this dataset, researchers can download it for free. Eleven traditional Chinese musical instruments and 55 traditional Chinese music excerpts are covered in this dataset. Then an evaluation standard is proposed based on ChMusic dataset. With this standard, researchers can compare their results following the same rule, and results from different researchers will become comparable.

VI. Acknowledgment

REFERENCES

[1] A. Azarang and H. Ghassemian, "Application of fractional-order differentiation in multispectral image fusion," *Remote sensing letters*, vol. 9, no. 1, pp. 91–100, 2018.

[2] A. Azarang and N. Kehtarnavaz, "A generative model method for unsupervised multispectral image fusion in remote sensing," *Signal, Image and Video Processing*, pp. 1–9, 2021.

[3] H. Wei, A. Sehgal, and N. Kehtarnavaz, "A deep learning-based smartphone app for real-time detection of retinal abnormalities in fundus images," in *Real-Time Image Processing and Deep Learning 2019*, vol. 10996. International Society for Optics and Photonics, 2019, p. 1099602.

[4] H. Zhu, H. Wei, B. Li, X. Yuan, and N. Kehtarnavaz, "A review of video object detection: Datasets, metrics and methods," *Applied Sciences*, vol. 10, no. 21, p. 7834, 2020.

[5] H. Wei and N. Kehtarnavaz, "Simultaneous utilization of inertial and video sensing for action detection and recognition in continuous action streams," *IEEE Sensors Journal*, vol. 20, no. 11, pp. 6055–6063, 2020.

[6] H. Zhu, H. Wei, B. Li, X. Yuan, and N. Kehtarnavaz, "Real-time moving object detection in high-resolution video sensing," *Sensors*, vol. 20, no. 12, p. 3591, 2020.

[7] X. Sun, L. Jiang, M. Zhang, C. Wang, and Y. Chen, "Unsupervised learning for product ontology from textual reviews on e-commerce sites," in *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, 2019, pp. 260–264.

[8] L. Jiang, O. Biran, M. Tiwari, Z. Weng, and Y. Benajiba, "End-to-end product taxonomy extension from text reviews," in *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*. IEEE, 2019, pp. 195–198.

[9] X. Li, L. Vilnis, D. Zhang, M. Boratko, and A. McCallum, "Smoothing the geometry of probabilistic box embeddings," in *International Conference on Learning Representations*, 2018.

[10] R. Su, F. Tao, X. Liu, H. Wei, X. Mei, Z. Duan, L. Yuan, J. Liu, and Y. Xie, "Themes informed audio-visual correspondence learning," *arXiv preprint arXiv:2009.06573*, 2020.

[11] Y. Zhang, Y. Long, X. Shen, H. Wei, M. Yang, H. Ye, and H. Mao, "Articulatory movement features for short-duration text-dependent speaker verification," *International Journal of Speech Technology*, vol. 20, no. 4, pp. 753–759, 2017.

[12] N. Alamdari and N. Kehtarnavaz, "A real-time smartphone app for unsupervised noise classification in realistic audio environments," in *2019 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2019, pp. 1–5.

[13] E. Benetos, S. Dixon, Z. Duan, and S. Ewert, "Automatic music transcription: An overview," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 20–30, 2018.

[14] R. Lu, K. Wu, Z. Duan, and C. Zhang, "Deep ranking: Triplet matchnet for music metric learning," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 121–125.

[15] P. Shreevathsa, M. Harshith, A. Rao *et al.*, "Music instrument recognition using machine learning algorithms," in *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)*. IEEE, 2020, pp. 161–166.

[16] G. Bandiera, O. Romani Picas, H. Tokuda, W. Hariya, K. Oishi, and X. Serra, "Good-sounds. org: A framework to explore goodness in instrumental sounds," in *Proceedings of the 17th International Society for Music Information Retrieval Conference*. International Society for Music Information Retrieval (ISMIR), 2016.

[17] E. Humphrey, S. Durand, and B. McFee, "Openmic-2018: An open data-set for multiple instrument recognition." in *ISMIR*, 2018, pp. 438–444.

[18] J. Shen, "Audio feature extraction and classification of the chinese national musical instrument," *Computer & Digital Engineering*, 2012.

[19] H. Shi, W. Xie, T. Xu, X. Zhu, and Q. Li, "Research and realization of identification methods on ethnic musical instruments," *Shanxi Dianzi Jishu*, 2020.

[20] J. Picone, "Fundamentals of speech recognition: A short course," *Institute for Signal and Information Processing, Mississippi State University*, 1996.

[21] H. Wei, Y. Long, and H. Mao, "Improvements on self-adaptive voice activity detector for telephone data," *International Journal of Speech Technology*, vol. 19, no. 3, pp. 623–630, 2016.

[22] Y. Anzai, *Pattern recognition and machine learning*. Elsevier, 2012.

[23] Y. Long, W. Guo, and L. Dai, "An sipca-wccn method for svm-based speaker verification system," in *2008 International Conference on Audio, Language and Image Processing*. IEEE, 2008, pp. 1295–1299.

[24] P. Herrera-Boyer, A. Klapuri, and M. Davy, "Automatic classification of pitched musical instrument sounds," in *Signal processing methods for music transcription*. Springer, 2006, pp. 163–200.

[25] H. Wei and N. Kehtarnavaz, "Determining number of speakers from single microphone speech signals by multi-label convolutional neural network," in *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2018, pp. 2706–2710.