

Neural Networks and Deep Learning: Pooling

Nicolas Thome

Conservatoire National des Arts et Métiers (Cnam)
Département Informatique

Pooling

- ▶ Pooling: statistical aggregation of a set of values, e.g. $\mathbf{x} = \{x_1, \dots, x_N\}$
- ▶ Output: a single scalar value
- ▶ Possible pooling functions:
 - ▶ Max pooling: $pool(\mathbf{x}) = \max_{i \in \{1; N\}} x_i$
 - ▶ Average pooling: $pool(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N x_i$
 - ▶ ℓ_p pooling: $pool(\mathbf{x}) = \left(\frac{1}{N} \sum_{i=1}^N x_i^p \right)^{\frac{1}{p}}$

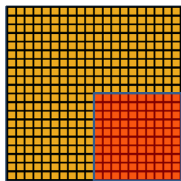


$max = 8, avg = 4.8$

- ▶ **Goal:** capture statistics of responses
 - ▶ Invariance *wrt* position of values
 - ▶ Permut values \Rightarrow same features

Pooling in Convolution Feature Maps

- ▶ **Spatial pooling:** aggregation over image (map) regions
- ▶ Pooling Input: each map (image)
- ▶ **Local Aggregation:** \Rightarrow local pooling receptive field
 - ▶ Local aggregation: keep spatial information
 - ▶ Global pooling & invariance through hierarchy



Convolved
feature

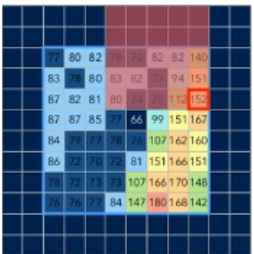
1	7
5	9

Pooled
feature

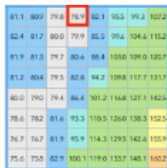
Spatial Max Pooling



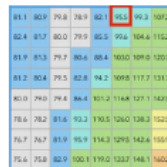
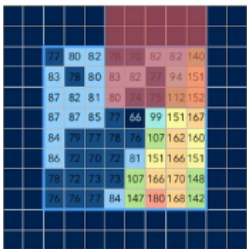
- ▶ For a given input map neuron, define:
 - ▶ Local pooling size
 - ▶ Pooling function
- ▶ **Ex: max pooling with 5×5 pooling area**
- ▶ binary input: pooling \Rightarrow presence / absence of feature in local pooling area
- ▶ (partial) Translation invariance \Rightarrow later



Spatial Average Pooling

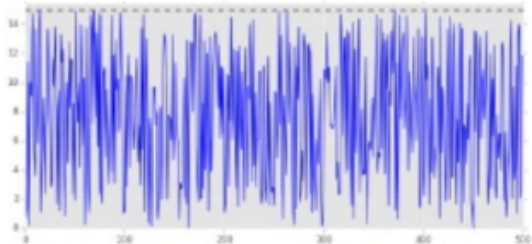


- ▶ **Ex: average pooling with 5×5 pooling area**
- ▶ binary input: pooling ~ count number of present features in local pooling area

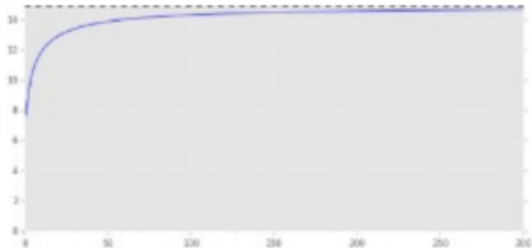


Spatial ℓ_p Pooling

Uniform, min=0, max=14.955, avg=7.712

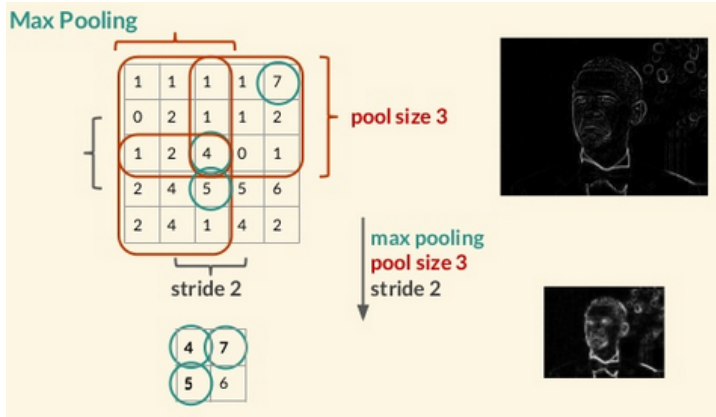


Lp pooling vs p



- ℓ_p pooling: smooth transition from average \rightarrow max (wrt p)

Spatial Pooling: Stride



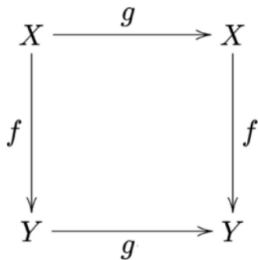
- ▶ Step s to which pooling areas centered
- ▶ $s > 1$: decreases spatial resolution \Rightarrow less parameters in Deep Models
 - ▶ \sim Downsampling

Spatial Pooling: from Equivariance to Invariance

- **Recap:** convolution equivariant to translation:

$$f[g(x)] = g[f(x)]$$

- f convolution, g translation



Map
representation
by one filter

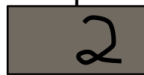
image



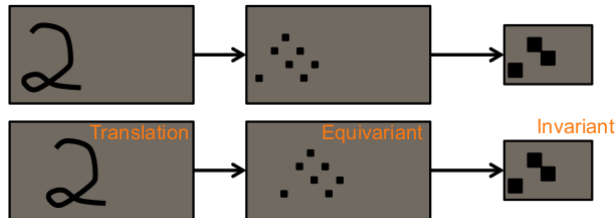
translated
representation



translated
image



Max Pooling & Translation Invariance



- Under some conditions, max pooling \Rightarrow translation invariance:

$$f' [g(x)] = f'(x)$$

- $f' = f \circ p$ with f convolution, p pooling

Max Pooling & Translation Invariance

- ▶ Translation invariance if:
 - ▶ Translation \nRightarrow new largest element at pooling region edge
 - ▶ Translation \nRightarrow remove max from pooling region
- ▶ Ex: 5×5 conv map, 3×3 max pooling centered at 15:
 $max = 15$,
- ▶ **Invariance OK:** \forall **translation** $(t_x, t_y) \in \pm 1$ **px**
 $\Rightarrow max = 15$

$$C = \begin{bmatrix} 11 & -5 & 1 & -2 & 0 \\ 1 & \boxed{3} & \boxed{0} & \boxed{0} & 5 \\ 8 & 4 & 15 & -10 & 4 \\ 8 & 6 & 5 & 3 & 7 \\ 3 & 0 & -2 & 9 & 3 \end{bmatrix}$$

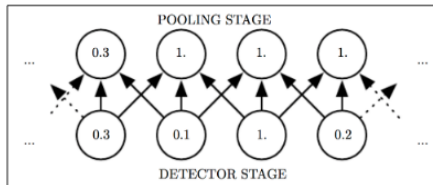
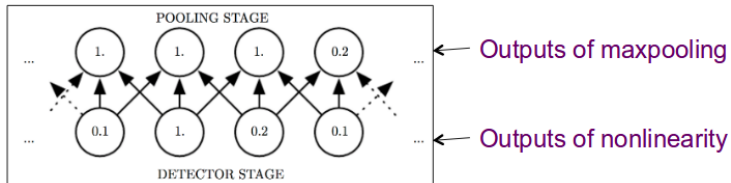
Max Pooling & Translation Invariance

- ▶ Translation invariance if:
 - ▶ Translation \nRightarrow new largest element at pooling region edge
 - ▶ Translation \nRightarrow remove max from pooling region
- ▶ Ex: 5×5 conv map, 3×3 max pooling centered at 15:
 $max = 15$,
- ▶ **Invariance KO: right translation** $t_x = +1$ px
 $\Rightarrow max = 7$

$$C = \begin{bmatrix} 11 & -5 & 1 & -2 & 0 \\ 1 & \boxed{3} & \boxed{0} & \boxed{0} & 5 \\ 8 & 15 & 4 & -10 & 4 \\ 8 & 6 & 5 & 3 & 7 \\ 3 & 0 & -2 & 9 & 3 \end{bmatrix}$$

Max Pooling & Translation Invariance

- ▶ Max pooling: partial translation invariance (under some conditions)
 - ▶ **At least local stability:** every value in the bottom row has changed, but only half of the values in the top row have changed
 - ▶ Distance after pooling decreases



Max Pooling & stability

distance decrease between pooled and convolved

Pooling: Conclusion

- ▶ Reducing spatial feature map size (stride)
- ▶ Partial translation invariance and stability
- ▶ **Invariance to richer deformations in deep models?**
- ▶ **Convolution on tensors** (color images / hierarchies)

⇒ following!

