

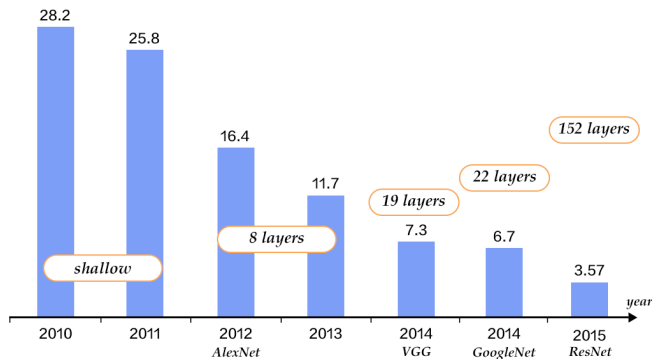
Neural Networks and Deep Learning: Modern Convolutional Neural Network Architectures

Nicolas Thome

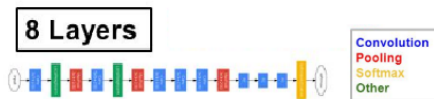
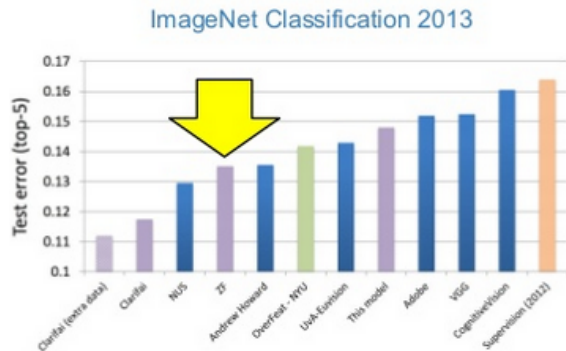
Conservatoire National des Arts et Métiers (Cnam)
Département Informatique

Deep Learning since 2012

More & more data (Facebook 10^9 images / day)
ILSVRC since 2012: larger & larger networks



Deep Learning since 2012: ImageNet'13



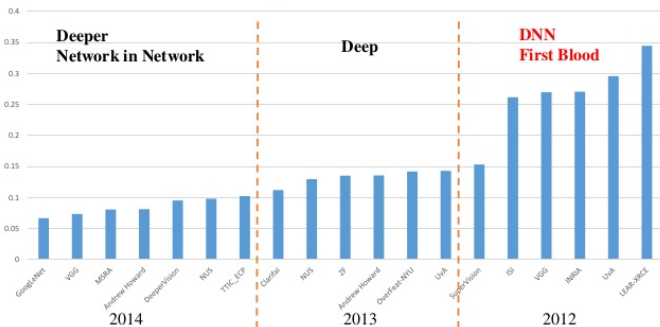
- ▶ ZF network [Zeiler and Fergus, 2014]: archi ~ AlexNet (2012), *i.e.* conv + FC
- ▶ With Local Contrast Normalization (LCN) after conv/pool

Deep Learning since 2012: ImageNet'14

ImageNet Classification

- 1000 categories and 1.2 million training images

ImageNet Classification Error



Li Fei-Fei: ImageNet Large Scale Visual Recognition Challenge, 2014 <http://image-net.org/>

9/3/2014

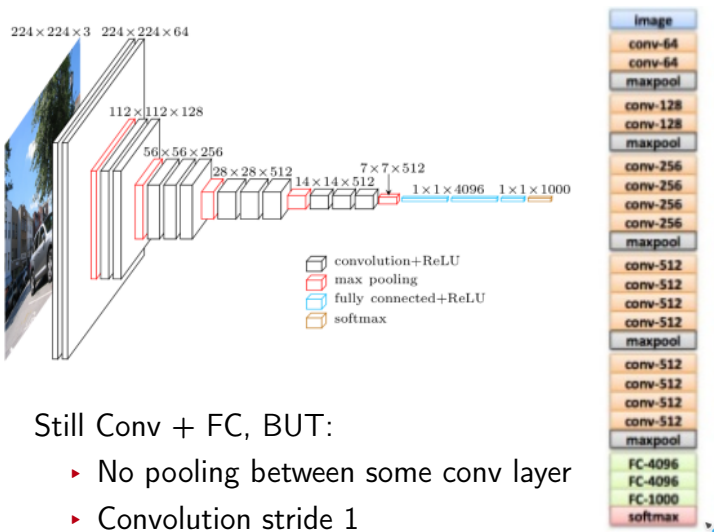
lipiji.pz@gmail.com

7

- Two leading nets: VGG & GoogLeNet

Modern ConvNet Architectures

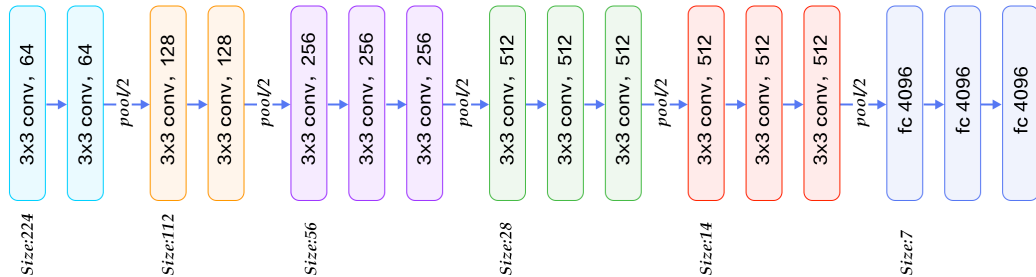
ImageNet'14: VGG [Simonyan and Zisserman, 2014]



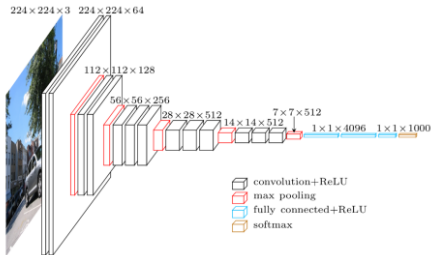
Still Conv + FC, BUT:

- ▶ No pooling between some conv layer
- ▶ Convolution stride 1

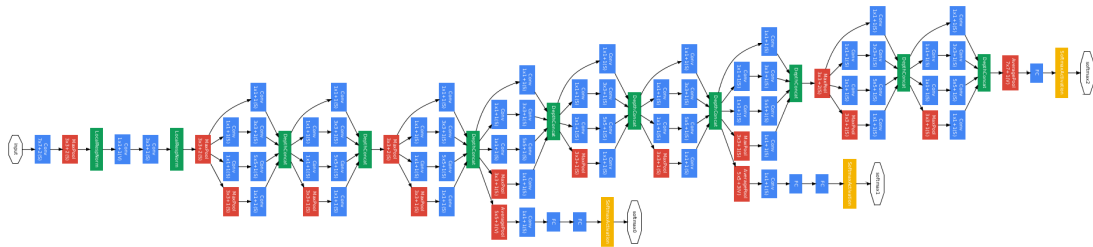
ImageNet'14: VGG



► 3x3 convolutions: two 3x3 conv \sim one 5x5 conv

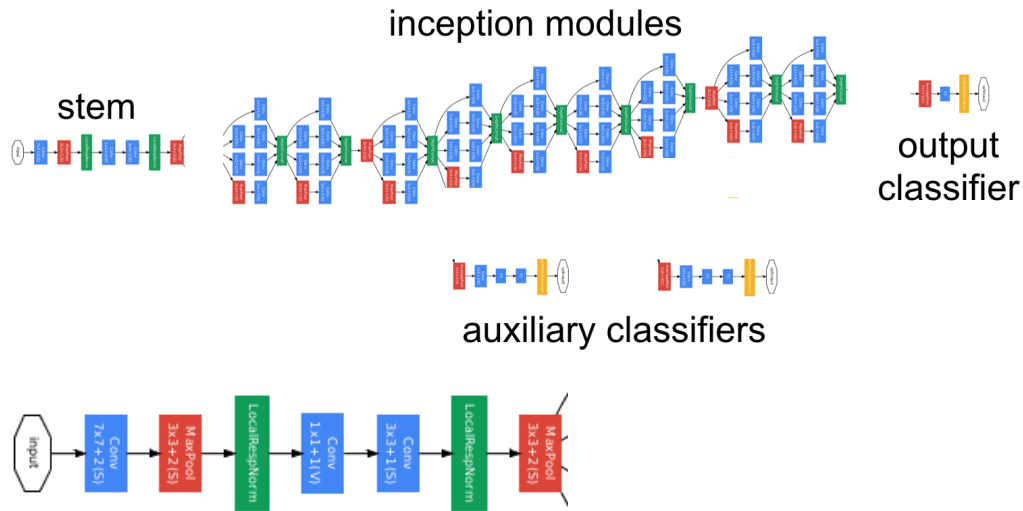


ImageNet'14: GoogLeNet [Szegedy et al., 2015]



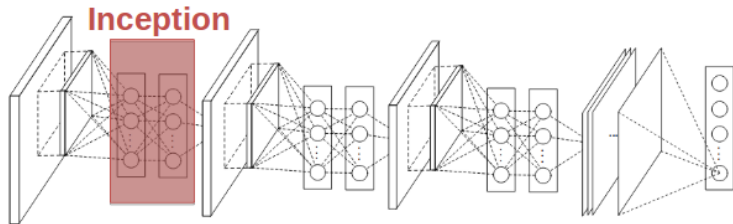
- ▶ GoogLeNet global archi: three main components
 1. 'Stem' ~ [Conv-Pool] Layer
 2. Inception modules: Networks in Networks
 3. Auxiliary classifiers

GoogLeNet: 'Stem' ~ [Conv-Pool] Layers



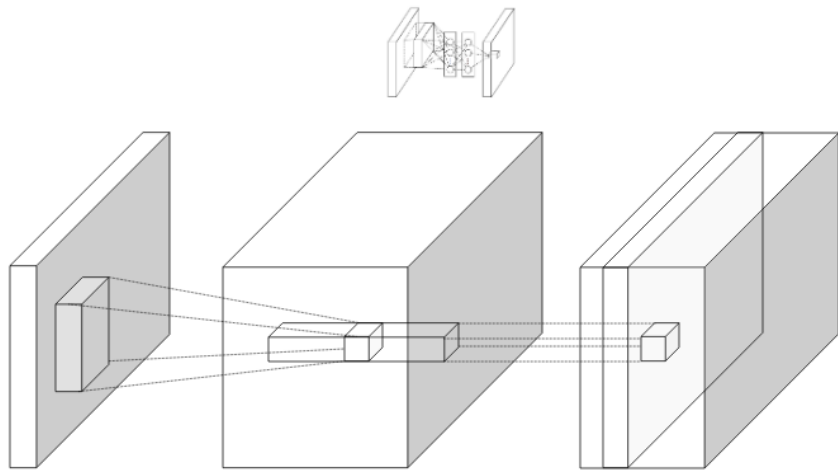
GoogLeNet: Inception Module

- ▶ Inspired from Network in Network (NiN) idea [Lin et al., 2013]



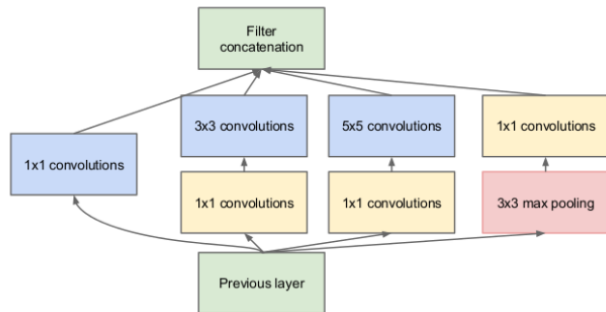
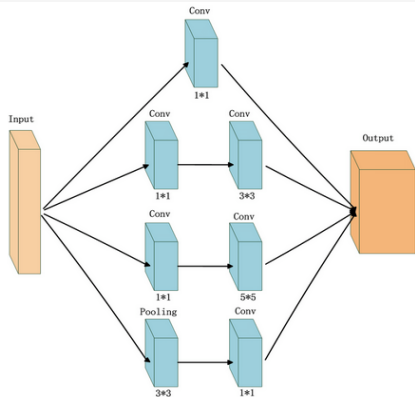
- ▶ Each conv layer: linear + non-linearity
- ▶ NiN block: hierarchy of conv layer
 - ▶ 1×1 conv \sim MLP \Rightarrow universal approximator, more expressive power for each patch

GoogLeNet: 1x1 Convolution



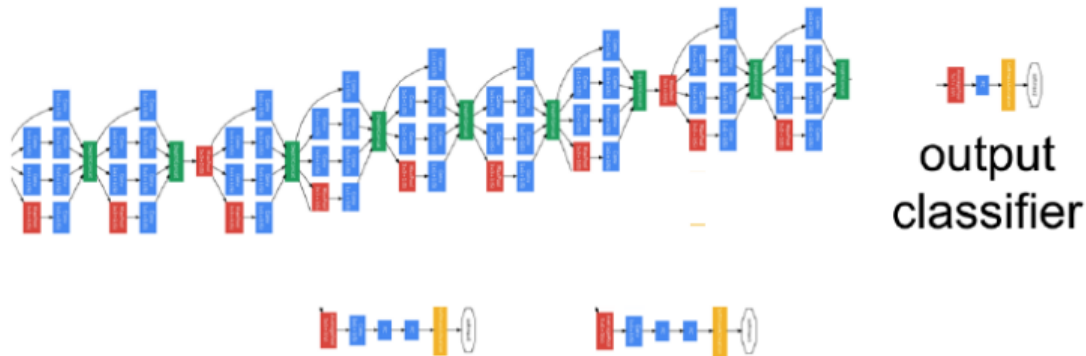
1x1 conv layer

GoogLeNet: Inception Module



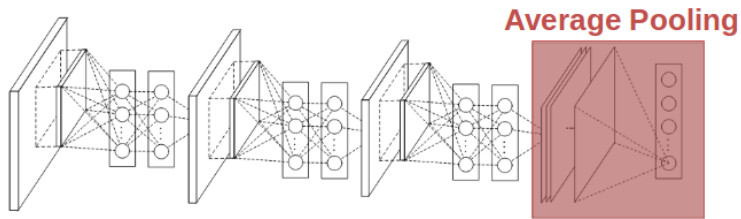
- ▶ 1×1 , 3×3 , $5 \times 5 \Rightarrow$ multi-scale
- ▶ Optimal filter size, pooling or not: learned
- ▶ 1×1 : dimensionality reduction
 \Rightarrow reasonable # parameters

ImageNet'14: GoogLeNet



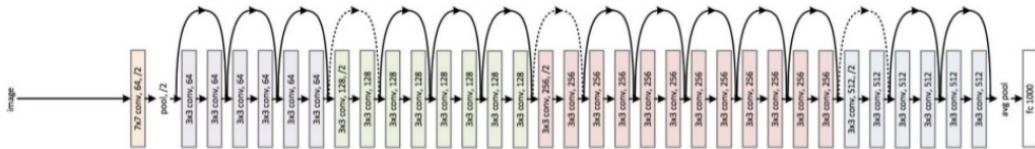
- ▶ **Two auxiliary classifiers**
 - ▶ \times vanishing gradient
 - ▶ Regularization during training

ImageNet'14: GoogLeNet



- ▶ **Output classifier: GLoabal Average Pooling (GAP)**
⇒ no FC!
 - ▶ Drastic ↓ # params!

Deep Learning since 2012: Conclusion



- ▶ Network size inflation
 - ▶ With [Conv-Pool] blocks + FC layers
- ▶ GoogLeNet
 - ▶ Networks in Network (NiN) block
 - ▶ Gradient at several depth levels
 - ▶ No fully connected layer
- ▶ Even deeper models? Residual connections?
⇒ Following!

References I



Lin, M., Chen, Q., and Yan, S. (2013).

Network in network.

CoRR, abs/1312.4400.



Simonyan, K. and Zisserman, A. (2014).

Very deep convolutional networks for large-scale image recognition.

CoRR, abs/1409.1556.



Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D.,

Vanhoucke, V., and Rabinovich, A. (2015).

Going deeper with convolutions.

In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9. IEEE.



Zeiler, M. and Fergus, R. (2014).

Visualizing and understanding convolutional networks, volume 8689 LNCS of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pages 818–833.

Springer Verlag, part 1 edition.