Transcribing Monophonic Audio with Deep-Learning Data

Evan Matthews, Kevin Zhou
Spring 2023

Overview

- The problem to solve
- Introduction to artificial intelligence
- The experiment
  - Setup
  - Gathering data
  - Results

Problem to solve

- We can convert sheet music -> audio (manually, MIDI, etc.)
- *Some* people can convert audio -> sheet music (manually by transcribing)
- Is it possible to *automate* the conversion from audio to sheet music?
- Reasons:
  - Transcribing process is tedious
  - Transcribing requires a good ear and perfect/relative pitch
  - Sheet music becomes a more open medium for singers/composers
  - The process of composing becomes much faster

How can we solve this problem?

Artificial Intelligence and deep-learning!

Introduction to Artificial Intelligence

- Suppose you are ChatGPT, and you are told **f(x) = 3x+2**
  - What is **f(1)**?
  - **f(3)**?
  - **f(ζ)**?

- Your accuracy: 66.67% ❌
  - *We can do a little better…*

Introduction to Artificial Intelligence

- Suppose you are ChatGPT, and you are told **f(x) = 3x+2, ζ = 3**
  - What is **f(1)**?
  - **f(3)**?
  - **f(ζ)**?
  - **f(ʊ)**?
- Your accuracy: 75.00% ❌
  - *Getting better…*

Introduction to Artificial Intelligence

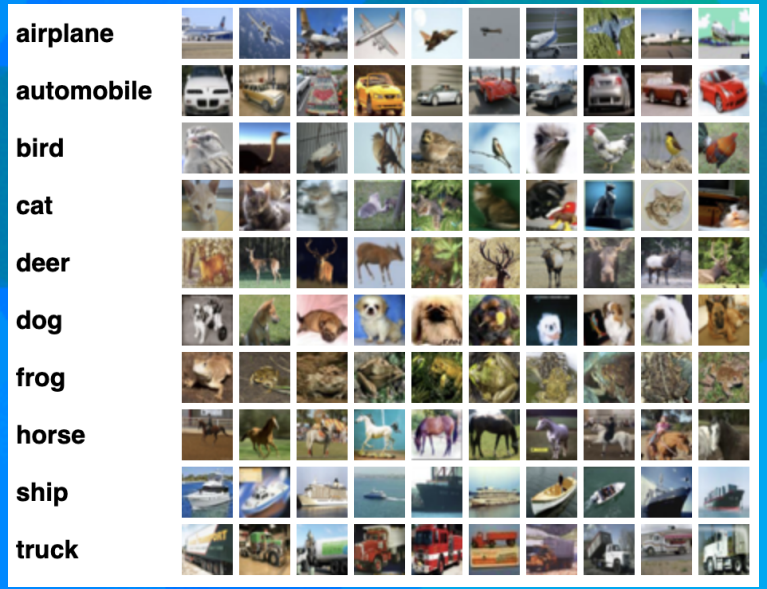
- Suppose you are ChatGPT, and you are told **f(x) = 3x+2, ζ = 3, ʊ = 0**
  - What is **f(1)**?
  - **f(3)**?
  - **f(ζ)**?
  - **f(ʊ)**?
  - **f(🪴)**?
- Your accuracy: 87.25% ✅
  - *… and so on.*

Deep-Learning Models

- Some types of artificial intelligence work by creating a statistical structure called a **deep-learning model**:
  1. Run a decent number of known relations through your model
  2. Test your model with more relations to obtain an accuracy
  3. Given a high average accuracy, implement your model

Why use deep-learning models?

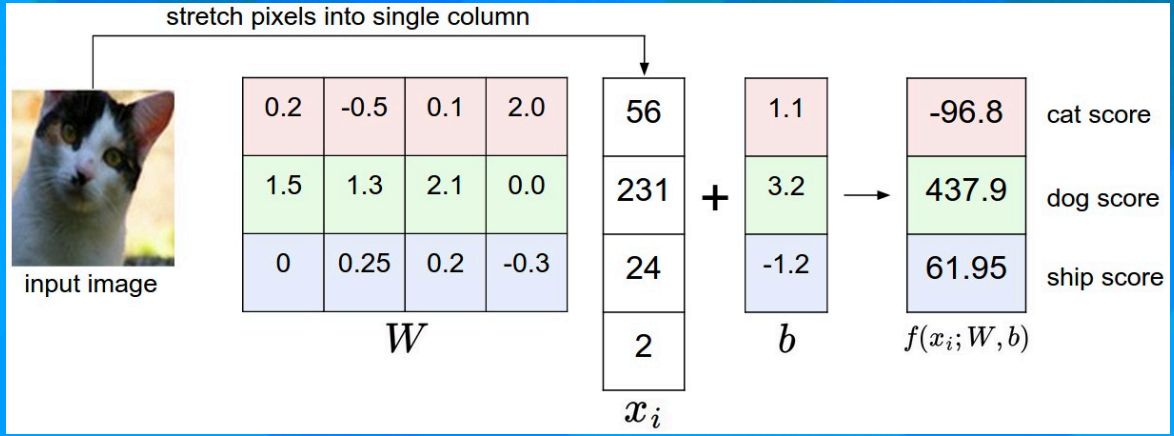
- A number of problems in computer science can't just be solved with existing relationships, (linear, exponential, etc.)
- Deep-learning models help estimate problems where the input and output are known, but the relationship is not.
- The **CIFAR-10 dataset**, for example, consists of 60,000 images of 10 "classes" of objects for classification.



https://www.cs.toronto.edu/~kriz/cifar.html

Implementation - the AI “black box”

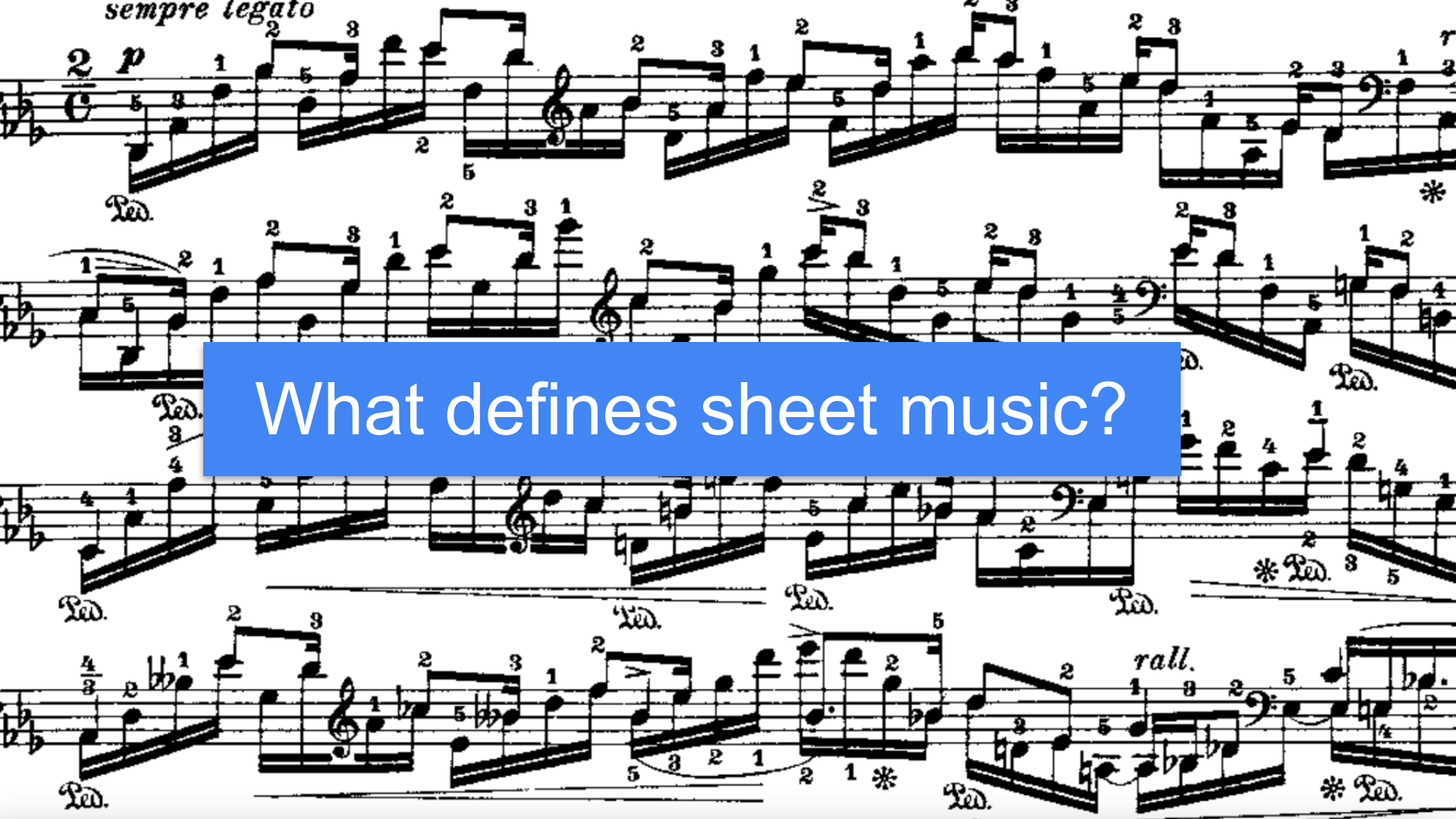
- How do deep-learning models make choices?
  - Answer: ***they don’t!***
  - The “black box” of AI works based on *statistical frequency*, where a pattern among data, or a **feature**, being linked to specific result (word/phrase, image, another pattern, etc.) becomes the “choice” that the deep-learning model makes.



https://courses.grainger.illinois.edu/cs440/fa2022/lectures/linear.html

Setup

- Problem: Can we *automate* the conversion from audio to sheet music?
- Goals:
  - Create a dataset of well-rounded MIDI files.
  - Create a web app for easily recording and uploading audio files to a database.
  - Collect several recordings per MIDI file.
  - Publish open-source database!
  - Running training/testing schemes on database *(in the future)*
  - Implement database as a deep-learning model *(in the future)*



What defines sheet music?

Varieties

- Elements which should be recognized by a deep-learning model:
  - Pitches:
    - Frequencies to MIDI note numbers
    - Enharmonics: model should understand that pitches have several note names, but it doesn't need to be accurate about which one to pick.
    - Accidentals: unless a key signature isn't given, notes should be reflective of their key signature, (proper use of sharps/flats).
  - Rhythms:
    - Lengths of notes with respect to the notes around them.
    - Have to include ties! (between notes, measures, etc.)
  - Dynamics:
    - Volume can be determined from recordings and scaled to pp -> ff.

Constraints

- Elements which are provided, given by the user, or aren't obvious enough:
  - Time Signature:
    - There exists an infinite number of ways to write 4/4, 3/4, etc.
  - Key Signature:
    - Maybe for some music, but not for avant-garde/modernist works!
  - Score type:
    - Provided: monophonic (single staff/voice)
  - Tempo:
    - Provided: slow, medium, fast tempos.
  - "Inflections":
    - Rhythmic: grace notes, mordents, trills, etc.
    - Pitches: microtones, poor tuning, etc.

The Web App!

Data gathering

- Funding received from Association for Computing Machinery at UIUC.
  - Used for gift cards, which help incentivized students to participate.
- Contacted vocal faculty, choral groups, School of Music students.
- Data collection is actively going on!

Results

- Data collection has been a success! ⭐
- Currently, not enough data to construct an accurate learning model.
- Data collection will persist throughout Summer-Fall 2023.
- Additional step to take: validate every recording according to sheet music.
- Results and database to be published open-source through ACM@UIUC, Github on our personal pages.

Recap

- Problem to solve
- Introduction to Artificial Intelligence, Deep-Learning Models
- The Experiment
  - Setup
  - Gathering data
  - Results

Thank you!