vintra

White Paper

Four Trends Shaping the Future of Video Surveillance Analytics

A revisited perspective from the co-founders of Vintra, a leader in AI-powered video analytics

By: Brent Boekestein and Ariel Amato, PhD

Introduction

Among all sensors, visual sensors like surveillance cameras are special because they can provide rich and versatile information. When we first published a white paper on emerging video analytics tech in 2017, it was considered advanced to do solid, deep learning-based object detection on typical surveillance footage. Since then, the combination of improving object detection, emerging classification capabilities, and new event detection technology now make it possible to reduce video review time by nearly 90%. For real-time alerting, the algorithms, with increasingly low false positive rates, are the new frontline security operator or analyst.

2021 IS SHAPING UP TO BE A SEMINAL YEAR IN THE ADVANCEMENT OF VIDEO ANALYTICS TECHNOLOGY

As advancements continue in object and event detection and classification, it's time to look at what's happening in 2021 and beyond. The next wave of AI-powered advancements in video analytics will focus on four things.

Condition Agnostic

Video analytics have long-struggled when environmental conditions are challenging. For example, if there is a rapid change in light conditions or weather conditions, traditional analytics have often produced high levels of false positives or reduced levels of true positives. An operator concerned with perimeter detection should not have to concern themselves with meaningful drops in detection accuracy each time a storm comes through. This is now changing due to a number of factors. First, the capabilities of the core object detectors continue to advance. Figure 1 below shows the rapid improvement in object detection accuracy across different open sourced detectors. Proprietary detectors, like those produced by Vintra, operate even more accurately across general and niche classes. As the detectors improve, the tech will work better on more types of video, during more hours of the day, to find more objects and events. Second, it is becoming easier to teach these systems to work in low level light conditions like at night or in the rain and to identify new and smaller and occluded objects in sophisticated scenes using technologies like synthetic data, transfer learning and domain adaptation. Training deep learning-based models has typically required lots of training data but the rise of synthetic data means we can shorten the time to produce training data and better control it's quality. As a recent MIT News article outlined, "Synthetic data is a bit like diet soda. To be effective, it has to resemble the 'real thing' in certain ways. Diet soda should look, taste, and fizz like regular soda. Similarly, a synthetic dataset must have the same mathematical and statistical properties as the real-world dataset it's standing in for." These new technologies are being leveraged right now and will unlock the ability for video analytics to increasingly work in a wider range of environmental conditions to find the widest range of objects and events so operators can spend more time on security and less time worrying about the weather.

Figure 1.0 - Object detection accuracy improvements

2.1X

Improvement in open-sourced object detector performance from 2014-2019.

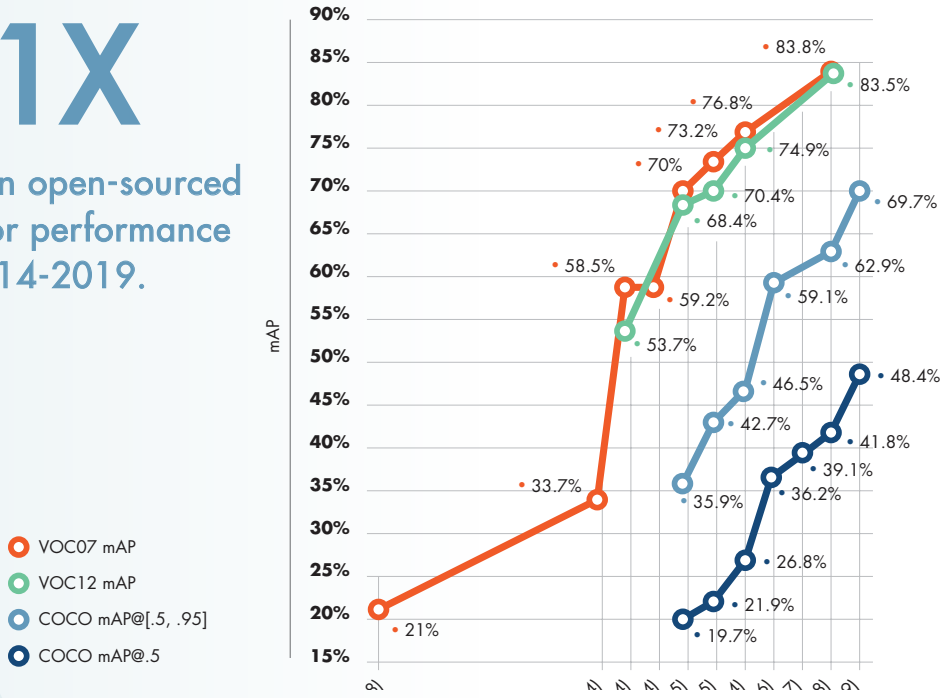

Fig. 1. The accuracy improvements of object detection on VOC07, VOC12 and MS-COCO datasets. Detectors in this figure: DPM-v1 [13], DPM-v5 [54], RCNN [16], SPPNet [17], Fast RCNN [18], Faster RCNN [19], SSD [21], FPN [22], Retina-Net [23], RefineDet [55], TridentNet[56].

1 https://www.google.com/url?q=https://pantelis.github.io/cs-gy-6613-spring-2020/docs/common/lectures/scene-understanding/object-detection/&sa=D&ust=1605561765180000&usg=AOvVaw1oAsAHSW7kpi8KNfR3IXgt

Bring Your Own (Moving) Camera

While more "smart" cameras are coming to market with advanced capabilities and functionality, what happens with the large investment in cameras with good images but no analytics capabilities? And what about the growing use of mobile surveillance cameras like dash cams, body cams, and drones? There are more than 100 million professionally installed cameras across the Americas and Europe and a low percentage of those have advanced intelligence features. With the increasing adoption of technologies like transfer learning (see graph below) and domain adaptation, deep learning-based tech supports 'retrofitting' analytics to practically any existing camera without any pre-learning or scene familiarization required. With a centralized GPU-based system, this means you can move the analytics around to the cameras that need them based on your unique environment and without getting locked in to only certain cameras having analytics. Transfer learning (TL) is a research approach in machine learning (ML) that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. For example, knowledge gained while learning to recognize cars could apply when trying to recognize trucks, which can speed up model development and improve accuracy. This same technology can be used to help make sense of the data coming from cameras that rarely

have analytics such as PTZ's and warped video from 180 degree cameras. Simply put, it is getting easier and easier to retrofit a "brain" to add intelligence to the "eyes" that are already installed and then easily move those analytics around to the cameras where they are needed. This includes all fixed and mobile cameras such as mobile phones, drones, dash cams, and even body cams. In mobile surveillance, many of the markets such as body cams and drones are expected to grow more than 30% annually for the next five years, leading to a tripling of those devices as part of the security infrastructure. Forward-thinking security teams will increasingly leverage mobile surveillance as part of their toolset and analytics will be required to keep up with this new deployment of cameras. State-of-the-art deep learning solutions are able to provide a full suite of video analytics to the customer using as little as a 480p stream from nearly any IP camera at just a few frames per second. These advancements mean any camera, including those that move and have been outside of the traditional analytics purview, can be a smart camera.

Figure 2.0 - Introduction To Transfer Learning

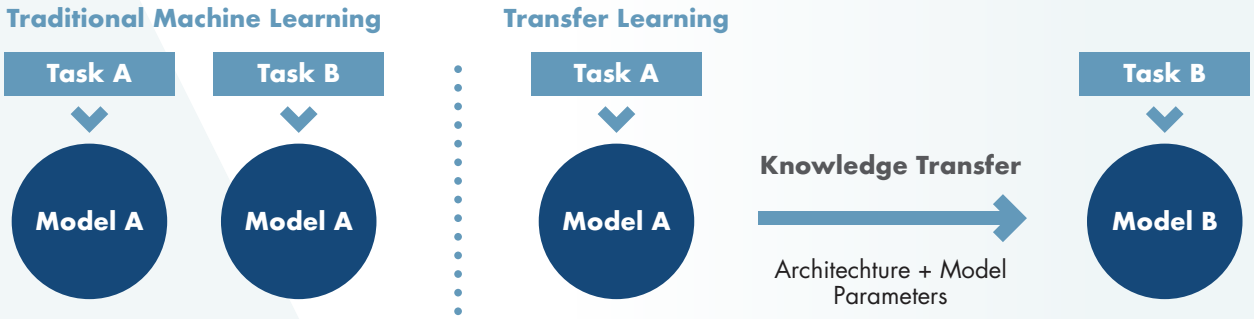

Fig. 2. Transfer learning means the use of previously acquired knowledge and skills in new learning or problem-solving situations. Figure 2 above shows the difference in approaches between Traditional Machine Learning and Transfer Learning.

2 https://www.oreilly.com/library/view/intelligent-projects-using/9781788996921/2231d733-c07c-4248-bdd9-18e66b758927.xhtml

Context-Aware

Analytics will increasingly take more scene and object context into account before making a determination. Think about the computer vision task of determining whether or not there is an object, like a long gun, present in a scene. Starting a few years ago, the initial technical approach was to build a long gun object detector. Today and looking ahead, that detector may be combined with additional information about the presence and location of a person (increasing the likelihood there is a long gun), the pose estimation of the person, whether the person has the object in their hands, and scene level data that may help determine the likelihood of the presence of the object. This same "context-aware" approach will enable variable determinations of the threat that an event poses. For example, a car that is moving faster through a scene than normal may be less of a risk when there are less cars and less people in the scene versus when the scene is crowded. Therefore, based on additional contextual information, the "threat score" of the car speeding event may be dynamically calculated based on the environment. There are two important outcomes of this "context aware" approach: more accurate detections from the technology and more use cases that will be unlocked.

Coalesced Intelligence

Humans are amazing at pattern recognition, with the ability to recognize many different data points and transform them into concrete, actionable steps. Being able to recognize patterns gave humans an evolutionary edge over animals. As machines, powered by robust GPU improvements, get better at recognizing patterns, the rich metadata pulled from sensors like video will increasingly be fused and leveraged to tackle the nearly impossible task of determining trends and patterns from large scale visual datasets. A pattern as it relates to activity in video consists of the following primitives for an object—presence, location, count, speed, and trajectory. It is one thing to see a suspicious vehicle (object) parked (presence) across (location) from a key facility once (count). But discovering a similar vehicle parked numerous times across different locations over long periods of time, and always shortly before or after those facilities have a particular activity going on, is another thing entirely. It requires the coalescing of various data points and the removal of "noisy data" from the analysis. This is not a task humans are particularly well-suited to do. But, powered by new GPU capabilities, deep-learning based algorithms are improving at pulling out this building block information from video in ever more accurate and faster leaps. While there is a logical move towards edge-based architectures for basic analytics, this type of multi-faceted analysis will continue to require centralized GPU-based architectures for the most important threat detection work. This increasing visual big data can also be stored in new types of databases that enable better searching and faster comparisons than before. As one example, Vintra's speed of comparing visually similar objects has improved by nearly 100x over the last 3 years. More advancements are ahead. Additionally, once patterns are established, anomalies can be detected with increasingly greater certainty. Operators can already set up rules to produce accurate alerts on things they care about that might happen. This is sometimes referred to as alerting on the "known unknowns" in a physical environment. But technology is improving at alerting them to the "unknown unknowns", that is, the situations that are so unexpected that they would not be considered at all when creating rules about how an environment should function. This requires many forms of metadata to be coalesced together to produce an accurate prediction about what is going on in the environment.

It's an exciting time to be solving challenges in the video analytics space as we continue to explore the possibilities of deep learning-powered technology. With 2021 in view, we are expecting condition agnostic, camera agnostic, and context aware video analytics that can coalesce new data points into better intelligence to take center stage in providing better security and safety outcomes for organizations.

If you'd like to learn more about how we are building for the future and addressing the key trends covered in this white paper, please send us a note to info@vintra.io or visit our website to get in touch.