

OBJECT DETECTION IN ROAD IMAGES

Name: Rory Egan

Supervisor: J.J. Collins

OBJECTIVES

This project explores the use of Convolutional Neural Networks (CNNs) to detect and classify objects in road scenes. The images used throughout this project were taken from the Berkeley Deep Drive dataset. Several CNN architectures have been investigated.

BERKELEY DEEP DRIVE

The dataset utilised for this project is the Berkeley Deep Drive dataset, consisting of 100,000 images and 100,000 HD video sequences annotated with 10 different classes. Due to the highly complex nature of the dataset it was manually subsampled down to 3 classes - cars, traffic lights and people. An example of an image being annotated is shown below:



AMAZON WEB SERVICES

Training was carried out using an AWS Deep Learning AMI to gain access to more computational power. This VM is also being utilised to serve a Flask application to demo the object detection abilities of the trained model.

CNNs EXPLAINED

CNNs are a type of deep neural network commonly applied to vision problems. Inspired by the visual cortex of the brain, CNNs identify primitive features of an object which are then mapped to more abstract concepts that make up the classes to be detected.

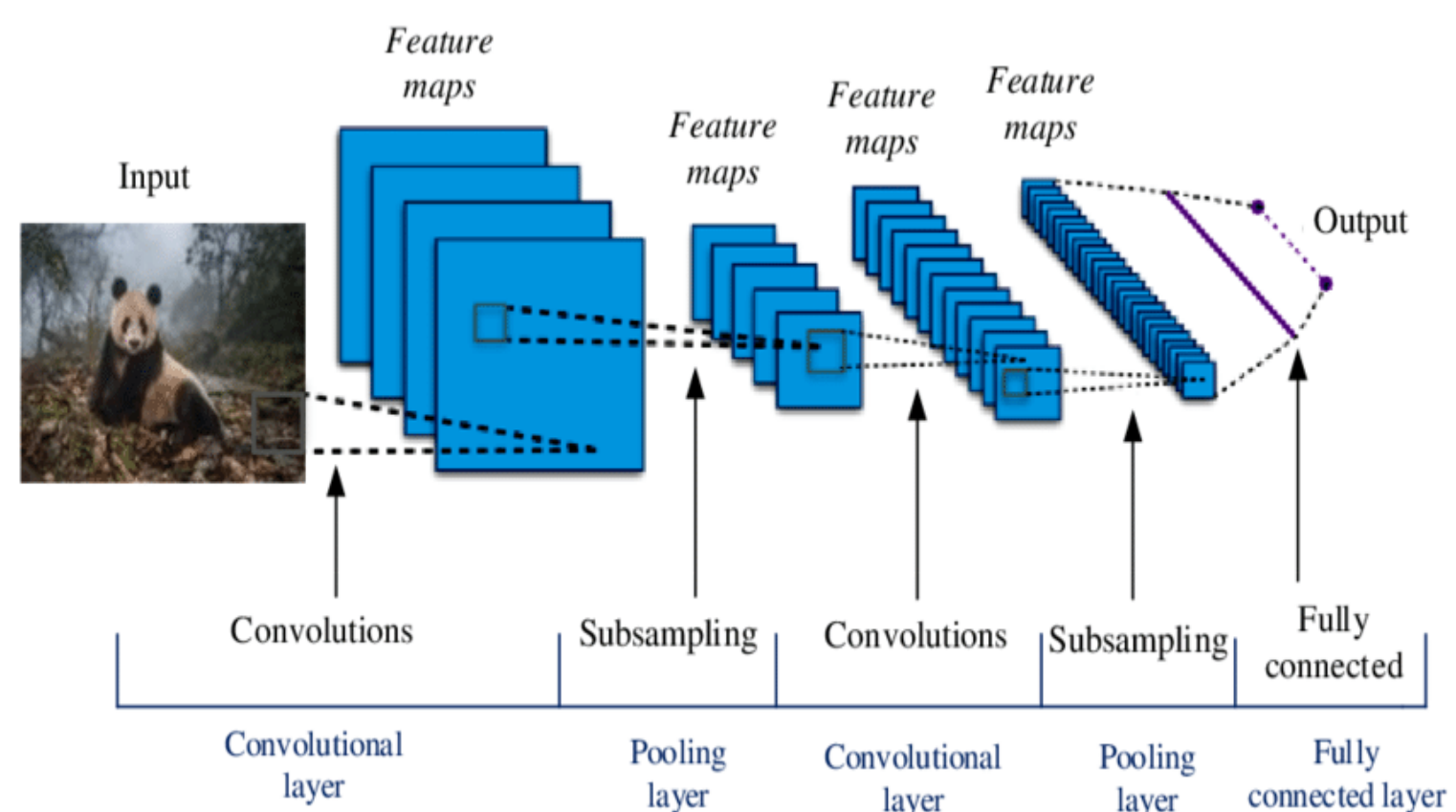


Image sourced from https://www.researchgate.net/figure/An-example-of-a-simple-CNN-architecture_fig6_327260166

INCEPTION ARCHITECTURE

The Inception v2 architecture was utilised for this project. The Inception architectures were developed to handle objects appearing at different focus levels within an image. The architectures are based on the concept of inception blocks. This is a combination of filters concatenated into a single output rather than having a layer for each filter. This allows the creation of "wider" rather than "deeper" networks, reducing the computational requirements of the network.