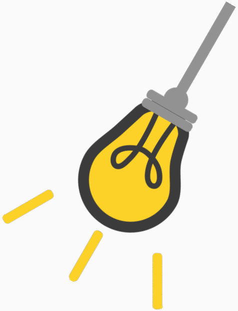
浙江大学爱丁堡大学联合学院
ZJU-UoE Institute

Lecture 13 - CNN structures

Nicola Romanò - nicola.romano@ed.ac.uk

Learning objectives

- Describe commonly used patterns in CNN architectures
- Describe and explain the advantages of different CNN architectures



Introduction

Introduction

Today we are going to discuss a few classic papers using CNN for image analysis.

We will analyse the following architectures:

- LeNET-5
- AlexNet
- VGG
- GoogLeNet
- ResNet

The idea is to get some **intuition** about these architectures and how they work.

LeNET-5

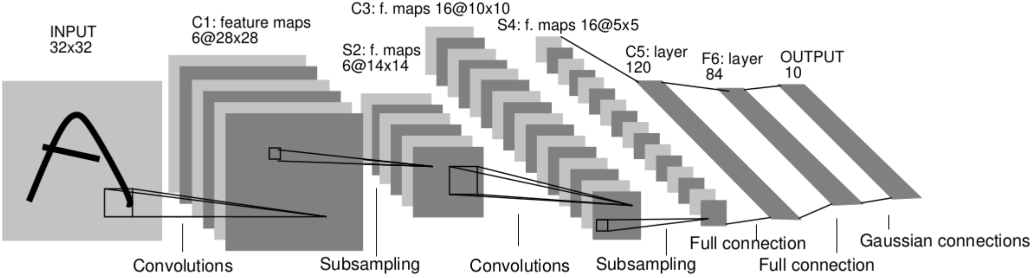
LeNET-5

- "Gradient Based Learning Applied to Document Recognition", Yann LeCun et al. 1998
- A seminal paper describing the use of CNN in image analysis
- Simple architecture with convolutional layers, average pooling and fully-connected layers
- Task: recognition of handwritten digits to be used for processing of bank cheques

LeNet-5 architecture

Gradient-Based Learning Applied to Document Recognition

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner



LeNet-5 take home points

- A simple architecture with convolutional layers, average pooling and fully-connected layers
- Introduced the $[\text{Conv} + \text{Pool}]_n + FC$ pattern
- This is mostly interesting from a historical perspective, not really used nowadays.

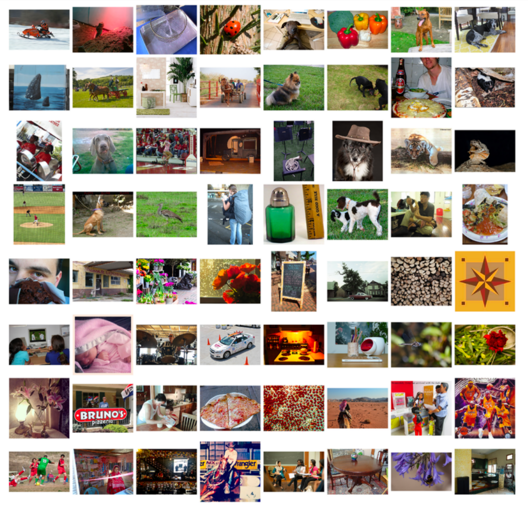
AlexNet

AlexNet

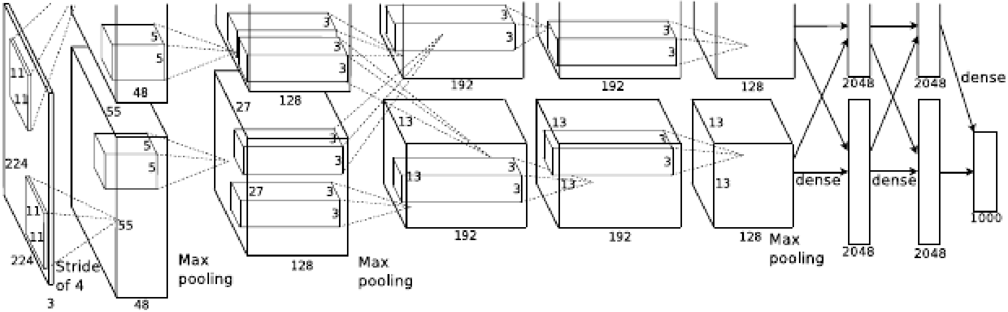
- "ImageNet Classification with Deep Convolutional Neural Networks", Alex Krizhevsky et al. 2012
- Widely considered as one of the most influential papers that boosted research in CNN for image analysis
- Similar architecture to LeNet-5, but with more convolutional layers
- Much bigger network (LeNet-5 60k parameters, AlexNet 60M parameters)
- Winner of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012.

The ImageNet Large Scale Visual Recognition Challenge

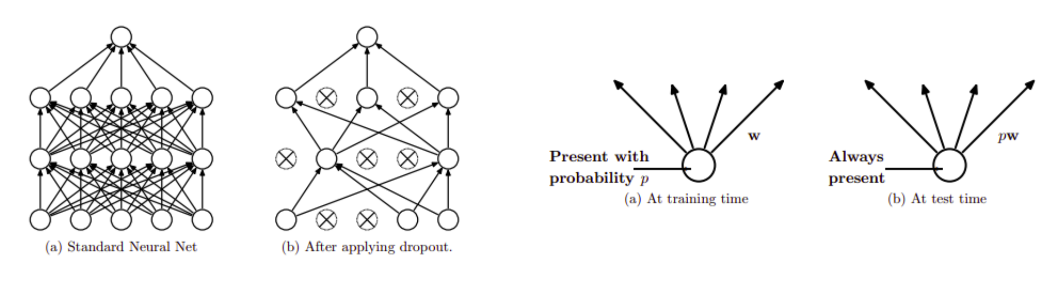
- ImageNet is a database of images of various objects, used for training and testing deep neural networks.
- Introduced in Deng et al., 2009 - ImageNet: A large-scale hierarchical image database
- It contains >14 million images of various objects, labelled with >20000 classes.
- The ILSVRC is a competition to define new algorithms for image classification.
- ILSVRC uses a subset of ImageNet, containing 1000 classes and 1.3M training images, 50k validation images and 100k test images.



ImageNet Classification with Deep Convolutional Neural Networks



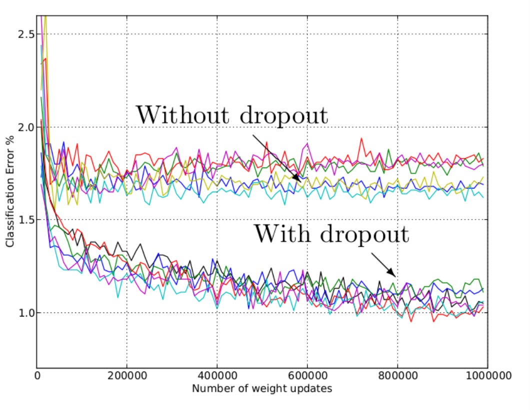
- A type of "regularization" technique, used to prevent overfitting
- A random subset of the weights is set to zero at each training step.
- Originally introduced in "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", Srivastava et al. 2014



Srivastava et al. 2014

Dropout

- A type of "regularization" technique, used to prevent overfitting
- A random subset of the weights is set to zero at each training step.
- Originally introduced in "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", Srivastava et al. 2014



Srivastava et al. 2014

The softmax activation function

- Generalization of the logistic function to multiple classes

The softmax activation function

- Generalization of the logistic function to multiple classes
- Common choice for classification problems in ANNs.

The softmax activation function

- Generalization of the logistic function to multiple classes
- Common choice for classification problems in ANNs.
- Defined as

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$$

The softmax activation function

- Generalization of the logistic function to multiple classes
- Common choice for classification problems in ANNs.
- Defined as

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$$

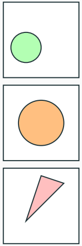
- It is used in the last layer of a CNN to compute the probability of each class.

The softmax activation function

- Generalization of the logistic function to multiple classes
- Common choice for classification problems in ANNs.
- Defined as

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$$

- It is used in the last layer of a CNN to compute the probability of each class.



Feedforward propagation $\longrightarrow$

Circle	Square	Triangle
8	4	1
9	3	2
5	4	10

Softmax $\longrightarrow$

Circle	Square	Triangle
0.98	0.018	0.002
0.99	0.002	0.008
0.006	0.002	**0.992**

AlexNet take home points

- Similar architecture to LeNet-5, but with more convolutional layers
- **ReLU activation functions** - faster computation, more efficient training
- **Dropout** to prevent overfitting
- Training on multiple GPUs

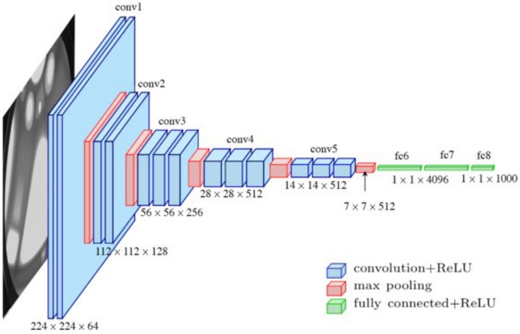
VGG

VGG

- "Very Deep Convolutional Networks for Large-Scale Image Recognition", Karen Simonyan and Andrew Zisserman, 2015
- Very popular architecture for image analysis
- Very deep network, with 16 layers (VGG-16) or 19 layers (VGG-19). 130M parameters
- Winner of ILSVRC in 2015.
- VGG-19 is slightly better, but more computationally expensive (in practice VGG-16 more common).

VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

Karen Simonyan[*] & Andrew Zisserman[+]
Visual Geometry Group, Department of Engineering Science, University of Oxford
{karen,az}@robots.ox.ac.uk



VGG take home points

- Very deep network, 130M parameters
- Uses small convolutions (3x3) with stride 1
- All layers have same configuration (simplified hyperparameter choice)
- 1×1 convolutions to increase non-linearity

GoogLeNet

- "Going Deeper with Convolutions", Szegedy et al. 2014
- Moves away from the structure we've seen so far
- Introduces "Inception" modules
- 12x less parameters than AlexNet but much more accurate!
- Newer versions (Inception v3, v4) have more powerful architectures

GoogLeNet architecture

Going deeper with convolutions

Christian Szegedy
Google Inc.

Wei Liu
University of North Carolina, Chapel Hill

Yangqing Jia
Google Inc.

Pierre Sermanet
Google Inc.

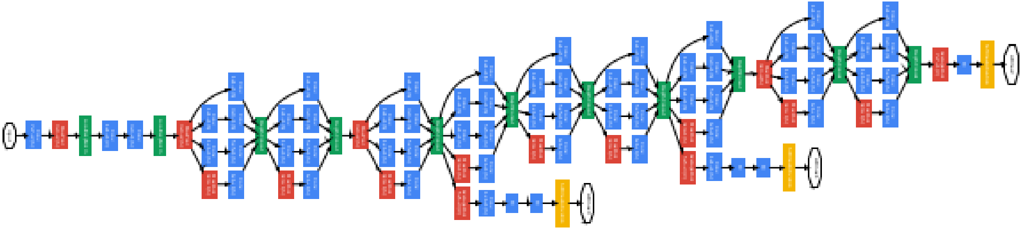
Scott Reed
University of Michigan

Dragomir Anguelov
Google Inc.

Dumitru Erhan
Google Inc.

Vincent Vanhoucke
Google Inc.

Andrew Rabinovich
Google Inc.



GoogLeNet take home points

- 22 layers
- Heavily relies on 1×1 convolutions
- Inception modules allow multi-scale feature extraction
- Drops FC layers
- Extra "side" classifications to improve gradient optimization in earlier layers

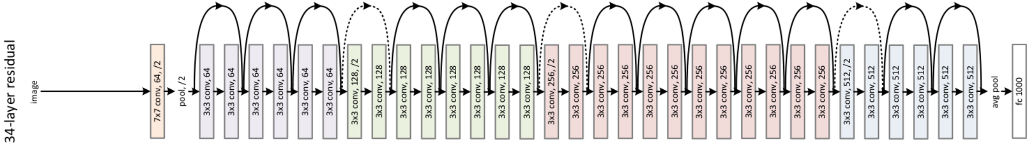
ResNet

ResNet

- He 2015 - Deep Residual Learning for Image Recognition.pdf
- Tackles the problem of degraded performance in larger networks
- Introduces *skip connections* between layers
- Up to 1000+ layers!

Deep Residual Learning for Image Recognition

Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun
Microsoft Research
{kahe, v-xiangz, v-shren, jiansun}@microsoft.com



ResNet take home points

- Very deep network (up to 1000+ layers)
- Uses *skip connections* between layers
- Uses *bottleneck* blocks (similar to GoogLeNet)

Comparison of CNN architectures