Journée GdR ISIS Bilan IRIM TRECVid / Deep Learning

Deep Learning (2nd edition)

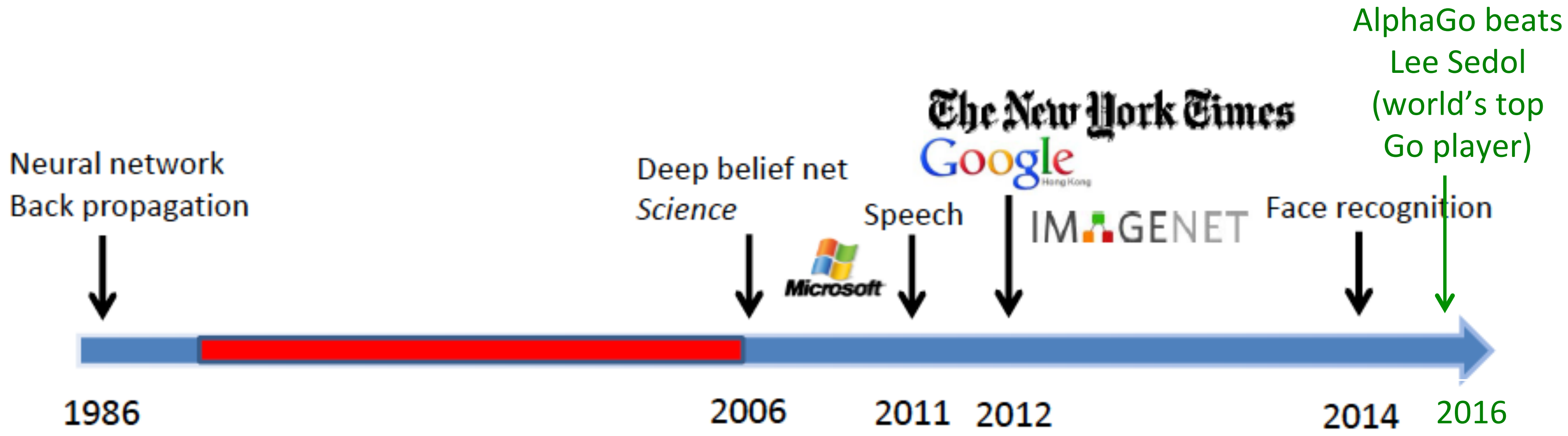
Introduction: Matthieu Cord (LIP6 UPMC)

14th of April 2016

Outline of my 2015 Deep Learning GdR ISIS Introduction talk

1. Key dates in deep learning
2. Deep learning for object recognition
   – Architecture
   – Results
   – Learning
3. Discussion
   – Deep vs Shallow (Why deep?)
   – Feature Learning vs Feature Engineering
   – Using deep in Computer Vision
4. Key issues for Deep&Vision – 2015
   – Talk S. Mallat (ENS)
   – Talk I. Laptev (INRIA)

And updates 2016



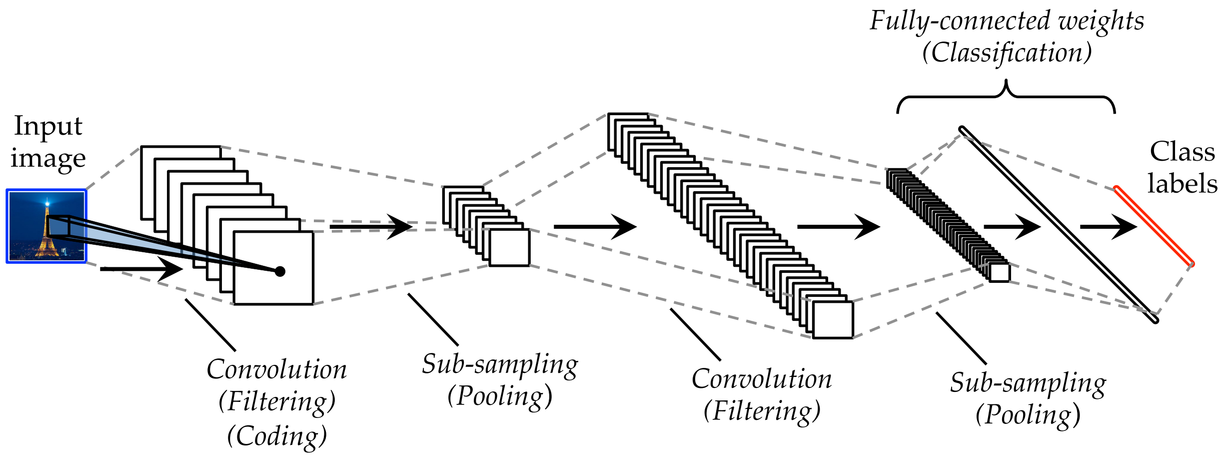
Neural network
Back propagation
1986
Deep belief net
Science
2006
Microsoft
Speech
2011
The New York Times
Google
IMAGENET
2012
Face recognition
2014
AlphaGo beats
Lee Sedol
(world's top
Go player)
2016

Outline of my 2015 Deep Learning GdR ISIS Introduction talk

1. Key dates in deep learning
2. **Deep learning for object recognition**
   - Architecture
   - Results
   - Learning
3. Discussion
   - Deep vs Shallow (Why deep?)
   - Feature Learning vs Feature Engineering
   - Using deep in Computer Vision
4. Key issues for Deep&Vision – 2015
   - Talk S. Mallat (ENS)
   - Talk I. Laptev (INRIA)

And updates 2016

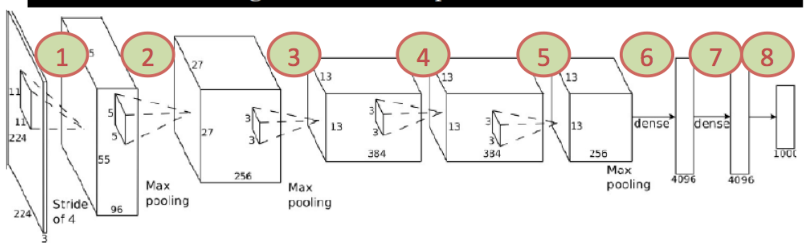
Deep Convolutional Neural Networks (CNN)



[LeCun-89]

Krizhevsky et al. [NIPS2012]

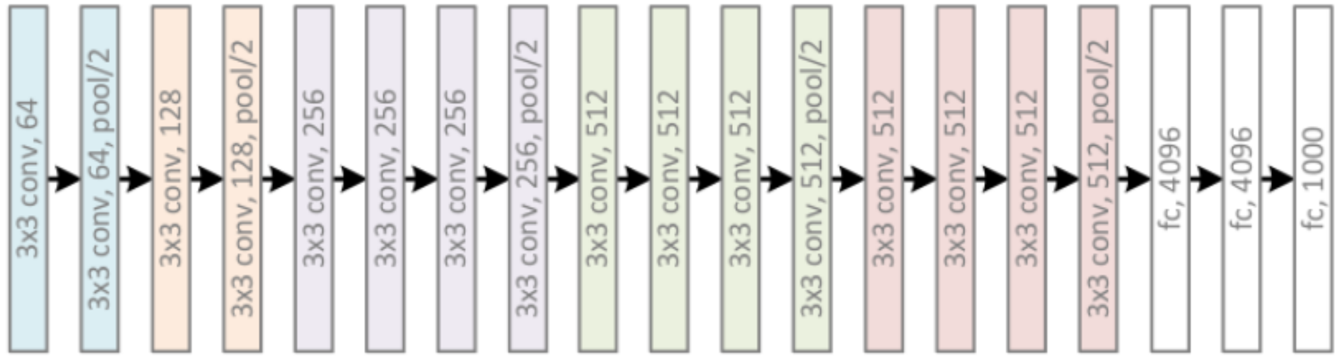
- Same model as LeCun'98 but:
  - Bigger model (8 layers)
  - More data (10^6 vs 10^3 images)
  - GPU implementation (50x speedup over CPU)
  - Better regularization (DropOut)



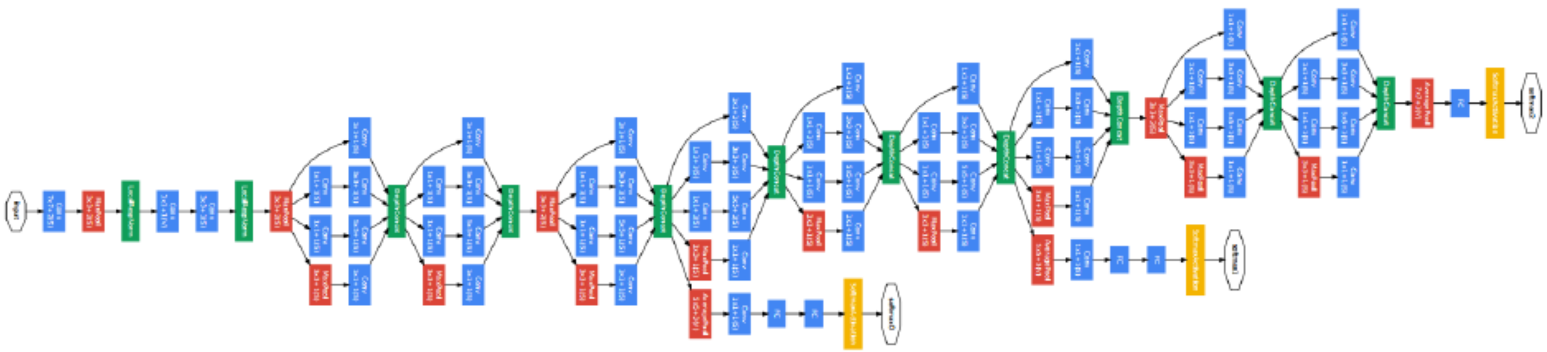
- 7 hidden layers, 650,000 neurons, 60,000,000 parameters
- Trained on 2 GPUs for a week

Very deep CNN 2015 Winners:

VGG, 16/19 layers, 2014



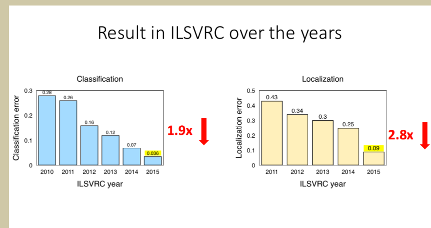
GoogleNet, 22 layers, 2014



ResNet, 152 layers, 2015

Outline of my 2015 Deep Learning GdR ISIS Introduction talk

1. Key dates in deep learning
2. **Deep learning for object recognition**
   - Architecture   from very deep to very very very … deep
   - Results   3.6% top5 error on ImageNet in 2015
   - Learning   Many tricks to boost SGD (like ADAM)



3. Discussion
   - Deep vs Shallow (Why deep?)
   - Feature Learning vs Feature Engineering
   - Using deep in Computer Vision
4. Key issues for Deep&Vision – 2015
   - Talk S. Mallat (ENS)
   - Talk I. Laptev (INRIA)

And updates 2016

Outline of my 2015 Deep Learning GdR ISIS Introduction talk

1. Key dates in deep learning
2. Deep learning for object recognition
   – Architecture
   – Results
   – Learning
3. **Discussion**
   – Deep vs Shallow (Why deep?)
   – Feature Learning vs Feature Engineering
   – Using deep in Computer Vision
4. Key issues for Deep&Vision – 2015
   – Talk S. Mallat (ENS)
   – Talk I. Laptev (INRIA)

And updates 2016

Using CNN representation in Vision

- Are CNN providing generic features ?
  - Yes! Deep features (from ImageNet) +SVM on PASCAL 07 => 10% better than best BoVW methods! [Chatfield]
- Transfer to many tasks [Razavian CVPRw2014]
  - Frozen features + SVM = solution to small datasets
  - Fine tuning not easy in that case (small datasets)
  - Which is the best layer cut for transfer?
    - Depending on the task
- How is it Transferable? [Yosinski NIPS 2014]

=> Many very good results in many vision contexts in 2015

Outline of my 2015 Deep Learning GdR ISIS Introduction talk

1. Key dates in deep learning
2. Deep learning for object recognition
   - Architecture
   - Results
   - Learning
3. Discussion
   - Deep vs Shallow (Why deep?)
   - Feature Learning vs Feature Engineering
   - Using deep in Computer Vision
4. **Key issues for Deep&Vision – 2015**
   - Talk S. Mallat (ENS)
   - Talk I. Laptev (INRIA)

And updates 2016

Key issues for Deep&Vision -

- Supervised/Unsupervised – learning generic data representation

- Weak on theoretical support:
  – Convergence bound, local minimum, …
  – Why it works ???
  $\Rightarrow$ Deep structure analysis/understanding

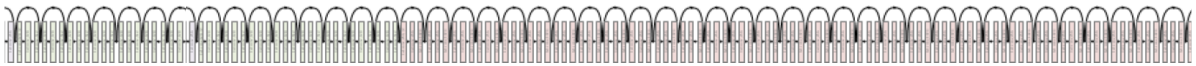
- ImageNet: Object recognition task
  – How to do for large and complex scenes ?
  – Localization: R-CNN [Girshick CVPR2014]

Key issues for Deep&Vision - 2015

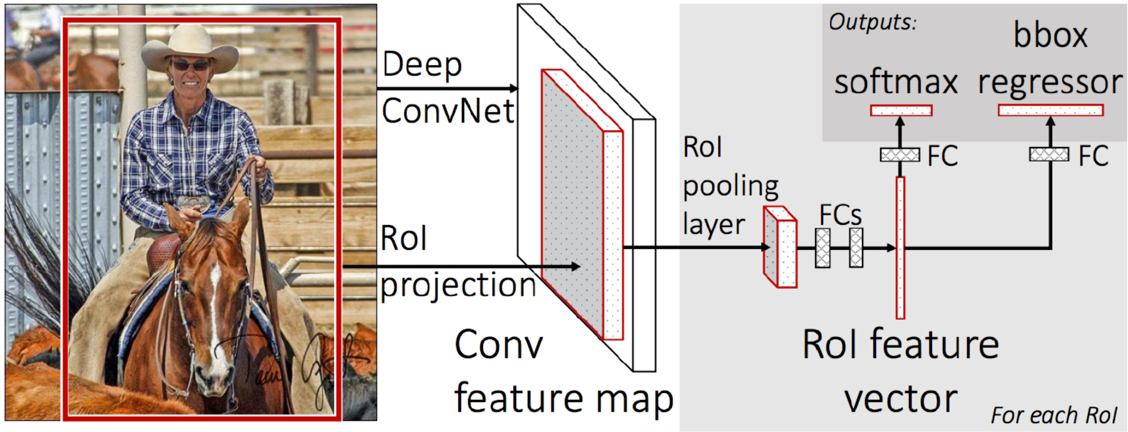
- Supervised/Unsupervised – learning generic data representation

- Weak on theoretical support:
  - Convergence bound, local minimum, …
  - Why it works ???
  - ⇒ Deep structure analysis/understanding
  - ⇒ Talk of S. Mallat --Filter banks / Scattering /contractive archi
- ImageNet: Object recognition task
  - How to do for large and complex scenes ?
  - Localization: R-CNN [Girshick CVPR2014]
  - ⇒ Talk of I. Laptev

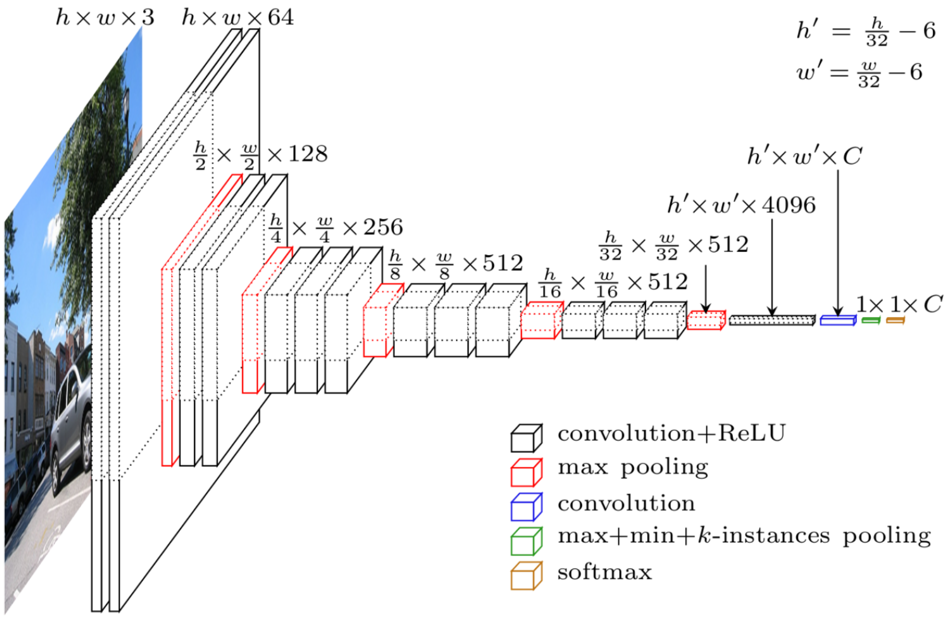
Key issues for Deep&Vision - 2016

- Supervised/Unsupervised(predictive) – learning generic data representation
  - ⇒ L'apprentissage profond non-supervisé : questions ouvertes, par Yann LeCun (Facebook AI Research, NYU, Collège de France)
- Weak on theoretical support:
  - – Convergence => math of deep learning tuto Vidal/Bruna ICCV 2015
  - – Why it works ???
  - ⇒ Deep structure analysis/understanding
  - ⇒ Talk of S. Mallat (Collège de F 2016): “on y comprend à peu près rien”
  - – How many layers ? => 
- ImageNet: Object recognition task
  - – How to do for large and complex scenes ?
  - – Localization: R-CNN [Girshick CVPR2014]
  - ⇒ Fast R-CNN [ICCV 2015], Faster R-CNN [NIPS 2015]

Key issues for Deep&Vision - 2016



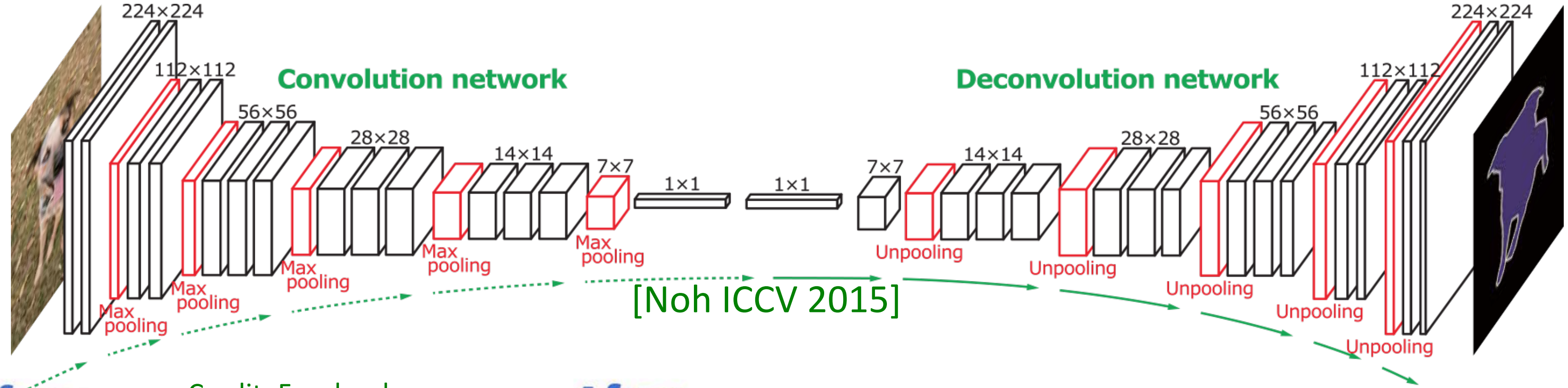
Girshick. *Fast R-CNN*. ICCV 2015



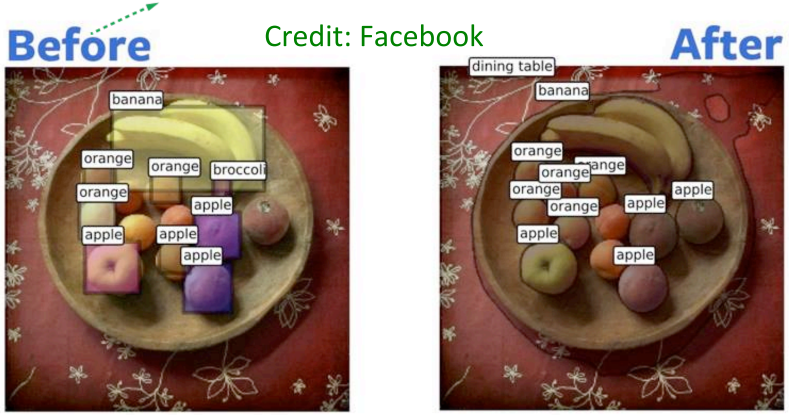
Durand. Weldon, CVPR 2016

Key issues for Deep&Vision - 2016

- Supervised Image Segmentation task



[Noh ICCV 2015]



Credit: Facebook

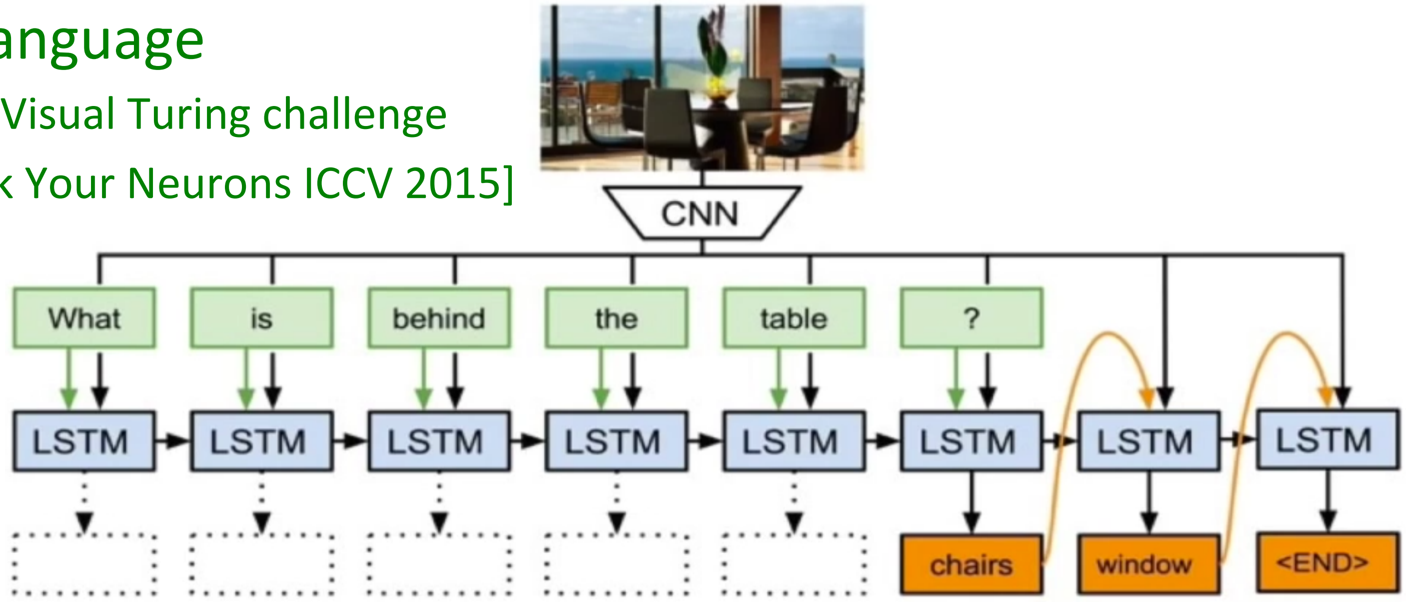
MS COCO Detection Challenge!

- Deep generative models
- Compression/Embedded/Green nets

Key issues for Deep&Vision - 2016

- Vision and Language
  - Visual Q&A, Visual Turing challenge

    [Malinowski Ask Your Neurons ICCV 2015]



  - Visual7W: Grounded Question Answering in Images [Yuke Zhu…Fei-Fei CVPR 16]

- Connection to sequential learning RNN, LSTM, memory nets, …
- Connection to Neurosciences

Posters - session 2016

- Caphee : A data-flow utility for Convolutional Networks implementation on FPGA, K. ABDELOUAHAB, UBP
- Deep Learning for Gender Recognition from Faces and Bodies, G. Antipov, Orange Labs
- Détection de visages sur un système embarqué faible consommation, O. Boisard, M. Paindavoine, LEAD, UB
- P-CNN: Pose-based CNN Features for Action Recognition (ICCV 2015), G. Chéron, I. Laptev and C. Schmid, INRIA
- LR-CNN for fine-grained classification with varying resolution, (ICIP 2015) M. Chevalier et al., LIP6 UPMC, Thales
- WELDON: Weakly Supervised Learning of Deep Convolutional Neural Networks (CVPR 2016), T. Durand, N. Thome, M. Cord, LIP6 UPMC
- Deep learning methods for image Super-Resolution, C. Peyrard, Orange Labs
- Context-aware CNNs for person head detection" (ICCV 2015), T.-H. Vu, A. Osokin and I. Laptev, INRIA
- bQBDC : batch Query By Dropout Committee for Deep Active learning, M. Ducoffe, I3S, UNS
- Deep CNN with multiple feature layers for saliency prediction in video, S Chaabouni, J. Benois-Pineau, LABRI
- Learning the structure of deep architectures using L1 regularization, P. Kulkarni, J. Zepeda, Technicolor
- Cost-Sensitive Adaptive Feature Acquisition with Representation Learning, G. Contardo, L. Denoyer, T. Artières, LIP6 UPMC, LIF UAM

LIP6 Team Ref. on deep learning and Visual representation:

Deep learning for Visual Recognition

- WELDON: Weakly Supervised Learning of Deep Convolutional Neural Networks, T. Durand, N. Thome, M. Cord, CVPR 2016
- MANTRA: Minimum Maximum Latent Structural SVM for Image Classification and Ranking, T Durand, N Thome, M Cord, ICCV 2015
- LR-CNN for fine-grained classification with varying resolution, M Chevalier, N Thome, M Cord, J Fournier, G Henaff, E Dusch, ICIP 2015
- Top-Down Regularization of Deep Belief Networks, H. Goh, N. Thome, M. Cord, JH. Lim, NIPS 2013
- Sequentially generated instance-dependent image representations for classification, G Dulac-Arnold, L Denoyer, N Thome, M Cord, P Gallinari, ICLR 2014
- Learning Deep Hierarchical Visual Feature Coding, H. Goh+, IEEE Transactions on Neural Networks and Learning Systems 2014
- Unsupervised and supervised visual codes with Restricted Boltzmann Machines, H. Goh+, ECCV 2012
- Biasing Restricted Boltzmann Machines to Manipulate Latent Selectivity and Sparsity, H. Goh+, NIPS workshop 2010

Bio-inspired Representation

- Cortical Networks of Visual RecognitionC Thériault, N Thome, M Cord, Biologically Inspired Computer Vision: Fundamentals and Applications, book chapter
- Extended coding and pooling in the HMAX model, C. Thériault, N. Thome, M. Cord, IEEE Trans. on Image Processing 2013

Visual representation

- Pooling in Image Representation: the Visual Codeword Point of View, S. Avila, N. Thome, M. Cord, E. Valle, A. araujo, CVIU 2013
- Dynamic Scene Classification: Learning Motion Descriptors with Slow Features Analysis, C. Thériault, N. Thome, M. Cord, CVPR 2013

Matthieu Cord
LIP6, Computer Science Department
UPMC Paris 6 - Sorbonne University
Paris FRANCE
http://webia.lip6.fr/~cord