

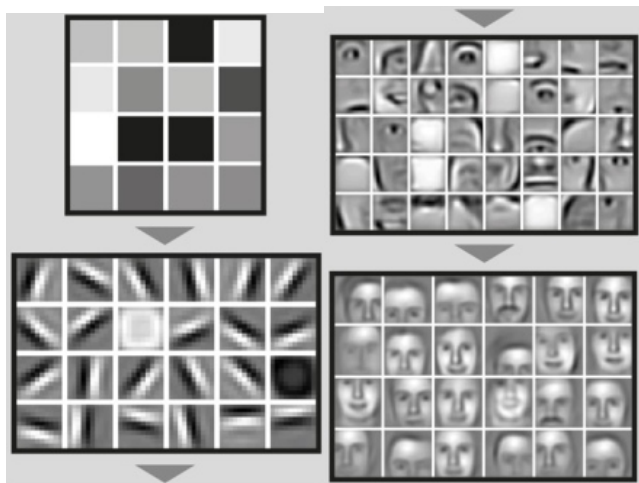
Convolution and Pooling

Benjamin Roth

CIS LMU München

December 12, 2017

Convolution and Pooling



Source: Computer science: The learning machines. Nature (2014).

Convolution and Pooling

- Convolution:
 - ▶ Sliding operator
 - ▶ Input: moving window over one instance.
 - ▶ Image/2D: patch (rectangle)
 - ▶ Text/sequence: subsequence
 - ▶ Preliminary output: Representation vector for each patch.
- Pooling
 - ▶ Operator to combine representations of all patches into vector of same size.
 - ▶ Operates component-wise (across all patches)
 - ▶ Most popular: Average pooling, Max-pooling
 - ▶ Max pooling: only select maximum value for each dimension
 - ▶ “Feature detector”, “Cat neuron fires”
 - ▶ Why maximum works? Because it is trained that way.

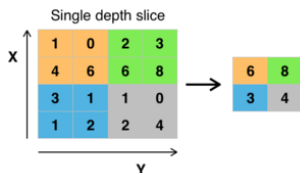
Convolution

- Tensor sizes:
 - ▶ **Input 3D tensor** (picture): colors (channels) \times input size (x-, y-direction)
 \Rightarrow Channels in hidden layers: feature activations instead of colors.
 - ▶ **Weight 4D tensor**:
number of convolutions (features/output channels) \times number of input channels \times patch size (x-, y-direction)
- The weight tensor is multiplied across all applicable patches, resulting in an **output 3D tensor** of size: output channels \times input size (x-, y-direction) ¹
- A non-linearity is typically applied on this output tensor before pooling.

¹if fully padded and stride= 1

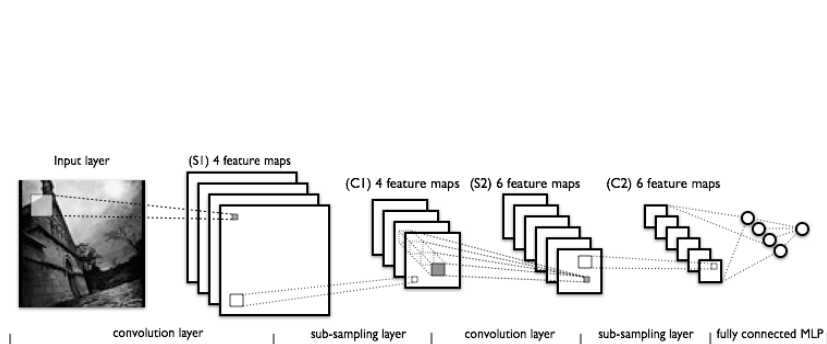
Pooling

- Note: Pooling is sometimes done only across a small number of succeeding patches.
- For example, to halve the size of the input image in each direction, pool across groups of $2 \times 2 = 4$ input patches (or rather, across their convolution outputs).



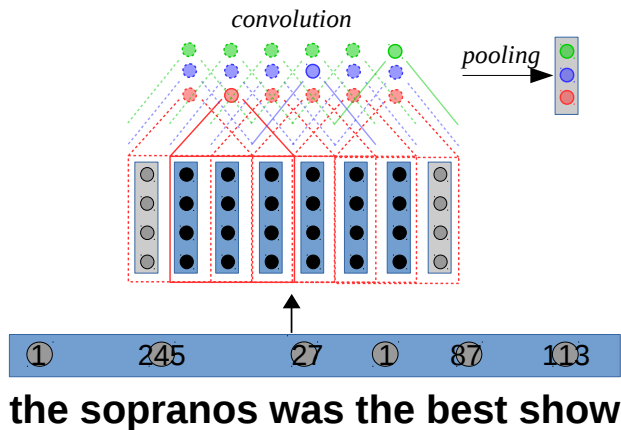
- Sometimes called “ 2×2 (downscale) stride”, “pool size / pooling region of 2×2 ”

Convolution and Pooling: LeNet



LeCun et al. (1998). Gradient-based learning applied to document recognition.

Convolution and Pooling for NLP

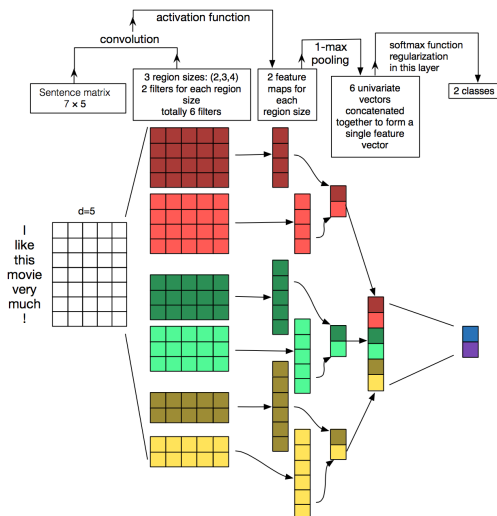


Convolution and Pooling for NLP

In the previous example:

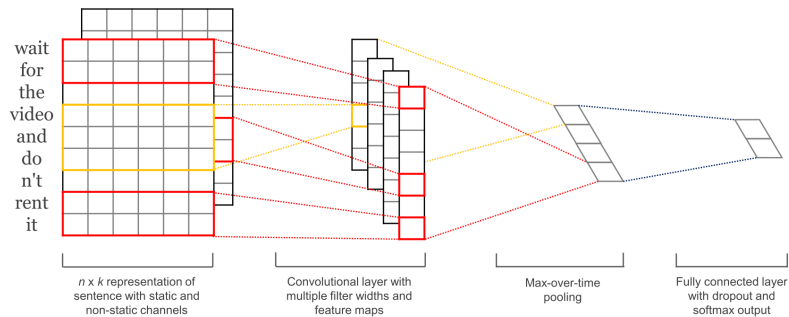
- How many filters?
- What is filter size (=filter width)?
- What stride (=step size)?
- What is padding size?
- which values would be selected by max-pooling?
- Where would non-linearity be applied?
- How many parameters are to be learned?

Convolution and Pooling for NLP



Source: Zhang, Y., & Wallace, B. (2015). A Sensitivity Analysis of ConvNets for Sentence Classification.

Convolution and Pooling for NLP



Source: Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification.