Karlsruher Institut für Technologie

AI Mat

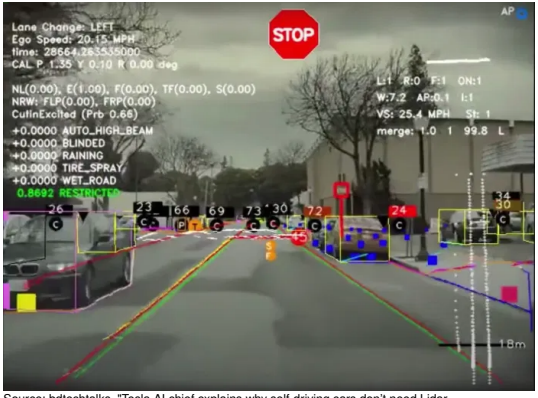
An Introduction to Convolutional Neural Networks(CNN)

Simon Schupp | 24.07.23

Table-Of-Content

Introduction

Visual Data and Algorithms



Source: bdtechtalks, "Tesla AI chief explains why self-driving cars don't need Lidar

Tesla Autopilot object Detection

Source: Google Blog "PaLM-E

Google PaLm-E robot

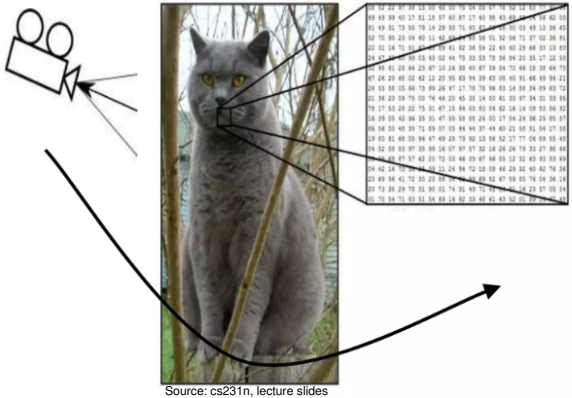
How Computer see

What I see

What a computer sees

Source: GitHub, hosamelsafty/Cats-VS-Dogs

Challenges: Viewpoint



Source: cs231n, lecture slides

When we change the viewpoint,
all the pixel change!

More challenges

- Viewpoint
- Illumination
- Occlusions
- Background Clutter
- etc.

Image Classifier

```
function CLASSIFYIMAGE(Image : Image)
    // What to do here? Magic?
    return label
end function
```

It's not obvious how to hard code this.

Image Classifier

```
function CLASSIFYIMAGE(Image : Image)
    // What to do here? Magic?
    return label
end function
```

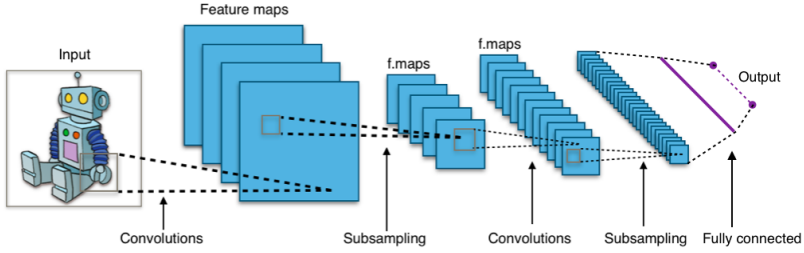
It's not obvious how to hard code this.

Idea: Learn how to classify the Images $\longrightarrow$ Use Convolutional Neural Networks(CNN)

Convolutional Neural Networks(CNN)

CNN Overview

- Used mostly to classify images
- A CNN is a specialized NN
- It can pick up patterns and make sense
- Consists of the following Components:
  - Convolutional Layer
  - Pooling Layer

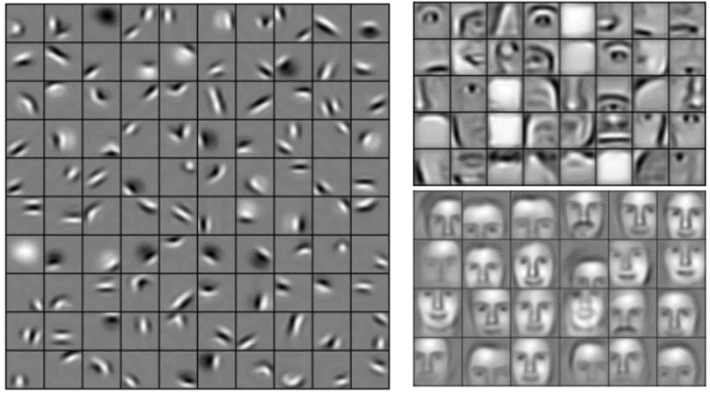


Source: Aphex34, Wikipedia CNN

A Typical CNN

Patterns

- Simple patterns: edges, shapes, textures
- More complex patterns: ears, eyes, noses
- Even more complex patterns: faces, humans

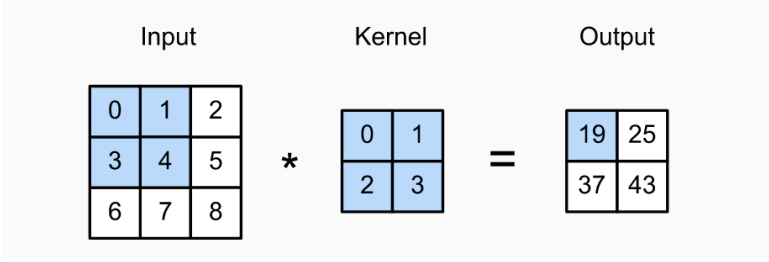


Source: Brandon Rohrer, course 193

Filters/Kernel

Filter/Kernel: A matrix

- initalized with random values
- values adjusted over time, to detect pattern



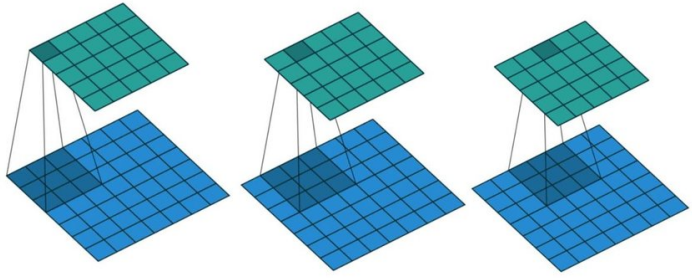
Source: d2l.ai, Chapter 7.2

Convolution Operation

Convolution

Two Important Hyperparameters:

- Stride
- Filter-Size



Source: Jelo Salomon, Lung Cancer Detection using Deep Learning

A Convolution with Stride 1

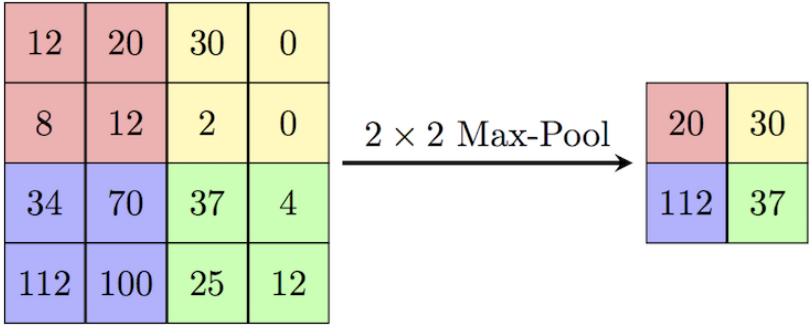
Convolution Animation

Convolution Animation

Pooling Layer

- A way to reduce the dimensionality of our Network
  - Requires less computation
- Many different ways to do this
  - Max-Pooling
  - Average-Pooling
  - 1x1 Convolutions

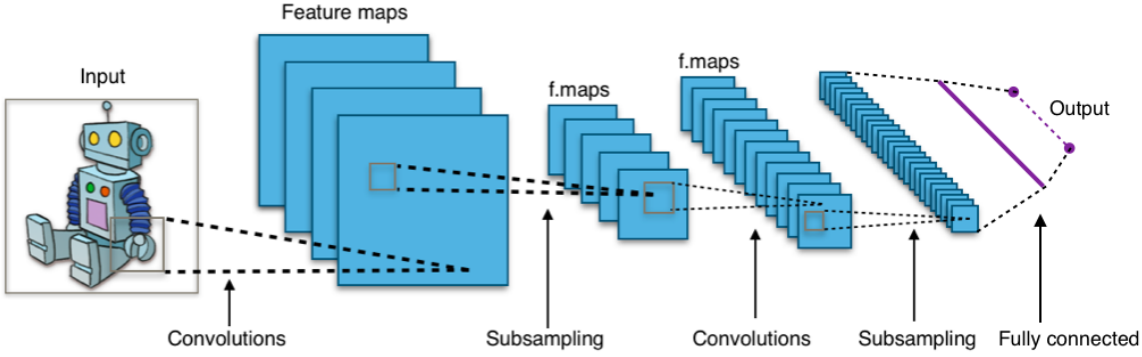
Pooling Layer



Source: cs231n lecture slides

A Max Pooling Layer

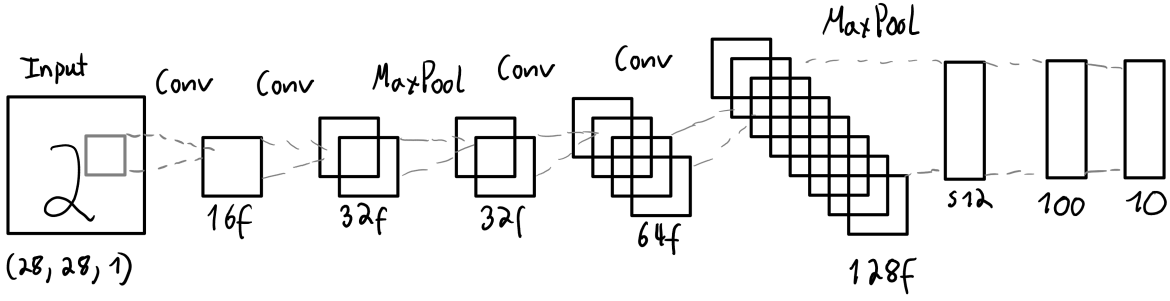
CNN: Bringing it Together



A Typical CNN

Self-trained CNN

Architecture

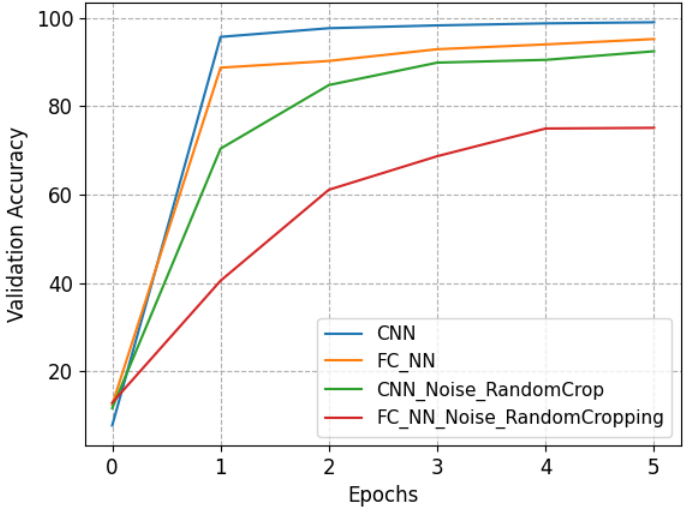


Trainings-Details

- Dataset: MNIST(60,000 Images)
- Trainings time: 5 Epochs
- Batch size: 1024
- Optimizer: Adam
- Learning-rate: 0.001
- Trained on Google Colab with GPU

Example Data from MNIST dataset

Results & Ablation



Summary

Summary

- Teaching a Computer the semantic meaning of Images is hard

Summary

- Teaching a Computer the semantic meaning of Images is hard
- But it's possible with the help of CNN

Summary

- Teaching a Computer the semantic meaning of Images is hard
- But it's possible with the help of CNN
- Nowdays its rather easy to get good performance on simple datasets

Literature

[1] Alex Krizhevsky, Ilya Sutskever und Geoffrey E Hinton. „ImageNet Classification with Deep Convolutional Neural Networks“. In: *Advances in Neural Information Processing Systems*. Hrsg. von F. Pereira u. a. Bd. 25. Curran Associates, Inc., 2012. URL: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[2] Y. LeCun u. a. „Backpropagation Applied to Handwritten Zip Code Recognition“. In: *Neural Computation* 1.4 (1989), S. 541–551. DOI: 10.1162/neco.1989.1.4.541.

[3] Y. Lecun u. a. „Gradient-based learning applied to document recognition“. In: *Proceedings of the IEEE* 86.11 (1998), S. 2278–2324. DOI: 10.1109/5.726791.

Literatur