

CV Case Study:
Complex Video
Annotation
For Loss Prevention

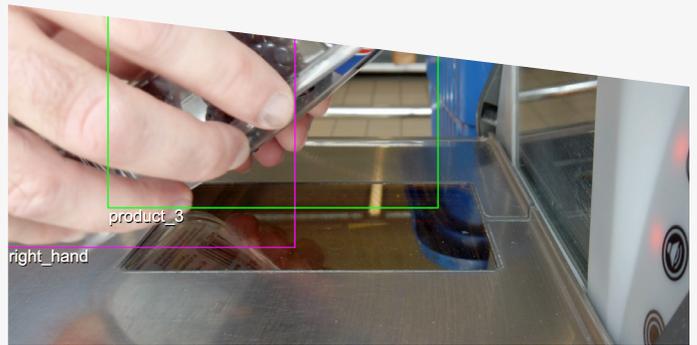
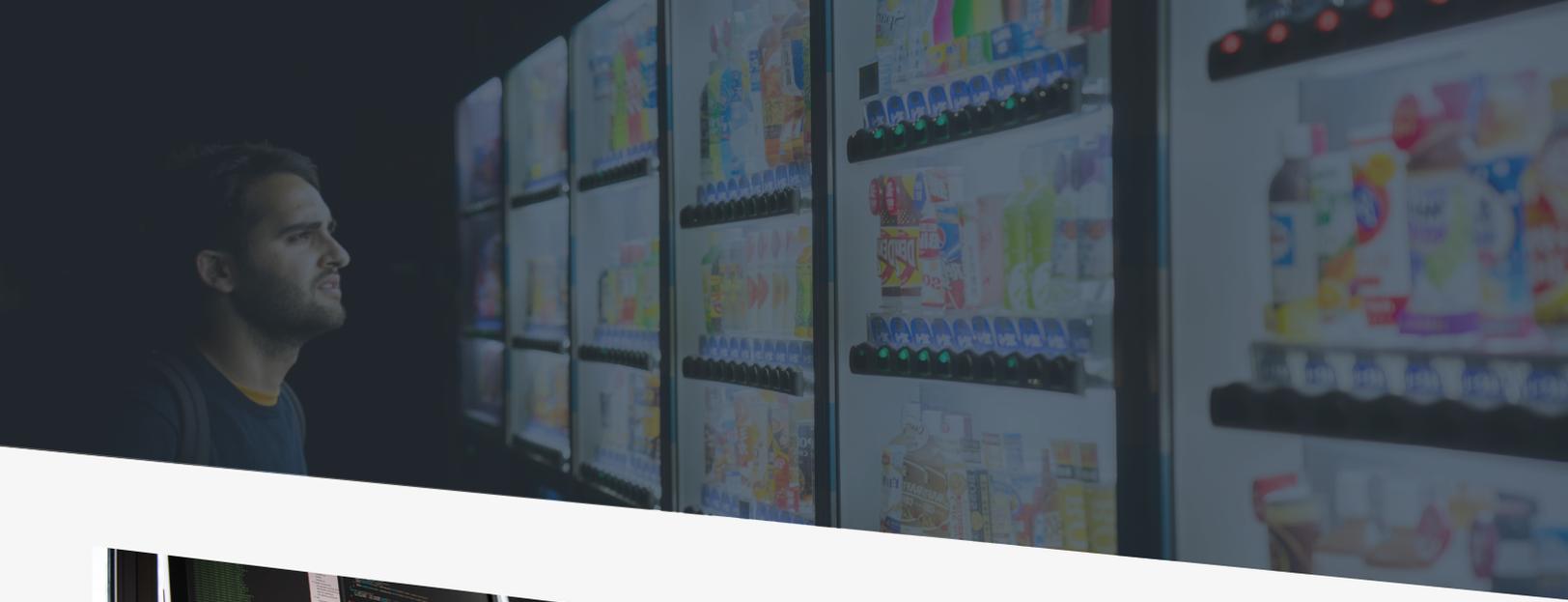


Project

A world leading retail company came to Alegion with a very complex video annotation use case for loss prevention. Loss of inventory due to shoplifting, employee theft, administrative errors, and fraud costs the retail industry nearly \$50 billion in 2018. According to the National Retail Security Survey, 36.5% of missing inventory was attributed to shoplifting and 33.2% was due to employee theft. According to the Loss Prevention Research Council, 58% of self-checkout shoplifters categorized stealing from a self-checkout machine as easy. With this being the case, the client’s goal was to develop a computer vision model that detected shoplifting at self-checkout.

Challenge

The client came to Alegion after finding that previous vendors could not meet their quality requirements. Each video being annotated varied in complexity and the annotation guidelines often had subtleties and non-intuitive directions to navigate. In addition, there were numerous classifications needed, resulting in an average of 29 judgements per frame across hundreds of thousands of frames in hours of video footage. A few examples include localizing with bounding boxes the left and right hands, hand scanners, heads, items once they’ve been picked up, the relationships between the items and the person, and the state of the hand and items if the hand is empty or holding an item.



Why ML:

With such a large volume of video and numerous annotations across hundreds of thousands of frames, utilizing integrated machine learning (ML) within the platform accelerated the annotation process while maintaining annotation quality. Object detection and proposal allowed Alegion to pre-label the head, hands, and hand scanner across each video, saving time across each task and speeding up delivery.

Approach:

Our customer success team met with the client in order to deeply understand the annotation guideline requirements and nuances of the use case. Due to the complexity and volume of judgements, a decision was made to utilize machine learning within the platform to pre-label some objects in the videos, with a focus on the head, hands, and hand scanners in each frame. Based on Alegion's expertise with similar use cases in the past, the platform was configured with a proven, optimized workflow and integrated machine learning to pre-label objects in the videos. The expertise of our team was critical for properly injecting ML into the labeling process without causing disruption that would affect quality. In addition, every part of the manual annotation workflow was optimized to ensure that ML pre-labeling was adding the proper efficiency gains. Our data science team was active throughout the process to monitor and improve the performance of ML pre-labeling.



Quality Management:

To fully measure and maximize the impact of ML pre-labeling for this specific use case, parallel workflows were created for the initial annotation work. One workflow leveraged ML pre-labeling with manual review and adjustment while the other had only manual annotations. The ML model for automated pre-labeling started at about 60% accuracy. Manual adjustments to the proposed annotations were then used to further train the model and improve its accuracy.

In addition to training the model, these parallel paths helped identify where automated pre-labeling provided the largest efficiency and quality gains. Continuous feedback from the client based on their model performance was also used to optimize the annotation workflows. As the pre-labeling model improved and the annotation workflow optimized, the parallel workflows were merged with confidence that both annotation throughput and quality were improved.

The annotation workflow also included a step for entity resolution. Entity resolution is used when

something identified in one frame is identified as a new entity or object in a later frame and the two instances need to be merged and annotated as the same entity. The incorrect identification can be due to automated ML pre-labeling or an incorrect human annotation. The Alegion platform includes Entity Resolution capabilities to quickly identify and merge misidentified entities throughout the entirety of a video. This step was used by quality control reviewers to correct both manual annotations from other annotators and ML pre-labeling proposals.

In addition to managing entity resolution, another element of ensuring quality output for such a complex task was the training and management of the annotation workforce. Annotators were selected based on experience with similar tasks and trained on specifics of the annotation guidelines, entity resolution, and the automated pre-labeling process. Throughout the process, a strong feedback loop between customer success and annotators was maintained to iterate and improve the entire annotation process.

Impact:

So far, we've annotated over one million judgements across 30,000 frames of video. During the early stages of automated pre-labeling, efficiency gains were moderate. Utilizing workflow configurations that managed the cognitive load of the workforce and by streamlining the annotation process in conjunction with ML pre-labeling, high quality standards were maintained, while achieving a 17% increase in efficiency of task completion after processing only one-third of the data. As the tasks continued, the accuracy of the ML pre-labeling continued to increase as well, ultimately realizing gains of up to 60% efficiency in task completion.

60% Efficiency Gains

- Manage cognitive load of the workforce
- Streamline the annotation process
- ML pre-labeling drove 60% gains in efficiency of task completion

5 Millions Judgments

- Each frame requires on average 29 judgements
- For example, for 30K frames of video over 1 million judgements are made

95% Accuracy

- 98-99% accuracy of object classification and bounding boxes
- 95% for subjective judgements (holding status, occlusion judgements, continuity of object IDs)