Sprocket:

A serverless video processing platform

Hanif and Pranav

Overview

- Intro to the Cloud Platform
- Why not video?
- Sprocket Framework
- Improvement in Performance
- Conclusion

What is the Cloud?

e.g. GCP, AWS, Azure

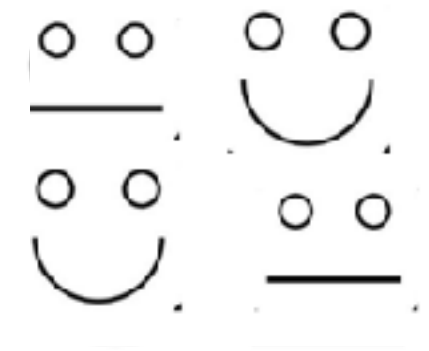
Not just for storage!

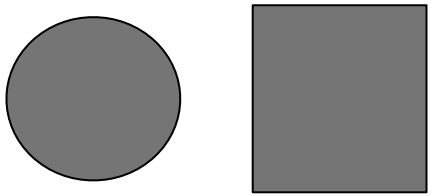
1. Hosting servers
2. Content Delivery Network

etc...

Machine Learning In The Cloud

- Provides pre-trained models
- Easily Trainable
- Makes it easy to incorporate ML into projects.





GOP 1 GOP 2

Why Not Videos?

- Notoriously hard:

1. Compression

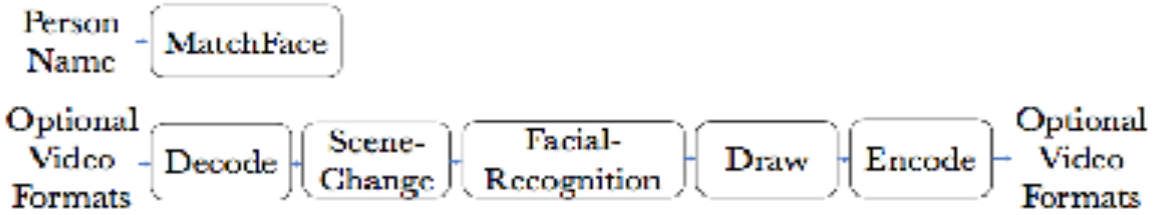
- Similar frames group together (GOP)

- One reference frame chosen

- Other frames stored as difference to reference

2. Huge Size (4K videos)

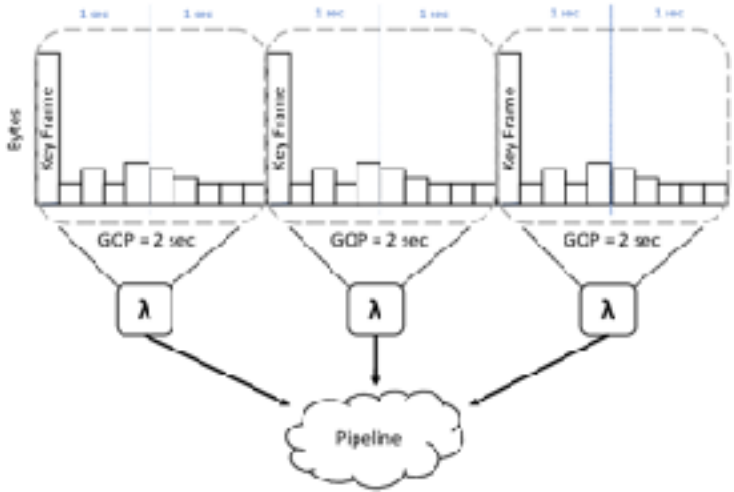
Sprocket Framework

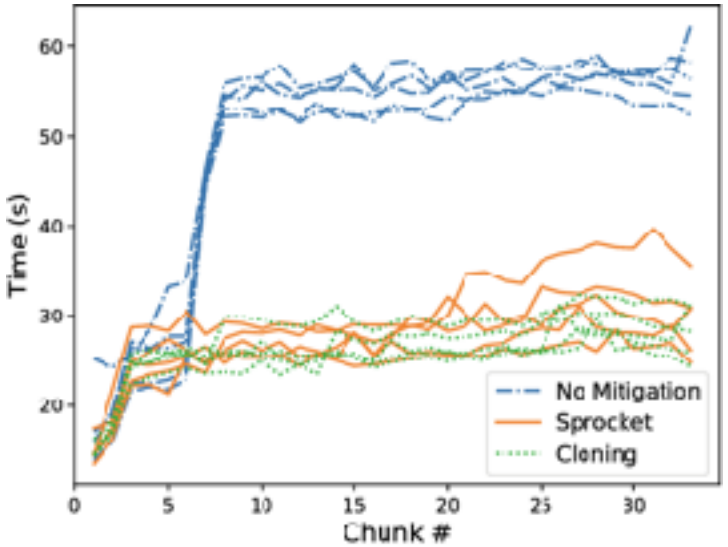


Leverages the cloud to quickly labels segment of the videos

Step 1 Decoding

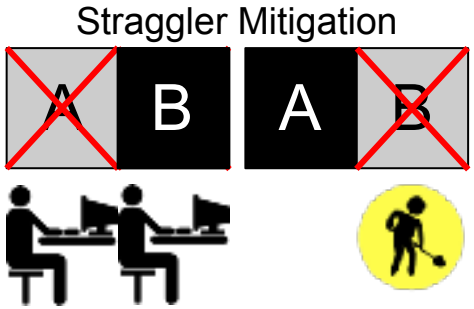
- Spawn thousands of cloud instances
- Parallely download and decode independent GOP
- Implement Straggler Mitigation





Straggler Mitigation

- One slow worker delays all later workers
- Proactive measures to minimize stragglers
- Sprocket Straggler Mitigation
  - Make GOP double the normal length
  - Send each GOP chunk to two workers
  - Each worker processes half of the chunk
  - First one to finish helps with other half
  - Less than 25% more data used

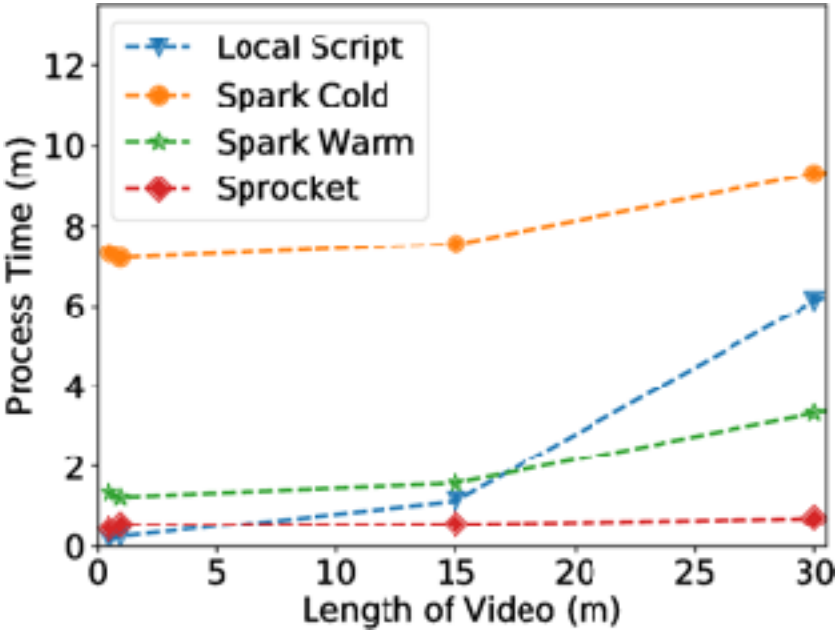


Step 2 Machine Learning

A. Extracting Semantically Meaningful Scenes
    - Sequence of frames must be preserved
    - Detecting the natural beginning and ending of scenes
    - Video is segmented into scenes
B. Recognizing Object That Users Want
    - Leverage on Cloud API, and detect frames where object exists
C. Stitching Up Scenes That Users Want
    - Filter out scenes where there are lots of frames without the object.

Is Sprocket better?

- Alternatives
  - More time to allocate workers
  - Fixed number of workers
  - More expensive
  - Servers
- Sprocket
  - Minimal startup delay
  - Variable number of workers
  - Cheaper
  - Serverless



Conclusion

- Serverless cloud environment
- High parallelism, low latency
- Less expensive
- Capable of advanced video processing