Survol des activités au CRIM en apprentissage profond

Présenté à Journée de la géomatique 2018

Par Samuel Foucher, Ph. D.
Chercheur senior et directeur de l'équipe Vision et imagerie

Contributeurs :
Mohamed Dahmane
Mario Beaulieu
Pierre-Luc St-Charles
Justine Boulent
Gilles Boulianne

Principal partenaire financier
Économie, Science et Innovation Québec

Plan de la présentation

- Brève présentation du CRIM
- Très brève introduction à l'apprentissage profond
- Exemple de réalisations
- Traitement de la parole
- Quelques conclusions

LE CRIM EN CHIFFRES (2016-2017)

OBNL PRIVÉ NEUTRE

FONDÉ EN 1985

BUDGET ANNUEL DE **8,2 M$**

CERTIFIÉ ISO **9001:2008**

- Centre de recherche appliquée
- Seul centre de cette nature dans l'Est du Canada, reconnu mondialement
- Expert en recherche appliquée **+** innovation collaborative **+** transfert en entreprises
- Nombreuses collaborations avec centres de recherche et universités
- Embauche et formation de maîtres, et doctorants universitaires

Annuellement…

50 COLLABORATEURS UNIVERSITAIRES

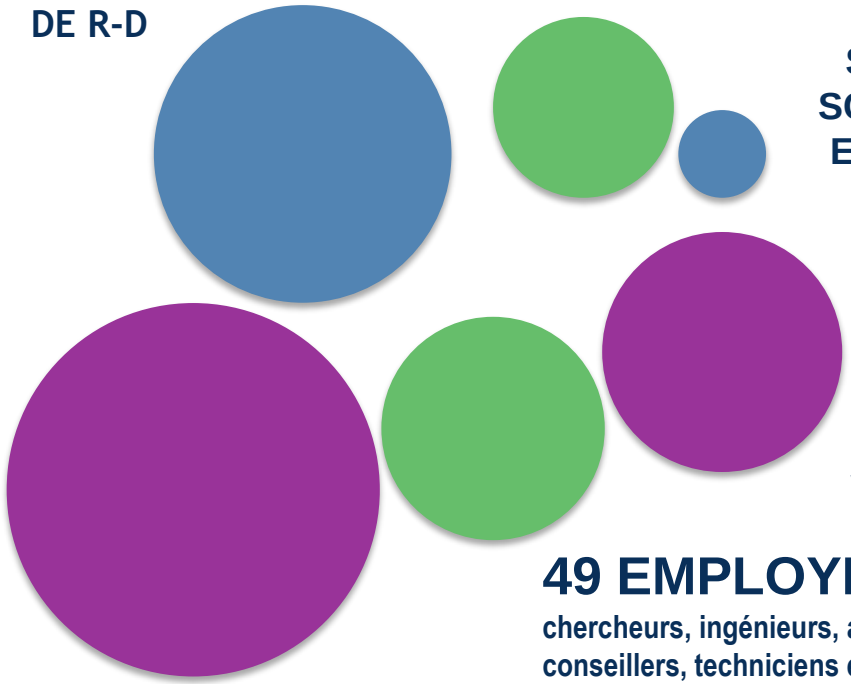
Provenant d'universités de partout dans le monde

95 PROJETS

DE R-D

11

SÉMINAIRES SCIENTIFIQUES ET JOURNÉES TECHNO



41

PUBLICATIONS SCIENTIFIQUES

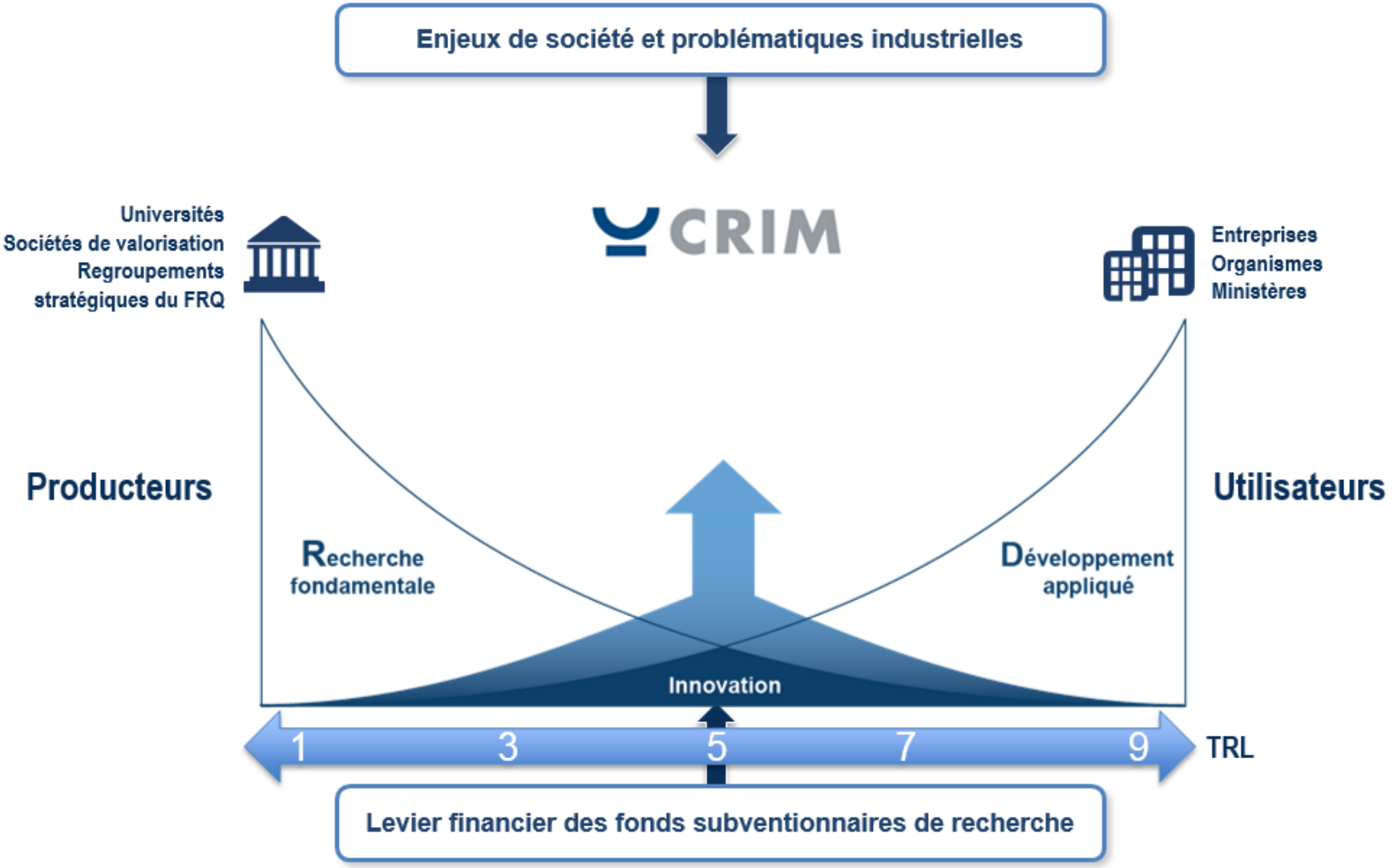
49 EMPLOYÉS

chercheurs, ingénieurs, agents de recherche, conseillers, techniciens en informatique, professionnels en accompagnement

166 CLIENTS

DESSERVIS

POSITIONNEMENT



QUATRE ÉQUIPES DE RECHERCHE

MODÉLISATION ET DÉVELOPPEMENT LOGICIEL AVANCÉ

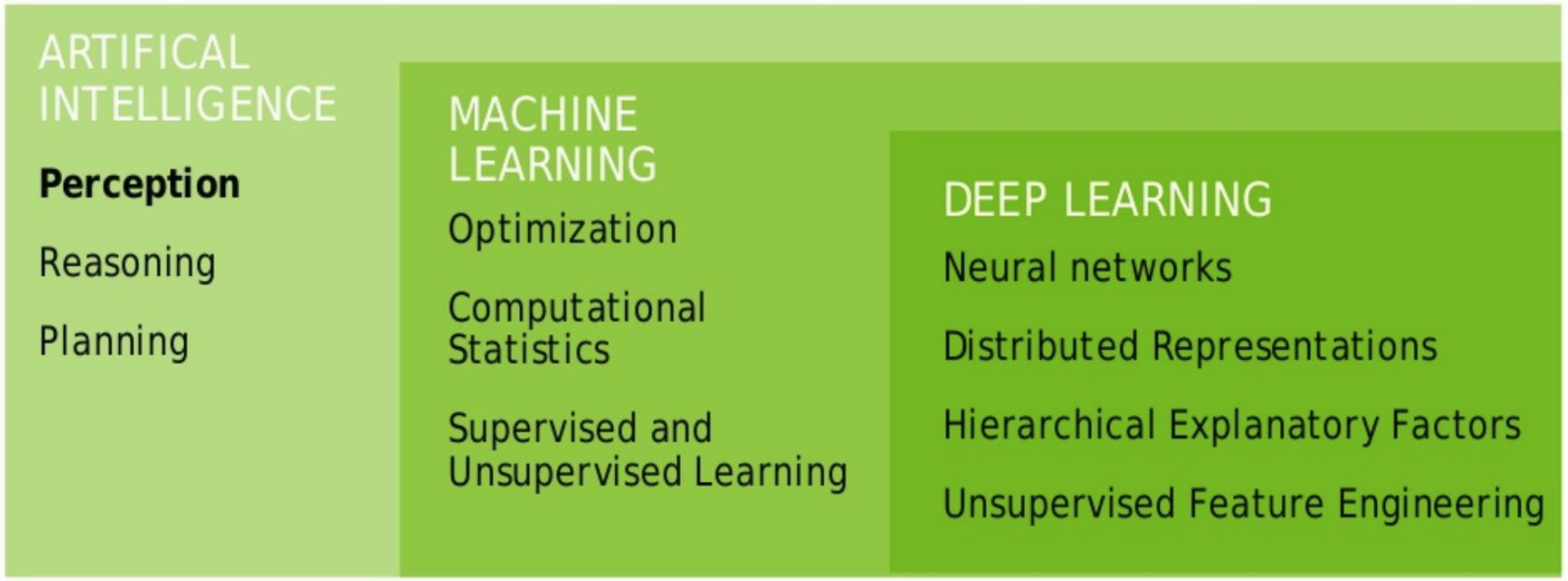
TECHNOLOGIES ÉMERGENTES ET SCIENCE DES DONNÉES

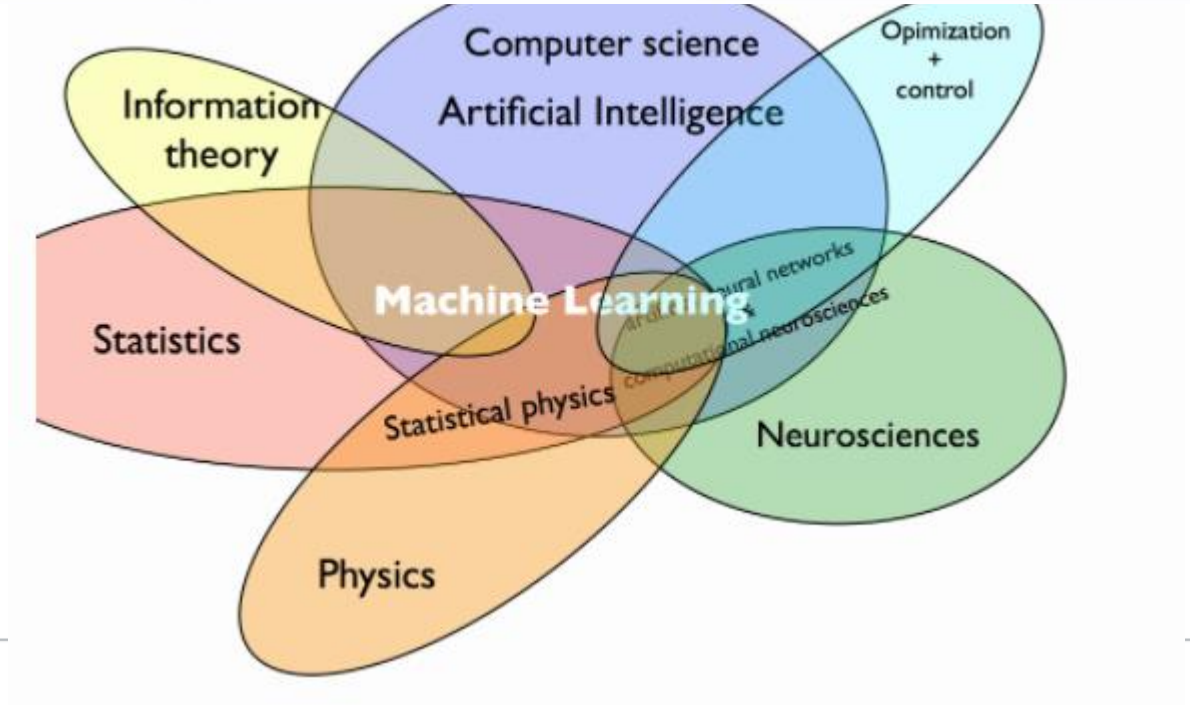
PAROLE ET TEXTE

VISION ET IMAGERIE

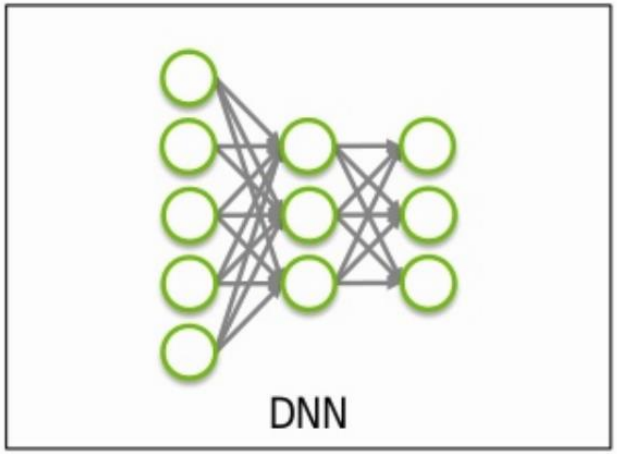
Très brève introduction à l'apprentissage profond

Machine Learning vs. AI vs. Deep Learning

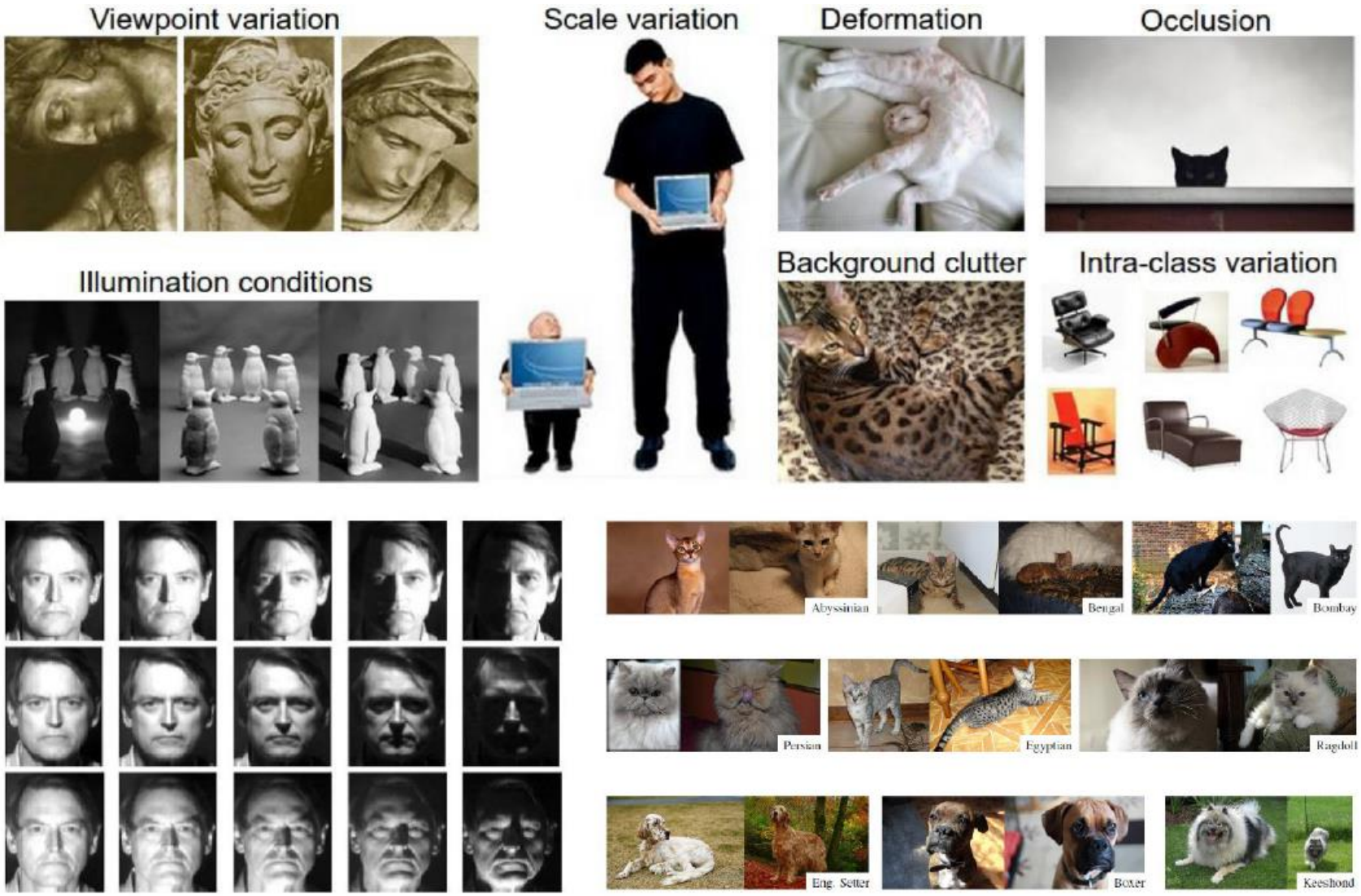




Big Learning



La reconnaissance/détection d'objets est un problème difficile en vision



Approches traditionnelles en vision par ordinateur

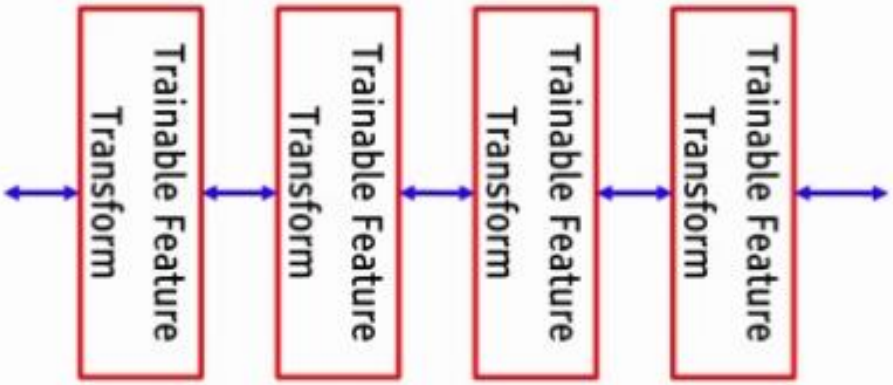
Traditional Pattern Recognition: Fixed/Handcrafted Feature Extractor

Mainstream Modern Pattern Recognition: Unsupervised mid-level features

Source: Y. LeCun

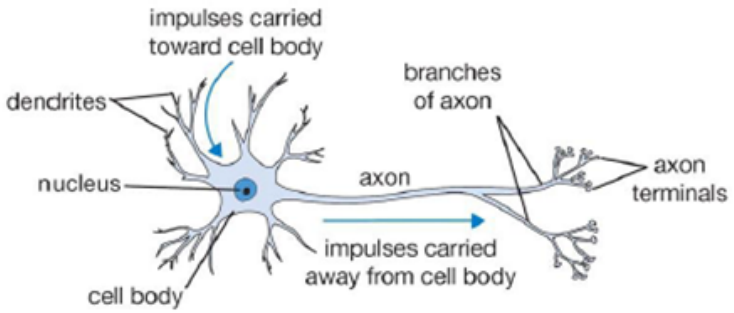
Représentation hiérarchique de l'information

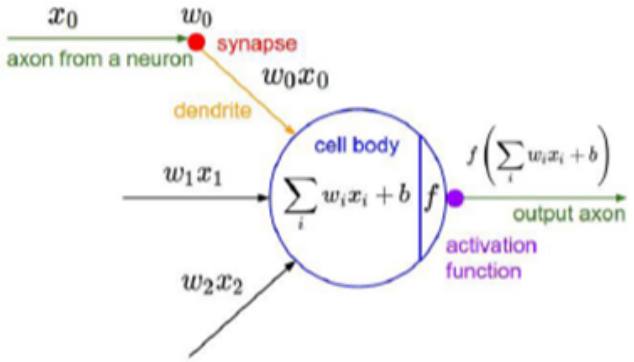
- Hierarchy of representations with increasing level of abstraction
- Each stage is a kind of trainable feature transform
- Image recognition
  - Pixel → edge → texton → motif → part → object
- Text
  - Character → word → word group → clause → sentence → story
- Speech
  - Sample → spectral band → sound → ... → phone → phoneme → word



Source: Y. LeCun

Neurones Artificiels

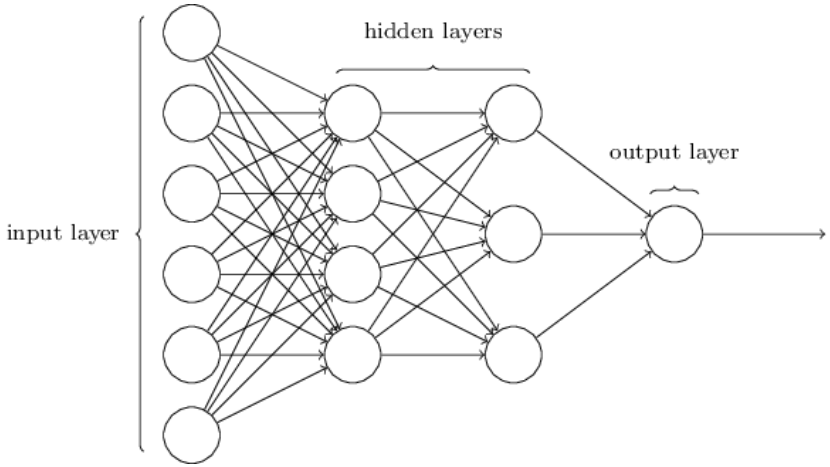


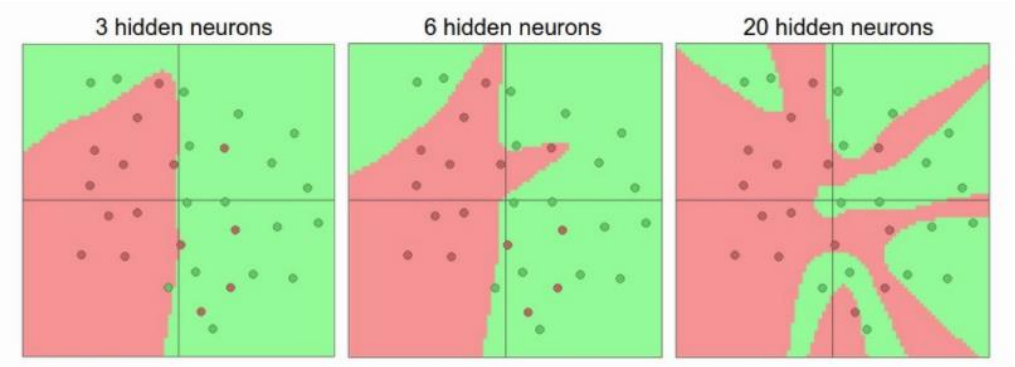


Lettvin, J.Y., Maturana, H.R., McCulloch, W.S., & Pitts, W.H. ; What the Frog's Eye Tells the Frog's Brain, (PDF, 14 pages) (1959) ; Proceedings of the IRE, Vol. 47, No. 11, pp. 1940-51.

Neurones Artificiels

Multi-Layer Perceptron (MLP)

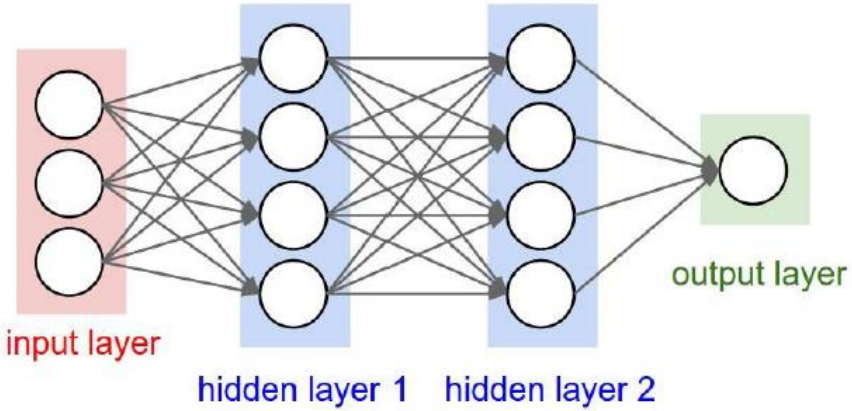




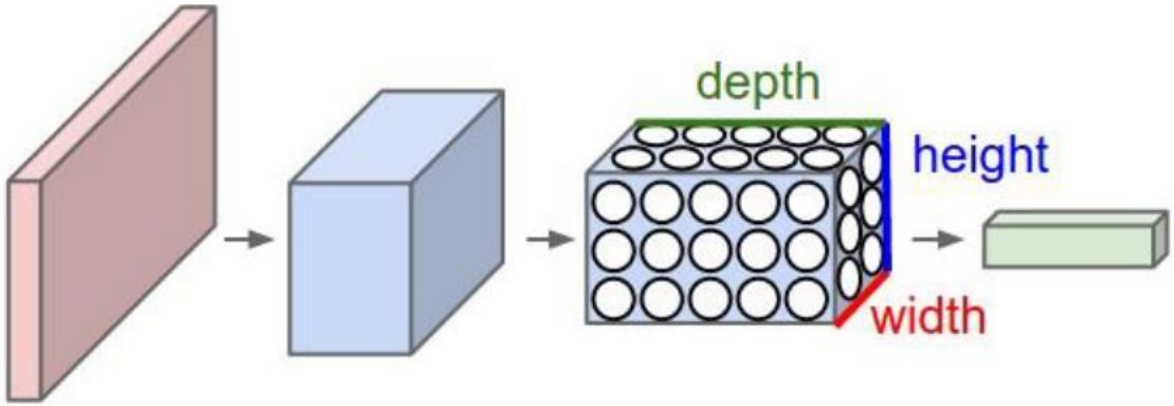
Larger Neural Networks can represent more complicated functions

Réseaux de neurones et convolutions

Réseaux de neurones régulier (complètement connecté)

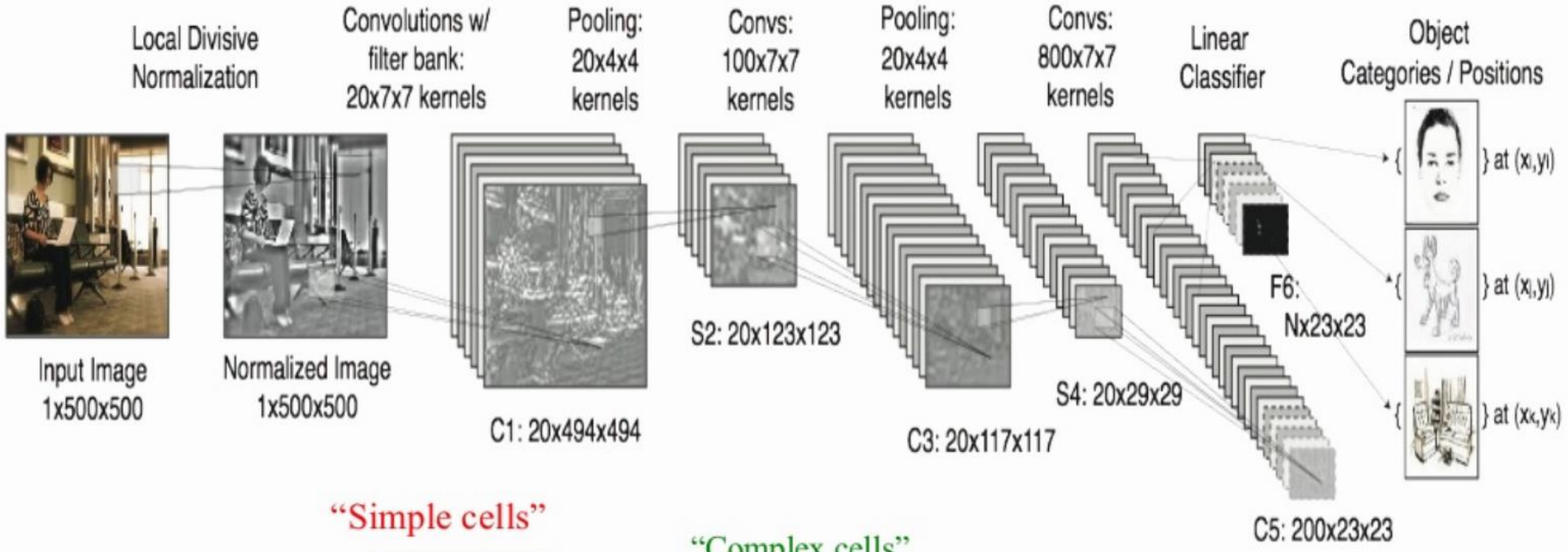


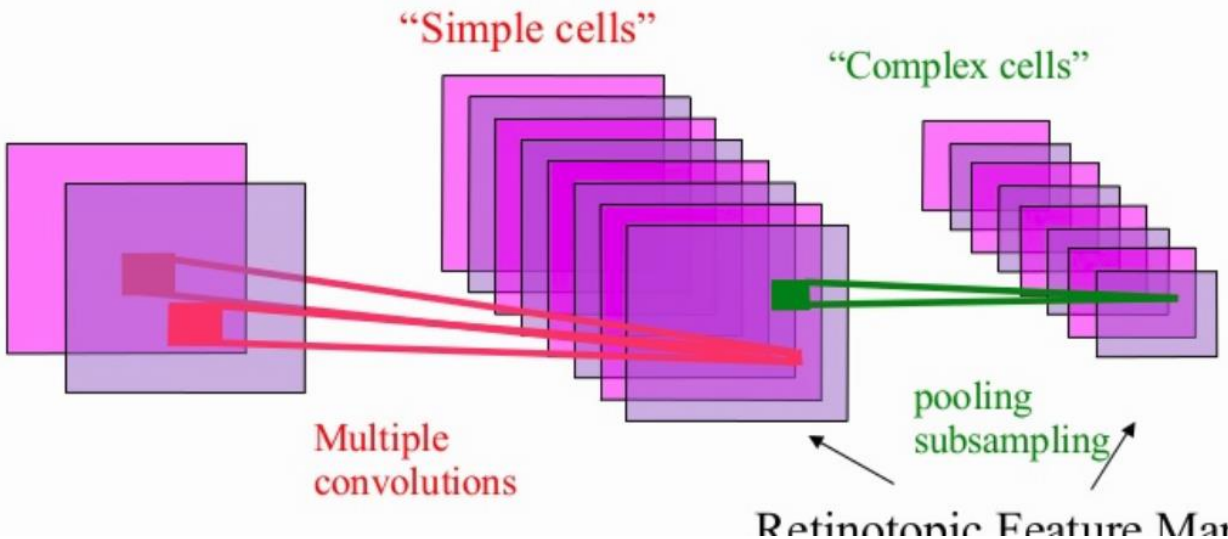
Réseaux de neurones à convolution (CNN) ou ConvNets



Each layer takes a 3d volume, produces 3d volume with some smooth function that may or may not have parameters.

Réseaux de neurones et convolutions





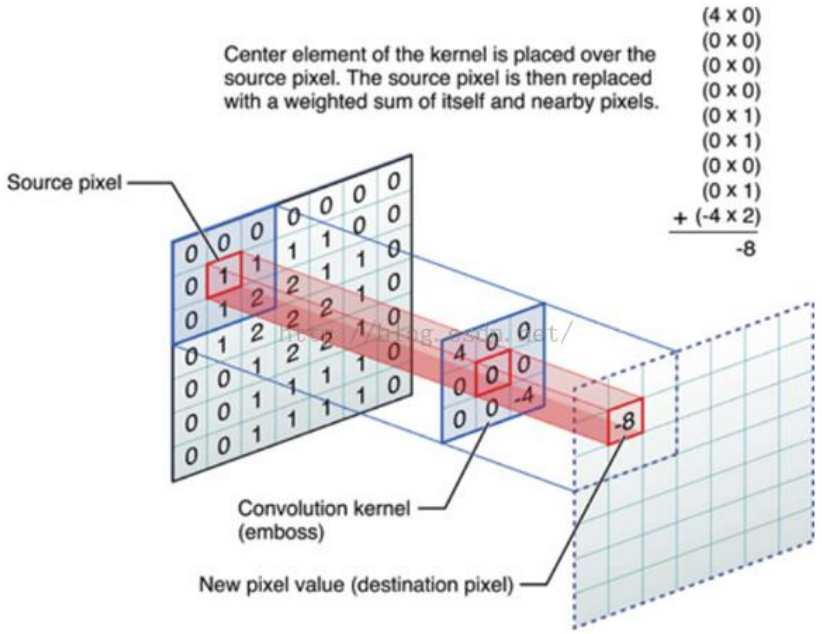
- Training is supervised
- With stochastic gradient descent

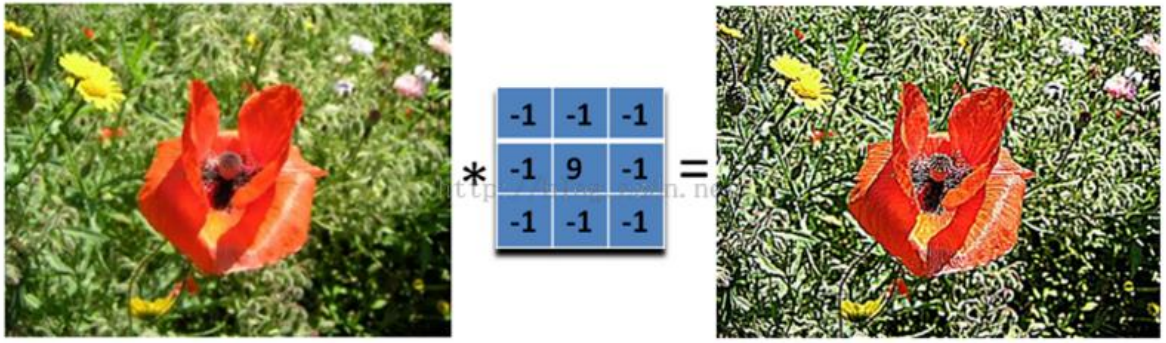
[LeCun et al. 89]

[LeCun et al. 98]

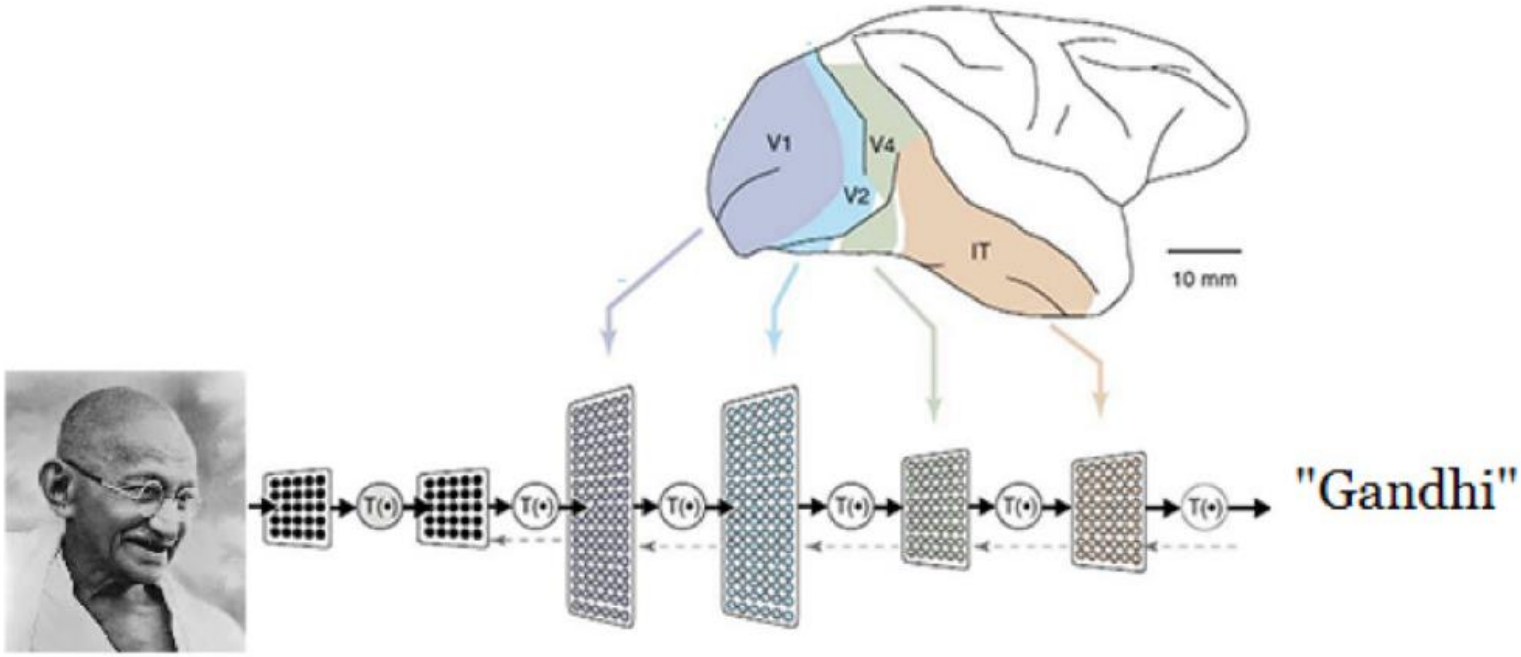
Principe de la convolution

- Permet de capturer la relation spatiale entre pixels





Cortex visuel humain



From: Large-Scale Deep Learning for Intelligent Computer Systems, Jeff Dean, WSDM 2016, adapted from Untangling invariant object recognition, J DiCarlo et D Cox, 2007

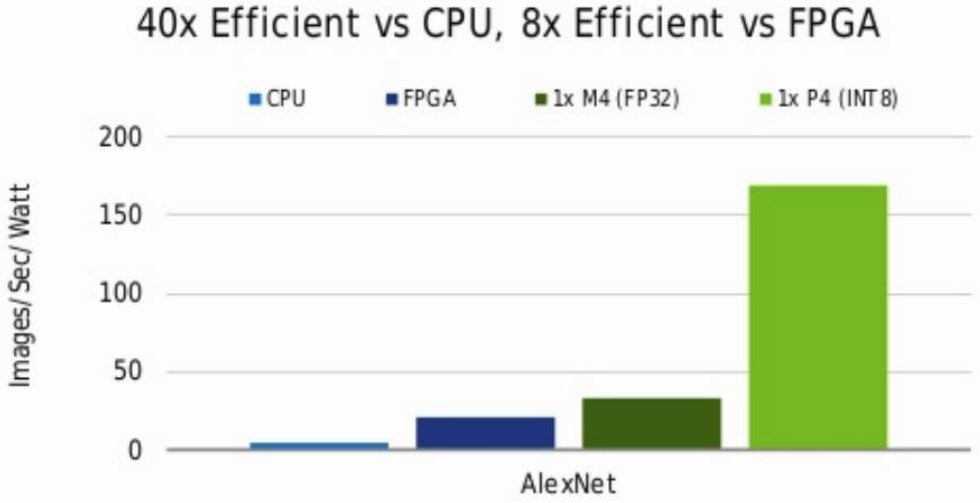
Calcul GPU

Très efficace pour des convolutions multi-bandes sur des fenêtres glissantes

TESLA P4

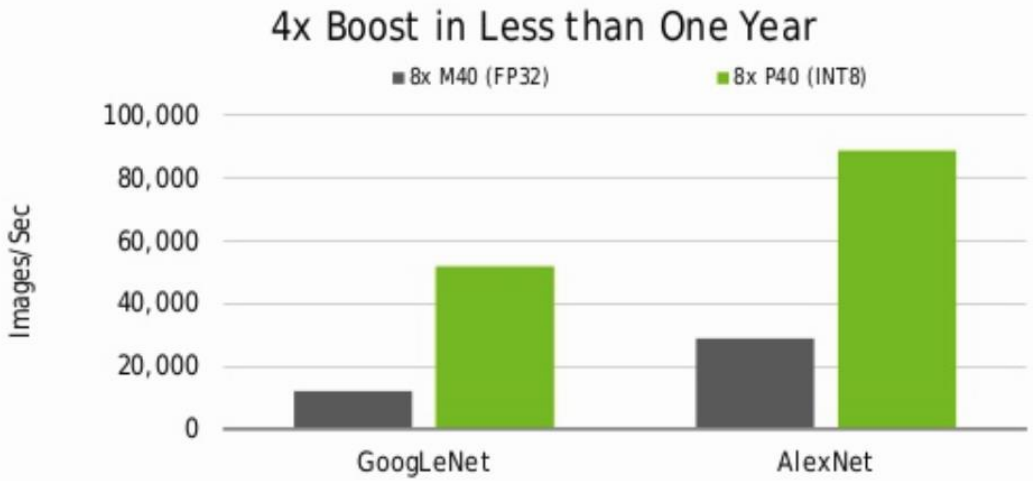
Maximum Efficiency for Scale-out Servers

5.5 TFLOPS



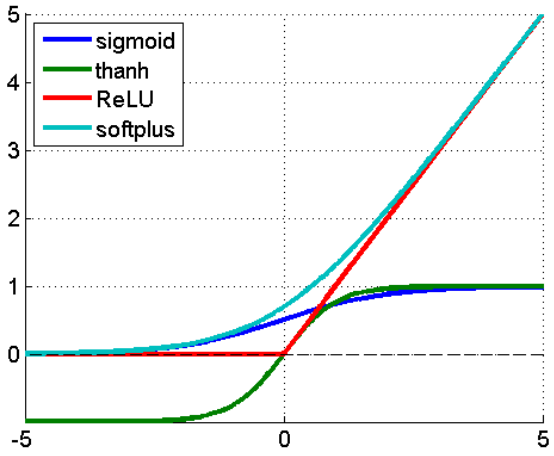
TESLA P40

Highest Throughput for Scale-up Servers

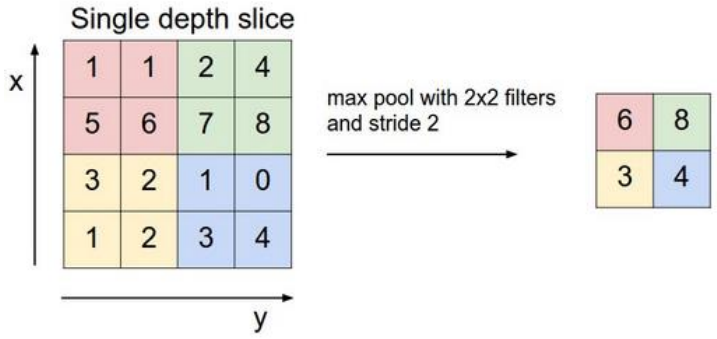


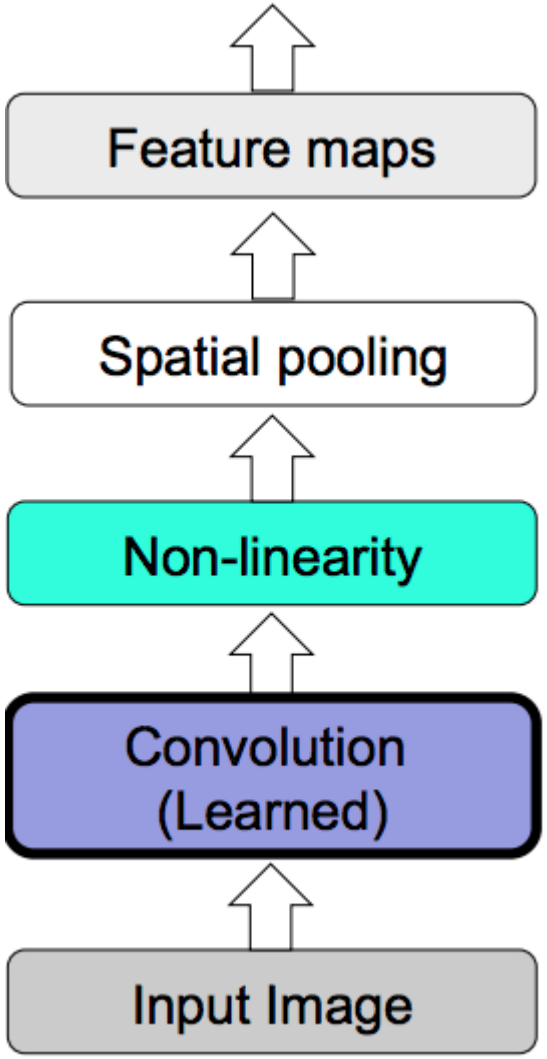
Couches non-linéaires

- Rectifier



- Pooling





Approches traditionnelles

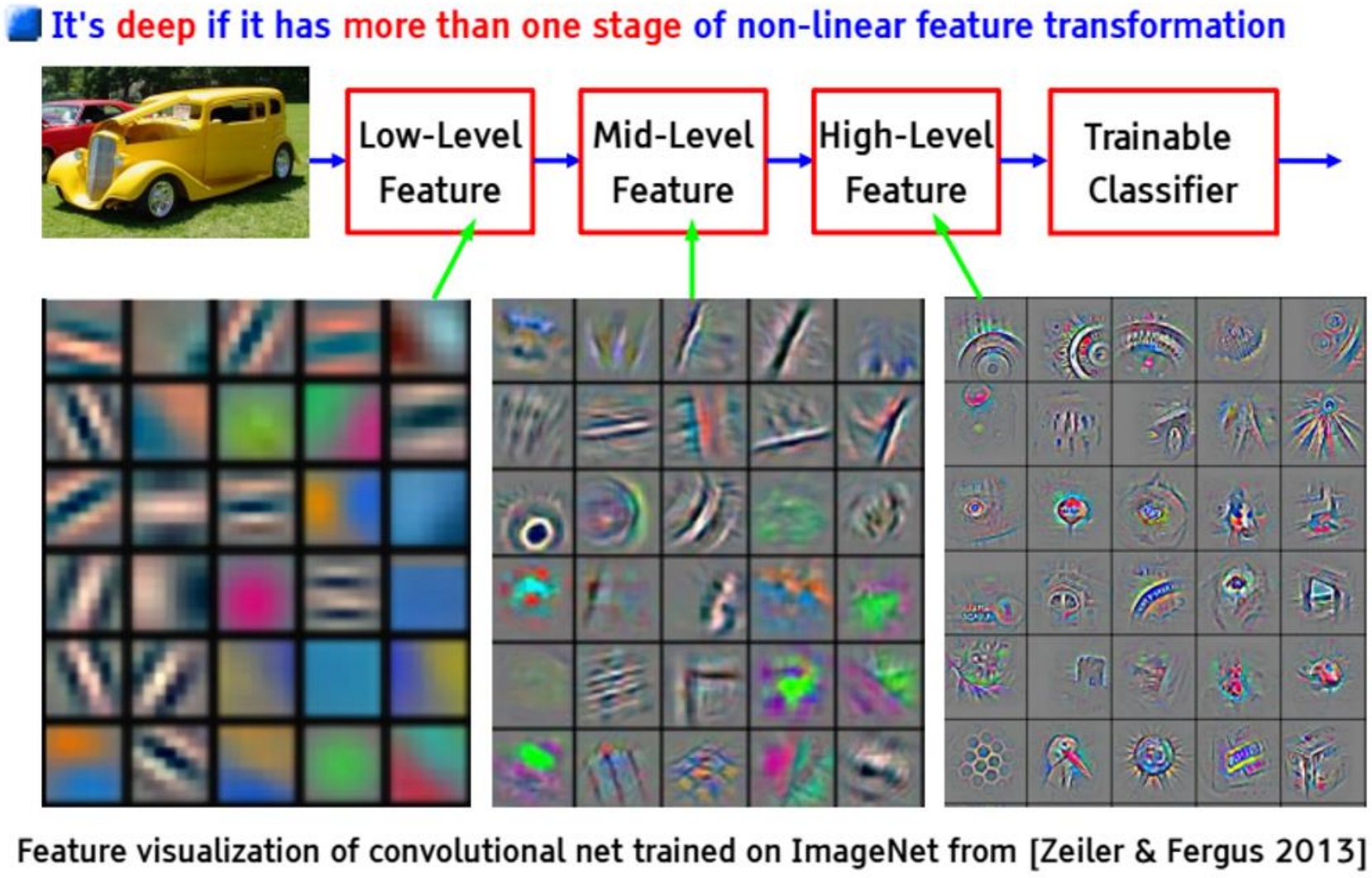
Traditional Pattern Recognition: Fixed/Handcrafted Feature Extractor

Mainstream Modern Pattern Recognition: Unsupervised mid-level features

Deep Learning: Representations are hierarchical and trained

Source: Y. LeCun

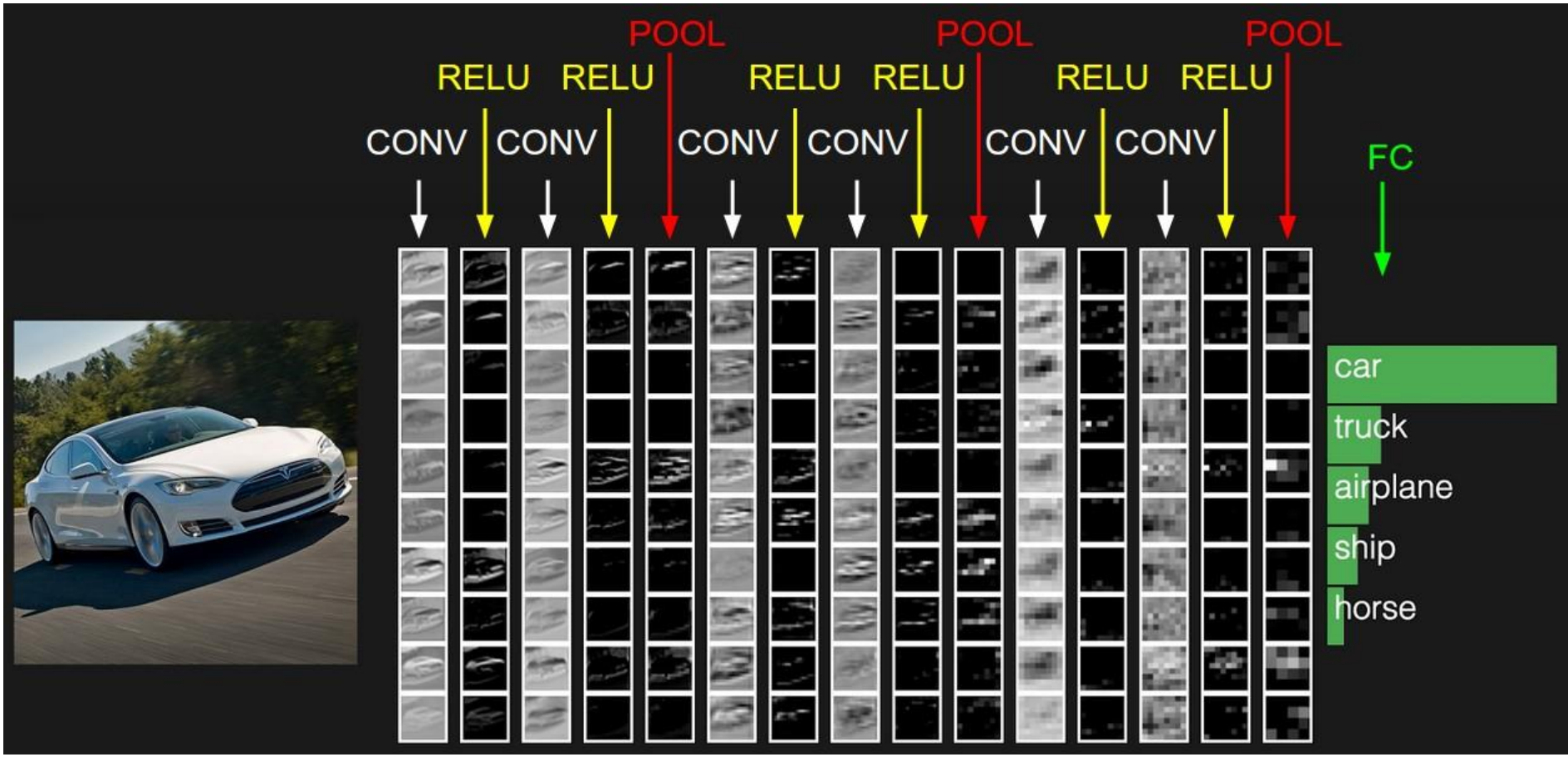
Apprentissage profond = représentation hiérarchique



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Exemple d'architecture ConvNet typique

- Alternance d'opérateurs de couches de convolution et d'opérateurs non-linéaires (ReLU, POOL, etc.)
- Décision finale par un réseau 'Fully Connected'



Source: Standford Univ.

A SHORT HISTORY

AI Winter (1987-1993)

Hinton

Backpropagation

Bengio *LeCun*

Convolutional Network handwriting recognition (digits)

Dave Steinkraus; Patrice Simard; Ian Buck (2005). "Using GPUs for Machine Learning Algorithms"

X30 gain in processing power

ImageNet Creation

DeepFace

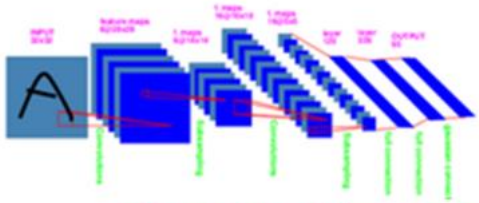
CNN outperforms human on ImageNet

DeepMind AlphaGo

AlphaGo Zero

First CNN on ImageNet

1986 *1998* *2005* *2010* *2012* *2014* *2015* *2016* *2017*

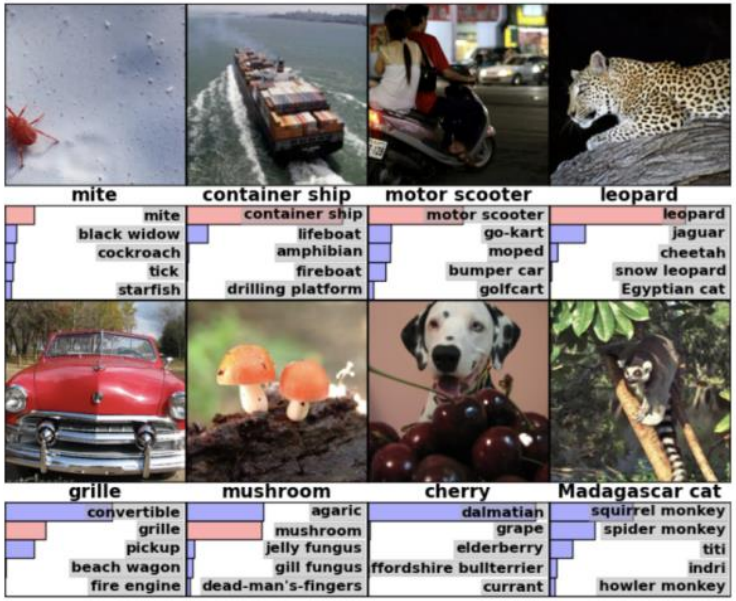


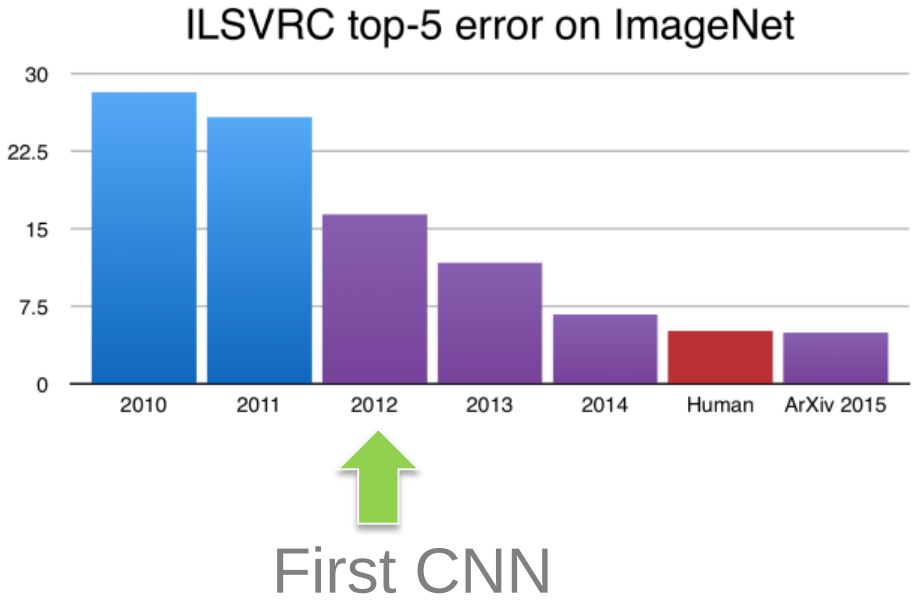
Convolutional layers are integrated into neural networks

Large annotated visual databases (~million)

The annual ImageNet Large-Scale Visual Recognition Challenge (ILSVRC)

- Image labeling (1,000 classes)
- 3.2 million labeled high-resolution images
- CNN outperforms human in 2015

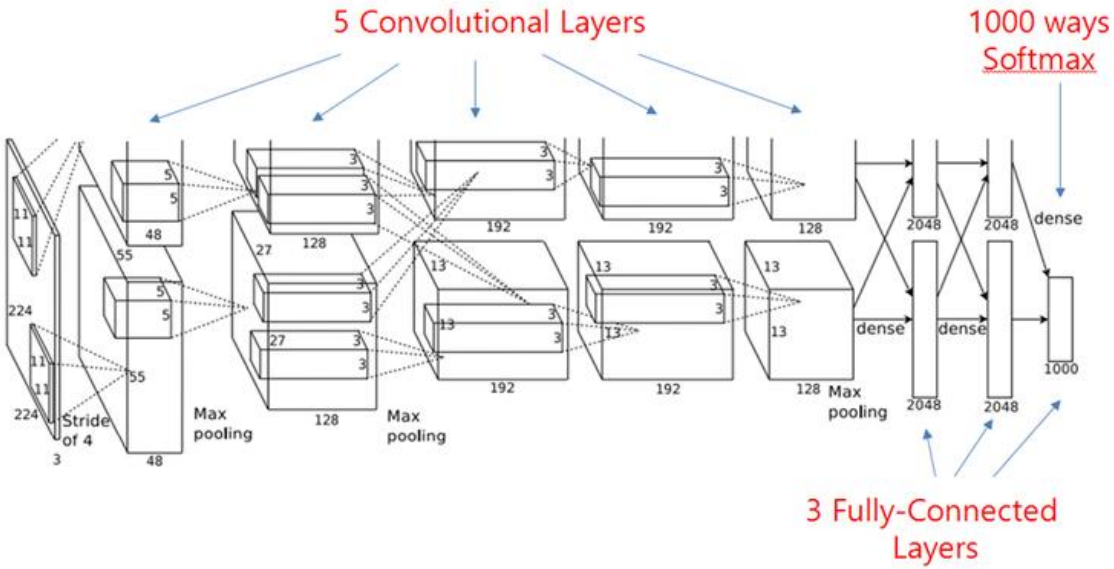


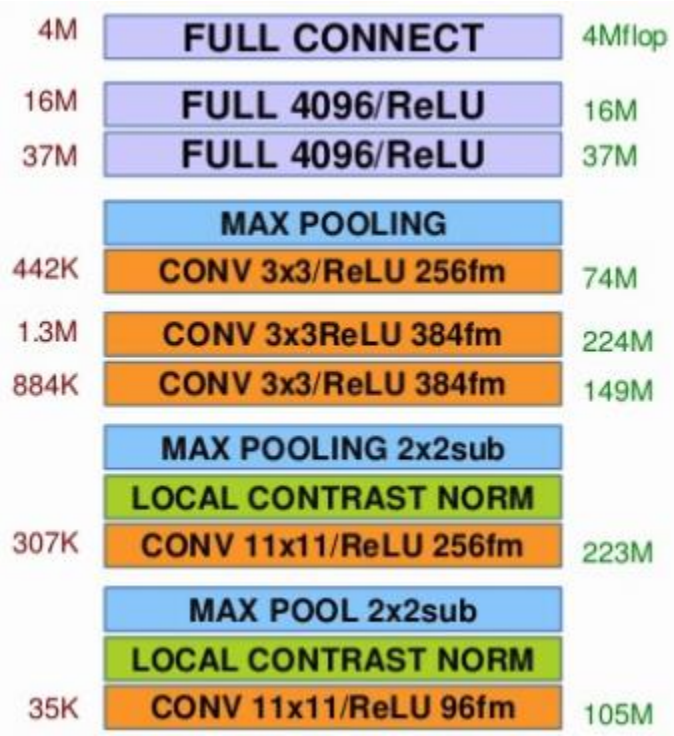


AlexNet (2012)

Innovations:

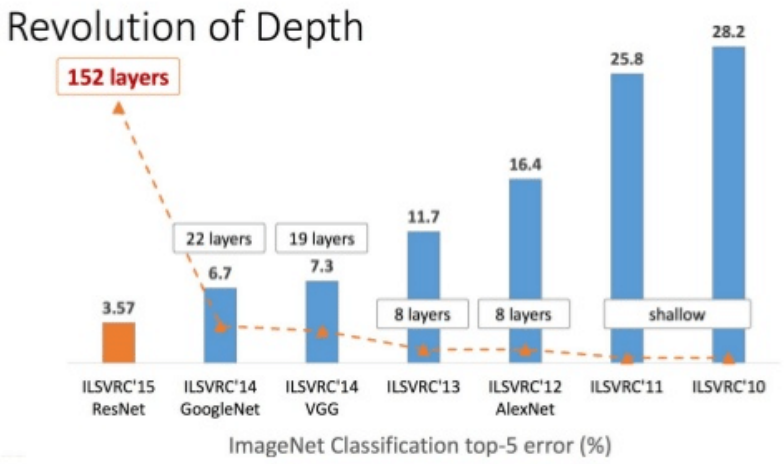
- use of rectified linear units (ReLU) as non-linearities
- use of dropout technique to selectively ignore single neurons during training, a way to avoid overfitting of the model
- overlapping max pooling, avoiding the averaging effects of average pooling
- use of GPUs NVIDIA GTX 580 to reduce training

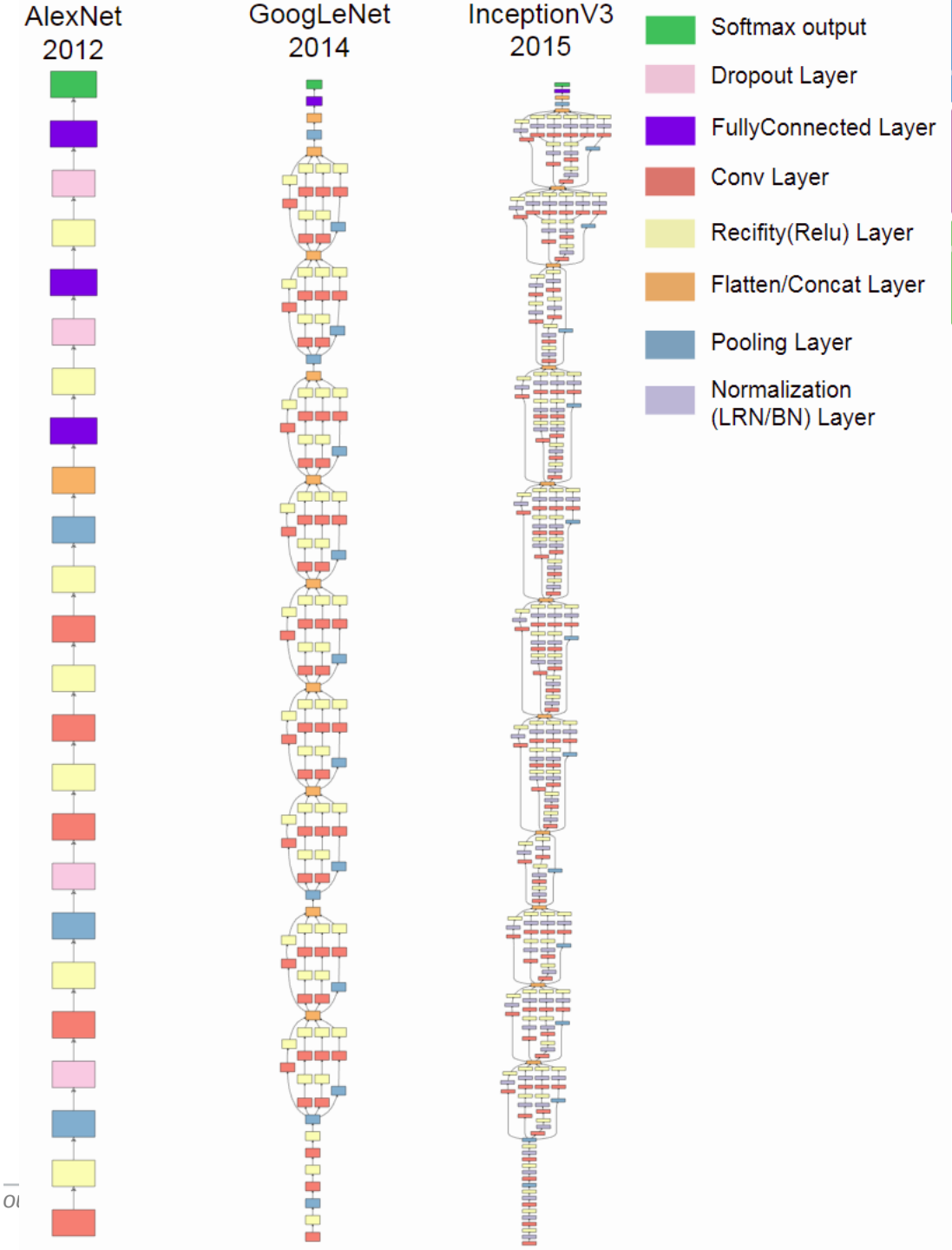




ImageNet Classification with Deep Convolutional Neural Networks, Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E., NIPS 2012

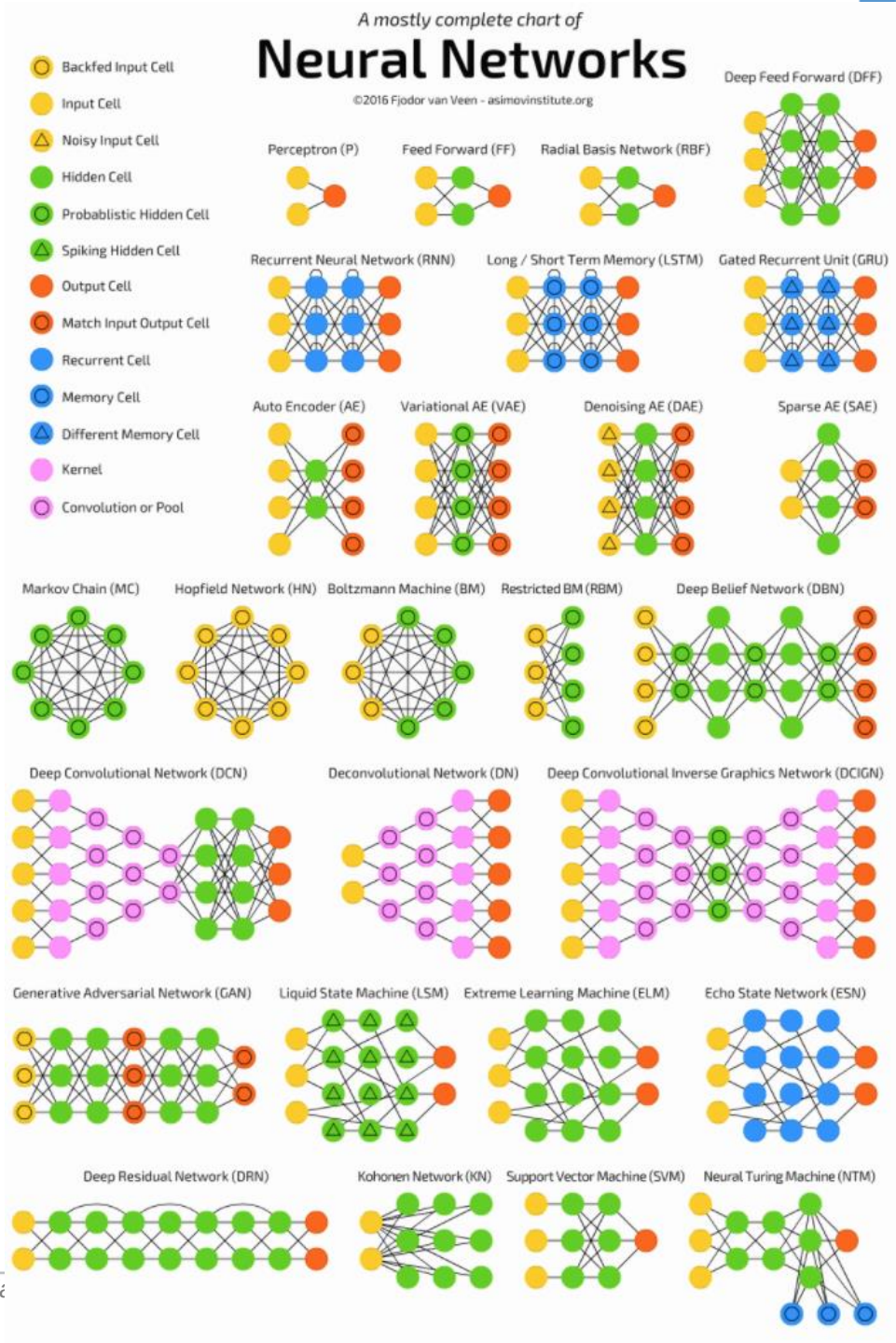
Complexité croissante





The Neural Network Zoo

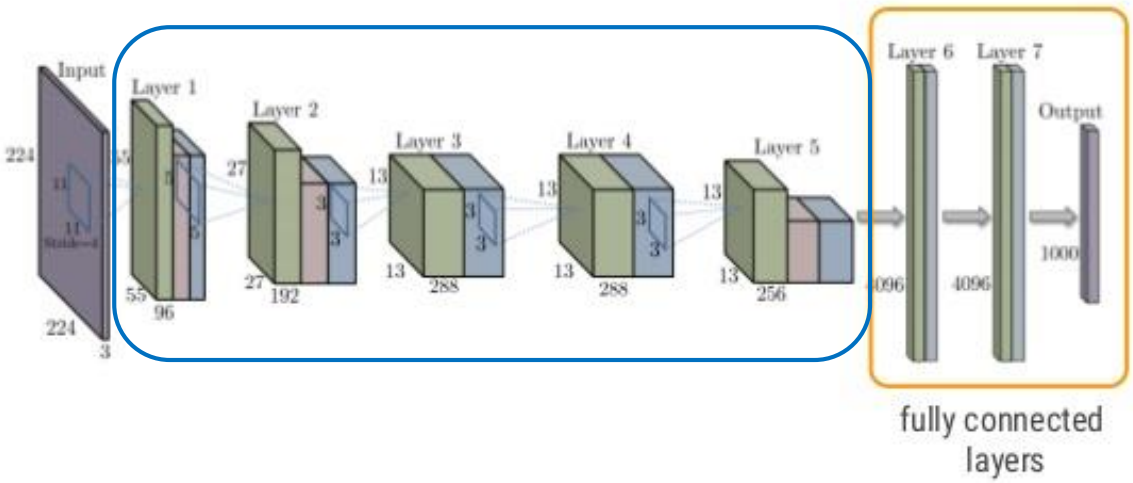
- Feed forward neural networks (FF or FFNN) and perceptrons (P) - 1958
- Hopfield network (HN) - 1982
- Boltzmann machines (BM) – 1986
- Restricted Boltzmann machines (RBM) – 1986
- Autoencoders (AE) – 1988
- Recurrent neural networks (RNN) – 1990
- Long / short term memory (LSTM) – 1997
- Convolutional neural networks (CNN) – 1998
- Deep belief networks (DBN) – 2007
- Variational autoencoders (VAE) – 2013
- Generative adversarial networks (GAN) - 2014
- Deep residual networks (DRN) - 2015



Quelques exemples de projets

Utilisation des CNN

- Approche ‘Deep Features’:
  - On utilise un réseau déjà entraineur comme producteur de ‘features’
  - Un classificateur est ajouté pour produire les classes d’intérêts
  - Différents CNN entrainés sur des données en vision (CaffeNet, GoogleNet, …)

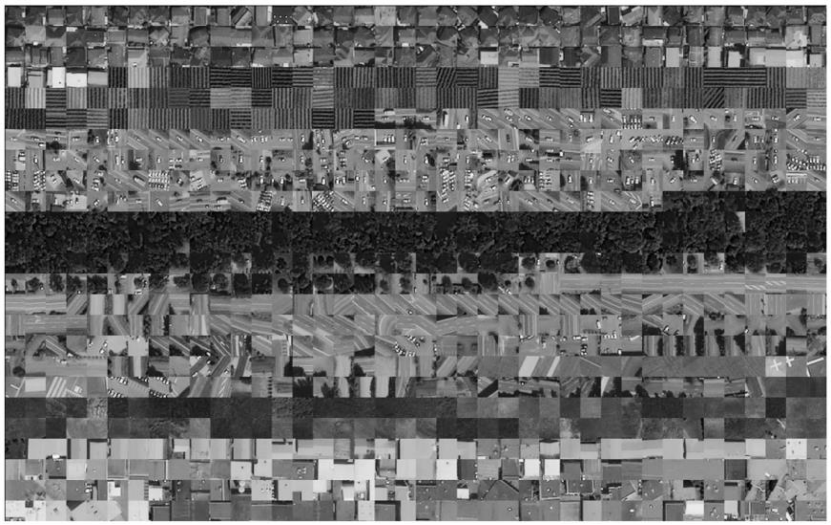


Pleiades imagery (50 cm)

- Training data

  40k images (32x32 R+G+B+NIR) from **Vancouver** area

  + data augmentation (7 rotations)



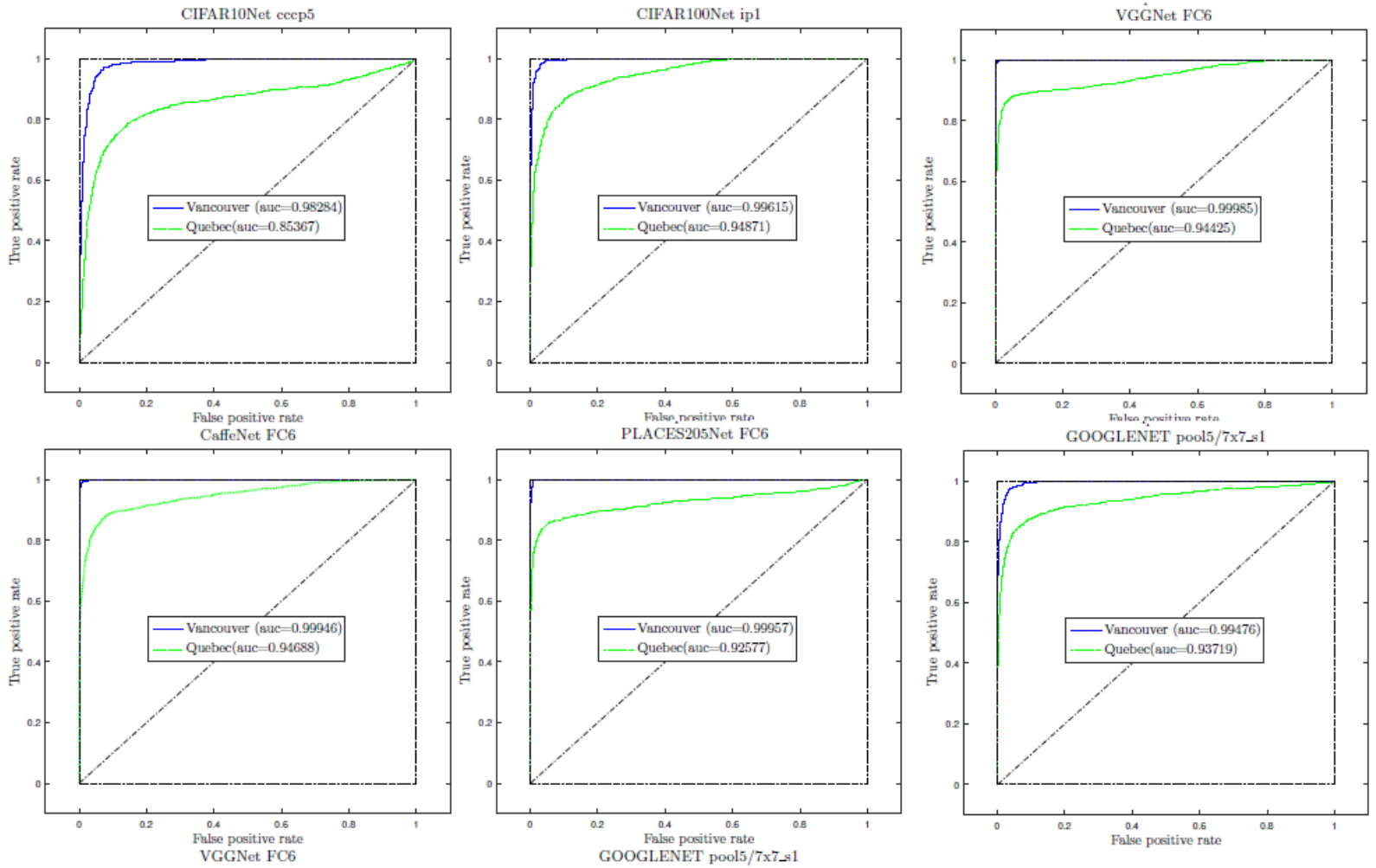
Data	Car	Non-car
Train-Vancouver	6,944	33,344
Test-Vancouver	872	4,176
Test-Quebec	2,565	5,670

Car/non-car database characteristics.

- Test data

  5k images (32x32) from **Vancouver** area
  8k images (32x32) from **Quebec** area

Résultats

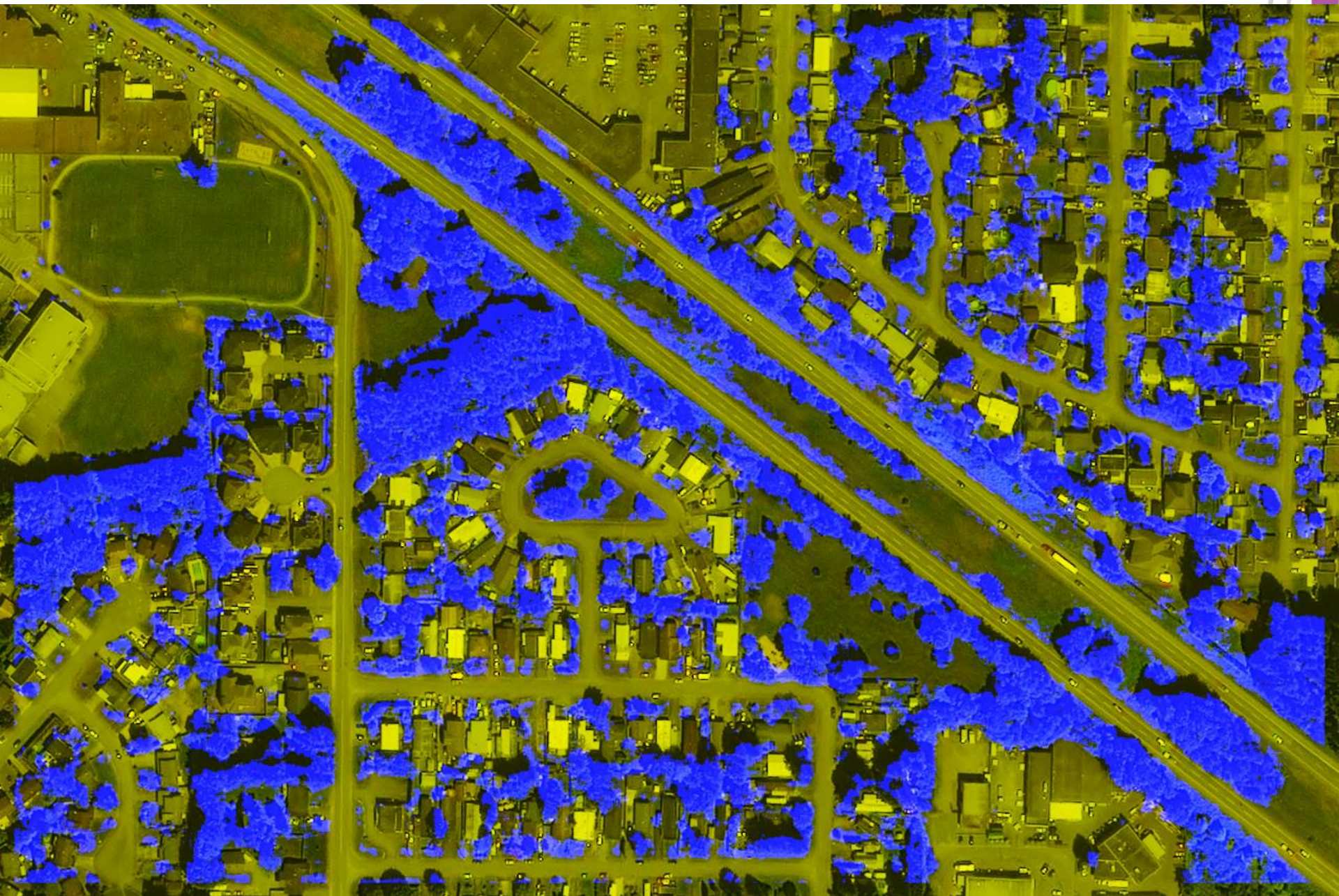


ROC curves for car detection on Vancouver- and Quebec-test regions.

Ref: Dahmane, Mohamed, Samuel Foucher, Mario Beaulieu, François Riendeau, Yacine Bouroubi et Mathieu Benoit. « The Potential of Deep Features for Small Object Class Identification in Very High Resolution Remote Sensing Imagery ». In : 14th International Conference on Image Analysis and Recognition (ICIAR). Montreal, 2017.

- Results of Deep features
- **Tree detection**

- Vancouver area



- Results ‘deep features’
- **Car detection**
- Vancouver area

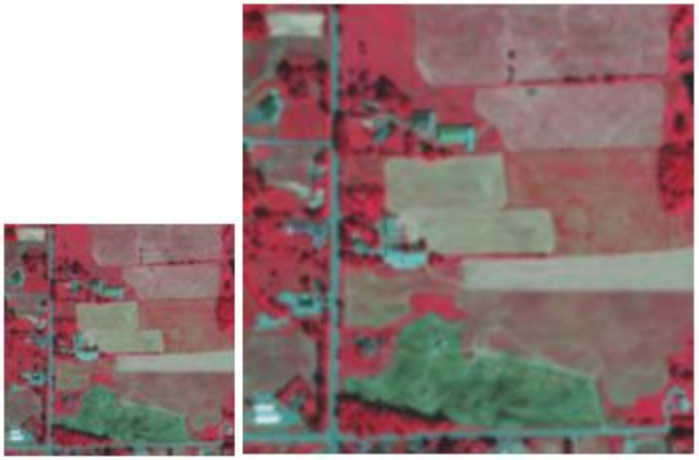
Plateforme GeoImageNet

- Une plateforme web collaborative pour l'annotation d'images Satellites à très haute résolution
- Annotation basée sur une grande taxonomie (classes de couverture terrestre et objets)
- Prévoir un grand jeux d'entraînement pour un système similaire à ImageNet
- Images Pleiade et WorldView (30-50cm) sur le Canada pour l'instant
- Fournir des services de référence en apprentissage machine
- Dépôt de modèles disponible pour téléchargement
- Informations sur les régions d'intérêt annotées disponibles au téléchargement
- Ouvert à la communauté des chercheurs (sur invitation)
- En ligne en juin 2019

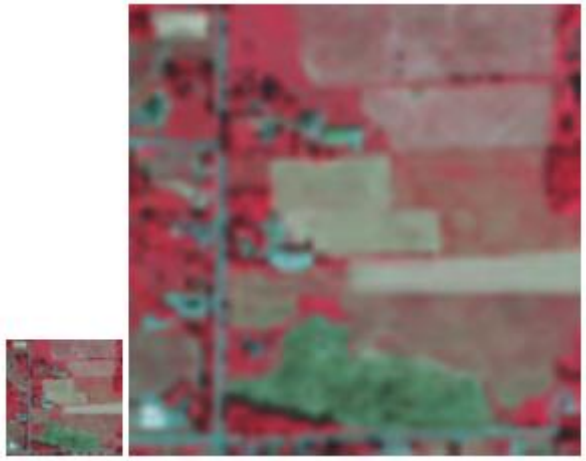
Techniques de super-résolution

Objectifs :

1. Améliorer la résolution d'une image Sentinel-2 à 10 m par un facteur x2 (5 m) ou x4 (2.5 m)
2. Performances en classification, détection d'objets, etc.



5 à 2.5 mètres

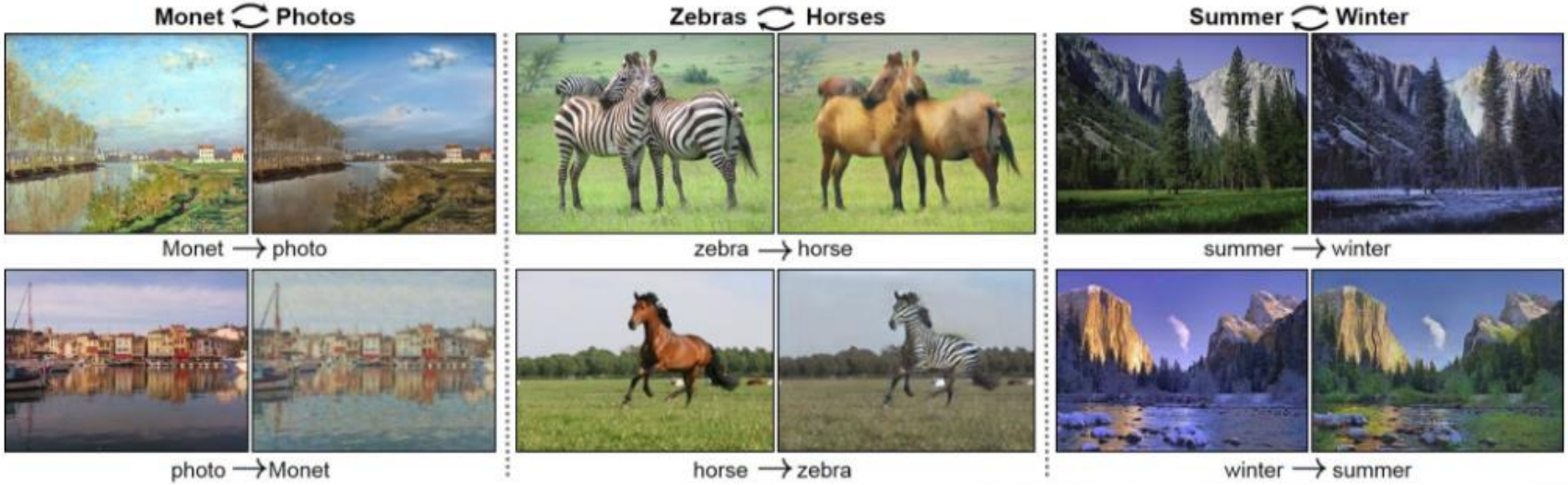


10 à 2.5 mètres

Image cible

Techniques de super-résolution

- *Transfert de catégorie d'objet, de style, etc.*

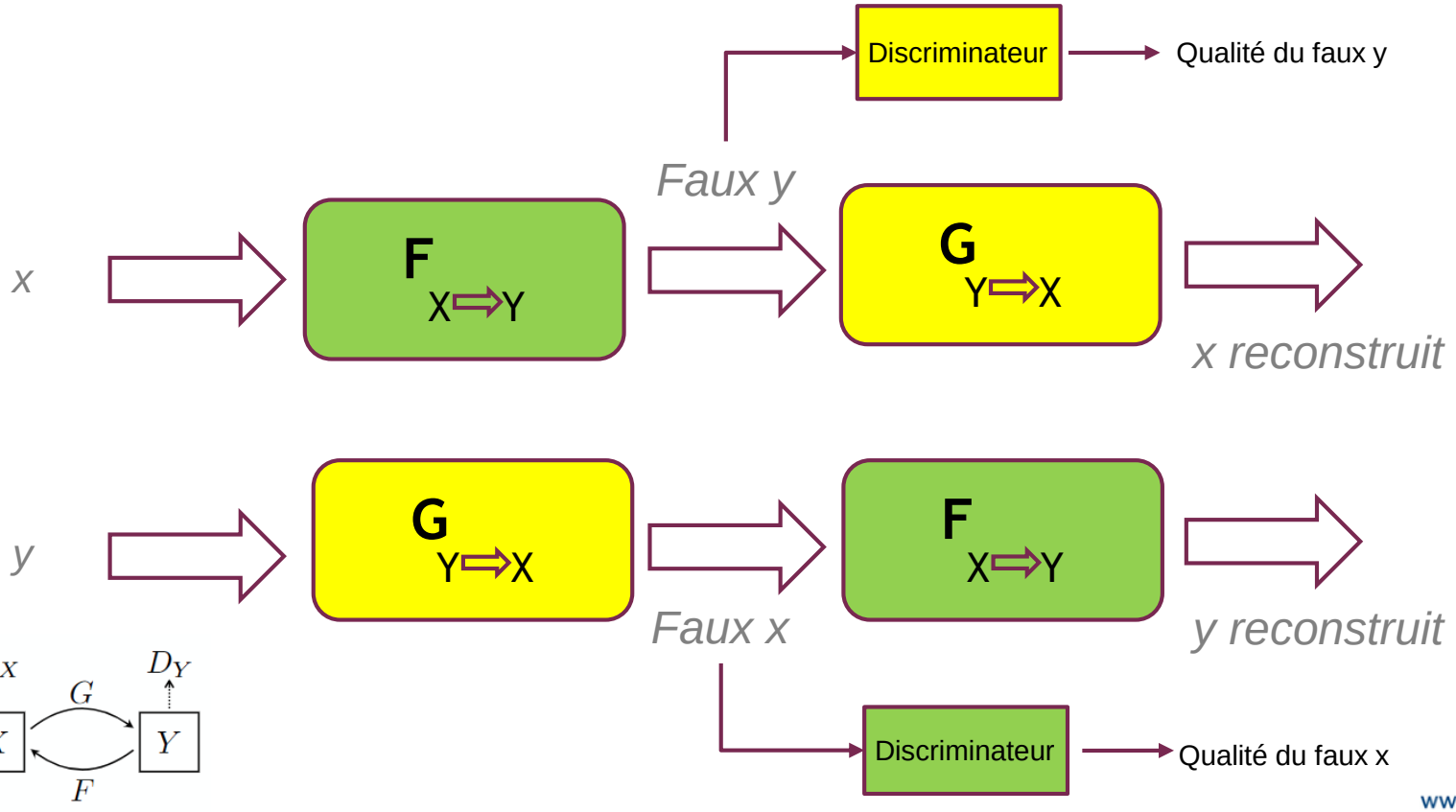


- *Transfert entre une carte et une image satellite*

Ref: Jun-Yan Zhu, Taesung Park*, Phillip Isola, and Alexei A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", in IEEE International Conference on Computer Vision (ICCV), 2017.*

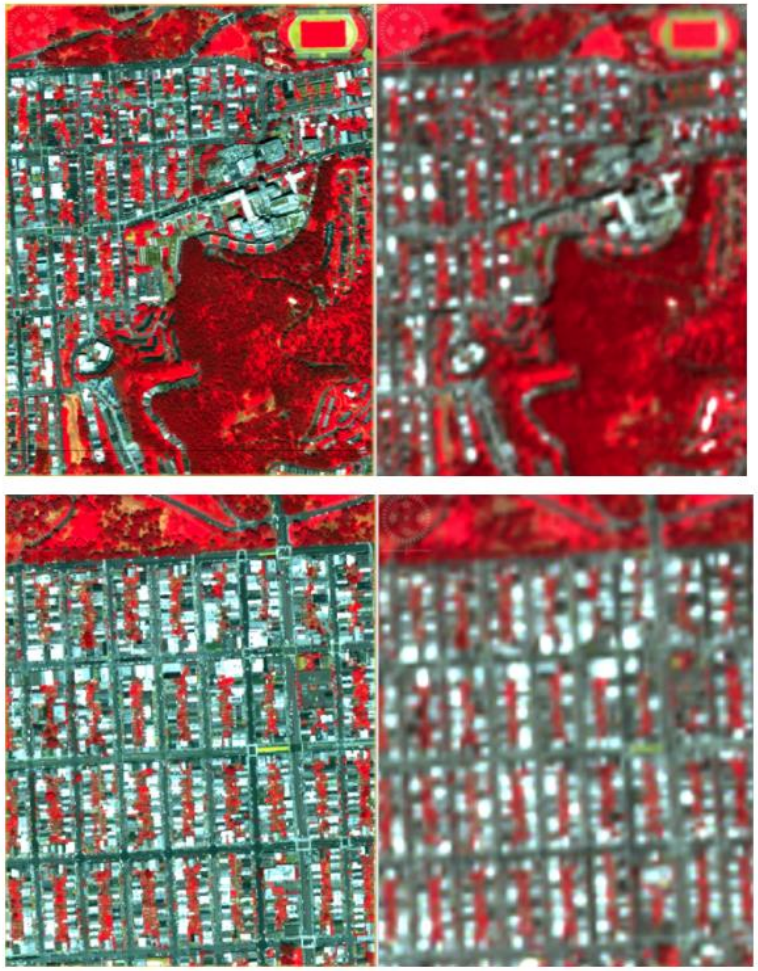
Techniques de super-résolution

- *Cycle - Generative Adverserial Neural Networks* (GANs)
- Transfert image à image entre deux domaines (X et Y)
- Par exemple: X= images en noir et blanc, Y= images couleurs
- Deux réseaux de neurones F et G entrainés pour passer d'un domaine à l'autre
- Deux réseaux de neurones (discriminateurs) qui jugent de la qualité des faux
- Itérations jusqu'à obtenir de bons 'faux': F(G(y))=y et bonne qualité des faux



Techniques de super-résolution

- Y = Woldview dégradée à 2.5m de résolution (Bicubique)
- X = Sentinel-2 à 10m et rééchantillonnée à 2.5m (Bicubique)

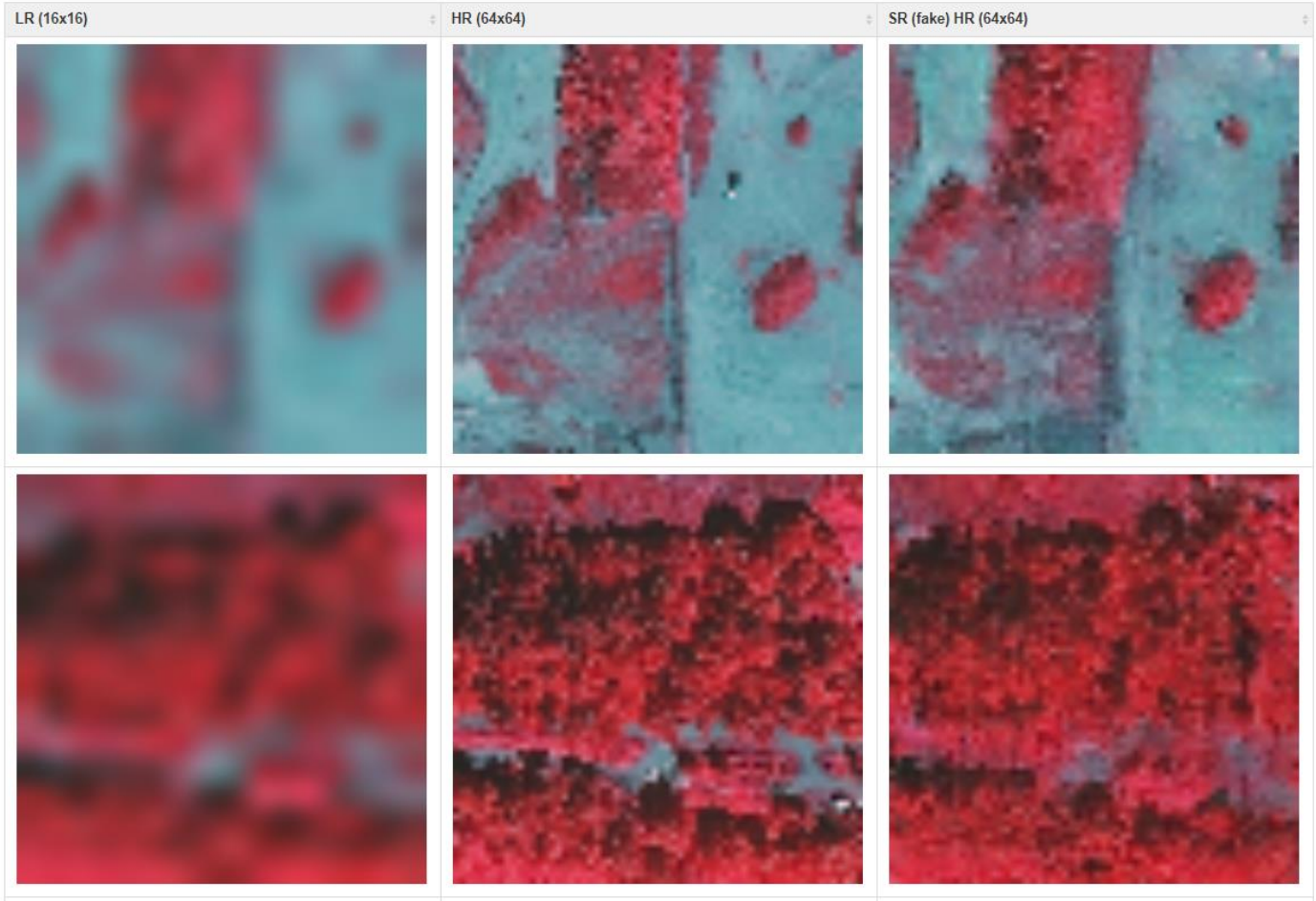


(*Proche-Infrarouge*, *Rouge*, *Vert*)

Techniques de super-résolution

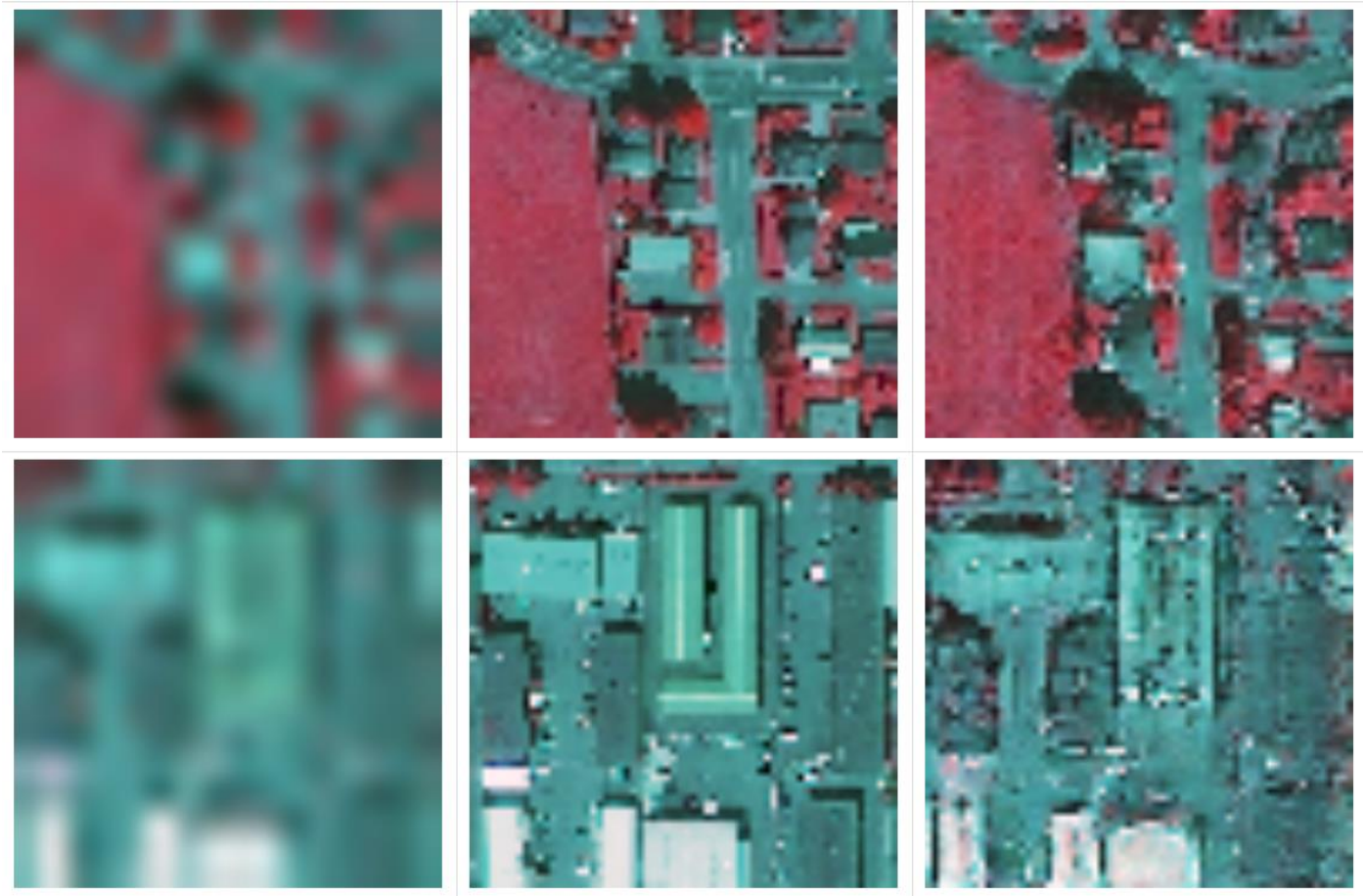
- Transfert de résolution spatiale?

x (Sentinel-2) *y (Worldview)* *Faux y*



Techniques de super-résolution

x (Sentinel-2) *y (Worldview)* *Faux y*



Données « BarkNet » de l'Université Laval

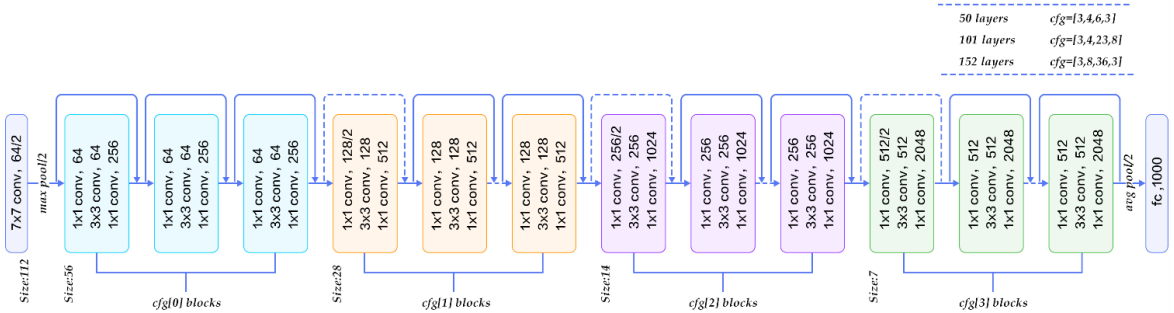
Bouleau jaune

Érable plane

Pruche du Canada

Pin Blanc

23 espèces, ~23k images

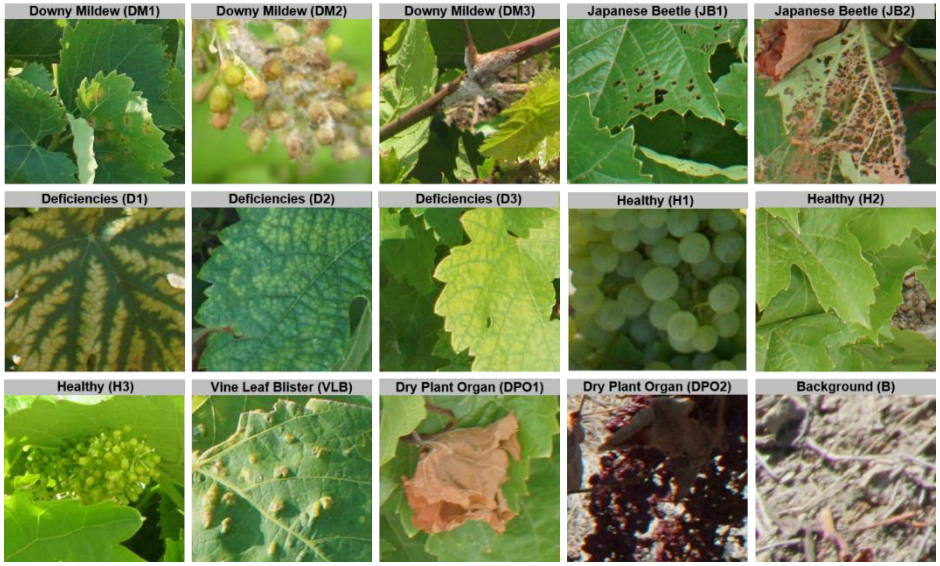


'ResNet-101' classique + data augmentation: **Accuracy = plus de 98%!** (pourriez-vous faire mieux?)

Ressources naturelles Canada
Service canadien des forêts
Natural Resources Canada
Canadian Forest Service

Identification automatique de problèmes phytosanitaires de la vigne

Analyse multi-classes : 7 labels



Grande variabilité intra-classe :

- Plusieurs organes (feuilles, grappes, tiges),
- Plusieurs stades de croissance,
- Plusieurs niveaux de développement de maladie.

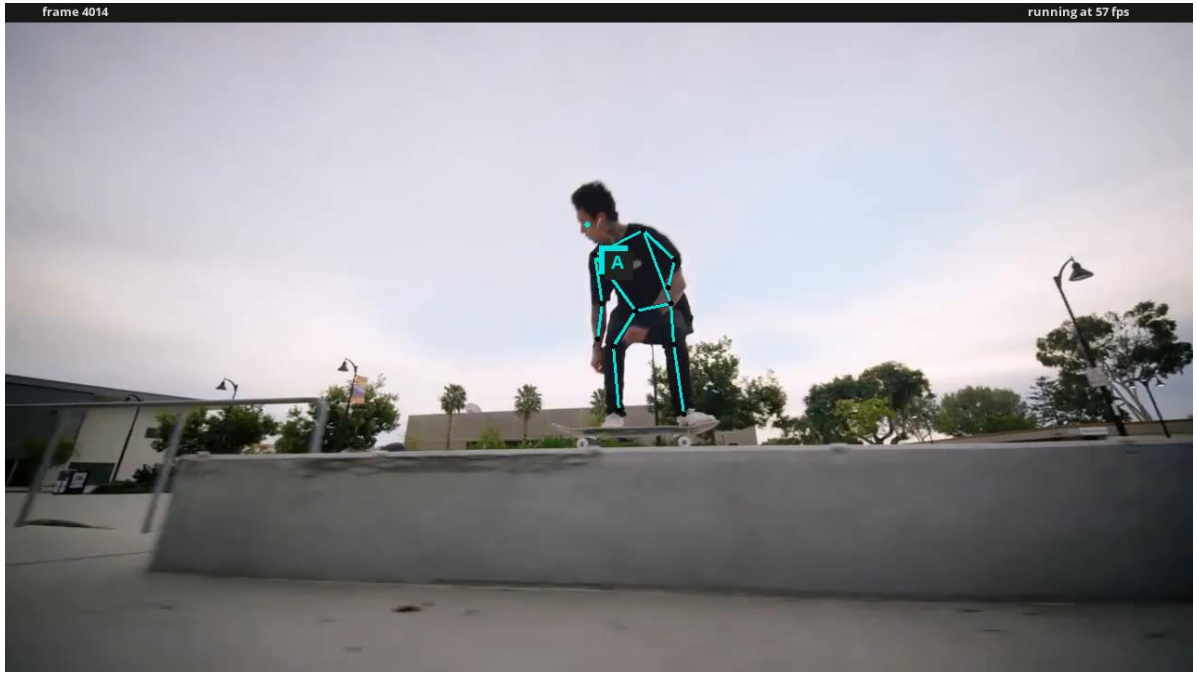
Précision globale sur un ensemble de test indépendant :

1. ResNet-101 : 95,75%,
2. ResNet-18 : 95,59%,
3. ResNet-152 : 95,53%
4. ResNet-50 : 95,43%,
5. ResNet-34 : 95,32%.

Matrice de confusion du ResNet-101 :

t/p	DM	JB	D	H	VLB	DPO	B	Total
DM	420	0	23	2	0	0	6	451
JB	3	222	0	1	0	0	1	227
D	3	0	224	15	1	0	5	248
H	2	1	2	706	2	0	3	716
VLB	0	0	0	0	42	0	0	42
DPO	3	0	0	2	0	38	2	45
B	0	0	0	3	0	0	151	154
Total	431	223	249	729	45	38	168	1883

Stage de Master, Hugo Bonnome: Estimation de Pose 3D
(https://drnoodle.github.io/fastpose_html/)



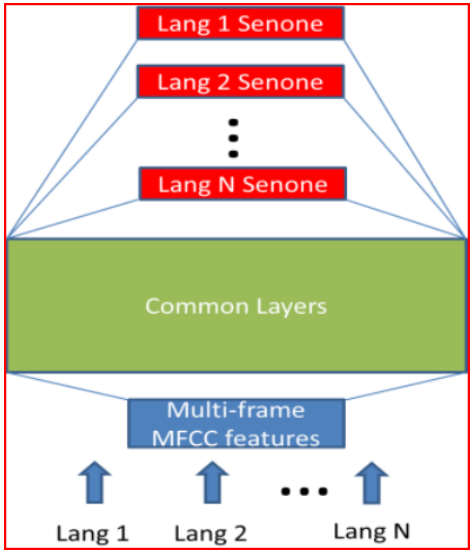
Un modèle pour estimer position des joints en 2D, et un modèle pour projeter en 3D…

Traitement de la parole

Sous-titrage télévision en direct

- Production des sous-titres en temps réel, en français (SOVO)
  - avec sous-titreurs
  - délai de 3 secondes ou moins
- Au départ : modèles acoustiques GMM-HMM (2010)
  - requièrent adaptation à chaque locuteur
  - mieux avec entraînement discriminatif (MMI)
- Réseaux neuronaux profonds (DNN)
  - mieux même sans adaptation au locuteur (multilocuteurs)
  - entraînés avec empreinte (i-vecteur) résumant les caractéristiques de locuteur : s'adaptent en aveugle en temps réel
  - entraînés multi-tâches avec données en anglais et français
- Impacts
  - Réduction des coûts
    - ➢ élimine le travail que devait faire chaque nouveau sous-titreur: corriger à la main les textes produits par le système pour permettre l'adaptation
  - Amélioration de la qualité
    - ➢ De 13.4 % à 8.0 % d'erreur : qualité subjective passe de « médiocre » à « bonne »

Système	WER
Point de départ	13.4%
GMM MMI, dép. locuteur	10.3%
DNN multilocuteurs	9.2%
Avec i-vecteurs	8.7%
DNN multilingues	8.0%



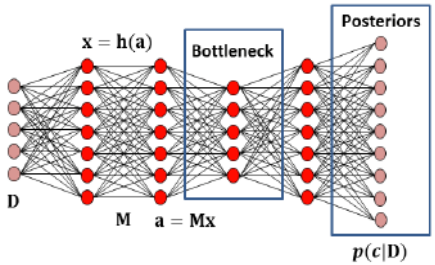
Biométrie vocale

– Vérification du locuteur (VL)

- **Évaluation NIST Speaker Recognition 2016**
- **Dernier NIST SRE 2012 : pas de deep learning**
- **Nouvelle approche : représentations profondes**
  - ➢ Compensation non-linéaire des i-vecteurs avec DNN (NWCN)
  - ➢ Classification de l'identité (SCN) : représentation « bottleneck »
- **Performances améliorées p/r état de l'art pour :**
  - ➢ Données langue étrangère, courtes durées, bruitées

– Détection d'usurpation

- **Réseau profond + représentation bottleneck**
  - ➢ 20 % d'amélioration p/r ASVspoof Challenge 2015



Tiré de Richardson et al. 2005

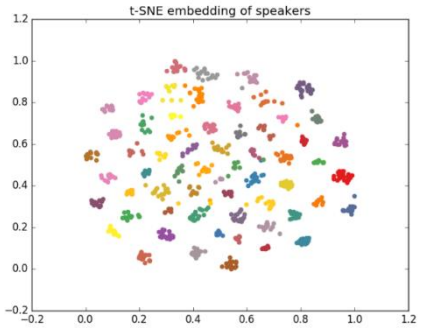


Figure 1: i-vector speaker space

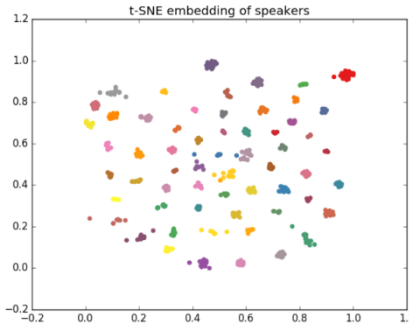


Figure 2: NWCN speaker space

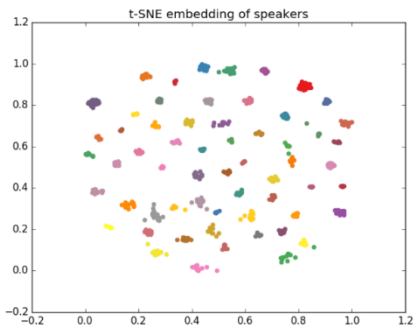


Figure 3: SCN speaker space

CONCLUSIONS

- **Véritable révolution dans de nombreux domaines d'application avec des gains en performance significatifs (classification d'images, reconnaissance de la parole, analyse du langage naturel, etc.)**
- **Beaucoup de ressources à disposition:**
  - **Librairies: Caffe, TensorFlow, etc.**
  - **Ressources de calcul: AWS, Google Cloud, Microsoft Azure, etc.**
  - **"Services cognitifs" maintenant mis à la disposition par Google, IBM et Microsoft**
- **Émergence dans le domaine du géospatiale et de la télédétection depuis 2015-2016**
- **Quelques défis:**
  - **Degré de liberté important dans le choix des architectures**
  - **Beaucoup de paramètres = beaucoup de données**
  - **Effet 'boite noir' qui peut être problématique dans certains domaines (ex: médecine)**
  - **Il semble relativement facile de tromper un réseau de neurones**
- **Le CRIM peut aider les entreprises à démystifier l'IA pour leur domaine par la réalisation de projets:**
  - **Longue expérience en accompagnement technologique en particulier pour les PME sur des problématiques de R-D appliquées**
  - **Transfert de savoir-faire pour accélérer l'appropriation de l'IA par les entreprises et augmenter leur autonomie**

WWW.CRIM.CA

Samuel Foucher

Équipe Vision et imagerie
CRIM – Centre de recherche informatique de Montréal

Samuel.Foucher@crim.ca

Le CRIM est un centre de recherche appliquée et d'expertise en technologies de l'information qui rend les organisations plus performantes et compétitives par le développement de technologies innovatrices et le transfert de savoir-faire de pointe, tout en contribuant à l'avancement scientifique.

Il permet aux organisations, principalement les PME, de démystifier et d'avoir accès aux technologies de pointe comme celles de l'intelligence artificielle afin de résoudre efficacement les problématiques technologiques auxquelles elles sont confrontées.

Ses chercheurs et professionnels en TI développent un large éventail d'applications dans des secteurs diversifiés et œuvrent dans des domaines d'expertises tels que l'apprentissage automatique, la vision par ordinateur, la reconnaissance de la parole, le traitement automatique des langues naturelles, la science des données et la recherche opérationnelle.

Le CRIM est un organisme sans but lucratif et sa neutralité et la force de son réseau en font une ressource incontournable. Son action s'inscrit dans les politiques et stratégies pilotées par le ministère de l'Économie, de la Science et de l'Innovation, son principal partenaire financier.