Deep Learning course

Session 10 – Convolutional Neural Networks (ConvNets)

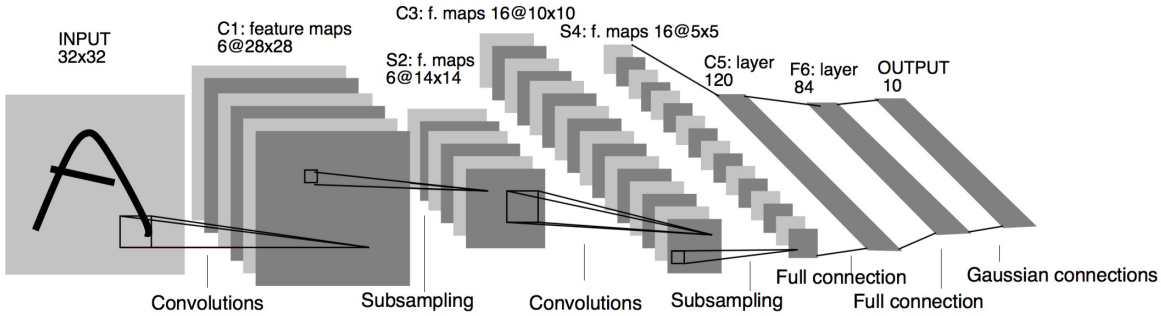
E. Francisco Roman-Rangel
edgar.roman@alumni.epfl.ch

CInC-UAEM. Cuernavaca, Mexico. September 29th, 2018.

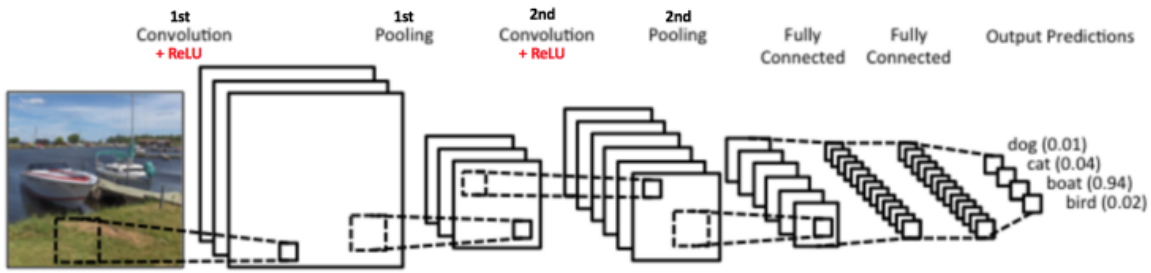
Outline

ConvNet (CNN)

Exploit spatial local structure applying (convolving) spatial filters.



LeNet



Pipeline

Convolution

Etymology

Convolution (*lat.* Convolvere): *volvere* (roll), *com* (together).

Definition

Roll together. Entwine. Merged shapes.
(German: *faltung*, i.e., folding).
Combine one function (e.g., *Image*) with another (e.g., *filter*).

$$(f * g)(t) = \int_0^t f(t-\tau)\, g(\tau)\, d\tau,$$

$$(I * k)[x, y] = \sum_{i,j} I[x-i, y-j] k[i, j],$$

Applications

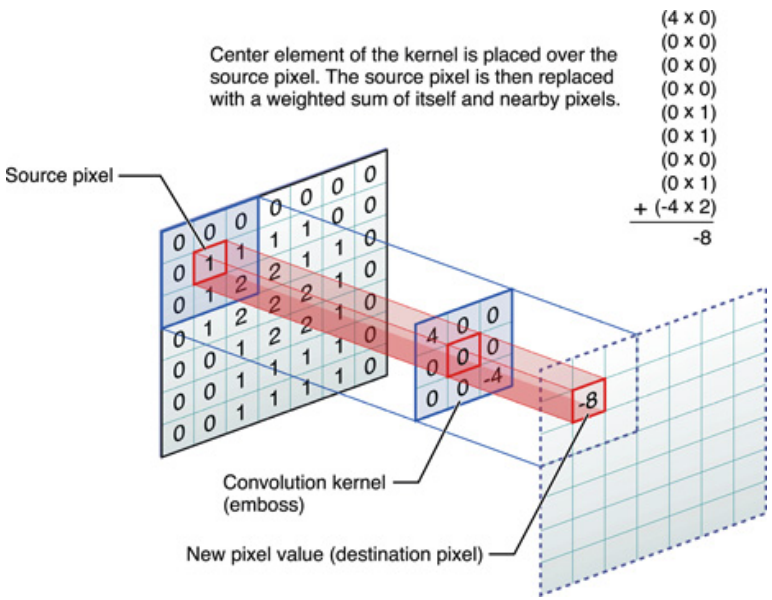
- *Statistics*: the probability distribution of the sum of two random variables is the convolution of each of their distributions.
- *Optics*: (1) shadow = convolution of the shape of the light source and an object; (2) out-of-focus photograph = convolution of a sharp image with a blur circle.
- *Acoustics*: echo = convolution of a sound with a function representing an object reflecting it.

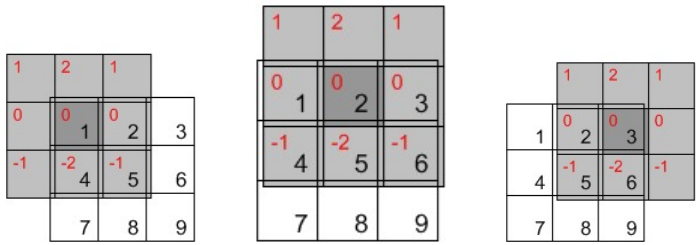
Convolution in computer vision

Convolve an Image with a filter.

$$(I * k)[x, y] = \sum_{i,j} I[x - i, y - j]k[i, j]$$

Different sizes:





Resulting size

$$I[M \times N] * k[m \times n] \rightarrow [M - m + 1, N - n + 1]$$

Padding

- Zeros:
  $[5|4|2|3|7] \rightarrow [0|5|4|3|2|7|0]$
- Extended:
  $[5|4|2|3|7] \rightarrow [5|5|4|3|2|7|7]$
- Cyclic:
  $[5|4|2|3|7] \rightarrow [7|5|4|3|2|7|5]$
- Undefined:
  $[5|4|2|3|7] \rightarrow [?|5|4|3|2|7|?]$

Definition

Convolution:

$$I * k = \sum_{i,j} I[x-i, y-j]k[i,j]$$

- Smoothing.
- Sharpening.

Correlation:

$$I \circ k = \sum_{i,j} I[x+i, y+j]k[i,j]$$

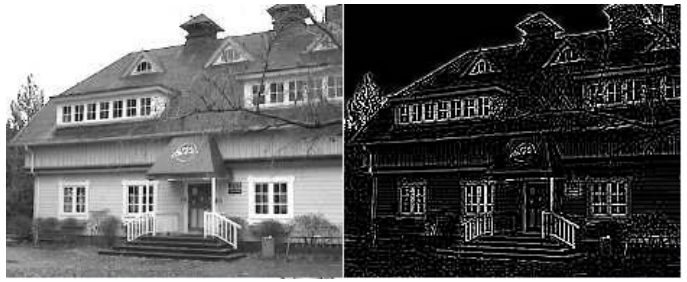
- Template matching.
- Edge detector.

Image Applications

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

Average blur.

0	0	-1	0	0		
0	-1	-2	-1	0		
-1	-2	16	-2	-1		
0	-1	-2	-1	0		
0	0	-1	0	0		

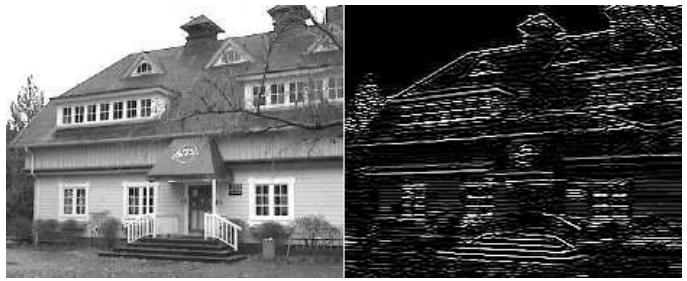


Laplacian of Gaussian (LoG).

0	0	0	5	0	0	0
0	5	18	32	18	5	0
0	18	64	100	64	18	0
5	32	100	100	100	32	5
0	18	64	100	64	18	0
0	5	18	32	18	5	0
0	0	0	5	0	0	0

Gaussian blur.

-1	-1	-1		-1	2	-1
2	2	2		-1	2	-1
-1	-1	-1		-1	2	-1
-1	-1	2		2	-1	-1
-1	2	-1		-1	2	-1
2	-1	-1		-1	-1	2



Horizontal line detection.

0	-1	0
-1	4	-1
0	-1	0

-1	-1	-1
-1	8	-1
-1	-1	-1

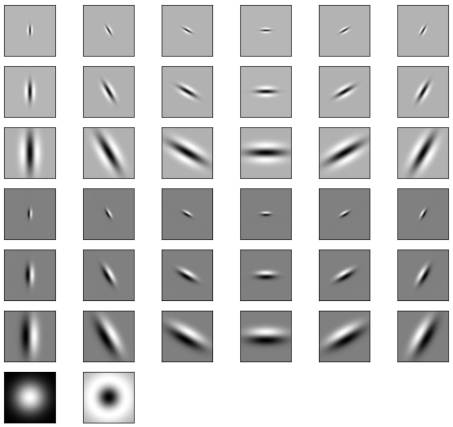
Laplacian.

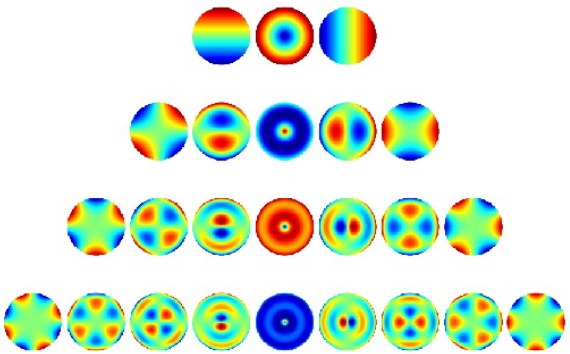
-1	-1	-1
-1	8	-1
-1	-1	-1

Edge detection.

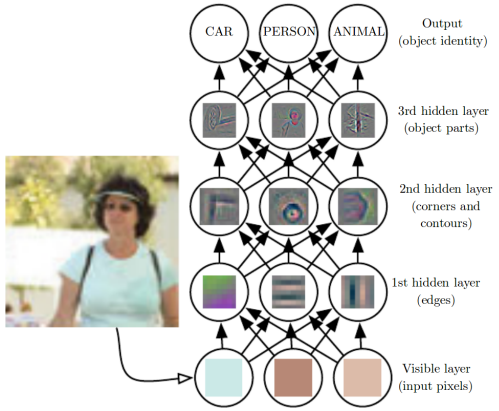
Hand-crafted vs Neural-learned filter

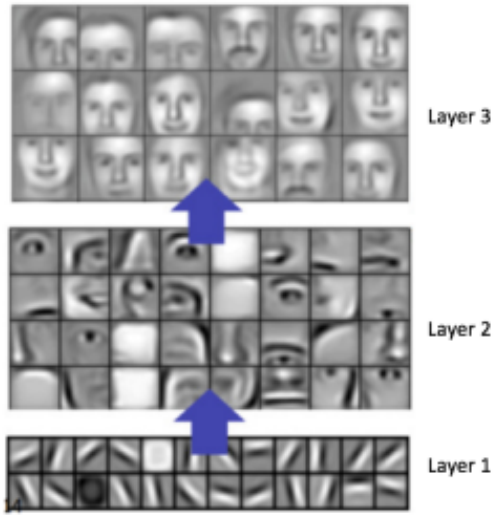
Old way:





Deep Learning way:





Convolution & Correlation

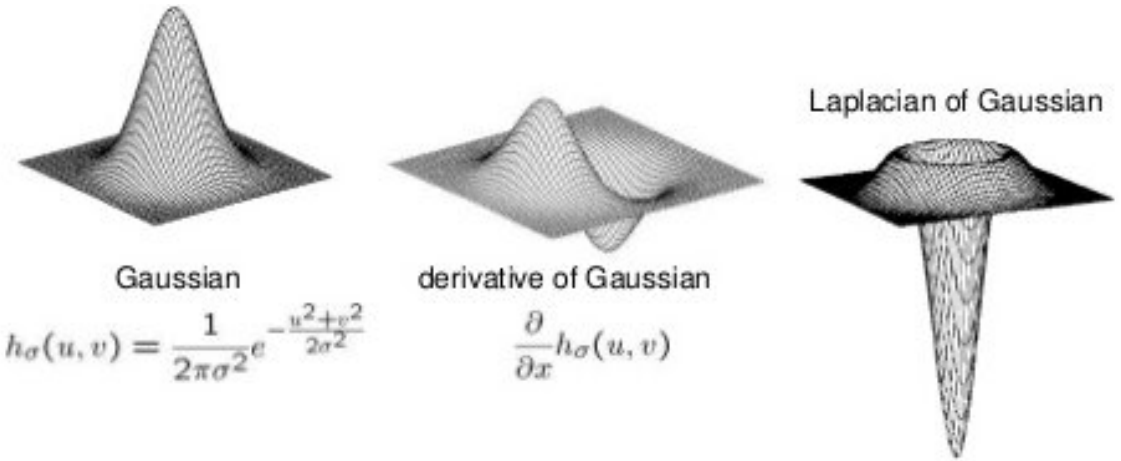
- *Both* are **shift invariant**.
- *Both* are **linear**.
- *Convolution* is **associative**. *Correlation* is not.
- *Convolution* is **commutative**. *Correlation* is not.
- No difference for symmetric filters.

Properties

e.g., associative:
Laplacian-of-Gaussian (LoG) by pre-convolved kernel

$$L * G * I = LoG * I$$

$$LoG = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.01 & 0.08 & 0.01 & 0.00 \\ 0.00 & 0.08 & 0.62 & 0.08 & 0.00 \\ 0.00 & 0.01 & 0.08 & 0.01 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix} * \begin{bmatrix} 0.17 & 0.67 & 0.17 \\ 0.67 & -3.33 & 0.67 \\ 0.17 & 0.67 & 0.17 \end{bmatrix} = \begin{bmatrix} 0.00 & 0.02 & 0.06 & 0.02 & 0.00 \\ 0.02 & 0.17 & 0.18 & 0.17 & 0.02 \\ 0.06 & 0.18 & -1.8 & 0.18 & 0.06 \\ 0.02 & 0.17 & 0.18 & 0.17 & 0.02 \\ 0.00 & 0.02 & 0.06 & 0.02 & 0.00 \end{bmatrix}$$



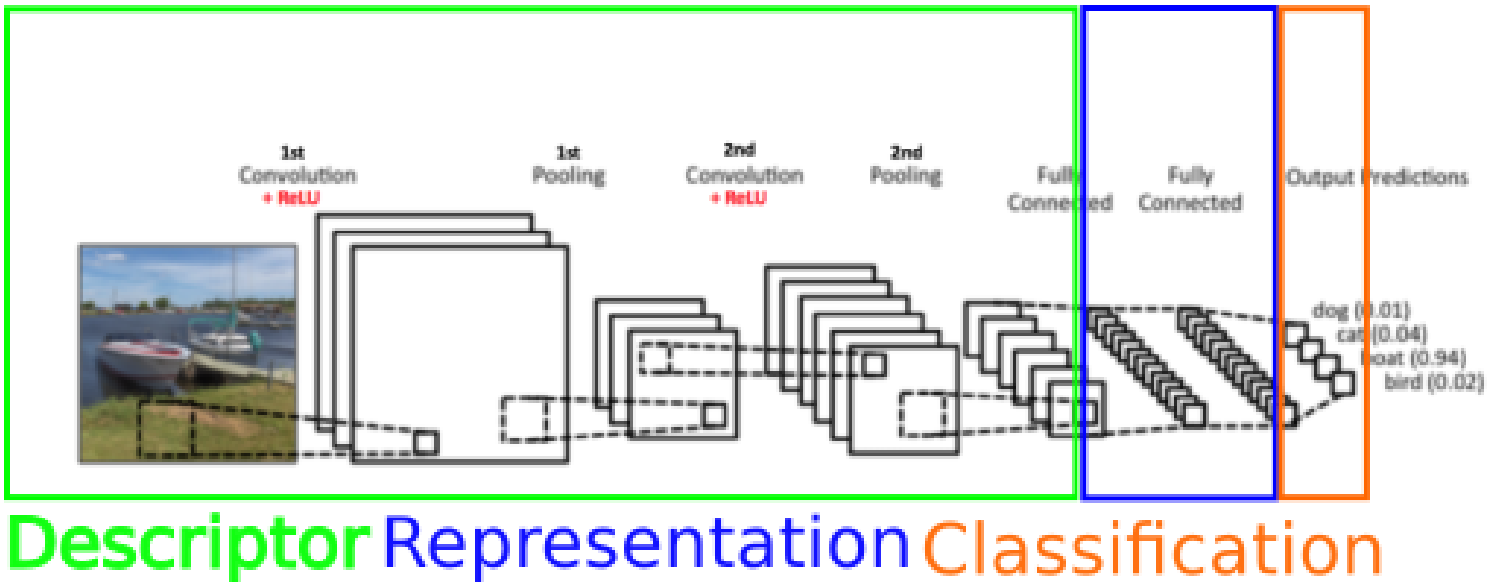
Outline

Characteristics

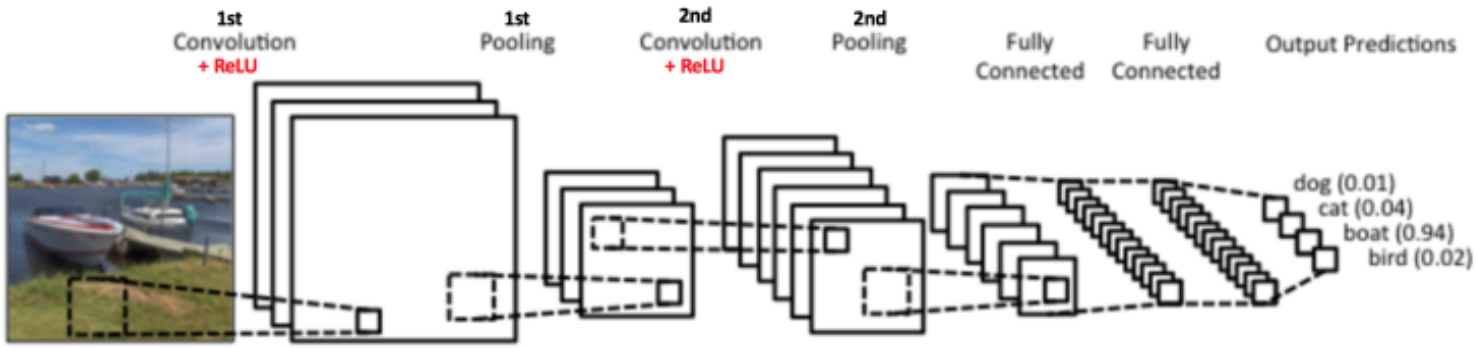
- Inspired by visual mechanisms in living organisms.
- Formed by small receptive fields.
- Receptive fields are **learned** instead of hand-crafted.
- **Sparse connectivity**: only a local section of the image is *seen*.
- **Parameter sharing**: the same receptive field can look at all local sections of the image, one at a time.
- **Equivariant representation**: if the input changes, the output changes in the same way, i.e., $f(g(x)) = g(f(x))$.

Structure

Based on the traditional Computer Vision pipeline (somehow).

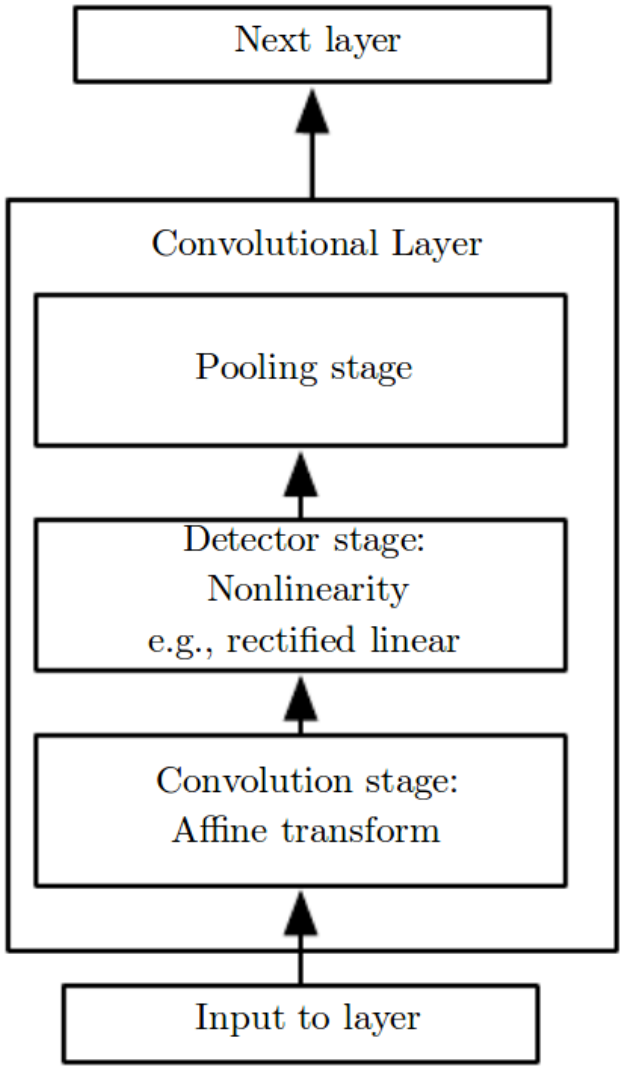


Structure



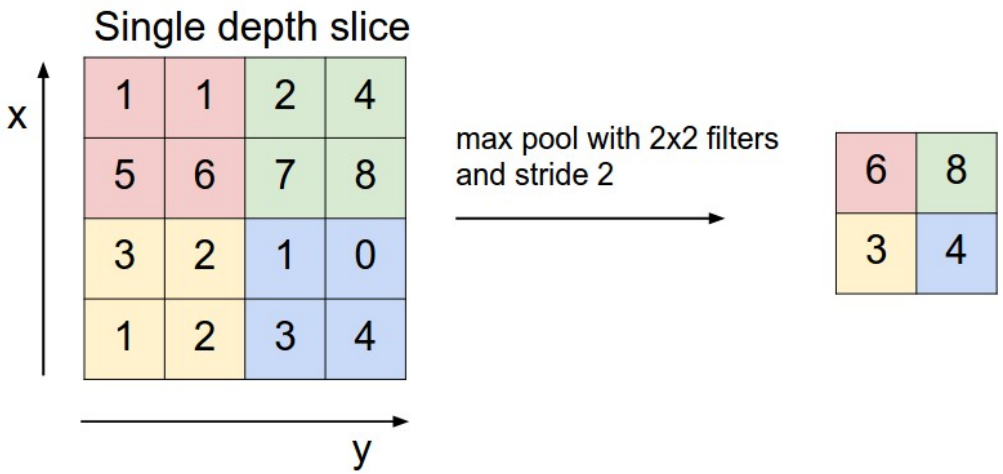
- Small receptive fields.
- Sparse connectivity.
- Parameters sharing.

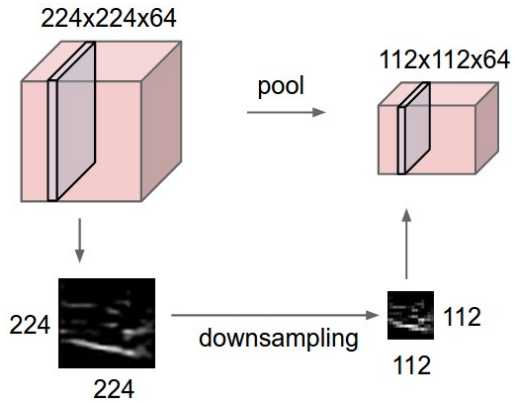
Convolutional layer



Pooling

Max pooling and average pooling.





Outline

Common practices

- Input layer divisible by 2.
- Small filters, with odd size, e.g., 3x3 to 9x9.
- Zero padding.
- Stride $= 1$.
- ReLu transfer function.
- Max pooling (2x2) vs stride $= 2$.
- BatchNorm instead of regularization.
- Increasing number of convolutional filters.

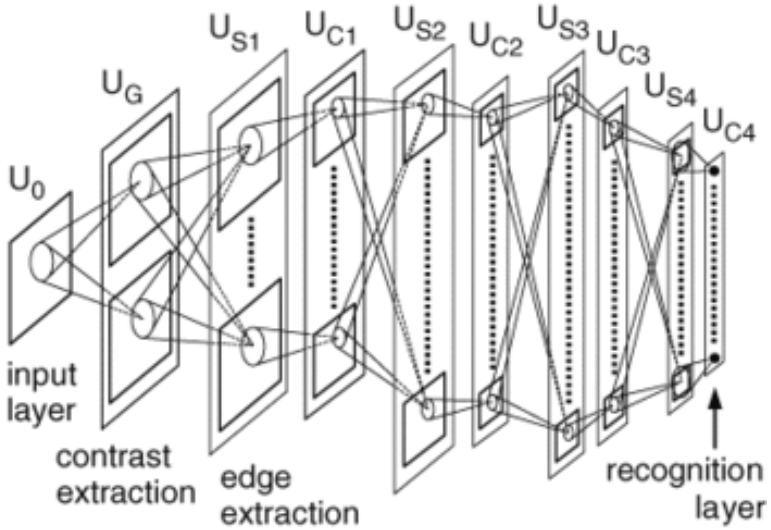
Other practices

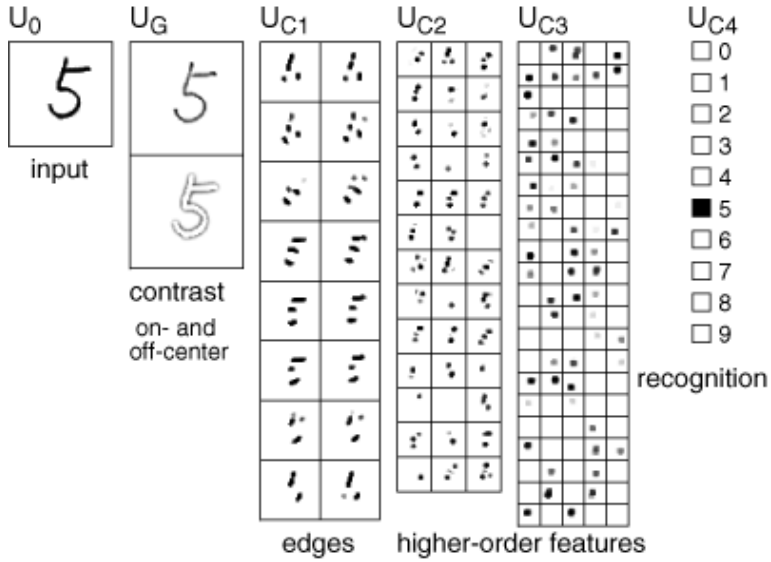
- 1-D convolution, e.g., conv2fc or fc2conv.
- Dilated convolutions
  $w[0] * x[0] + w[1] * x[2] + w[2] * x[4]$.
- Compromising stride in favor of memory (specially for GPU).

Outline

Neocognitron

Fukushima, 1980

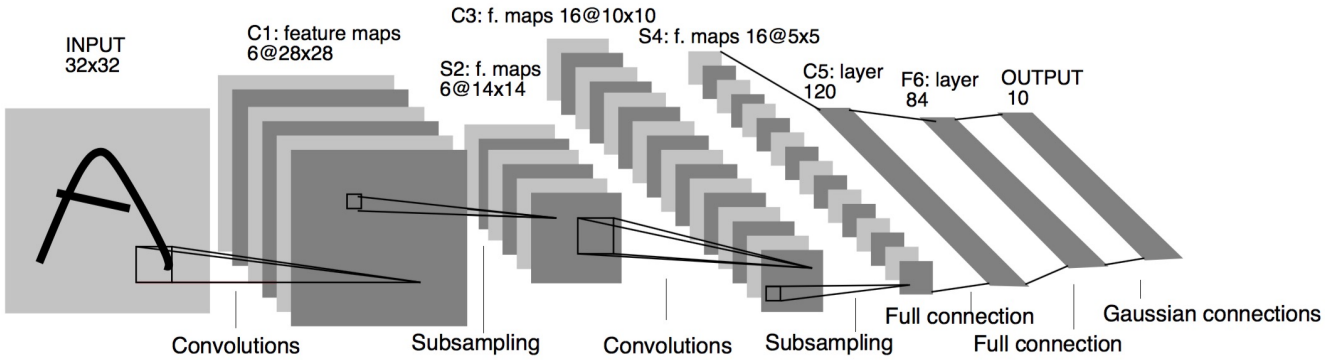




- S-cell and C-cells.
- Sublayers, aka cell-planes,
  i.e., same filter looking at different locations.
- Recognition layers.

LeNet5

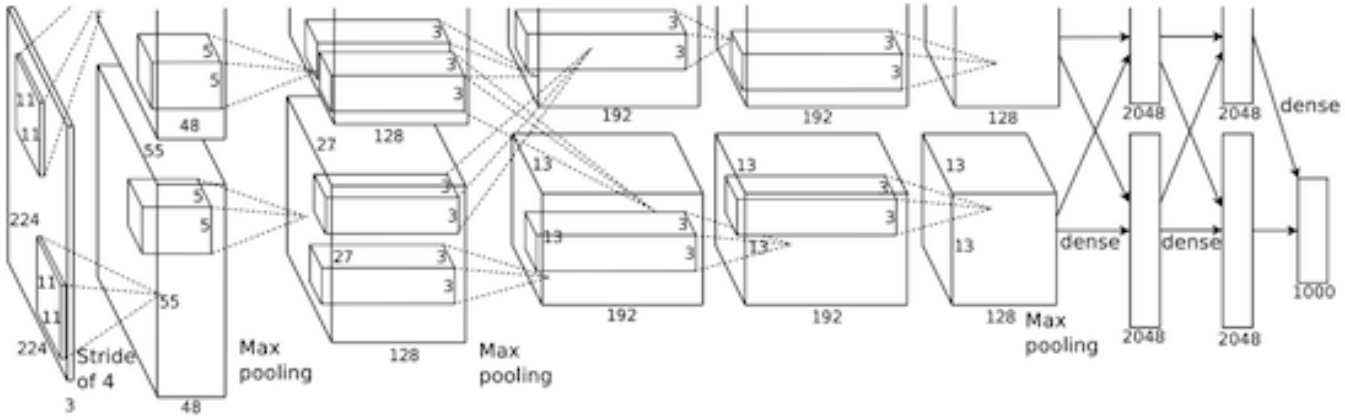
Yann LeCun et al., 1998



- Convolution to exploit local correlations.
- Nonlinearity *tanh* or *sigmoid*.
- Multilayer perceptron for classification.
- Sparse connectivity between layers.
- Trained con CPU.

AlexNet

Alex Krizhevsky et al. 2012

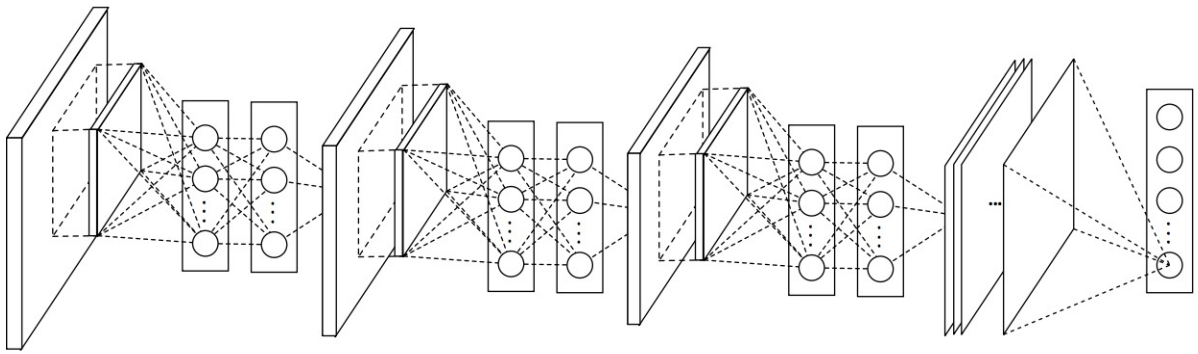


- Expanded LeNet.
- ReLu.
- Dropout.
- Max-pooling.
- Trained con GPUs*.

Variants: ZFNet (ILSVRC 2013 winner), VGGNet (Depth analysis).

Network-in-network

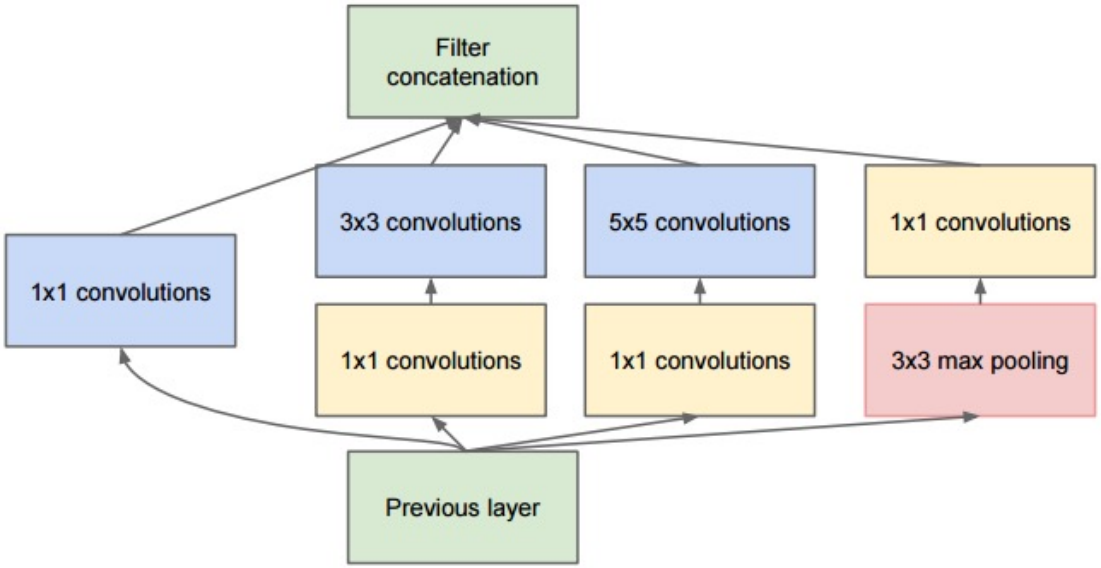
Li et al. 2014



- MLP between convolutions.
- 1x1 convolutions.

GoogleLeNet

Szegedy et al., 2014

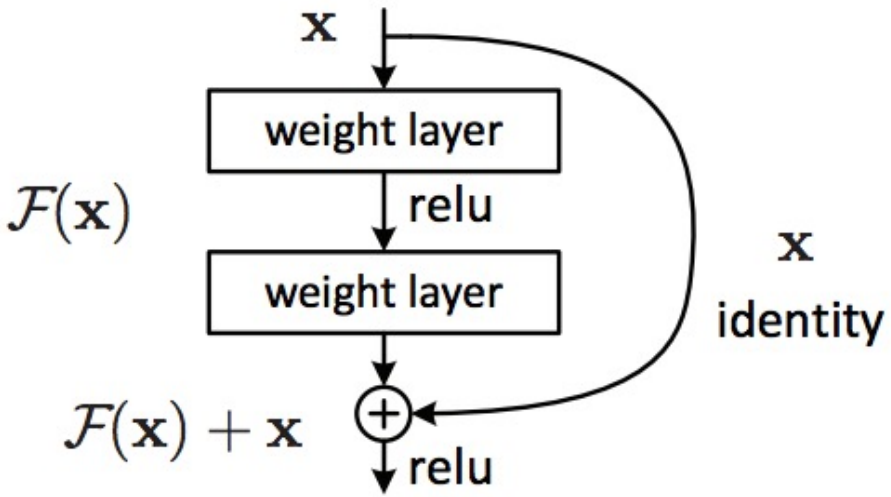


- ILSVRC 2014 winner.
- Inception module: reduce parameters.

Variants of the inception module.

ResNet

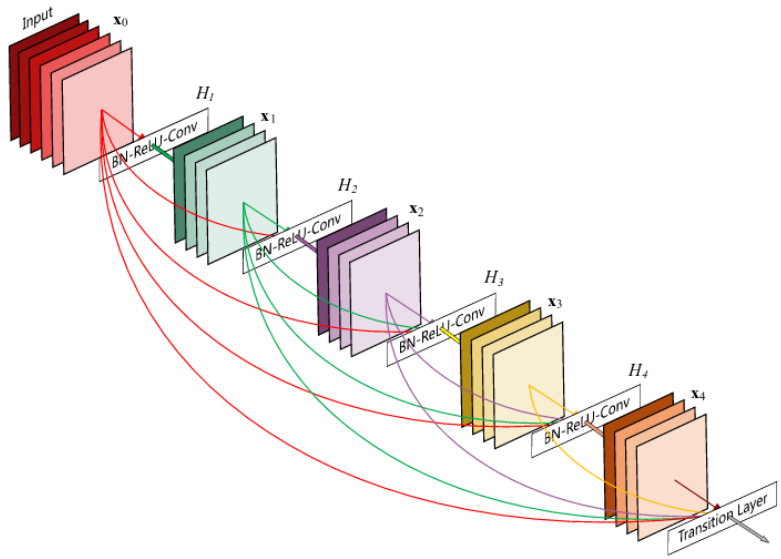
He et al., 2015

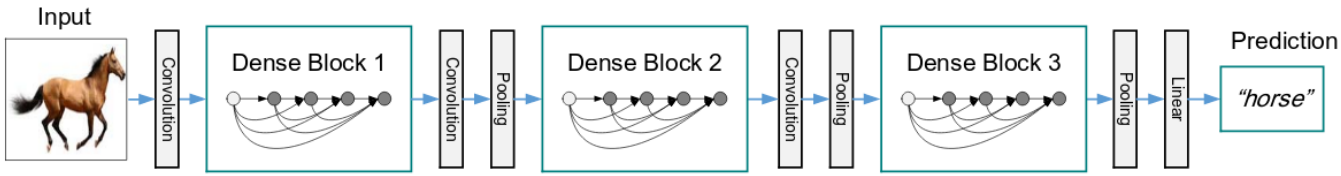


- Residual learning.
- Bypass for a sort of classifier.
- 1st network of ¿ 100 layers.
- Current state-of-the-art.
- A research topic on its own (e.g., bias?).

DenseNet

Huang et al., 2016





- Improved Residual network.
- Frustration!!!

Other architectures

- Region-based CNN (R-CNN) (localization).
- Fully connected ConvNets (semantic segmentation).
- Multi-modal ConvNets (for depth images, optical flow in videos).
- Conv AE (Local descriptors and dimensionality reduction).
- CNN + RNN (sequential data, action recognition).

To know more

- http://cs231n.github.io/convolutional-networks/
- Goodfellow, DL Book. Chapter 9.
- http://colah.github.io/posts/2014-07-Understanding-Convolutions/
- http://colah.github.io/posts/2014-07-Conv-Nets-Modular/
- https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

Thank you.

Q&A