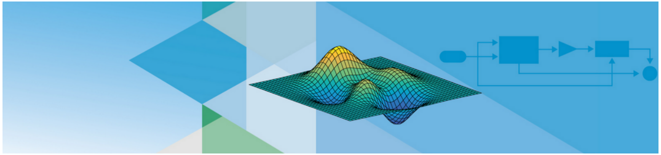
Deep Learning for Visual Recognition

MATLAB EXPO 2016 France



MathWorks

Nicolas Thome
Université Pierre et Marie Curie (UPMC)
Laboratoire d'Informatique de Paris 6 (LIP6)

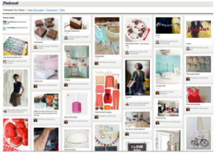
UPMC SORBONNE UNIVERSITÉS

LIP6 Dpt : Données et Apprentissage Artificiel

Context

Big Data: Images & Videos everywhere

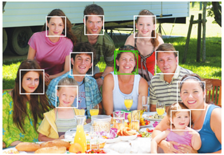
BBC: 2.4M videos

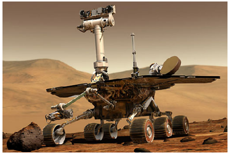


Facebook: 140B images

100M monitoring cameras

- Obvious need to access, organize, search, or classify these data: **Visual Recognition**
- Huge number of applications: mobile visual search, robotics, autonomous driving, augmented reality, medical imaging *etc*
- Leading track in major CV conferences during the last decade

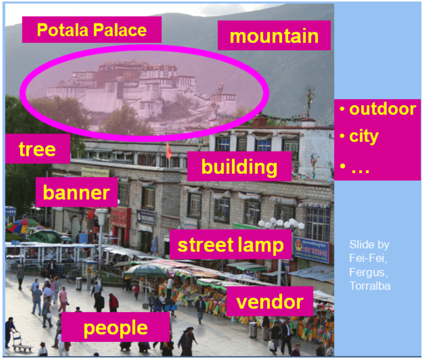




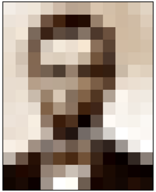
Outline

Visual Recognition: Perceiving Visual World

- Scene categorization
- Object localization
- Context & Attribute recognition
- Rough 3D layout, depth ordering
- Rich description of scene, language, *e.g.* sentences



Challenge: filling the semantic gap



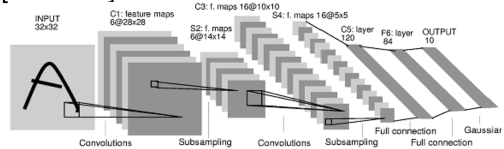
What we perceive *vs*
What a computer sees

243	239	240	225	206	185	188	218	211	206	216	225
242	239	218	110	67	31	34	152	213	206	208	221
243	242	123	58	94	82	132	77	108	208	208	215
235	217	115	212	243	236	247	139	91	209	208	211
233	208	131	222	219	226	196	114	74	208	213	214
232	217	131	116	77	150	69	56	52	201	228	223
232	232	182	186	184	179	159	123	93	232	235	235
232	236	201	154	216	133	129	81	175	252	241	240
235	238	230	128	172	138	65	63	234	249	241	245
237	236	247	143	59	78	10	94	255	248	247	251
234	237	245	193	55	33	115	144	213	255	253	251
248	245	161	128	149	109	138	65	47	156	239	255
190	107	39	102	94	73	114	58	17	7	51	137
23	32	33	148	168	203	179	43	27	17	12	8
17	26	12	160	255	255	109	22	26	19	35	24

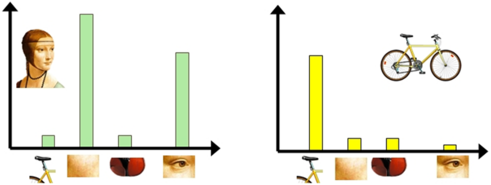
- Illumination variations
- View-point variations
- Deformable objects
- intra-class variance
- *etc*

Outline

- 80's: training Convolutionnal Neural Networks (CNN) with back-propagation ⇒ postal code reading [LBD+89]

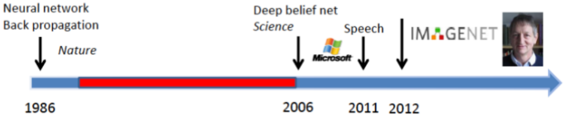


- 90's: golden age of kernel methods, NN = black box
- 2000's: BoW + SVM : state-of-the-art CV



Visual Recognition History: Trends and methods in the last four decades

- Deep learning revival: unsupervised learning (DBN) [HOT06]



- 2012: CNN outstanding success in ImageNet [KSH12]

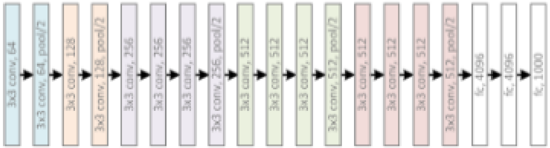
Rank	Name	Error rate	Description
1	**U. Toronto**	0.15315	Deep learning
2	U. Tokyo	0.26172	Hand-crafted features and learning models. Bottleneck.
3	U. Oxford	0.26979	
4	Xerox/INRIA	0.27058	

  - Huge number of labeled images (10^6 images)
  - GPU implementation for training

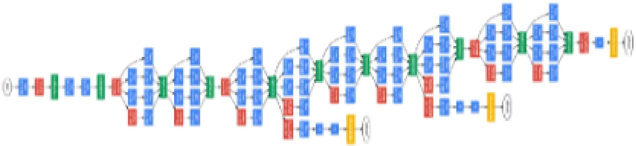
Deep Learning since 2012

More & more data (Facebook 10^9 images / day), larger & larger networks

VGG, 16/19 layers, 2014



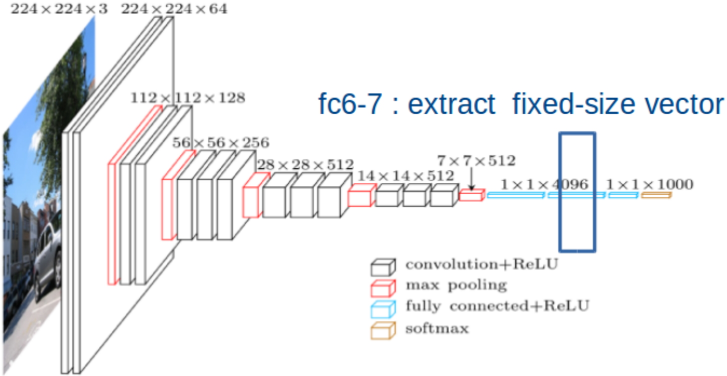
GoogleNet, 22 layers, 2014



ResNet, 152 layers, 2015

Deep Learning since 2012

Transferring Representations learned from ImageNet

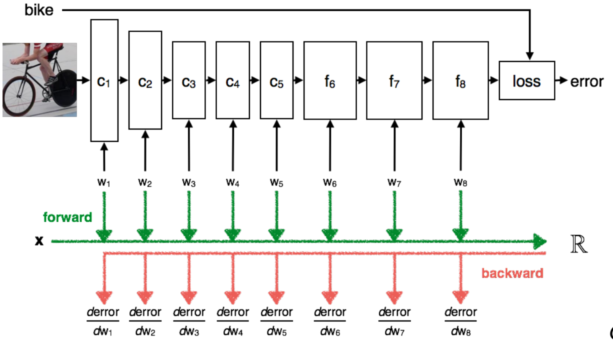


- Extract layer ⇒ fixed-size vector: "Deep Features" (DF)
- Now state-of-the-art for any visual recognition task

MATLAB

MatConvNet: MatLab toolbox for CNN processing

- Developed by Oxford team (Vedaldi, Lenc), `http://www.vlfeat.org/matconvnet/`
- Using it for processing & training (chain) feedforward CNNs
  - Efficient CNN implementation far from trivial



Credits: Vedaldi, Zisserman

Resource for the community: MatConvNet

Forward run of a network

- Wide range of available pre-trained networks: VGG, Googlenet, ResNet
- Fast execution : easy-to-use GPU implementation
- Input: image, output: one ImageNet class

```
 run matlab/vl_setupnn
% Load the (online available) CNN
net = load('imagenet-vgg-m.mat');

 % Load and normalize image
im = single(imread('peppers.png'));
im = imresize(im, net.meta.normalization.imageSize(1:2));
im = im-net.meta.normalization.averageImage;
 % Run the CNN
res = vl_simplenn(net, im);

 % Scores for the 1,000 ImageNet classes
scores = squeeze(gather(res(end).x)) ;
[bestScore , bestClass] = max(scores) ;
```

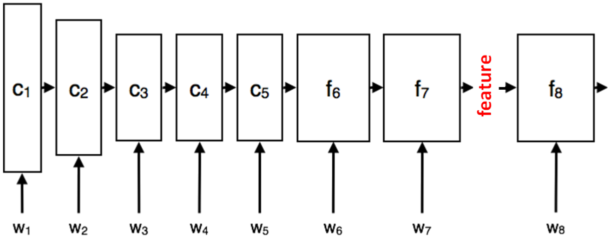
bell pepper (ImageNet class #735), score 0.924

Resource for the community: MatConvNet

- Transfer: CNN as a feature extractor

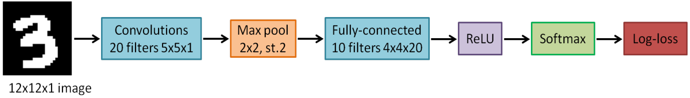
```
 % Load the (online available) CNN
% Load and normalize image, Run the CNN
 res = vl_simplenn(net, im);

 % Extract features
features = squeeze(gather(res(20).x)) ;
 % Learn / test an SVM on these features
```



- Design your own network: architecture

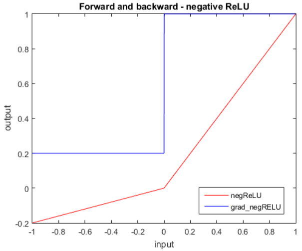
```
 % Convolution
net.layers{1} = struct('type', 'conv',
'weights', {0.01*randn(5,5,1,20,'single'),
zeros(1,20,'single')}, 'stride',1,'pad',0);
```



Resource for the community: MatConvNet

- Design your own block: custom layer functions
  - Custom layer: one Matlab file with forward/backward functions

```
function out = vl_negReLU(x,dzdy,opts)
 if nargin <= 1 || isempty(dzdy)
  out = x.*(x>0) + 0.2*x.*(x<0);
 else
  out = dzdy .* ((x>0) + 0.2.*(x<0));
 end
```



- Training a CNN model

Efficient implementation, Optimized for GPU
Use GPU = boolean option

opts.gpus = 1;
stats = cnn_train(net, imdb, @get_batch_function, opts);

model	batch sz.	CPU	GPU	CuDNN
AlexNet	256	22.1	192.4	264.1
VGG-F	256	21.4	211.4	289.7
VGG-M	128	7.8	116.5	136.6
VGG-S	128	7.4	96.2	110.1
VGG-VD-16	24	1.7	18.4	20.0
VGG-VD-19	24	1.5	15.7	16.5

Table 1.1: ImageNet training speed (images/s).

Resource for the community: MatConvNet

MatConvNet: a use case [CTC+15]

- Context: fine-grained recognition on low-resolution images
  - Varying image size
  - 6667 training images
- Evaluated frameworks:
  - Pre-trained deep features + SVM
  - Custom network learned from scratch on small images

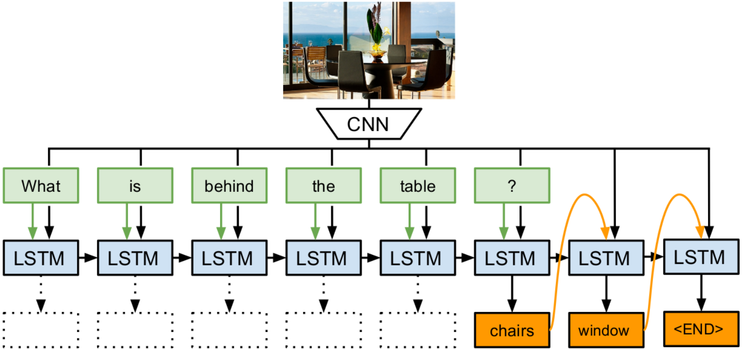
Method	Accuracy
CNNM (1^{st} fc)	32.7%
CNNM (2^{nd} fc)	27.2%
Our LRCNN	44.8%

Outline

Deep Learning since 2012

Breakthroughs with CNNs

- Deep learning, DF: very powerful intermediate representations
  - Semantic relationship wrt various categories, *e.g.* 10^3 ImageNet
  - Open the way to unreachable applications: image captioning, visual question answering, image generation, *etc*



Breakthroughs with CNNs

Modern data & annotations

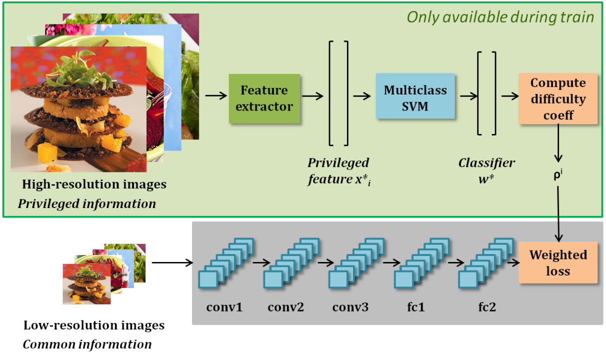
- Privileged information (PI) = additional example-specific information only available during training
- Goal: benefit from this additional data to improve the classifier

Breakthroughs with CNNs

Privileged information (PI)

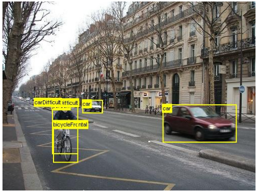
- SVM+ [VV09] / Margin Transfer [SQL14]: (PI) $\Leftrightarrow$ difficulty level
- Curriculum learning [BLCW09]: start easy / increase difficulty

$\Rightarrow$ Our deep+: end-to-end training of a deep CNN with (PI)

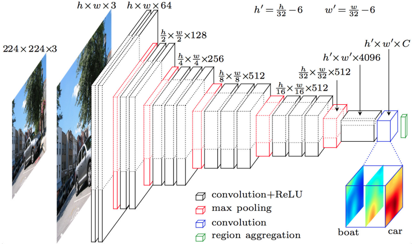


Open Issues in Deep Learning for Visual Recognition

- Deep CNNs: breakthrough, large scale data and Transfer ⇒ solved problem ?
- Limited invariance (conv layers): OK for centered objects, KO for "natural" photos



- Weakly Supervised Learning of deep CNNs [DTC16, DTC15], region localization



- Architecture, compression, learning formulation (unsupervised training)
- Formal understanding: model [BM13], optimization [HV15, DPG$^+$14], over-fitting

Thank you for your attention !

- Sorbonne Universités - LIP6, MLIA Team (P. Gallinari)
- Machine learning for vision: M. Cord, N. Thome, PhD Students:
  - M. Chevalier: Learning Using Privileged Information (LUPI)
  - T. Durand: Structured prediction and Weakly Supervised Learning
  - X. Wang: Visual Recognition with Eye-Tracker
  - M. Blot: Deep Architectures for Large-Scale Recognition
  - M. Carvalho: Deep Networks Compression

T. Durand, N. Thome, and M. Cord. Weakly Supervised Learning of Deep CNNs, CVPR 2016.
M. Chevalier *et. al*. LR-CNN For Fine-grained Classification with Varying Resolution, ICIP 2015.
T. Durand, N. Thome, and M. Cord. Minimum Maximum Latent Structural SVM for Image Classification and Ranking, ICCV 2015.
X. Wang, N. Thome and M. Cord. Gaze Latent Support Vector Machine for Image Classification, ICIP 2016.
M. Blot, N. Thome and M. Cord. MaxMin convolutional neural networks for image classification, ICIP 2016.
M. Carvalho, M. Cord, N. Thome, S. Avila and E. Valle. Deep Neural Networks Under Stress, ICIP 2016.

References I

- Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston, *Curriculum learning*, International Conference on Machine Learning (ICML) (2009).
- Joan Bruna and Stephane Mallat, *Invariant scattering convolution networks*, IEEE Trans. Pattern Anal. Mach. Intell. **35** (2013), no. 8, 1872–1886.
- Marion Chevalier, Nicolas Thome, Matthieu Cord, Jérôme Fournier, Gills Henaff, and Elodie Dusch, *Lr-cnn for fine-grained classification with varying resolution*, IEEE International Conference on Image Processing (ICIP) (2015).
- Yann Dauphin, Razvan Pascanu, Çaglar Gülçehre, Kyunghyun Cho, Surya Ganguli, and Yoshua Bengio, *Identifying and attacking the saddle point problem in high-dimensional non-convex optimization*, CoRR **abs/1406.2572** (2014).
- Thibaut Durand, Nicolas Thome, and Matthieu Cord, *MANTRA: Minimum Maximum Latent Structural SVM for Image Classification and Ranking*, International Conference on Computer Vision (ICCV), 2015.
- Thibaut Durand, Nicolas Thome, and Matthieu Cord, *WELDON: Weakly Supervised Learning of Deep Convolutional Neural Networks*, Computer Vision and Pattern Recognition (CVPR), 2016.
- Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh, *A fast learning algorithm for deep belief nets*, Neural Comput. **18** (2006), no. 7, 1527–1554.
- Benjamin D. Haeffele and René Vidal, *Global optimality in tensor factorization, deep learning, and beyond*, CoRR **abs/1506.07540** (2015).

References II

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, *Imagenet classification with deep convolutional neural networks*, Advances in neural information processing systems, 2012, pp. 1097–1105.

Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel, *Backpropagation applied to handwritten zip code recognition*, Neural computation 1 (1989), no. 4, 541–551.

Viktoriia Sharmanska, Novi Quadrianto, and Christoph H. Lampert, *Learning to transfer privileged information*.

Vladimir Vapnik and Akshay Vashist, *A new learning paradigm: Learning using privileged information*, Neural Networks (2009).