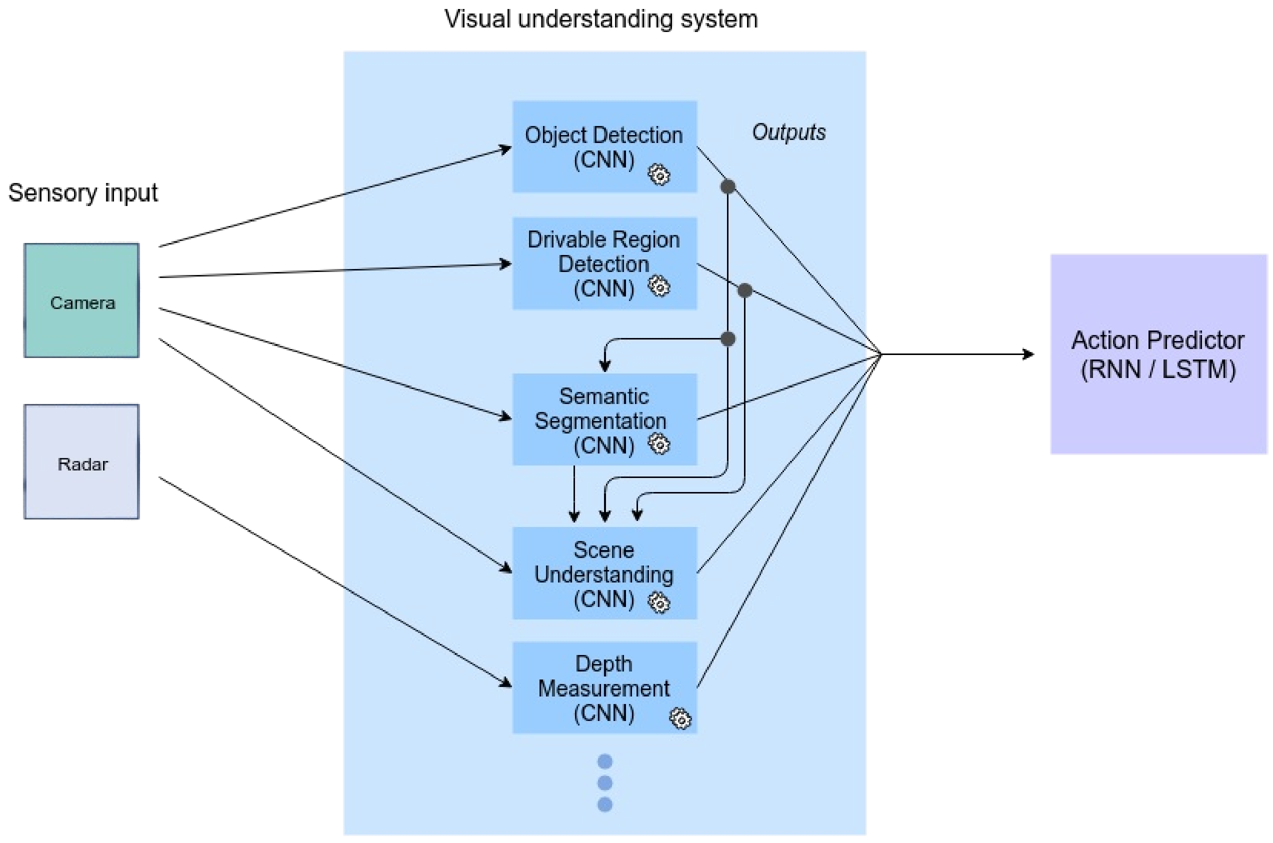
Architectural Components

Let's take a look at the architectural components of the self-driving car system.

We'll cover the following

Overall architecture for self-driving vehicle

Let's discuss a simplified, high-level architecture for building a self-driving car. Our discussion entails the important learning problems to be solved and how different learning models can fit together, as shown below.


Visual understanding system
Sensory input
Camera
Radar
Object Detection (CNN)
Drivable Region Detection (CNN)
Semantic Segmentation (CNN)
Scene Understanding (CNN)
Depth Measurement (CNN)
Outputs
Action Predictor (RNN / LSTM)

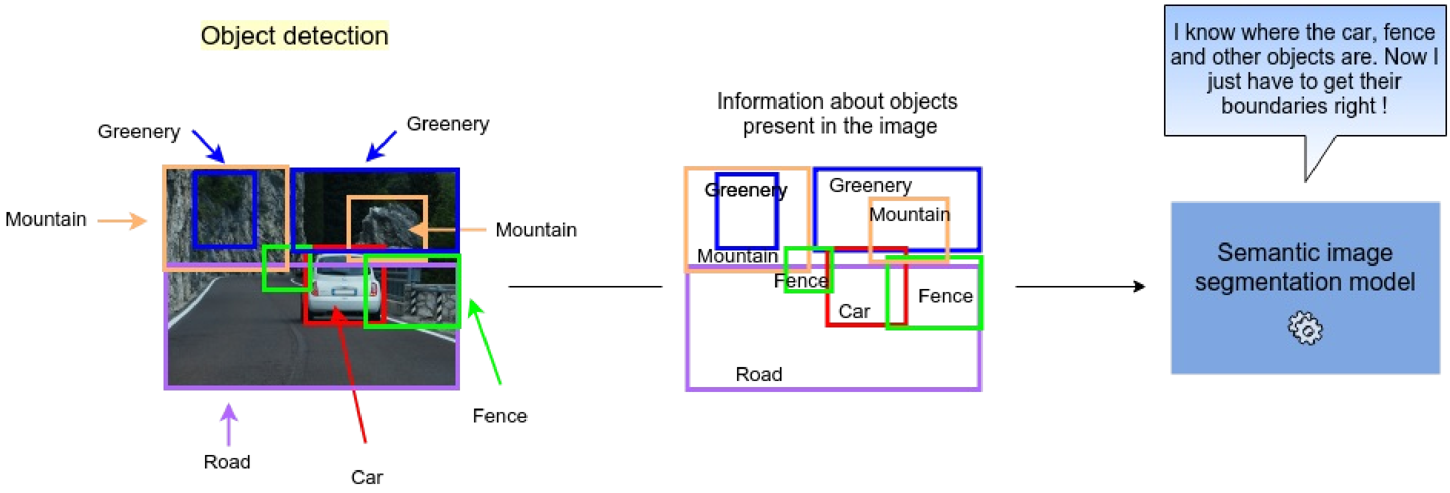
High-level architecture (CNN: Convolutional neural network, RNN: Recurrent neural network, LSTM: Long short-term memory)

The system is designed to receive sensory inputs via cameras and radars, which are fed to the *visual understanding system* consisting of different convolutional neural networks (CNN), each for a specific subtask. The output of the *visual understanding system* is used by the *action predictor RNN or LSTM*. Based on the visual understanding of the environment, this component will plan the next move of the vehicle. The next move will be a combination of outcomes, i.e., applying brakes, accelerating, and/or steering the vehicle.

> We won't be discussing input through Lidar here. However, it can also be used for scene analysis similar to a camera, especially for reconstructing a 3D view of an environment.

Now, let's have a look inside the visual understanding system. The object detection CNN detects and localizes all the obstacles and entities (e.g., humans and other vehicles) in the vehicle's environment. This is essential information because the action predictor RNN may predict to slow down the vehicle due to the impending obstacle (e.g., a person crossing the road). However, the most crucial information for the action predictor RNN is information that allows it to extract a drivable path for the vehicle. Therefore, due to the significance of this task, we will train a separate model for this purpose: the drivable region detection CNN. It will be trained to detect the road lanes to help the system decide whether the pathway for the vehicle is clear or not.

Moreover, as the object detection CNN identifies the key objects in the image (predicts bounding boxes), it is further required to share its output along with the raw pixel data for *semantic image segmentation* (i.e., draw pixel-wise boundaries around the objects). These boundaries will help in navigating the autonomous vehicle in a complex environment where overlapping objects/obstacles are presented.


Object detection
Greenery
Greenery
Mountain
Mountain
Fence
Road
Car
Information about objects present in the image
Greenery
Greenery
Mountain
Mountain
Fence
Car
Fence
Road
I know where the car, fence and other objects are. Now I just have to get their boundaries right !
Semantic image segmentation model

Object detection model's output is fed into the segmentation model

Machine learning models used in the components

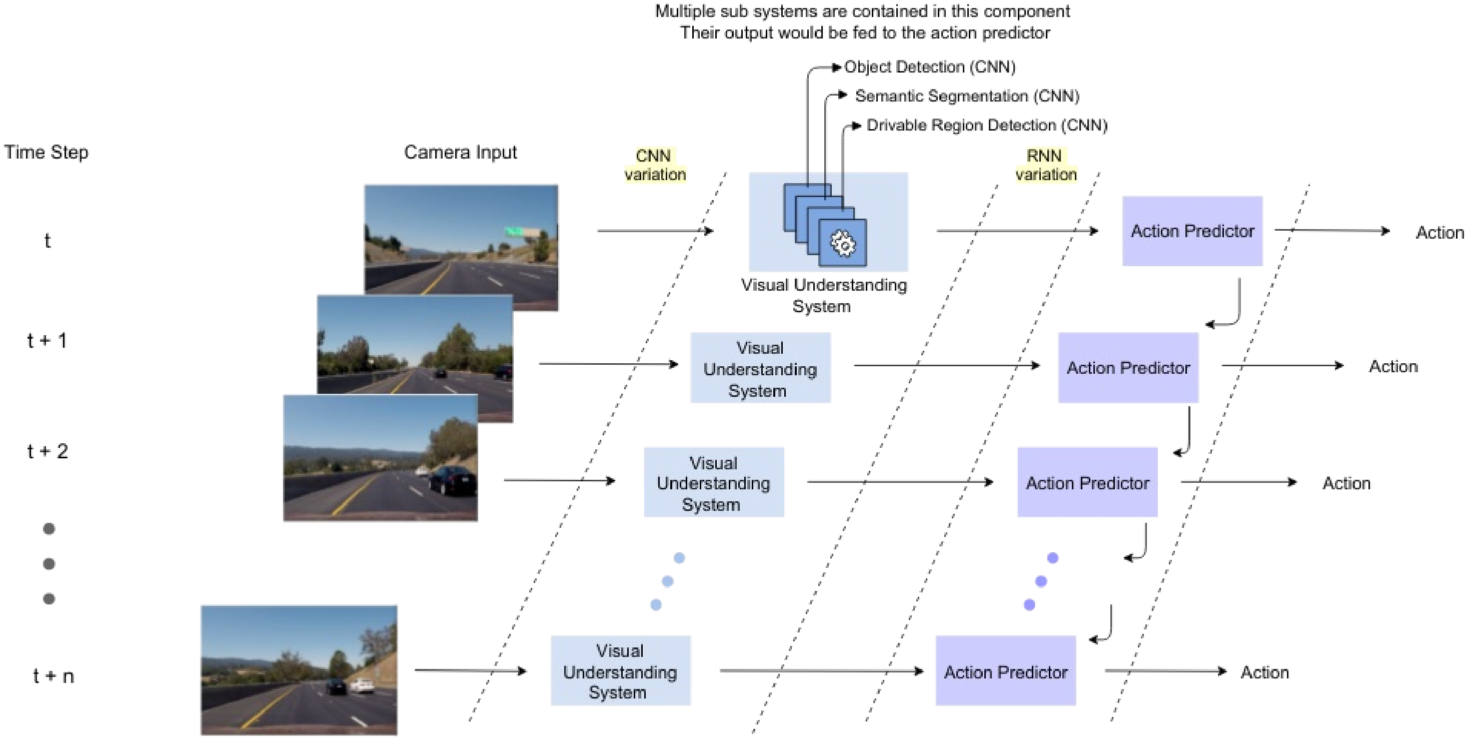
Many of the subtasks in the visual understanding component are carried out via specialized CNN models that are suited for that particular subtask.

The action predictor component, on the other hand, needs to make a movement decision based on:

1. Outputs of all the visual understanding sub-tasks
2. Track/record of the vehicle's movements based on previous scene understanding

This can be best learned through a recurrent neural network (RNN) or long short-term memory (LSTM) that can utilize the temporal features of the data, i.e., previous and current predictions from the scene segmentation as inputs.

Let's dive a little deeper into how this is happens, focusing on the input from the camera.


Multiple sub systems are contained in this component
Their output would be fed to the action predictor
Object Detection (CNN)
Semantic Segmentation (CNN)
Drivable Region Detection (CNN)
Time Step
Camera Input
CNN variation
RNN variation
t
t + 1
t + 2
t + n
Visual Understanding System
Visual Understanding System
Visual Understanding System
Visual Understanding System
Action Predictor
Action Predictor
Action Predictor
Action Predictor
Action
Action
Action
Action

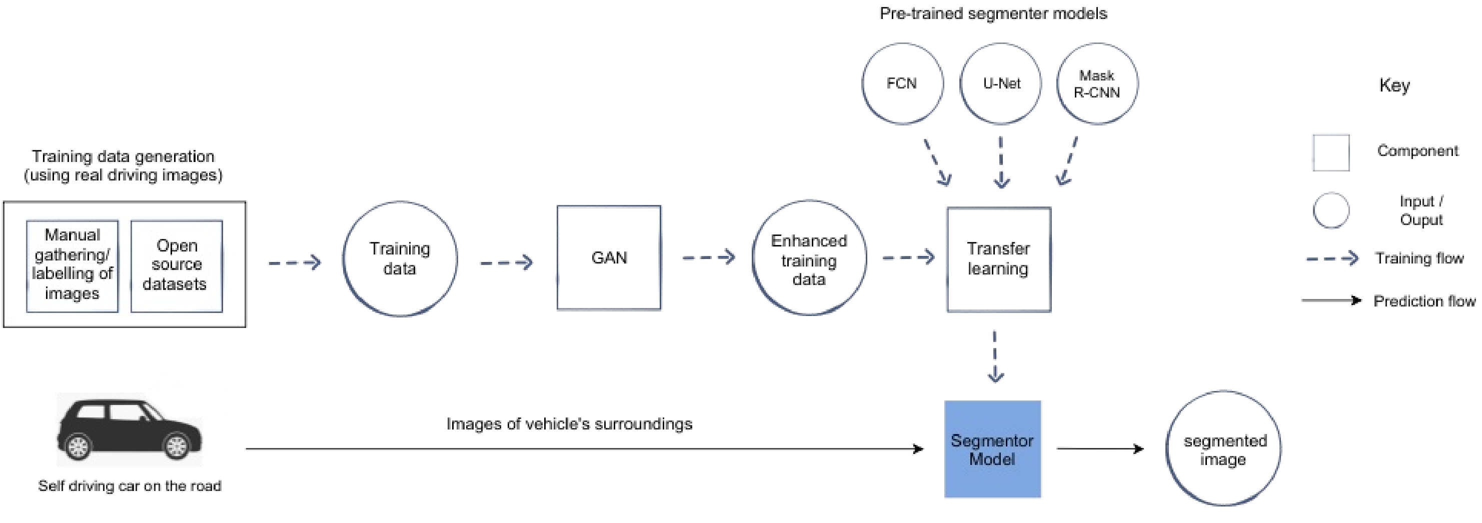
How the self-driving vehicle architecture works in a block diagram

You will receive the video frames covering the vehicle's surroundings from the camera as input. This video is nothing more than a sequence of images (frames per second). The image at each time step (t, t+1,...) will be fed to the visual understanding system. The multiple outputs of this component will form the input to the self-driving vehicle's action predictor. The action predictor will also use the previous time step information (t-1) while predicting the action for the current time step (t).

> RNN will utilize the past information to make "informed" predictions in real-time.

System architecture for semantic image segmentation

You just saw how the semantic image segmentation task fits in the overall architecture for the self-driving vehicle system architecture. Now, let's zoom in on the architecture for the semantic image segmentation task.


Pre-trained segmenter models
FCN
U-Net
Mask R-CNN
Key
Component
Input / Ouput
Training flow
Prediction flow
Training data generation (using real driving images)
Manual gathering/ labelling of images
Open source datasets
Training data
GAN
Enhanced training data
Transfer learning
Images of vehicle's surroundings
Self driving car on the road
Segmentor Model
segmented image

Architectural diagram for semantic image segmentation

The above diagram shows the training flow and prediction flow for the semantic image segmentation system architecture.

The training flow begins with training data generation, which makes use of two techniques. In the first technique, real-time driving images are captured with the help of a camera and are manually given pixel-wise labels by human annotators. The latter technique simply uses open-source datasets of self-driving vehicle images. This training data is then enhanced or augmented with the help of generative adversarial networks (GANs). Collectively all the training data is then used to train the segmenter model. Transfer learning is applied with the segmenter model to utilise powerful feature detectors from pre-trained models (FCN, U-Net, Mask R-CNN). The pre-trained models are optimized on your datasets to get the final model.

In the prediction flow, your self-driving vehicle would be on the road. You would receive real-time images of its surroundings, which would be given to the segmenter model for semantic image segmentation.

← Back
Metrics

Next →
Training Data Generation

☑ Mark as Completed

Report an Issue

Ask a Question
(https://discuss.educative.io/tag/architectural-components__self-driving-car-image-segmentation__grokking-the-machine-learning-interview)