Hortense Fong, Columbia University

Applying “Explainable” AI: Using Theory to Understand AI Emotion Models

Applying "Explainable" AI: Using Theory to Understand AI Emotion Models

Hortense Fong

MSI Analytics Conference

May 2023

Columbia Business School

AI has made quite the splash … through successful prediction and generation

CNBC

Google CEO: A.I. is more important than fire or electricity

Sundar Pichai says it is
artificial intelligence.

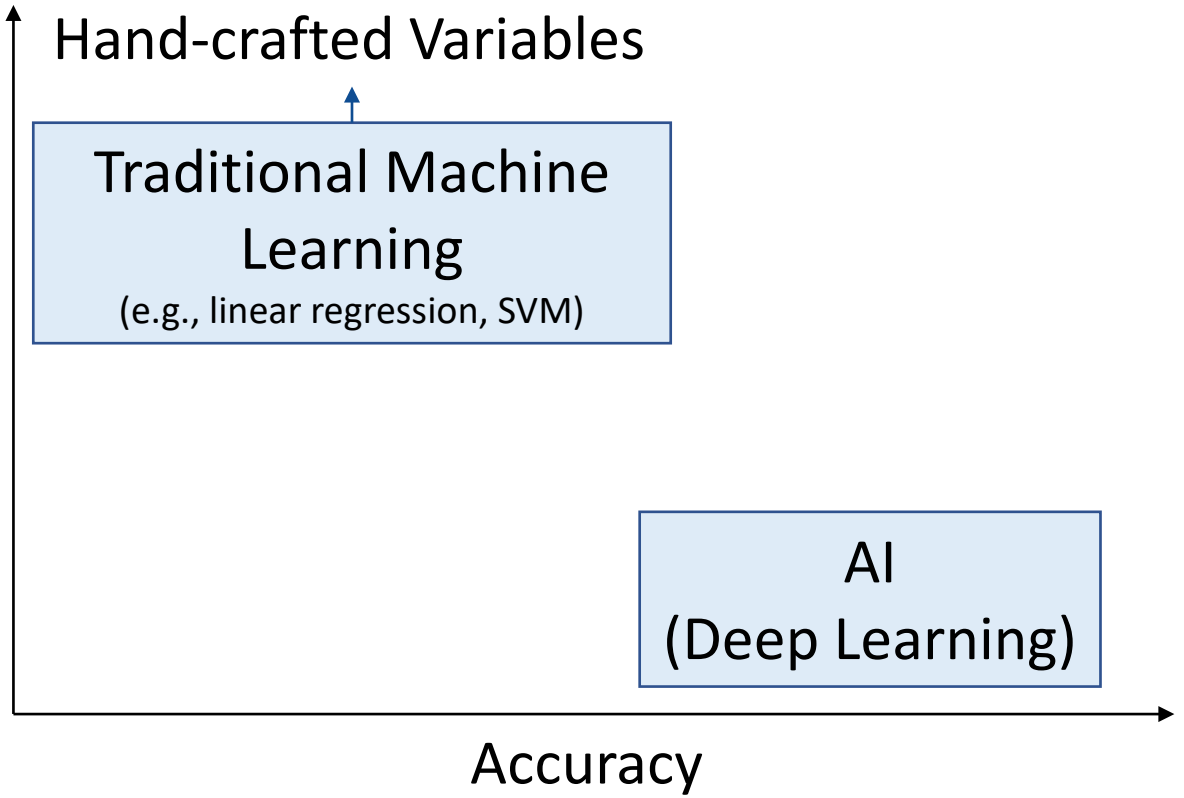
Simulated exams	GPT-4 estimated percentile
Uniform Bar Exam (MBE+MEE+MPT)[1]	298/400 ~90th
LSAT	163 ~88th
SAT Evidence-Based Reading & Writing	710/800 ~93rd
SAT Math	700/800 ~89th
Graduate Record Examination (GRE) Quantitative	163/170 ~80th
Graduate Record Examinat	169/170

Newstalk

AI art piece wins international photography competition

A German artist has turned down an international photography award as he revealed the photograph was AI-generated.

The improved performance comes at the cost of explainability



Ability to provide a qualitative understanding of the relationship between the input variables and the response (Ribeiro et al. 2016)

Why do we care about explainability?

Explainability is important for:

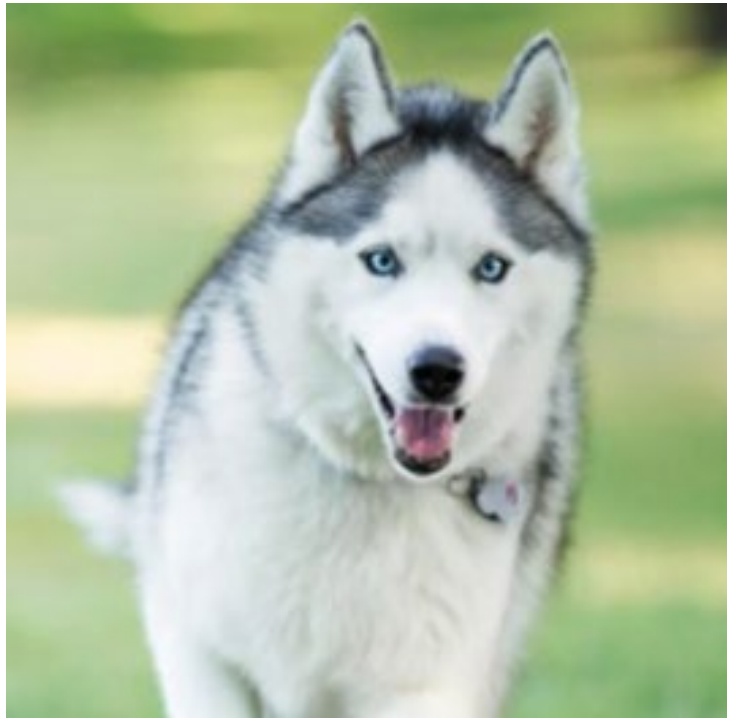
1. Managers to have trust in predictions → deploy model at scale
2. Generalizability/robustness of model in other settings
3. "If your system doesn't work and you don't know why it's quite hard to improve it." – Uber AI researcher
4. Ethical and fairness concerns

Forbes

Nobody Can Explain For Sure Why ChatGPT Is So Good At What It Does, Troubling AI Ethics And AI Law

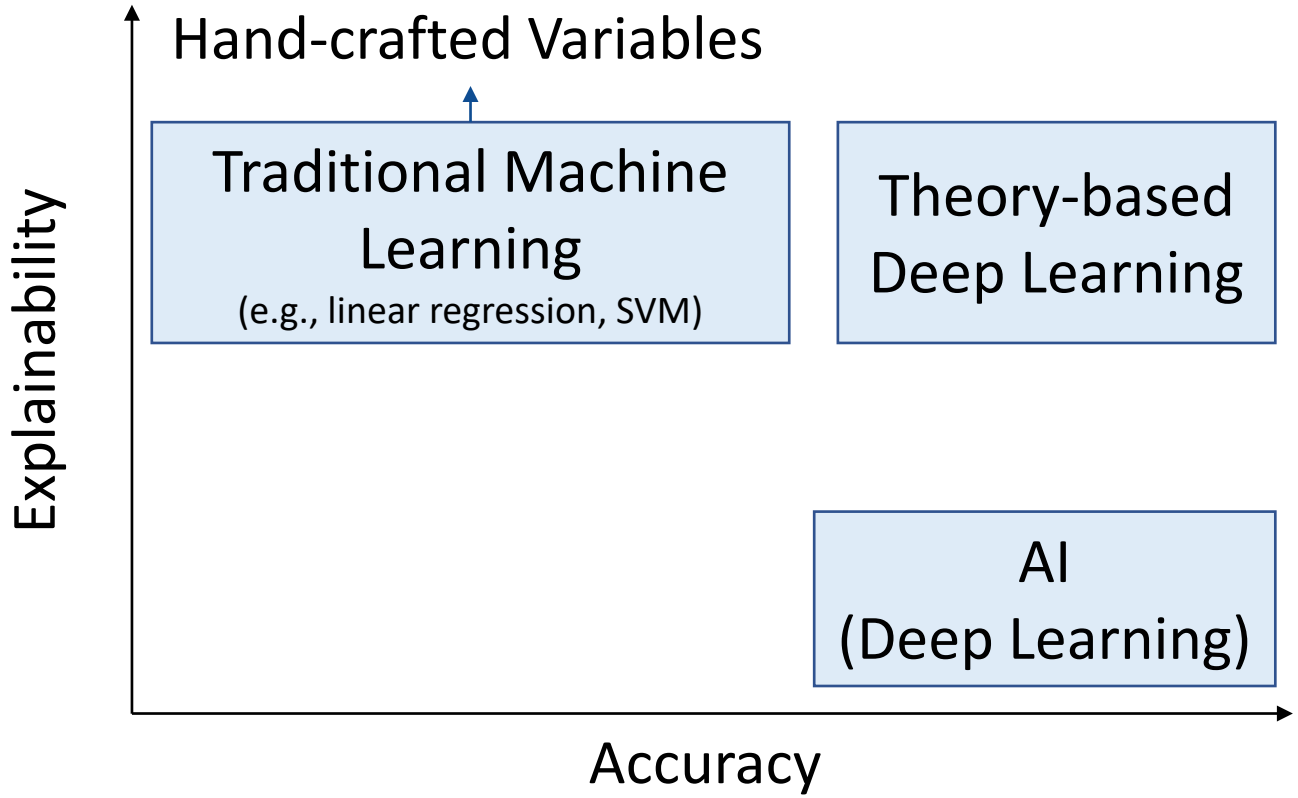
Wondered how it is that ChatGPT and other generative AI are so good at what they do? AI researchers and AI makers are also unsure and unable...

Why do we care about explainability?
Can we classify wolves vs. huskies (breed of dog)?



Source: Ribeiro et al. (2016)

Incorporate theory to gain explainability



- Deep learning improves accuracy but loses explainability
- Theory enables explainability
- Ideally without losing predictive accuracy

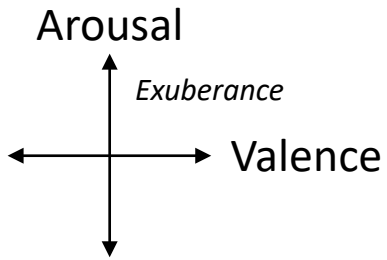
Application: A Theory-Based Explainable Deep Learning Architecture for Music Emotion

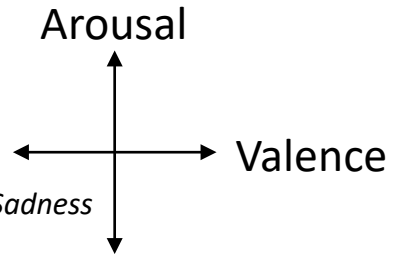
Hortense Fong, Vineet Kumar, K. Sudhir

“Music is the language of emotion”

It can elicit a wide range of emotions

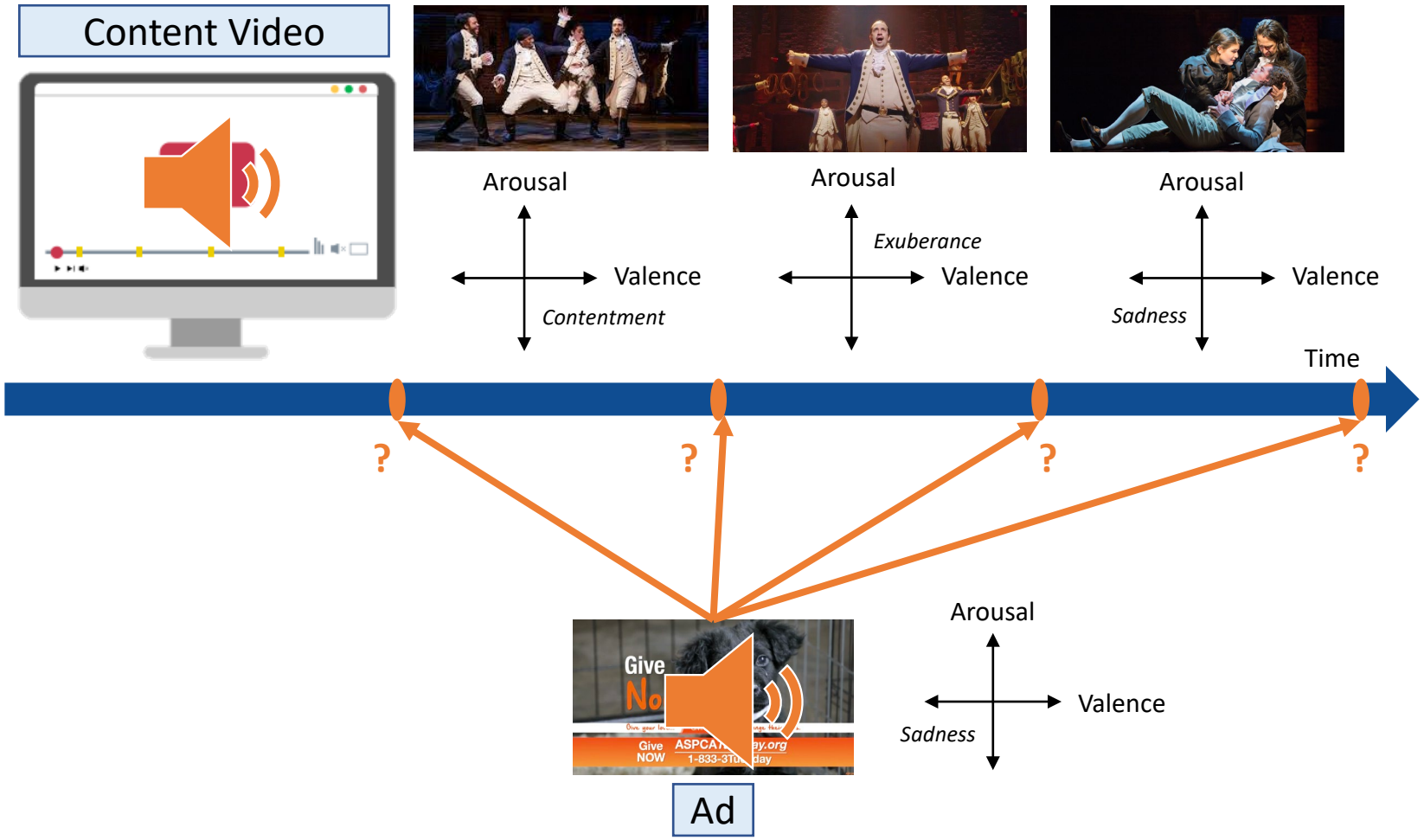
ASPCA Giving Tuesday





Emotion induced by content impacts ad effectiveness

Where to insert an ad based on content emotion?



- Music is designed to elicit the intended emotion in video
- Content emotion varies over time
- Interaction of content emotion and ad emotion impacts ad effectiveness
- Billions of videos on YouTube

→ Use music emotion as proxy for video content

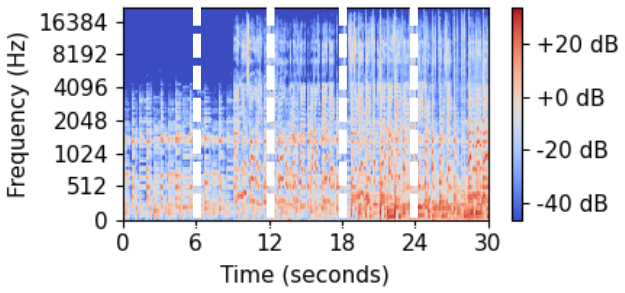
→ Need model to determine optimal ad insertion positions at scale

→ Where to insert ad?

Predicting Emotion from Music

Input

Convert raw music into spectrogram images



Model

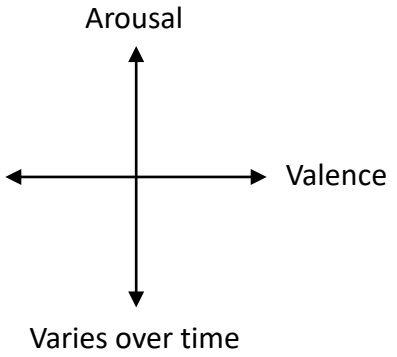
Deep learning model (CNN)

Music theory-based filters provide explainability without reducing performance

Our Contribution

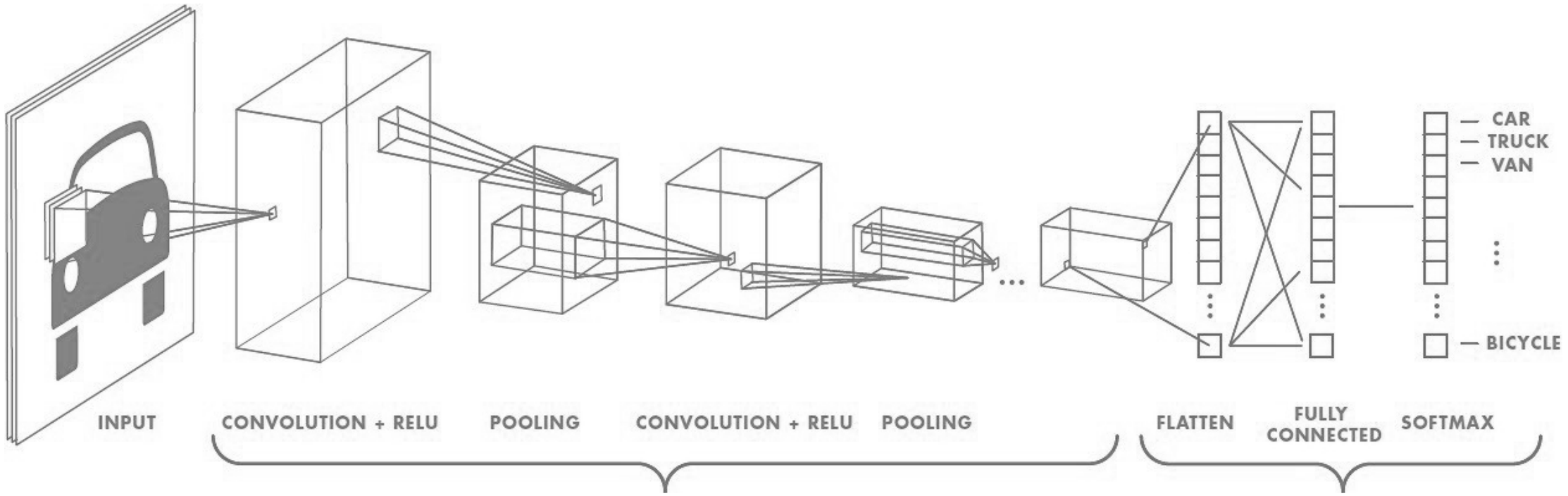
Output

Predicted emotion (valence + arousal) evoked by music



Convolutional neural network (CNN) was designed for computer vision

Goal of model: Classify vehicle in each image



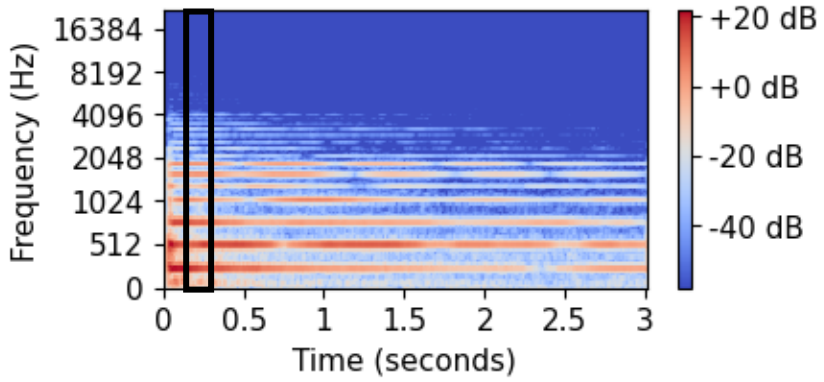
Feature Learning
Automatically learns features of data useful for classification

Classification
Uses learned features to estimate probability of each class

Source: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

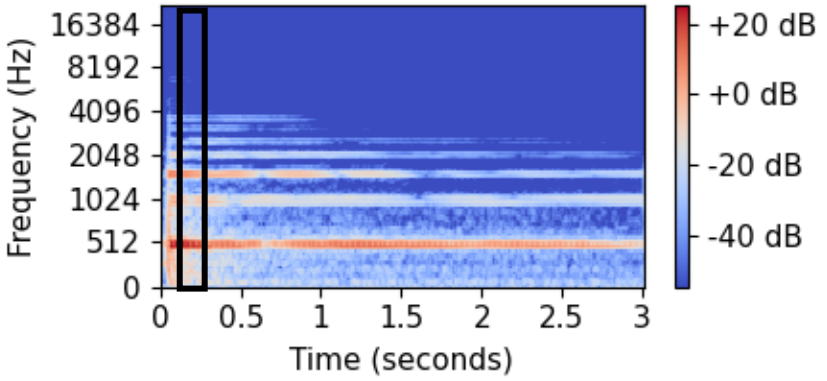
Deep learning for music uses vision convolution filters

Octaves: f_0 & $2f_0$



Pleasant / Consonant

Minor second: f_0 & $(25/24)f_0$



Jarring / Dissonant

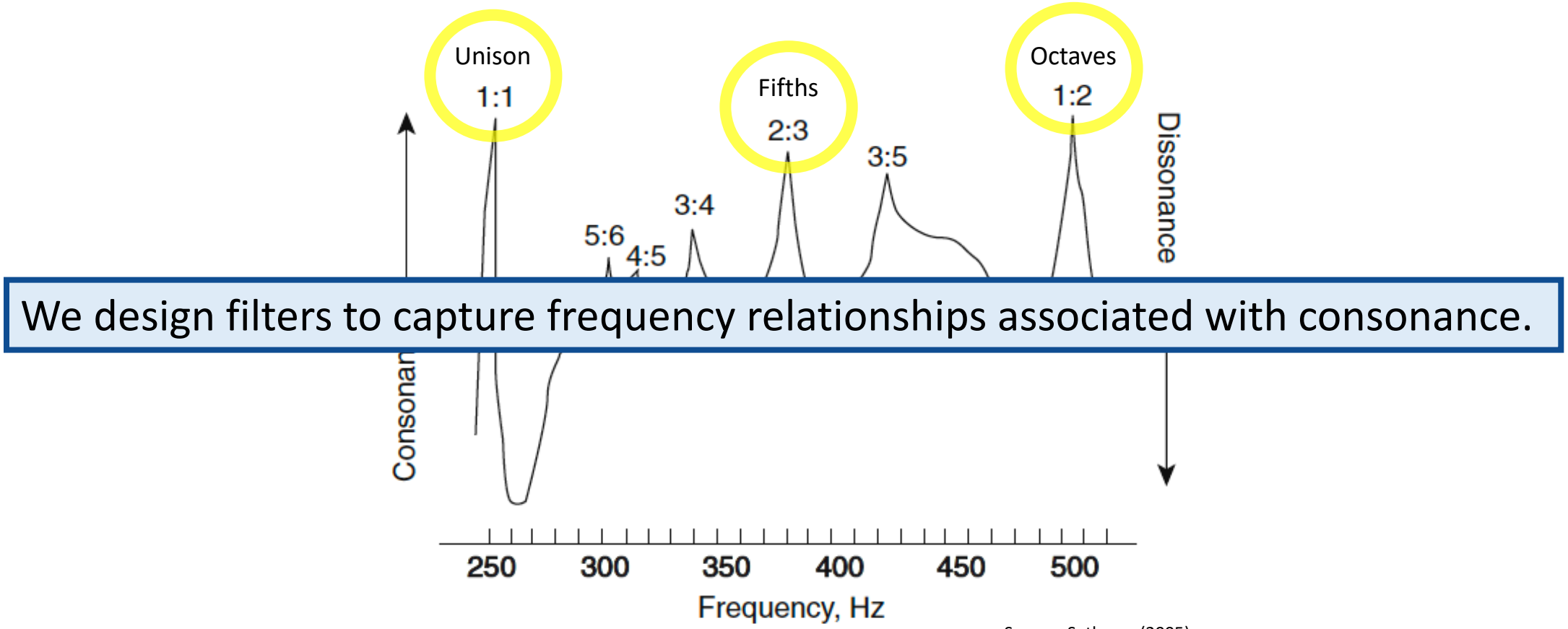
Important musical features rely on non-local information!

Music Theory Background

Emotion is related to consonance and dissonance

Consonance: A combination of notes that sound pleasant when played together
→ Positive valence, low arousal

Dissonance: A combination of notes that sound jarring when played together
→ Negative valence, high arousal



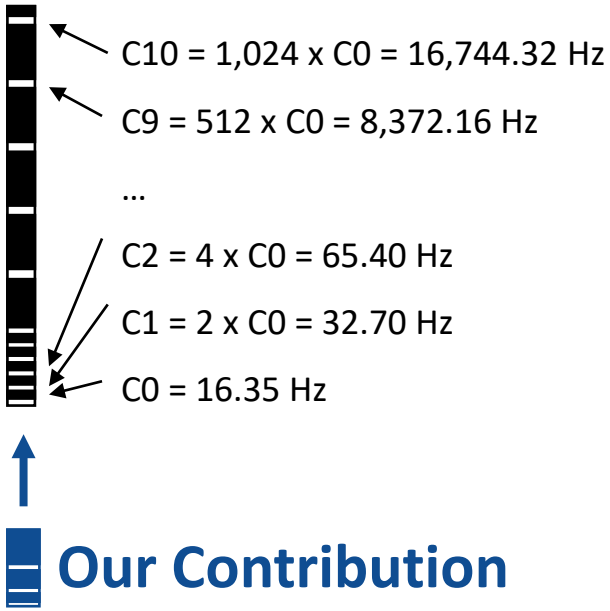
We design filters to capture frequency relationships associated with consonance.

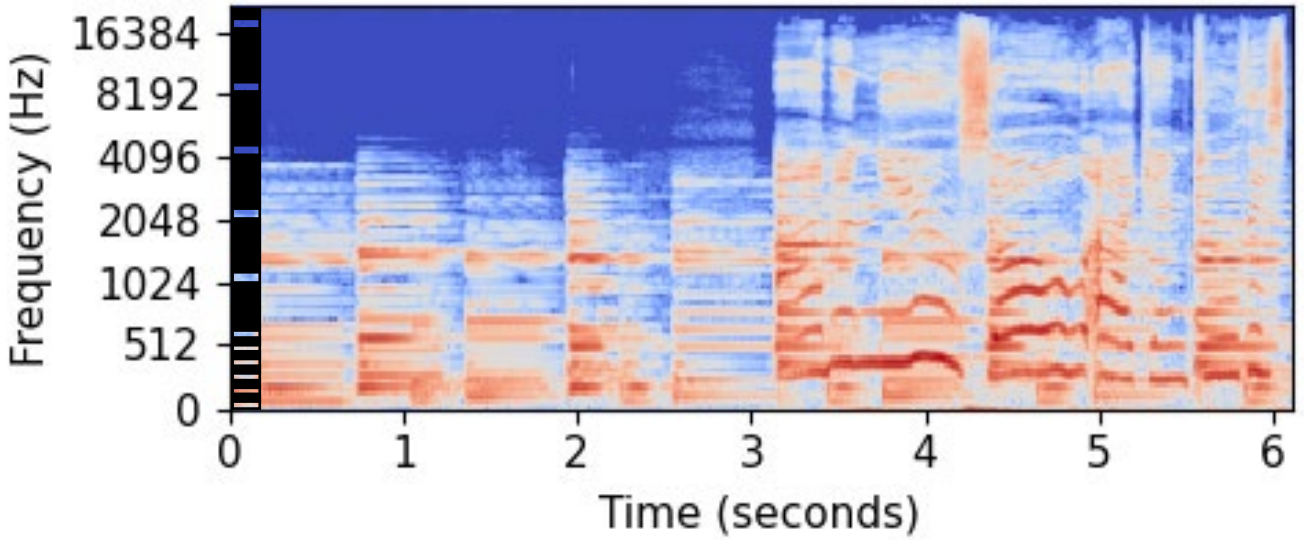
Source: Sethares (2005)

Our contribution: Designing theory-based filters from physics of sound

Octaves for pitch class C:

frequency $C_n = 2^n C0$





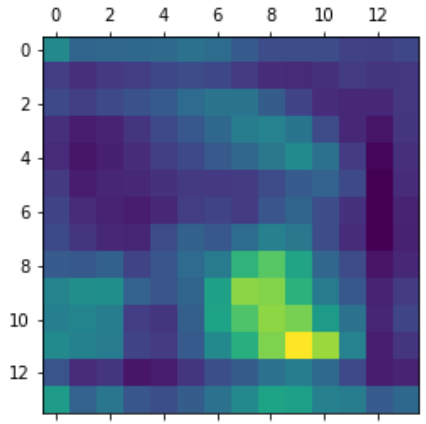
Our consonance filters are based on non-contiguous frequency ratios.

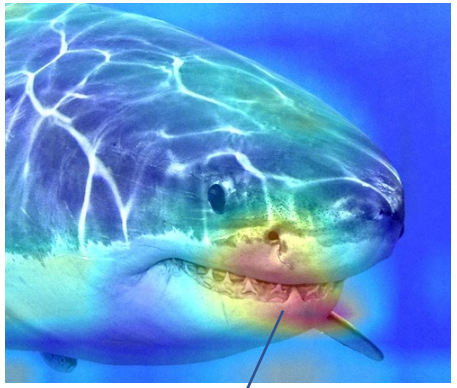
Explainability: Grad-CAM visual explanation for image CNN

Input

Grad-CAM Heatmap

Output





Shark

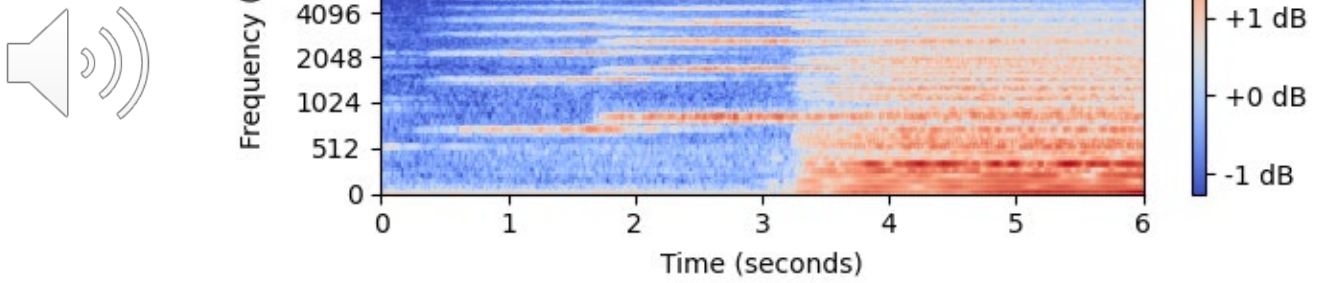
Sharp teeth important in classification!

Whale

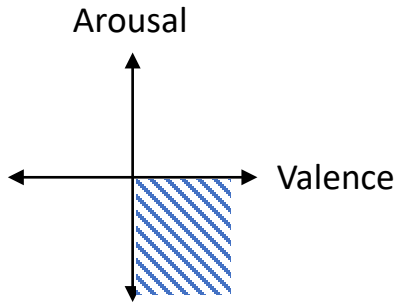
Source: https://medium.com/@stepanulyanin/implementing-grad-cam-in-pytorch-ea0937c31e82

Explainability: Why does the model predict what it predicts?

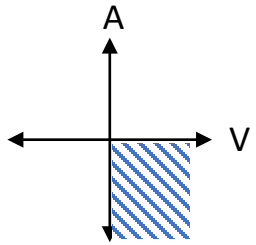
Input: mel spectrogram



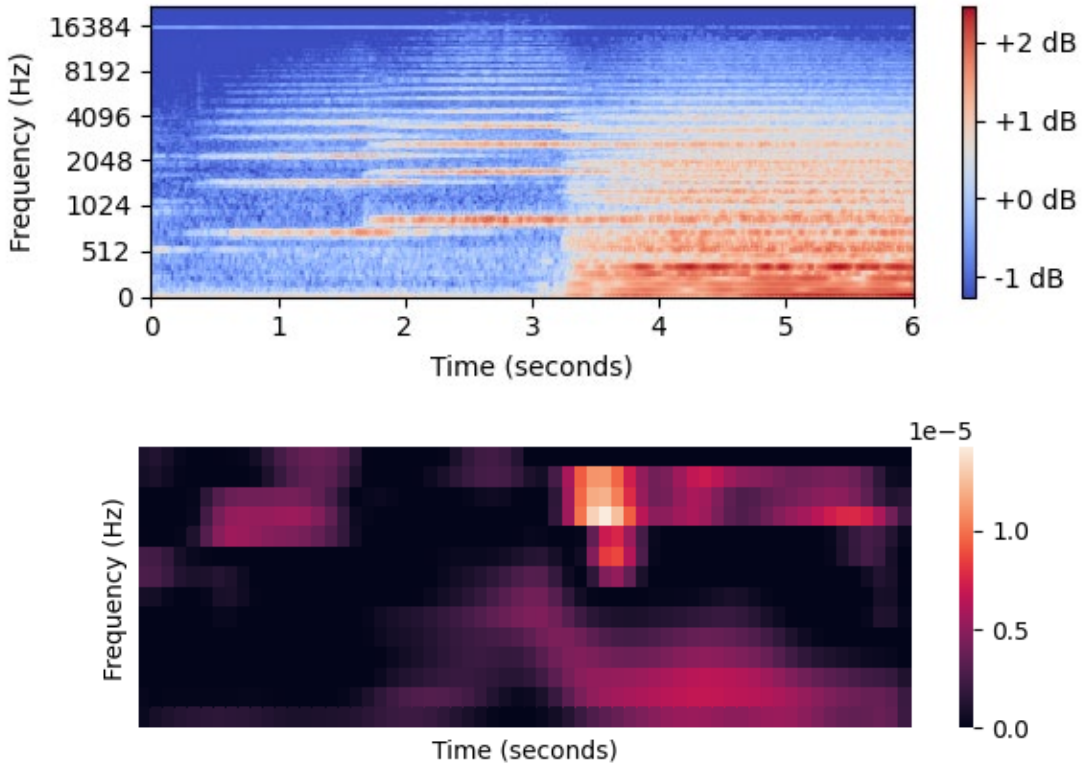
Output: emotion



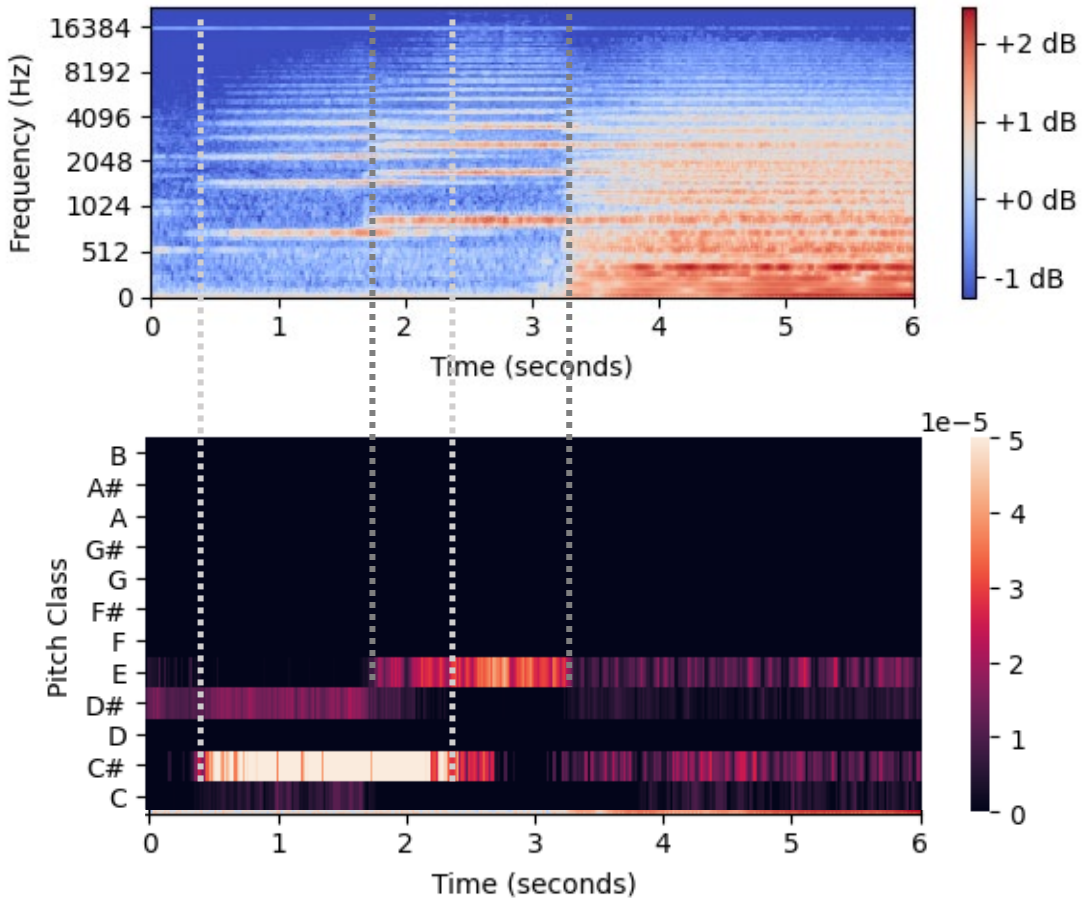
Explainability: Our theory-based filters generate explainable Grad-CAM heatmaps



Atheoretical Filter Grad-CAM Heatmap



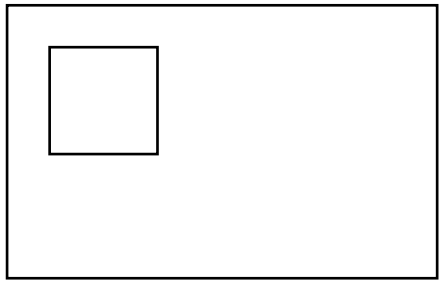
Our Theory-Based Filter Grad-CAM Heatmap



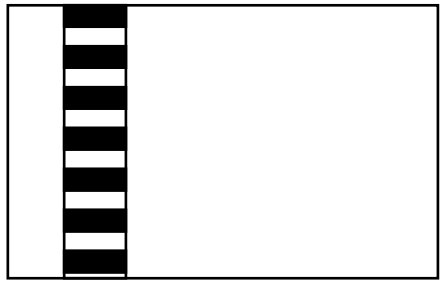
Brightness identifies points of consonance

In addition to explainable, our model is more parsimonious

Atheoretical DL: CNN + Square Filter
(Chowdhury et al. 2019)



Theory-based DL: CNN + Consonance Filter
(Our model)

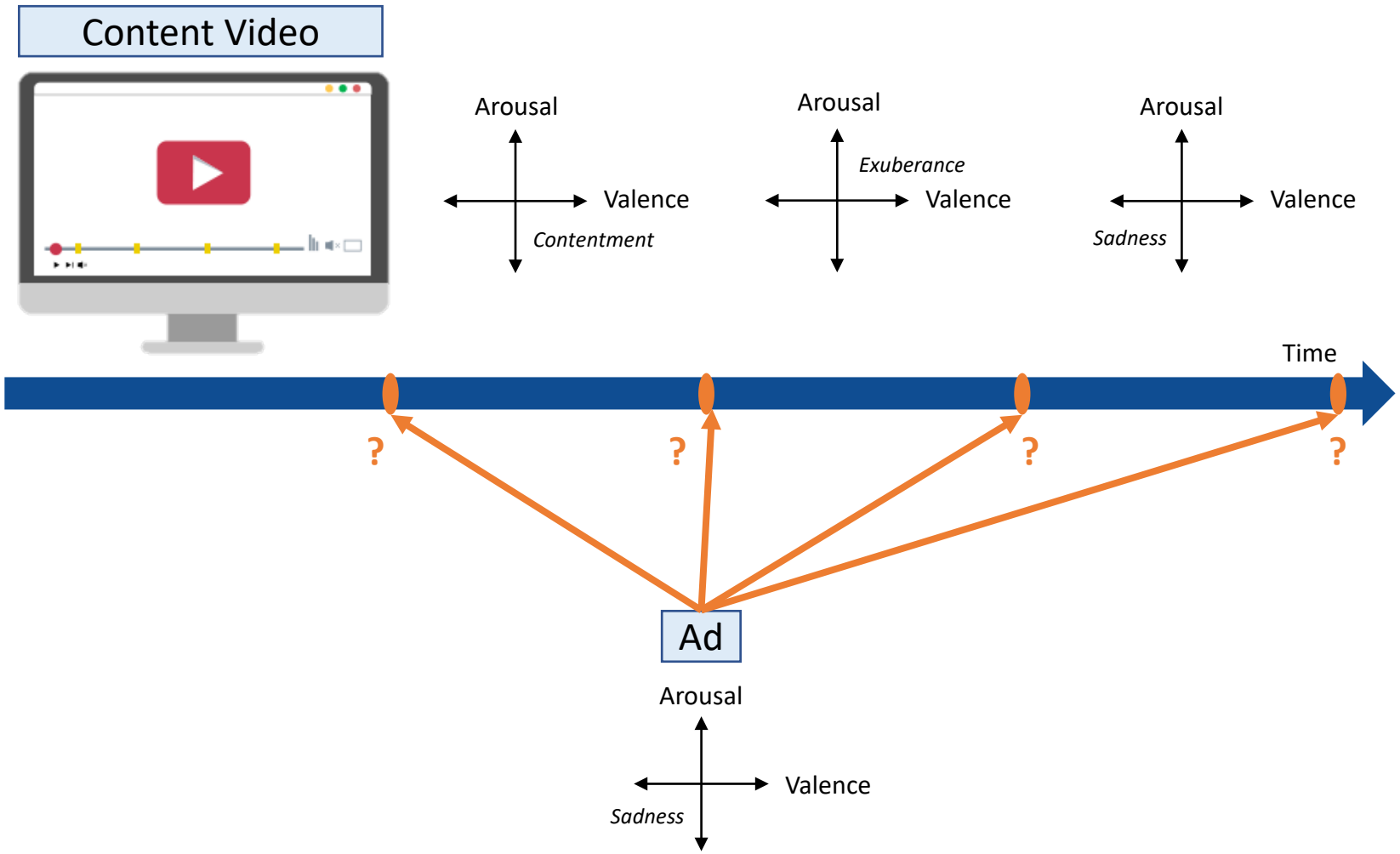


Number of parameters

~5,000,000

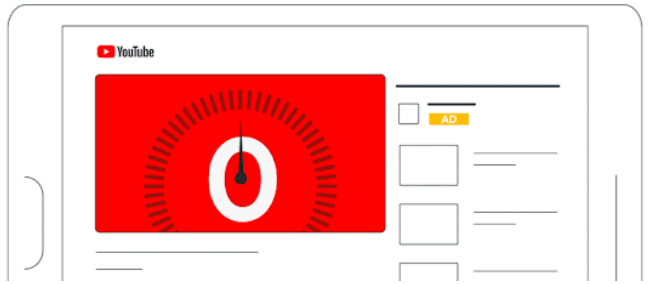
50,900

Ad Insertion Application: YouTube Mid-Roll Ads



Ad Outcomes

1. Ad skip



2. Brand recall

Use Model for Emotion-based Ad Position

Tag Source	Avg. JS Distance	Avg. Recall Rate
Human	0.21	31%
Traditional ML	0.64	16%
Atheoretical CNN	0.36	29%
MusicEmoCNN (Our Proposed Model)	0.38	30%

Our proposed deep learning model (MusicEmoCNN):

- works in real-time
- is scalable
- is explainable

What is theory?

- Theories from natural science
  - Physics of sound
  - Human vision
- Theories from social science
  - Music theory
  - Prototype theory
- Managerial knowledge

Thank you!

hf2462@gsb.columbia.edu