Innovation Drivers and Data Labeling Trends for Video Machine Learning
Your Presenters

Chip Ray
Chief Technology Officer

Kathleen Walch
Principal Analyst
Q: What video-specific trends are you seeing in data labeling across the industry?
Trends in Data Labeling

- Movement from easier data labeling tasks to more complex and sophisticated tasks as use cases continue to grow
- In particular, a growing increase for video annotation including:
  - Image classification in large taxonomies
  - Tracking across multiple images / frames
  - Instance segmentation
  - Object tracking with persistence even when missing
- Also seeing growth in areas related to sensor fusion such as
  - Complex entity segmentation
  - Determining intent
Growth of Data Types in Data Labeling

Data Labeling Market Forecast (2019-2027) [$ Billion]
Source: Cognilytica

Total 2022 Estimated Market by Data Type
Source: Cognilytica © 2022

Structured
1.2%
Sensor Fusion
9.8%
Audio
11.7%

Video
22.9%

Image
31.8%
Q: What use cases are you seeing in the industry that utilize video-based ML?
Use Case Diversity
Consumer Interactions Annotation Use Cases

- Self-checkout fraud monitoring
- Retail shopper behavior monitoring
- Television viewer behavior monitoring
- Sports-focused pose estimation and monitoring

34% Retailers expecting a “major increase” in use of self-service kiosks.
Security Video Annotation Use Cases

● Home security monitoring
● Premise security
● Drone-based traffic monitoring
● Hyperspectral drone-based security monitoring
Video Annotation in Sports
Transportation Video Annotation Use Cases

- Driver awareness monitoring
- Autonomous last-mile package delivery

Consumer Reports adds 2 points to a vehicle’s overall score if it includes a driver awareness monitoring system.
Healthcare Video Annotation Use Cases

- Robotic surgery
- Precision medicine and treatment protocols
- Radiology and diagnostics
- Operating theater monitoring
- Sports medicine and rehab

86% Healthcare companies utilizing some form of AI.
Industrial Video Annotation Use Cases

- Autonomous pipe inspection
- Construction site safety monitoring
- Factory safety monitoring
- Refuse collection vehicle monitoring

2020 Reduction in building costs by utilizing Robotics, AI, and the Internet of Things in construction.
Video Annotation in Retail: Frictionless Shopping
Factors driving Growth of Video Applications

- **Economic factors**
  - Companies are looking to gain more insights with their unstructured video data

- **Workforce factors**
  - Companies are using video applications to augment humans. Retail is heavily using video applications and examples include: Computer vision to deter theft and reduce losses incurred from theft and product loss, enhance self-checkout, analyze in-store customer behavior to identify product interest, and help keep track of inventory.

- **More sophisticated/educated workforce around AI**
  - As AI continues to mature we’re seeing an increasing number of companies companies using AI based approaches and sophistication and increased level of education/knowledge around AI applications

- **More complex use cases**
  - Seeing increased uses of Inference at the edge
Innovation Drivers

- Explosion of papers and models
- Inference at the edge
- Data curation

Explosion of Papers & Models

**NUMBER of AI-RELATED PUBLICATIONS on arXiv by FIELD of STUDY 2015-20**

Source: arXiv, 2020 | Chart: 2021 AI Index Report

- 11,011 cs.CV
- 11,098 cs.LG
- 5,573 cs.CL
- 2,571 cs.RO
- 1,923 cs.AI
- 1,818 stat.ML
- 743 cs.NE

Chart: Artificial Intelligence Index Report, 2021, Pg 8
Inference At The Edge

Benefits
- Real-time insights
- Increased privacy
- High availability

Drivers
- Maturation of neural networks
- Advances in compute infrastructure
- Adoption of IoT devices
Data Curation for Video
Advantages of Video over Image-Based Apps

1. Moving beyond just Bounding boxes and polygons for more complex use cases
2. Moving beyond just Image classification to apply object tracking
3. Moving beyond just Keypoint tagging to Scenario, Gesture, & Action Classification
Video Advantages

1. Context - identify objects, actions, attributes, relationships

2. Temporal understanding - action recognition, scene understanding

3. Object persistence - aggregate activity over time, track states
Human Pose Estimation
Human Pose Estimation

COCO (17)  OpenPose (18)  BlazePose (33)  COCO WholeBody (133)
Human Pose Estimation

Performed exercises counter based on detected body pose. Left: Squats; Right: Push-Ups

Challenges to video annotation

- Objects are continuously moving in the visual file
- Frame-by-Frame Annotation of a Video is Complicated
- Finding the workforce required to generate large amounts of training data
- Finding a vendor to label videos at the accuracy needed/expected
Video Is Complex—
It’s More Than Just a Collection of Images

- Entity relationships and persistence
- Situational relationships
- Sequence
- Past and future context
- Time based annotation (temporal tagging)
Video Annotation: A Unique Challenge

1. Media Challenges
2. Annotation Efficiency Challenges
3. Workforce Challenges
Video Annotation Challenges: Media

Media Challenges

- High resolution
- Long video lengths
- Proprietary media formats
- Compression
- Video size
- Frame rate
- Lighting variations
- Frame loss or corruption
Video Annotation Challenges: Annotation Efficiency

Annotation Efficiency Challenges

- Annotation density
- Erratic subject movements
- Background noise
- Complex entity relationships
- High precision requirements
- High quality requirements
Video Annotation Challenges: Workforce

**Workforce Challenges**

- Sensitivity of data
- Training and re-training
- Specialization
- Certification
- Content concerns
- Indeterminate accuracy
- Poor feedback
- Edge/Corner cases
Advanced Trends

- Multi-camera
- Multi-modal
- Advanced HPE

Traffic-Aware Multi-Camera Tracking of Vehicles Based on ReID and Camera Link Model, Hsu, Wang, Hwang, 2020
HPE Body Modeling Types

Type Overview

- **Kinematic** - advantage of flexible graph-representation, limited in representing texture and shape information.

- **Planar** - used to represent the shape and appearance of a human body

- **Volumetric** - used for human reconstruction
### What’s Lagging?

- Framework support
- Quality video datasets
- Standardization

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>COCO</strong> <em>(Microsoft Common Objects in Context)</em></td>
<td>The MS COCO dataset is a large-scale object detection, segmentation, key-point detection, and captioning dataset. The dataset consists of...</td>
</tr>
<tr>
<td><strong>UCF101</strong> <em>(UCF101 Human Actions dataset)</em></td>
<td>UCF101 dataset is an extension of UCF50 and consists of 13,320 video clips, which are classified into 101 categories. These 101 categories can be classified into 5 types (Body motion,...</td>
</tr>
<tr>
<td><strong>Kinetics</strong> <em>(Kinetics Human Action Video Dataset)</em></td>
<td>The Kinetics dataset is a large-scale, high-quality dataset for human action recognition in videos. The dataset consists of around 500,000 video clips covering 600 human action...</td>
</tr>
<tr>
<td><strong>HMDB51</strong></td>
<td>The HMDB51 dataset is a large collection of realistic videos from various sources, including movies and web videos. The dataset is composed of 6,849 video clips from 51 action catego...</td>
</tr>
</tbody>
</table>
Alegion is the Data Labeling Solution for Enterprise-Grade AI

- Delivers accurately-annotated, model-ready data to train and validate ML models
- Provides both the platform and workforce to operate with quality at scale
- Processes structured and unstructured data including image, video, text, audio, Lidar
- Leads the industry in streaming, high-resolution, high-density video annotation
- Supports complex use cases, powered by robust ontologies and entity relationships
Q&A Session

Ask Us Anything!
Thank You!

Chip Ray
Chief Technology Officer
cray@alegion.com

Get A DEMO
alegion.com/demo

150 FREE Annotation Hours
alegion.com/control