## ///ALEGION

# Putting Video Annotation Solutions To The Test Alegion

unner

### How Does Your Video Annotation System Stack Up?

When we released our new video annotation (VA) solution last fall (available via <u>Alegion Control</u> and within <u>our managed platform</u>), we decided to run a couple of tests to see how its efficiency and performance measured up to the leading open source and commercial offerings.

In addition to enhanced support for long videos with dense annotations, Alegion's latest VA system features an entirely new user experience, refined over two years of work with complex, large scale video annotation projects.

We were particularly interested in assessing how well this new user experience reduced task time and increased annotation quality. We chose a couple of key competitors, built a moderately complex use case, and conducted some unscientific but very enlightening time trials.

Here's how we did it, and what we learned:



### What We Tested

### **The Competition**

In addition to Alegion Video Annotation, we tested the open source Computer Vision Annotation Tool (CVAT) and another well-known tier 1 commercial labeling platform (let's call them 'Brand X' for short). All three are self-service systems that allow a team to quickly upload assets and set up an ontology.

#### The Use Case

When we selected the video and defined the annotation guidelines, we made choices that modeled some of our real world projects, but that avoided bias toward any of the products. Because we were primarily interested in worker efficiency and time on task, rather than scalability, we chose to keep the clip short in length, but long enough to provide a good test of efficiency and playback performance. We had annotators already familiar with each annotation system complete 200 frames of a 4K video. Appropriately, the video is a clip of a sparring match between boxers.

#### The guidelines for annotators were simple:



Localize each boxer with a bounding box



Maintain 20 skeletal keypoints for each

Maintain a reference between the keypoints and each boxer



### What We Tested

While we could have post processed these relationships based on bounding box containment, this is not possible in many use cases. We wanted to include this important reference requirement as we see this request in many of our projects.

#### **The Metrics**

Time on task was our primary metric, and we also focused on quality assurance (QA) efficiency. Effort to verify labeling quality is a part of every Alegion project, but it can be a distinctly different experience from annotation.

The QA process was kept simple: take the time to verify the localization quality, ensure the accuracy of any classifications, and ensure the keypoints were associated with the correct boxer. No corrections were made in QA.



### What We Learned

The difference in time on task for workers between Alegion and the competition was substantial. Annotator notes recorded after each annotation session confirmed some key differences that drove efficiency gains in Alegion's favor.

### **Average Annotation Time**

Alegion	CVAT	'Brand X'
102 minutes	192 minutes	185 minutes

### Average Quality Assurance (QA) Time



### Annotation time was **45 to 47 percent faster** with Alegion Video Annotation.

QA time was **28 to 36 percent faster** with Alegion Video Annotation.



### **Key Findings**

In addition to the timing results, we used annotator interviews to gain further insight into the tooling capabilities that made the difference between Alegion's VA solution and competitors.

### A User Experience Purpose Built For Video

We found the largest time savings and reduction in fatigue came through reductions in clicks, mouse travel, and context switching. As the number of frames and annotations increases, small UX inefficiencies add up - particularly with cases that require dense annotation.

Some key drivers of efficiency are obvious. Well designed shortcut keys and viewing options are essential, and each product has its strengths in this area.

However, only Alegion has a user experience designed from the ground up for working with multiple entities over time. For example, the Alegion timeline view is the only one designed for working with multiple entities simultaneously. The ability to browse and edit multiple entities and their keyframes in one view, drastically reduces the number of clicks and mouse travel. Context switching and cognitive load are lessened when relationships between entities are always in view.





### **Key Findings**

Keyframe and context-sensitive classification tools allow annotators to perform most tasks inside the annotation window. Some of these capabilities delivered the greatest benefits in QA where verifying proper classification and associations between entities is the bulk of the work.

#### **A Real World Classification Model**

Associations between labeled entities is a common requirement of a wide variety of use cases. Sports, security, and retail commonly feature associations between persons, body parts, and objects like shopping items. In addition, the number of these relationships can be open ended in complex use cases.

Of the three platforms tested, Alegion was the only system with a built-in model for associations. Hierarchical relationships like skeletal points are also built into the Alegion user interface so that they can be defined and verified according to their logical structure. In the competitive systems we tested, we had to define arbitrary ID fields as lookups between entities. These felt like workarounds, but mostly they were error prone and time consuming because they required manual entry and updates across the frames of video. This inefficiency was compounded in QA when these field values had to be rechecked.

#### **Performant Playback and Annotation Synchronization**

Even with the modest number of frames in this test, Alegion's ability to smoothly stream 4K video and keep dense annotations in sync was a key factor that increased worker efficiency. In the annotation task and in QA, the precise application of classifications and localization quality is much faster to verify. With the use of the timeline scrubber and smooth playback, annotators can view localizations with speed and precision without compensating for lag time inaccuracy.

### Conclusion

Most video annotation systems are derivative of image annotation tools, and this leads to inefficiency and worker fatigue as annotation density and scale increases.

When we designed the Alegion Video Annotation system, we treated video as a first class data type. Video is a domain with a unique set of challenges, but also some well-understood and proven solutions. Using a system built specifically for video, combined with a rich classification model, gives workers a solution that excels as the complexity and scale of a video annotation project increases.

The Alegion video annotation system delivers high quality labeled data, reduces worker fatigue, and saves teams time and money.

Want to know more about Alegion's next generation video annotation capabilities and how they compare with your current solution? Contact us today at solutions@alegion.com



