

Tutorial

ismr
2018

19th International Society for Music Information Retrieval Conference
September 23–27, Paris, France

Deep Learning for Music Information Retrieval

September 23, 2018

Tutorial on Github

https://github.com/slychief/ismir2018_tutorial

or

<http://tiny.cc/dlismir18>

Clone or download ▾

**Download the data sets linked in the README!
(prepared subset of MagnaTagATune dataset)**

Deep Learning for Music Information Retrieval

Presenters:



Thomas Lidy

Head of Machine Learning, Musimap

tom@musimap.com
www.musimap.com



Alexander Schindler

Research Scientist, AIT & TU Wien

Alexander.Schindler@ait.ac.at
<http://ifs.tuwien.ac.at/~schindler>



Sebastian Böck

Post Doc, OFAI & TU Wien

sebastian.boeck@tuwien.ac.at
<http://ifs.tuwien.ac.at/~schindler>



Tutorial Outline

I. Convolutional Neural Networks:

- Instrumental vs. Vocal
- Genre Recognition
- Mood Recognition

II. Similarity Retrieval and Representation Learning:

- Similarity Retrieval
- Siamese Networks
- Learning Music Similarity from Tags

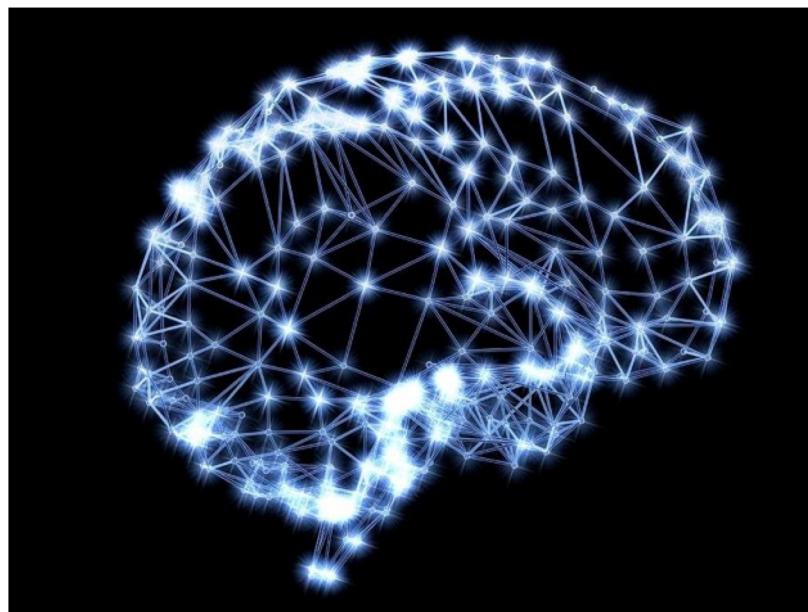
III. Onset and Beat Detection with RNNs:

- Recurrent Neural Networks
- Onset and Beat Detection

Deep Learning

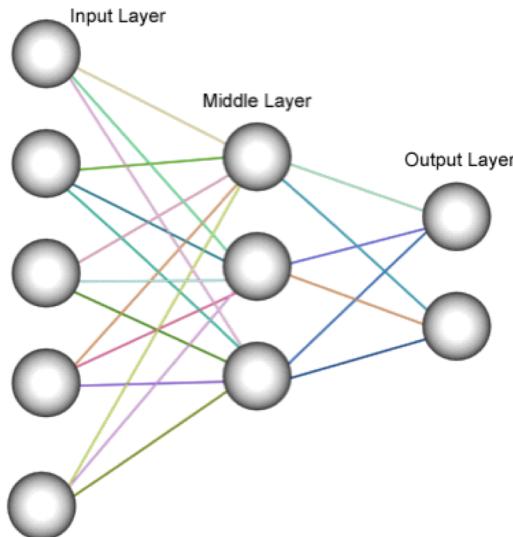
= (Deep) Artificial Neural Networks (ANNs)

Neural Networks are loosely inspired by biological neurons that are interconnected and communicate with each other



Neural Networks

In reality, a neural network is a **mathematical function**:



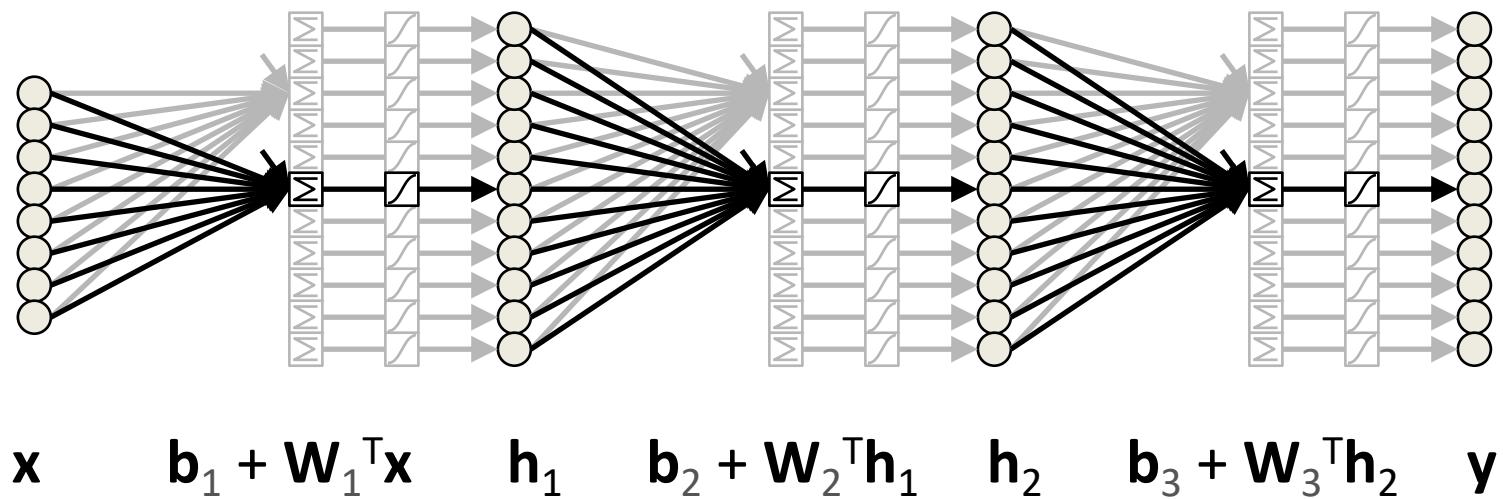
- in which the “**neurons**” are **sets of adaptive weights**, i.e. numerical parameters that are **tuned by a learning algorithm**
- which has the capability of approximating non-linear functions of their inputs

What are Artificial Neural Networks?

Mathematical expressions, such as:

$$y = \sigma(\mathbf{b}_3 + \mathbf{W}_3^T \sigma(\mathbf{b}_2 + \mathbf{W}_2^T \sigma(\mathbf{b}_1 + \mathbf{W}_1^T \mathbf{x})))$$

expression can be visualized as a graph:



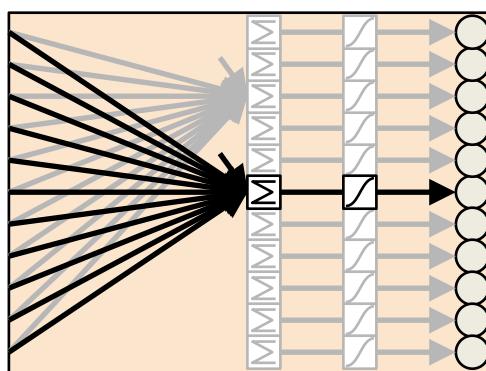
What are Artificial Neural Networks?

Mathematical expressions, such as:

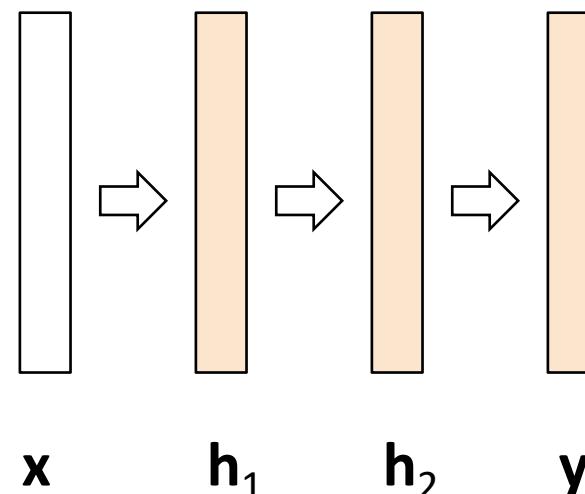
$$f_{W,b}(x) = \sigma(b + W^T x)$$

$$y = (f_{W_3, b_3} \circ f_{W_2, b_2} \circ f_{W_1, b_1})(x)$$

expression can be visualized as a graph – as connected layers:



“dense layer”

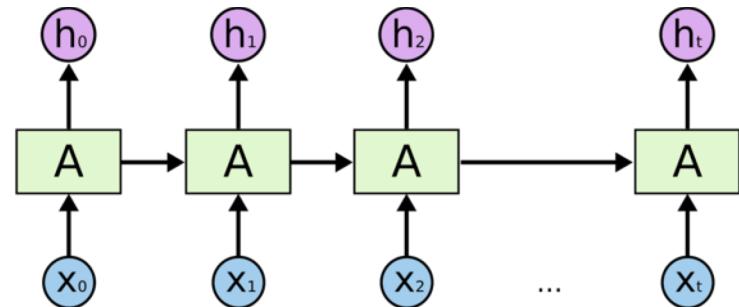
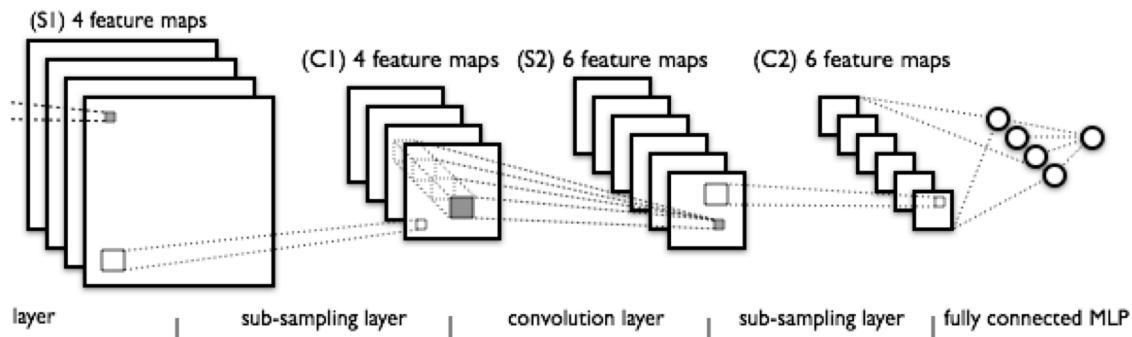


composed of simpler **functions**, commonly termed “**layers**”

Neural Network Architectures

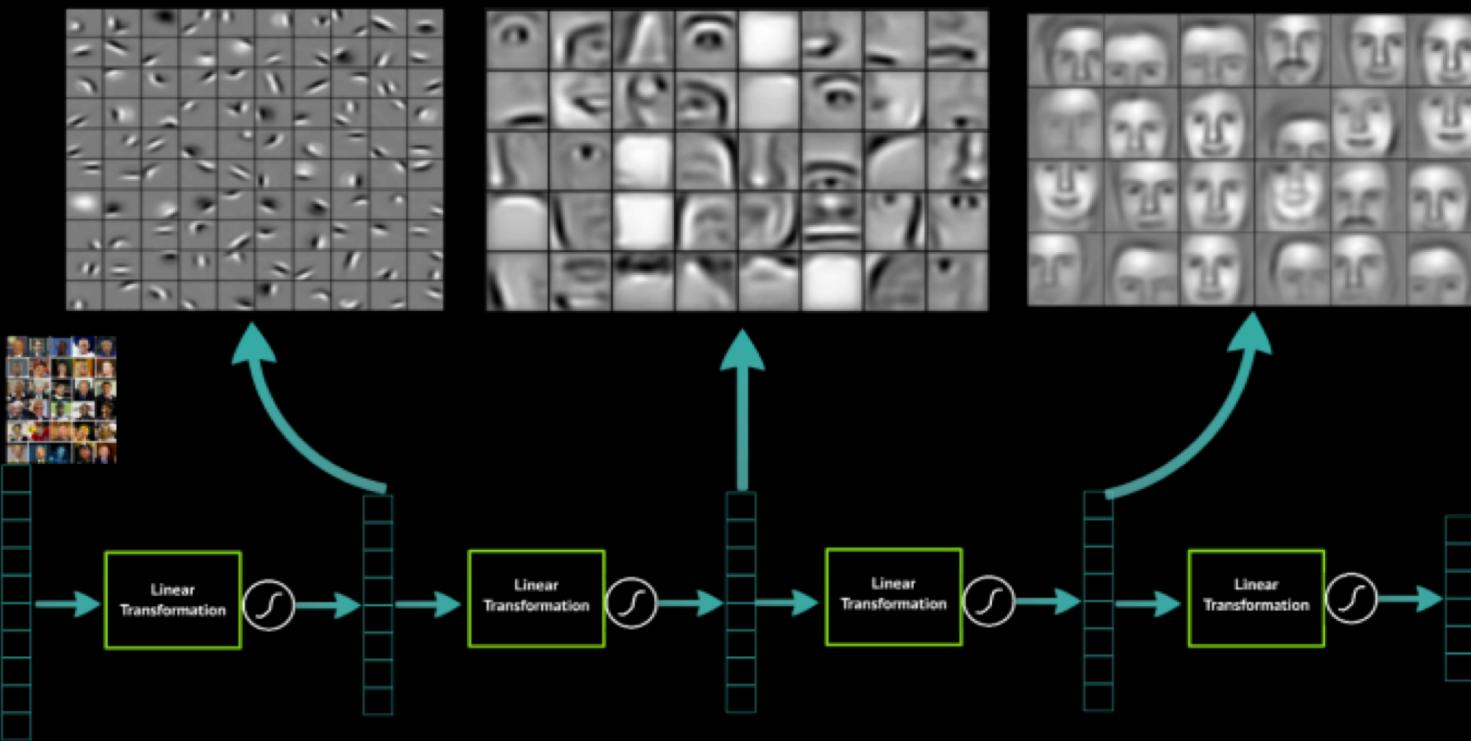
Two main Neural Network types in use today:

- **Convolutional Neural Networks** (ConvNets or CNN)
- **Recurrent Neural Networks** (RNN, LSTM, GRU)



Convolutional Neural Networks (CNN)

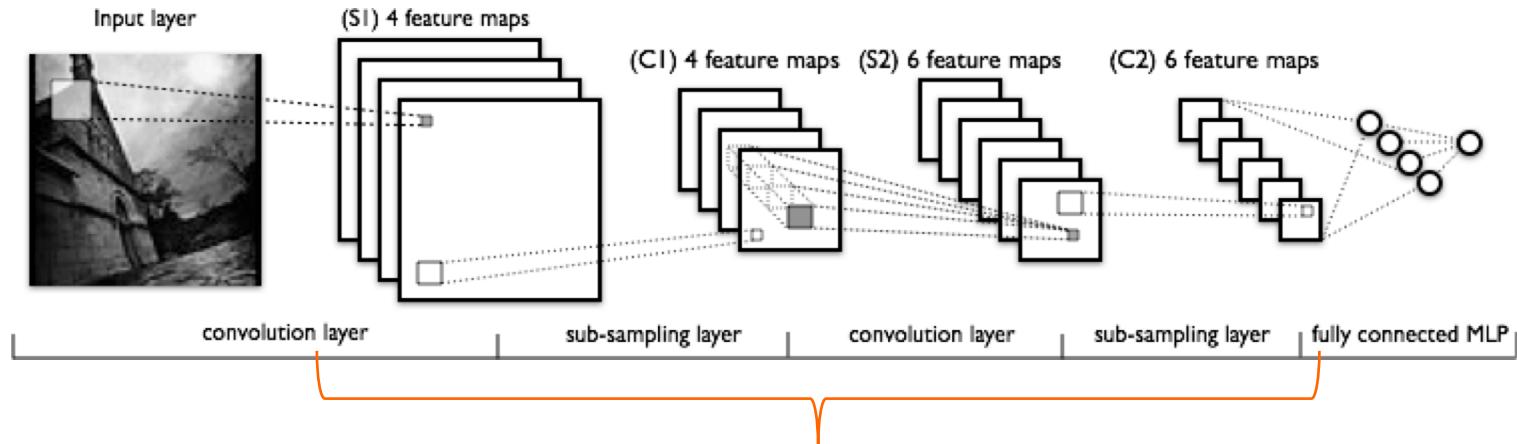
Deep Learning learns layers of features



Note: the images are conceptual here and do not represent the actual output of the neurons.

Img: <https://blog.datarobot.com/a-primer-on-deep-learning>

Convolutional Neural Network (CNN)

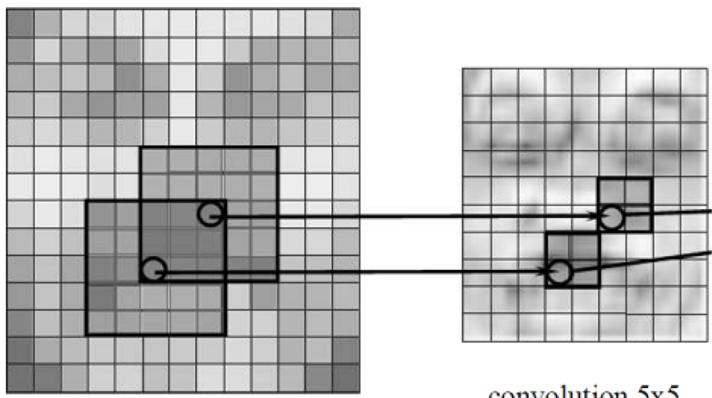


Combines three types of layers:

- **Convolutional layer:** performs 2D convolution of 2D input with multiple learned 2D kernels – **learns shapes**
- **Subsampling layer:** replaces 2D patches by their maximum (“max-pooling”) or average (“average-pooling”) – **reduces resolution**
- **Fully-connected layer:** computes weighted sums of its input with multiple sets of learned coefficients – **maps to output**

What is a Convolution?

- Apply local filter kernels and slide them over the input
- Instead of using predefined kernels, these kernels are the neurons that are learned!



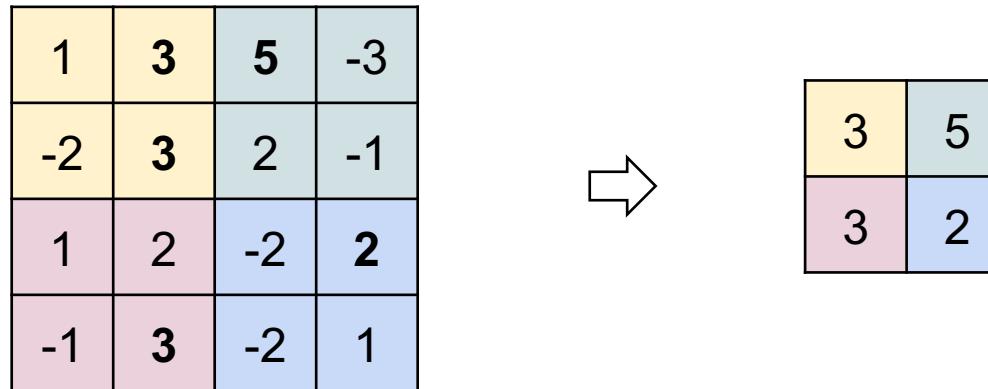
Operation	Kernel	Image result
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	

Images: <http://sanghyukchun.github.io/75/>
[https://en.wikipedia.org/wiki/Kernel_\(image_processing\)](https://en.wikipedia.org/wiki/Kernel_(image_processing))

What is Pooling?

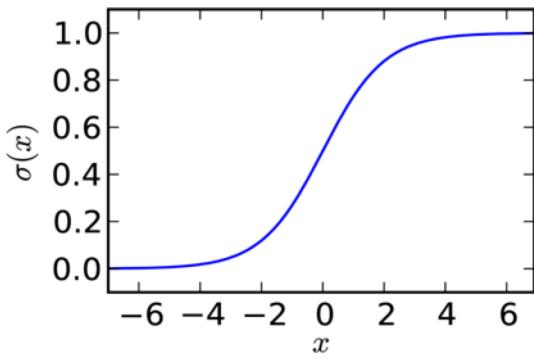
Second very important aspect of a CNN:
(also called subsampling or downsampling)

A **pooling layer** reduces the size of feature maps (i.e. output of a CNN layer and thus the input to the next layer)

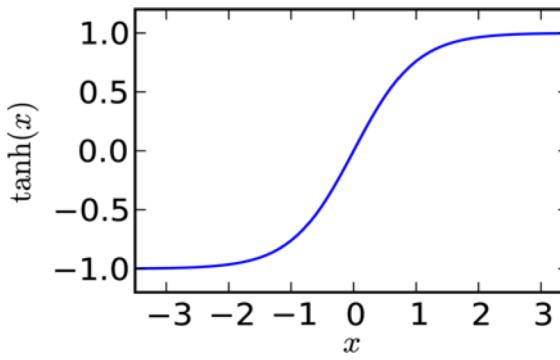


Max pooling: take the max. activation across small regions
(e.g. 2x2, as in the example above)
it can also be considered as an aggregation step

Activation Functions: Linear Rectifier (ReLU)



Sigmoid

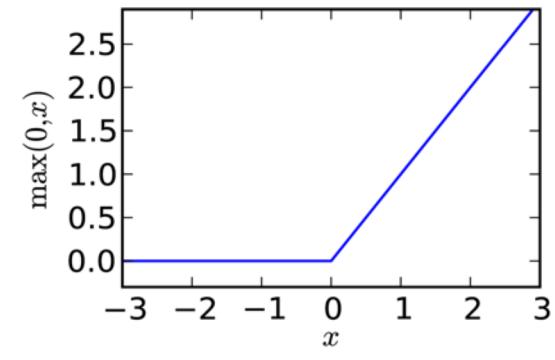


TanH

saturates for large inputs
(small slope, weak gradient!)

saturates for large inputs
(small slope, weak gradient!)

has nonzero mean
(slows learning)



ReLU

has nonzero mean
(slows learning)
has zero gradient for
negative input

Variants of ReLU: Leaky Rectifier (LReLU),
Parametric Rectifier (PReLU), ...

Benefits:

- no saturation
- low computational costs



Coding Tutorial:

Jupyter Notebook

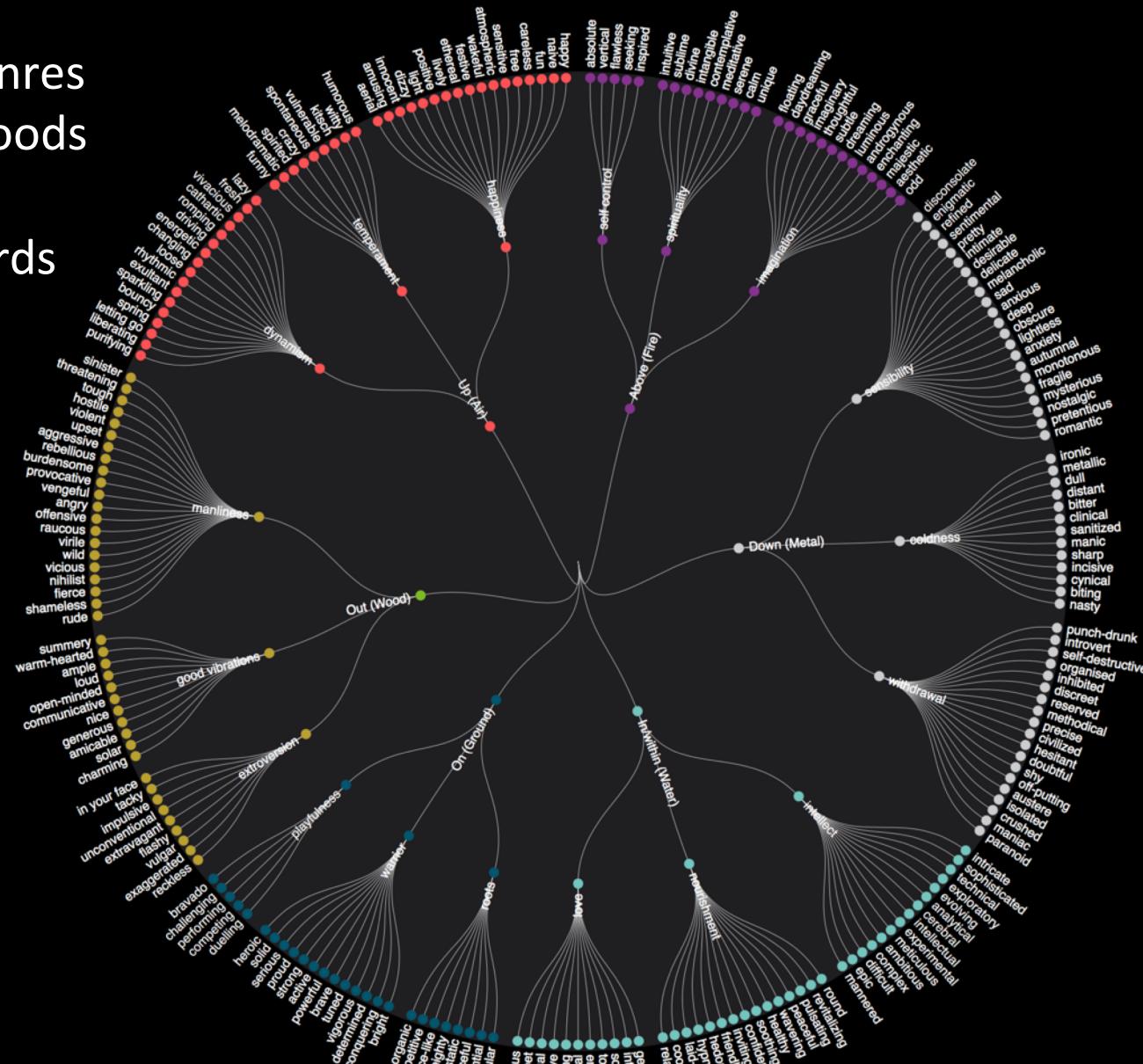
Part_1_Convolutional_Neural_Networks.ipynb



MOODS, GENRES, SUB-STYLES and more ...



400 genres
256 moods
11,000
keywords



 musimap
35M+
tracks
analyzed

Industry Meetup: Thursday 14:00 – 18:00

Come and find out what person you are given your taste in music :-)

PROFILING
BY MUSIMAP

musimap

WHICH ONE ARE YOU ?

To know it, tell us about your favourite music below

Enter below 12 of your favorite songs!
Use the featured demos to see what happens next...

DEMO 1

DEMO 2

DEMO 3

Use your Spotify to easily pick your favorite playlist!
And see what happens next...



Organizers and Hosts: Vienna Deep Learning Meetup

1400
Members

Vienna Deep Learning Meetup

Startseite Mitglieder Sponsoren Fotos Seiten Diskussionen Mehr Gruppenverwaltung Mein Profil



Foto bearbeiten

Wien, Österreich
Gegründet 17. Dez 2015

Über uns...

Freunde einladen

Deep Learners 799

Gruppenreview 1

Anstehende
Meetups 2

Vergangene
Meetups 13

Unser Kalender

13th Deep Learning Meetup in Vienna: Google Tensorflow

Bearbeiten Absagen Feature Kopieren Ticket Export
Informiere deine Freunde Teilen

Dienstag, 24. Oktober 2017
18:00 bis 22:00

Marx Palast
Maria-Jacobi-Gasse 2, Vienna (Karte bearbeiten)

<http://www.marxrestauration.at/Anfahrt>

After our exceptional AI Summit Vienna in September, we planned to continue with our regular monthly Vienna Deep Learning Meetup series. But not quite...

We are proud to have:

Yufeng Guo
Developer Advocate for Machine Learning at Google Cloud, New York

Dein RSVP: Ja

Ändern

Freunde einladen

Tools

250 nehmen teil

Antwort bis zum:
23. Okt um 22:00



Tom Lidy
Organisator,
Event-Koordinator

Music & Machine
Learning since
2004. Researcher
on Deep Learning
in Music and Head
of Machine... [mehr](#)

Dein Intro
bearbeiten

<https://www.meetup.com/Vienna-Deep-Learning-Meetup>

~200
monthly
participants

Slides from 20 past meetups + further resources:
<https://github.com/vdlm/meetups>