At Waymo, we use machine learning to detect and classify different types of objects and road features. The powerful neural nets that make up our perception system learn to recognize objects and their corresponding behaviors from labeled examples of everything our Waymo Driver encounters, from joggers and cyclists, to traffic light colors and temporary road signs, or even trees and shrubs. Over the past decade we have built up an enormous collection of objects captured by our powerful custom-designed hardware.

This wealth of experience is invaluable, but it also poses a challenge: how to find the most useful examples in this sea of sensor data. Trying to locate specific examples – such as when our vehicles have observed a person carrying a skateboard – could be like looking for the proverbial needle in a haystack. However, through a collaboration between Waymo and Google Research, we have drawn from Google's expertise in web search to develop our Content Search tool. By using technology similar to what powers Google Photos and Google Image Search, Waymo's Content Search lets our engineers quickly locate just about any object in our driving history and data logs — essentially turning Waymo's 20 million miles of on-road experience into a searchable catalog of billions of objects.

Approaching data mining as a search problem

Our Waymo Driver gets smarter with time and experience: everything that one vehicle learns can be shared amongst the entire fleet. This ability to learn both quickly and collectively is especially important because our driving environments are constantly changing. For example, with personal mobility solutions gaining popularity (especially in urban areas), the Waymo Driver regularly encounters new forms of transportation. So we want to continually train our system to ensure that we can not only distinguish between a vehicle and a cyclist, but also between a pedestrian and a person on a scooter.

In the past, to find these distinct examples in our driving logs, our researchers relied on heuristic methods that parsed our data based on various features, such as an object's estimated speed and height. For instance, to locate examples of people riding scooters, we might have looked through our log data for objects of a certain height traveling between 0 and 20 mph. While this method yielded relevant examples, the results were often too broad, since many objects share those attributes.

With Content Search, we can now approach this type of data mining as a search problem. The core principle underlying this new tool is "knowledge transfer" — applying knowledge gained from solving one problem (such as finding all the "dogs" in your Google Photos album) to a different, but related problem (such as searching through our driving logs to identify all the times our Waymo Driver has driven past a dog). By indexing our massive catalog of driving data, our engineers can find relevant data to train and improve our neural networks much faster.

Bringing the power of image recognition to self-driving cars

With Content Search, our engineers can search the world via our sensor logs in multiple ways: We can conduct a "similarity search," hone in on objects in ultra fine-grained categories, and search by text in the scene.

Saguaro
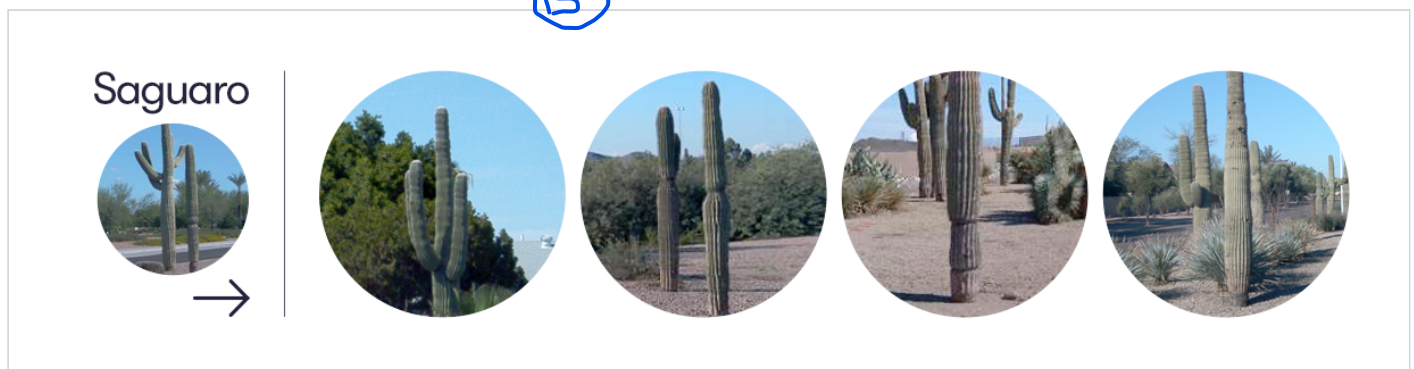
True to its name, similarity search allows us to easily find items similar to a given object in our driving logs by running an image comparison query. For example, to improve our models around cacti as vegetation, we can start with any image of a cactus, whether it's an example we have already found in our driving history, a photo of a cactus we found online, or even a drawing of a cactus. Content Search then returns instances where our self-driving vehicles observed similar-looking objects in the real world.

This core image search model works by converting every object in Waymo's driving logs, whether a park bench, a trash can near the side of the road, or a moving object, into embeddings, a machine learning technique that makes it possible to rank objects by how similar they are to each other. By creating embeddings for each object based on attributes and deploying a process similar to Google's real-time embedding similarities matching service, we can efficiently compare any query against the images in our driving logs and locate objects similar to our query in a matter of seconds.

Category

A single class of object can vary in shape, form, and type. For instance, road debris can be anything from a plastic bag or a tire scrap to a cardboard box or a lost pair of pants. To build robust machine learning models that can generalize and detect the different articles we might come across on the road – even ones our vehicle has never seen before – our engineers train our neural nets with a diverse range of examples.

To do this, we utilize ultra fine-grained search to find objects within a specific category. The backend for this search is a categorical machine learning model that helps our Content Search tool understand whether a specific object category is found in an image or not. This deep level of understanding opens up the ability to perform extraordinarily niche searches on objects that share a particular trait such as the make and model of a car, or even specific breeds of dogs.

text

Lastly, many objects on the road contain text which is relevant to driving, such as road signs, or the "oversized" notice on a large truck. Content Search uses a state-of-the-art optical character recognition model to annotate our driving logs based on text and words found in a scene, enabling us to read road signs, emergency vehicles, and other cars and trucks with signage.

Data labeling at scale

With Content Search, we're able to automatically annotate billions of objects in our driving history which in turn has exponentially increased the speed and quality of data we send for labeling. The ability to accelerate labeling has contributed to many improvements across our system, from detecting school buses with children about to step onto the sidewalk or people riding electric scooters to a cat or a dog crossing a street. As Waymo expands to more cities, we'll continue to encounter new objects and scenarios. Content Search allows us to learn even more quickly and help us achieve our goal of bringing self-driving technology to more people, in more places.

*Acknowledgements

Special thanks to Nichola Abdo, Junhua Mao, Guowang Li of Waymo; Adrian Smarandoiu and Solomon Chang of Google Infrastructure; and Zhen Li, Howard Zhou, Neil Alldrin, Tom Duerig, Ruiqi Guo, Alessandro Bissacco and Yasuhisa Fujii of Google Research.

Join our team and help us build the World's Most Experienced Driver.
Waymo is looking for talented software and hardware engineers, researchers, and out-of-the-box thinkers to help us tackle real-world problems, and make the roads safer for everyone. Come work with other passionate engineers and world-class researchers on novel and difficult problems — learn more at waymo.com/joinus.