華中科技大學
HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

MATT: A Multiple-instance Attention Mechanism for Long-tail Music Genre Classification

Huazhong University of Science & Technology
Xiaokai Liu, Shihui Song, Menghua Zhang, Yafan Huang*

Outline

- Background
- Related Works
- Approach
- Experiments
- Conclusion

Music Genre Classification

- Music Genre Classification(Genre)
  - Identify the genres of given music segments, which is an essential task in the research field of Music Information Retrieval (MIR)

What genre is this music?

Classical Music

Long-tail Music Genre Identification

- Most Genres have only a few training instances.

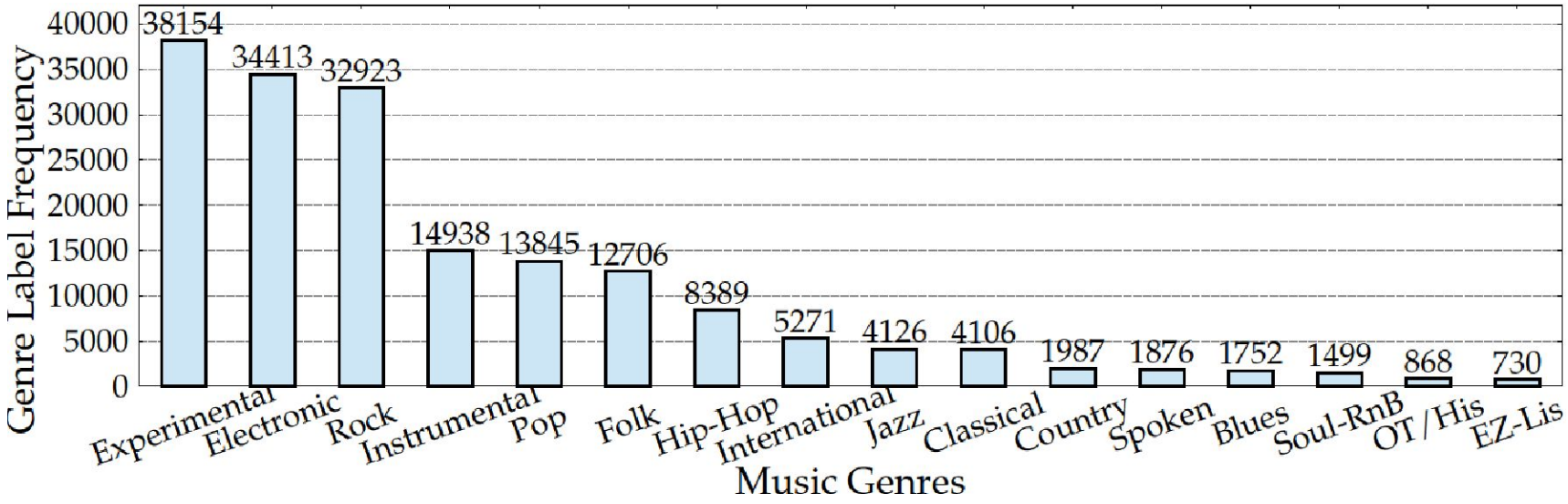


Fig. 1. Music Genre Distribution of FMA Dataset.

Related Works

dataset[1]	#clips	#artists	year	audio
RWC [12]	465	-	2001	yes
CAL500 [45]	500	500	2007	yes
Ballroom [13]	698	-	2004	yes
GTZAN [46]	1,000	∼ 300	2002	yes
MusiClef [36]	1,355	218	2012	yes
Artist20 [7]	1,413	20	2007	yes
ISMIR2004	1,458	-	2004	yes
Homburg [15]	1,886	1,463	2005	yes
103-Artists [30]	2,445	103	2005	yes
Unique [41]	3,115	3,115	2010	yes
1517-Artists [40]	3,180	1,517	2008	yes
LMD [42]	3,227	-	2007	no
EBallroom [23]	4,180	-	2016	no[2]
USPOP [1]	8,752	400	2003	no
CAL10k [44]	10,271	4,597	2010	no
MagnaTagATune [20]	25,863[3]	230	2009	yes[4]
Codaich [28]	26,420	1,941	2006	no
FMA	**106,574**	**16,341**	**2017**	**yes**
OMRAS2 [24]	152,410	6,938	2009	no
MSD [3]	1,000,000	44,745	2011	no[2]
AudioSet [10]	2,084,320	-	2017	no[2]
AcousticBrainz [32]	2,524,739[5]	-	2017	no

Comparison between FMA and alternative datasets.

dataset	clips	genres	length	size	
			[s]	[GiB]	#days
small	8,000	8	30	7.4	2.8
medium	25,000	16	30	23	8.7
large	106,574	161	30	98	37
full	106,574	161	278	917	343

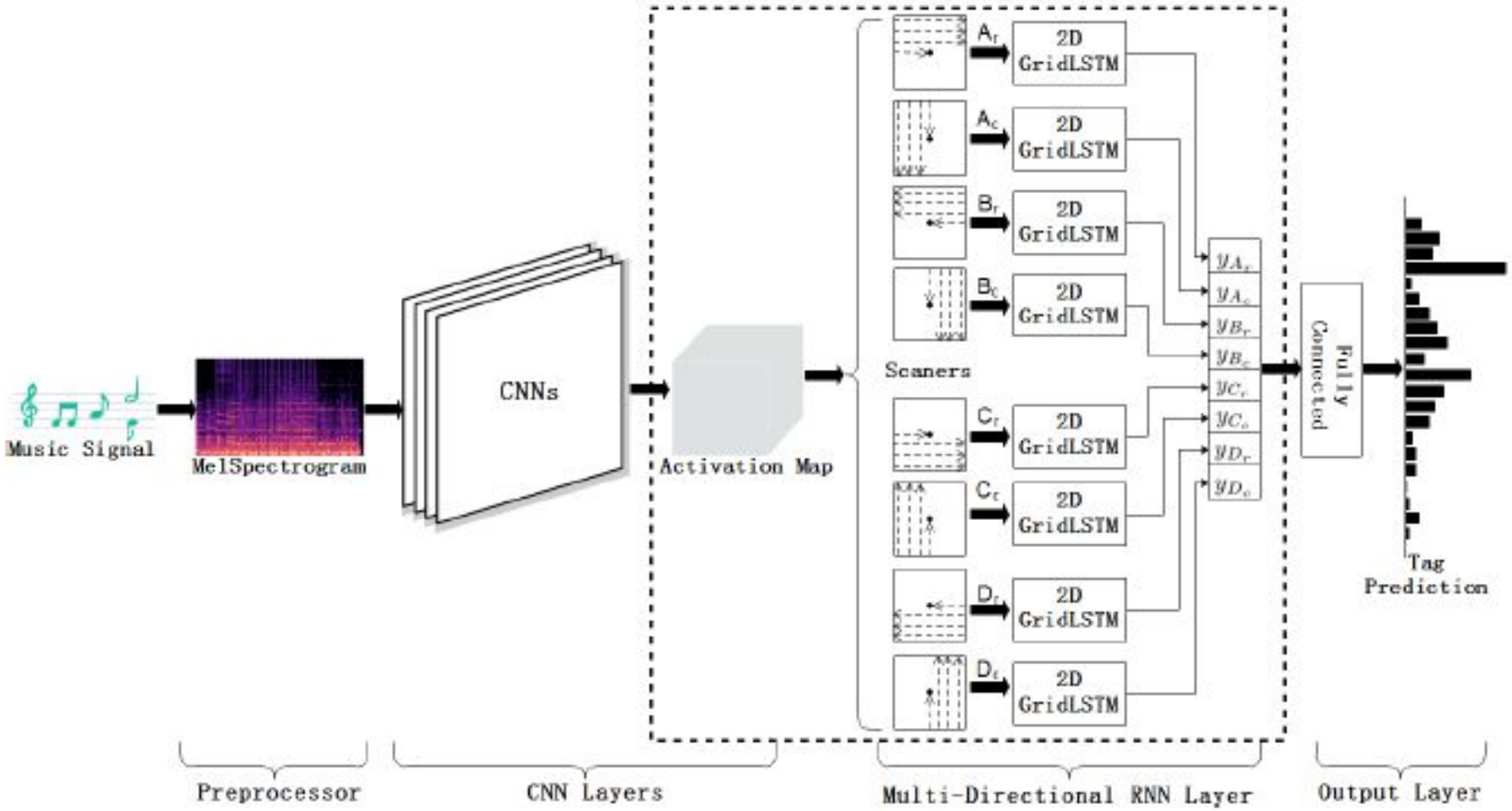
Proposed subsets of the FMA.

feature set	dim.	LR	kNN	SVM	MLP
1 Chroma [11]	84	44	44	48	49
2 Tonnetz [14]	42	40	37	42	41
3 MFCC [33]	140	58	55	61	53
4 Spec. centroid	7	42	45	46	48
5 Spec. bandwidth	7	41	45	44	45
6 Spec. contrast [17]	49	51	50	54	53
7 Spec. rolloff	7	42	46	48	48
8 RMS energy	7	37	39	39	39
9 Zero-crossing rate	7	42	45	45	46
3 + 6	189	60	55	63	54
3 + 6 + 4	273	60	55	63	53
1 to 9	518	61	52	63	58

Test set accuracies of various features and classifiers for top genre recognition on the medium subset.

Defferrard, Michaël, et al. "FMA: A DATASET FOR MUSIC ANALYSIS."

Related Works



Wang, Zhen, et al. "Music classification using an improved CRNN with multi-directional spatial dependencies in both time and frequency dimensions." *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019.

Approach

Music Instance Preprocessor

Given an instance and its related album ID and artist ID pair, we employ multi-feature extraction methods to embed music audio segments into continuous vector space.

Music Instance Encoder:

The encoding layer aims to convert given instances into their corresponding vector representation and maintains their musical semantics at the same time.

Multiple-instance Attention Mechanism

Under the guidance of the final audio embeddings, MATT can identify the most informative music segment exactly matching the relevant genre.

Music Instance Preprocessor

We split all musical audio segments into multiple album-artist bags and denote them as $\{S1, S2, \ldots\}$.

Each bag S contains music instances $\{s1, s2, \ldots\}$ with the same artist ID $\mathcal{P}$ and album ID $\mathcal{A}$.

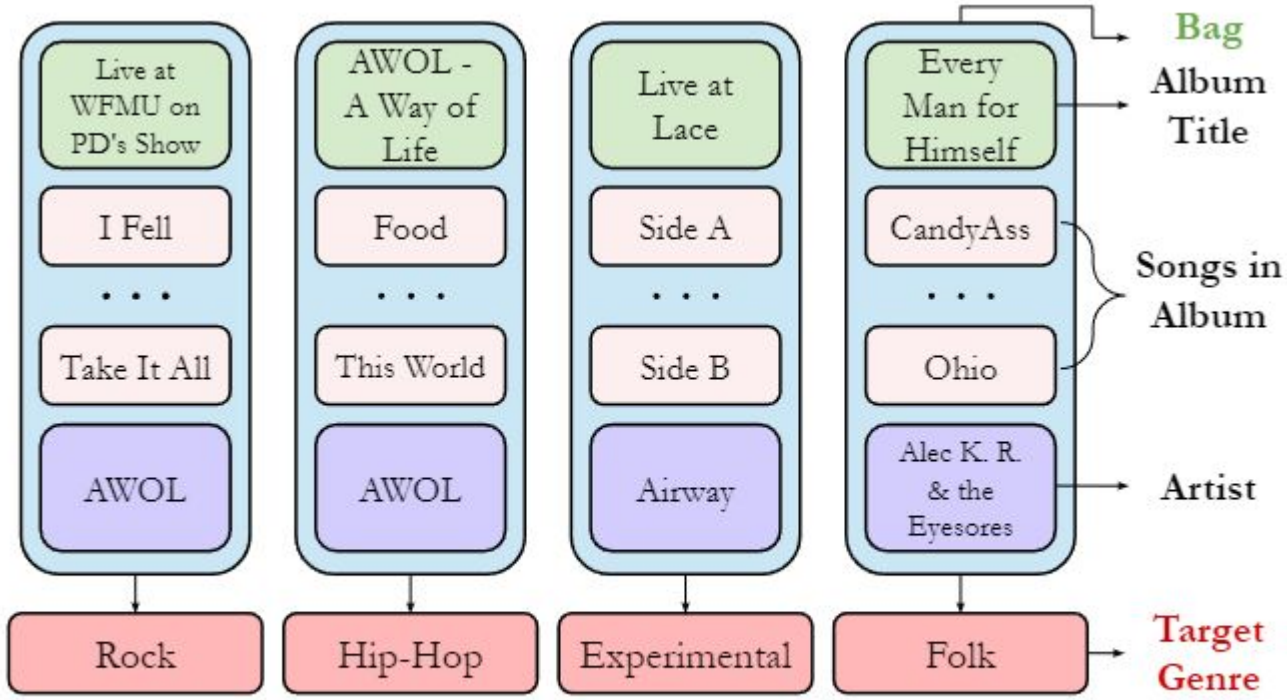
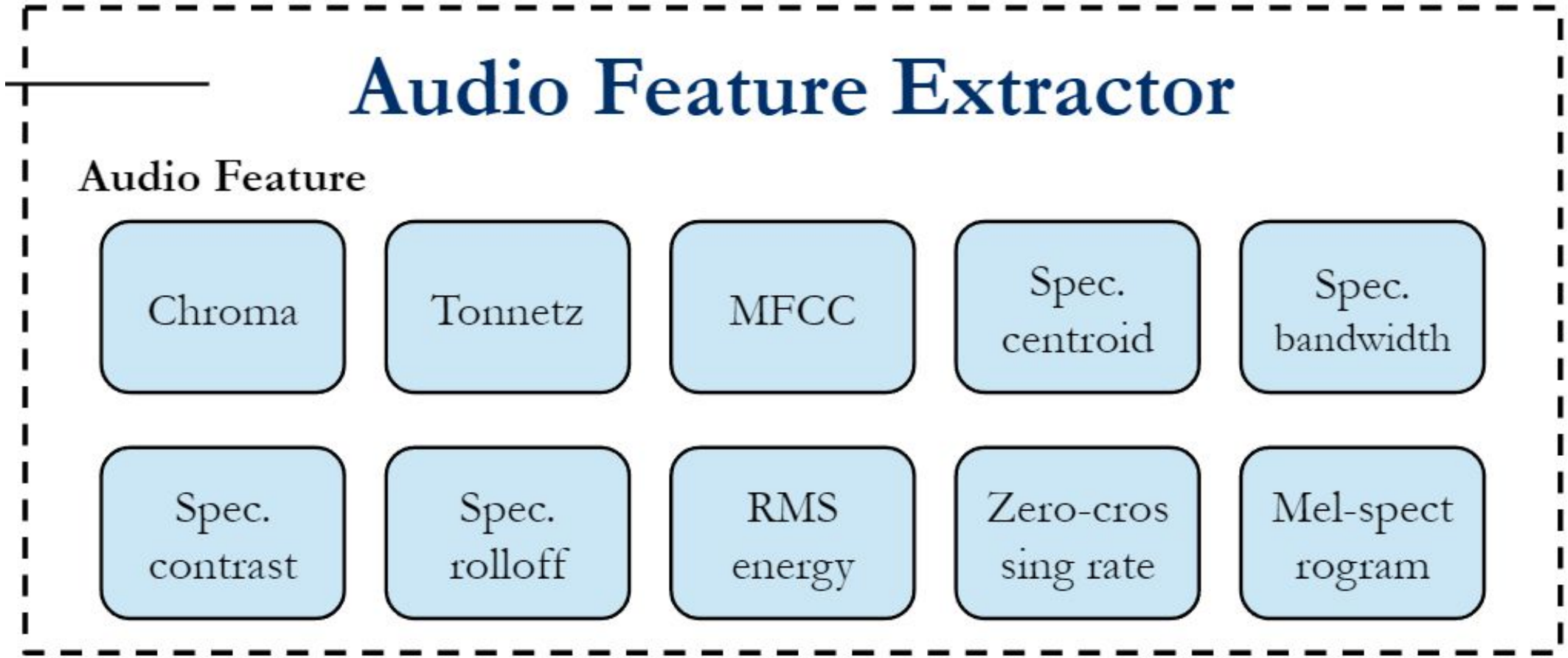
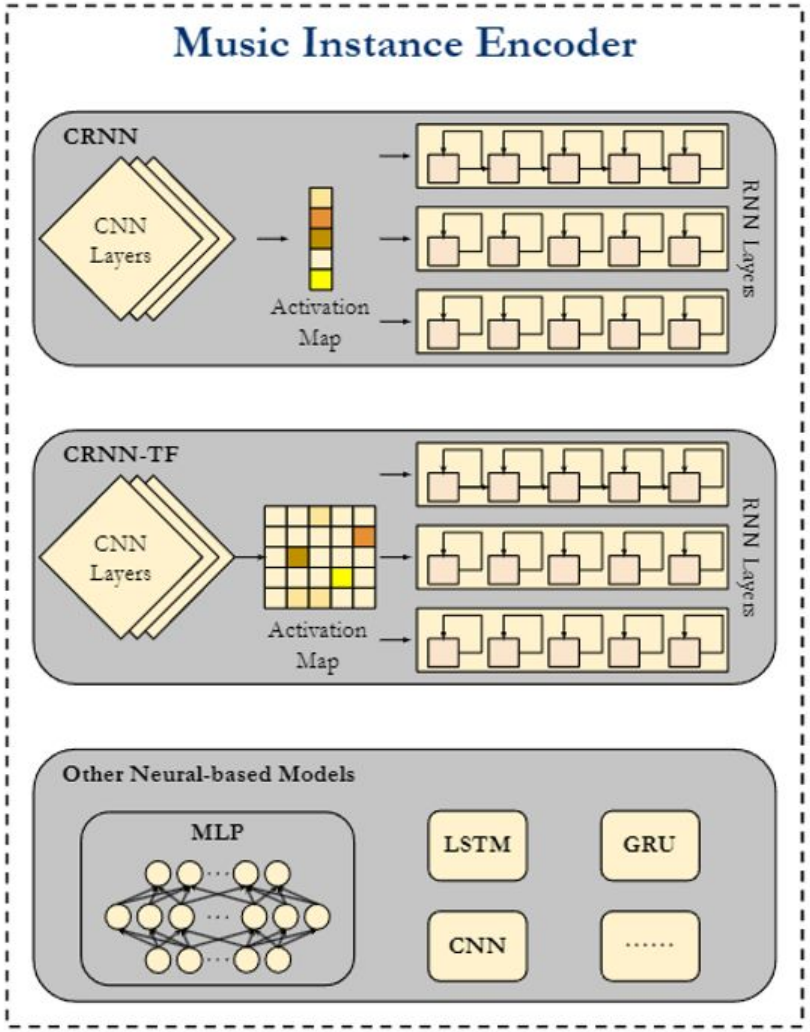


Fig. 2. An Example of MIL for Music Genre Classification.

Audio Feature Extraction



Music Instance Encoder

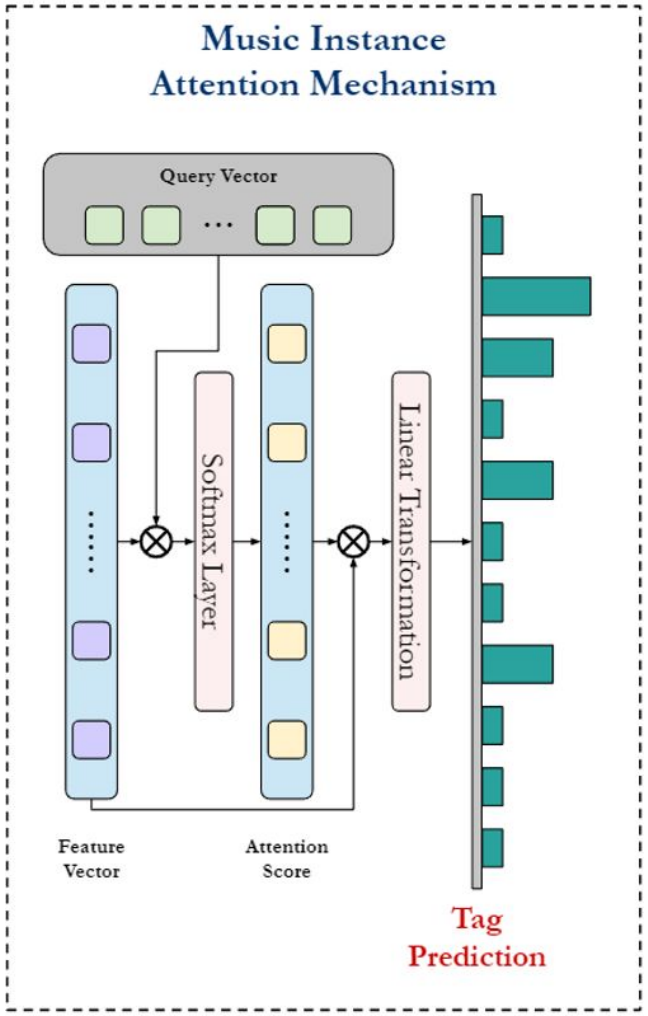


We choose neural networks with convolutional layers, e.g. the CRNN and CRNN-TF, to encode input embeddings extracted from the log-amplitude mel-spectrogram feature extraction method.

The convolutional recurrent neural network (CRNN) consists of convolutional layers and Gated recurrent unit (GRU) layers.

The convolutional recurrent neural network in Time and Frequency dimensions (CRNN-TF) is a variant of CRNN. It can extract spatial dependencies in both the Time and Frequency dimensions of music signals. CRNN-TF has achieved promising performances in several state-of-the-art deep learning-based music models.

Multiple-instance Attention Mechanism



Under the guidance of the final audio embeddings, MATT can identify the most informative music segment exactly matching the relevant genre.

Attention Score:

$$e_k = \tanh(W_s[s_k; q_g]) + b_s \quad (1)$$

$$a_k = \frac{\exp(e_k)}{\sum_{j=1}^{m} \exp(e_j)} \quad (2)$$

Global Representation:

$$g_{p,a} = \mathrm{ATT}(q_g, s_1, s_2, ..., s_m) \quad (3)$$

Output Classification Score

$$P(g|h, t, S_{p,a}) = \frac{\exp(o_r)}{\sum_{\hat{g} \in \mathcal{G}} \exp(o_{\hat{g}})} \quad (4)$$

$$o = M g_{p,a} \quad (5)$$

Overall Architecture

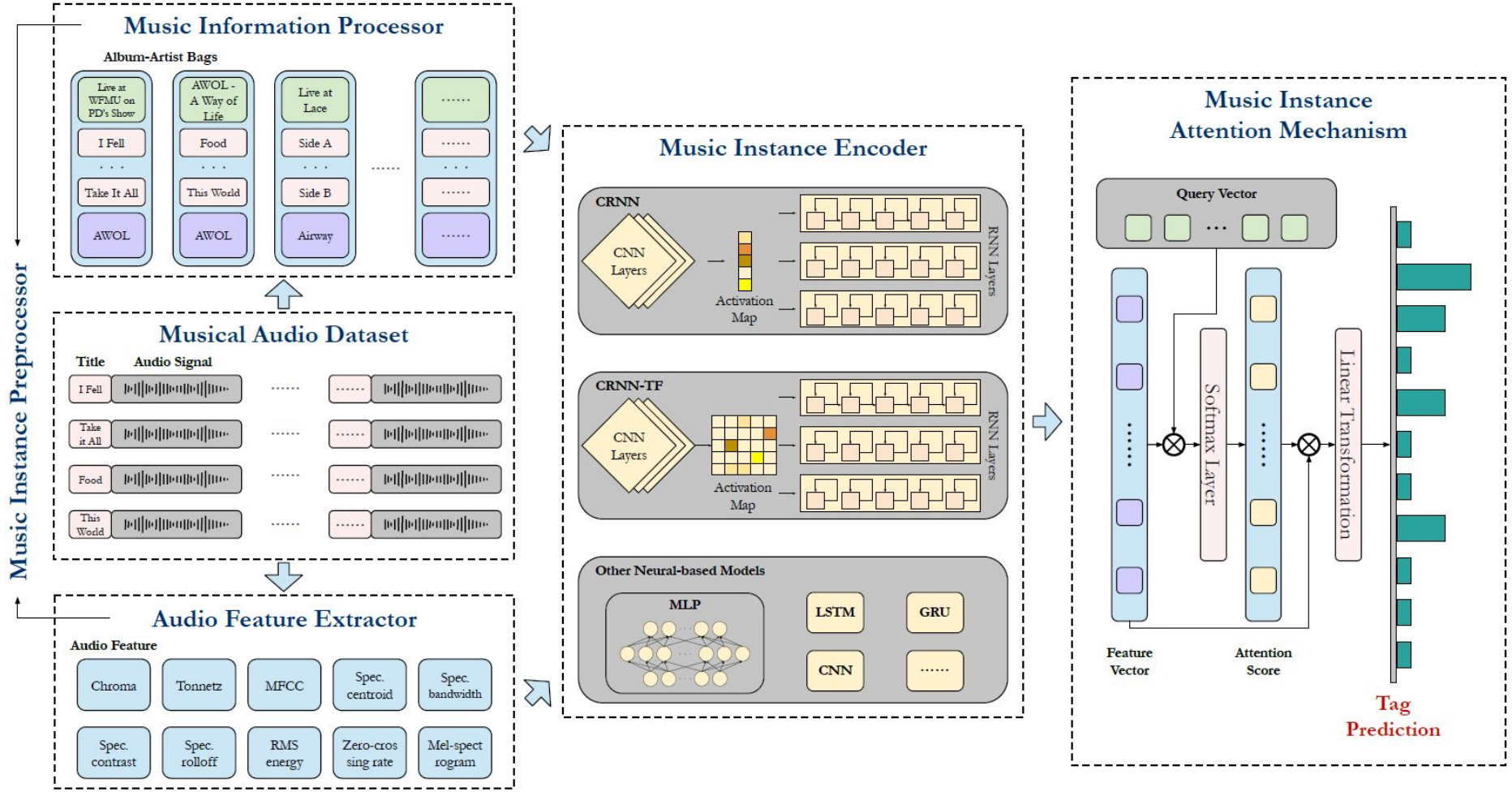
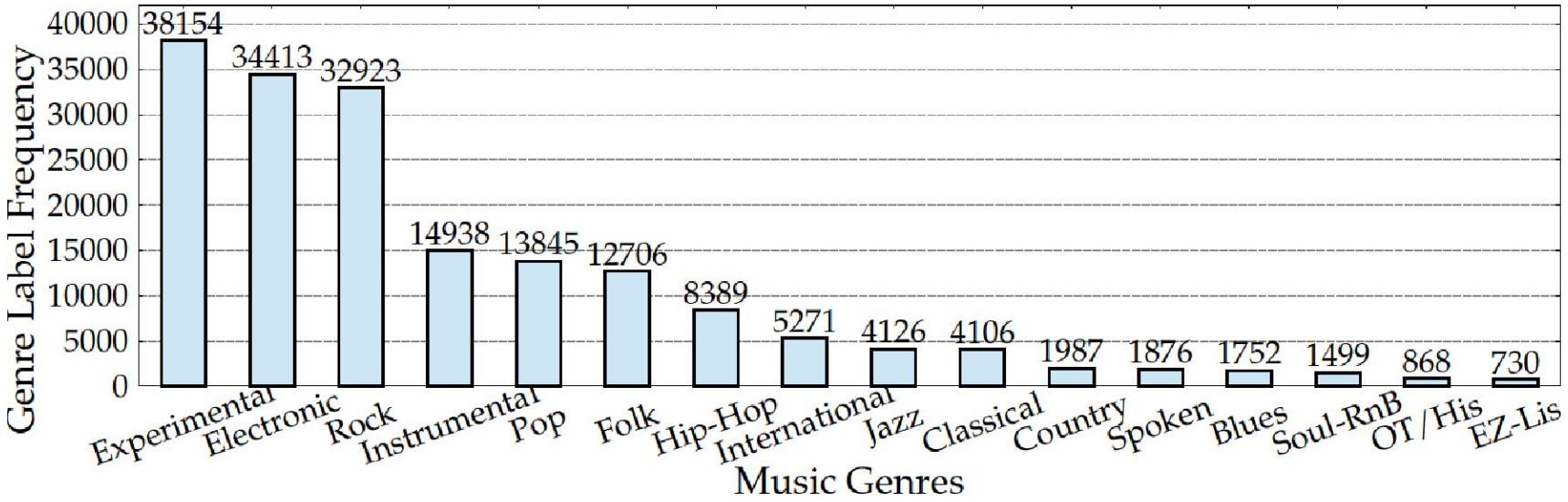


Fig. 3. The Overall Workflow of the MATT-based Music Genre Classification Model.

Experiments

FMA dataset (Medium Size: 16 Music Genres)

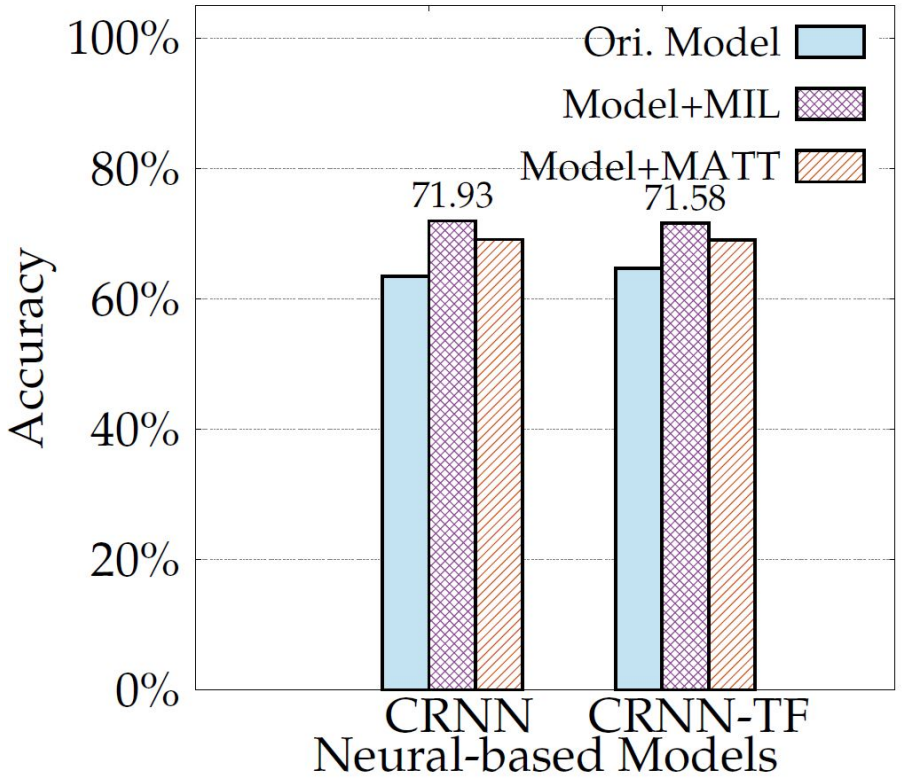


To make the results reproducible, we adopt the default data split schema proposed by FMA and the default hyperparameter settings from baseline methods.

Overall Evaluation Results

TABLE I
TESTING ACCURACY (%) OF VARIOUS FEATURES AND MODELS ON THE FMA DATA MEDIUM SUBSET

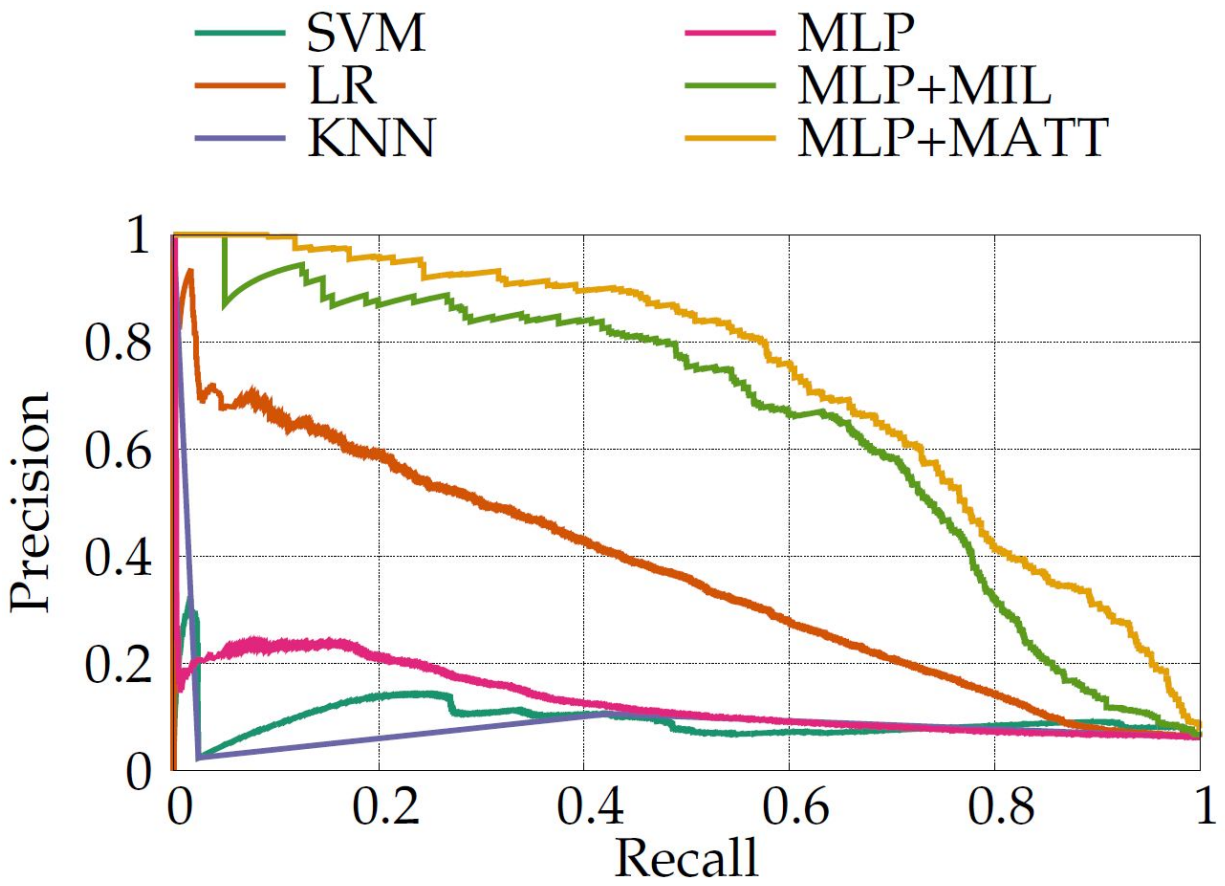
Feature Set	Dim.	LR	KNN	SVM	MLP	MLP+MIL	MLP+MATT
1 Chroma	84	44	44	48	49	**47.74**	39.53
2 Tonnetz	42	40	37	42	41	**45.90**	43.45
3 MFCC	140	58	55	61	53	**64.71**	63.19
4 Spec. Centroid	7	42	45	46	48	**49.36**	44.93
5 Spec. BW..	7	41	45	44	45	43.26	**44.23**
6 Spec. Contrast	49	51	50	54	53	**55.69**	49.59
7 Spec. Rolloff	7	42	46	48	48	49.16	**49.20**
8 RMS Energy	7	37	39	39	39	**41.08**	38.67
9 Zero-Crossing	7	42	45	45	46	**47.96**	46.91
3+6	189	60	55	63	54	65.68	**68.91**
3+6+4	196	60	55	63	53	**65.84**	63.31
1 to 9	518	61	52	63	58	**69.22**	68.40



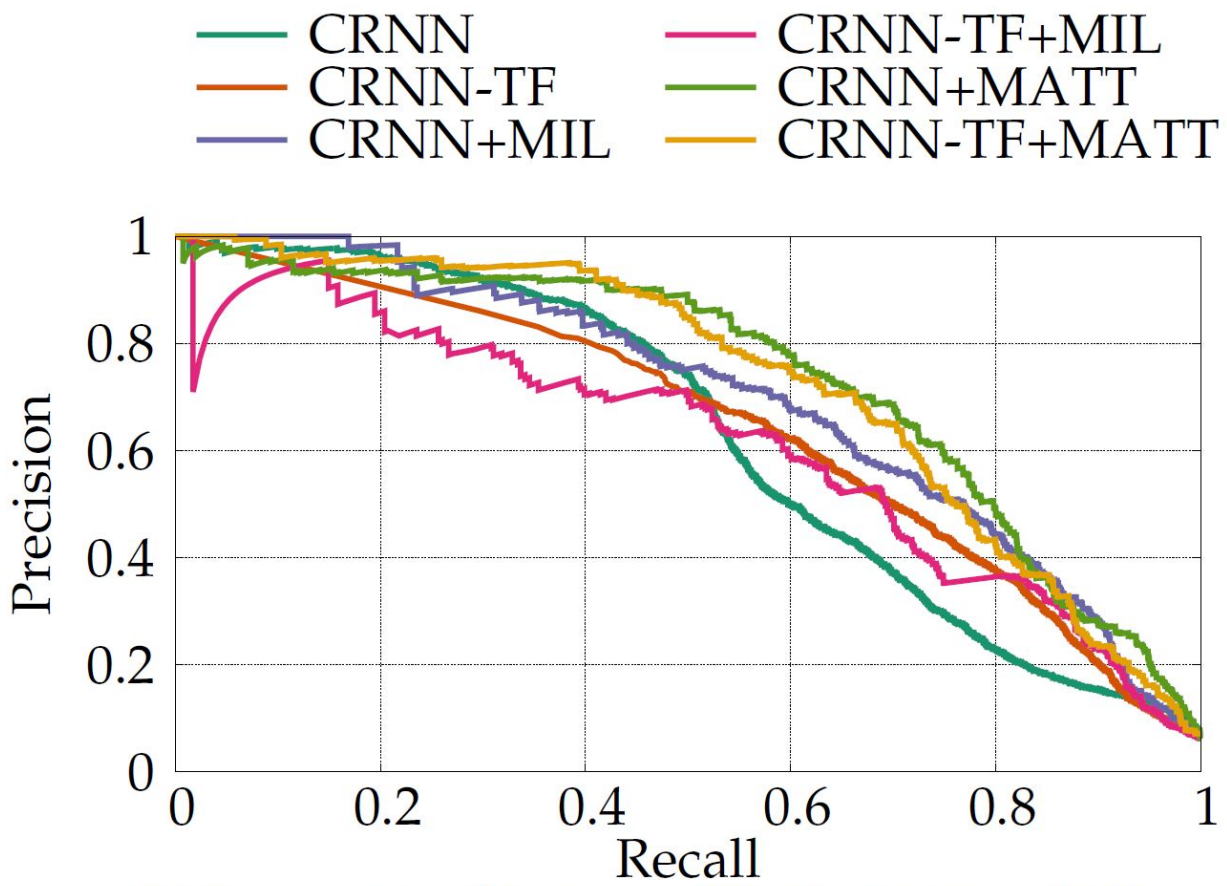
(c) Testing performance of neural-based models on the FMA Dataset.

Fig. 4. PR-Curve and Accuracy of Our Proposed Models against Baselines

Overall Evaluation Results



(a) Precision-recall curves of the proposed methods vs feature-based baselines.



(b) Precision-recall curves of the proposed methods vs neural-based baselines.

Fig. 4. PR-Curve and Accuracy of Our Proposed Models against Baselines

Evaluation Results for Long-tail Genres

TABLE II
TOP@K ACCURACY (%) ON LONG-TAIL CLASSES.

Number of Training Instances	<100			<200		
Top@K	**K=2**	**K=3**	**K=5**	**K=2**	**K=3**	**K=5**
1 LR	<5.0	7.06	25.29	<5.0	6.38	22.87
2 KNN	17.06	25.88	45.29	15.43	23.94	41.49
3 SVM	11.76	19.41	44.71	10.64	17.55	40.42
4 MLP	<5.0	10.59	28.23	<5.0	10.11	27.13
5 MLP+MATT	**28.24**	**44.71**	**47.06**	**26.60**	**41.49**	**44.15**
6 CRNN	9.41	13.53	34.71	8.51	12.23	38.83
7 CRNN + MATT	**54.12**	**55.88**	**57.65**	**49.47**	**51.60**	**61.7**
8 CRNN-TF	<5.0	9.41	20.59	<5.0	8.51	18.62
9 CRNN-TF + MATT	**54.12**	**63.53**	**65.29**	**59.04**	**61.17**	**71.21**

Case Study

- In the training stage, we train the models using the bag-level dataset. In the testing phase, we evaluate the models without using the multi-instance learning strategy. In other words, the model is only equipped with a plain attention mechanism in the evaluation stage.

- We conduct experiments using MLP with the whole FMA feature sets. By training the MLP model on the bag-level dataset using features 1 to 9 from the feature set, the testing accuracy degrades from 68.83% to 63.69%.

- However, we found that the testing accuracy is still greater than that of the training model at the segment level, whose testing accuracy is only 58%.

Conclusion

- In this paper, we propose the MATT mechanism to identify music genres, especially the long-tail genres, in a large-scale music benchmark. And the CRNN-TF MATT reaches SOTA performance in identifying long-tail music genres.

- Comprehensive experimental results demonstrate that the MATT can significantly improve the performance of the MGC models for identifying long-tail music genres.

- Our work points out a new direction for the future work of music classification——classify music genres based on the bag-level dataset.

QA

liuxk@hust.edu.cn

Code :
https://github.com/johannesliu/Music-Genre-Classifcation
Personal Website: https://johannesliu.github.io
Music Work Website: https://www.weekendcomposer.com