

ActiQ

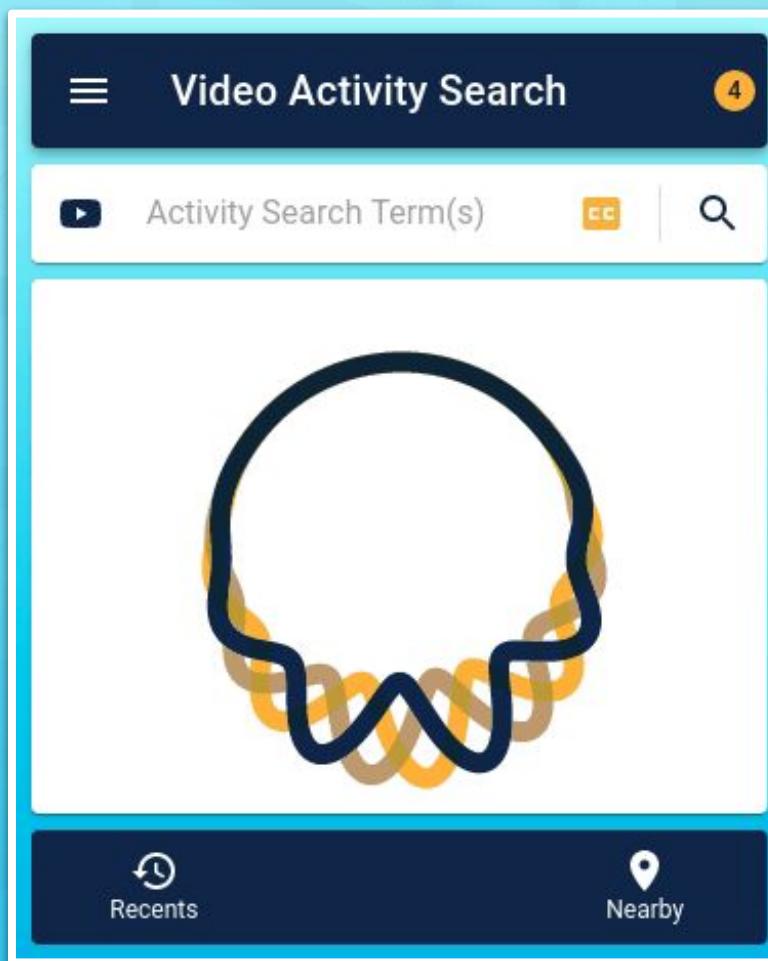
Video Activity Query



ActiQ | Story Time

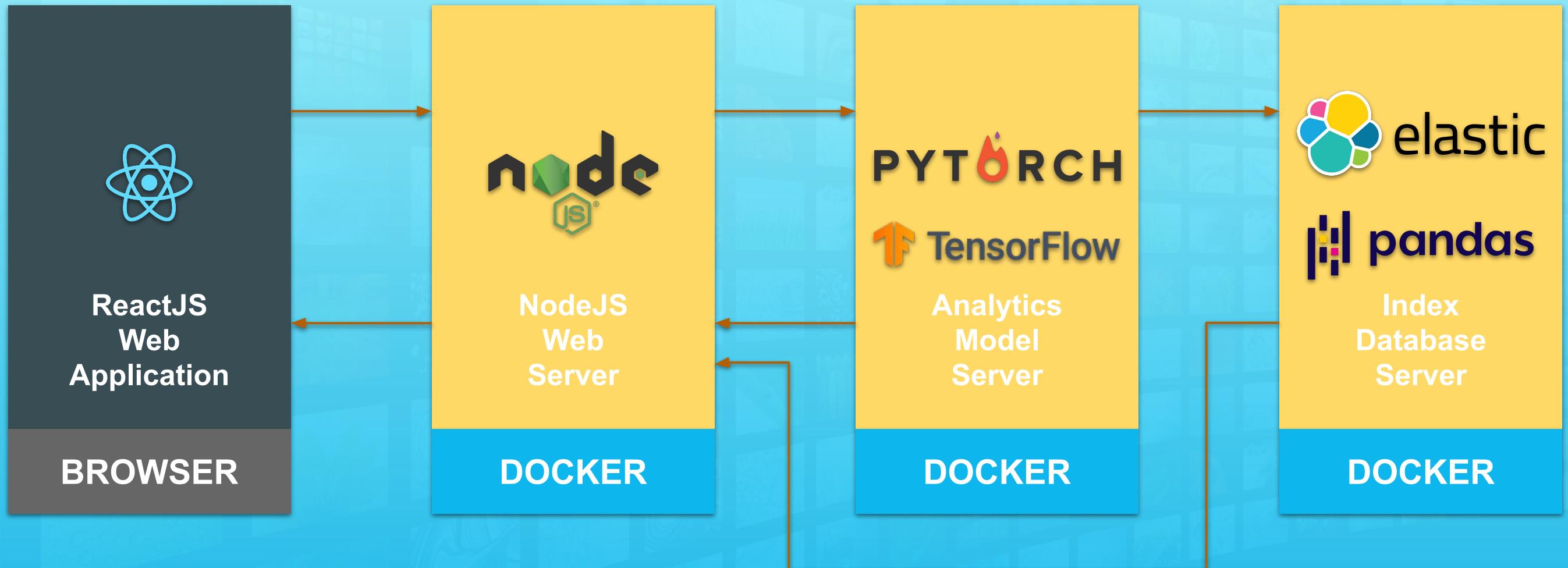


ActiQ | Live Demo of Application



<http://actiq.biz>

ActiQ | Product Architecture

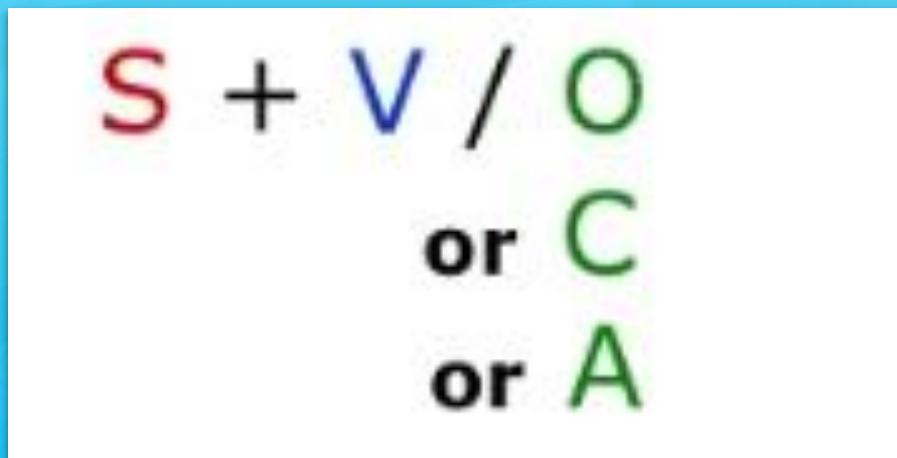


ActiQ | Video Processing Pipeline



ActiQ | Modeling Overview

- **Multiple DataSets Provide Different Levels of Understanding**
 - General Context
 - Actions and Postures
 - Human Object Interactions
- **Transcribed Captions Further Augments Accuracy**
 - Provides temporal speech localization

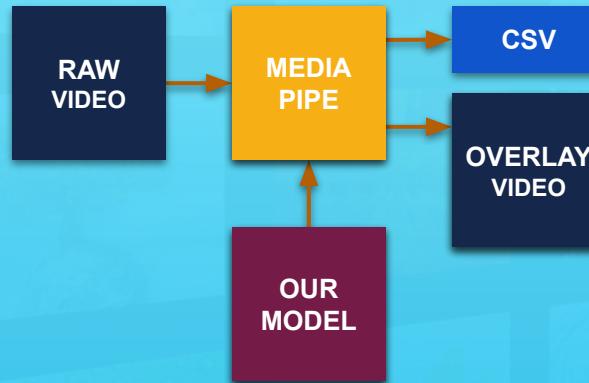


ActiQ | MediaPipe Architecture

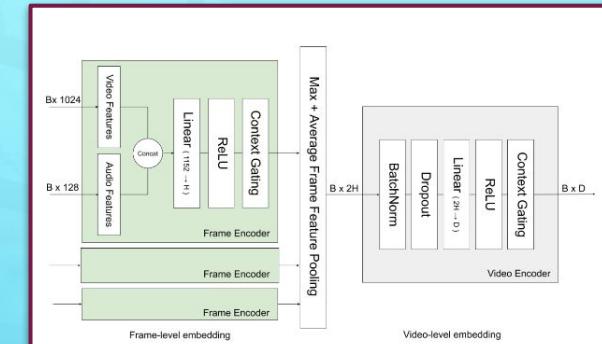
Google

How it works

- Combine Computer Vision + ML in Complex Workflows
- Uses Bazel Build System for Cross Language Graph Compilation



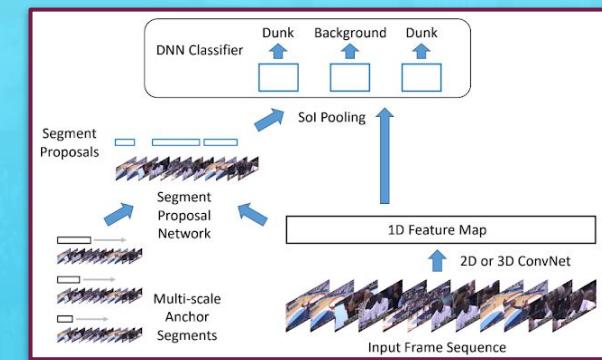
Context Gating



MediaPipe



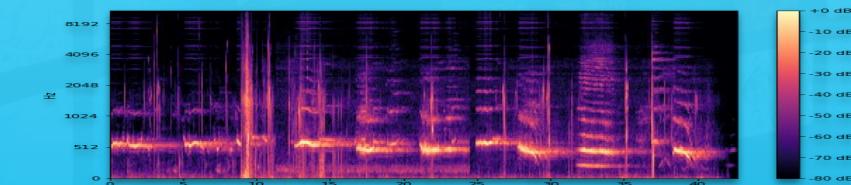
Deep Bag of Frames



Bazel



Audio Spectrogram

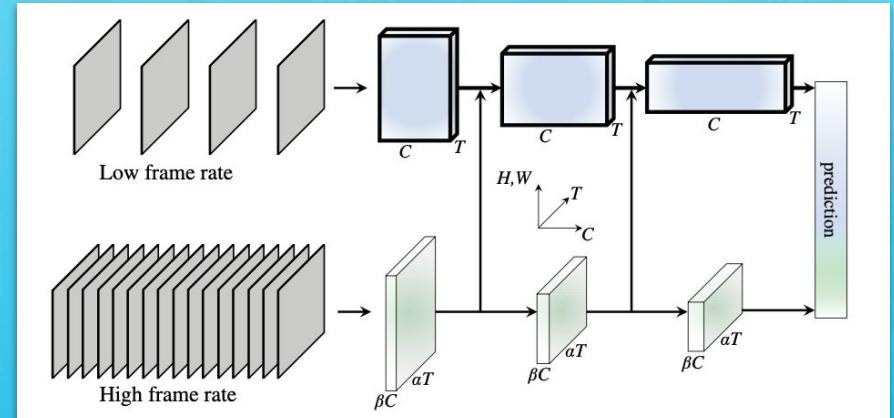
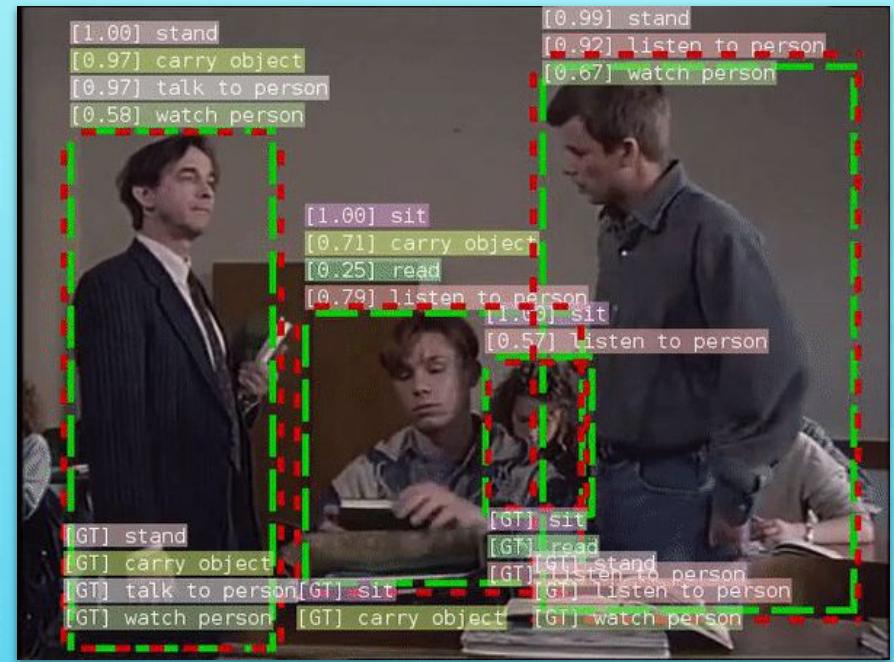


ActiQ | SlowFast Architecture

facebook research

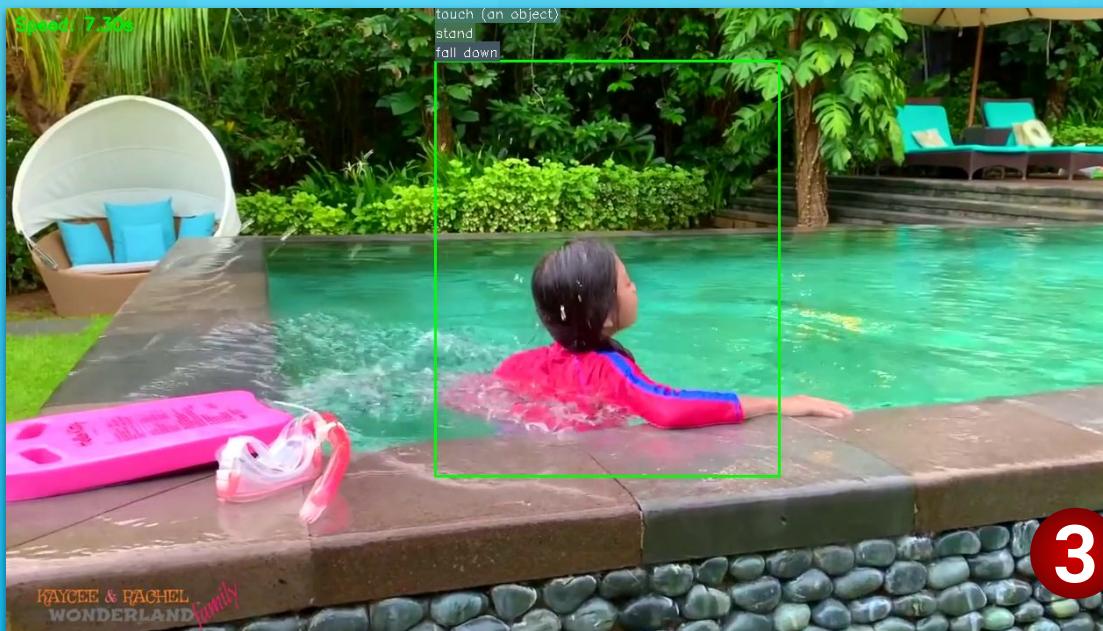
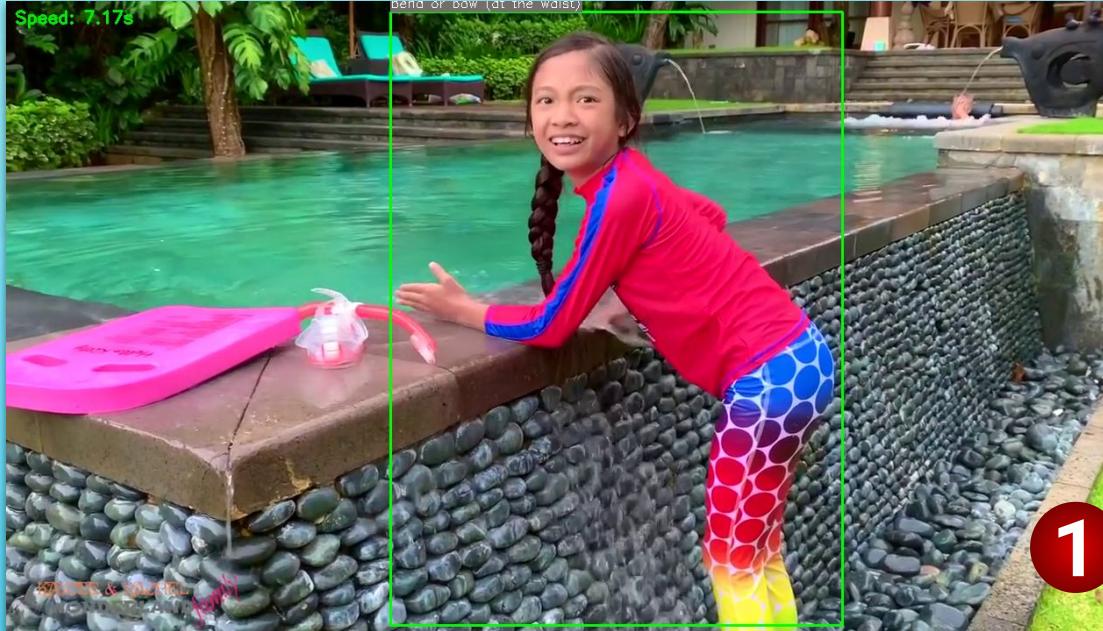
How it works

- Two pathways develop their own video-modeling expertise
- **Slow Pathway**
 - Low frame rate, slow refreshing speed
 - Captures spatial semantics
- **Fast Pathway**
 - High frame rate, fast refreshing speed
 - Captures motion at fine temporal resolution
- Pathways are connected laterally



ActiQ | Prediction Overview

"Girl Dives into Pool"



- Predictions for 1 long second video segments
 - 60 Frames
- Three predictions per segment

EXAMPLE

1. girl, bend or bow, pool
2. girl, jump or leap, pool
3. girl, stand, fall, pool
4. girl, swim, underwater

ActiQ | Prediction Samples



AFTERNOON
Routine



BEACH
ROUTINE



AFTERNOON
Routine

"taking a bath"

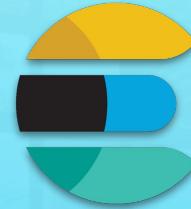
"swimming at beach"

"snorkeling underwater"

ActiQ | Search Architecture

How it works

- Using ElasticSearch for a highly scalable solution
- Create dynamic temporal buckets with Pandas
 - Action last more than one second!
 - Search 30 seconds, etc.
- Pre-Process
 - Create windows of 15 to 30 seconds and aggregate probabilities
 - Score each bucket using a threshold
- What we are searching?
 - Once we find the most probable bucket, we find the first instance of the action

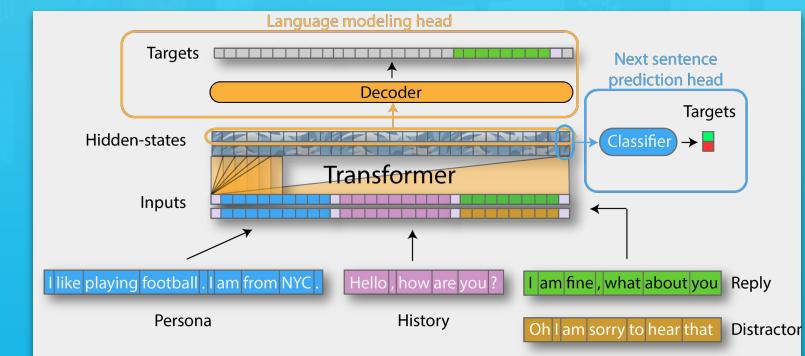


elasticsearch



ActiQ | Next Steps

- **Improve inference performance**
 - SlowFast currently at 0.2x realtime
- **Incorporate additional models**
 - Highly extensible architecture
- **Incorporate additional metadata**
 - GPS location
 - Camera Sensor / Lighting
- **Language model integration**
 - Make contextual sense of multiple labels
- **Build Camera App**
 - Perform inference in real-time on phone
 - Inject metadata stream into video container



ActiQ | Questions for the Team



Vinicio De Sola

- DBoF Model Training
- Search Algorithm Design
- ElasticSearch Integration



Pri Nonis

- Product Architecture/Engineering
- Video Processing Pipeline
- Responsive Web Application



Connor Stern

- Dataset Research, Exploratory Data Analysis
- SlowFast Pipeline and Model Training
- Narrative Synthesis