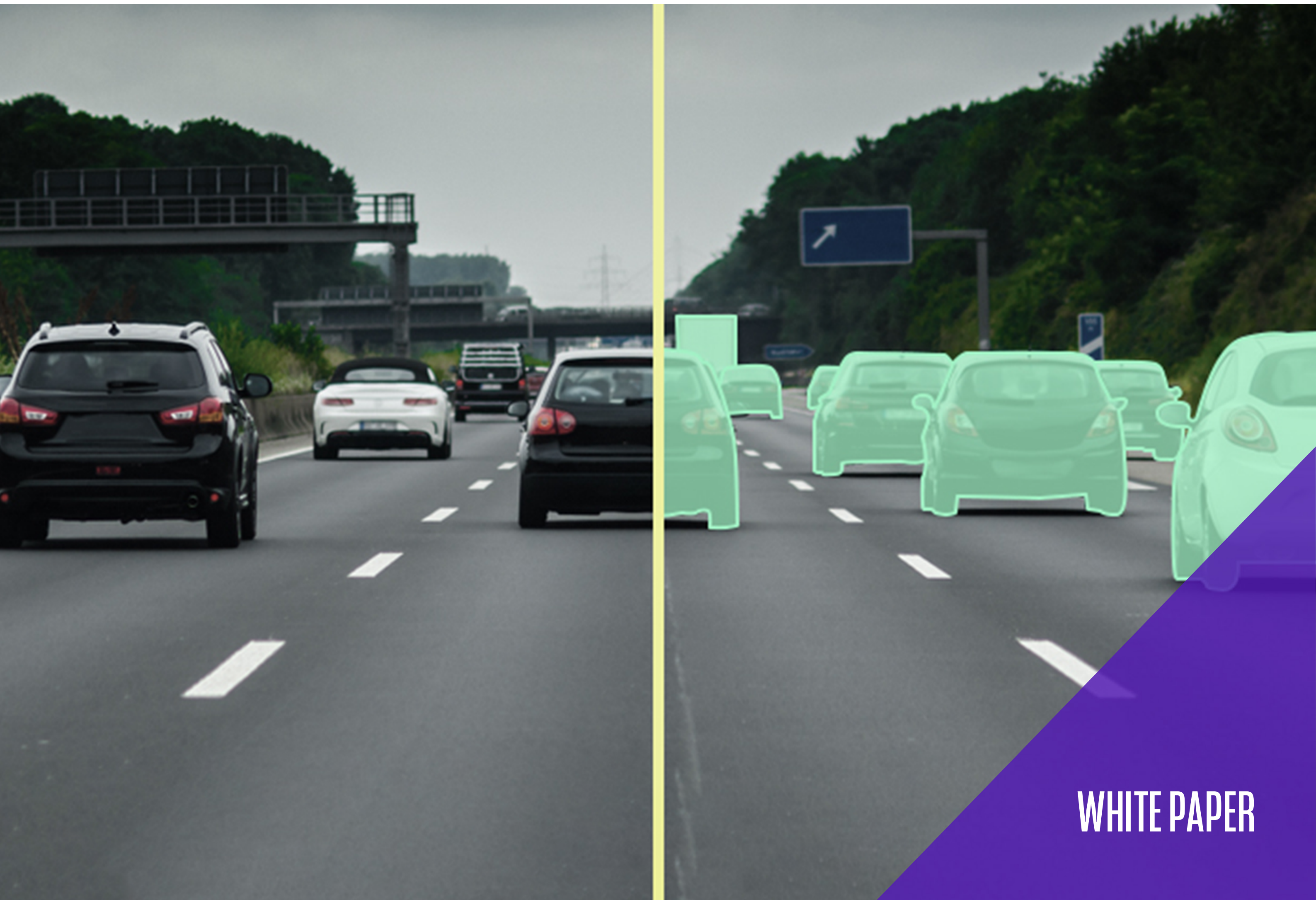
iMerit

DATA ANNOTATION FOR AUTONOMOUS VEHICLES



WHITE PAPER

1.35 MILLION!

If you want to understand the rationale for autonomous vehicles (AVs), pay attention to that number. That's how many people the World Health Organization says die every year from automobile crashes, with a person at the wheel. In fact, road traffic accidents are the leading cause of deaths for people aged 5-29.

Fully 90 percent of the roughly 30,000 traffic related deaths in the United States each year are, according to the Automobile Association of America (AAA), the result of (human) driver error.

Drastically reducing that statistic represents arguably the most compelling grand challenge of 21st Century transportation. On the strength of some early successes with basic AV building blocks, car makers are investing tens of billions of dollars annually to develop true autonomous vehicles, with no human needed at the wheel. Globally, car makers, software companies and others spent more than $54 billion on AV development in 2019, and market researchers expect that to grow more than tenfold to $556 billion by 2026. By that time, according to ABI Research, more than eight million cars with some level of autonomous vehicle technology will be on the road.

In fact, the race to a truly autonomous vehicle has expanded the world of car and car component manufacturers through revealing actions. At the 2020 International Consumer Electronics Show, for example, Sony surprised attendees and the world's technology press with its first concept car – an electronic vehicle with 33 sensors inside and out. While the car was designed for human drivers, the sensor array and always-on connectivity suggested a milestone on the way towards a network-controlled AV.

It's no wonder that the sector has attracted many of the global leaders in artificial intelligence, software development and device engineering. Among those developing AI software and/or other aspects for autonomous vehicles are Alphabet, Microsoft, Amazon, NVIDIA, Waymo, and Apple (through its acquisition of drive.ai).

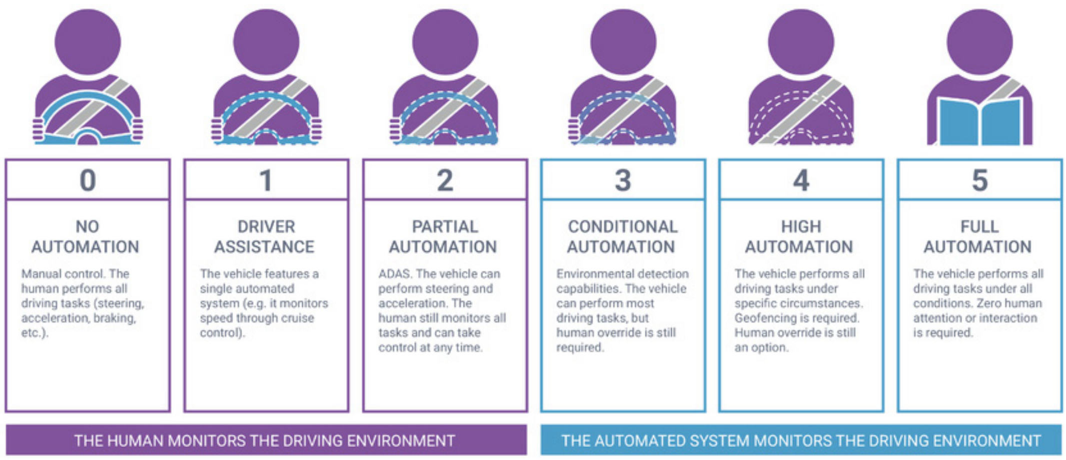
While some of the work has remained under the radar, various developers have hit meaningful milestones over the past few years. Baidu prototype autonomous vehicles, for example, have already driven more than 2 million kilometers. AV system developer Aptiv, meanwhile, has completed more than 50,000 autonomous vehicle taxi rides (in Las Vegas) in collaboration with the ride hailing service, Lyft.

A recent global survey of consumer attitudes found that more than three in four adults expect to ride in an autonomous vehicle in their lifetime. Almost 70 percent of American adults have that expectation – and the numbers rise to 80 percent and above in major Asian markets. More than 70 percent of adults globally believe that by 2029 autonomous vehicles will be safer than those driven by humans; their number one concern today: a threat to their personal safety by a failure of technology.

Safety, avoiding traffic accidents caused by humans, and engineering confidence that machines will do far better is but one of several compelling reasons for the massive investment in superfast, "smart" onboard computers that can think – and see – like a human, but without the downside of human frailties, such as a propensity to fall victim to distraction, or to fall asleep at the wheel. A Nobel Prize might go, for example, to the AV visionary who finally ends rush hour traffic jams through intelligent car sequencing.

The first development milestones have shown tremendous promise on the journey from yesterday's cars to those that, like George Jetson's favorite commuting transportation, do the driving for you. Some basic "driver assist" enhancements already available from a few manufacturers, such as Tesla, Kia, Mercedes Benz, and Cadillac, include so called "lane keeping assist," systems that use car-mounted cameras and an onboard computer to keep the car in lane when it starts to drift without the use of a turn signal, a self-parking feature, and collision avoidance systems that employ radar and automatic braking to slow or stop a car before it hits another object. Or vice versa.

Levels of driving automation



Source: **synopsys.com**

The Society of Automotive Engineers defines six levels of automobile driving automation, ranging from Level 0 – fully driver-operated vehicles – to Level 5 – fully autonomous vehicles. Most partially automated vehicles on the road today are Level 1: a single driver assist function, such as adaptive cruise control or lane assist. This level assumes the driver is primarily in control of the vehicle and relies on automation to deliver a basic level of enhanced safety or convenience.

Partial automation systems apply basic AI machine learning algorithms tied to a tightly bounded set of circumstances. For example, the car camera sees the car start to drift over the lines between one lane and another. It's been trained to recognize those white lines and to characterize them as safety barriers. The onboard computer's algorithm reasons that when the car drifts and the driver has not used the turn signal (A+B), that means the car (and driver) could be in danger, and it "knows" to correct the error (A+B->C) and return the car to its proper alignment.

Level 2 automation exponentially adds complexity by enabling the vehicle to control two primary functions – steering and acceleration/braking – but stop short of a fully autonomous ability to react to environmental factors, such as a pedestrian crossing against traffic or an oncoming vehicle. A handful of car makers have released cars with Level 2 partial driving automation systems, such as Tesla's Autopilot and Cadillac's Super Cruise.

Much of the near-term investment in AV development centers on the technologically aggressive leap from Level 2 partial driving automation to Level 3 conditional driving automation. This is the point in AV evolution where cars begin to actively see the world around them and begin to think for themselves.

Level 3 car systems employ multiple sensors to provide environmental data on, static and moving objects, such as cars, pedestrians, barriers, and such. These cars will be capable of significant independent operation albeit with an expectation of human driver intervention in a crisis.

The next step beyond, Level 4 vehicles – some of which have made it to pilots, and other limited rollouts – are fully autonomous, capable of acting independently in a crisis, such as an impending crash, but have the controls (steering wheel, brake and accelerator pedals) that enable human intervention. Most of the initial commercial applications center on ride sharing and other commercial ventures.

A fully autonomous vehicle – Level 5 – just takes that scenario to its logical conclusion: managing every aspect of driving than a human would – except without the all-too-human failures that result in car crashes and worse. Level 5 autonomous vehicles won't include a steering wheel, or brake and accelerator pedals, and represent the end goal for AV development.

CRASH TEST DUMMY

What happens when a fully autonomous vehicle encounters another vehicle with an all-too-human driver? When there's a collision, how should we assess blame? In an infamous 2017 incident in Las Vegas, let's say 80-20, advantage AV.

In November 2017 an autonomous shuttle pilot project had just launched to great fanfare when, less than two hours in, the shuttle collided with a truck in a minor fender bender. Or, more accurately, the truck collided with the shuttle, according to local police officials who said the truck driver was at fault. As with most accidents, though, assessing blame was an inexact process, one that highlights some of the limitations, and challenges for AV development.

The good news for AV adherents is that the shuttle performed as designed – it detected the approaching truck and stopped. In essence, the shuttle's AV algorithm opted for a default reaction: it sensed impending danger and stopped the vehicle. But at least a few shuttle passengers would have liked a take charge onboard computer. Could it have honked the shuttle horn to alert the truck driver? Or taken the riskier maneuver of backing the shuttle away from the approaching truck? How should it have used its milliseconds of situation analysis to think like a human, only better?

Even with cutting edge technology, there's always a backseat driver. It's not easy being a driving computer.

EDUCATION

ROTE LEARNING

Before an autonomous vehicle can reason its way out of trouble on the road it needs to understand what it's seeing. AV car makers use some combination of three forms of sensors – cameras, laser-based 3-D Light Detection and Ranging (LiDAR) and traditional 2-D radar – that work in tandem to sense the road, identify objects, and discern the difference between passing scenery and a threatening object. LiDAR is becoming particularly attractive as a way to help the AV computer judge distance and speed. A single LiDAR camera is typically installed on the roof of the car, while radar, required to identify objects very close to the car where the roof-mounted LiDAR camera can't see, are typically mounted front and back.

To safely navigate a city or any other streetscape, an autonomous vehicle needs to see the road as a human being does. It needs to be able to understand – and react to:

1 **Objects on the road and the difference between static and moving objects.**

2 **The nature of those objects.**

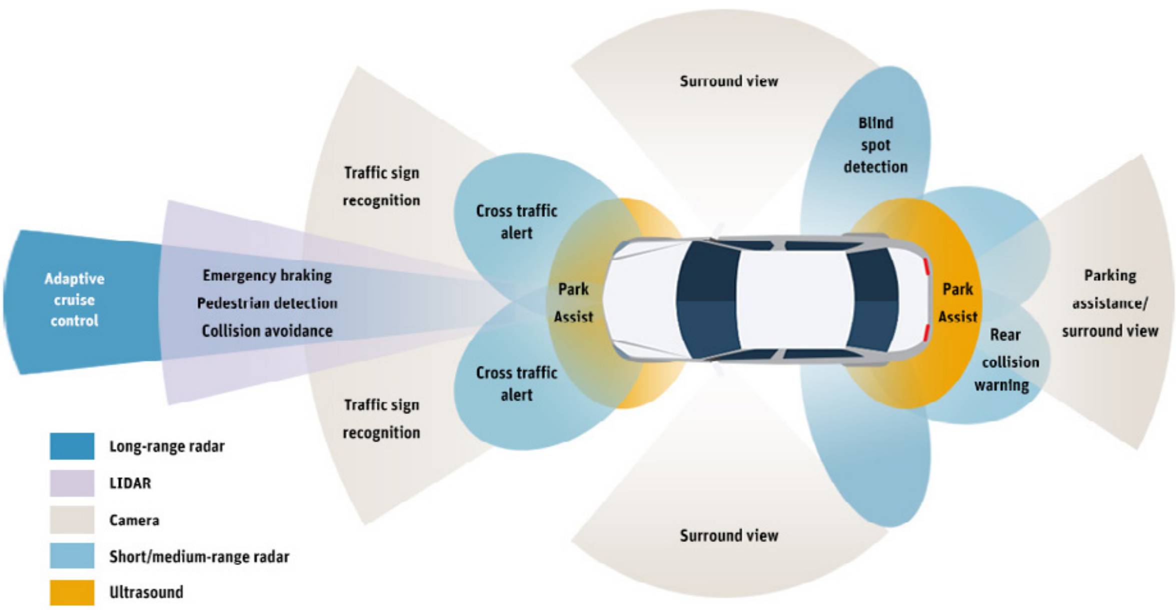
3 **The speed and direction of those objects.**

4 **Any potential threat caused by those objects.**

The algorithms powering autonomous vehicles need to, in milliseconds, absorb and process data from multiple sensor arrays and sensor types, evaluate shifting environmental conditions in real time, and issue instructions to the car's steering, braking, and acceleration systems as needed. All of the "If A then B" decision making relies on the algorithm's ability to infer real-time scenarios from the training data used to help it recognize and interpret the significance of sensor input.

The graphic below describes the various views – and corresponding sensors – an autonomous vehicle relies on to navigate the road, monitor traffic and other objects, and react to changes in its immediate environment, everything from a "simple" curve in the road to an exponentially more complex chain reaction series of vehicle collisions.

Autonomous Vehicle Radar Systems & Sensors



Source: [ansys.com]

Processing all of that data is built on a complex training program that takes sample data from each type of input, annotates it to break each data image down to its component objects, and then correlates it across all of the many scenarios the car could encounter on the way from Point A to Point B.

Training that system of onboard sensors and AV computers encompasses thousands upon thousands of hours of video (training data) broken down into its component parts of different objects, analyzed, categorized, and then fed into the algorithm to test whether or not the lesson has been learned. It takes many years to train a child to recognize and react to potentially dangerous objects before a parent will let go of its hands; the task is no less daunting for a machine.

As autonomous vehicle developers progress from basic Level 1 driver assist features towards full Level 5 automated driving, the requirements for algorithm training and data process increase steadily up a series of stepped priorities and capabilities. As the graphic below illustrates, the increasing complexity of each task results in exponentially more time and effort – and skills – to deliver the level of training data the algorithm requires to think and see like a human.

That ascending level of complexity explains why AV developers have come to rely on third-party data annotation firms. As with other specialized tasks that require an advanced, highly trained workforce and technologically focused tools, data annotation has evolved into its own machine learning niche.

As autonomous vehicle developers progress from basic Level 1 driver assist features towards full Level 5 automated driving, the requirements for algorithm training and data process increase steadily up a series of stepped priorities and capabilities.

The growing focus on annotation tool development, in fact, reflects the evolution in data annotation complexity and task requirements driven directly by AV developers. While early AV training data projects primarily employed basic annotation tools, such as bounding boxes and polygon identification, data segmentation requirements have grown steadily more complex and detailed. That has driven a correspondingly aggressive effort on the part of leading annotation firms to develop ever more sophisticated tools, processes and workforce training in order to keep pace.

In something of a continuing development loop, the results of those investments in more powerful and functional annotation tools will likely include ever more demanding data annotation efforts aimed at accelerating the machine learning cycle, which in turn will put a premium on further tool development.

The final exam: testing whether or not the algorithm can apply all that data to recognize patterns – and accurately infer the identity of new objects through its “experience.” Except this is the Ivy League of machine learning. Sentiment analysis firms, for example, are ecstatic if their natural language processing engines can achieve 75-80 percent accuracy. AV algorithms have to score in the very high 90s if they want to avoid flunking out.

CLIMBING STEPS OF COMPLEXITY

Secs/task

1 **Bounding boxes**

Marking location using simple boxes around objects.

Mins/task

2 **Polygons**

Marking location by drawing precise boundaries around some objects.

+30 Mins/task

3 **Segmentation**

Marking all the boundaries of all objects in an image precisely.

+30 Mins/task

4 **PanOptic Segmentation**

All objects are precisely marked & differentiated by class of object (car, tree, fire, hydrant etc.)

+30 Mins/task

5 **Tracking**

Objects marked and tracked across frames of video, included occluded sections, by correlating frames back and forth.

+30 Mins/task

6 **LIDAR**

Combining thousands of points to mark the shape of objects. Labeled LIDAR data helps the computer judge distance and speed.

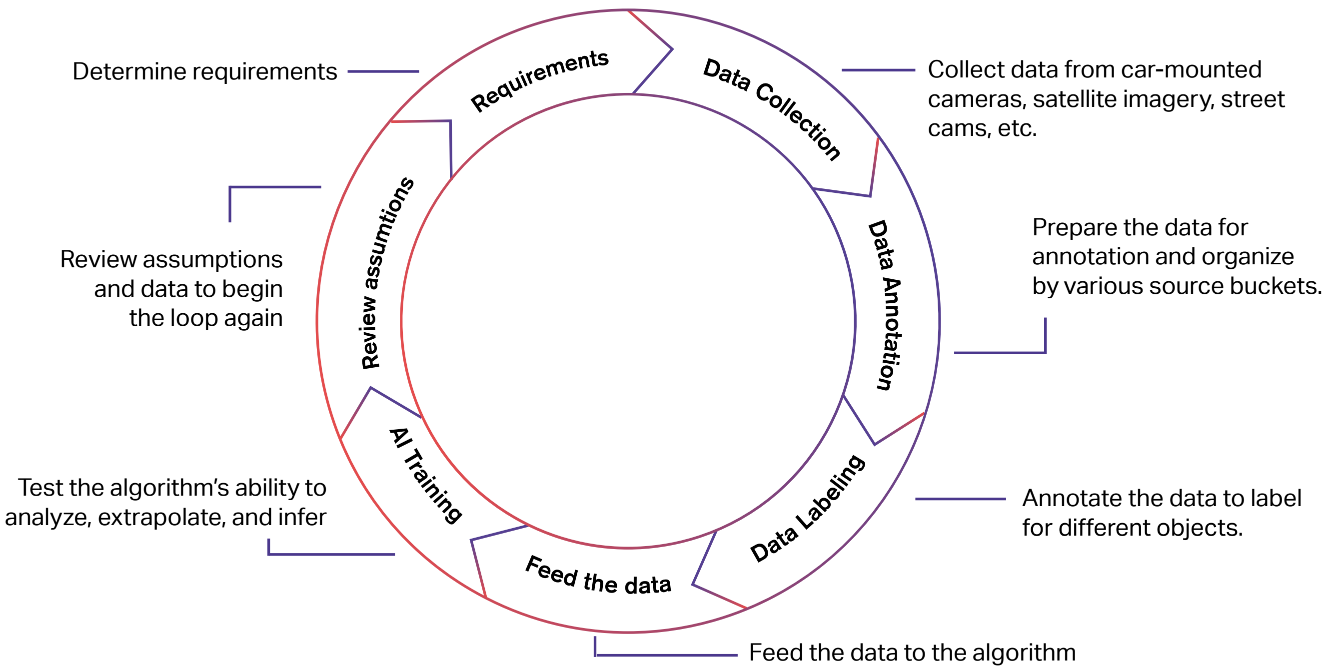
+30 Mins/task

7 **MultiSensor Fusion**

Combining LIDAR and images from multiple angles captured from diffrent sensors, to reduce uncertainty in navigation.

But it's become increasingly clear that the only truly effective means of teaching autonomous vehicle computers is through a combination of human intervention, applying acquired experience and judgment to select the best training data, and software tools developed to do the actual annotation from that material. Companies with advanced workforces and multi-variant tooling capabilities are best positioned to facilitate this process.

ML Cycle for Autonomous vehicles



Source: **imerit.net**

A recently published paper in the MIT Sloan Management Review on the future of artificial intelligence notes, ironically, the impact of human-machine interaction as a critical factor in enhanced machine learning. Of utmost importance, the authors note, is the value of teaming the predictive capabilities of algorithms with "the expertise and intuition of humans, especially in decision-framing." In the field of autonomous vehicle development, this approach applies directly to the use of AI-powered tools built to support autonomous vehicle development.

MOUNTAINS OF AV DATA

Ask pretty much any autonomous vehicle developer and they will quickly admit that sorting through and annotating millions of street images is the Mount Everest of product development; a daunting task, requiring specialist skills, an awful lot of experience, and a keen eye for potentially treacherous gaps along the way.

Data annotation specialists need to present objects and images in simple, clearly defined packages that an algorithm can use to understand the world it is starting to "see." We look out on a street and easily break the image in our mind's eye into all of its component parts – the woman in a bright red dress approaching the crosswalk, the bicycle messenger hurtling towards the intersection from the opposite direction, the parked cars sitting silently along both curbs, and the trees, buildings, and mailboxes that dot the street in a seemingly haphazard pattern.

The AV onboard computer, though, has to be fed the equivalent of years – decades – of human experience in the form of polygons that accurately describe every object in the real world. And it needs to categorize them in ways that, eventually, will allow that computer (and its AV algorithm) to extrapolate variations on each one in order to recognize the same woman in shorts, a Volkswagen Beetle from a Cadillac Escalade, a four-wheeled shopping cart from a mailbox.

As any good carpenter will tell you, owning a set of tools and knowing how to use them are two very different concepts. That's why car manufacturers and AV system developers typically engage data annotation companies to apply a specialist's skill set to preparing and annotating thousands of hours of video data at a time.

DATA ANNOTATION TOOLBOX

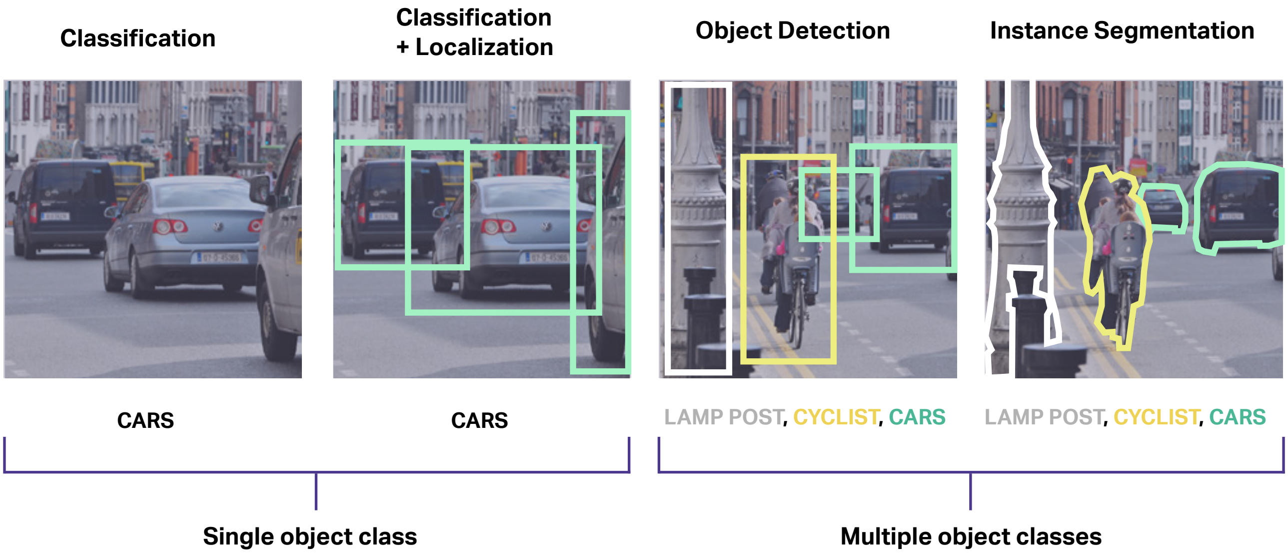
All machine learning involves the parsing of some type of language to embed human understanding and reasoning into an algorithm. The natural language processing behind Alexa, chatbots and sentiment analysis tools are all based on text analysis. In some respects, that is a less challenging process than autonomous vehicle machine learning – Global service providers have to apply NLP across dozens of languages, or more. But the stakes are higher with autonomous vehicles, and the success metrics are correspondingly far more stringent.

Since autonomous vehicle development is very much a visual interpretation challenge, all of the training data is some form of video – everything from still images to full motion video. The cameras and other sensors installed in an autonomous vehicle are bombarded by constantly shifting streams of information. Some of it is static – buildings, fields, lampposts and the like as the car travels down one street or highway or another – while the rest represent seemingly random events that can require immediate intervention by the car's onboard AV computer – a pedestrian darting out between parked cars, a bicyclist cutting in front or around the car, or another car veering into the same lane.

In each instance, the AV algorithm controlling the car must make split-second decisions about the nature of the object and the danger it represents to the vehicle – or vice versa.

Before the AV can react to the surrounding environment, it must first know the nature of each of the separate objects that comprise its immediate world. In fact, it first needs to understand the road as an environment, what constitutes an individual object, the difference between individual and often momentarily overlapping objects, and the dynamic changes driven by shifting perspectives and motion – in essence the "ground truth" as captured in video that will then be used as a reality check against the algorithm's inferences and extrapolation.

After AV data experts have prepped the imagery, they examine individual frames of video to segment it into objects that can be categorized according to the objectives and priorities of the car manufacturer or AV system developer: all of the different forms of buildings, men, women, children, cars, trucks and so on. Data annotation specialists use software tools to identify the component objects (and their positioning) of each image through "semantic segmentation" that breaks down a frame of video pixel by pixel into its separate objects, and apply "bounding boxes" that (as the term implies) set the parameters for each of those objects as polygons that can be fed to the AV algorithm. The software can be used to track objects from one frame to another, to teach motion vectors, speed and acceleration.



Semantic segmentation is a labor intensive, three-step process that delivers a fully formed description of an image and the objects within. First, each of the objects within the image must be identified and classified, such as the cars in the graphic above. Once the object has been classified, the algorithm needs to be trained to detect it in various images (and differentiate it from other objects). Data annotation experts will use a simple bounding box to identify and position each object within an image. Then the annotation expert will apply pixel-by-pixel image segmentation to set the boundaries for each object, including the borders between one object and another. Think of a digital paint-by-numbers image in reverse, where the complete painting is reduced to its component objects down to single pixel accuracy.

Then repeat tens of thousands of times.

For some sense of the scale of the work, consider that iMerit data experts have been performing dense pixel semantic segmentation on almost 100,000 street images each day for just one global car manufacturer. And each image contains multiple objects that must be systematically categorized to enable the AV algorithm to not just see but also to interpret its environment.

It's not just the volume of data that must be managed, processed, and interpreted. Annotation must also take into account the qualitative aspects of the data and its ability to provide a sufficiently robust view of all the situations that can develop as an autonomous vehicle navigates everything from a busy urban intersection to a crowded parking garage. The bounding boxes used to identify and define individual objects also need to define motion vectors – for example, another car pulling out of a parking spot – to enable the AV algorithm to accurately predict its path and speed.

COMPREHENSIVE
APPROACH AND SOLUTION

Simply feeding raw data through an annotation engine, though, ignores the other steps that must be in place for an effective process. A comprehensive, end-to-end solution for data annotation encompasses three essential pillars – data preparation, annotation, and testing – that support the overall result: each adds a crucial element to the process integrity that drives AV machine learning.

DATA PREPARATION

The stream of unstructured data created and/or acquired by the AV developer must be structured according to categorization buckets, such as source – car cameras, street cams and even satellite imagery. At this point, an experienced data expert might identify gaps in the raw data (missing scenes, perspectives) or even a shortage of specific jects or scenes that could prevent the algorithm from understanding the truth of what it "sees." An experienced annotation expert will know, for example, when to request additional data sources adjacent to the first source. For example, different street views.

DATA ANNOTATION

The heart of the machine learning process and arguably the most significant hurdle for autonomous vehicle developers depends on the quantity and quality of the raw data. It's been said that identifying every conceivable object for an autonomous vehicle's education is a practical impossibility. So successful development hinges on identifying those objects – and variations – that can enable an AV to navigate safely, and successfully infer its surroundings from its algorithmic training.

The annotated data forms the essential course material for machine learning. AV developers will feed that output through their AV algorithm to train it to process sensory data from its immediate environment much as a human brain would: (1) recognize objects and terrain; (2) calculate distance, direction, and speed; (3) determine its own direction and speed based on environmental factors; and (4) infer missing data from its own accumulated experience and "wisdom."

DATA TESTING

This is where many AV developers move from simulations to real-world testing – can a car, controlled by an AV algorithm, safely navigate a closed course, or even a carefully controlled series of streets? Much of the examination concerns the inevitable testing handovers between car AV and driver. Each handoff is an event, triggered either by the driver or the onboard AV computer, that reveals the algorithm was confused and unable to cope with something the vehicle saw or sensed. While AV manufacturers and system developers perform live road and simulation tests, the data annotation company plays a vital role in the "post-mortem " investigation into the cause of each handover.

In each instance, the development team must determine the cause of the handover – and a path to filling that gap in processing capabilities. If the algorithm failed to infer new objects from historical data, will a fix require more of the same data to reinforce machine learning, or will developers need to apply additional data categories, for example billboards, or lawnmowers?

In most cases, the AV developer will run through a simulation with the algorithm to determine what would have happened without the handover. The annotation company, working with AV developers, needs to then help determine whether the car performed as designed, but failed because of a too-limited data set, or was exposed by a flaw in the algorithm.

Employing annotation software alone is the equivalent of working in a vacuum. An effective, efficient process requires human intelligence – institutional knowledge developed through widespread work with multiple clients, experienced annotation experts who understand how to identify and prioritize the most valuable raw data, how to spot holes in the data library and how to help investigate anomalies that arise in the testing process.

Moreover, as with any successful collaboration, the data annotation firm and client must be able to take the specific data requirements of the project and develop robust guidelines the annotation solutions team and client use to support QA goals. Those guidelines become "live" documents that are updated through a feedback loop that is used to improve quality and accuracy metrics.

WHAT CAN iMERIT DO FOR YOUR AV PROJECT?

iMerit helps some of the world's largest companies deploy artificial intelligence accurately, securely, and with operational efficiency. One of the innovators in the field, we work with three of the world's largest automotive manufacturers. We're a global leader in data annotation for autonomous vehicles, having enriched more than 150 million data points for autonomous vehicle development. In addition to our state-of-the-art annotation tools, we employ approximately 650 full-time data experts working on AV projects, with leading car manufacturers.

Our secret sauce? All of the processes and software we have developed that lead to greater operational efficiency for your organization. Over the course of our engagements with you we constantly work to increase your organization's efficiency.

The payoff for you? More accurate data, and lower effective costs to you for the data we process. The data is enriched securely in one of our nine monitored delivery centers and realizes quality metrics in excess of 98%.

We always bring a comprehensive approach to our clients' projects. Once you engage iMerit for a project or ongoing initiative we assign a solutions architect to work with you on establishing goals and objectives, and a plan to achieve them.

Need help with "ground truth" use cases? What about support for your data acquisition and pre-annotation process where our experienced data experts can help sort through different data sets to decide which are worth annotating? Are you gathering the right types and sufficient quantity of data to develop accurate predictions? These are all projects where iMerit brings its knowledge and experience with autonomous vehicles, the institutional knowledge of a stable, highly motivated workforce, and skilled experts in all aspects of data facilitation for machine learning.

iMerit

DATA ANNOTATION FOR

AUTONOMOUS VEHICLES

imerit.net