

FMA: A Dataset For Music Analysis

Kirell Benzi, Michaël Defferrard, Pierre Vandergheynst, Xavier Bresson
LTS2, EPFL, Lausanne, Switzerland

November 29, 2016

Abstract

We present a new music dataset that can be used for several music analysis tasks. Our major goal is to go beyond the existing limitations of available music datasets, which are either the small size of datasets with raw audio tracks, availability and legality of the music data, or the lack of meta-data for artists analysis or song ratings for recommender systems. Existing datasets such as GTZAN, TagATune, and Million Song Dataset suffer from the previous limitations. It is however essential to establish such benchmark datasets to advance the field of music analysis, like the ImageNet dataset that made possible the large success of deep learning techniques in computer vision. In this paper, we introduce the Free Music Archive (FMA) dataset which size is 86,992 songs with raw audio tracks and meta-data including artist name, song title, music genre, and track counts. For research purposes, we define 2 additional datasets from the original one; a small genre-balanced dataset of 1,000 data and a medium genre-unbalanced dataset of 14,570 data, both with audio tracks and Echonest features. For all datasets, we provide a train-test splitting for future algorithms' comparisons.

1 Introduction and Motivation

Music Analysis Field. Developing new mathematical models and algorithms to solve challenging real-world problems is obviously of first importance in any field of research. But such novel techniques must be evaluated and compared to the existing state-of-the-art techniques to be adopted as new standards by research communities. Evaluation and comparison require benchmark datasets to achieve their purpose, which are sometimes challenging to collect and share. In computer vision, the community has developed over the years well-established benchmark datasets such as MNIST [16] publicly accessible at [17], CIFAR [13] available at [12], or ImageNet [8] provided at [7]. All these image datasets are free, legal and easily available on Internet. Such datasets have proved essential to advance the field of computer vision. The most celebrated example is the ImageNet dataset and challenge in 2012. This unprecedented dataset of 1.5M image data allowed to

demonstrate the power of deep learning techniques, which was able to win the competition by a large margin over the second best. In music analysis, such benchmark datasets are not easily available and they also lack of essential information. For examples, developing deep learning techniques for music applications require the use of raw audio data, and designing recommender systems need meta-data like artist features and song ratings. Besides, the most influential competition in the field of music analysis organized every year is the Music Information Retrieval Evaluation eXchange (MIREX) [6]. MIREX proposes several important music analysis challenges such as song identification, tag classification, music similarity and retrieval, etc. However, participants do not have access to the dataset; neither the test set, nor the important train set. They must upload their code that will be evaluated by the organizers. In other words, participants cannot train their analysis models on any part of the dataset. These existing shortcomings make essential the development of new music datasets to advance the field of music analysis algorithms.

Available music datasets and their limitations. In music analysis, three datasets have been widely used by the international society for music information retrieval (ISMIR) [10]: the GTZAN, TagATune, and MSD (Million Song Dataset) datasets. However each dataset come with some limitations:

- GTZAN [21] is a collection of 1000 songs with 10 music genres. Each song is represented by a 22050Hz Mono 16-bit wav audio file of 30sec. It is the most popular music dataset with 6,600 answers on Google search, and 600 on Google scholar on Nov 2016. It has been widely used for different music applications. A complete study of this dataset has been done in [19], and the dataset is available at [20]. The main limitations of GTZAN is the legality of the dataset, the small size with 1000 data, no complete meta-data regarding artist names and song titles, and no additional meta-data like ratings.
- TagATune [15] was a popular dataset of 5,405 source songs of 230 artists. It was used for music tagging and music metric learning. Its website is [14], but it is not anymore publicly available since 2009.
- The Million Song Dataset (MSD) [5] is a free and legal collection of 1 million music songs. Each song data is composed of high-level and medium-level audio features provided by Echonest service and meta-data like artist name, song title, track genre, etc. It is available at [4]. The very large scale of this music dataset would make it an ideal candidate for deep learning, which works best with large datasets. However, deep learning techniques work directly on the raw data (audio tracks in this case, like raw image pixels in computer vision) and not on pre-processed audio features given by Echonest. Unfortunately, this prevents the application of deep learning on this dataset. Finally, Echonest reassigned the indexes of the MSD songs, which today makes (almost) impossible the use of the most music recent features developed by Echonest.

What would be the requirements for an ideal music dataset?

- (1) Availability of raw audio data like 30sec or more
- (2) Availability of meta-data such as genres, artist, title, year, lyrics, users' ratings, users' comments, etc

- (3) Large-scale dataset to be able to sample the high-dimensional distribution of music genres
- (4) Free and easy access of the dataset
- (5) Legality of the dataset, no copyright issue with e.g. open access licence

2 Introducing the FMA dataset

The website. The FMA website [3] is run by WFMU, the longest-running free-form radio station in the United States. The website provides a large catalogue of artists and high-quality mp3 songs. Each song is legally free to download as artists decided to work under open licence such as Creative Commons license. This is directly inspired by the open source movement, and the goal of the FMA site is to offer a legal and technological platform for artists and listeners to harness the full potential of online music sharing [18]. FMA has been funded by the New York State Music Fund, the MacArthur Foundation, and the National Endowment for the Arts.

Crawling the data. The FMA organization provides an API [2] to collect the meta-data associated to each song such as artist name, track title, track genre, and track listens. The collect of the mp3 audio files was done with a web scraper. The collection of FMA was done in April 2016, and the size of the dataset at this time was 86,992 songs. It was recently announced on July 2016 that the website had reached the landmark of 100,000 songs.

The raw FMA dataset. The fully collected dataset on April 2016 contained 86,992 music data. Each music data is composed of an mp3 audio file and some meta-data provided on the FMA website for each song, and crawled by the API. All mp3 audio tracks are coded with sampling rate of 44,100Hz, bit rate 128kb/s, and in stereo. However, not all meta-data is available for each song, and we thus decided to remove the data which has not the following complete meta-information; artist name, track title, track genres, and track listens. This reduced the dataset to 82,113 data. Then, we created a top genre by simply picking the first genre in the original list of track genres. A few examples of data are given in Table 1. Figure 2 also presents a part of the distribution of genre sizes of the 82,113 data, which obeys to a power law distribution as often found in natural data collection. We further cleaned the dataset by removing songs which belong to "non-standard" music genres like 'Noise', 'Garage', 'Sound Collage', 'Singer', 'Audio Collage', 'Glitch', 'Unclassifiable', 'Lo-Fi', 'Spoken', 'Poetry', 'Talk Radio', 'Avant-Garde', 'Experimental', 'Ambient', and 'Field Recording'. We also suppressed the songs with the title 'Untitled'. After this filtering, the remaining size is 37,741. Then, we decided to cut the long tail of the power law distribution by removing songs of genres with less than 300 songs, which gave us 30,109 data. Finally, we kept the songs that have the raw audio and all Echonest features. Echonest features are medium-level and high-level audio features provided by Echonest [9], a music platform for developers and media companies which has a collection of 3.5M artists and 37.5M songs.

Echonest features can be either computed by uploading the songs to the Echonest website, which are then analyzed and send them back to the user, or simply retrieved from the Echonest database if the songs were analyzed before. All Echonest operations are done through an API [1]. Our final filtering consists in keeping songs that have both raw audio file and complete Echonest features. The final size of the dataset is 14,570. Figure 3 presents a part of the distribution of the genre sizes for this final dataset.

Proposed benchmark datasets. For research purposes, we propose 3 datasets:

- (1) **FMA_small:** This dataset contains 1000 songs with 10 genres equally balanced, that is 100 songs per class. For each song, we extracted 30sec precisely in the middle of the song. The idea is to create an alternative to the popular GTZAN dataset with respect to the size parameter. But unlike GTZAN, all 1000 songs have not only raw audio tracks but also meta-data composed of artist name, track title, track genres, track listens, and Echonest features. The GTZAN data may not have for example song title or artist name, and also no copyright permission. For reproducibility research, we also provide a split 75-25 of the dataset for future techniques’ comparisons. Eventually, the genre distribution of this dataset is given in Figure 3.
- (2) **FMA_medium:** This dataset has 14,570 songs of 30sec with 20 genres. Unlike FMA_small, the size distribution of genres is not balanced, but all songs have raw audio files and all meta-data. We also provide a split 75-25 of this dataset. The genre distribution is plotted in Figure 2.
- (3) **FMA_large:** This is the original full dataset collected on April 2016. It has 86,992 data. No post-processing was applied. This is as real-world data as it can get. The genre distribution is given by Figure 1.

3 Conclusion and Perspectives

Availability and reproducible research. For the sake of open science, the proposed 3 benchmarked datasets are made available for free at <https://lts2.epfl.ch/datasets>. In addition to the existing ones, we hope the music analysis community will benefit from this new dataset.

Potential applications. We foresee multiple applications of this new dataset. The most obvious one is for music analysis with deep learning techniques. As raw music tracks are available, deep learning techniques like convolutional neural networks [16] and recurrent neural networks [11] can be directly applied. Besides, standard data analysis tasks such as classification, clustering, regression, text analysis, visualization can also be applied to this new music dataset.

References

- [1] Echonest API. <http://developer.echonest.com/docs>. *Spotify*.
- [2] Free Music Archive API. <http://freemusicarchive.org/api/docs>.
- [3] Free Music Archive. <http://freemusicarchive.org>. *WFMU, Jersey City, NJ 07303, U.S.*
- [4] T. Bertin-Mahieux, D. Ellis, B. Whitman, and P. Lamere. <http://labrosa.ee.columbia.edu/millionsong>. *Columbia University*.
- [5] T. Bertin-Mahieux, D. Ellis, B. Whitman, and P. Lamere. The Million Song Dataset. *International Society for Music Information Retrieval Conference (ISMIR)*, 2011.
- [6] MIREX challenges. <http://www.music-ir.org/mirex/wiki>. *International Music Information Retrieval Systems Evaluation Laboratory (IMIRSEL), University of Illinois at Urbana-Champaign (UIUC)*.
- [7] J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. Fei-Fei. <http://image-net.org>.
- [8] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. *Computer Vision and Pattern Recognition (CVPR)*, 2009.
- [9] The Echonest. <http://the.echonest.com>. *Spotify*.
- [10] International Society for Music Information Retrieval (ISMIR). <http://www.ismir.net>.
- [11] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 1997.
- [12] A. Krizhevsky. <https://www.cs.toronto.edu/~kriz/cifar.html>. *University of Toronto*.
- [13] A. Krizhevsky. Learning Multiple Layers of Features from Tiny Images. *Technical Report, Department of Computer Science, University of Toronto*, 2009.
- [14] E. Law, L. Von Ahn, R. Dannenberg, and M. Crawford. <https://musicmachinery.com/tag/tagatune>.
- [15] E. Law, L. Von Ahn, R. Dannenberg, and M. Crawford. TagATune: A Game for Music and Sound Annotation. *ISMIR*, 2007.
- [16] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-Based Learning Applied to Document Recognition. In *Proceedings of the IEEE*, 86(11), pages 2278–2324, 1998.
- [17] Y. LeCun, C. Cortes, and C. Burges. <http://yann.lecun.com/exdb/mnist>. *Courant Institute, NYU*.
- [18] Ceci Moss. Interview with Jason Sigal of the Free Music Archive, link. *Rhizome.org*.

- [19] B. Sturm. An Analysis of the GTZAN Music Genre Dataset. *ACM workshop on Music information retrieval with user-centered and multimodal strategies*, pages 7–12, 2012.
- [20] G. Tzanetakis and P. Cook. http://marsyasweb.appspot.com/download/data_sets.
- [21] G. Tzanetakis and P. Cook. Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

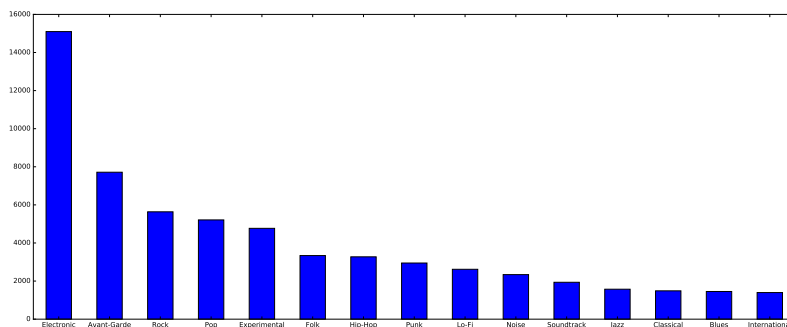


Figure 1: Histograms of the top 15 music genres for the large FMA dataset of 82,113 songs.

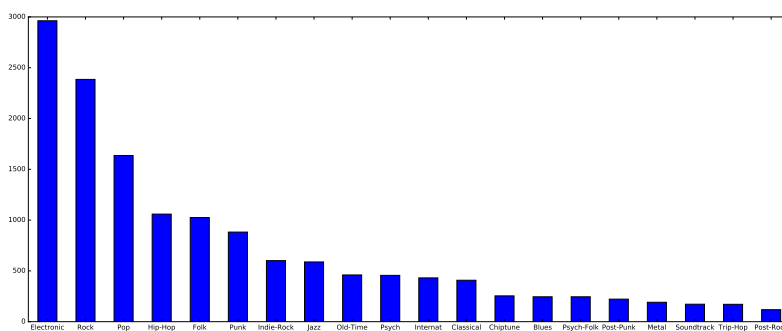


Figure 2: Histograms of the 20 music genres for the medium FMA dataset of 14,570 songs.

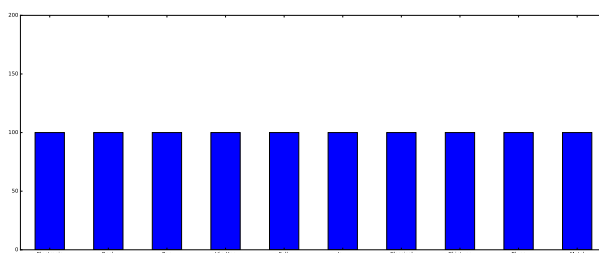


Figure 3: Histograms of the 10 music genres for the small FMA dataset of 1000 songs.

fma_id	artist	play_count	title	genres	top_genre
10	Kurt Vile	42936	Freeway	[Pop]	Pop
134	AWOL	880	Street Music	[Hip-Hop]	Hip-Hop
141	Alec K. Redfearn & the Eyesores	590	Ohio	[Folk]	Folk
142	Alec K. Redfearn & the Eyesores	670	Punjabi Watery Grave	[Folk]	Folk
144	Amoebic Ensemble	901	Wire Up	[Jazz]	Jazz
145	Amoebic Ensemble	682	Amoebiasis	[Jazz]	Jazz
188	Ed Askew	973	Piano 1	[Folk]	Folk
206	Ed Askew	73	Bella Crane	[Folk]	Folk
236	Banana Clipz	6695	Push Am (Left, Right)	[Electronic, African]	Electronic
237	Banana Clipz	1056	Garbage and Fire	[Jazz]	Jazz
238	Barnacled	961	France Attacks	[Jazz]	Jazz
461	Cantonement Jazz Band	3933	Besmer	[Blues, Jazz: Vocal]	Blues
462	Cantonement Jazz Band	3256	Has Been Blues	[Blues, Jazz: Vocal]	Blues
463	Cantonement Jazz Band	3238	I'll Be Blue	[Blues, Jazz: Vocal]	Blues
824	Here Comes A Big Black Cloud!!	391	Black Mold	[Rock, Loud-Rock, Psych-Rock, Indie-Rock]	Rock
825	Here Comes A Big Black Cloud!!	1847	Death March	[Rock, Loud-Rock, Psych-Rock, Indie-Rock]	Rock
837	Heroin UK	146	DopeSick	[Rock]	Rock
889	Illusion of Safety	163	Wasteland	[Electronic]	Electronic
896	Impediments	1542	2012	[Punk, Power-Pop]	Punk
992	Jason Willett	560	Beautiful Song w/ kick drum solo	[Rock]	Rock

Table 1: A few song data in the FMA dataset.