Convolutional neural networks

E. Decencière

MINES ParisTech
PSL Research University
Center for Mathematical Morphology

Contents

Contents

A picture is worth a thousand words

Definition

- Classically, an image is a matrix of values belonging to $[0, \ldots, 255]$ (grey level images) or to $[0, \ldots, 255]^3$ (color images).
- More generally, an image is a q-dimensional array of values belonging to R^d.

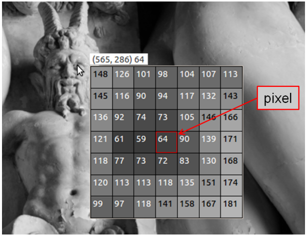


Figure: Grey level values around the left eye of the faun.

Designing computer vision systems that are able to extract semantic information from an image is a difficult task. How can we build systems that extract meaning from an image?

Extracting semantic information from an image

- Where is the phone? (localization task)
- How many mugs are there? (quantification task)
- Is there a window in the room?
- At what time of the day was the photograph taken?

Applying machine learning to images

Classical approach

- Compute features from the image
- Apply machine learning to those features

Modern neural networks approach

- Directly take as input the image pixels
- The network is supposed to build its own features

Some accomplishments

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

2012: *AlexNet* [Krizhevsky et al., 2012] won this challenge by a large margin

The database contains more that 1 million training images, belonging to 1000 different classes (including 120 dog breeds!).

Some accomplishments (cont.)

- 2011: first super-human performance, IJCNN 2011 traffic sign recognition contest [Cireșan et al., 2011]
- 2012: visual object detection (Mitosis detection in breast cancer histology [Cireșan et al., 2013])
- 2012: segmentation competition (neuronal membranes in electron microscopy images [Ciresan et al., 2012])
- 2016: AlphaGo beats Lee Sedol, one of the best go players, in a 5-game match

Convolutional neural networks in deep learning

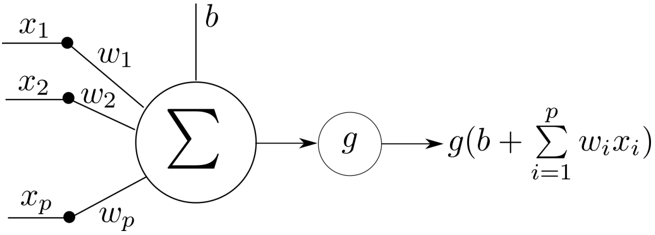
- They are pivotal to many of the successes achieved by neural networks these recent years
- They are interesting for dealing with regular structured data, such as images (or board games!)

Two acronyms are used for convolutional neural networks in the literature: *CNN* and *ConvNet*.

Contents

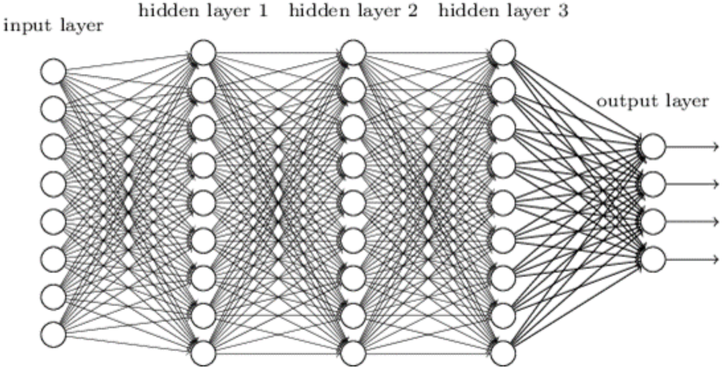
Contents

Most common artificial neuron



- $b, w_1, \ldots, w_n$ are the neuron parameters, to be learnt
- α is the activation or transfer function

Neural network



(from http://www.jtoy.net)

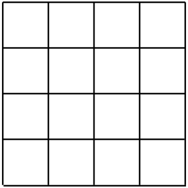
Dealing with images

Fully connected layers:

- scale badly to large size images
- do not take into account the local structure of images

Input image, input neurons

In the scalar case, each input pixel is considered as an input neuron.



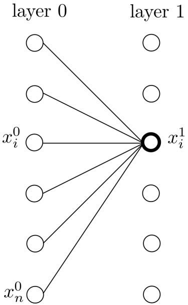
In the following slides, for illustration purposes, we will consider one-dimensional images

Input image, input neurons

In the scalar case, each input pixel is considered as an input neuron.

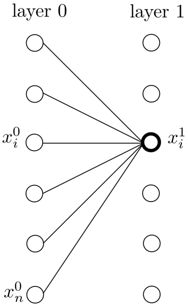
In the following slides, for illustration purposes, we will consider one-dimensional images

Towards convolutional layers

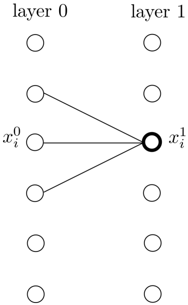


Fully connected layer: $n(s+1)$ weights

Towards convolutional layers

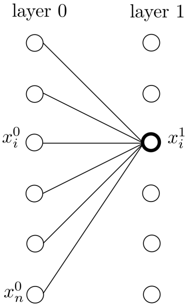


Fully connected layer: $n(s+1)$ weights

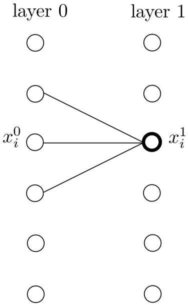


Locally conn. layer: $n(s+1)$ weights

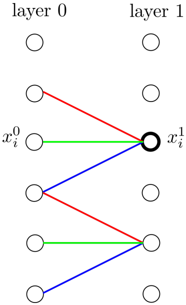
Towards convolutional layers



Fully connected layer: $n(s+1)$ weights

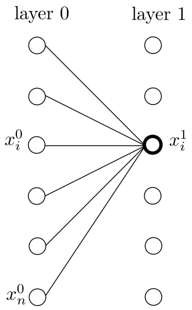


Locally conn. layer: $n(s+1)$ weights

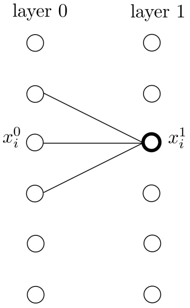


Weight replication: $s+1$ weights

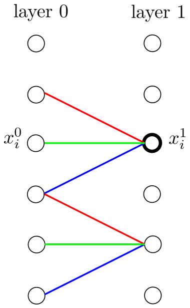
Towards convolutional layers



Fully connected layer: $n(s+1)$ weights



Locally conn. layer: $n(s+1)$ weights

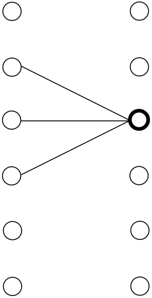


Weight replication: $s+1$ weights

A convolutional layer computes a convolution, plus a constant, of the precedent layer.

Stride

A convolutional layer can at the same time downsample the image by applying a sampling step, or *stride*.

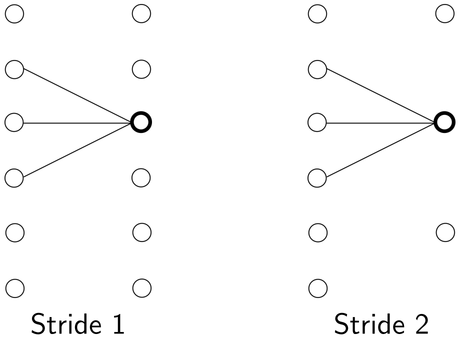


Stride 1

Note that today, for reducing the layers size, max pooling layers are often preferred.

Stride

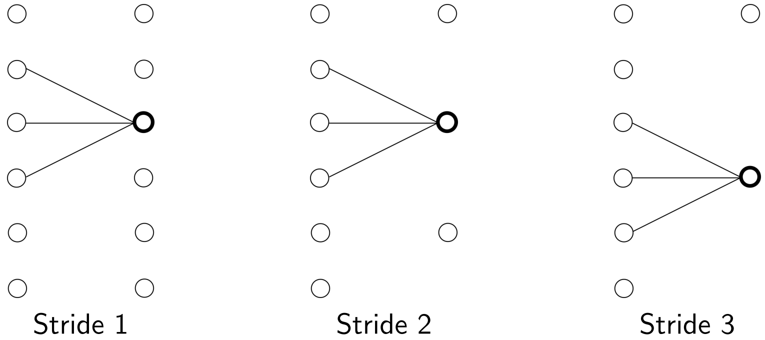
A convolutional layer can at the same time downsample the image by applying a sampling step, or *stride*.



Note that today, for reducing the layers size, max pooling layers are often preferred.

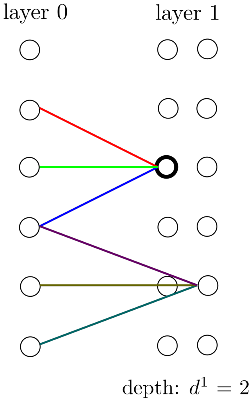
Stride

A convolutional layer can at the same time downsample the image by applying a sampling step, or *stride*.

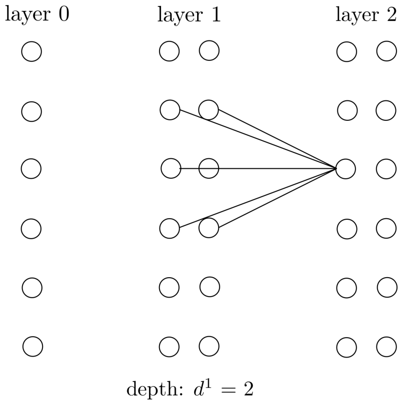


Note that today, for reducing the layers size, max pooling layers are often preferred.

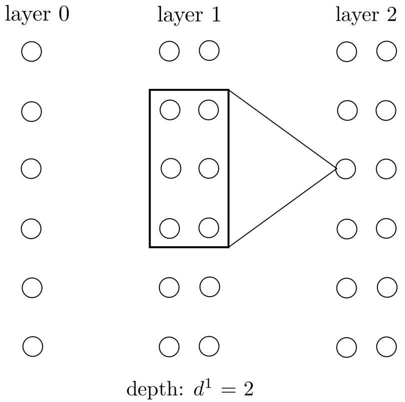
Several filters in the same convolutional layer



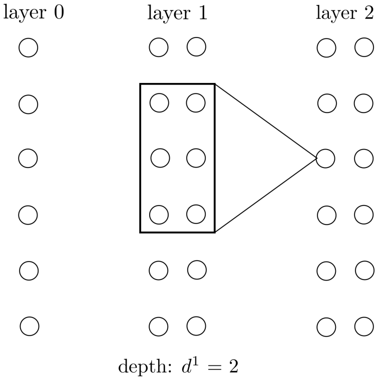
Several filters in the same convolutional layer



Several filters in the same convolutional layer

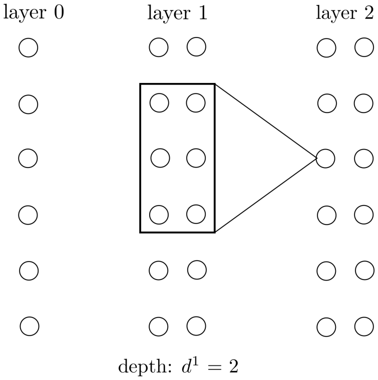


Consequences on the parameter number



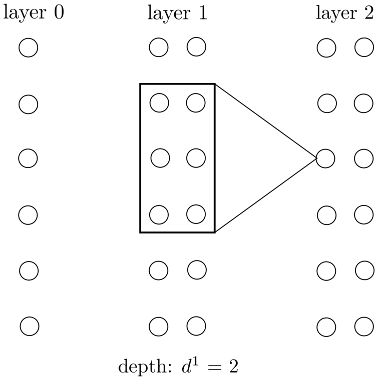
- How many parameters do we have in layer 1?

Consequences on the parameter number



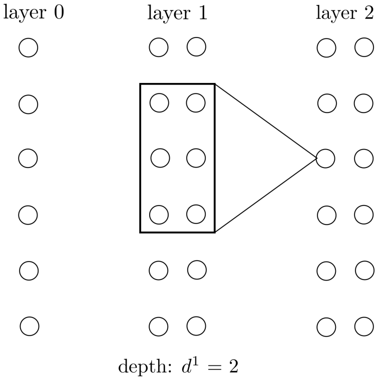
- How many parameters do we have in layer 1?
- $d^1 \times (s + 1)$

Consequences on the parameter number



- How many parameters do we have in layer 1?
- $d^1 \times (s+1)$
- In layer 2?

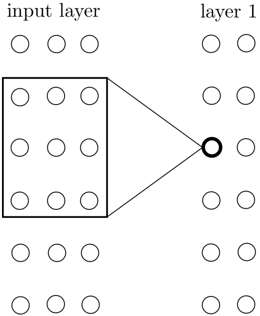
Consequences on the parameter number



- How many parameters do we have in layer 1?
- $d^1 \times (s + 1)$
- In layer 2?
- $d^2 \times (d^1 \times s + 1)$

Multi-valued images

An input image with p channels (for instance à colour image with 3 channels) can be represented by an input layer of depth 3



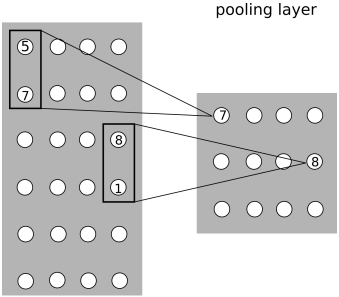
Some properties

- Translation invariance
- Efficient implementation using matrix operations and graphical processing units

Contents

Max pooling

- Convolutional networks often contain subsampling steps. These can be done using strides of 2 or more within convolutional layers or, as it is common practice today, using *max pooling* layers.
- Sampling is only applied along the spatial dimensions, not along the dimension of the filters.

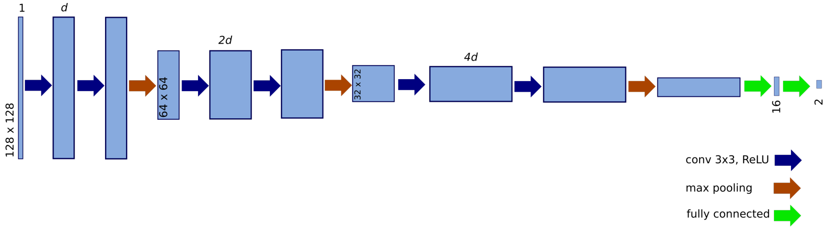


Main components of a convolutional neural network

Many successful architectures, especially for image classification, follow the same pattern:

1. Several iterations of: One or several convolutional layers, with increasing depth, followed by max pooling
2. A few fully connected layers

Example networks: fovea detection



Example: VGGnet

- Proposed by K. Simonyan and A. Zisserman from the University of Oxford [Simonyan and Zisserman, 2014]
- Number of parameters (VGG16): 138 million.

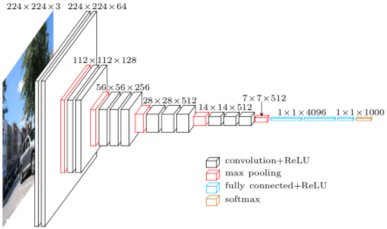
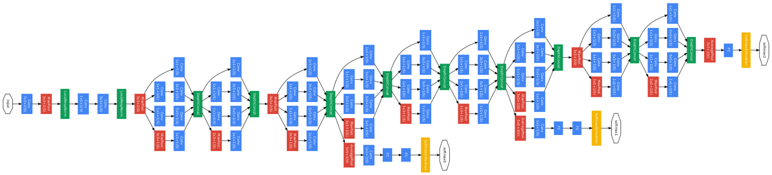


Figure: VGG16 (From https://www.cs.toronto.edu/ frossard/post/vgg16/)

Example: GoogLeNet

- This network won the ImageNet Large Scale Visual Recognition Competition (ILSVRC) in 2014.
- Number of parameters: *only* 5 million.



ResNet won the following year. One of its versions contains more than 1000 layers.

Current trends

- Small convolutions (3×3)
- Increasing number of layers - more than 1000
- Skip connections

Contents

Designing a convolutional neural network

- Network architecture
- Hyper-parameter setting
- Optimization method and parameters

Learning, specially for complex networks, can be difficult. It is always time consuming.

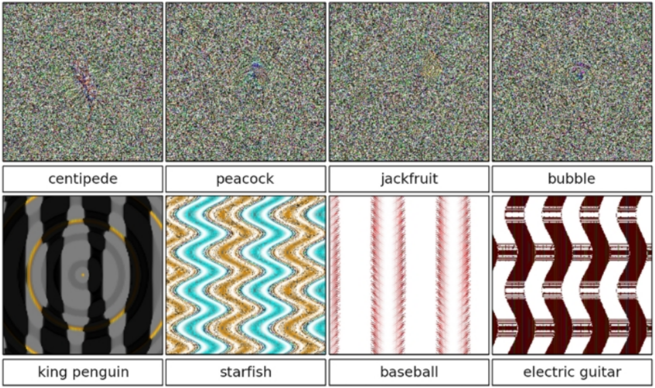
Using an existing network

- Replicate network architecture and hyper-parameters
- Load weights

With standard libraries this is very simple. Prediction time is usually fast.

ConvNets can be fooled

Deep learning can produce astonishing results [Nguyen et al., 2015]...



Main deep learning libraries

Deep learning is a very competitive domain, where code sharing is very common.

- Theano
- Torch
- Tensorflow
- Caffe
- MatConvNet

Keras

Keras is a very easy to use interface to Theano and Tensorflow. See demonstration.

Contents

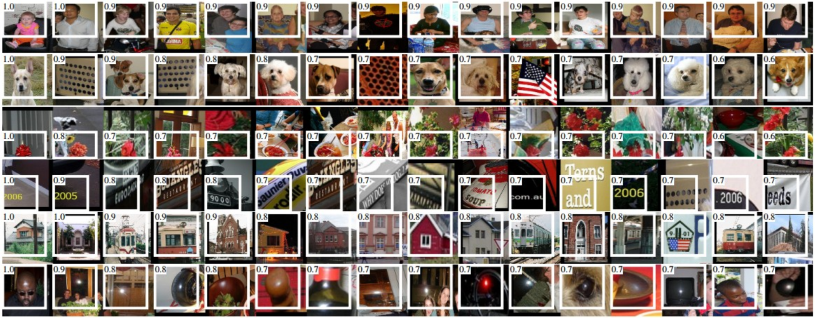
Visualizing at least what we don’t understand

We do not fully understand today where the good performances of deep learning come from. But we can at least have a look at what is going on inside the networks.

- Neuron outputs (activations)
- Filter values
- “Important” pixels

Maximal neuron activation

Which images maximally activate a given neuron?
[Girshick et al., 2014]



Contents

Limitations

For a deep-learning solution to work, you need:

- A lot of annotated data
- A lot of fiddling (different architectures; hyper-parameters)
- One (or, even better, several) powerful GPUs

Moreover, these models lack interpretability, and fitting them can be very time-consuming.

Some missing topics

- Transfer learning / domain adaptation
- Data augmentation
- Generative adversarial networks (GANs)
- Optimization, gradient descent, regularization
- ...

(These are not specific to ConvNets)

Some useful pointers

- Course by G. Hinton on neural networks (coursera, youtube)
- Stanford course on convolutional neural networks: http://cs231n.github.io/
- Hardware recommendations: https://timdettmers.wordpress.com/2014/08/14/which-gpu-for-deep-learning/

Conclusion

References I

[Ciresan et al., 2012] Ciresan, D., Giusti, A., Gambardella, L. M., and Schmidhuber, J. (2012). Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 25*, pages 2843–2851. Curran Associates, Inc.

[Cireșan et al., 2011] Cireșan, D., Meier, U., Masci, J., and Schmidhuber, J. (2011). A committee of neural networks for traffic sign classification. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 1918–1921. IEEE.

[Cireșan et al., 2013] Cireșan, D. C., Giusti, A., Gambardella, L. M., and Schmidhuber, J. (2013). Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, pages 411–418. Springer, Berlin, Heidelberg.

[Girshick et al., 2014] Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *CVPR 2014*, pages 580–587.

[Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc.

References II

[Nguyen et al., 2015] Nguyen, A., Yosinski, J., and Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 427–436.

[Simonyan and Zisserman, 2014] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.