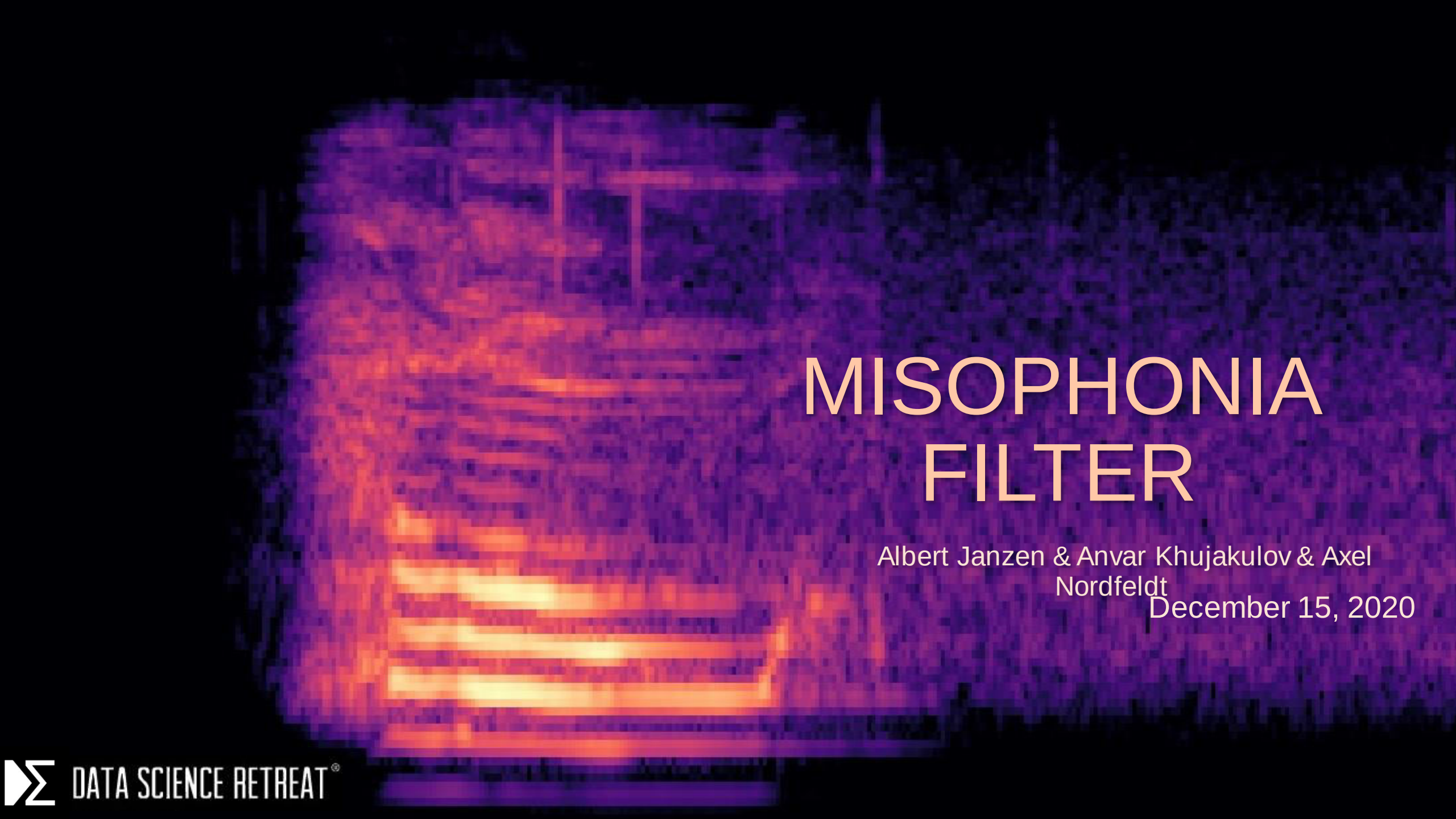


MISOPHONIA FILTER

Albert Janzen & Anvar Khujakulov & Axel Nordfeldt

December 15, 2020

DATA SCIENCE RETREAT®

contents

Miso- (“hatred”) + -phonia (“sound”)

misophonia is a proposed neurological condition, where particular sounds elicit a strong emotional response, e.g. hate, anger, anxiety, rage, panic, physiological distress.

The most common triggering sounds are oral sounds such as

- chewing
- smacking,
- speaking with a dry mouth
- swallowing


SAS-ASMR

Why talking about misophonia?

- up to 20% of the population may have some degree of misophonia
- It impacts the lives of both people with misophonia and the people around them

The role of machine learning

- There are no evidence-based methods for treatment
- Machine learning can help filtering out particular sounds

The Misophonia Filter

is a device on your computer, on your phone and even in your ear that filters out particular sounds. For this, a machine has to learn to

1. classify “disgusting” sounds,
2. recognize the start and the end of these sounds and
3. predict whether such a sound will appear in the future, e.g. in 2 milliseconds

The device then filters out the sounds by cancelling out the relevant frequencies.

the approach

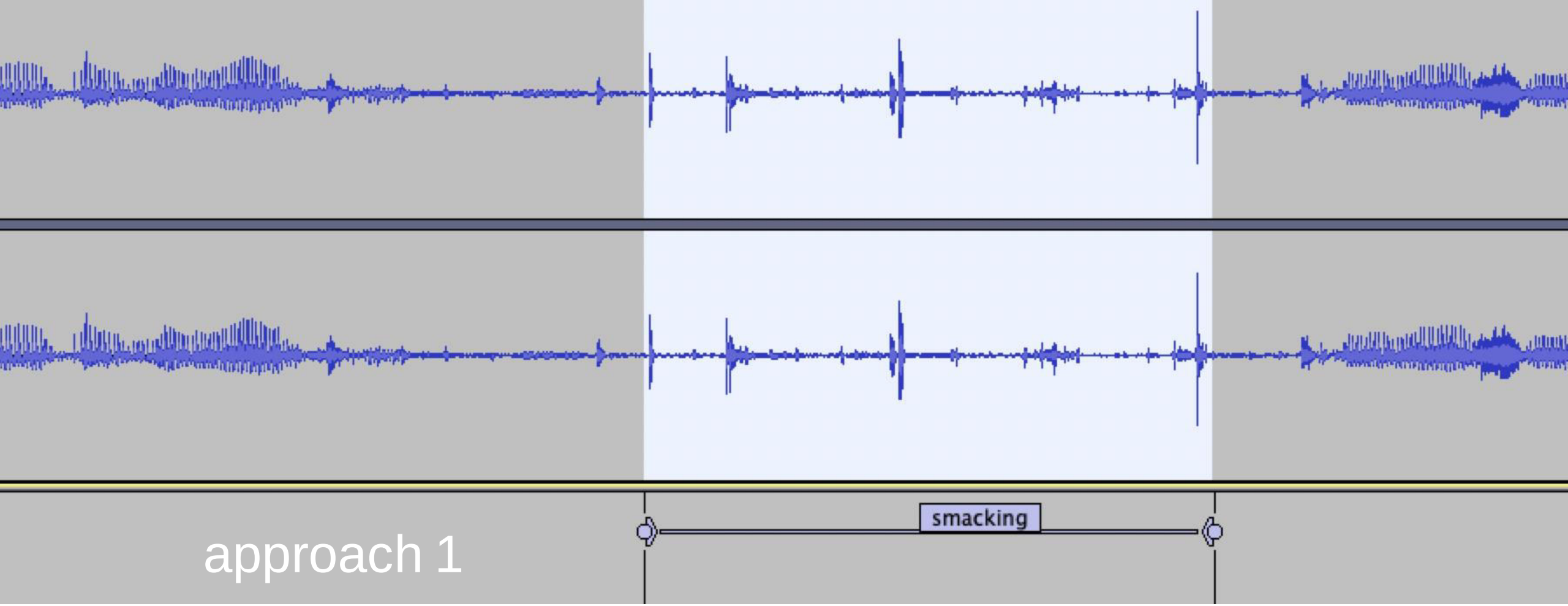
- start with step 1, making a machine recognise whether or not there is a “disgusting” sound in a given video.
- step 2 and step 3 depend on the success of step 1

collect and label data in two different approaches:

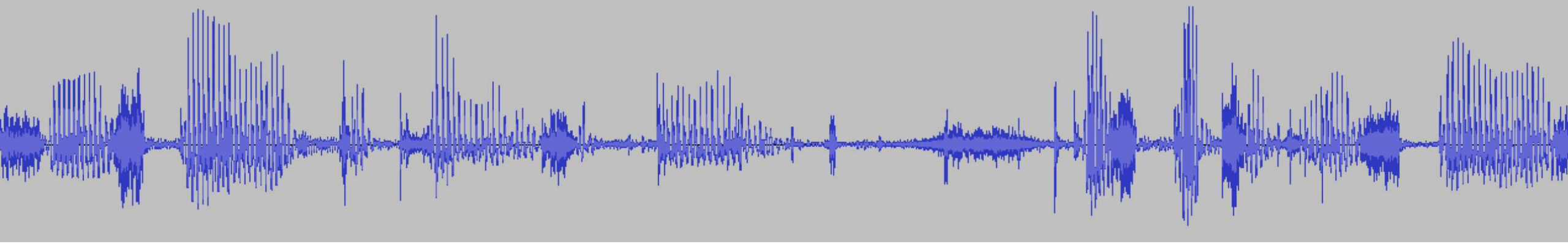
1 label data manually

2 use different sources for positive (with “disgusting” sound)

and negative class (without “disgusting” sound)

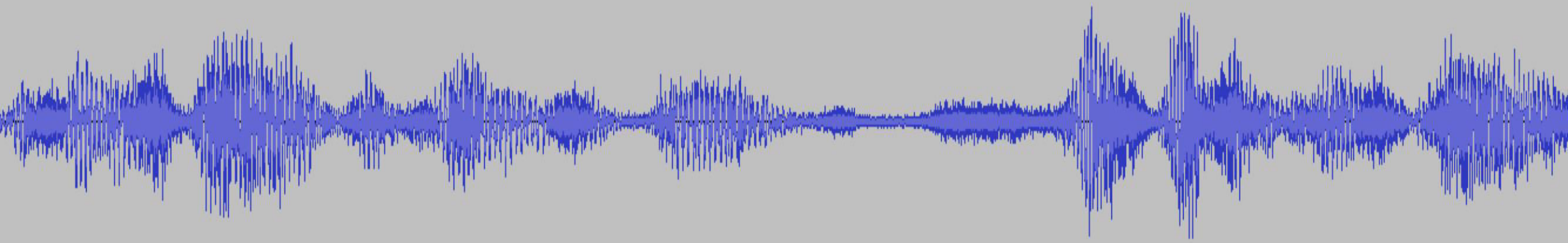


listen to a lot of videos and label the start and the end of each “disgusting” sound



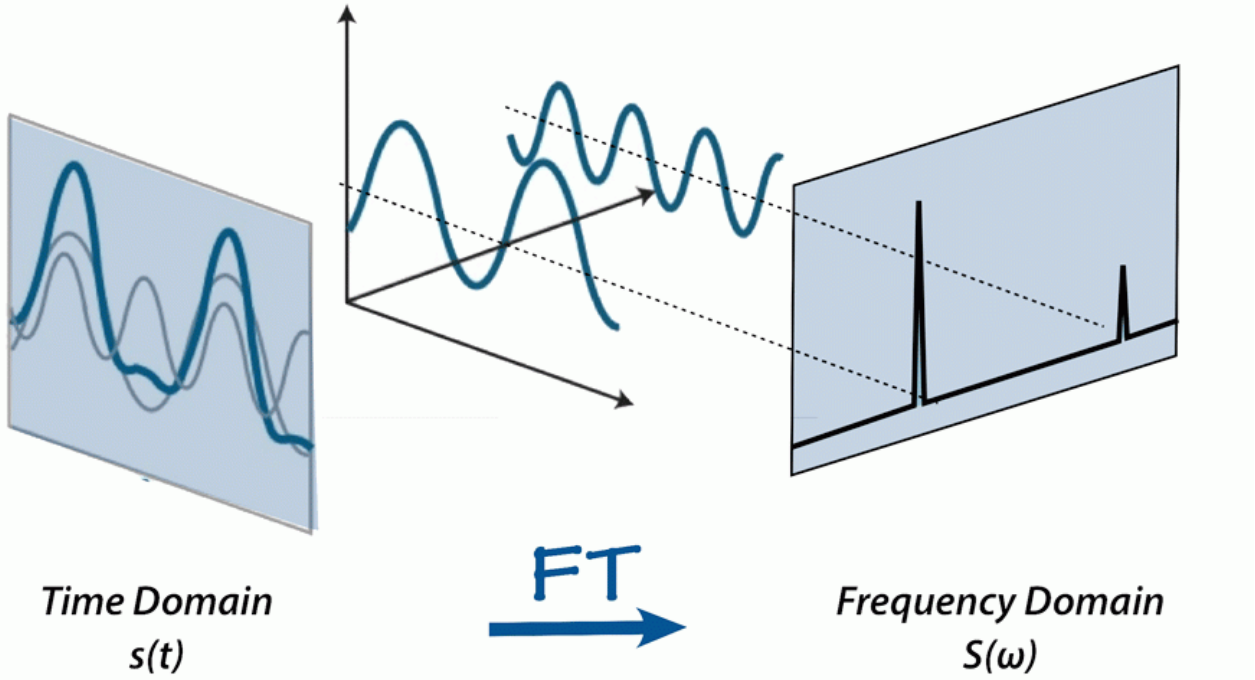
approach 2

- slice up video containing only “disgusting” sounds
- apply data augmentation



processing the data

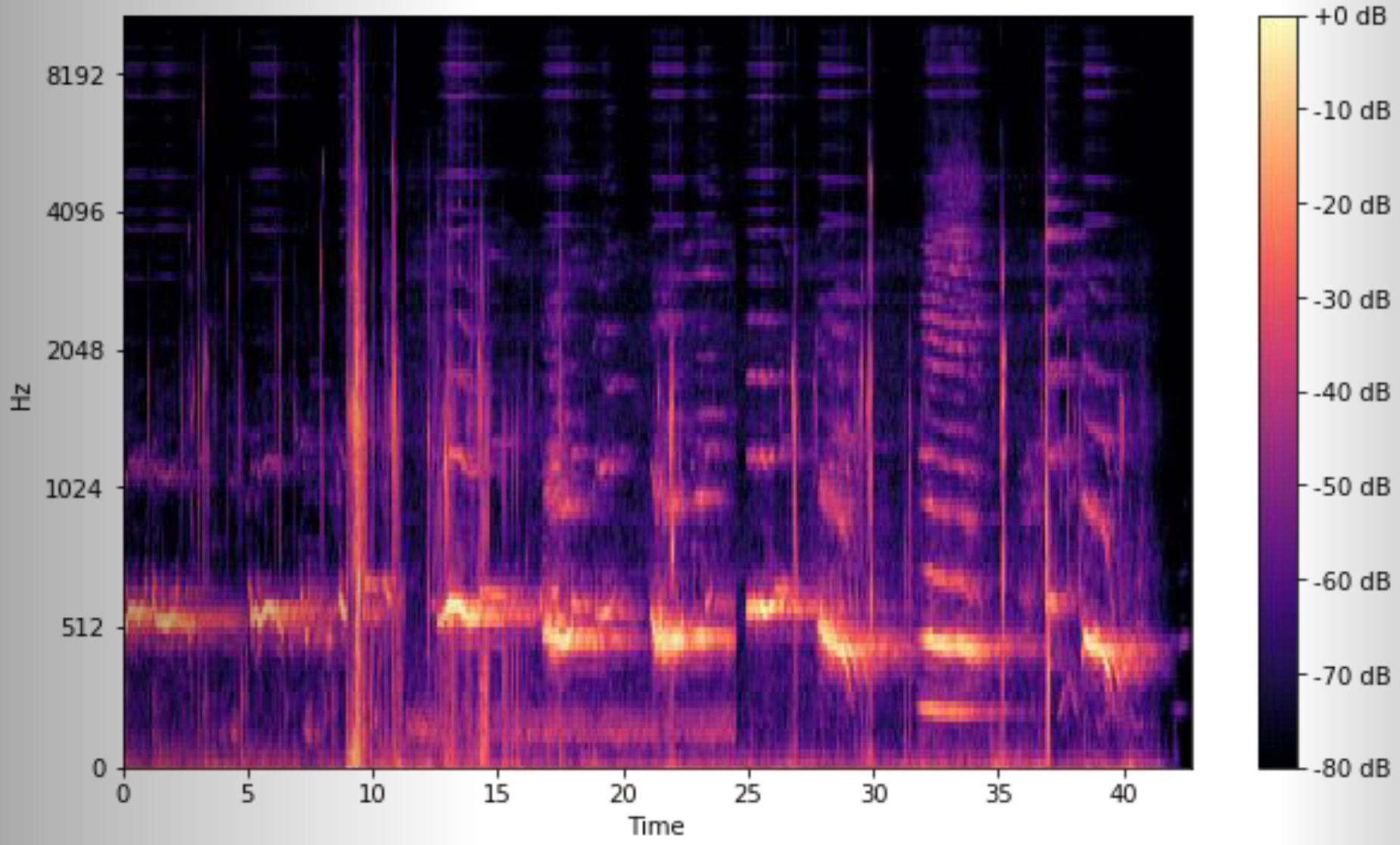
audio files are stored as a series of numbers, where each number denotes the amplitude at a point in time



with a Fourier-Transformation, the underlying frequencies can be detected

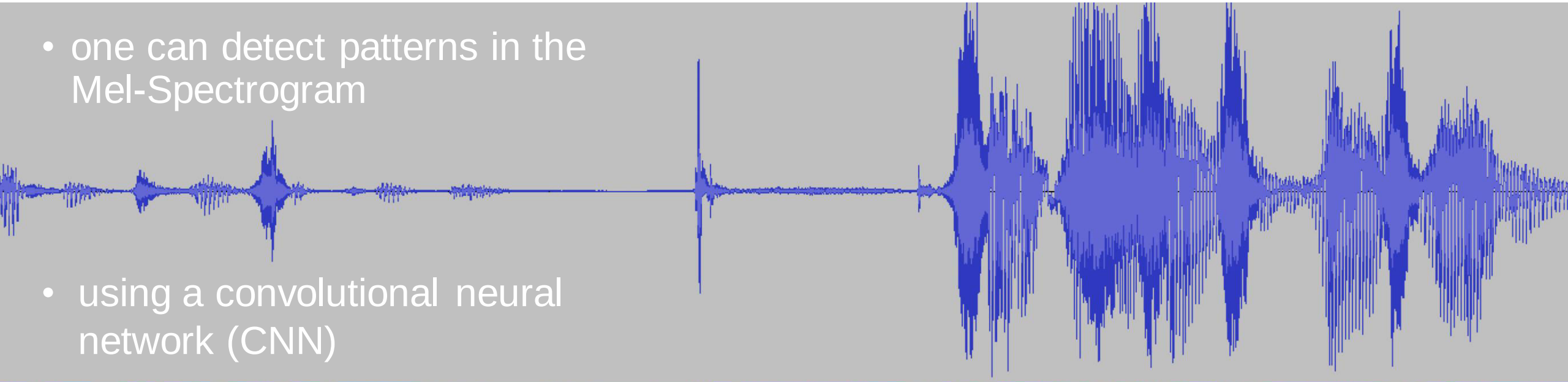
the Mel-Spectrogram

shows the amplitude of each frequency at each point in time

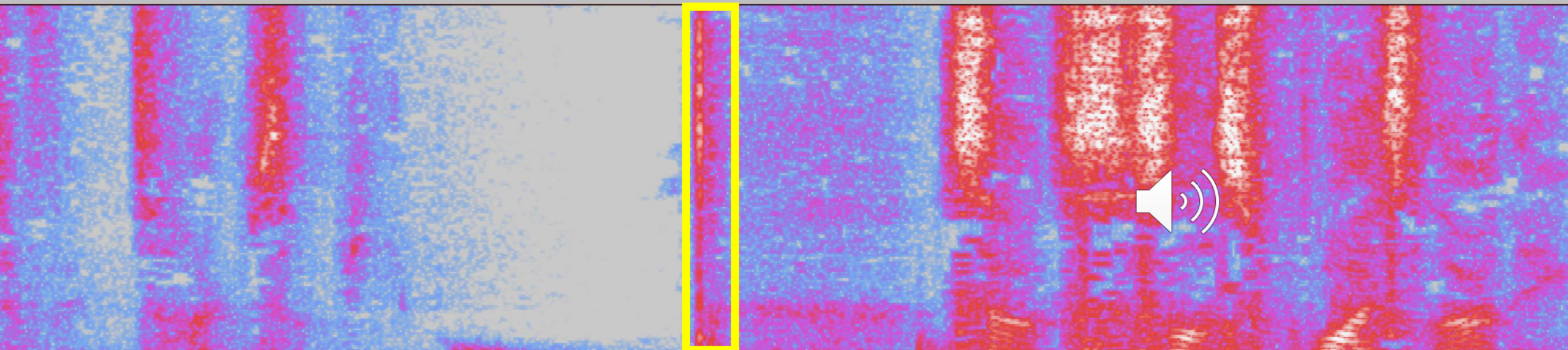


the machine learning technique

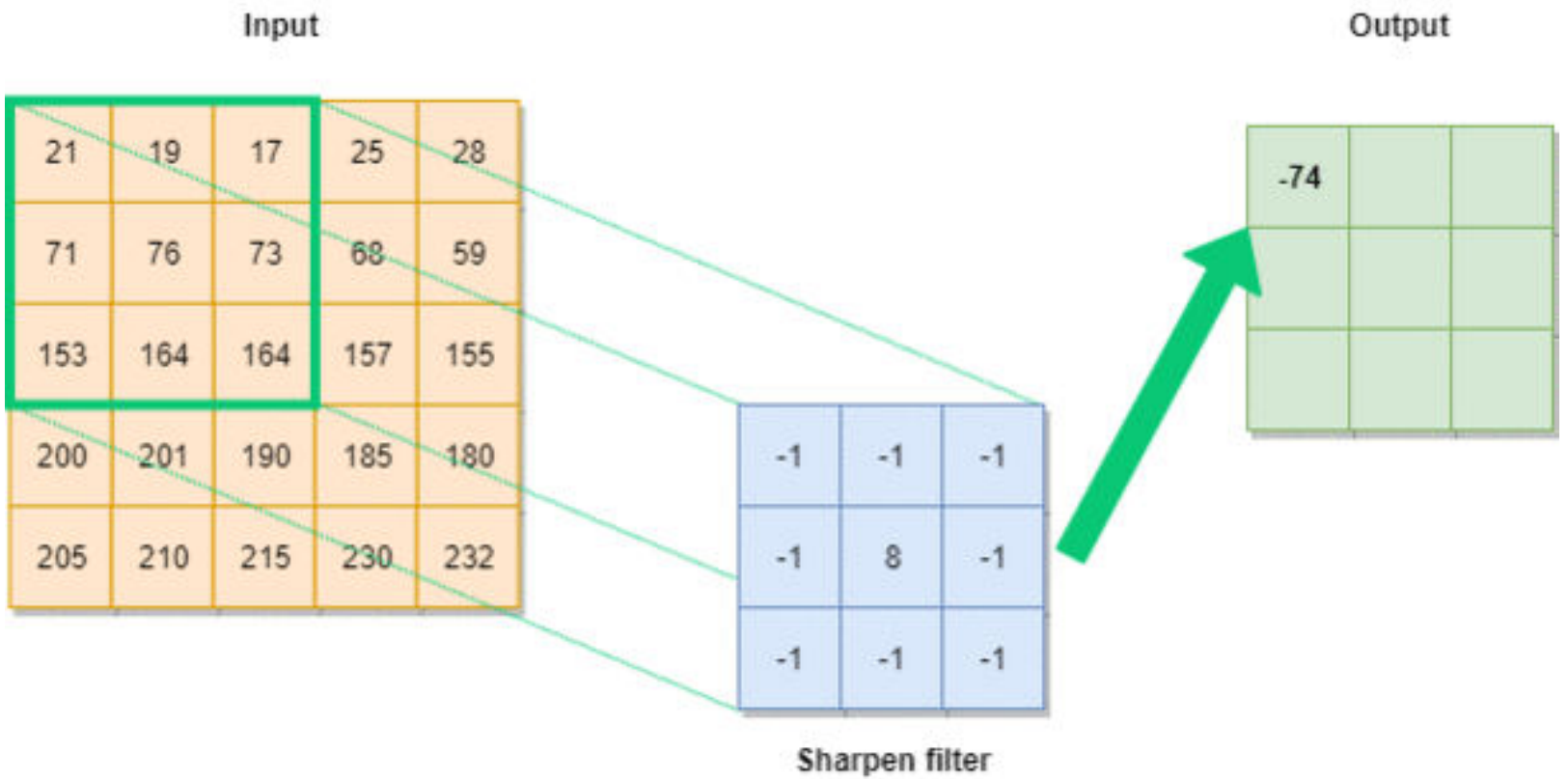
- one can detect patterns in the Mel-Spectrogram

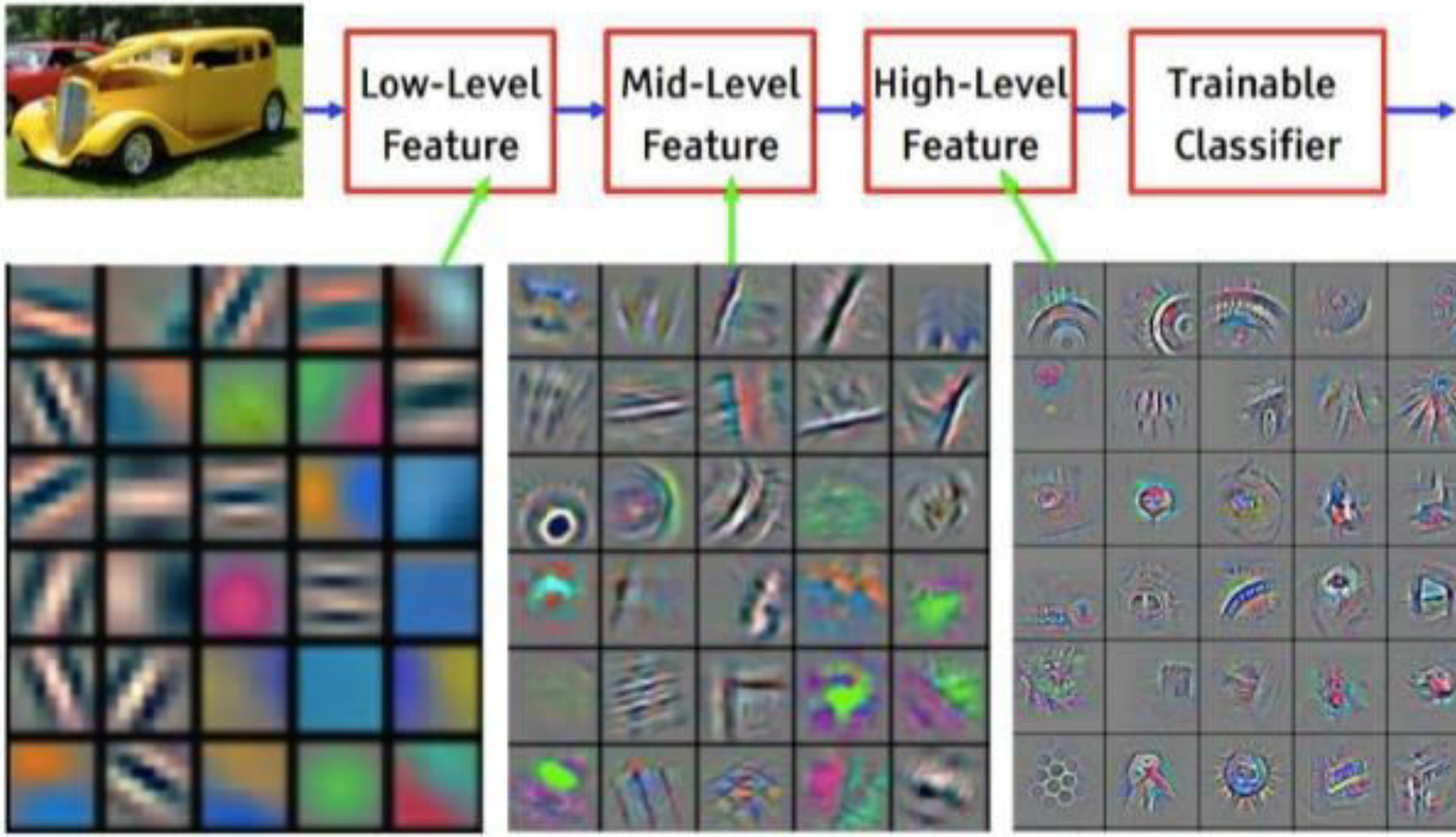


- using a convolutional neural network (CNN)



how do CNNs work?

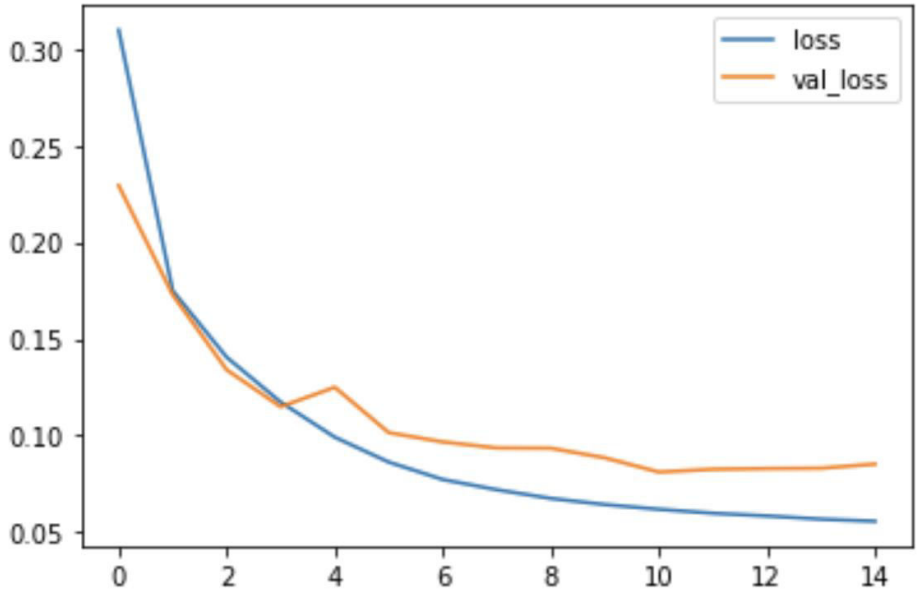


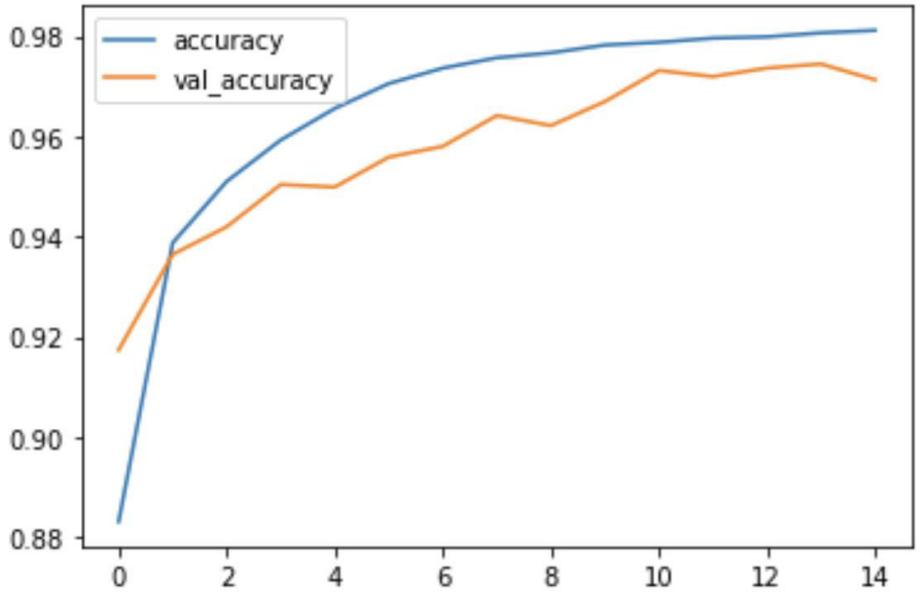


Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

results

- a CNN with one 2-dimensional convolutional layer yields the best results





limitations

- collecting enough data that represents the variety of triggering sounds in all the different ways to record them
- predicting the sounds in advance depends on there being clues that humans might not pick up
- apply CNNs on raw audio data: a 1-dimensional vector

scope for improvement

knowing that a machine can recognize “disgusting” sounds is a promising fundament for teaching a machine to recognize their start and end, and when they appear in the future

There is a long and exciting way ahead!