AI Data

can code communities

Review Questions

- What is natural language?
- What are some natural language processing tasks and what do they do?

Take Home Exercise Review

Have students locate and explain the functionality of all the user feedback features in Chat GPT that can help optimize the model.

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are inspired by the human visual system. They are particularly effective for tasks involving images, such as object recognition and classification.

At their core, CNNs consist of layers of neurons that learn to identify patterns within images. The key component of a CNN is the convolutional layer, where small filters slide across the input image, extracting features such as edges, textures, and shapes. These features are then passed through additional layers, to make predictions or classifications.

Computer Vision

There are many facets of computer vision (also known as image processing).

Image Classification
- Assigns a label or class to the entire image.

Object Detection
- Identifies and labels each and every noteworthy object in an image.

Text to Image
- Generates an image based on the text description.

Image to Text
- Describes the image. More detailed than image classification.

Image to Image
- Edits or generates a new image based on the image input.

Image Processing Models

Dall E 3 is OpenAI's most recent image processing model. It is mainly used for text to image generation, but can also do image to image generation.

MidJourney is another popular image processing model.

Computer Vision (Continued)

Video processing falls under computer vision due to the fact that they are both using image processing, but with more complexity.

Image to Video

- Generates a video based on the image input.

Text to Video

- Generates a video based on the text description.

Video Classification

- Assigns a label or class to the entire video.

Object Detection can be a task for both image and video.

Video Processing Models

- Sora is OpenAI’s text to video model. One of the first models that can generate video up to a minute long.

Audio Processing

AI audio processing is oddly similar to image processing due to the way the audio data is preprocessed.

- Audio has 3 dimensions
  - Amplitude (volume)
  - Frequency (pitch)
  - Time
- Displayed using spectrograms
  - Shows the amplitude of all the frequencies within the human hearing range.
  - Difference of amplitude is shown by a change of color so that we can display all the necessary information all within a 2D image.

Audio Processing Tasks

Text to speech

- Generates audio with a voice which says the text.

Text to audio

- Generates audio based on the text description (music, sfx, etc.)

Speech recognition

- Extracts the speech (as text) in the input audio.

Audio Processing Tasks

Audio to Audio

- Performing audio effects to the input audio (de-noising, stem splitting, etc.)

Audio Classification

- Assigns a label or class to the input audio.

Voice Activity Detection

- Determines whether speech is present or not in an audio input.
- Usually a preprocessing part for speech recognition.

Audio Processing Models

OpenAI has a text to speech model, called TTS, as well as a speech recognition model, called Whisper (Python usage).

Suno is a text to audio model that specializes in music generation.

Voice.ai is an audio to audio generator that has a couple of different effects.

Take Home Exercise

Use either Dall E 3 or Suno to generate an image or audio. Use prompt engineering techniques to get the most quality response.