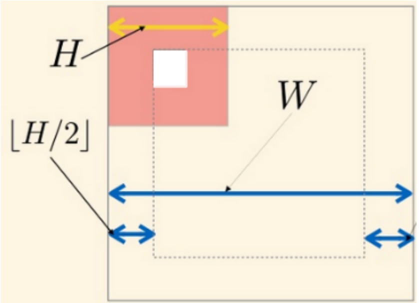
Neural Networks and Deep Learning: Other Modern Deep Learning Components

Nicolas Thome

Conservatoire National des Arts et Métiers (Cnam)
Département Informatique

Padding Modules

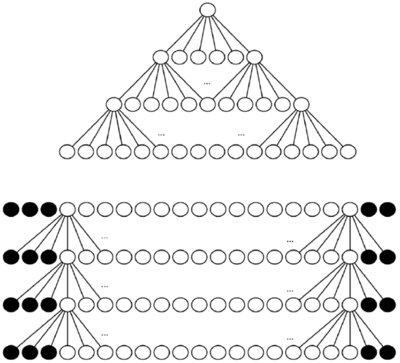


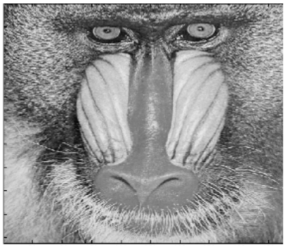
- Filter of size H: problem with the borders $\lfloor \frac{H}{2} \rfloor$
- <u>**Solution 1**</u>: reduce processing to computable area $\Rightarrow$ decreased output size
- <u>**Solution 2**</u>: padding, *i.e.* fill missing info with (arbitrary) values
  - Zero-padding, recopy, mirror, *etc*

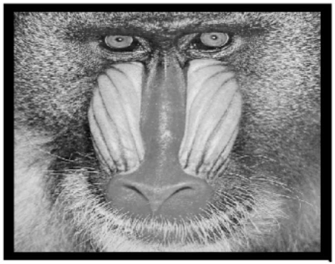
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	2	5	6	3	6	7	3	0	0
0	0	2	3	4	6	7	5	1	8	4	0	0
0	0	8	7	6	5	7	6	3	3	4	0	0
0	0	2	3	5	6	7	8	2	7	3	0	0
0	0	4	5	3	2	1	6	8	7	2	0	0
0	0	1	4	5	3	2	6	7	8	1	0	0
0	0	2	3	4	5	6	8	9	2	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0

2	2	2	3	4	6	7	5	1	8	4	4	4
1	1	1	1	2	5	6	3	6	7	3	3	3
1	1	1	1	2	5	6	3	6	7	3	3	7
3	2	2	3	4	6	7	5	1	8	4	4	8
7	8	8	7	6	5	7	6	3	3	4	4	3
3	2	2	3	5	6	7	8	2	7	3	3	7
5	4	4	5	3	2	1	6	8	7	2	2	7
4	1	1	4	5	3	2	6	7	8	1	1	8
3	2	2	3	4	5	6	8	9	2	1	1	2
3	2	2	3	4	5	6	8	9	2	1	1	2
3	1	1	4	5	3	2	6	7	8	1	1	2

Zero-Padding



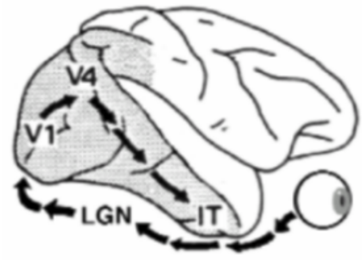




- To avoid shrinking the spatial extent of the network rapidly

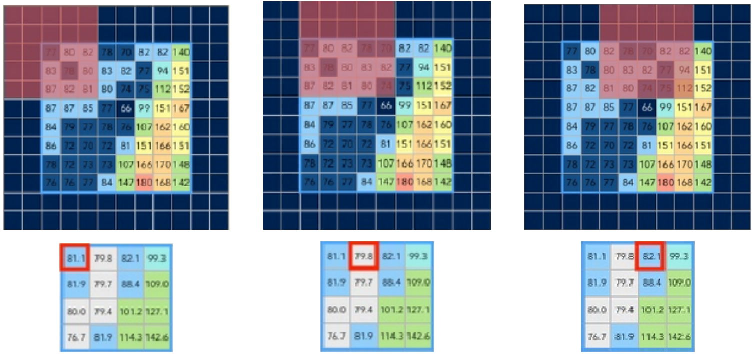
Normalization: Local Feature Normalization

- **Normalize value wrt neighbors in different feature maps**
- Operates at each spatial position independently
- Feature groups: subset of maps (sliding window)
- $\sim$ Lateral inhibition



Pooling Modules

Overlapping Pooling - Ex: pooling size: 5, stride s=2



Credit: K. Matsui

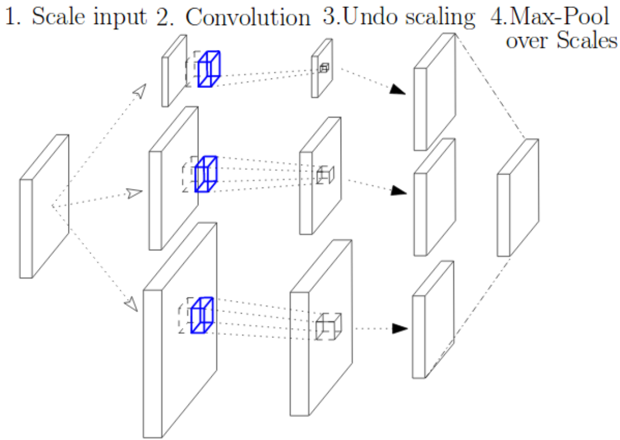
Pooling Modules

Pooling across feature maps

- Aggregation for a given spatial position, between different tensor maps
- Tensor maps (filter output) associated to a given transformation $\Rightarrow$ invariance with max pooling

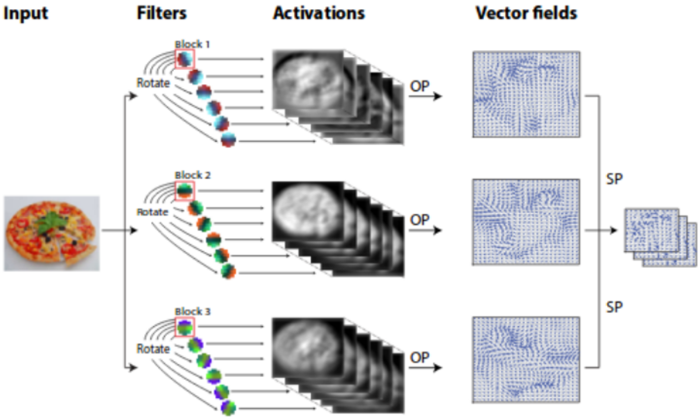
Pooling across feature maps

- Ex: scaling [Kanazawa et al., 2014]

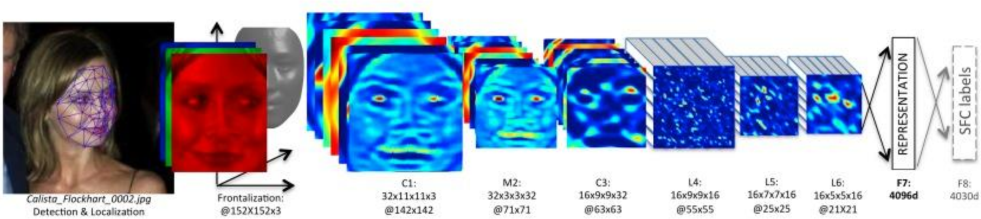


Pooling across feature maps

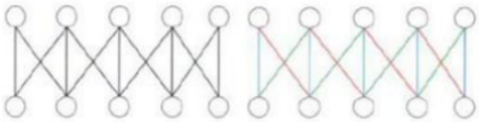
- Ex: rotation [Marcos et al., 2017]



Locally Connected *vs* Convolution Layers

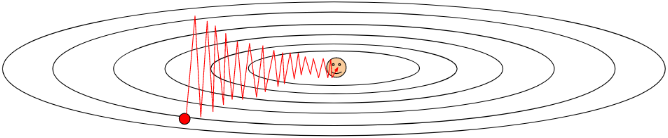


- Locally connected: different features detected across image positions
- Sucessful in specific context, *e.g.* DeepFace [Taigman et al., 2014]



Other Modern ConvNet Modules: Conclusion

- **Spatial resolution:** padding, overlapping pooling
- **Specificity and invariance:** normalization, locally connected layers, pooling across maps
- **Optimization issues?**
  $\Rightarrow$ **following!**



References I

Kanazawa, A., Sharma, A., and Jacobs, D. W. (2014).
Locally scale-invariant convolutional neural networks.
In *Deep Learning and Representation Learning Workshop: NIPS 2014*.

Marcos, D., Volpi, M., Komodakis, N., and Tuia, D. (2017).
Rotation equivariant vector field networks.
In *The IEEE International Conference on Computer Vision (ICCV)*.

Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014).
Deepface: Closing the gap to human-level performance in face verification.
In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.