



PCL2-Tutorial 09

Lucas Suomela, Rong Li, Tosca Peruzzi-Vieli, Isabelle Cretton

Welcome!



Today's Content

Recap- Code Optimization and Multimodal Data Processing

- Importance of Code Optimization
- Runtime and Memory Optimization

Multimodal Data Processing

- Importance and Applications
- Different Modalities: Audio, Image, Video

Recap: Hopefully You Learned... Exercise 8

1. Identifying and Fixing Bottlenecks

How? tools like cProfile and memory-profiler to spot the slow parts of your code.

Goal: Focus on optimizing the most time-consuming parts, use efficient algorithms, and reduce redundant work.

2. Understanding Hash Functions

Hash functions turn any input into a fixed-size value, making lookups in dictionaries and sets super-fast!

Why They Matter: They keep your data retrieval speedy

3. Implementing Reservoir Sampling

How to randomly sample k items from a giant dataset without using too much memory?

Solution:

- Start with a reservoir of size k .
- Fill it with the first k items.
- For each new item i , maybe replace a random item in the reservoir, with a twist of probability k/i .

4. Using Caching

- Save time by storing results of expensive computations for reuse.
- **LRU Cache:** It keeps the most recently used results handy and discards the old ones.

These tools and techniques make your code faster and more efficient. Keep experimenting and optimizing!

Multimodal Data Processing

Why Care About Multimodal Data?

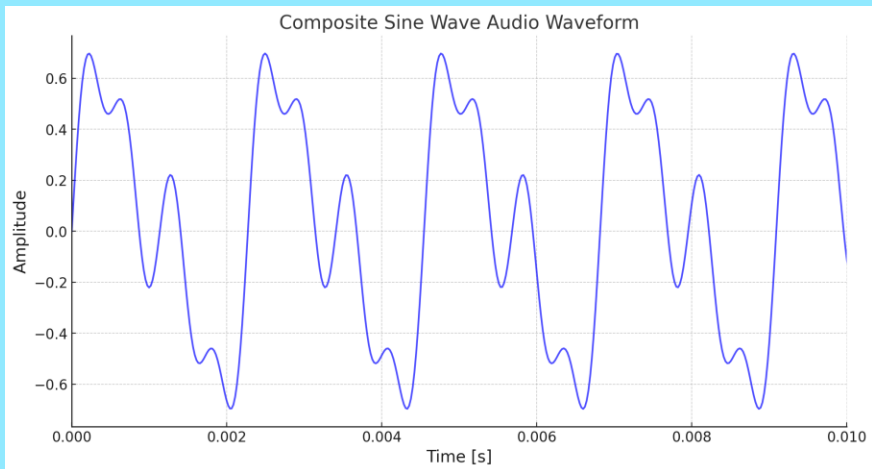
Human communication naturally involves multiple modalities beyond just text. Understanding and processing these different forms of data is crucial for creating comprehensive and effective computational models.

MODALITY	AS INPUT	AS OUTPUT
Audio	Speech Recognition	Text-to-Speech
Image	OCR (Optical Character Recognition)	Image Generation
Video	Sign Language Recognition	Sign Language Translation

Audio Data

Audio data is a sequence of sound pressure levels recorded over time, often represented as a series of numerical samples, we can read these files easily

We can visualise and look for patterns!



```
import numpy as np
import matplotlib.pyplot as plt
import soundfile as sf
```

```
# Read audio file
audio, sampling_rate = sf.read('audio_file.wav')
```

```
# Print basic information
print(f"Number of samples: {len(audio)}")
print(f"Sampling rate: {sampling_rate} Hz")
```

```
# Plot waveform
plt.figure(figsize=(10, 4))
plt.plot(np.arange(len(audio)) / sampling_rate, audio)
plt.title('Composite Sine Wave Audio Waveform')
plt.xlabel('Time [s]')
plt.ylabel('Amplitude')
plt.grid(True)
plt.show()
```

[Link](#) to a nice intro article
of Multimodal in ML

Images

OCR

OCR is the process of converting different types of documents, such as scanned paper documents, PDFs, or images captured by a digital camera, into editable and searchable data.

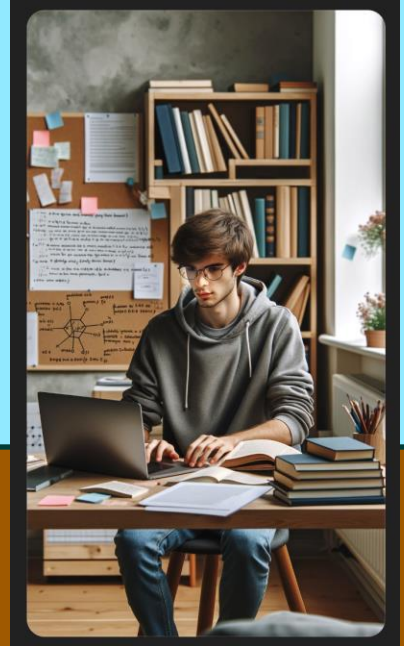
Optical Character
Recognition

Image Generation

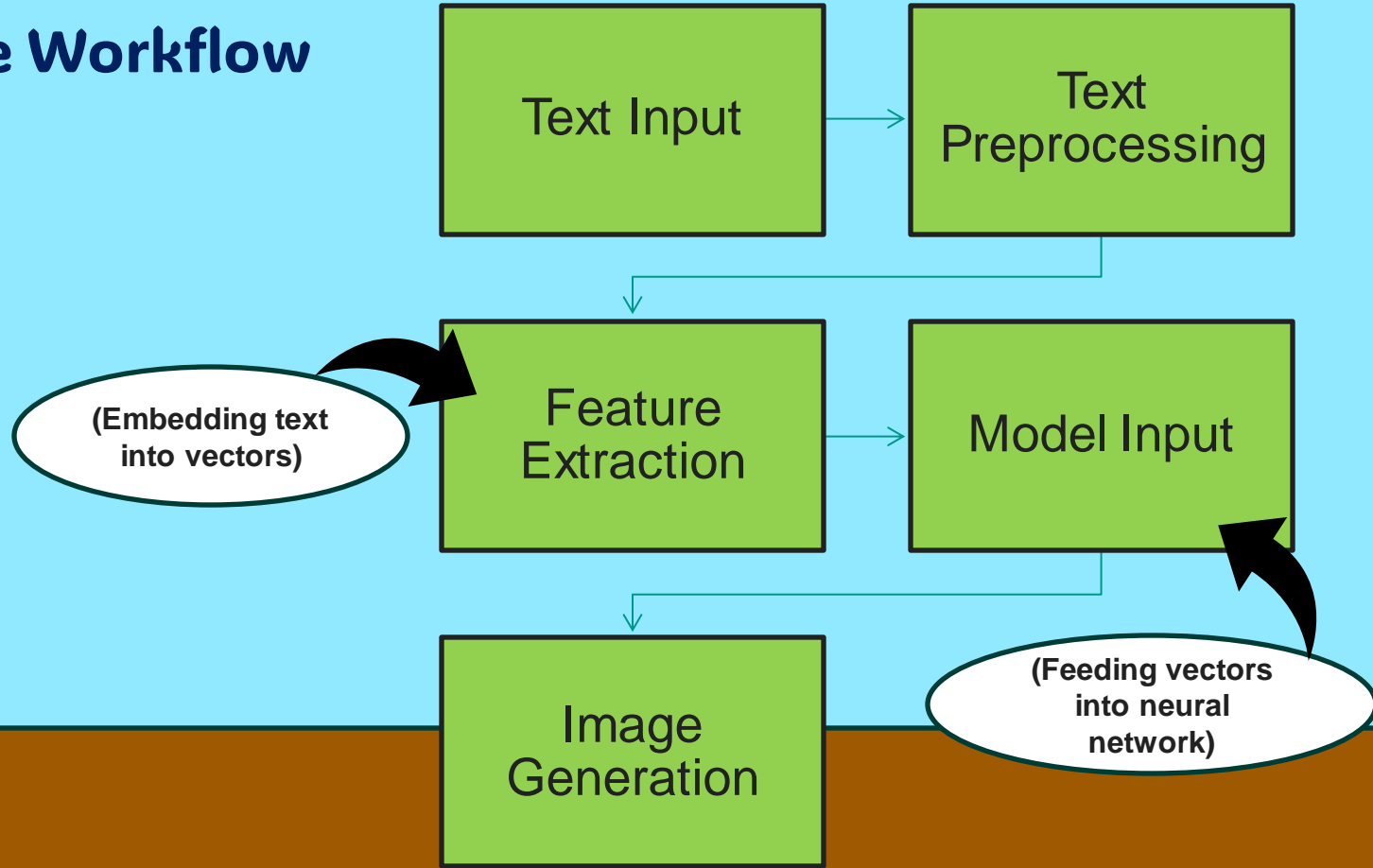
Image generation from text involves creating visual content based on textual descriptions.

Create an image of a CL student studying

It's literally me!

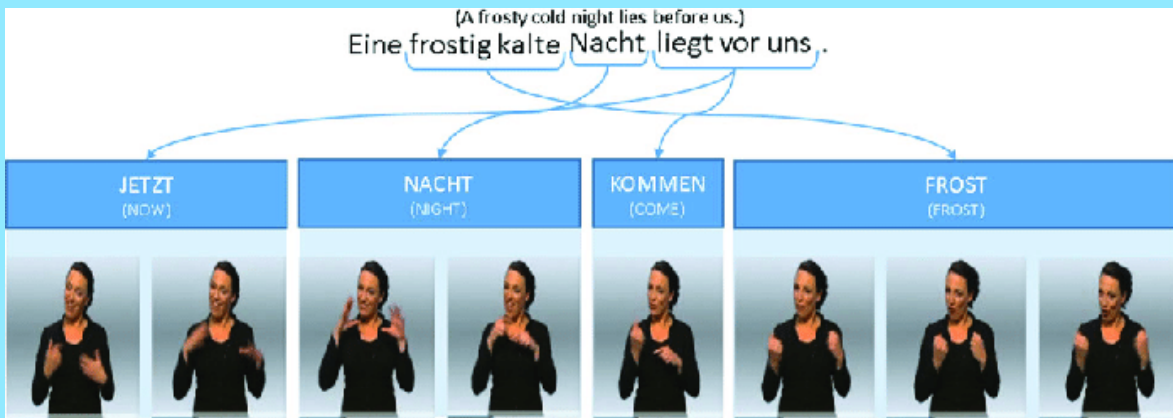


Text-to-Image Workflow (schematic)



Video

- Video data combines visual and auditory information, making it a rich source of multimodal data.
- In the context of NLP, video processing can enhance applications like sentiment analysis, lip-reading, sign-language translation and multimodal translation.



Components of Video Data:

- **Frames**: A sequence of images that make up the video.
- **Audio Track**: Sound that accompanies the video.
- **Frame Rate**: The number of frames displayed per second (fps).
- **Resolution**: The dimensions of each frame in pixels.

Now It's Your Turn...

To-Do

- ☐ Don't worry too much! Fill out this form for next week's lecture
- ☒ ~~Start panicking about the finals...~~
- ☐ Enjoy the upcoming weekend!