Building a Video Annotation Platform

By Chip Ray, CTO Alegion

Video annotation is exploding. The work is challenging and time consuming. ML teams need a superior training data platform to ensure project success.

Over the past year Alegion has taken on the challenge of building the best possible video annotation experience. Having assessed the tooling available on the market, we knew there just had to be a better way.

We recently launched Alegion Control, our next generation annotation platform built explicitly to simplify and accelerate video annotation. We think it's pretty great! It boasts an amazing U/X with unmatched support for video length, annotation density, and ontology complexity without sacrificing performance.

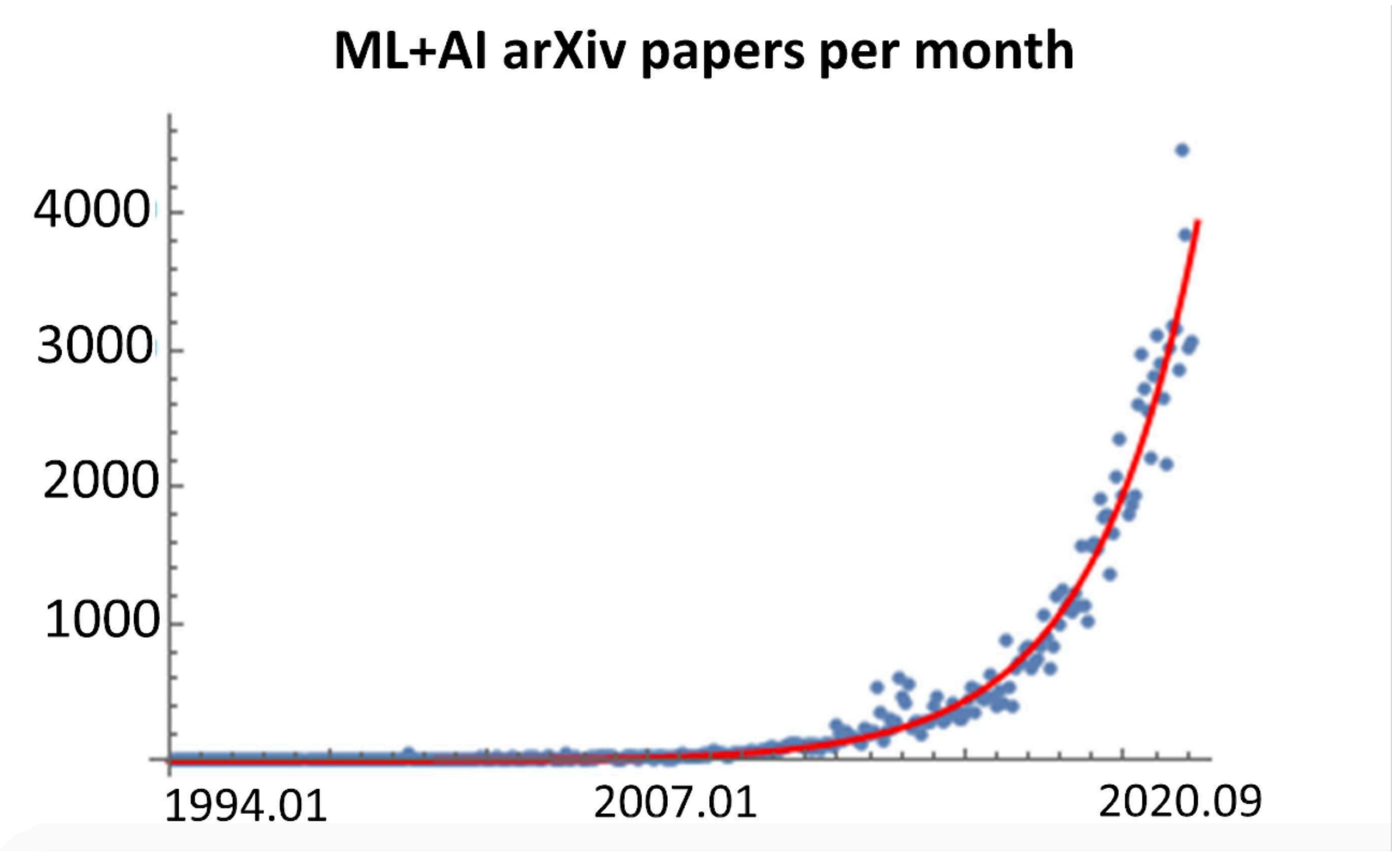
This is part 1 of a 3 part series that examines the challenges of building a super-responsive user experience for video annotation (VA), one that must operate across numerous annotators with potentially unstable internet connections, and scale on the back-end.

Part 1 Outline:

- Why focus on video annotation?
- Key differences between image annotation and video annotation
- What exactly is so hard about video as a data type?

Why focus on video annotation?

Major advances in computer vision (CV) are common and frequent compared to other fields; research is growing exponentially. Video is clearly the fastest growing data type in the CV space. As an exercise, scroll through papers on "Arxiv Sanity" (great name) and look at the release dates on the papers (the link defaults a search for "video"). It's hard to comprehend the variety of use cases that data scientists have found for video analytics and processing.



Mario Krenn: @MarioKrenn6240

So we dove into what those challenges were, and as a follow-up to the posts on our Design thinking for Alegion Control, we thought we'd share some of the hard problems we've had to solve on the engineering side of the house as well.

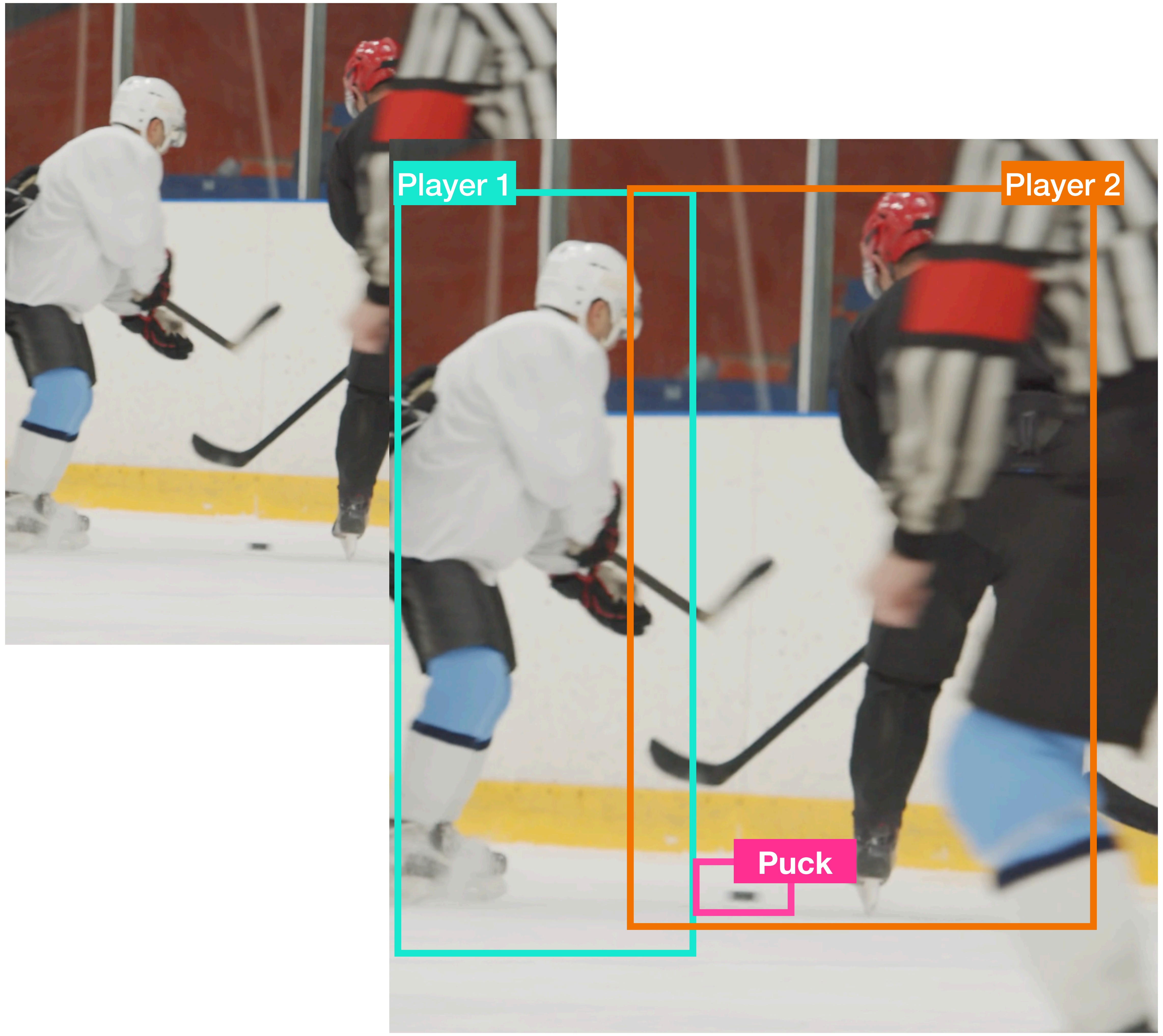
Key differences between image annotation and video annotation

Most video annotation solutions were built upon image annotation solutions. Early implementations processed videos as a sequence of images and used image annotation techniques. An incoming video was expanded to a sequence of frames and distributed, usually in parallel, to be annotated. Afterwards, the annotations were zipped back up based on the frame number. This approach worked... until it didn't.

Video brings temporal context to build semantic scene understanding

In recent years, we've seen perception models move from object detection to object tracking. This involves identifying individual instances of objects and giving them a unique identifier that remains constant over time -- `car_7`, `player_3`, `hotdog_18` -- what we call "entity persistence". You can imagine the challenges around "localizing", drawing a bounding box, or placing a keypoint, all of the players in a 1000-frame hockey game clip. Afterwards, because there were twenty annotators working in parallel, you'd have to figure out that Chip's `player_3` is the same as Kenneth’s `player_17` … for every frame. Sounds like the worst game of Pass the Parcel ever.

Video also gives temporal context to identify events (periods of time) in which an object is in different states. This information is needed to build scene understanding and event recognition solutions.



What exactly is so hard about video as a data type?

There are so many complexities to video annotation that we had to break this series into 3 parts, focusing on the different classes of problems we had to work through. Part 2 will look at the scalability challenges required to support 100k+ annotations in a single video and the back-end challenges of supporting 1K+ simultaneous annotators. Part 3 will focus on playback and real-time streaming, i.e. optimizing browser performance.

Challenges of Video:

- **More data**
- **Video variables**
- **Number of annotations**

More data

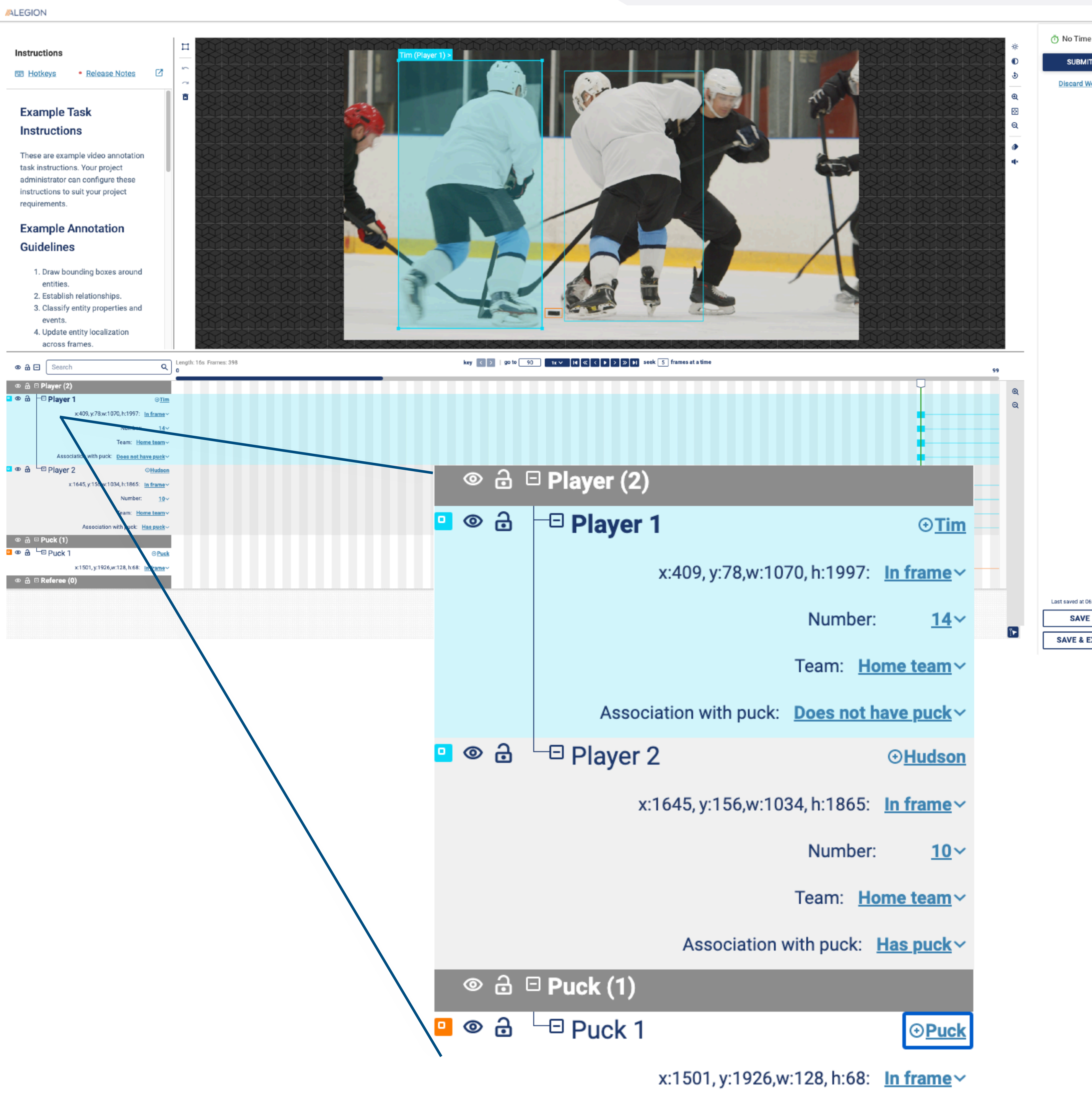
This one is fairly simple. Video assets are simply larger and more information dense than images. Copying, processing, viewing, and storing all become much more expensive operations. It requires a ton of work to make a video annotation tool feel as responsive as an image annotation tool.

Other video annotation platforms store all the frame data in the browser. Consequently users would experience their browsers crashing and lose any unsaved work. This was not acceptable to us and more importantly our customers. We will get into this more in part 3 of this series.

Number of annotations

As we have outlined, the most difficult part of video annotation is the amount of labeled data customers need. Typically, every single frame of the video is labeled with multiple localizations. Most customer videos are thousands of frames in length, so memory and performance problems are easy to encounter.

Imagine building a perception model based on analyzing hockey games. A two minute hockey video captured at 60 frames per second could easily have 100K+ annotations. Players need to be tracked, each with a localization (bounding box, keypoint, etc.). Each player has multiple labels, some that are static (player number, team), some that change over time (in-frame/out-of-frame, offense/defense). The puck's relationship attribute needs to be continually updated to reflect which player is in current possession of it. Additionally, scene classifications are captured in order to provide semantic richness. For example, these scene classifications can feed features that detect the state of play (regular play vs. time-out vs. foul).



Example Task Instructions

These are example video annotation task instructions. Your project administrator can configure these instructions to suit your project requirements.

Example Annotation Guidelines

1. Draw bounding boxes around entities.
2. Establish relationships.
3. Classify entity properties and events.
4. Update entity localization across frames.

Player (2)

Player 1 ⊕Tim
x:409, y:78,w:1070, h:1997: In frame
Number: 14
Team: Home team
Association with puck: Does not have puck

Player 2 ⊕Hudson
x:1645, y:156,w:1034, h:1865: In frame
Number: 10
Team: Home team
Association with puck: Has puck

Puck (1)

Puck 1 ⊕Puck
x:1501, y:1926,w:128, h:68: In frame

Video variables

There are many, many things that can affect the quality of annotations deriving from the quality and characteristics of the video to be annotated. Because annotations need to be accurate, it's paramount to ensure that we have a robust video processing pipeline that accounts for all of the variables without adding unnecessary burdens on our customers.

In part 3 we will discuss:

- The expansiveness of the MPEG spec
- Variability of formats, codecs, and encoder/decoders
- The scourge of RTSP (and other streaming encoders)
- Frame rates woes
- Compression artifacts and encoding errors
- The contradictory goals for encoders and decoders

Conclusion

To conclude part 1 of this series, with rapid growth in the volume of video collection and storage, the demand for video annotation is on the rise, and will only continue to grow. Alegion has researched and developed solutions to overcome the complexities of video annotation including supporting the sheer size of the data to be worked and the number of annotations needed, as well as delivering the contextual details that video inherently provides. Look out for part 2 of this series where we will discuss the scalability challenges required to support 100K+ annotations in a single video and the back-end challenges of supporting thousands of simultaneous annotators. Part 3 will focus on the user experience of our platform including playback and real-time streaming, i.e. optimizing browser performance.

About Alegion:

Alegion, based in Austin, TX, is the training data solution for enterprise-grade machine learning, enabling efficient and accurate annotation of video, images, and text. Alegion’s platform and services accelerate time to value, empowering the development of highly accurate machine learning models for Fortune 100 companies, systems integrators, business process outsourcers, and AI-driven organizations worldwide.

Want to learn more about Alegion's self service annotation platform?

Reach out to solutions@alegion.com

Follow us on

//ALEGION