NTNU | Norwegian University of Science and Technology

DEEP LEARNING CONVOLUTIONAL NEURAL NETWORKS

Tor Andre Myrvoll
2020-10-16

Introduction - Outline

Convolutional Neural Networks (CNNs)

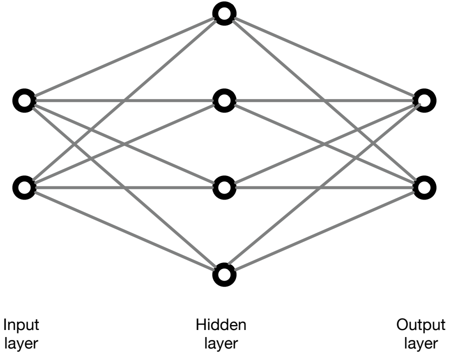
- Motivation
- Basic CNNs
- Transfer learning

Introduction - Motivation

Introduction - Motivation

MLP Specification

- Binary classifier
- 512 × 512 images
- One hidden layer with as many nodes as there are inputs



Introduction - Motivation

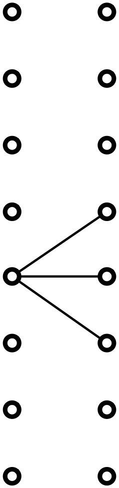
Reality cathces up...

- We stack the image into a columnvector of dimension 512 × 512 = 262144. Large, but doable
- The number of hidden nodes will also be 262144 dimensions
- The number of connections from input to output will be $262144^2 = 68719476736$
- Which is roughly 268 GB of data is using float
- Game over?

Introduction - Solutions

Only use a subset of the connections

- Using $k \ll N$ inputs will decrease number of parameters
- Ex. $k = 3 \Rightarrow k \cdot N = 786432$, a reduction of factor 90000



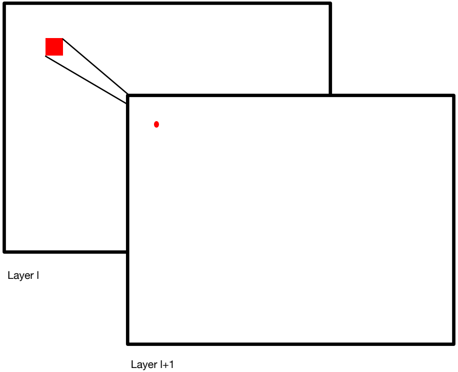
Introduction - The CNN

For images it makes sense to use local information

- The excitation in the $l + q$ layer is given as

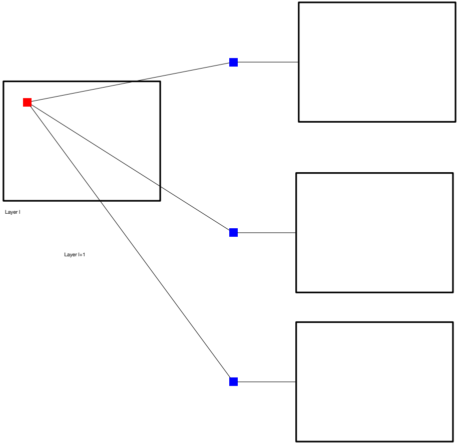
$$z_{i,j}^{l+1} = \sum_{m,n=-L/2}^{L/2} k_{m,n} v_{i-m,j-n}^{l}$$

- We see that this corresponds to a *convolution*, hence the naming scheme



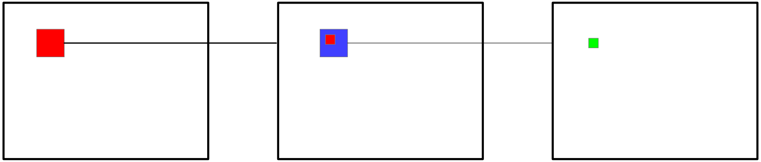
Introduction - The CNN

We can use multiple filters



Introduction - The CNN

In addition to convolution we do non-linear processing and *pooling*



Convolution

Non-linear processing

Pooling

Introduction - The CNN

The three processing steps that defines a CNN *layer* are:

- Convolution

$$z_{i,j}^{l+1} = \sum_{m,n=-L/2}^{L/2} k_{m,n} v_{i-m,j-n}^{l}$$

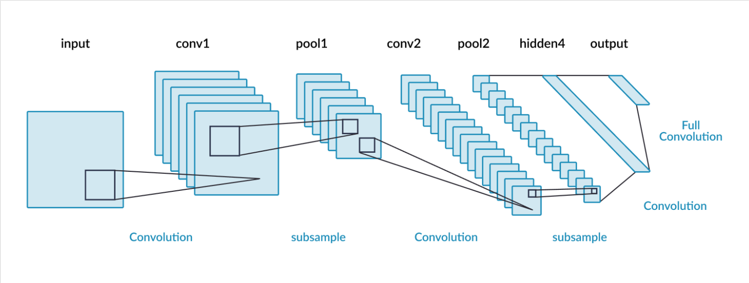
- Non-linear proxcessing: ReLU, Sigmoid, etc.
- Pooling: Representing local areas by some statistic
  - Examples: Mean value, median, maximum value, power

$$z_{i,j}^{l} = \max\{v_{i-m,j-n}\}_{m,n=-L/2}^{L/2}$$

  - Pooling leads to dimensionality reduction, and implicit translation and rotational invariance

Introduction - CNN example – LeNet

Early example of CNN – LeNet-5. Handwritten charater recognition:



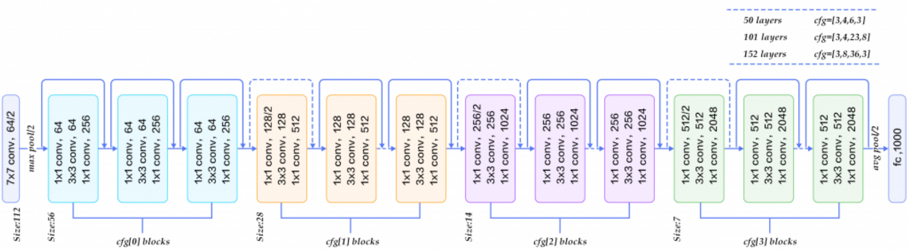
Introduction - CNN example – LeNet

Structure:

- Layer 1: Convolution. Six 5×5 kernels
- Layer 2: Pooling: $32 \times 32 \rightarrow 14 \times 14$
- Layer 3: Convolution. 16 5×5 kernels
- Layer 4: Pooling: $10 \times 10 \rightarrow 5 \times 5$
- Layer 5: Convolution: 120 5×5 kernels
- Layer 6: Fully connecter. 84 features output.

Introduction - CNN example – ResNet

Massive 152 layer network for image classification



Introduction - Transfer learning

Training a large CNN is a massive undertaking. Use *transfer learning* to use a previously training CNN.

- Only use the CNN layers of a model
- This part of the model is often called a *feature extractor*
- Define and train your own fully connected layers after the CNN (don't update the CNN parameters)
- Available from for instance the Keras Applications API
- Available from Apple when using the COre ML API

Thank you for your attention

NTNU | Norwegian University of Science and Technology