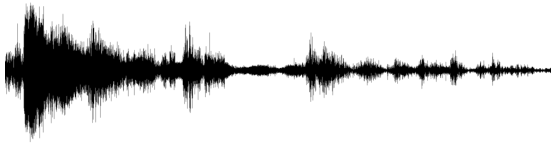
Automatic Sound Recognition in 2020

A Small Introduction

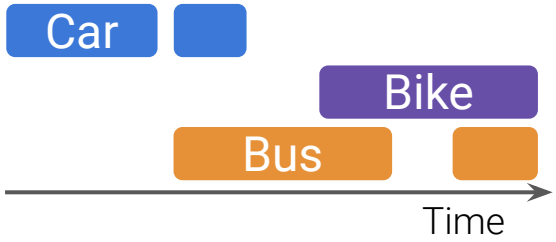
Franz Anders

ASR: Automatic Speech Recognition

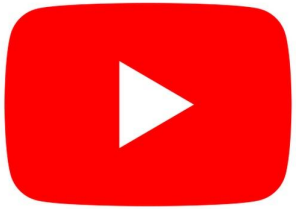


“Hello AI Monday”

ASR: Automatic Sound Recognition



Applications



Video-Tagging

Interpret Surrounding

Music Genre Recognition

Roadmap for this Presentation

Goal: Give you a starting-guide

1. What are the tasks?
2. How does it work?
3. What are current challenges?
4. Where to start?

1. What are the tasks?

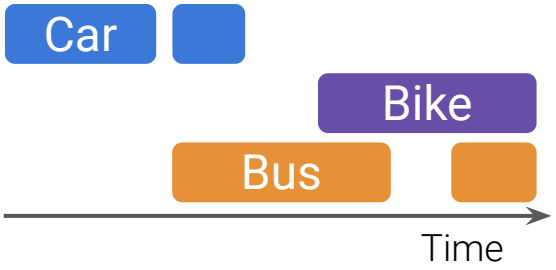
Classification / Tagging

Recognizer

- Street
- Cafe
- Party

Detection

Recognizer



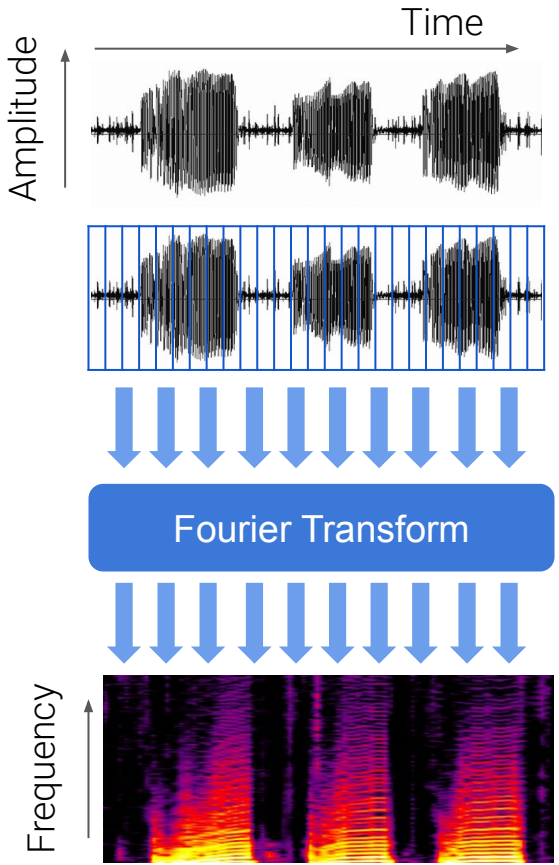
2. How does it work: Audio Representation

Timedomain Signal

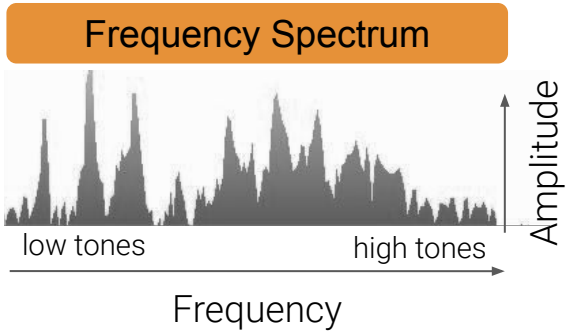
Chop it into windows

Transform each window into frequency domain

Spectrogram!

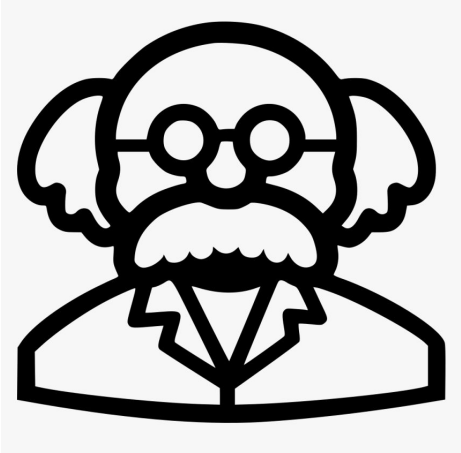


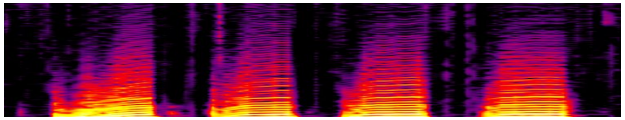
FT



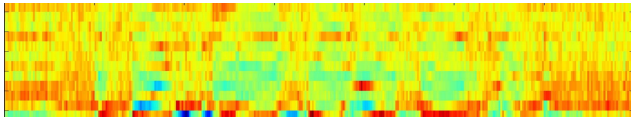
1. How does it work: Approach pre 2016

Pre 2015:
MFCCs + GMMs + HMMs

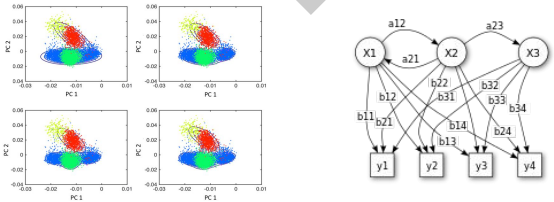




Spectro



Featuremap:
MFCCs and whatnot



Classifier:
GMMs + HMMs

Street

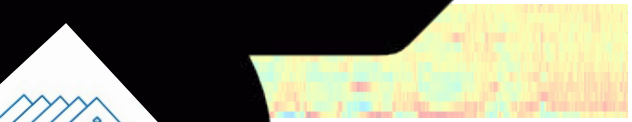
Cafe

Output

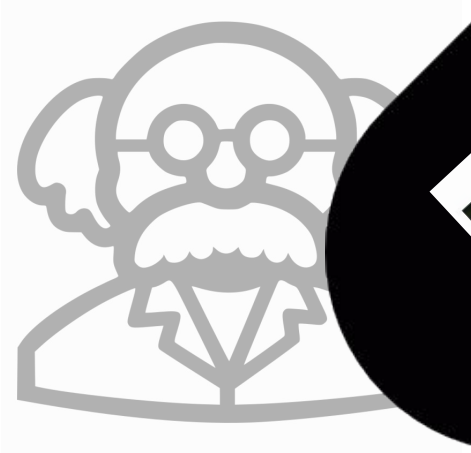
1. How does it work: Approach pre 2016

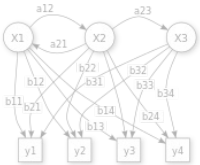
Pre 2015:
MFCCs + GMMs + HMMs

Spectro



Featuremap:
MFCCs and whatnot





Classifier:
GMMs + HMMs

Street

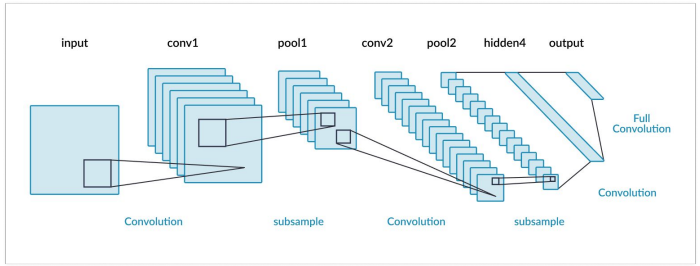
Cafe

Output

2. How does it work: Approach post 2016

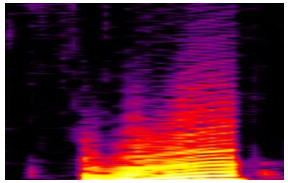
Computer Vision since 2012

Conv. Neural Network (CNN)

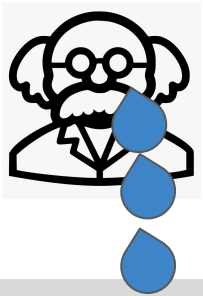


Cat

Dog

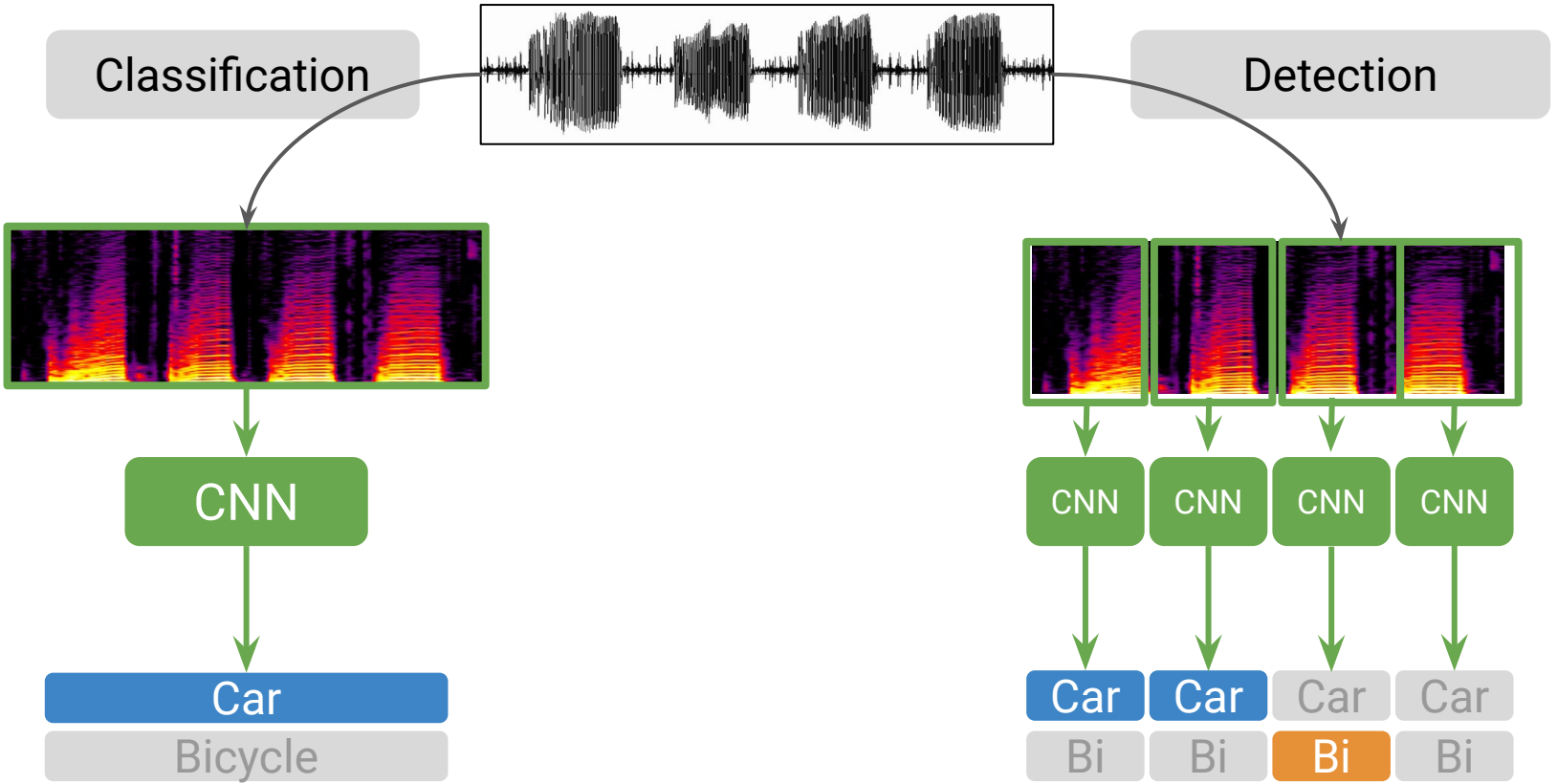


Lol, thats just an image



Sound Recognition 2016

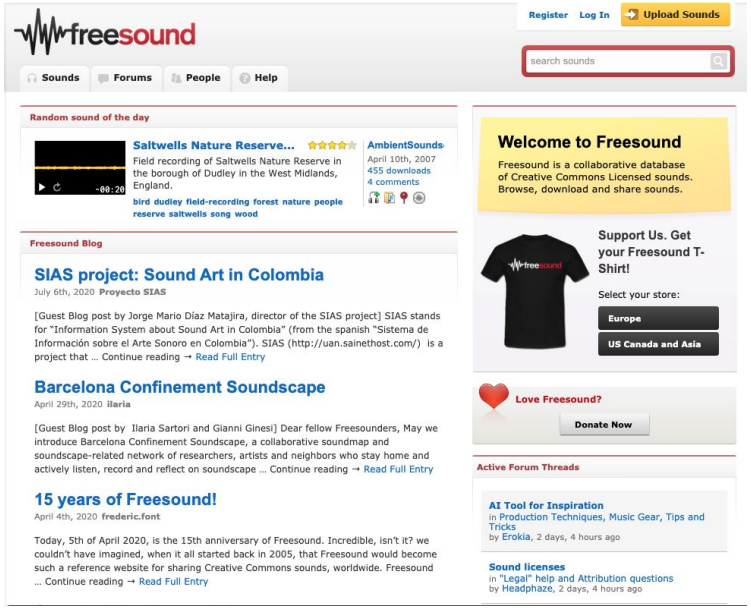
2. How does it work: Approach post 2016

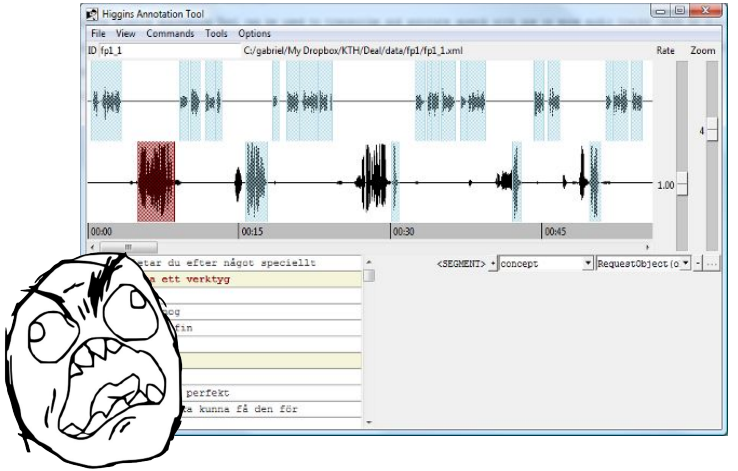


2. How does it work: Approach post 2016

3. Current Challenges and Trends

Problem





3. Current Challenges and Trends

Solution: “Embeddings” (Pretrained Networks)

Supervised

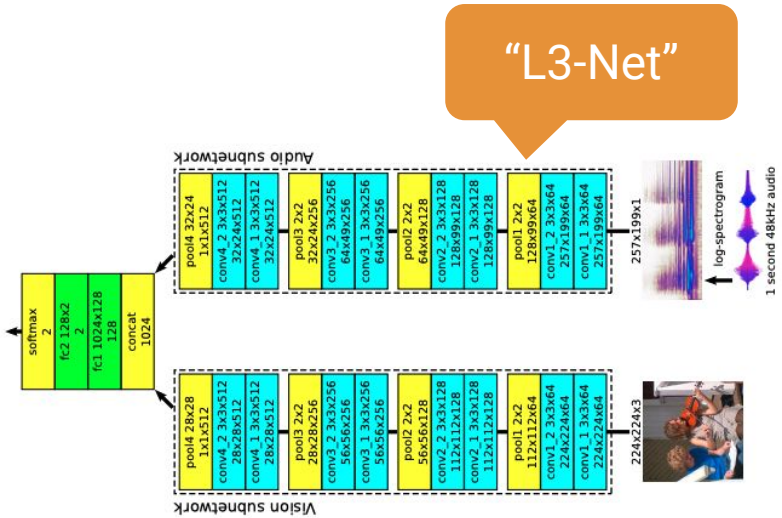
IMAGENET

“VGG-ish”

AudioSet

Unsupervised

“L3-Net”

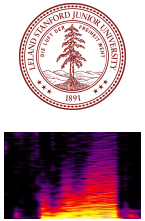


2. Current Challenges and Trends

How to name sounds?

vs

4. Where do I start?



Understand the Basics:

- Take a basic course in CNNs for Computer Vision: http://cs231n.stanford.edu/
- Wrap your head around the spectrogram: https://towardsdatascience.com/getting-to-know-the-mel-spectrogram-31bca3e2d9d0

Solve your Task:

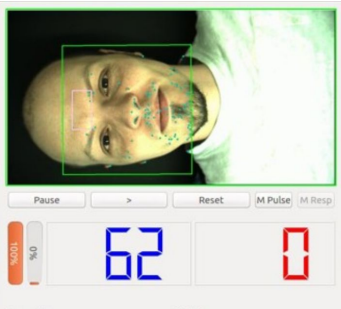
- Check out the anual “DCASE” Challenge: http://dcase.community/
- Search for the task which mostly sounds like yours (starting in 2016)
- Look at the respective baseline system
- Freak out because it’s not that easy after all

The devil is in the detail….

- A thousand spectrogram types
- Dataset Partitioning
- Data-Augmentation completely different
- Convolutional Recurrent Neural Networks
- …..

Shameless self promotion

- Me: PhD at HTWK Leipzig
  - Automatic Recognition of Vocalizations in Infants
  - Contact me: franz.anders@gmx.de
- Department: HTWK Laboratory for Biosignal Processing
  - Measurement of Vital Parameters and Movement with Cameras
  - Research & Application Development for Industry Projects
  - Contact us: www.labp-leipzig.de , mirco.fuchs@htwk-leipzig.de



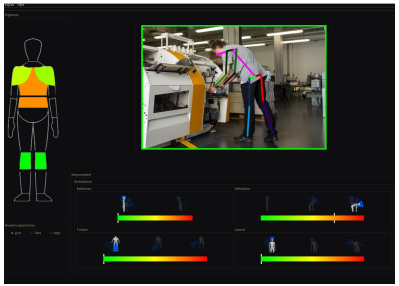


Image Sources

- Bomb-CNN: ps://www.researchgate.net/publication/322303457_Face_Recognition_Based_on_Convolutional_Neural_Network
- Cat: https://de.wiktionary.org/wiki/cat
- CNN: https://missinglink.ai/guides/convolutional-neural-networks/convolutional-neural-network-architecture-forging-pathways-future/
- Baby: https://www.flaticon.com/de/kostenloses-icon/baby_822123
- Trash: https://icons-for-free.com/trash+bin+icon-1320086460670911435/
- Imagenet-Logo: http://www.image-net.org/
- Creation of Adam: https://en.wikipedia.org/wiki/The_Creation_of_Adam
- Mr. Bean Meme: https://imgflip.com/memetemplate/156789093/Mr-bean-copying
- Annotation Tool: http://www.speech.kth.se/hat/
- Audioset-Logo: http://dcase.community/documents/workshop2017/presentations/the_story_of_audioset.pdf
- Stanford Logo: https://en.wikipedia.org/wiki/Stanford_University
- DCASE Logo: https://www.audioanalytic.com/audio-analytic-sponsor-dcase-2018/
- LaBP-Logo: https://github.com/labp