CV / VLM

Unit 1: Introduction to Computer Vision (CV)

1.3.2

Intro to Image Classification

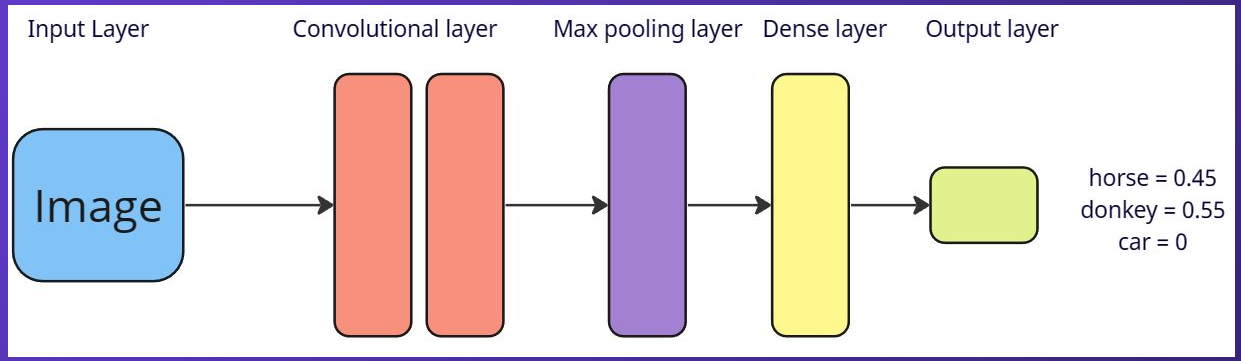
Introduction to CNNs for Image Classification

TIL-AI
TODAY I LEARNED AI

CNN Introduction and Overview

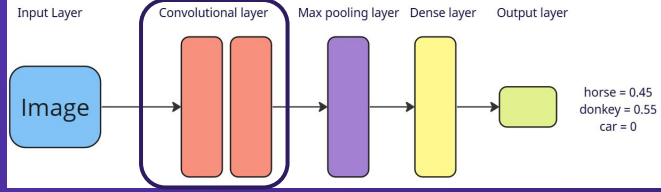
Components:

1. Convolutional Layer
2. Pooling Layer
3. Activation Functions
4. Dense (Fully Connected) + Output Layers

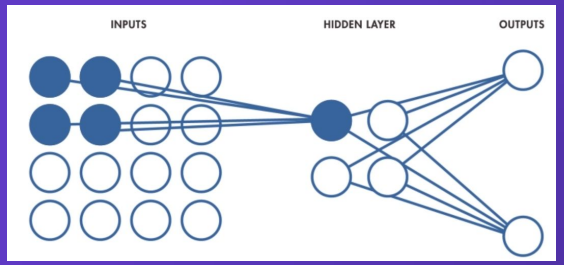


TIL-AI
TODAY I LEARNED AI

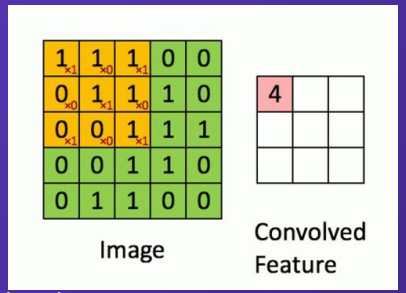
Convolution Layer



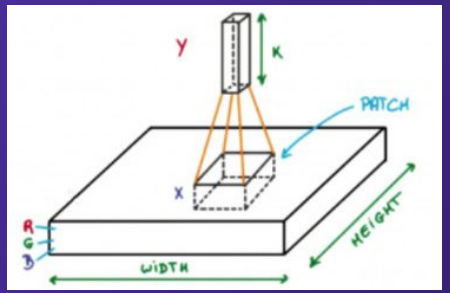
- Slides a small filter across the image; the convolution layer can be used to extract features from the image (edges, shapes, lines, etc) after training.



Convolutional Layer



(top) An example of a 3x3 kernel being applied on Image

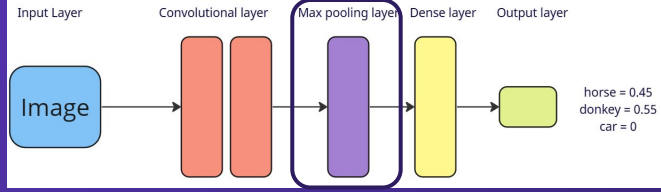


(top) 'Sliding window' a physical representation of how a filter traverses the image

- The parameters of the filter (kernels) are to be learned during the training through back propagation.

TIL-AI
TODAY I LEARNED-AI

Pooling Layer



- Takes the feature map and simplifies it (resize).
- Keeps the most important information (with max/average pooling)
- Reduces image size making calculations faster and preventing the network from getting confused by tiny details.

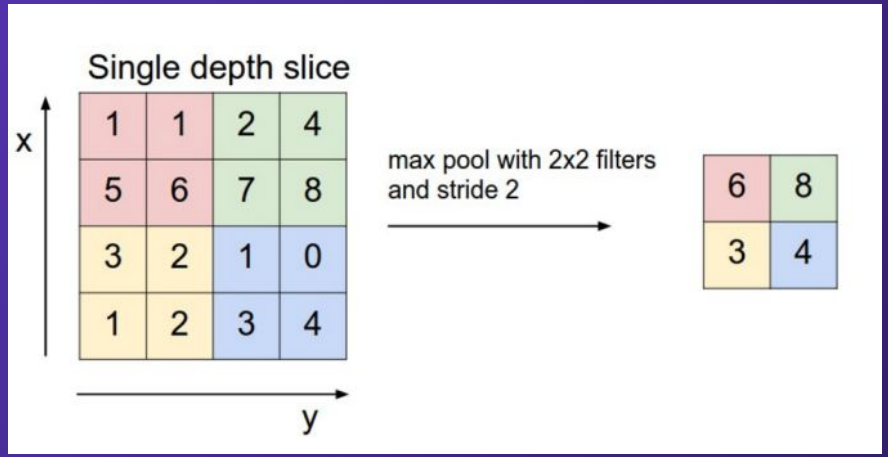
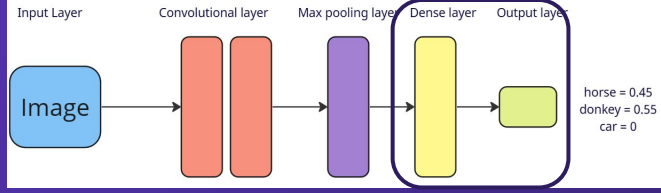


Image source: cs231n.stanford.edu

TIL-AI
TODAY I LEARNED-AI



Dense (Fully Connected) Layers + Output Layers

- Similar to regular neural networks, Dense layers connect all the features from previous layers.

- Finally, The output layers (for classification) would corresponds to the number of class for output.

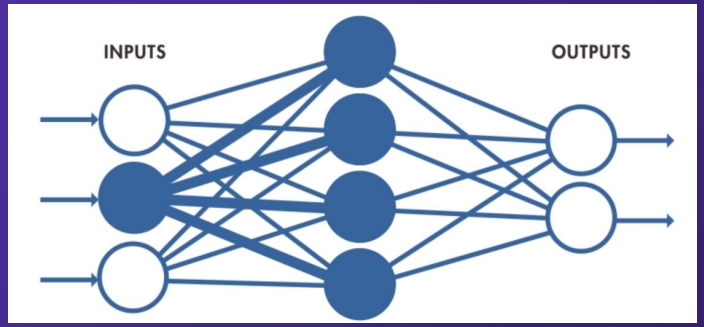


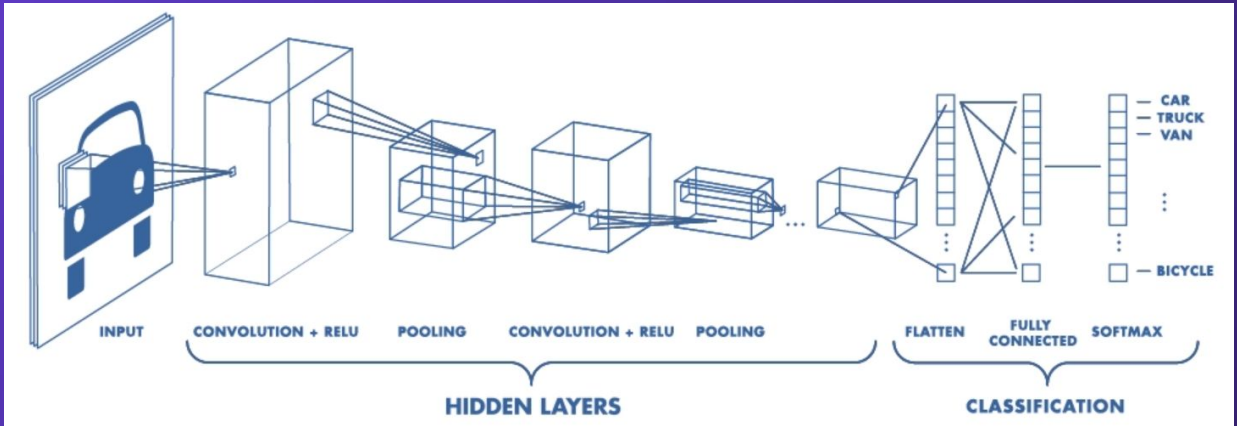
Image source: cs231n.stanford.edu

Notes:
- Both Dense Layers and Fully connected Layers bears the same meaning.
    - Torch adapted Fully connected nodes as FC layers, while Keras call its FC layers as Dense Layers.

TIL-AI
TODAY I LEARNED-AI

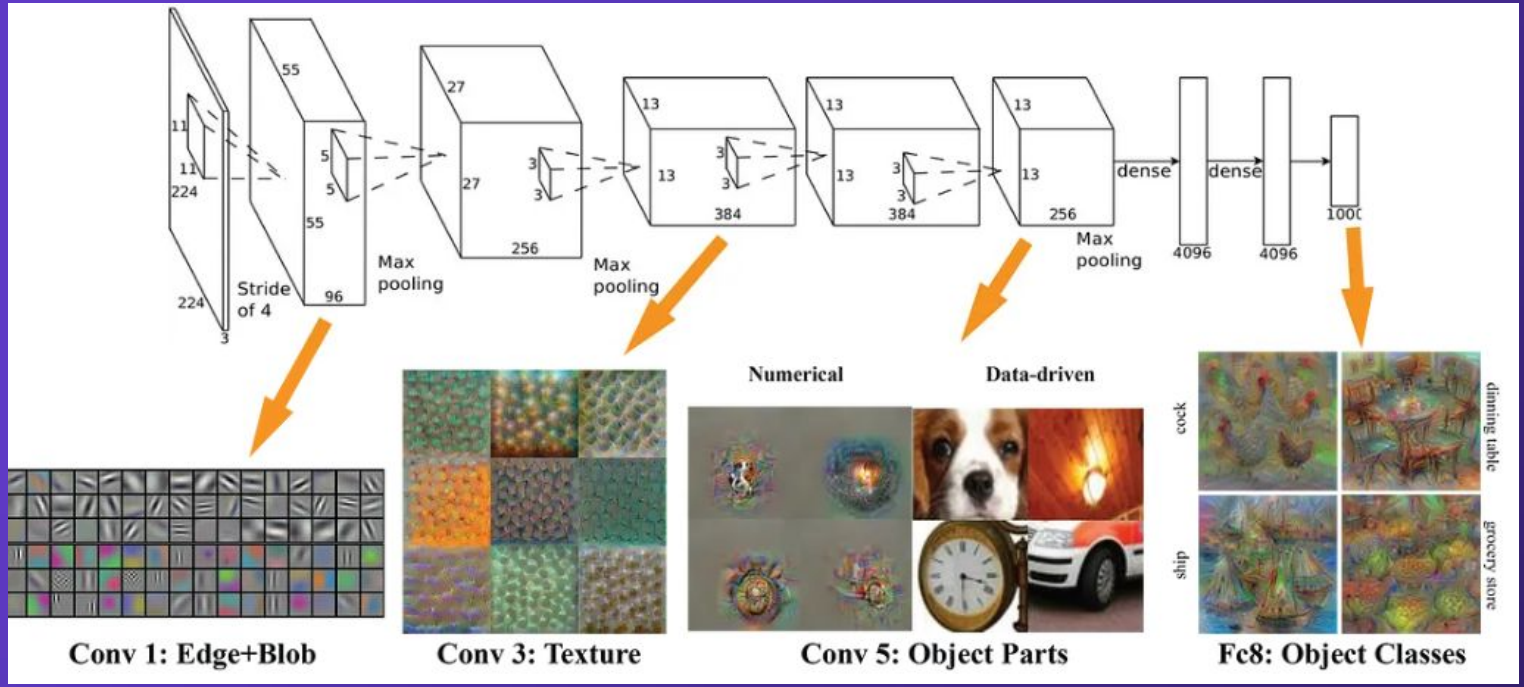
CNN Layers

- Each layer allows a CNN to extract various features in an image
- A typical CNN has many sets of the previously described layers, with each successive layer extracting higher level features.



TIL-AI
TODAY I LEARNED-AI

CNN Layers (cont.)



Hierarchy of the Features Extraction

Image source: VGG Feature Map Visualisation (from Brown Uni)

CNN Layers (cont.)

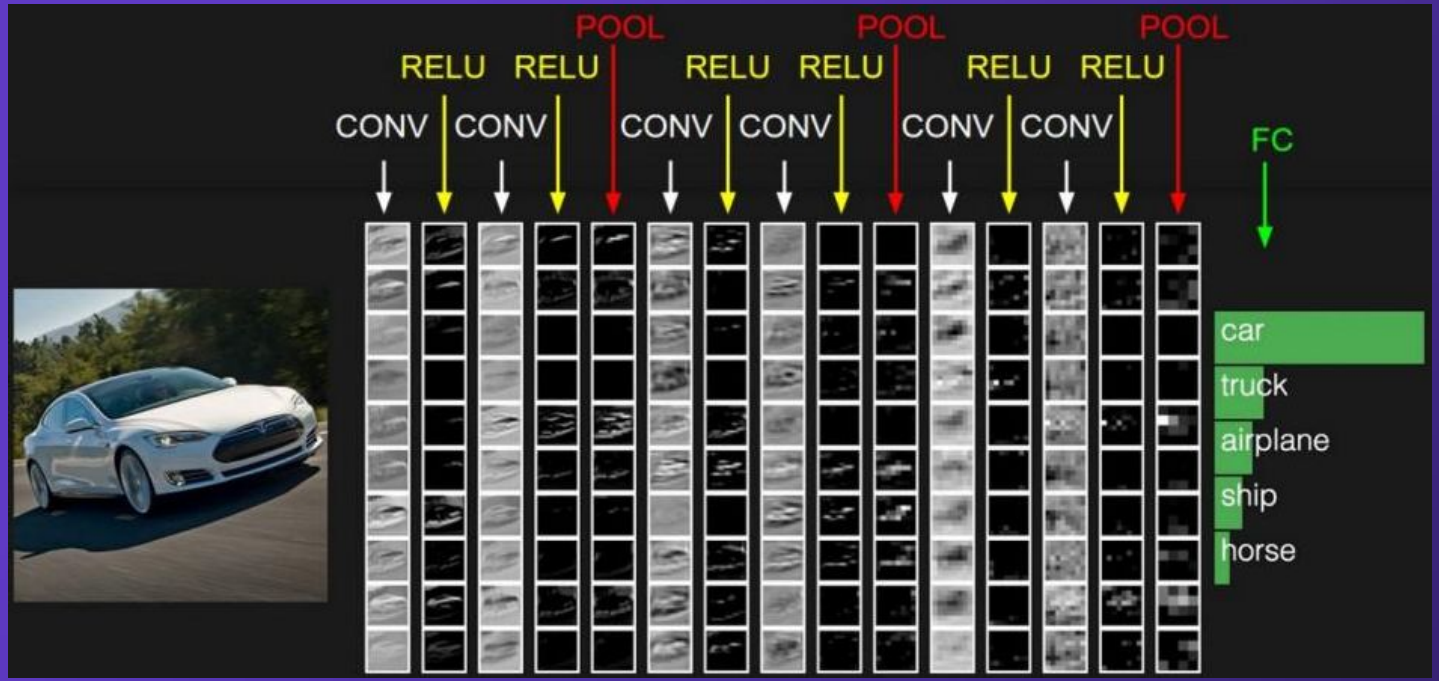


Image source: cs231n.stanford.edu

Overview

1. Convolutional Layer: Feature Extractions

2. Activation Functions: ReLU, providing non-linearity functionality

3. Pooling Layer: Simplifies/resize

4. Dense (Fully Connected) + Output Layers: Classifier

TIL-AI
TODAY I LEARNED-AI

Advantages of CNNs

- Good at detecting patterns and features in images, videos, and audio signals.
- Robust to translation, rotation, and scaling invariance.
- End-to-end training - no need for manual feature extraction.
- Can handle large amounts of data and achieve high accuracy.

Disadvantages of CNNs

- Computationally expensive to train and require a lot of memory.
- Can be prone to overfitting if there is lack of data or proper regularization.
- Require large amounts of labeled data.
- Interpretability is limited - it’s hard to understand what the network has learned.