Music Genre Recognition summary

G. Fountoukidou, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 1.0.1

Artificial Intelligence & Information Analysis Lab

Music Genre Recognition

- **Introduction**
- Audio Feature Extraction
- Music Spectrograms
- Sound Texture Selection
- Machine Learning Algorithms
- Gaussian Processes
- Support Vector Machine
- Music Recognition using Deep Neural Networks

Introduction to Music Genre Recognition

- Music information retrieval (MIR) is an interdisciplinary field that combines content from music theory, signal processing, and machine learning to analyze musical material.
- MIR uses computer algorithms to detect and intelligently manage musical material.
- Music Genre Recognition is one of the most important subfields in MIR (MIR).

Introduction to Music Genre Recognition

- Automatic music is a fascinating subject in MIR since it allows systems to communicate with one other, discovering media collections, organizing musical databases and perform content-based music recommendation.
- There are two primary processes in music genre categorization: **feature extraction** and c**lassification**. The first captures audio signal data, while the second categorizes the music into different genres based on the extracted attributes.

Music Genre Recognition

- Introduction
- **Audio Feature Extraction**
- Music Spectrograms
- Sound Texture Selection
- Machine Learning Algorithms
- Gaussian Processes
- Support Vector Machine
- Music Recognition using Deep Neural Networks

Feature Extraction

Feature extraction obtains the audio signal information from the music. Feature extraction from an audio signal used chord recognition methods.

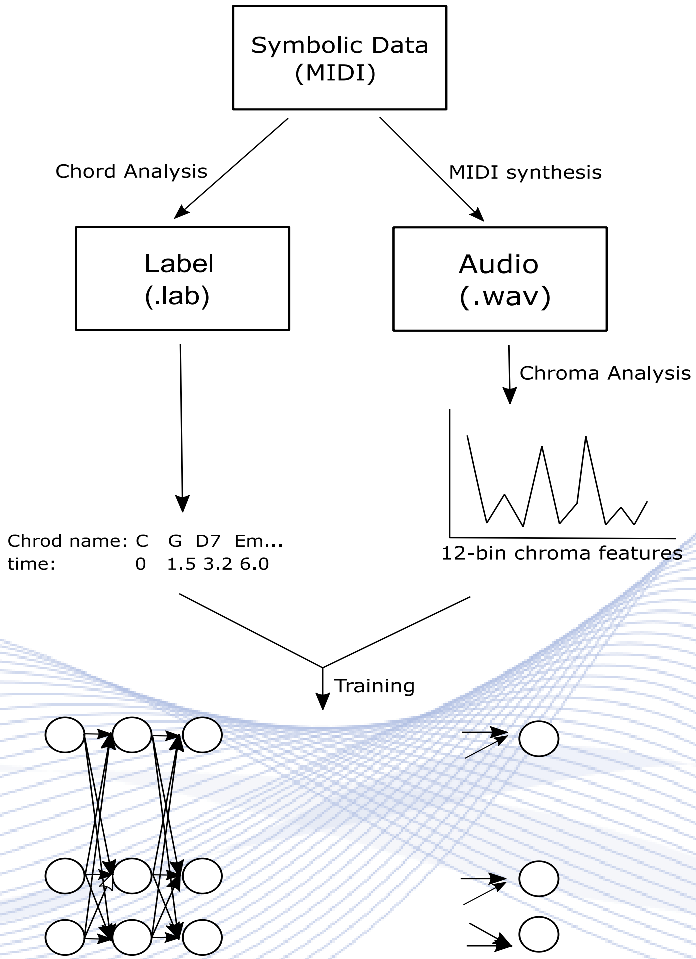
- From a certain music-related characteristic, the chord identification job generates a chord label. A chromagram is used as an input to these systems, with a chord label as an output for each chromagram frame.
- The most popular models used in chord recognition are the hidden Markov models.

Feature Extraction – Markov Models

The frequency domain is first segregated from the input signal. It's then transferred to the Pitch Class Profile domain, where vectors are utilized as features to train a hidden Markov model with one state for each chord. The Expectation Maximization technique is then used to construct chord models using these features. Finally, the Viterbi algorithm is used to complete the chord recognition.

- Duration-explicit hidden Markov models is one of the chord-recognition methods. The transition matrix's length constraints are broken up by the models. The method then creates different models for duration distributions that show time signatures in order to boost the duration constraint in each model.

Feature Extraction – Markov Models



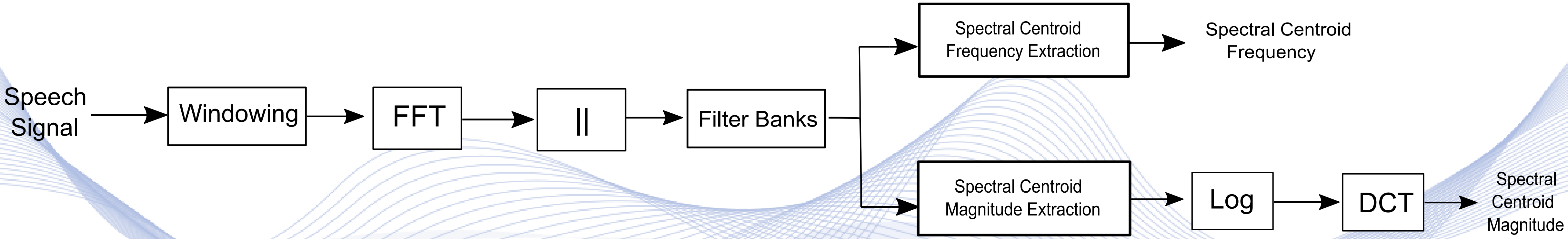
Using a hidden Markov filter to automatically recognize chords from audio.

Feature Extraction – Frequency Domain Features

The Fourier Transform can be used to translate an audio signal into the frequency domain. After that, these characteristics are retrieved:

- **Mel-Frequency Cepstral Coefficients** (MFCC): MFCCs are considered very helpful features for tasks like the recognition of speech.
- **Chroma Features:** The entire energy of the signal per each of the 12 pitch classes .(C, C#, D, D#, E ,F, F#, G, G#, A, A#, B) is represented as a vector. The mean and standard deviation was calculated using the sum of the chroma vectors.

Feature Extraction – Frequency Domain Features



Process to create Spectral Centroid features

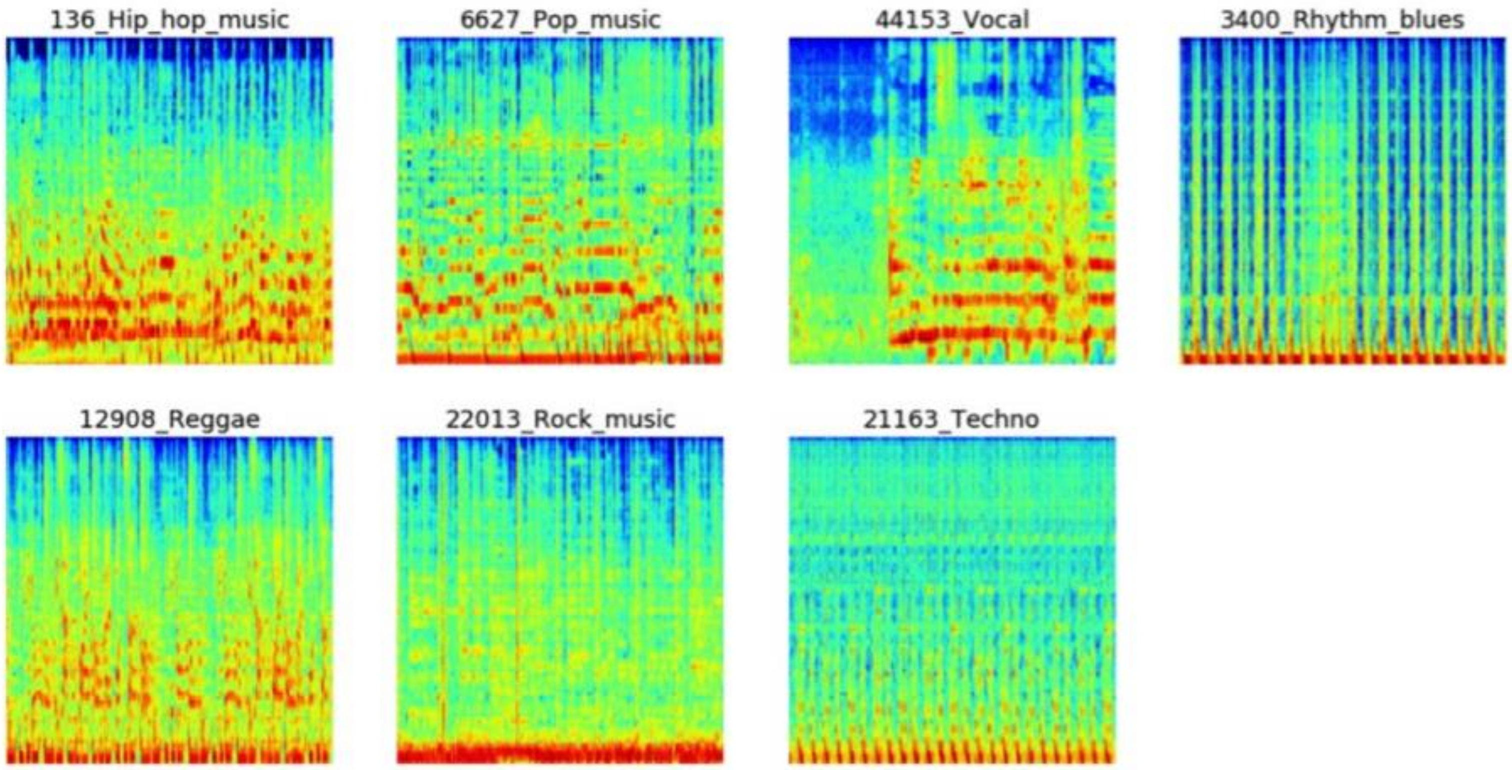
Music Genre Recognition

Spectrograms

The time axis is represented by the x-axis, while the frequency axis is represented by the y-axis in a **spectrogram**, which is a two-dimensional representation of a signal. To quantify the amplitude of a specific frequency in a certain time interval, a colormap is employed.

- The representation of audio in the time domain for neural network input is not particularly precise due to the high sampling rate of audio data.

Spectrograms



Spectrograms for a single audio source from each musical genre [3].

Spectrograms

Since texture is the main visual content found in the spectrogram, different types of texture representations have been used to describe the content of these images, such as Gray-Level Co-Occurrence Matrix (GLCM), Local Binary Patterns (LBP) Local Phase Quantization Gabor Filters and Weber Local Descriptor (WLD).

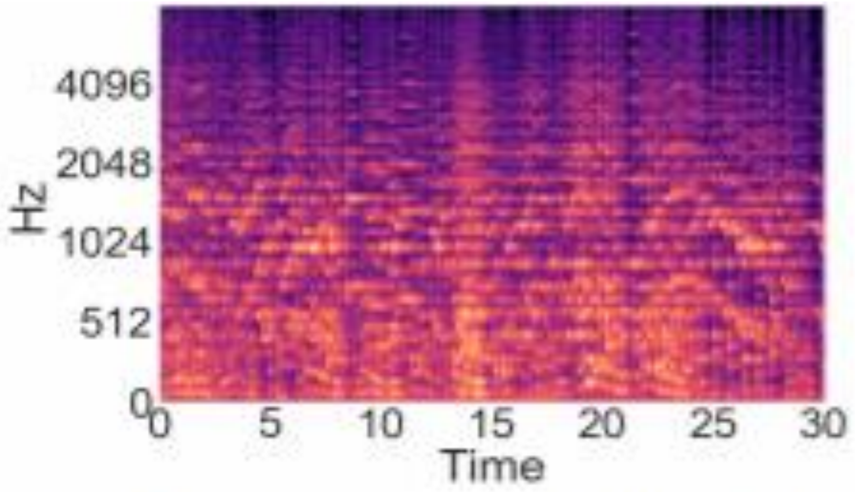
Mel Spectrogram: Mel-frequency cepstrum (MFC) representations are widely used in automatic speaker and speech recognition. The mel spectrogram produces a time frequency representation of a sound imitating the biological auditory systems of human beings. We compute the magnitude spectrum from the time series musical data and then map it on to the mel scale.

Spectrograms

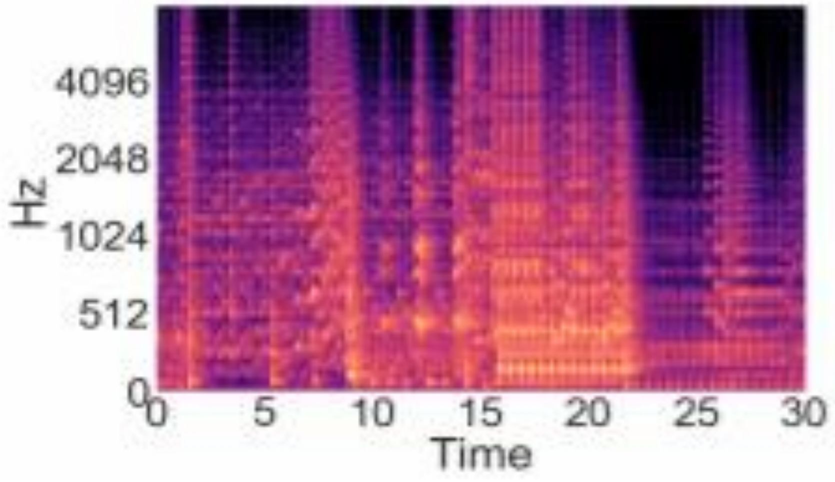
Mel Frequency Cepstral Coefficient (MFCC) is employed for this technique of characteristic extraction because of its dependability in producing good features, and the notion of this feature is combined with the **Spectral Centroid Feature (SCF).**

- They're made by applying Fourier transforms to the signal, then logarithmic power values, and lastly cosine transformations.

Spectrograms



(a) *Mel Spectrogram Classical*



(e) *Mel Spectrogram Jazz*

Mel spectrogram [1].

Music Genre Recognition

- Introduction
- Audio Feature Extraction
- Music Spectrograms
- **Sound Texture Selection**
- Machine Learning Algorithms
- Gaussian Processes
- Support Vector Machine
- Music Recognition using Deep Neural Networks

Texture Extraction

The issue of assigning genre-related labels to digital music files is known as Classification of Music Genre.

Music tracks are sometimes represented as a collection of sound textures influenced by timbre.

The total number of sound textures per track in shallow-learning systems is typically too high, necessitating texture down sampling to make training tractable. In the context of the bag of frames track descriptions, texture selection helps to classify genre.

Texture Extraction – K-Means

K-Means Texture Selection

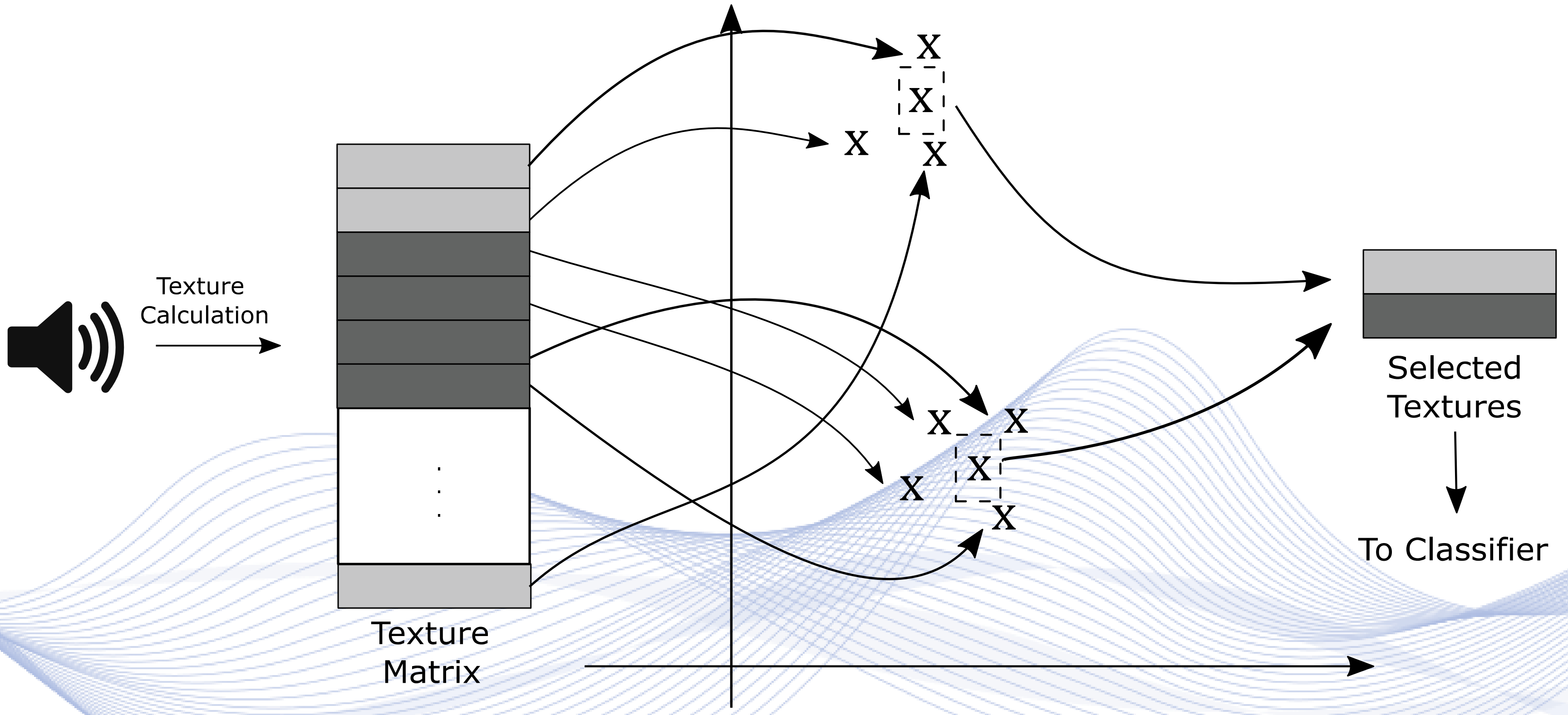
The clustering algorithm K-Means is very common. It works by iteratively computing centroids, which are locations that form different patterns in a dataset. Those points will be accustomed to search the data for further points that match the pattern.

Every audio texture is expressed by a vector in space $\mathbb{R}^n$, where n is the amount of features in the space.

- A texture matrix $T \in \mathbb{R}^{m \times n}$ can be used to characterize a music track, where m is the total amount of textures in the song and n is the amount of characteristics.

Artificial Intelligence & Information Analysis Lab

Texture Extraction – K-Means



K-Means Texture Selection.

Music Genre Recognition

- Introduction
- Audio Feature Extraction
- Music Spectrograms
- Sound Texture Selection
- **Machine Learning Algorithms**
- Gaussian Processes
- Support Vector Machine
- Music Recognition using Deep Neural Networks

Machine Learning Algorithms

- The music genre may be classified using a variety of classifiers. Traditionally, a single classifier has been used to categorize the feature vectors that are created after the features have been extracted. For audio classification, Naive Bayesian, VFI, PART, J48 NNge, and JRip are often used classifiers.
- The Naive Bayes classifier analyzes the training data statistically, generates maximum likelihood estimators, and uses conditional probabilities on observed attribute values as decision criterion to optimize conditional probabilities.

Music Genre Recognition

- Introduction
- Audio Feature Extraction
- Music Spectrograms
- Sound Texture Selection
- Machine Learning Algorithms
- **Gaussian Processes**
- Support Vector Machine
- Music Recognition using Deep Neural Networks

Gaussian Processes

CT Gaussian Processes (GPs) are Bayesian nonparametric models that are gaining in popularity due to their higher-level ability to apprehend strongly nonlinear data interactions in a variety of tasks, including dimensionality reduction, time series analysis, novelty identification, and traditional regression and classification.

Gaussian Processes

GP models, like SVM, are based on kernel functions and Gram matrices. They, on the other hand, generate completely probabilistic outcomes with a specified level of prediction uncertainty. Furthermore, there are algorithms for GP hyperparameter learning, which the SVM architecture does not provide. Experiments specifically demonstrated that the GP outperformed the SVM in both **music genre recognition** and music emotion prediction tasks.

Music Genre Recognition

- Introduction
- Audio Feature Extraction
- Music Spectrograms
- Sound Texture Selection
- Machine Learning Algorithms
- Gaussian Processes
- **Support Vector Machine**
- Music Recognition using Deep Neural Networks

Support Vector Machine

There has been a lot of study and experimentation on the classification of music genre, and the most common method is feature extraction accompanied by supervised machine learning.

- In the world of music information retrieval (MIR), these are two of the most important activities. The **support vector machine** (SVM) has been the most common model in MIR systems so far.

Support Vector Machine

When using Support Vector Machine in music genre classification, feature extraction process is done toward music files in the data set. The features are computed for every short-time frame of audio signal for time-domain features and are calculated according to short time fourier transform (STFT) for frequency-domain features.

- The process of classification in the experiments will use SVM classifier algorithm with kernel method.
- The results demonstrate that the polynomial kernel is the best kernel for automated musical genre categorization. The combination of musical surface, MFFC, tonality, and LPC is the greatest audio feature combination.

Music Genre Recognition

- Introduction
- Feature Extraction
- Spectrograms
- Texture Selection
- Support Vector Machine
- **Music Recognition using Deep Neural Networks**
- Machine Learning Algorithms
- Gaussian Processes

Music Genre Recognition using Deep Neural Networks

Without the use of hand-crafted features, we can perform music genre classification using **deep learning**. For image recognition, convolutional neural networks (CNNs) have been commonly used.

A CNN is equipped to guess the image type using the 3-channel (RGB) matrix deception of an image. A spectrogram may be used to reflect a **sound wave**, which can then be treated as an **image**. The CNN's job is to forecast the genre label using the spectrogram.

Music Genre Recognition using Deep Neural Networks

Labeled input audio signals

↓

Wav file Preprocessing

↓

Feature Extraction/spectrogram

↓

Dimensionality Reduction/Manifold learning

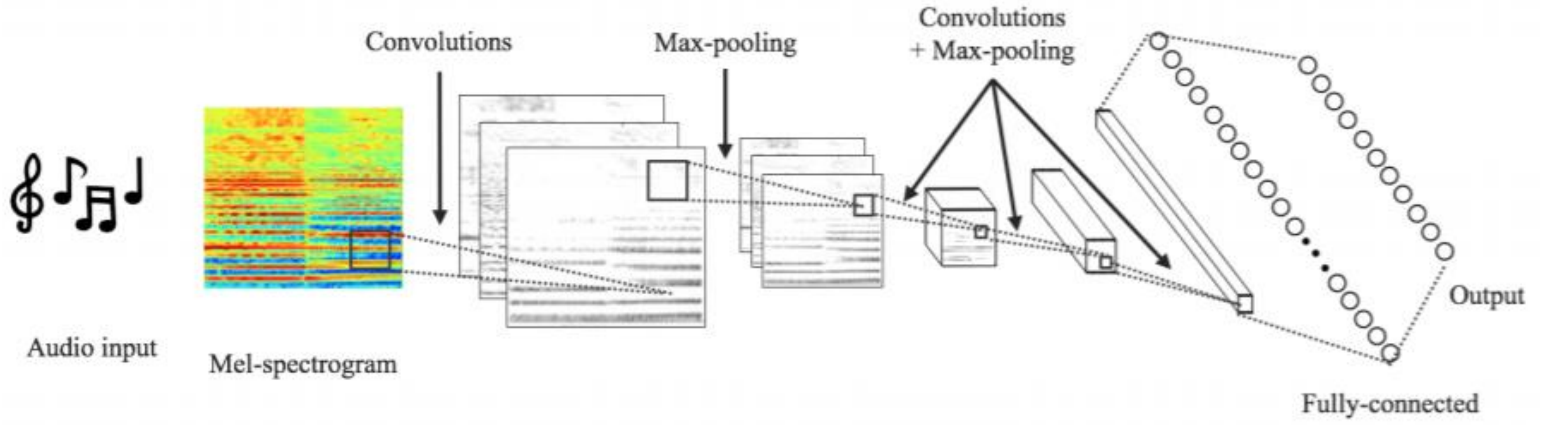
↓

Model(SVM, Deep nets, ...)

↓

Music Genre Classification

Music Genre Recognition using Deep Neural Networks



Music Genre Recognition [7].

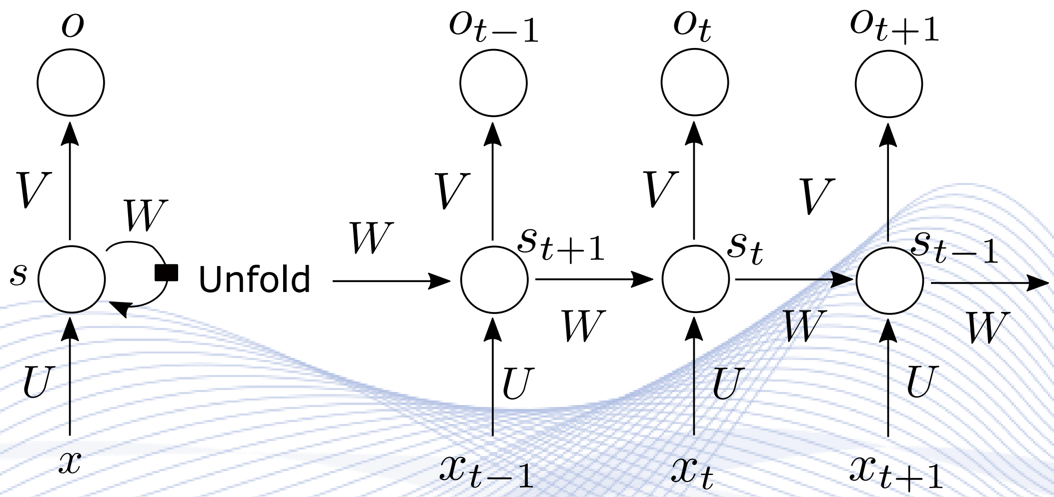
Music Genre Recognition using Deep Neural Networks

RNNs are a type of artificial neural network that identifies patterns in data streams**.** This algorithms have a temporal dimension since they consider time and sequence. Also, images that can be broken down into a succession of patches. and viewed as a sequence, can be used with RNNs.

- Gjerdingen and Perrott (2008) conducted experiments with human subjects, demonstrating that humans could confidently identify the type of music by listening to only a 3-second snippet of music.

Music Genre Recognition using Deep Neural Networks

Taking this into account, each clip in the dataset is divided into three-second chunks.



A recurrent neural network unit is depicted as a block diagram.

Bibliography

[1] Deepanway Ghosal, Maheshkumar H. Kolekar, Music Genre Recognition using Deep Neural Networks and Transfer Learning, Indian Institute of Technology Patna, India, 2018.

[2] Aziz Nasridinov, Young-Ho Park, A Study on Music Genre Recognition and Classification Techniques, School of Computer Engineering, Dongguk University at Gyeongju, Department of Multimedia Science, Sookmyung Women's University, 2014.

[3] Hareesh Bahuleyan University of Waterloo, ON, Music Genre Classification using Machine Learning Techniques,Canada, 2018.

[4] Yandre M.G. Costaa,∗, Luiz S. Oliveira b, Carlos N. Silla Jr. c a PCC/DIN, State University of Maringá (UEM), Maringá, PR, Brazil b PPGInf, Federal University of Paraná, Curitiba, PR, Brazil c PPGIa, Pontifical Catholic University of Paraná, Curitiba, PR, Brazil, 2016.

Bibliography

[5] Juliano H. Foleiss, Tiago F. Tavares,Department of Computing Federal, University of Technology,Paraná, School of Electrical and Computer Engineering, University of Campinas, Campinas, SP – Brazil, 2019.

[6] A.B. Mutiara , R. Refianti , and N.R.A. Mukarromah Faculty of Computer Science and Information Technology, Gunadarma University, Musical Genre Classification Using Support Vector Machines and Audio Features, Indonesia, 2016.

[7] Yoonchang Han, Jaehun Kim, and Kyogu Lee, Senior Member, IEEE, Deep convolutional neural networks for predominant instrument recognition in polyphonic music, 2016.

[8] Roberto Basili, University of Rome Tor Vergata, Alfredo Serafini, Armando Stellato University of Rome, Classification of musical Genre: a machine learning approach, 2004.

Bibliography

[9] Konstantin Markov The University of Aizu, Tomoko Matsui The Institute of Statistical Mathematics, Music Genre and Emotion Recognition Using Gaussian Processes, 2014.

[10] Noraziahtulhidayu Kamarudin, S.A.R Al-Haddad, Shaiful Jahari Hashim, Mohammad Ali Nematollahi, Abd Rauf Bin Hassan, Department Computer and Communication System Engineering, University Putra Malaysia. Selangor, Malaysia. Feature Extraction Using Spectral Centroid and Mel Frequency Cepstral Coefficient for Quranic Accent Automatic Identification, 2014.

[11] Arjun Raj Rajanna, Kamelia Aryafar, Ali Shokoufandeh† and Raymond Ptucha, Electrical Engineering Department, Deep Neural Networks: A Case Study for Music Genre Classification, 2015

Bibliography

[12] VISHNUPRIYA S., BIGDATA ANALYTICS SRM UNIVERSITY, Chennai,india K.MEENAKSHI INFORMATION TECHNOLOGY SRM UNIVERSITY, Automatic Music Genre Classification using Convolution Neural Network, 2018

[13] Dipjyoti Bisharad, Rabul Hussain Laskar Department of Electronics and Communication Engineering, National Institute of Technology, Silchar, Silchar, India, Music genre recognition using convolutional recurrent neural network architecture, 2019

Bibliography

[OPP2013] A. Oppenheim, A. Willsky, Signals and Systems, Pearson New International, 2013.

[MIT1997] S. K. Mitra, Digital Signal Processing, McGraw-Hill, 1997.

[OPP1999] A.V. Oppenheim, Discrete-time signal processing, Pearson Education India, 1999.

[HAY2007] S. Haykin, B. Van Veen, Signals and systems, John Wiley, 2007.

[LAT2005] B. P. Lathi, Linear Systems and Signals, Oxford University Press, 2005.

[HWE2013] H. Hwei. Schaum's Outline of Signals and Systems, McGraw-Hill, 2013.

[MCC2003] J. McClellan, R. W. Schafer, and M. A. Yoder, Signal Processing, Pearson Education Prentice Hall, 2003.

Bibliography

[PHI2008] C. L. Phillips, J. M. Parr, and E. A. Riskin, Signals, Systems, and Transforms, Pearson Education, 2008.

[PRO2007] J.G. Proakis, D.G. Manolakis, Digital signal processing. PHI Publication, 2007.

[DUT2009] T. Dutoit and F. Marques, Applied Signal Processing. A MATLAB-Based Proof of Concept. New York, N.Y.: Springer, 2009

Bibliography

[PIT2000] I. Pitas, “Digital Image Processing Algorithms and Applications”, J. Wiley, 2000.

[PIT2021] I. Pitas, “Computer vision”, Createspace/Amazon, in press.

[PIT2017] I. Pitas, “Digital video processing and analysis” , China Machine Press, 2017 (in Chinese).

[PIT2013] I. Pitas, “Digital Video and Television” , Createspace/Amazon, 2013.

[NIK2000] N. Nikolaidis and I. Pitas, “3D Image Processing Algorithms”, J. Wiley, 2000.

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

Contact: Prof. I. Pitas
pitas@csd.auth.gr