**Computer Vision**

You’ve been building up a solid foundation of programming and data analysis skills and, in this section, I’m excited to introduce you to one more self-driving car tool: **computer vision**.

Computer vision is how machines, like self-driving cars, visually perceive the world and respond to it. Imagine that you are driving a car; you have to use your senses to notice pedestrians, other cars, bicyclists, and all the road and traffic markings around you.

This is similar to how a self-driving car “sees” the world. It gathers data through cameras and other sensors, and then uses that input to safely navigate and move through the world.

A car parked on the side of a road

Description automatically generated

Color coded image of how a self-driving car "sees" lanes and surrounding vehicles.

In this lesson, we’ll learn about common computer vision techniques that are used to analyze camera images and extract important information from them like the color or shape of different objects.

We’ll also briefly talk about machine learning and how it’s used in combination with computer vision to give machines a way to learn from data and recognize patterns in images.

## Computer Vision in Industry

To help us learn about computer vision techniques and applications, we have with us an industry expert: Tarin Ziyaee, the co-founder and Chief Technology Officer of [Voyage](https://voyage.auto/), a self-driving car company.

At Voyage, computer vision is used in a myriad of applications, for example, the recognition of what state a traffic light is in, or, the detection of lanes, and so on.

One cool thing about computer vision though, is that the techniques that you'll learn about, need not only be used with camera images, but also images created with other sensors. The techniques that you'll learn, will be useful for any data that has **spatial coherency**.

And spatially coherent data can be thought of as any data that predictably varies over space, like sound, for example. If you hear sound from a speaker close up it will sound very loud, but the farther you get away, the softer the sound will get. And so the volume of a sound can give you spatial information!

So, in addition to cameras, self-driving cars use sensors like radar and LiDAR, which use sound waves and lasers to gather data about a car’s surroundings.

**Sensors**

Traditionally, RADAR is used for long-range detections, while cameras are used for rich sensory inputs.

When it comes to sensor configurations for SDCs, we have to note the details therein; LiDARs and RADARs are what are known as **active sensors**. That is, they sense the environment based on transmissions of energy.

Cameras on the other hand, are **passive sensors**; they can only sense the environment based on energy (in this case, photons) already in the scene.

These sensor details have serious repercussions with respect to the types of algorithms we end up using to analyze this data. Computer vision has many powerful tools, but part of good design had to do not only with knowing what to do, but also what not to do when we use sensor data from multiple sources.

Now, as we mentioned, computer vision is a very powerful tool, but it should not necessarily be associated with *only* camera images. It's possible to also construct LiDAR images with a LiDAR sensor, thereby giving you measured-depth alongside well as classified pixels. Next you'll see some LiDAR sensor output and resulting data!

**LiDAR Sensor Data**

LiDAR stands for Light Detection and Ranging, and it is a type of sensor that uses light (a laser) to measure the distance between itself and objects that reflect light. It does this by sending a series of laser pulses out and measuring the time it takes for an object to reflect that light back to the sensor; the longer the reflection takes, the farther an object is from the sensor. In this sense, LiDAR is **spatially coherent** data, and can be used to create a visual world representation.

Below, you can see the output of a LiDAR that Voyage uses, it sends out pulses of light and detects the car's surroundings.

A picture containing dark, night, standing

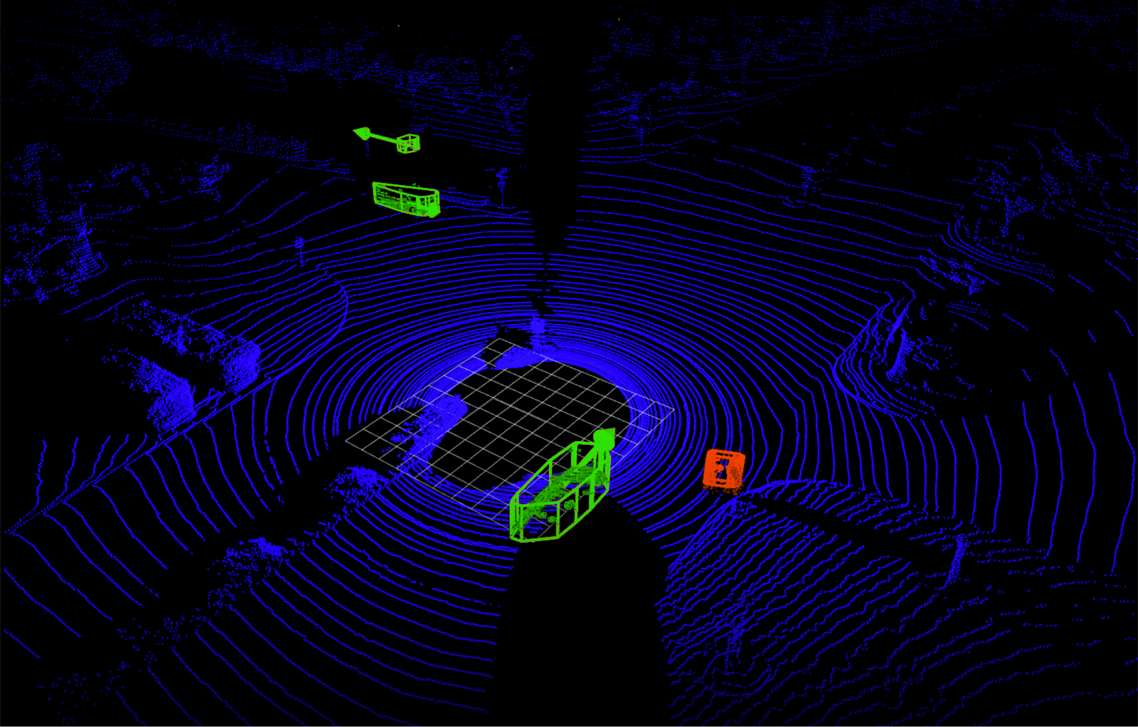
Description automatically generated

LiDAR sensor output. The blue indicates the detected, close-together points and the grid in the center is the space that the self-driving car occupies in the scene.

**Point Clouds**

Since LiDAR uses laser light, which sends out a thin light beam, the data it collects ends up being many single points also called **point clouds**. These point clouds can tell us a lot about an object like its shape and surface texture. By clustering points and analyzing them, this data provides enough information to classify an object or track a cluster of points over time!

Below, you can see the result of a classification algorithm performed on these points. The red are pedestrians and green indicates other cars.



That same LiDAR data, sent through a classifier that clusters the points together and classifies them as a human (red) or as a car wth some velocity (green).

**Visual World**

As you can see, LiDAR data provides enough spatial information to create a visual world representation, and in industry, computer vision techniques can be used to classify objects using not only camera images, but also point clouds and other types of spatially coherent data!

## Image Classification Pipeline

An image classifier is an algorithm that takes in an image as input and outputs a label or “class” that identifies that image. For example, a traffic sign classifier will look at different of roads and be able to identify whether that road contains humans, cars, bikes and so on. Distinguishing and classifying each image based on its contents.

There are many types of classifiers, used to recognize specific objects or even behaviors — like whether a person is walking or running — but they all involve a similar series of steps...

1. First, a computer receives visual **input** from an imaging device like a camera. This is typically captured as an image or a sequence of images.
2. Each image is then sent through some **pre-processing** steps whose purpose is to standardize each image. Common pre-processing steps include resizing an image, or rotating it, to change its shape or transforming the image from one color to another - like from color to grayscale. Only by standardizing each image, for example: making them the same size, can you then compare them and further analyze them in the same way.
3. Next, we **extract features**. Features are what help us define certain objects, and they are usually information about object shape or color. For example, some features that distinguish a car from a bicycle are that a car is usually a much larger shape and that it has 4 wheels instead of two. The shape and wheels would be distinguishing features for a car. And we’ll talk more about features later in this lesson.
4. And then, finally, these features are fed into a **classification model**! This step looks at any features from the previous step and predicts whether, say, this image is of a car or a pedestrian or a bike, and so on.

Image classification pipeline and specific pipeline example applied to classifying an image of a car. A screenshot of a cell phone

Description automatically generated

You’ll be programming each of these classification steps manually so that you really understand each step.

And by the end of this lesson, you’ll have all the skills you’ll need to complete the final project: building a traffic light classifier, which takes in images of traffic lights and separates them into three classes: red, yellow, or green lights.

## Image Classification Pipeline

A standard classification pipeline uses the following steps to classify a given image.

A screenshot of a cell phone

Description automatically generated

## Classification Techniques

You’ve just seen a complete classification pipeline. For an image classifier, starting with some input images, computer vision techniques are used to process those images, and extract features, like distinguishing colors or shapes in that image.

Then, a classifier looks at these features and outputs a class, which is a label that describes the image.

**A classifier should predict that images with similar shapes or colors have the same class.**

We usually tell a classification model what to look for. For example, say we are looking at a bunch of images and we want to classify them into two classes: car and not-car! To classify a car, we might write a program that looks for the different parts of a car: wheels, lights, windows, and so on, and then if those things are found, we’ll classify an image as a car. We decide what traits are important to look for.

## Machine Learning

However, there's another way to create a classifier and that’s with machine learning.

Machine learning allows a computer to figure out things on its own by giving it lots of examples. So, instead of telling a model what traits to look for, with machine learning, we’d just give it lots of images of cars and not-cars and let it learn to recognize traits that differentiate them! It can learn to recognize wheels and windows and which classification algorithm is best for accurately classifying any given image as car or not-car!

Now, you might be wondering: how exactly does a model like this learn to classify different images?

Next, we’ll look at how machine learning techniques can actually be trained to classify sets of images.

## Separating Data

When we talk about machine learning, you’ll often hear the terms “deep learning” and “neural networks” come up. This might conjure up an image of a brain or some strange graphic of layers of mathematical formulas and data. But at their core, all of these learning techniques are about separating data into different classes.

The image that this conjures in my head is of a child playing on a beach. The child sees some blue and yellow shells in the sand.

Then, someone says to the child: “put these shells into groups and draw a line between them.”

Without me telling you anything else, how would you group these shells?

A close up of graphics

Description automatically generated

Image of yellow and blue shells

You might separate them by color and shape and draw a line between these two types of shells. Now, for a computer, the scenario might be that we give it individual images of shells, and just like a child, a neural network can learn how to separate these images of shells based on similarities or differences in the given examples.

A close up of a flower

Description automatically generated

Separated shells

After this separation step, if a network sees a new image, one it hasn’t seen before, it sorts and classifies it based on which side of the line it falls!

Realistically, data is often a lot more complex than this, but neural networks just layer separation on top of separation layer to create more complex boundaries and group all kinds of data!

## Training a Model

Earlier, I gave a simple example of a classification model that classifies all images as either **car** or **not-car**. And I said that using machine learning techniques, you can give a model lots of examples of cars and not-cars until it learns to recognize them. But, how exactly does this work?

Similar to how humans learn, a model has to learn by its successes and mistakes, and we often call this the training phase. At the beginning of a training phase, a classification model typically performs very badly.

The model will look at an image, try to classify it as a car or not-car and monitor the errors that it makes. For example, if a model mistakenly classifies a car as not-car, it will learn from this mistake, tweak its classification parameters and improve its performance each time it sees more images.

After many iterations, the model converges on the right set of parameters and the error rate becomes low. That’s when we consider the model to be trained.

A training flow is pictured below. This is a [convolutional neural network](http://cs231n.github.io/convolutional-networks/) that learns to recognize and distinguish between images of cars and not-cars.

A screenshot of a cell phone

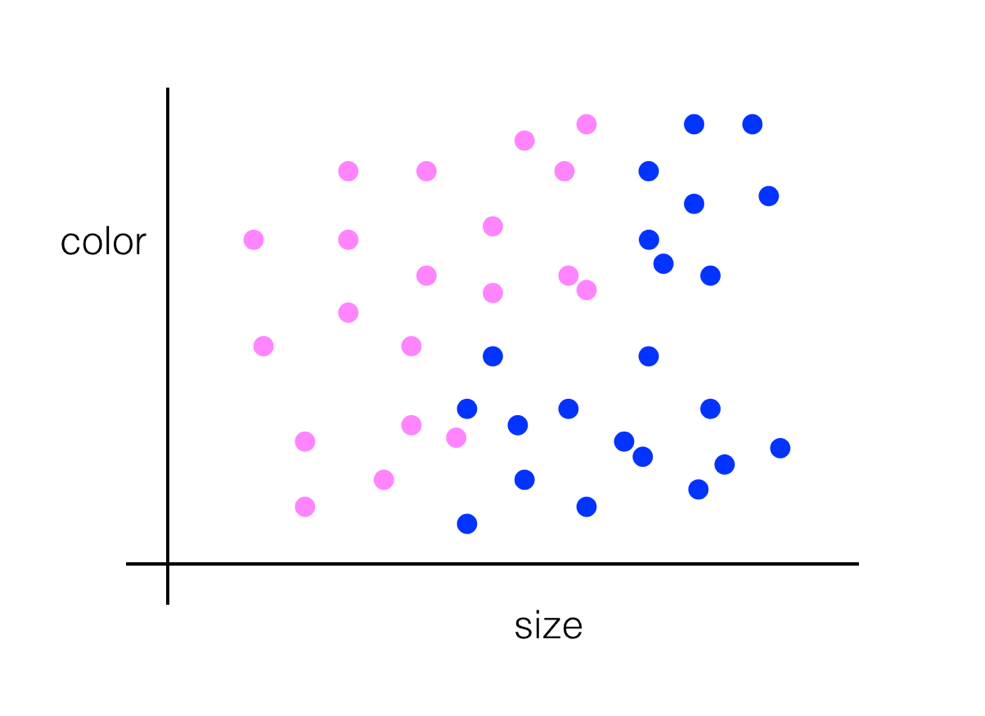
Description automatically generated

A convolutional neural network will adjust its pattern recognition algorithm until it learns to accurately classify a set of images.

Now, this is a very high-level view of how to train any classification model. And the details will vary based on the type of model you use and the training algorithm you choose.

## Separating Data

Say you want to separate two types of image data: images of bikes and of cars. You look at the color of each image and the apparent size of the vehicle in it and plot the data on a graph. Given the following points (pink dots are bikes and blue are cars), how would you choose to separate this data?



Pink and blue dots representing images of cars and of bikes.

A close up of a map

Description automatically generated

## Layers of Separation

What if the data looked like this?

A close up of a logo

Description automatically generated

Pink and blue dots representing images of cars and of bikes.

You could combine two different lines of separation! You could even plot a curved line to separate the blue dots from the pink, and this is what machine learning learns to do — to choose the best algorithm to separate any given data.

A picture containing table, clock, group, computer

Description automatically generated

A picture containing photo, sitting, different, full

Description automatically generated

Separated data.

## Images as Numerical Data

Every pixel in an image is just a numerical value and, we can also change these pixel values. We can multiply every single one by a scalar to change how bright the image is, we can shift each pixel value to the right, and many more operations!

**Treating images as grids of numbers is the basis for many image processing techniques.**

Most color and shape transformations are done just by mathematically operating on an image and changing it pixel-by-pixel.