QUIZ QUESTION

Before you get started, do some research onApollo's github repo and find out which of the following are outputs for Apollo's perception module?

- [ ] Image collected from camera
- [x] Three-dimensional bounding boxes of other vehicles
- [ ] The velocity, angular velocity and heading of our autonomous vehicle
- [x] The velocity, angular velocity and heading of detected vehicles
- [x] Lane lines and other road markings
- [ ] Information collected by LiDAR and Radar

Detection Classification Tracking Segmentation

such as road, car, or sky.

Pipline

Input Data	Pre-processing	Feature	Classification Model
Images	Resizing Images Rotating Color Transformation	Shape Wheels	Determine the Class

Car

bicycle, pedestrian or no such object at all.

QUIZ QUESTION

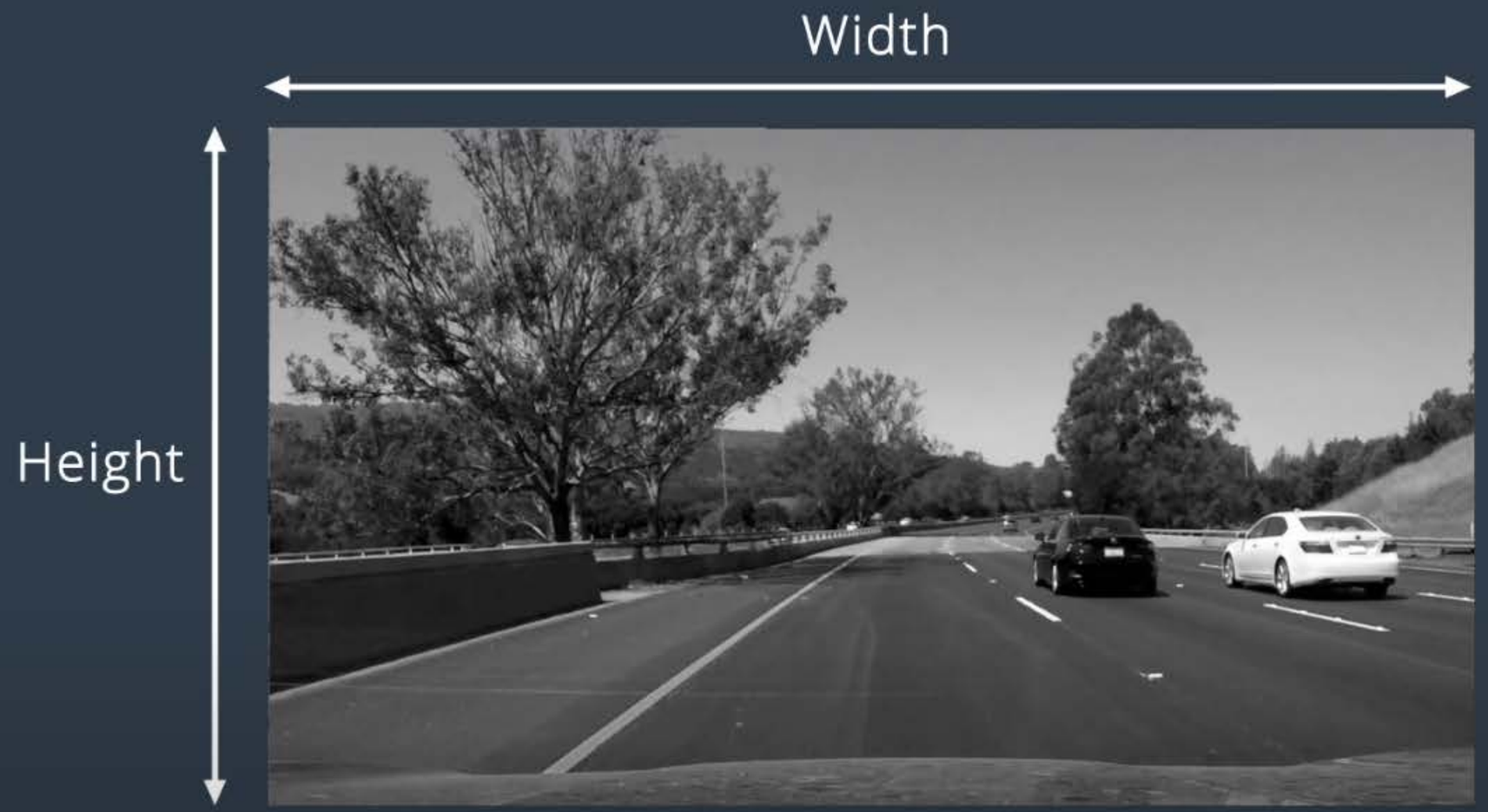
In the following quiz, your task is to identify the corresponding computer vision tasks of each of the perception scenarios on the road.

Tracking | Classification | Detection | Tracking | Detection

SCENARIOS	TYPE OF COMPUTER VISION TASK
Determine which pixels in the image captured by the camera correspond to the travelable area.	Segmentation
Determine if the obstacle in a picture is a pedestrian or a biker.	Classification
Recognize the status of a traffic light.	Segmentation
Differentiating multiple cars on the road in a sequence of continuous driving record frames.	Classification

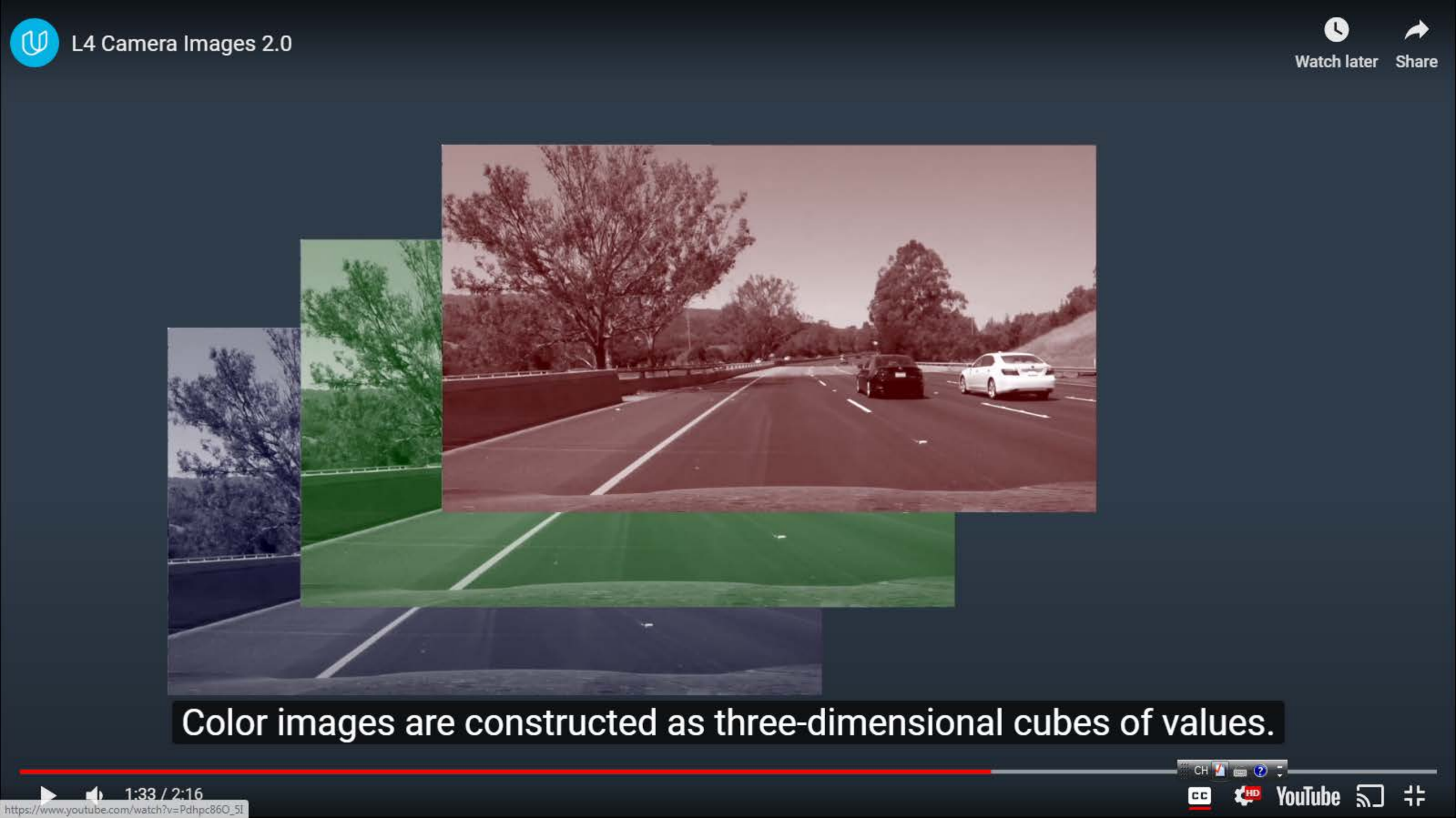
Distinguish lanes and road signs in driving record images	Classification
Mark a dangerously driven vehicle on the road	Tracking
Find all pedestrians in the driving record image	Detection

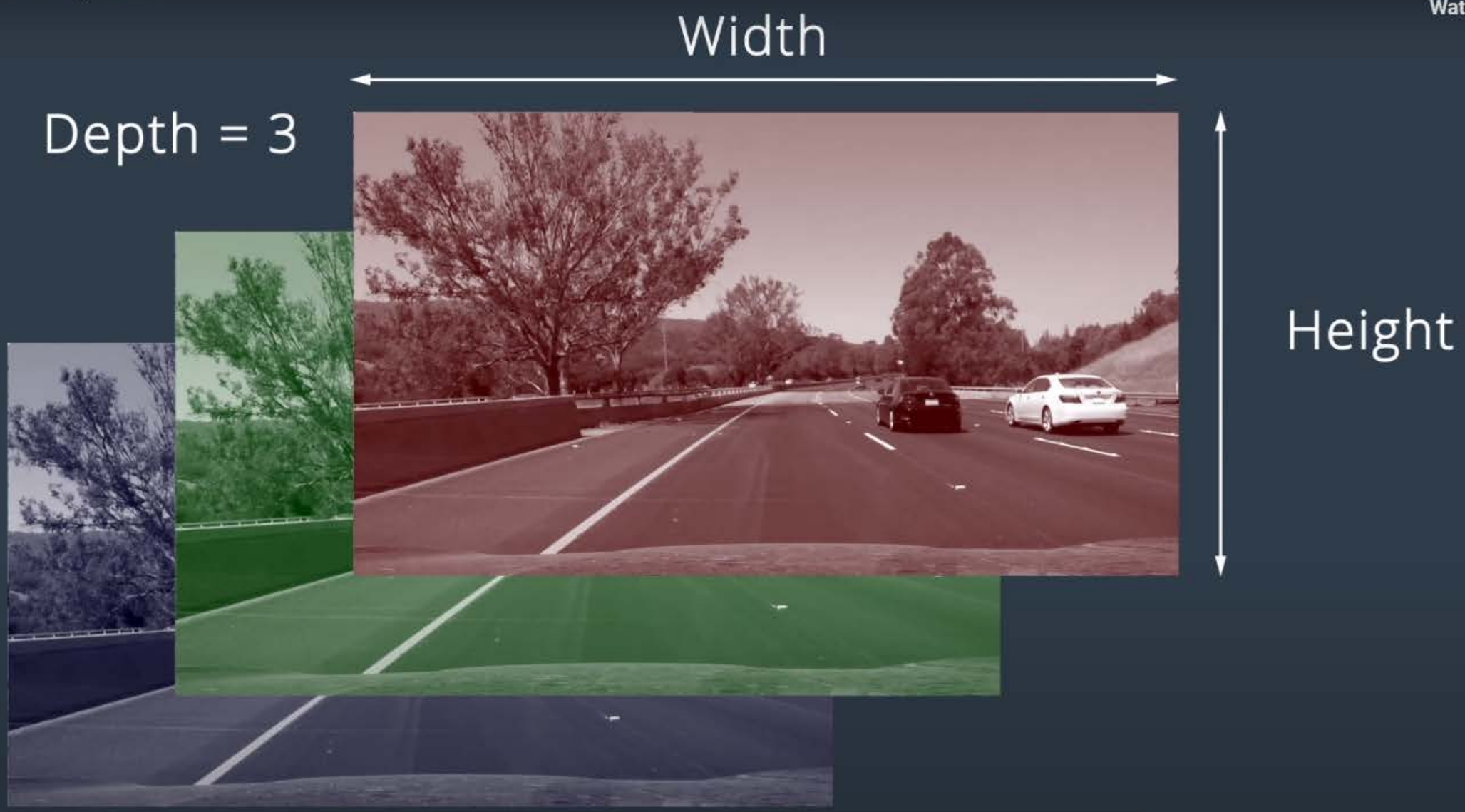

"Oh, there is a car in the image!"
also known as a matrix.
0:26 / 2:16
YouTube



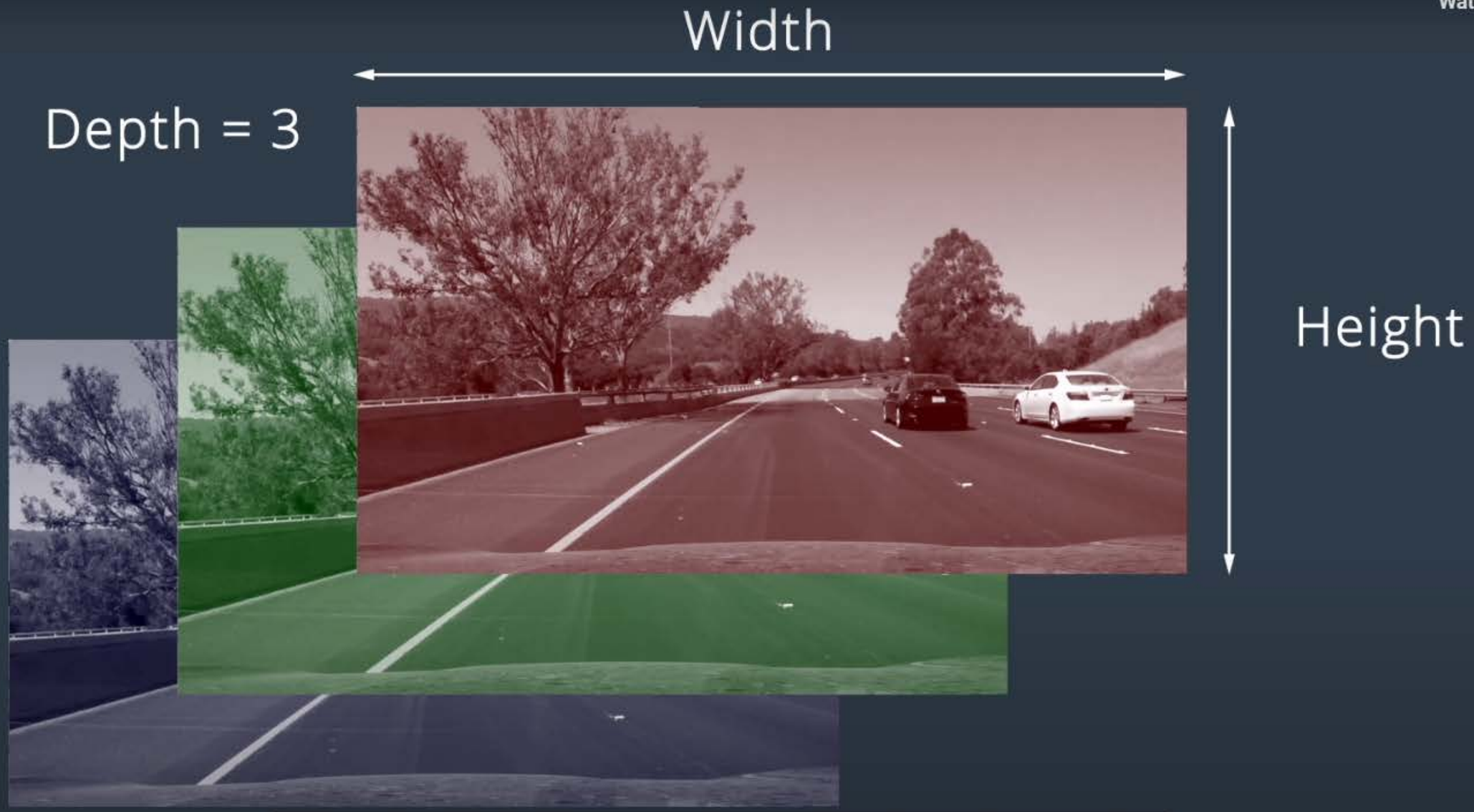
Grayscale Image

Color images are similar but a little more complex.

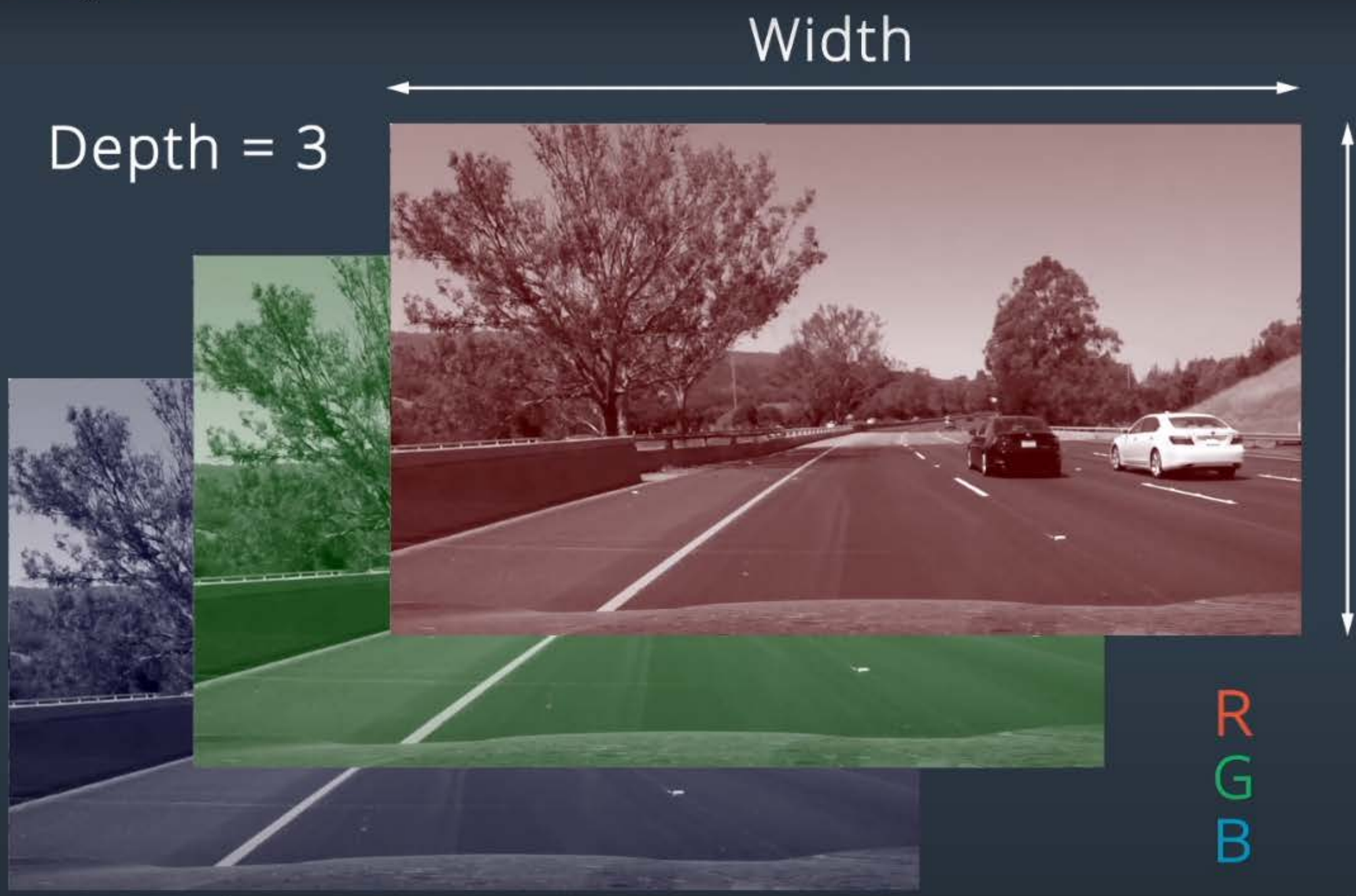

L4 Camera Images 2.0
Watch later
Share
Color images are constructed as three-dimensional cubes of values.
1:33 / 2:16
YouTube



Each cube has a height, width and depth.



The depth is the number of color channels.

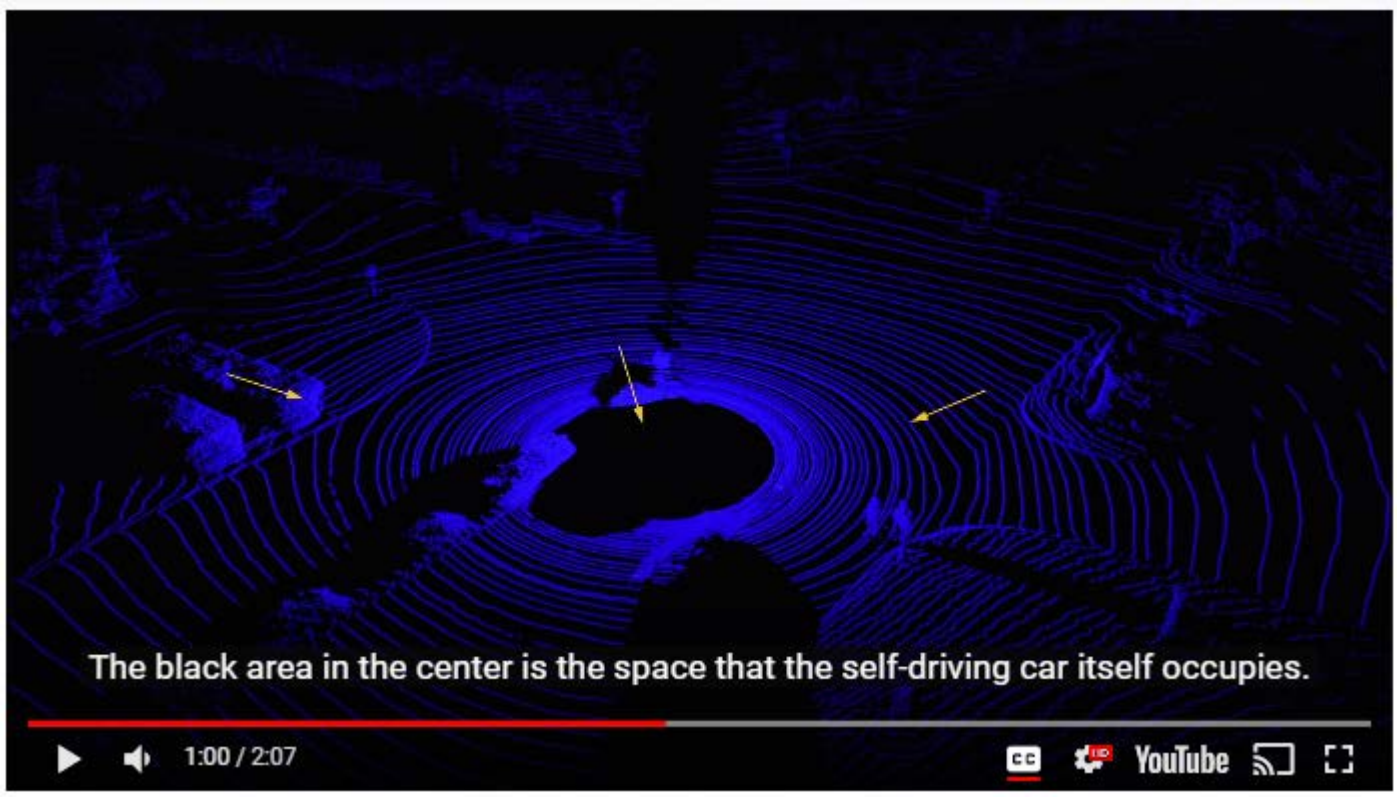


combinations of three colors: red, green and blue.

QUIZ QUESTION

A combination of RGB values can represent a color. Which of the following RGB values is the most likely to represent the following colors?

- (2, 179, 228)
- (236, 200, 26)
- (2, 194, 107)


The black area in the center is the space that the self-driving car itself occupies.
1:00 / 2:07
YouTube

QUIZ QUESTION

How many dimensions do we need to describe the position of a point in a LiDAR image?

- 2
- ✓ 3
- 4

SUBMIT

Footnote on Lidar

There are other possibilities to scan the laser beams. Instead of rotating the lasers or having a rotating mirror, we can scan the lidar with a vibrating micromirror. Those lidars are in development but none are commercially available now (as of March 2017).

Instead of mechanically moving the laser beam, a similar principle to phased array radar can be employed. Dividing a single laser beam into multiple waveguides, the phase relationship between the waveguides can be altered and thereby the direction of the laser beam shifted. The advantage is that the form factor can be much smaller and that there are no moving parts.

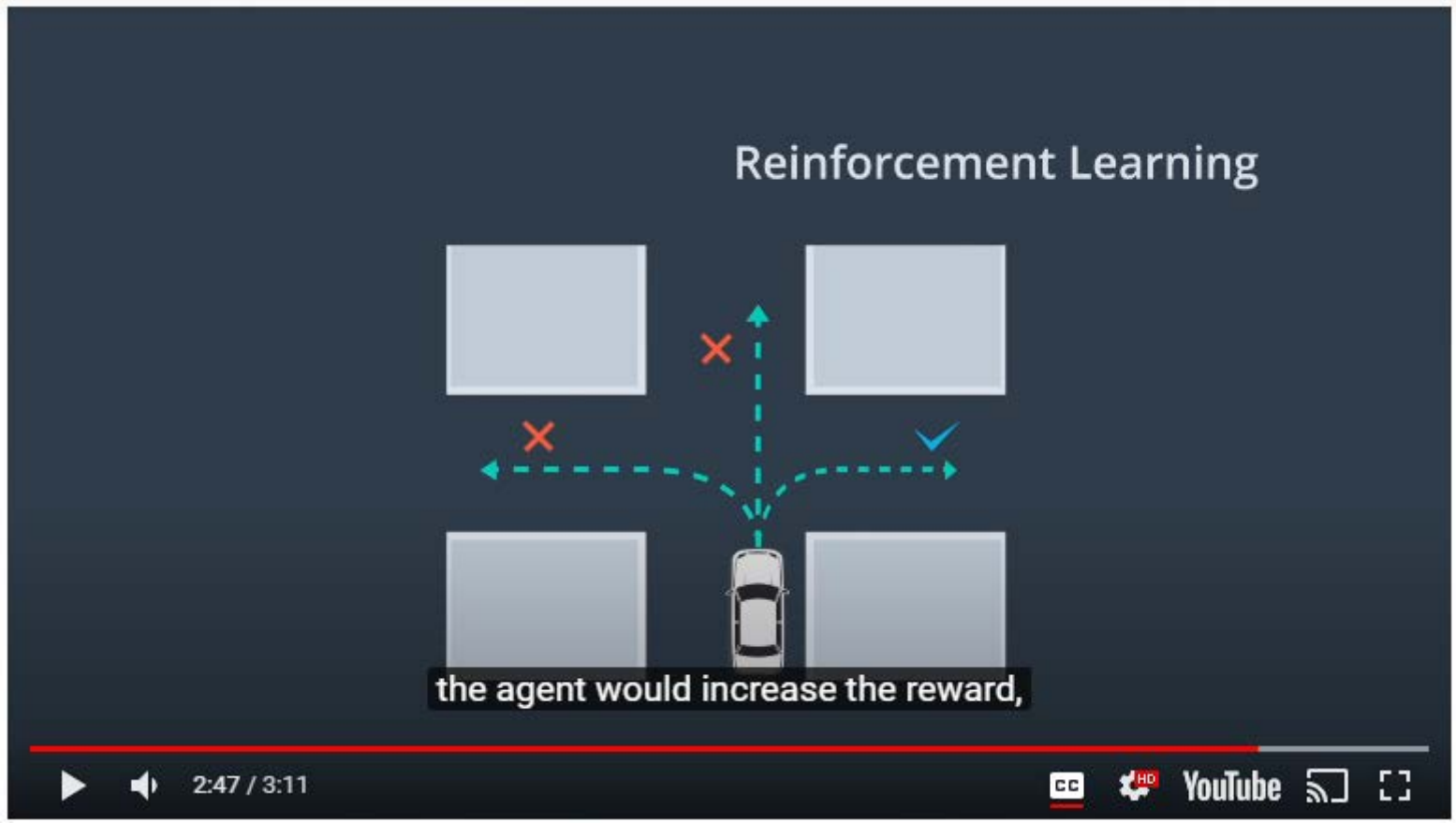
Another possibility is to use the laser as a gigantic flash like with a camera and then measuring the arrival times for all the objects with one big imaging photodiode array. This is in effect a 3D camera. The components are currently very expensive and currently this is used more in space and in terrain mapping applications.


Machine Learning Application
and doctors even use machine learning to assist with medical diagnoses.
0:47 / 3:11
YouTube

the model is taking advantage of ground truth labels created by humans.


Unsupervised Learning
Here we don't provide the ground truth labels,
1:39 / 3:11
YouTube


Unsupervised Learning
the computer learns on its own to recognize distinctions.
1:50 / 3:11
YouTube


Reinforcement Learning
the agent would increase the reward,
2:47 / 3:11
YouTube

QUIZ QUESTION

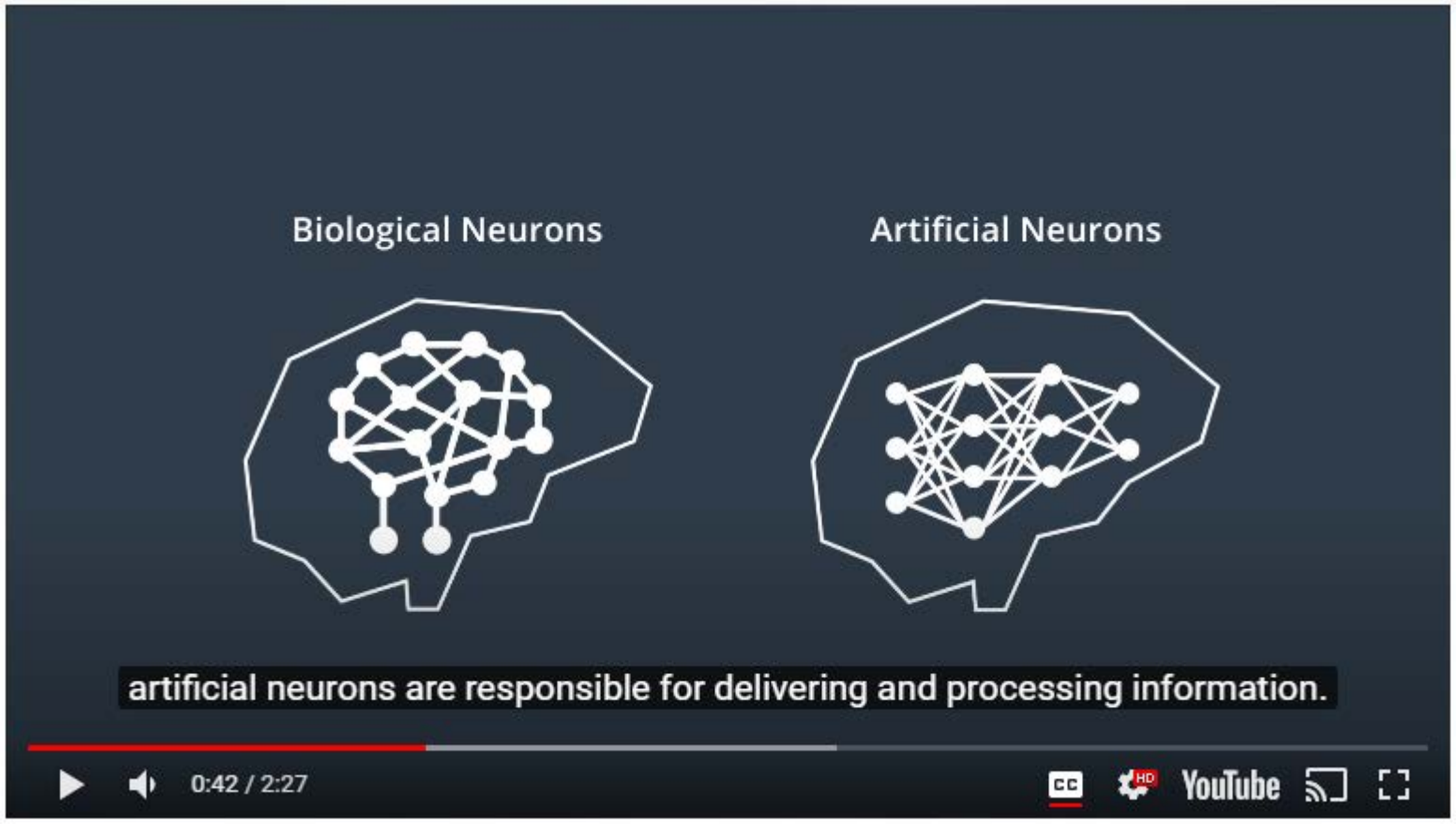
Which of the following choices could be best matched ?

Submit to check your answer choices!

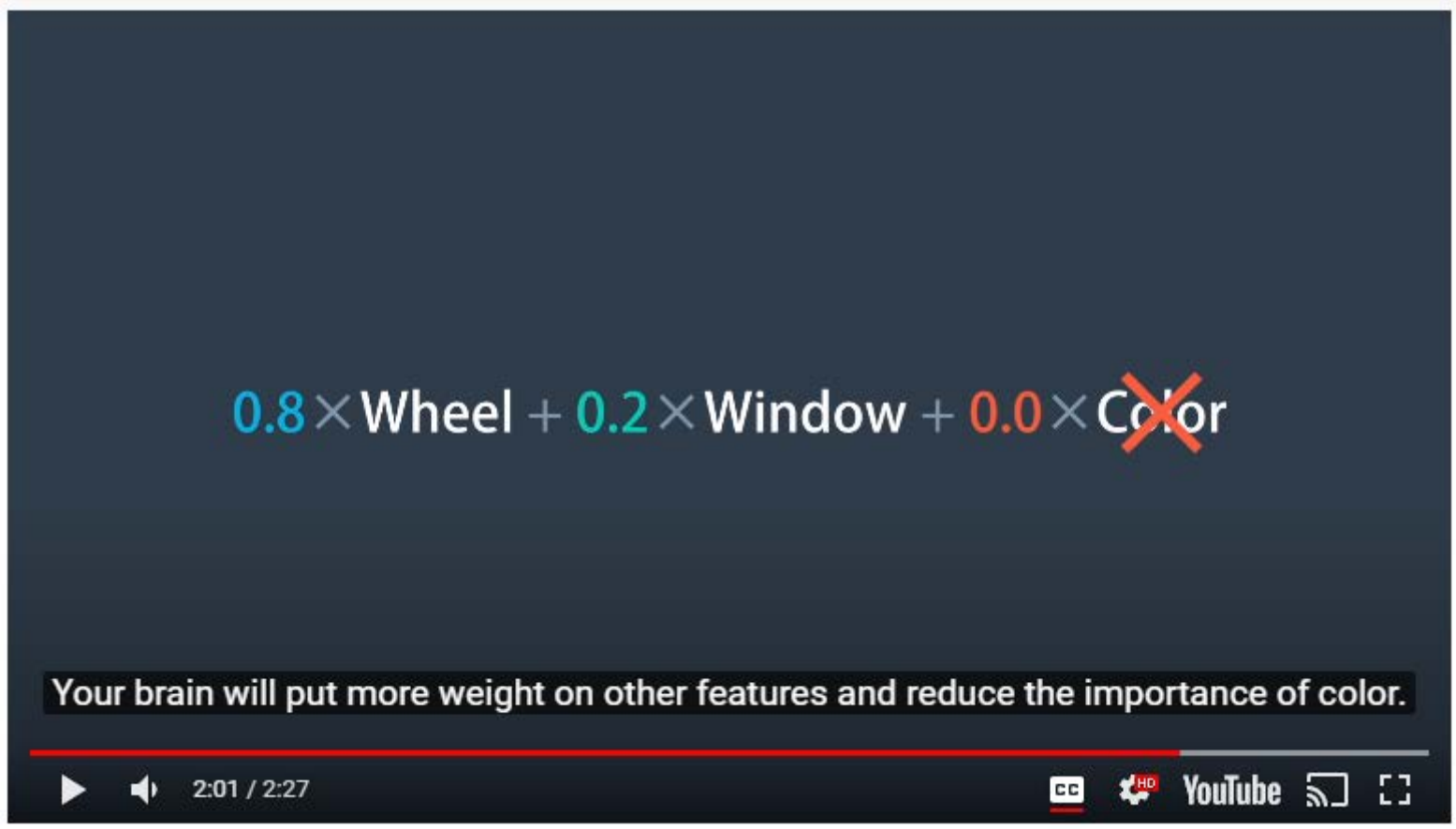
OPTIONS	TYPES OF MACHINE LEARNING
Data without labels	Unsupervised Learning
Summarized and improved from the Experience	Reinforcement learning
Data with corresponding labels	Supervised Learning
Parts data with labels and the other without labels	Semi-supervised learning

Biological Neurons

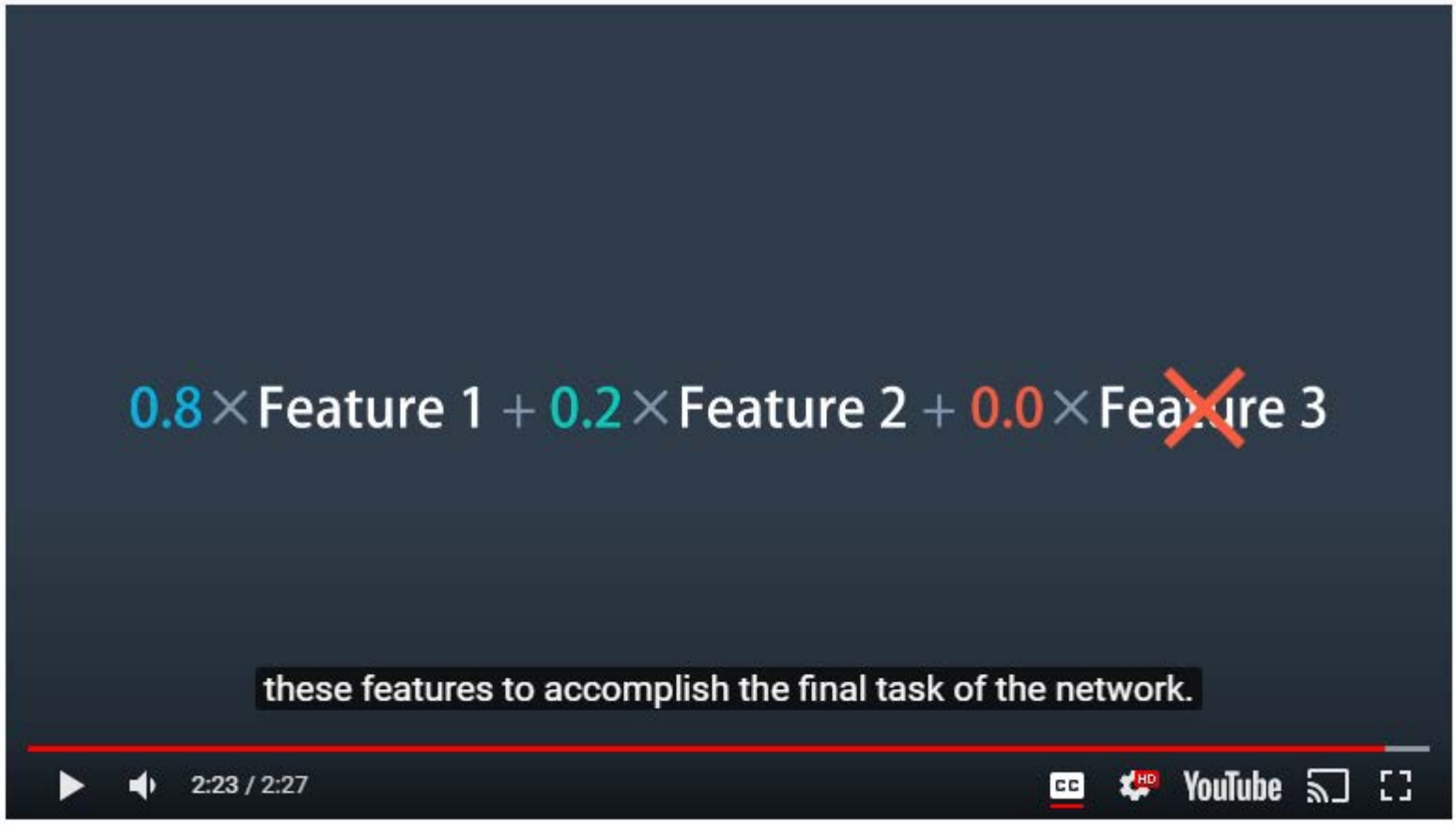
artificial neurons to create artificial neural networks for machine learning.


Biological Neurons
Artificial Neurons
artificial neurons are responsible for delivering and processing information.
0:42 / 2:27
YouTube


Window
Wheel
Color
Blue
windows and colors, then your brain would
1:41 / 2:27
YouTube


0.8 × Wheel + 0.2 × Window + 0.0 × Color
Your brain will put more weight on other features and reduce the importance of color.
2:01 / 2:27
YouTube

an image but these might be features that we humans can't describe or even understand.


0.8 × Feature 1 + 0.2 × Feature 2 + 0.0 × Feature 3
these features to accomplish the final task of the network.
2:23 / 2:27
YouTube

Thanks for completing that!

Awesome! We take pixel values as the most basic features, and extract higher level features from them!

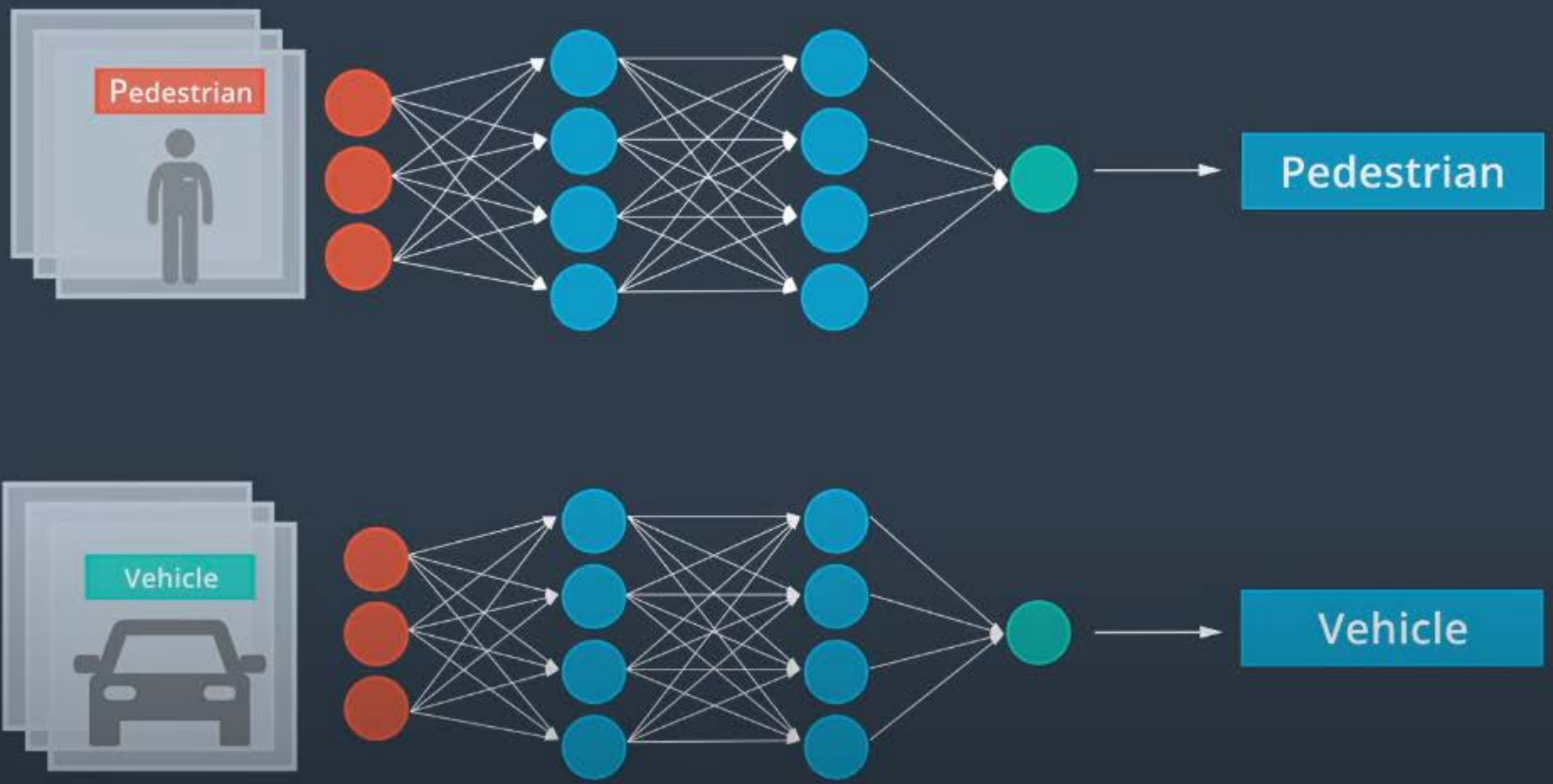
CONTINUE

QUIZ QUESTION

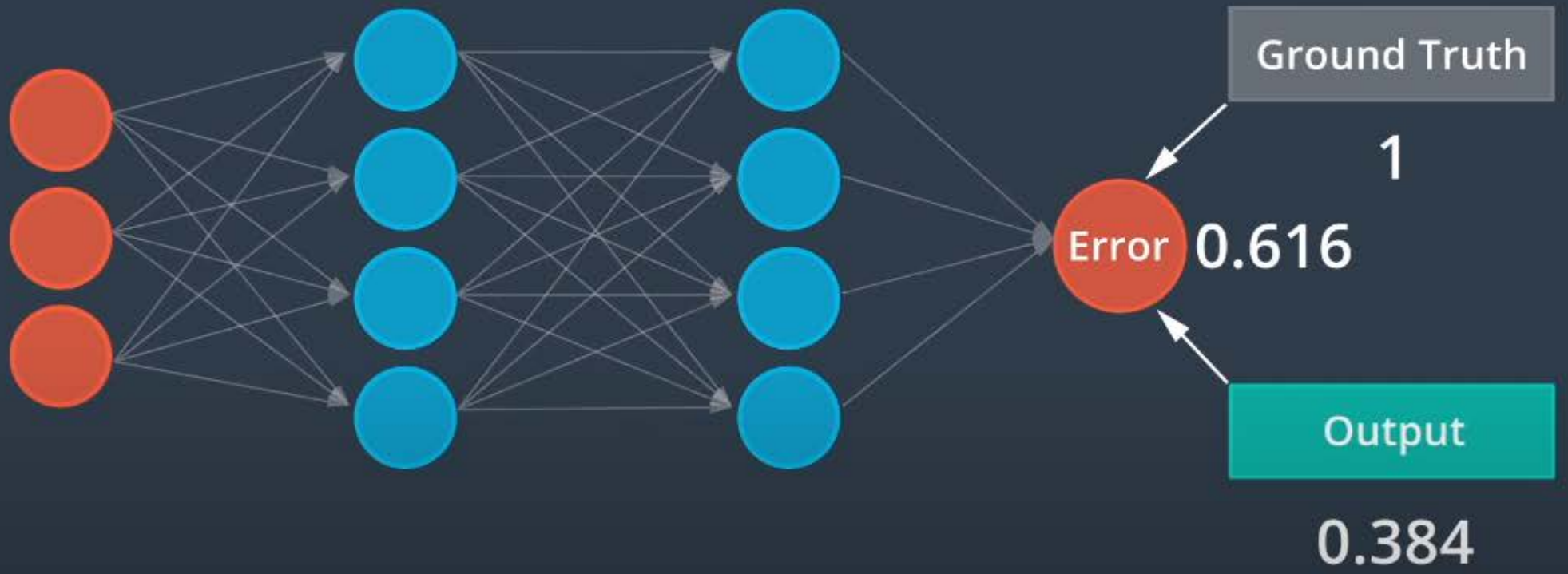
For neural networks, what is the most basic representation of features of an image?

- The contour lines of objects in the image.
- ✓ The pixel values of the image.
- Every bit of the image.
- You can always find smaller features

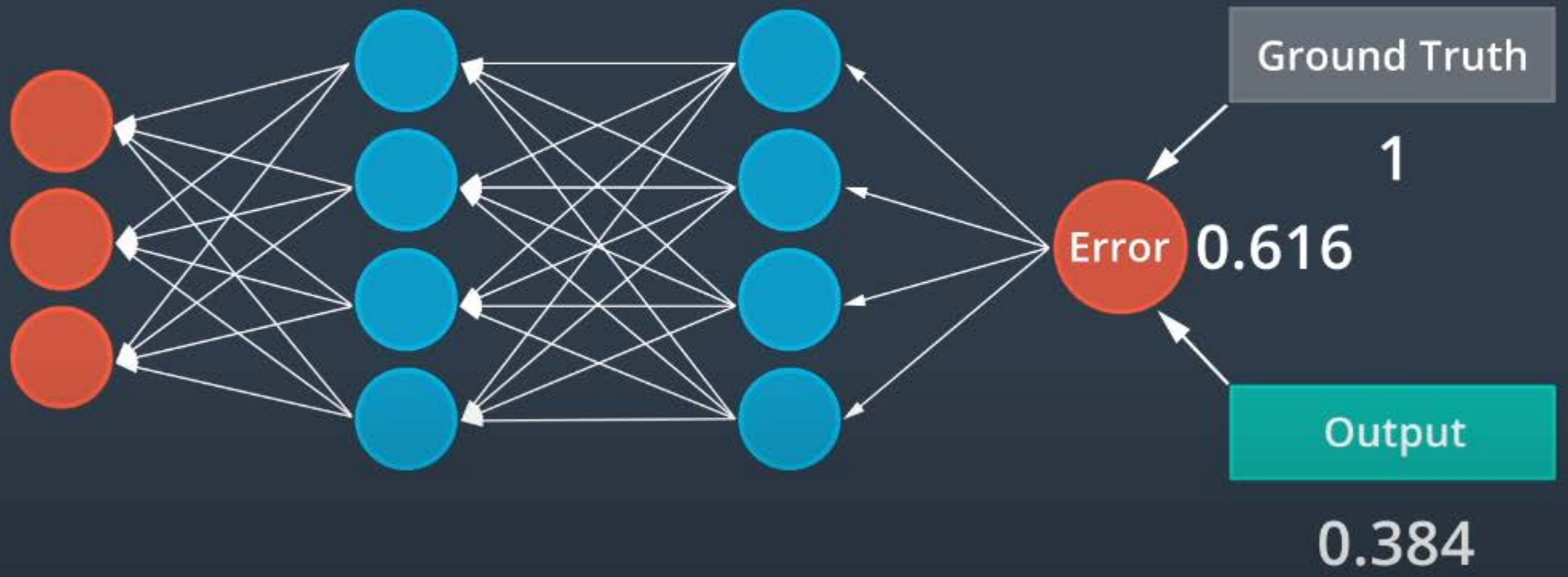

Feed forward, error measurement and backpropagation.
0:13 / 1:24
YouTube



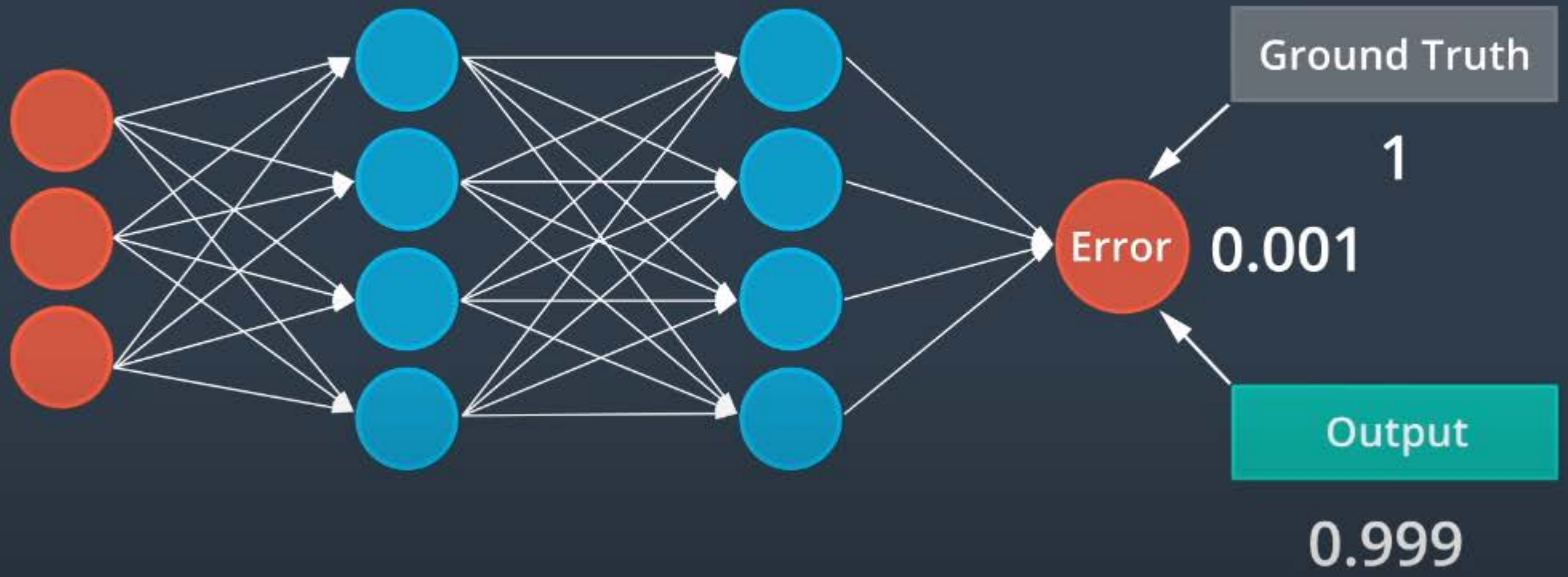
This is called feed forward.



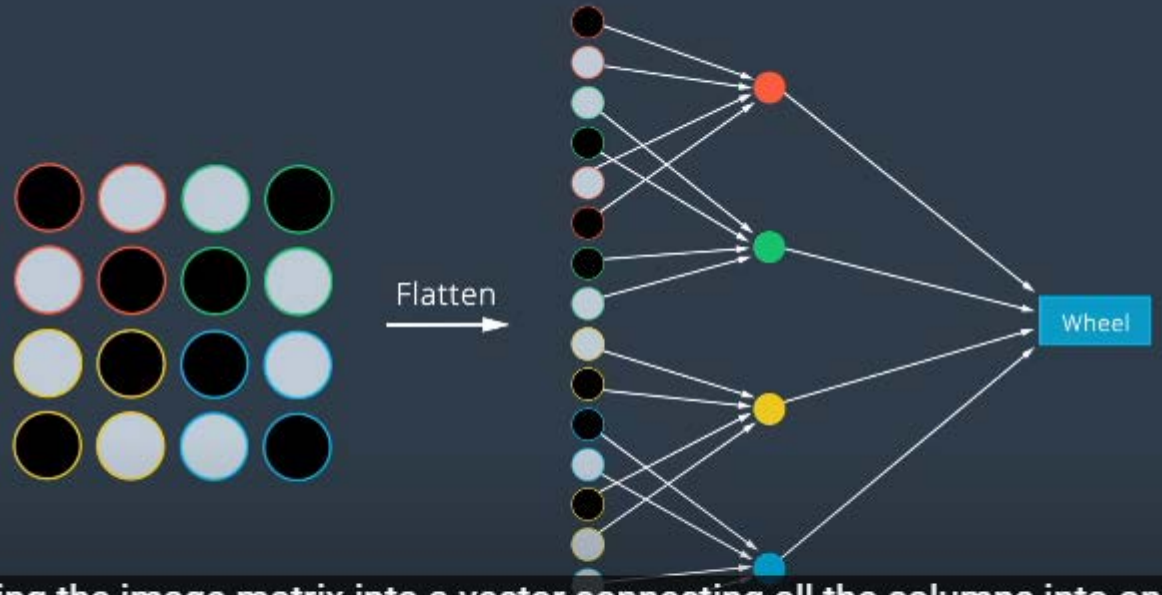
the ground truth label and the output generated by the feed-forward process.



the neural network like the feed-forward process, but in reverse.

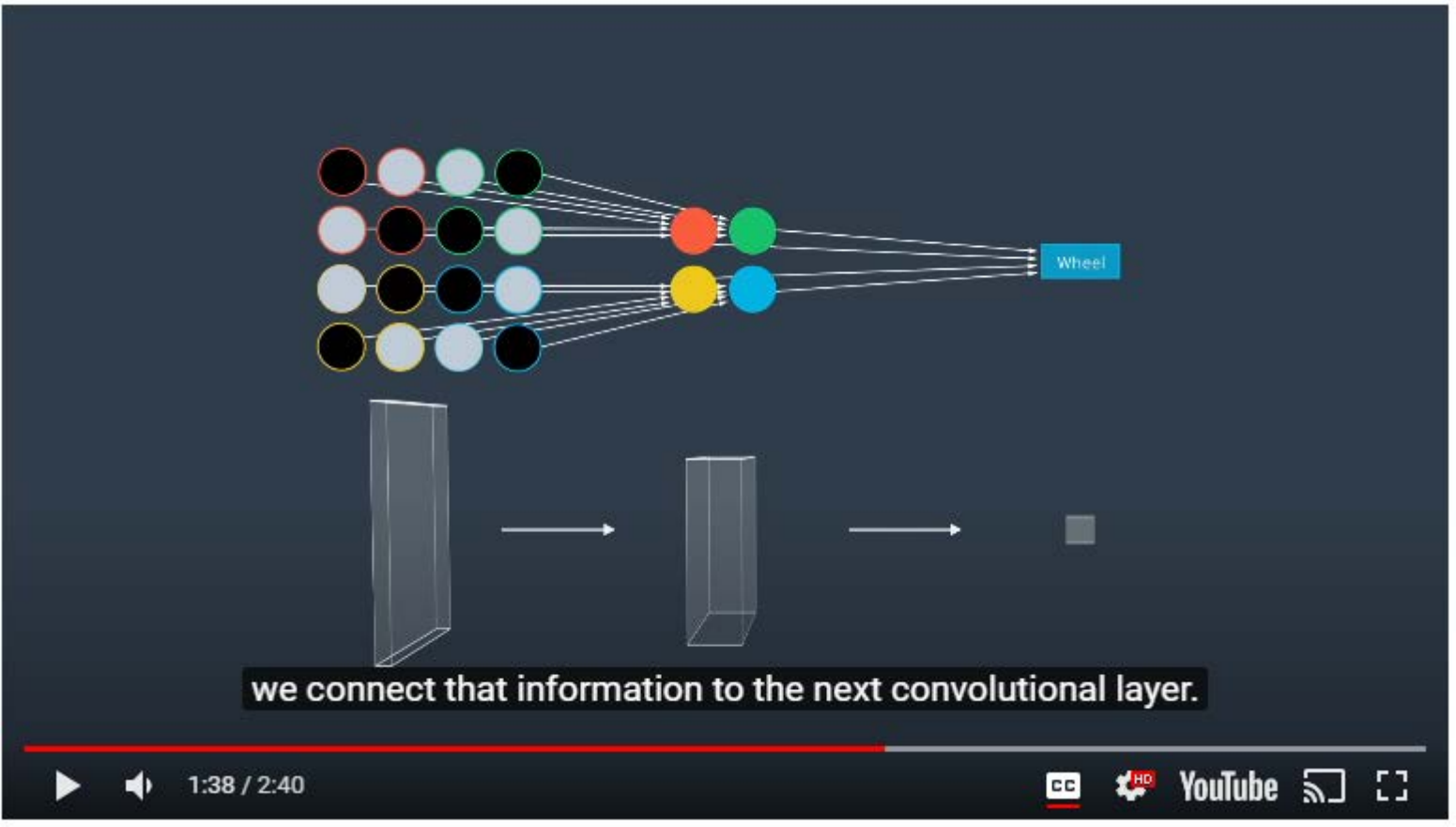


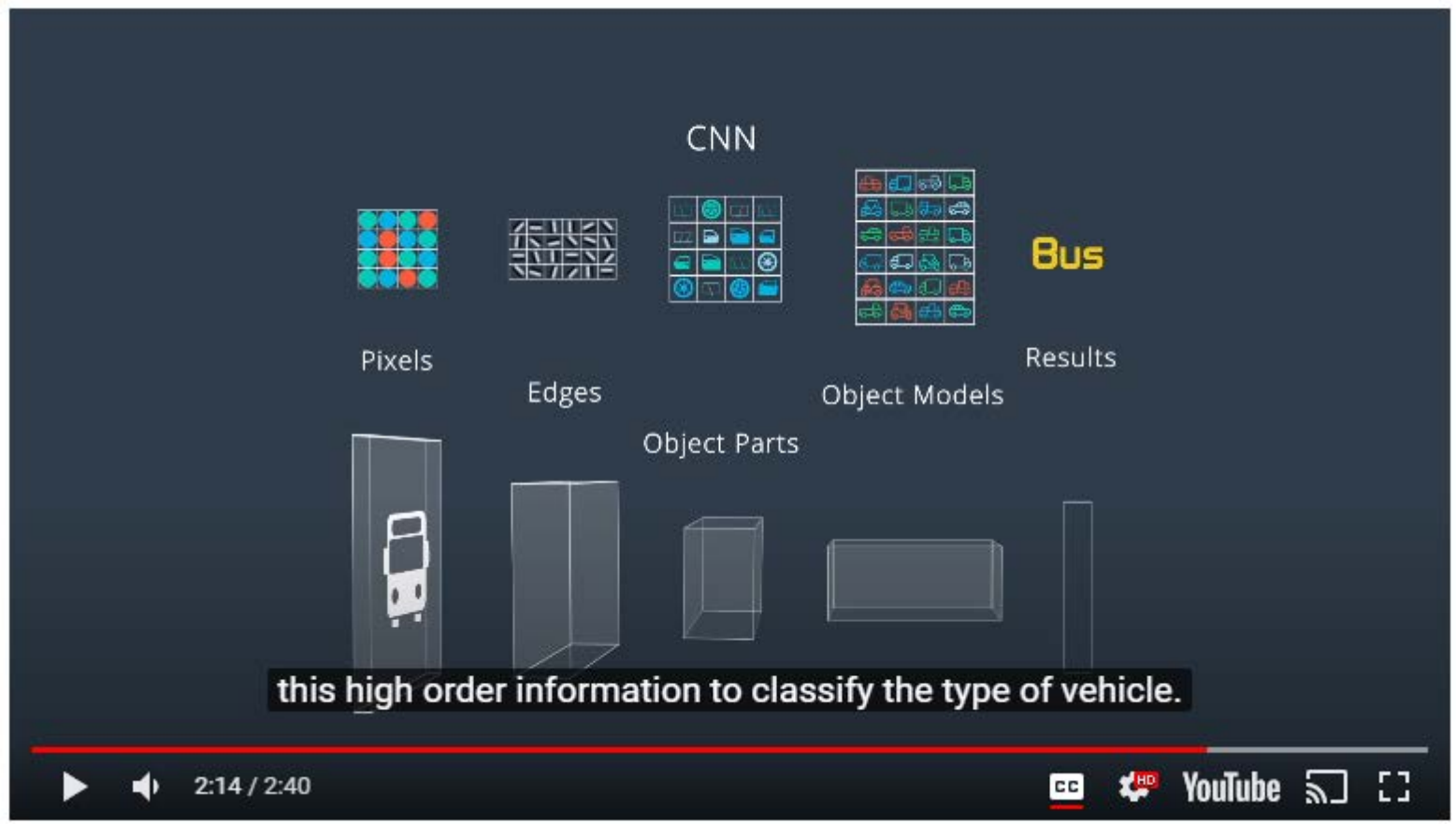
make accurate predictions based on new data.



reshaping the image matrix into a vector connecting all the columns into one giant row.

As we convolve a filter over the entire input image,


Wheel
we connect that information to the next convolutional layer.
1:38 / 2:40
CC
HD
YouTube


CNN
Bus
Pixels
Edges
Object Parts
Object Models
Results
this high order information to classify the type of vehicle.
2:14 / 2:40
YouTube

High-level Features

When you see the picture below, what features do you use to recognize it as a car?

Your reflection

a

Things to think about

In general, we recognize a car by looking at its shape and various components, such as tires, lights, windows, doors, etc.

High-level Features

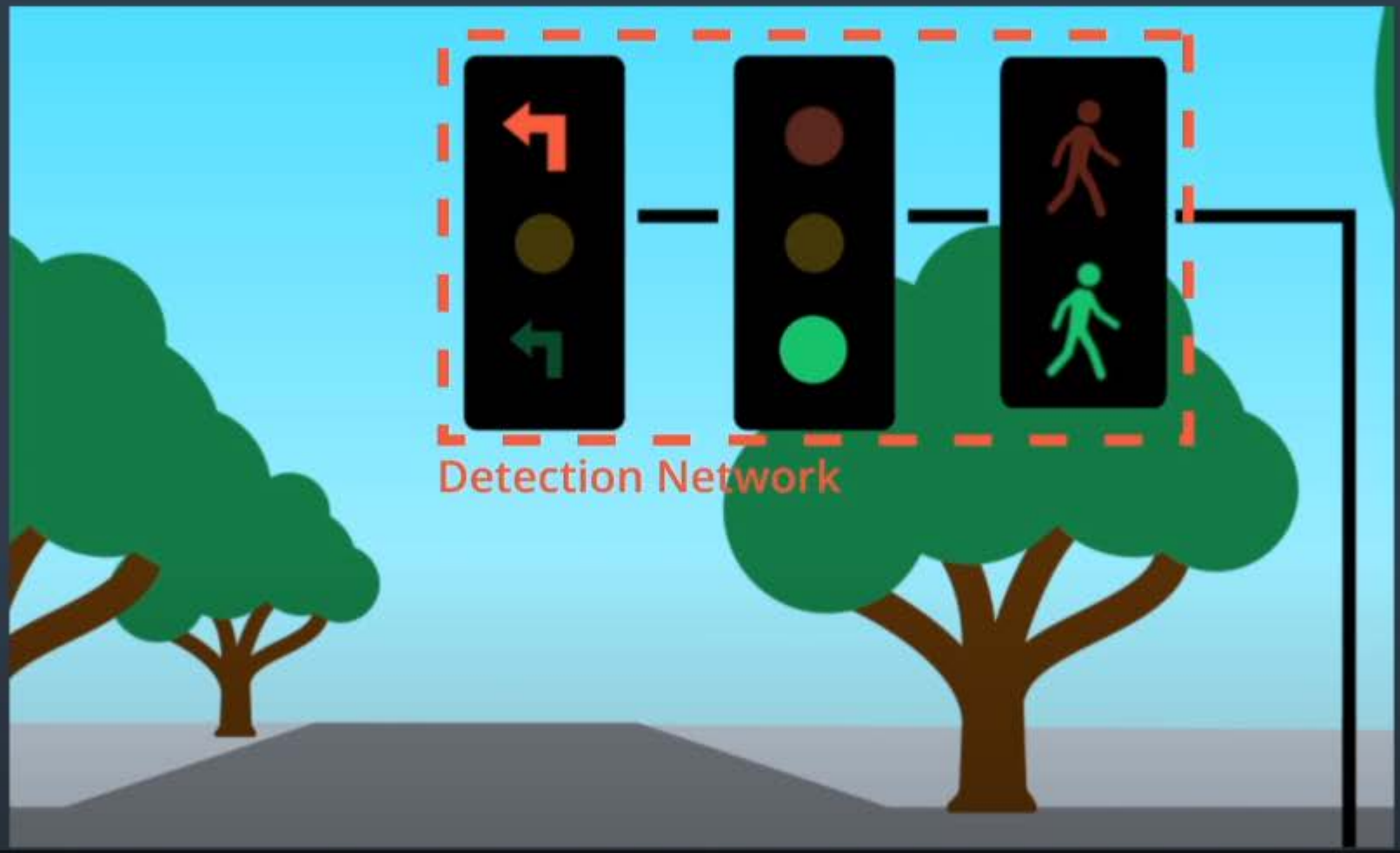
The image is a matrix of pixel values. How can we extract features such as shapes, tires, lights, windows, doors, etc. from a matrix of pixel values?

Your reflection

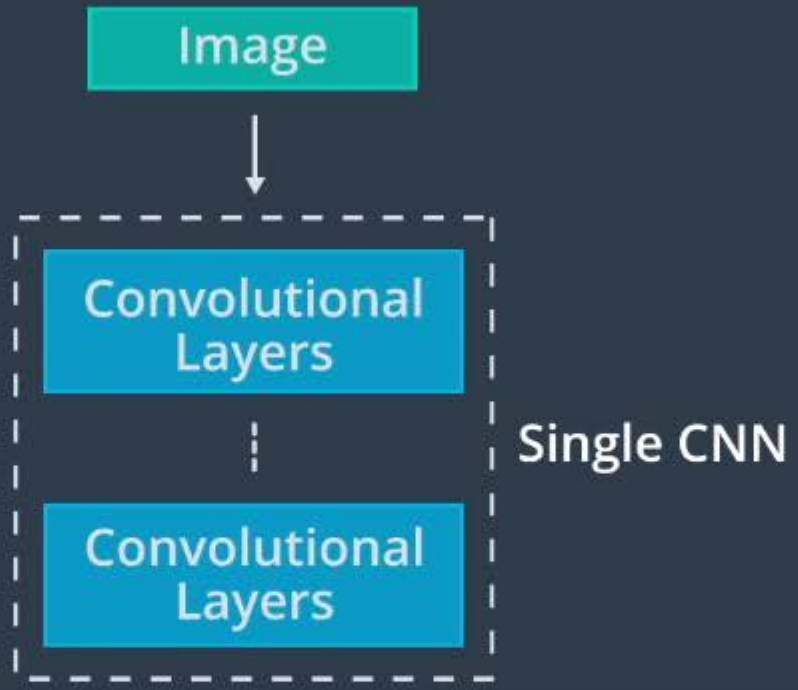
a

Things to think about

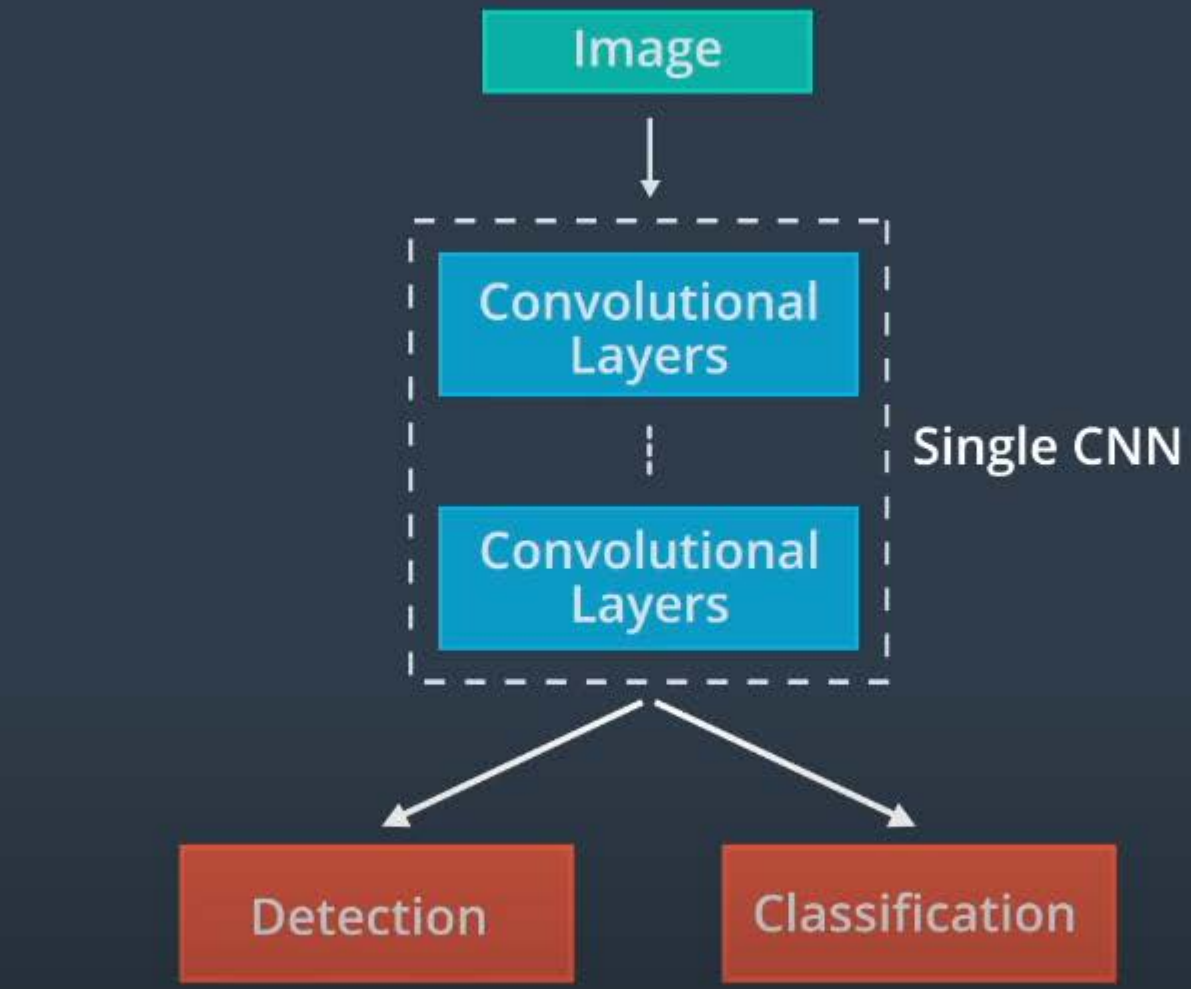
Thank you for your answer. Next, we will discuss a structure for extracting high-level features called convolutional neural networks. However, high-level features are not necessarily be descriptive like tires, windshield, etc.



First, we'll use computer vision to localize the traffic light within an image.



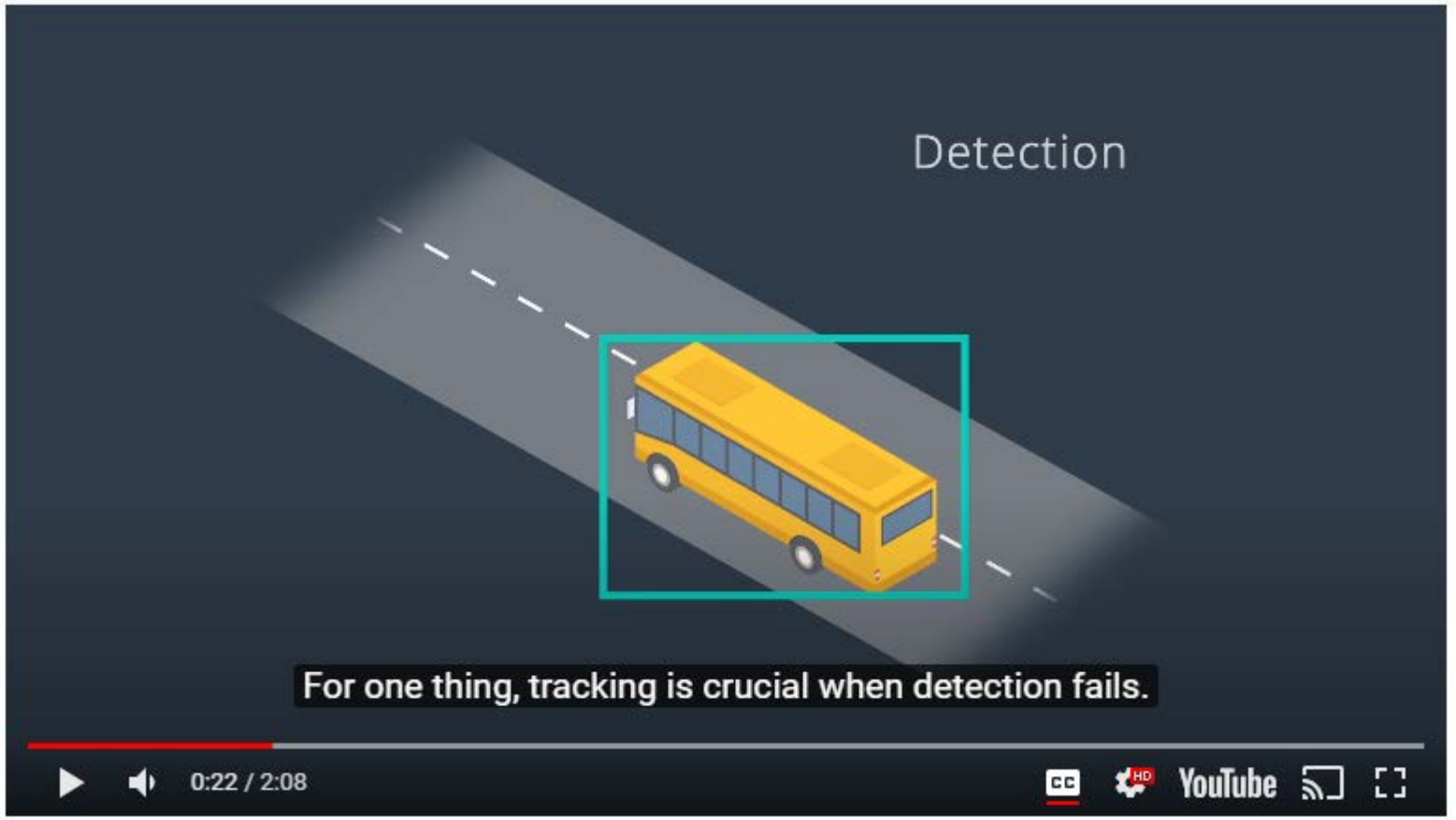
or we could do detection and classification using a single CNN architecture.

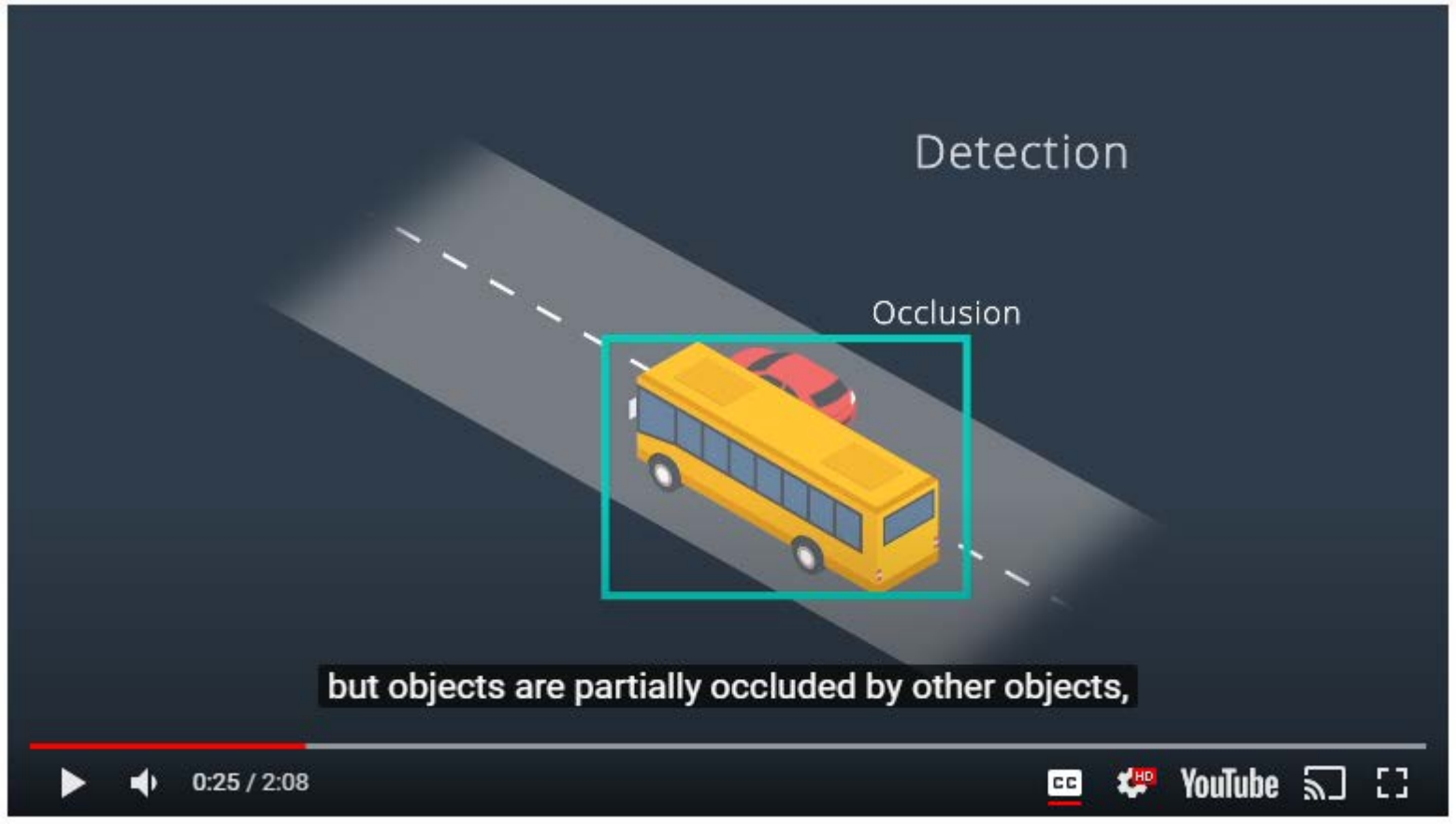


and another might perform classification.

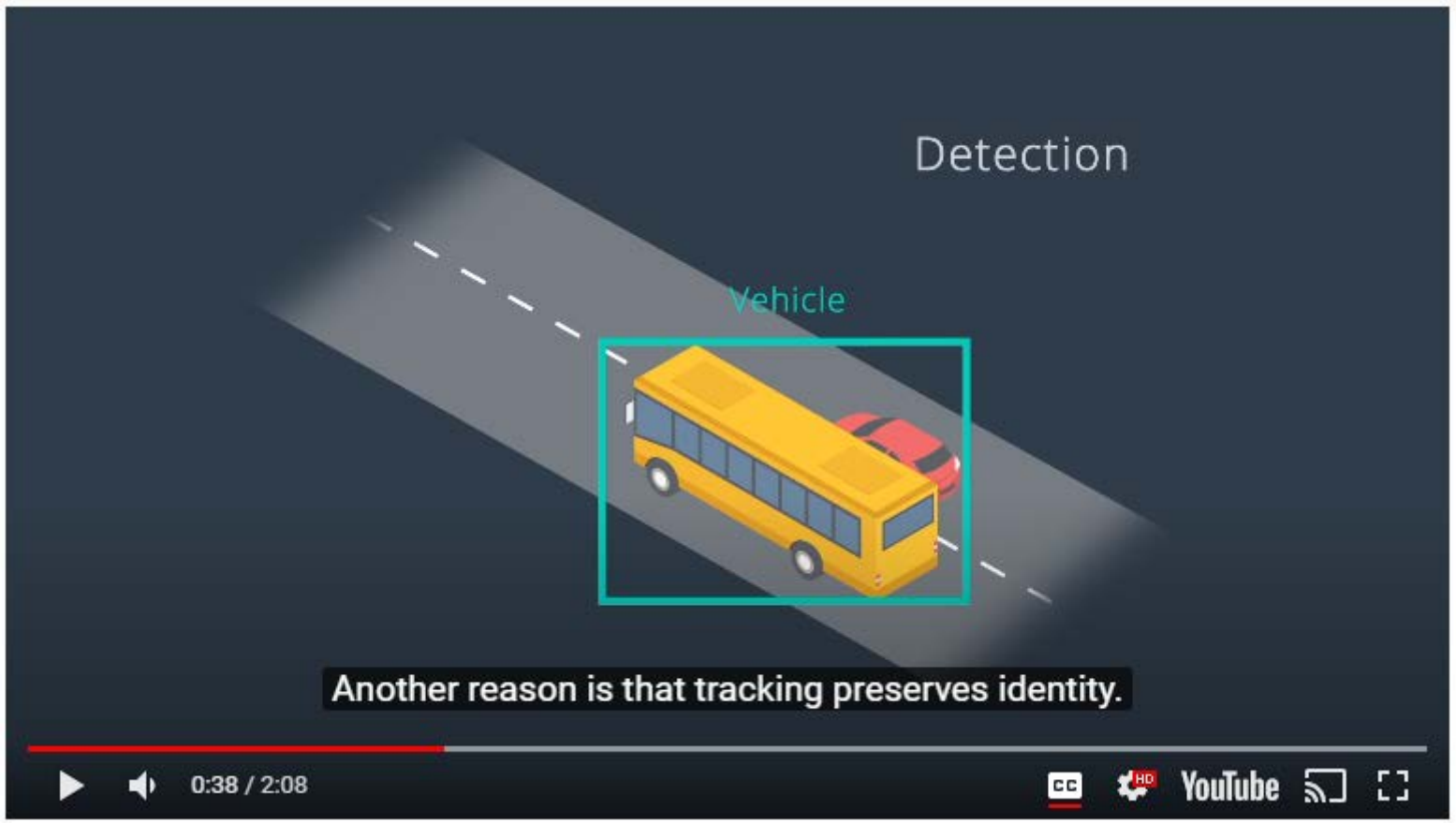

A classic architecture is R-CNN and its variants: Fast R-CNN and Faster R-CNN.
2:01 / 2:07
YouTube


YOLO and SSD are different architectures with similar forms.
2:06 / 2:07
YouTube

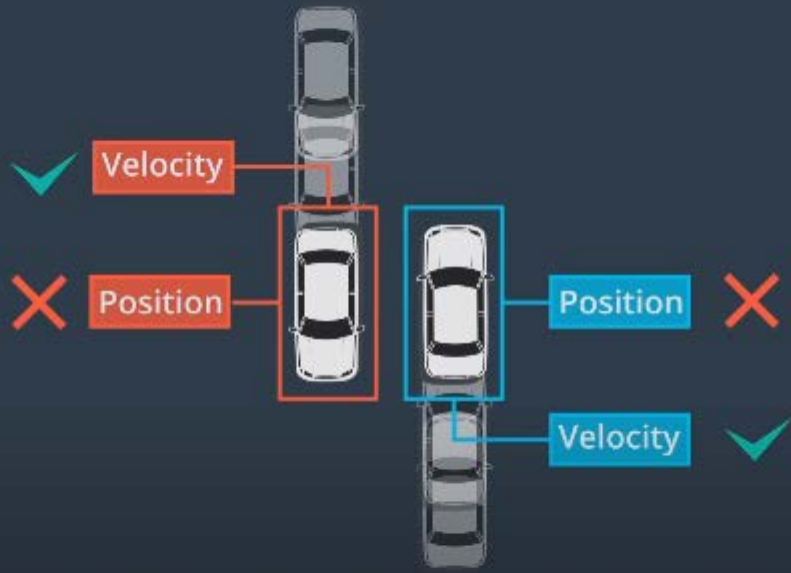

Detection
For one thing, tracking is crucial when detection fails.
0:22 / 2:08
YouTube


Detection
Occlusion
but objects are partially occluded by other objects,
0:25 / 2:08
YouTube


Tracking
Tracking handles occlusion.
0:33 / 2:08
YouTube


Detection
Vehicle
Another reason is that tracking preserves identity.
0:38 / 2:08
YouTube


The first step in tracking is identity conformation.
1:04 / 2:08
YouTube

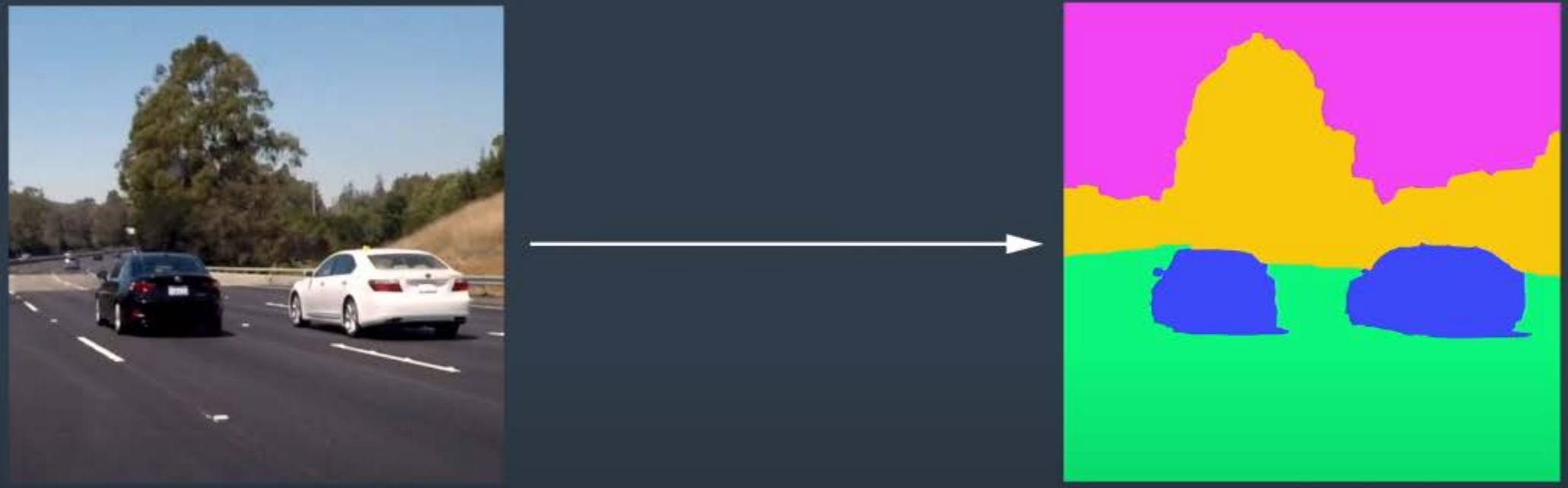


this information can also help us find matching objects quickly.

QUIZ QUESTION

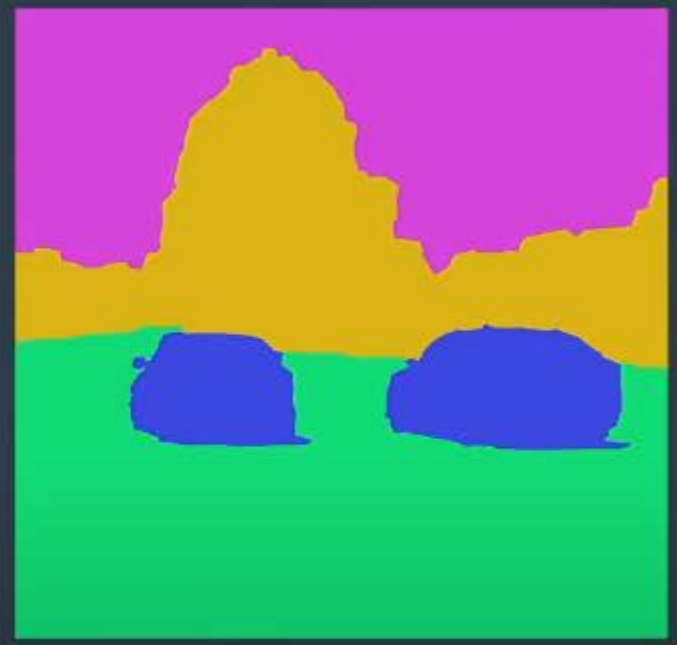
Which of the following is **NOT** handled by tracking?

- Measure similarity of detected objects across multiple frames
- Assign identifications for detected vehicle
- ✓ Predict the future behavior of a vehicle
- Estimate the position of a vehicle detected in previous frame but occluded in the current frame

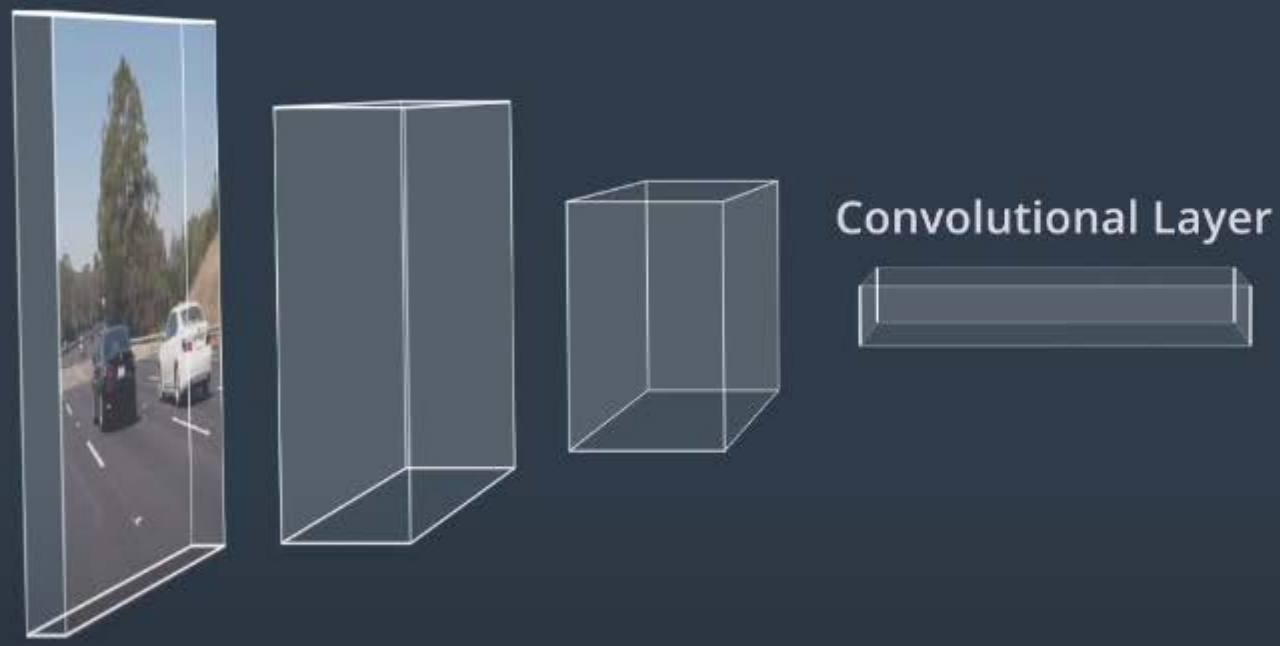


Semantic segmentation relies on a special type of CNN

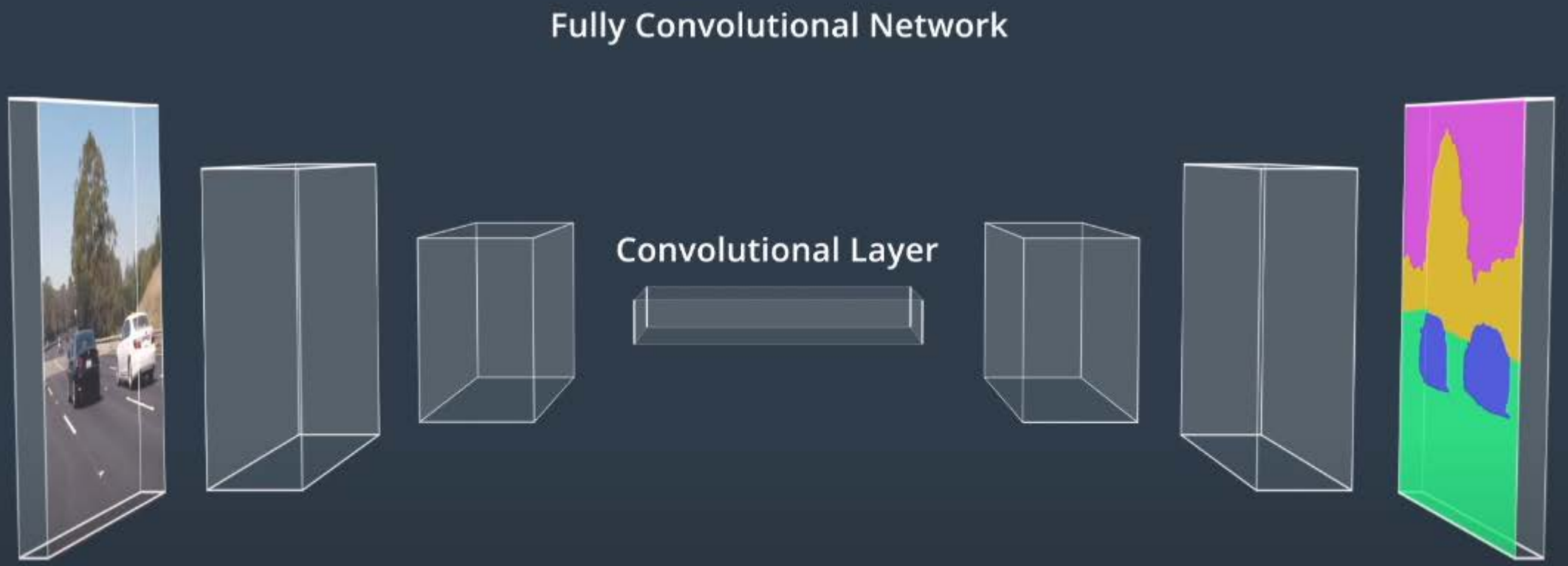
Fully Convolutional Network



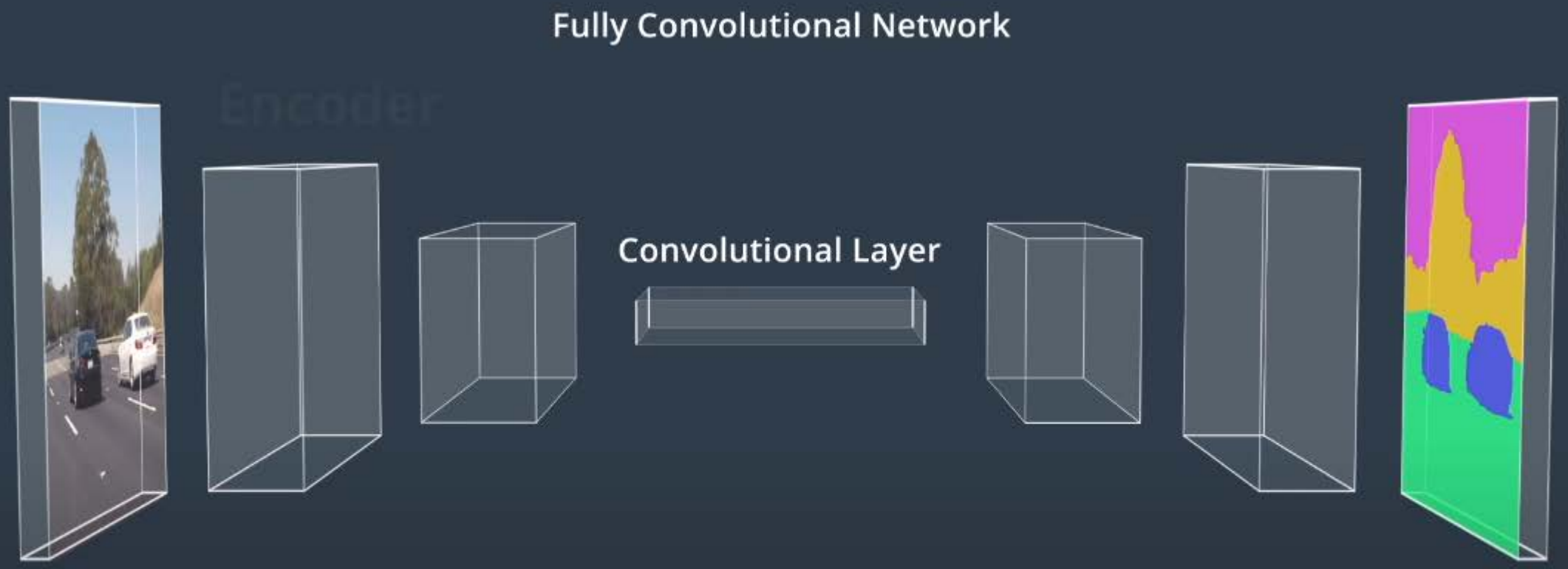
called a Fully Convolutional Network or FCN.



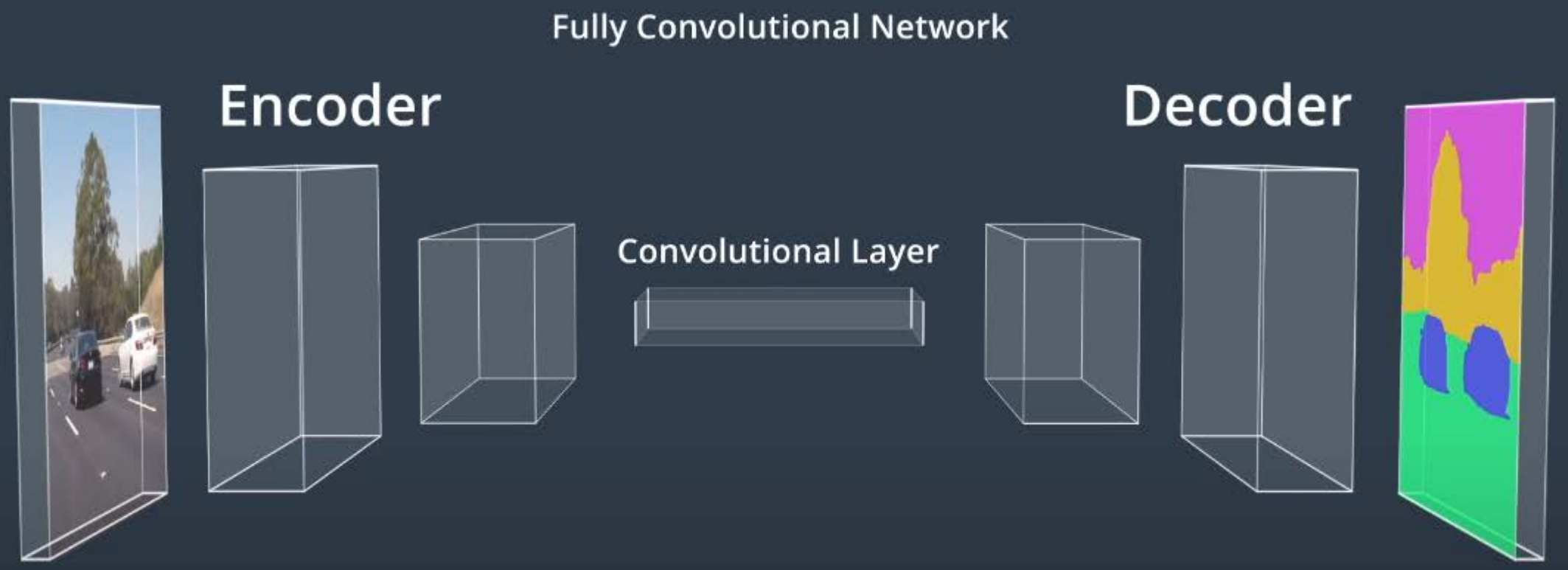
hence the name fully convolutional network.



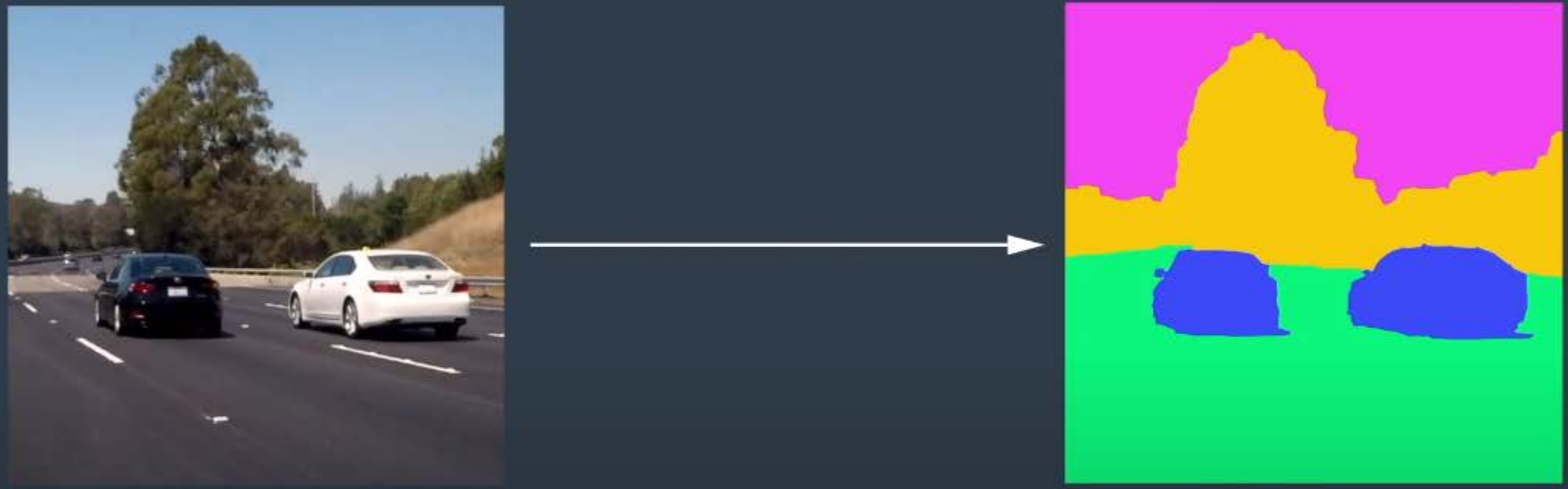
until the size of the final output matches the size of the original input image.



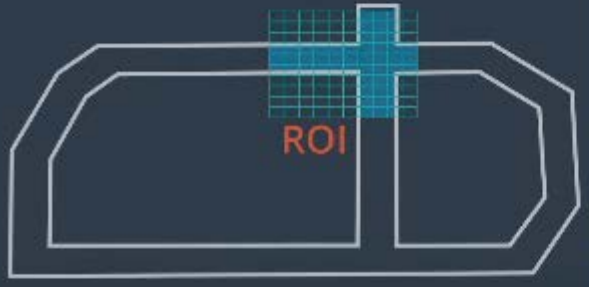
The first half of the network is often called the encoder because

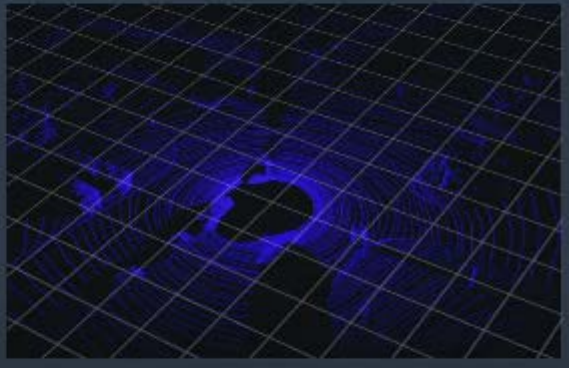


decoding those features and applying them to the output.

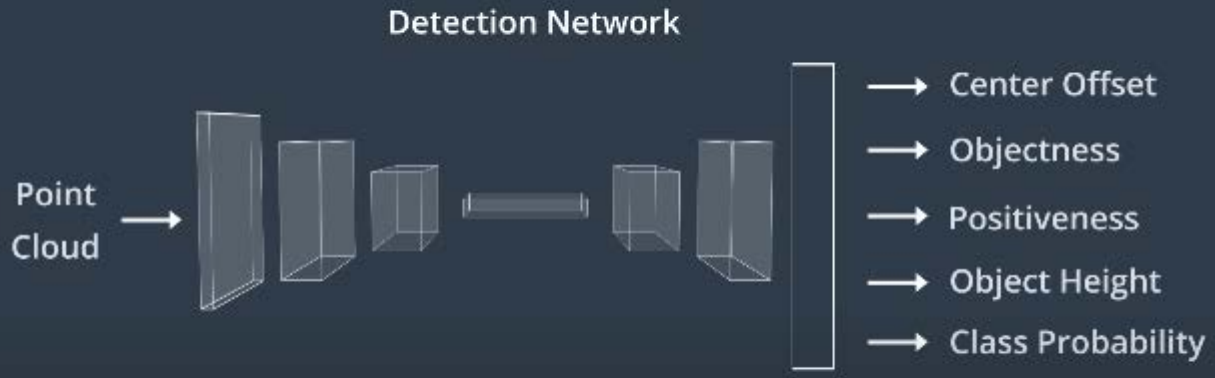


Semantic segmentation involves classifying each pixel in the image.



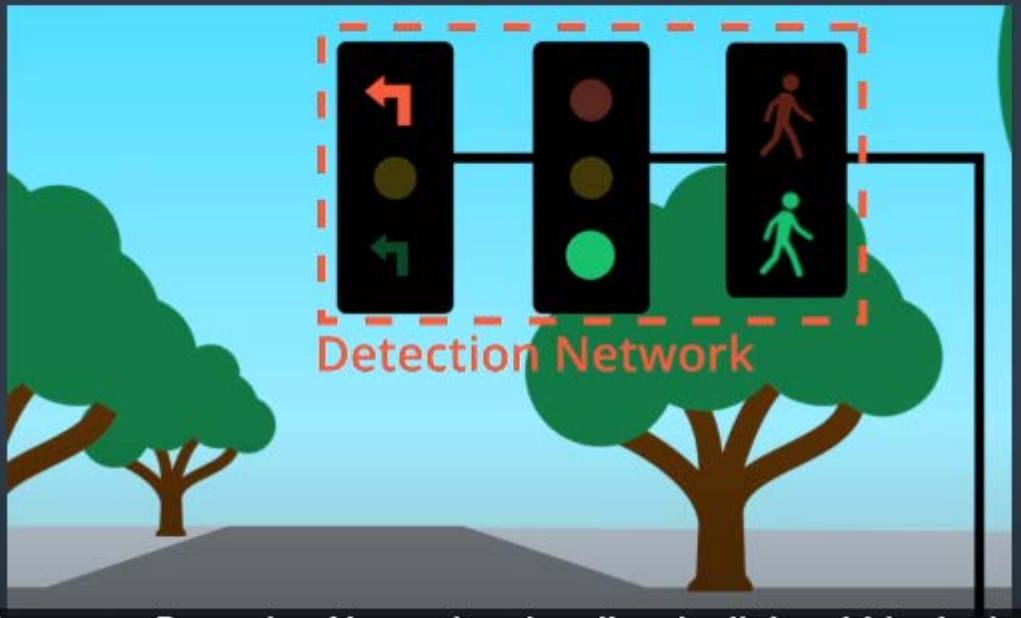


narrowing the search scope and accelerating perception.



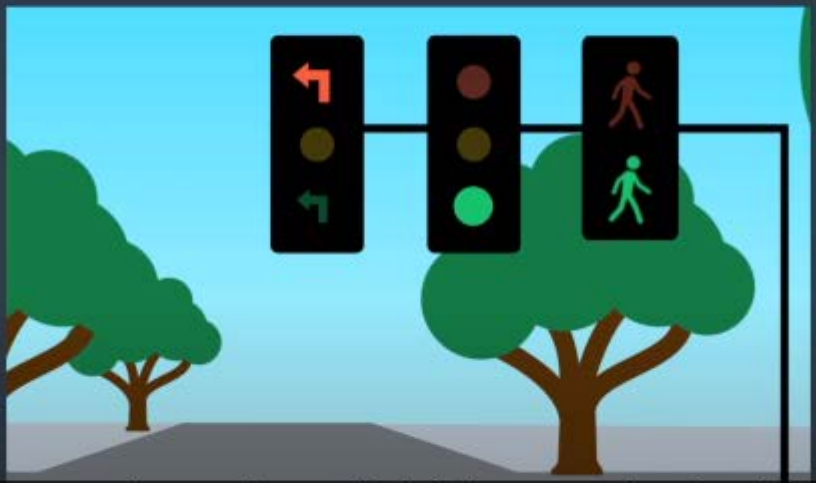
The output is used to construct three-dimensional bounding boxes around the objects.


Detection to Track Association in order to identify individual objects across timesteps.
0:45 / 2:27
YouTube

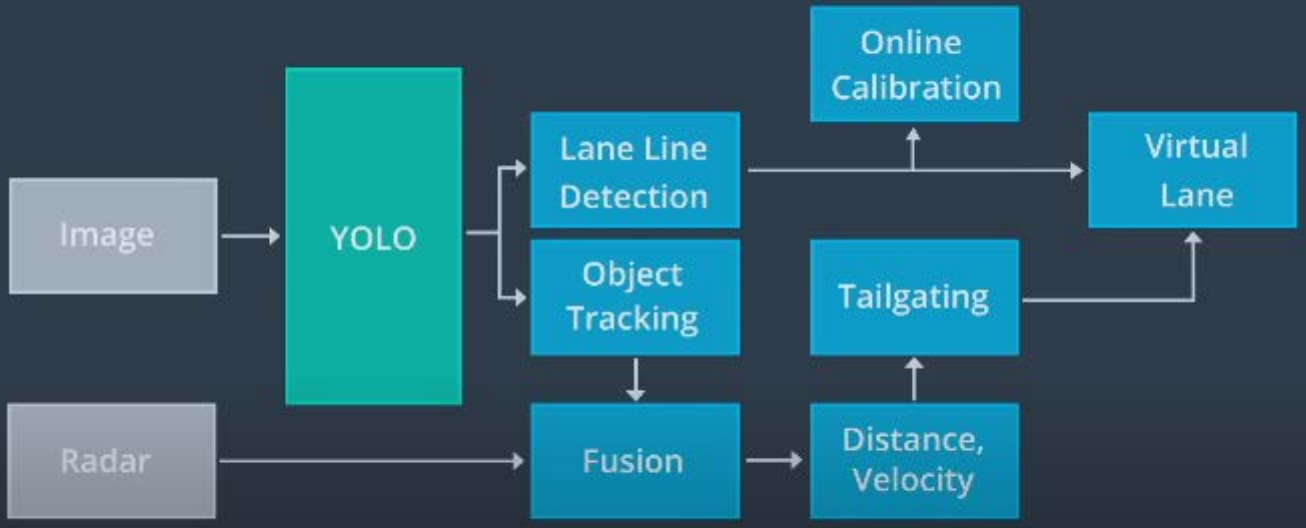


Apollo uses a Detection Network to localize the light within the image.

Which Lights Pertain to the Lane?



the system needs to select which lights pertained to its own lane.



the dynamic objects are passed along to the planning and control modules.

QUIZ QUESTION

Which of the following tasks can be accelerated by ROI

- [ ] To detect vehicles, pedestrians, bicycle riders
- [x] Traffic signs detection
- [x] Traffic lights detection
- [ ] Lane line detection

this is called sensor fusion.

Radar and LiDAR

Radars have in automobiles for years. You can find them in systems like adaptive cruise control, blind spot warning, collision warning and collision avoidance. Even though Radar is a mature technology, it still gets improved all the time to make it even more powerful. While other sensors measure velocity by calculating the difference between two readings, Radar uses something called the Doppler effect to measure speed directly. The Doppler effect measures the change in frequency of the Radar waves based on whether the object is moving away from you or toward you. This is kind of like how a fire engine siren will sound differently depending on whether the fire engine is moving away from you or toward you. The Doppler effect is important for sensor fusion because it gives us the velocity as an independent measure parameter, and it makes the fusion algorithms converge much faster. Radar can also be used for localization by generating Radar maps of the environment. Because Radar waves bounce off hard waves bounce off hard surfaces, they can provide measurements to objects without direct line of flight. Radar can see underneath other vehicles, and spot buildings and objects that would be obscured otherwise. Of all the sensors on the car, Radar is the least affected by rain or fog and can have a wide field of view, about 150 degrees, or a long range, 200 plus meters. Compared to LiDARs and cameras, Radars have a low resolution. Especially in the vertical direction, the resolution is very limited. The lower resolution also means that reflections from static objects can cause problems. For example, manhole covers or a soda can lying on the street can have high Radar reflectivity even though they are relatively small.This is called Radar clutter, and it's why current automotive Radars usually disregard static objects.

LiDAR stands for Light Detection and Ranging, just as Radar stands for Radio Detection and Ranging. Unlike Radar, which uses radio waves, LiDAR uses an infrared laser beam to determine the distance between the sensor and a nearby object. Most current LiDAR use light in the 900 nanometer wave length range, although some LiDARs use longer wave lengths, which perform better in rain and fog. In current LiDARs, a rotating swivel scans the laser beam across the field of view. The lasers are pulsed, and the pulses are reflected by objects. These reflections return a point cloud that represents these objects. LiDAR has a much higher spatial resolution than Radar because of the more focused laser beam, the larger number of scan layers in the vertical direction, and the high density of LiDAR points per layer. The current generation of LiDARs cannot measure the velocity of objects directly. And have to rely on the differing position between two or more scans. LiDARs are also more affected by weather conditions and by dirt on the sensor, which requires keeping them clean. They are also much bulkier than other sensors and therefore, more difficult to integrate unless one just wants to mount a big scanner on the roof of the vehicle.

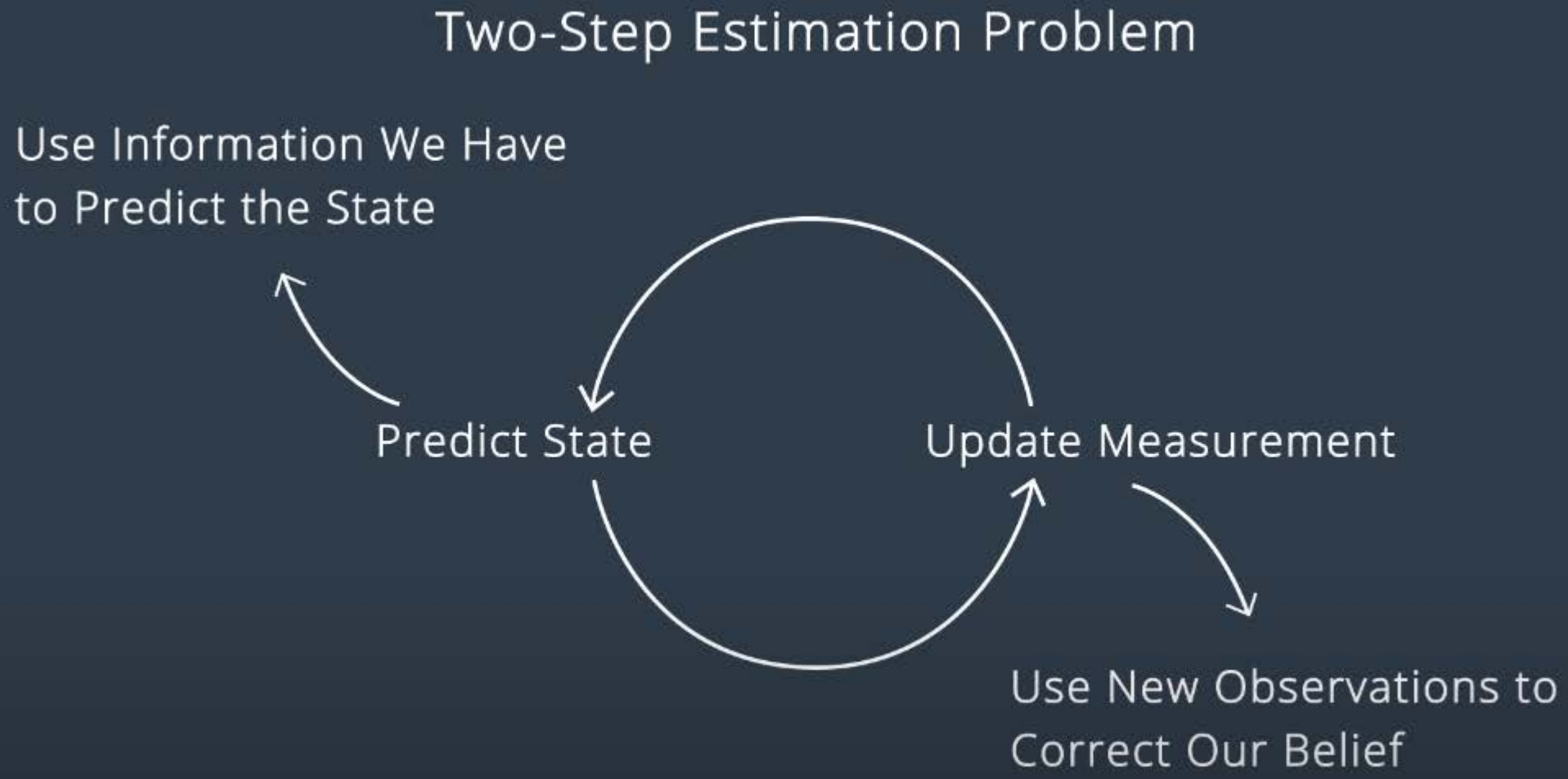
QUIZ QUESTION

Here are a few properties of Radar and LiDAR sensors. Drag and drop the sensor labels to the properties that they describe.

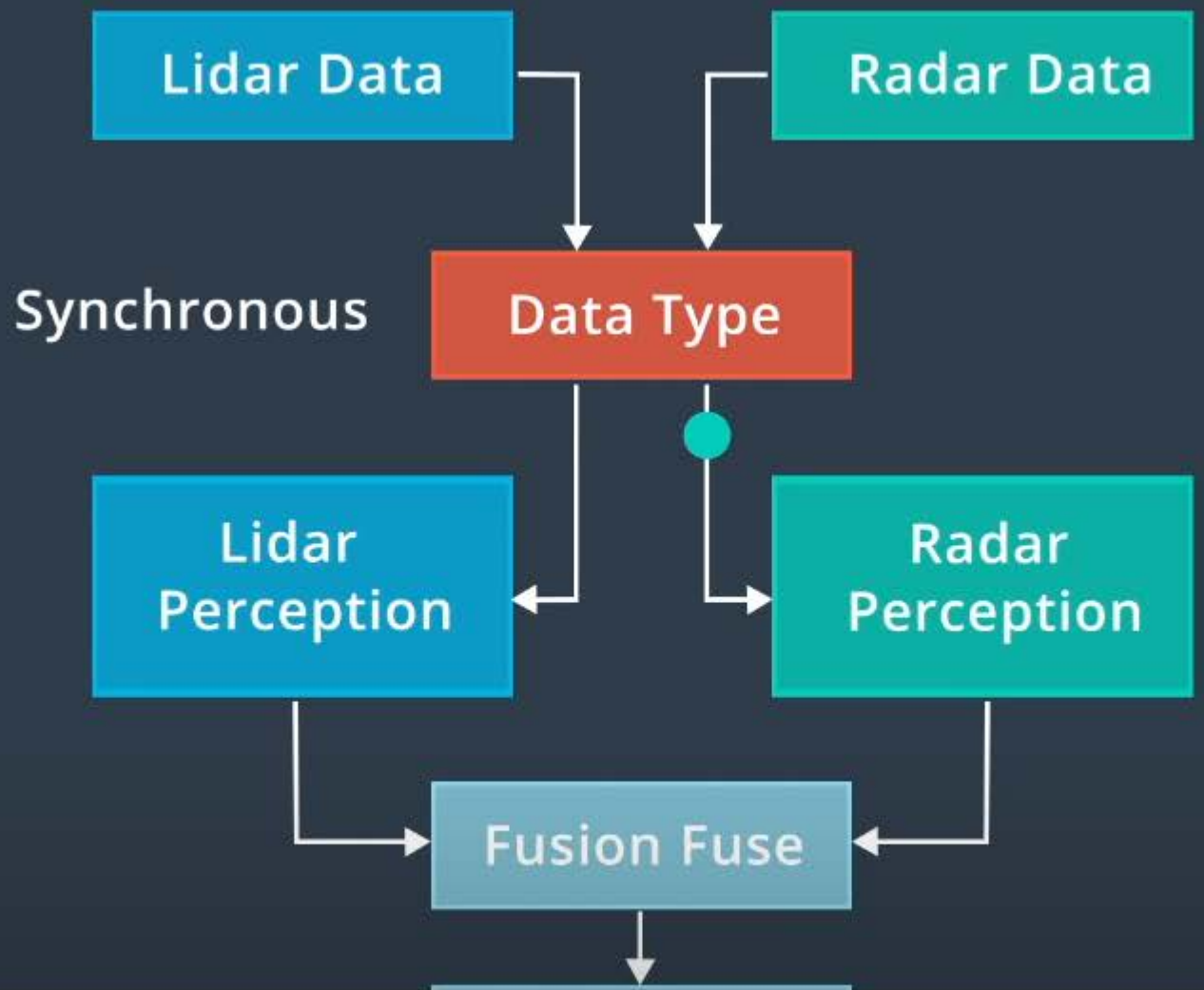
Submit to check your answer choices!

PROPERTIES	LIDAR OR RADAR?
Wavelength in mm	Radar
Wavelength in infrared	LiDAR
Can sense non-line of sight objects	Radar

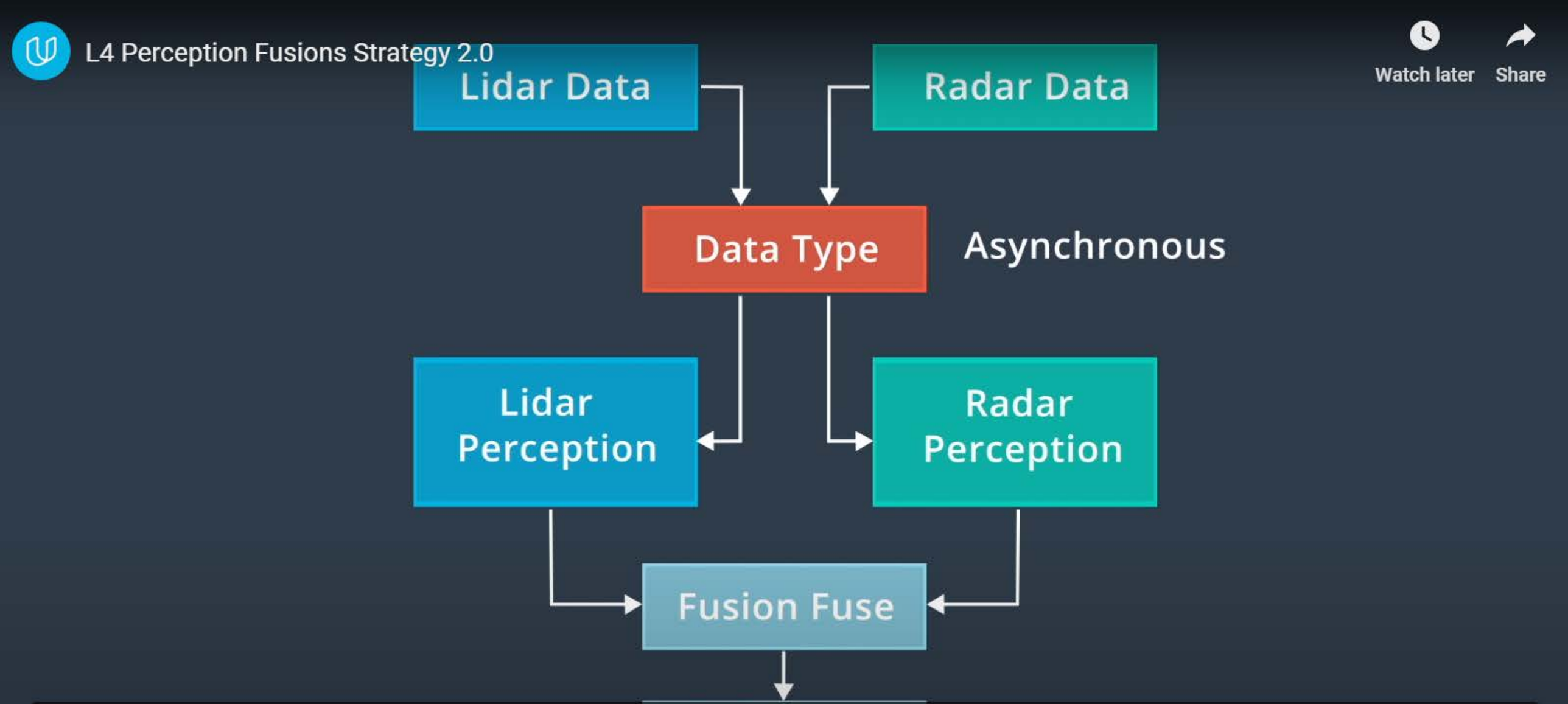
Most affected by dirt and small debris	LiDAR
Creates a point cloud	LiDAR
Can currently directly measure velocity	Radar
Higher Resolution	LiDAR



The Kalman filter algorithm is an endless loop of prediction and update steps.



Synchronous fusion updates the measurements from different sensors all at the same time,


L4 Perception Fusions Strategy 2.0
Watch later
Share
Lidar Data
Radar Data
Data Type
Asynchronous
Lidar Perception
Radar Perception
Fusion Fuse
Fused Objects
whereas asynchronous fusion updates the sensor measurements one at a time as they arrive.
1:08 / 1:28
YouTube