Les architectures CNN en Computer Vision

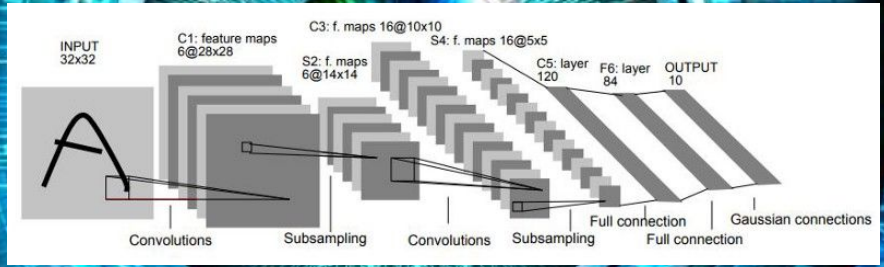
AlexNet, VGG, GoogLeNet, and ResNet

Introduction

LeNet-5

- Application du Deep Learning à la Computer Vision
- Reconnaissance de chiffres manuscrits avec des CNN

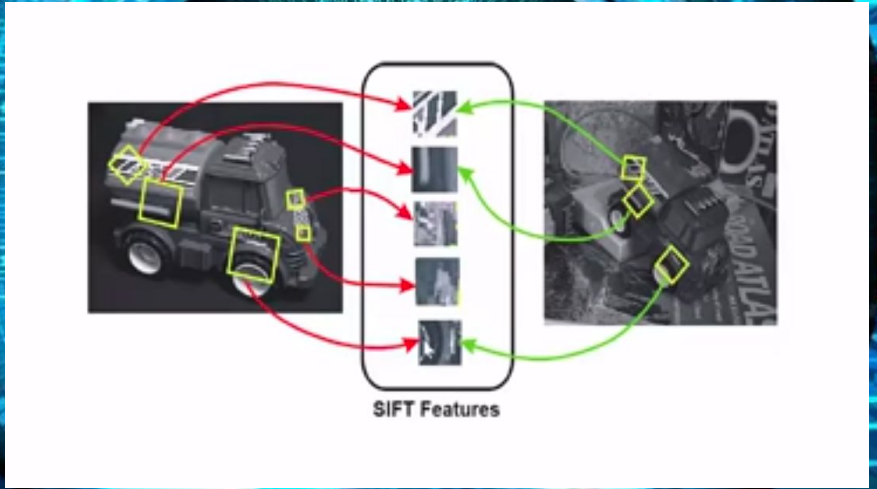
Gradient-based learning applied to document recognition (1998)



Introduction

SIFT (Scale-Invariant Feature Transform)

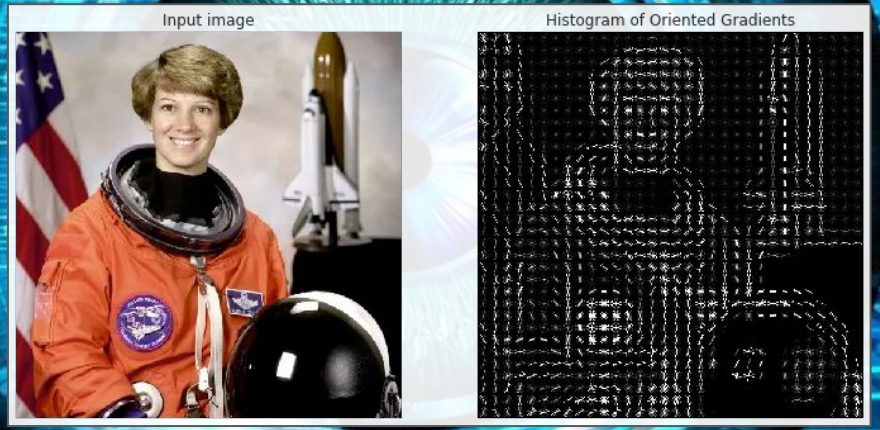
Distinctive image features from scale-invariant keypoints (2004)



Introduction

HOG (Histogram of Oriented Gradients)

Histograms of oriented gradients for human detection (2005)



Introduction

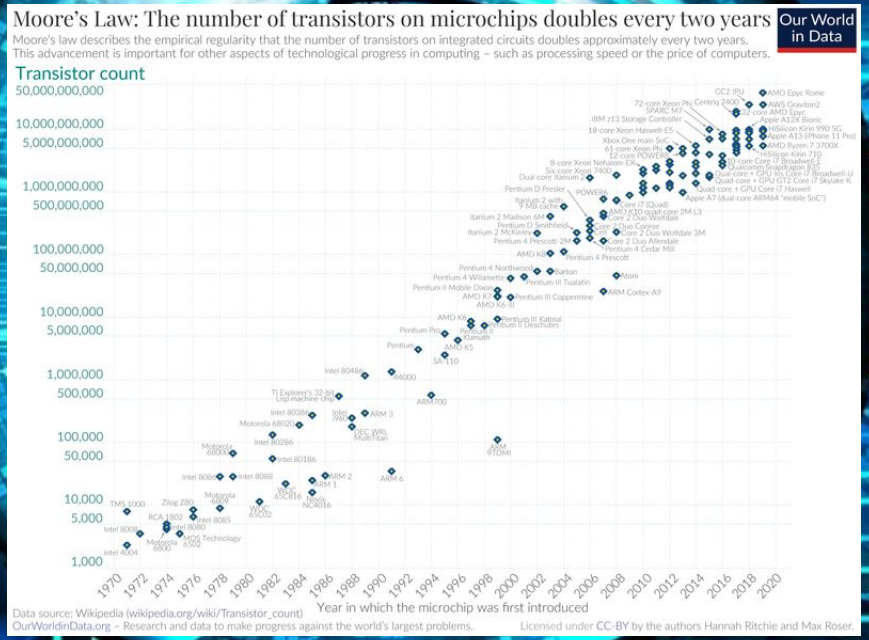
ImageNet

- Conçu en 2009 par Fei-Fei Li et son équipe de Stanford
- 3,2 millions d'images, en moyenne 500 à 1000 images par classes

ImageNet: A Large-Scale Hierarchical Image Database (2009)

Introduction

Loi de Moore

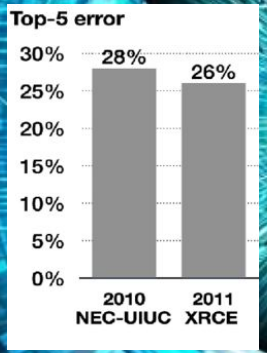


Introduction

ILSVRC

- 2010 : ImageNet Large Scale Visual Recognition Challenge

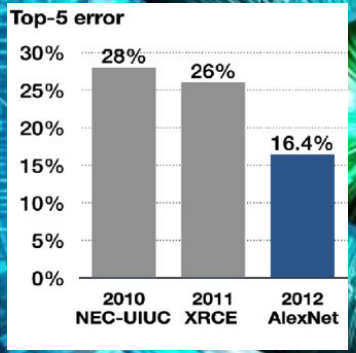
ImageNet: A Large-Scale Hierarchical Image Database (2009)



AlexNet

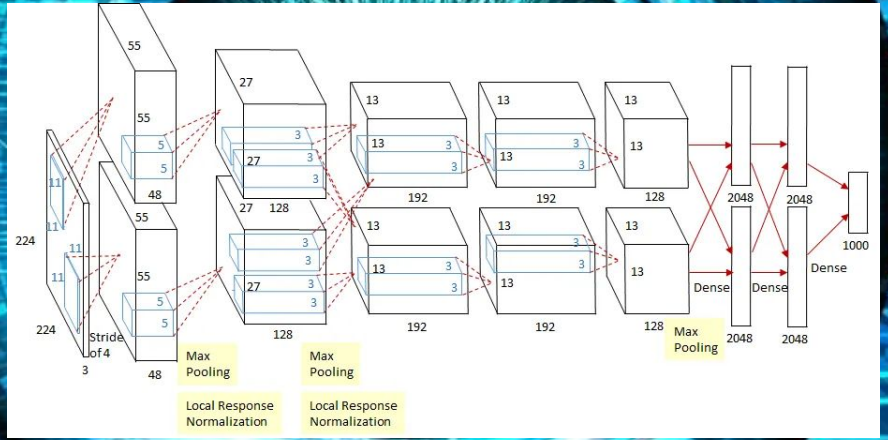
2012 : AlexNet remporte le challenge ImageNet (ILSVRC), 1er réseau de neurones profond à gagner le challenge

ImageNet Classification with Deep Convolutional Neural Networks (2012)



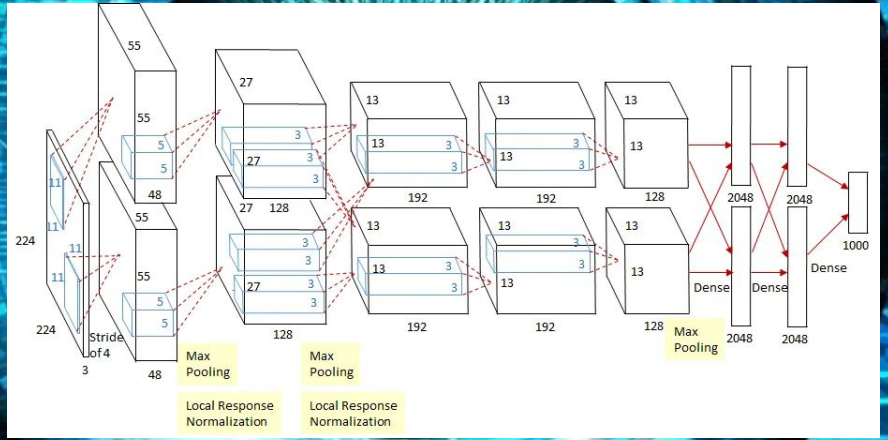
AlexNet

Architecture :
5 couches convolutives
+ 3 couches fully connected
= 8 couches



AlexNet

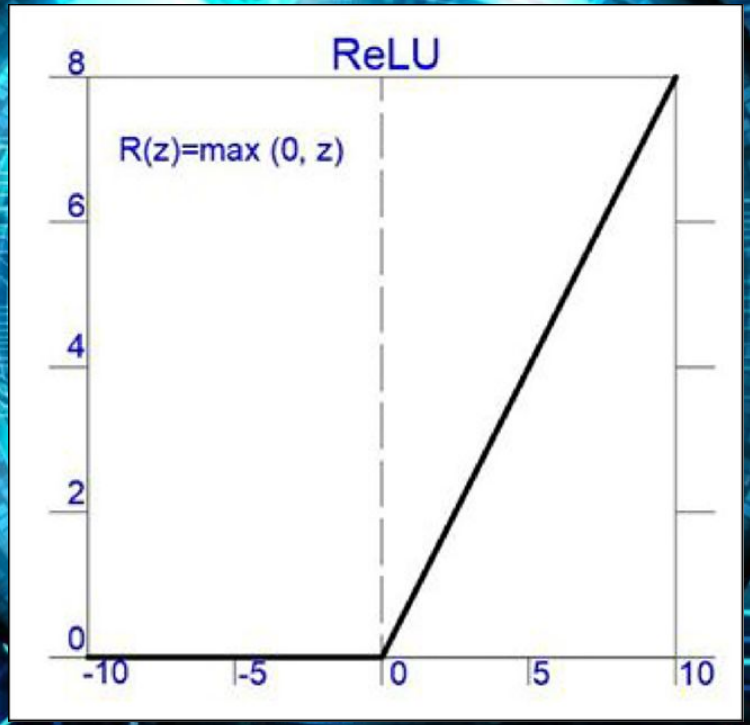
Parallélisation GPU



AlexNet

Rectified Linear Units (ReLU) :

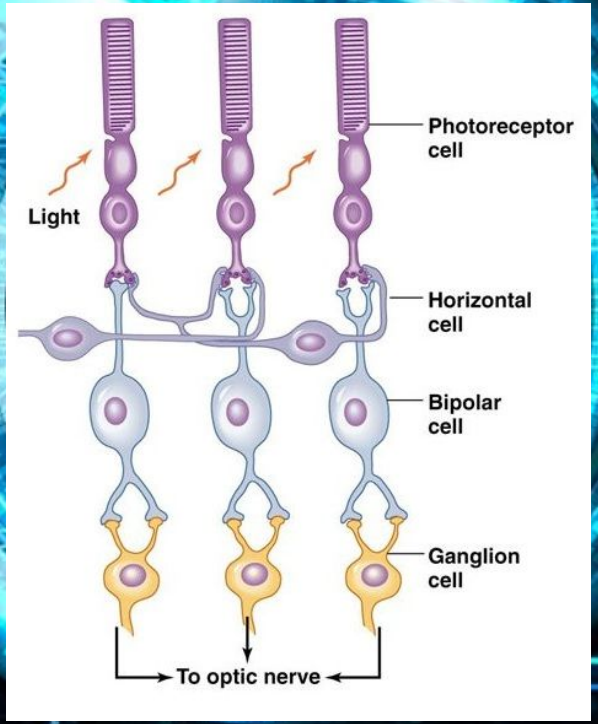
Fonctions d'activation ReLU pour introduire la non-linéarité sans gradient vanishing



AlexNet

Local Response Normalization (LRN) :

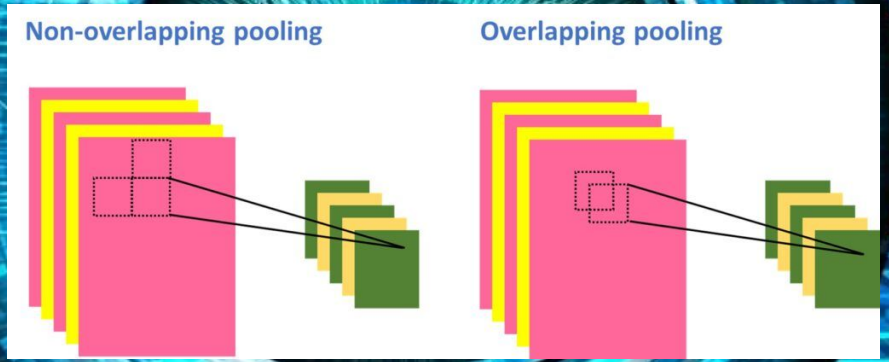
Normalisation des neurones par rapport aux neurones voisins : améliore la sélectivité des neurones



AlexNet

Overlapping pooling :

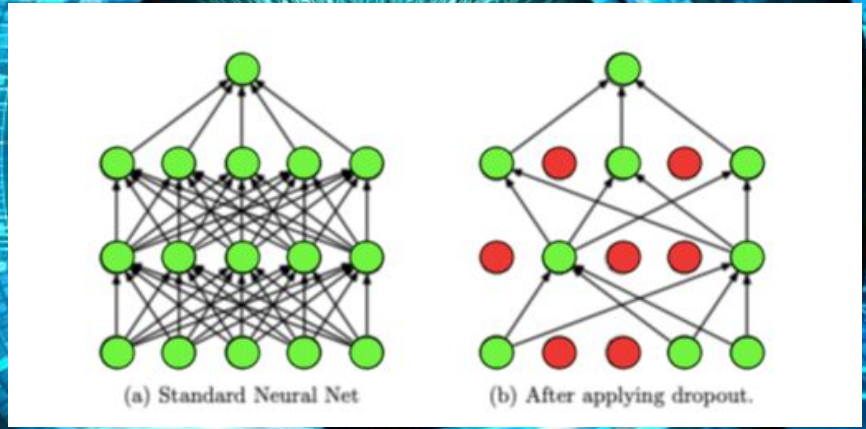
taille des filtres > taille du stride



AlexNet

Régularisation dropout :

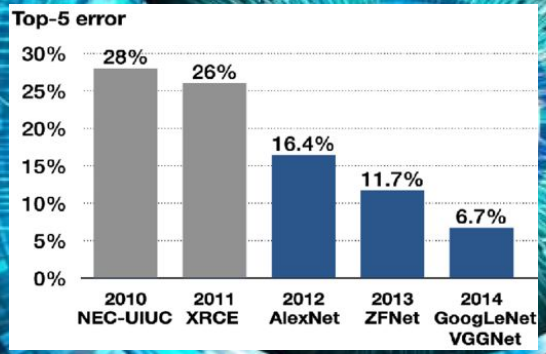
Dropout dans les couches fully connected pour éviter l'overfitting



VGG

2014 : VGGNet est classé parmi les meilleurs au challenge ImageNet (ILSVRC)

Very Deep Convolutional Networks for Large-Scale Image Recognition (2014)

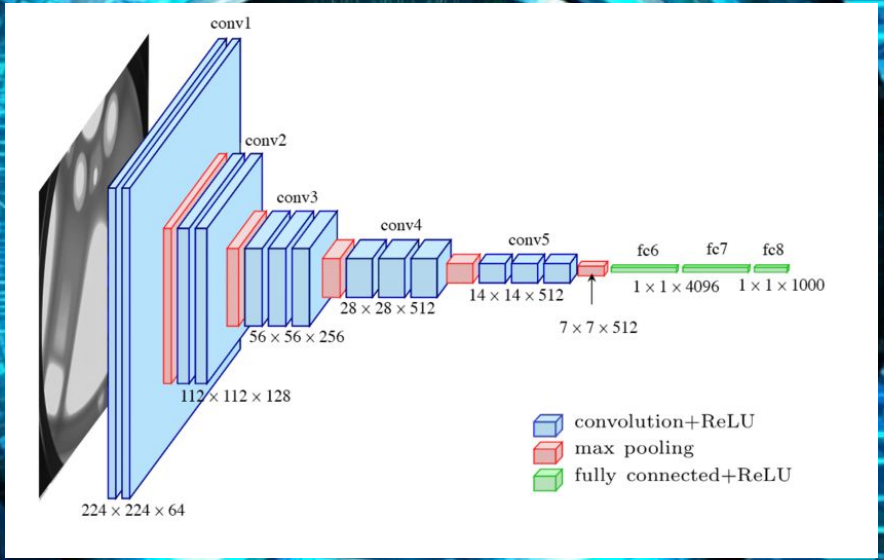


VGG

Exploration en profondeur :

Augmenter la profondeur du réseau augmente la précision : jusqu'à 16-19 couches de poids

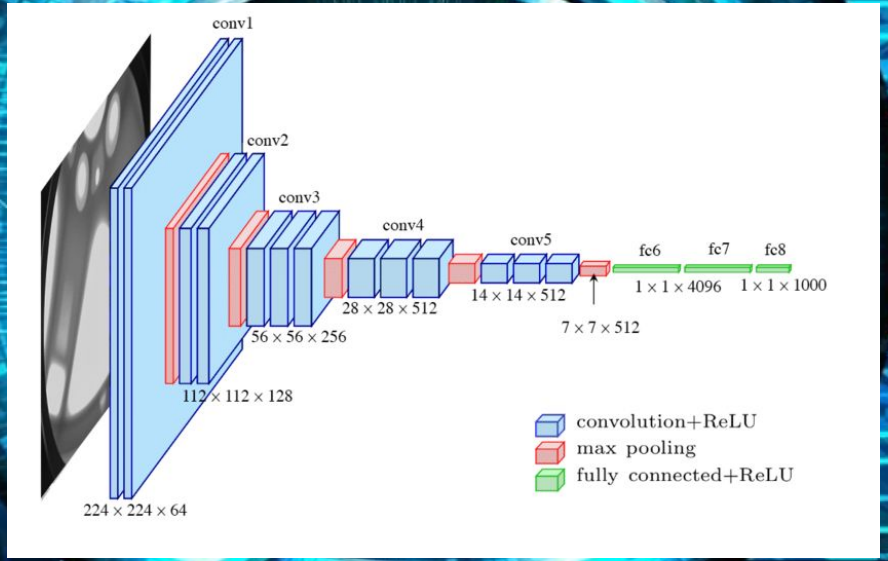
Very Deep Convolutional Networks for Large-Scale Image Recognition (2014)



VGG

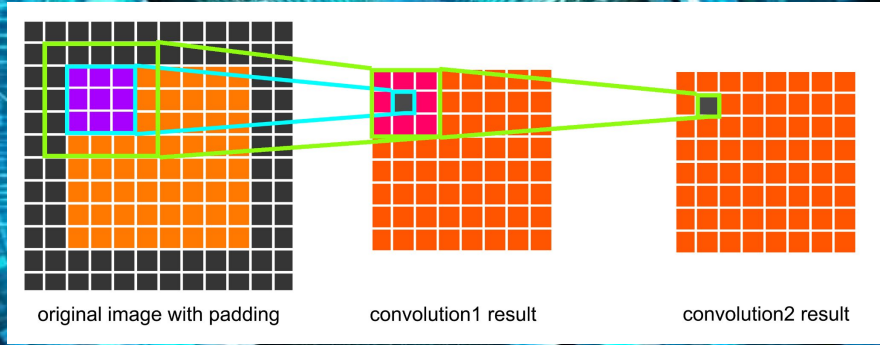
Max-Pooling pour le sous-échantillonnage :

Couches de max-pooling avec une fenêtre 2x2 et un stride de 2 pour sous-échantillonner



VGG

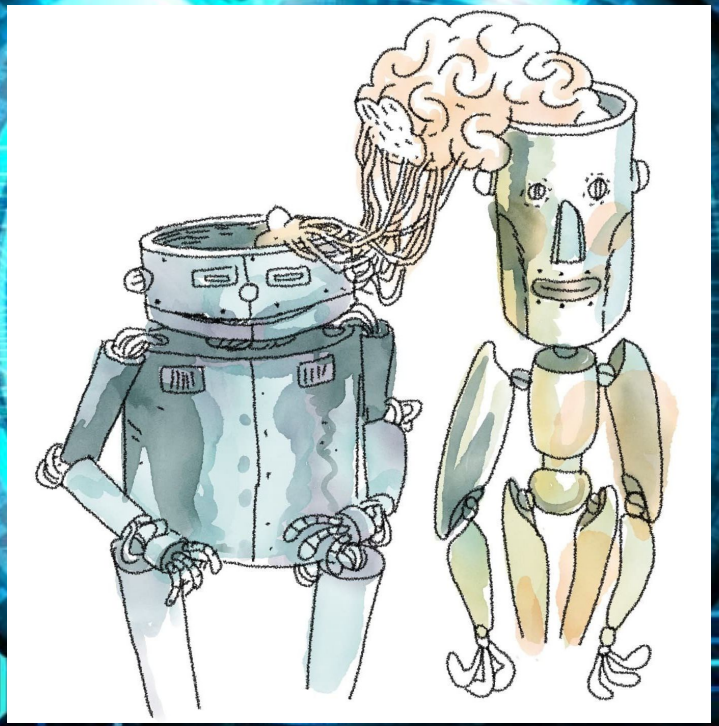
Petits filtres convolutifs (3x3)



VGG

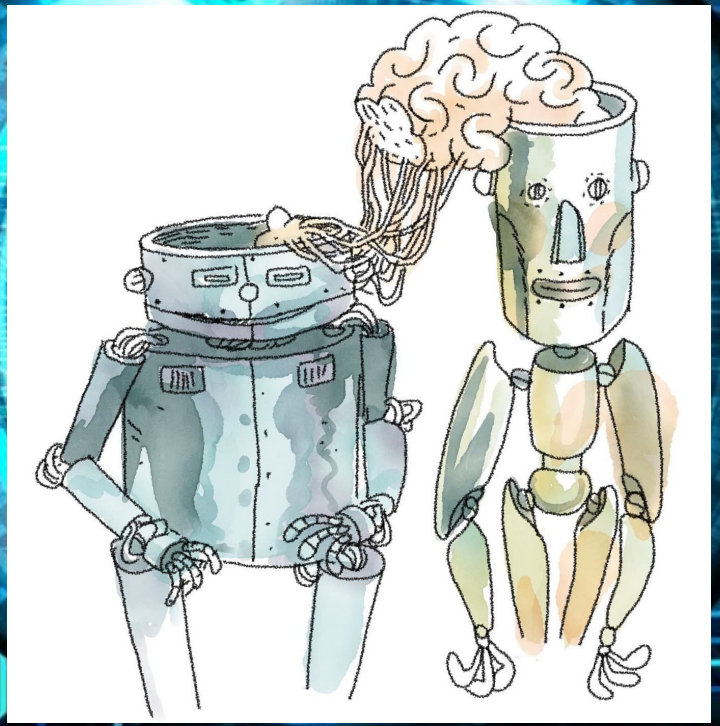
Transfert learning :

L’architecture simple et polyvalente permet l'apprentissage par transfert



VGG

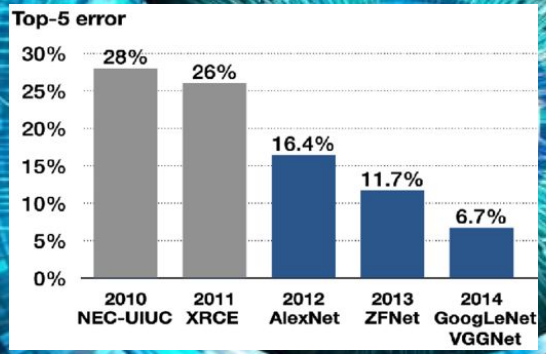
Modèles ConvNet de VGG disponibles en open source



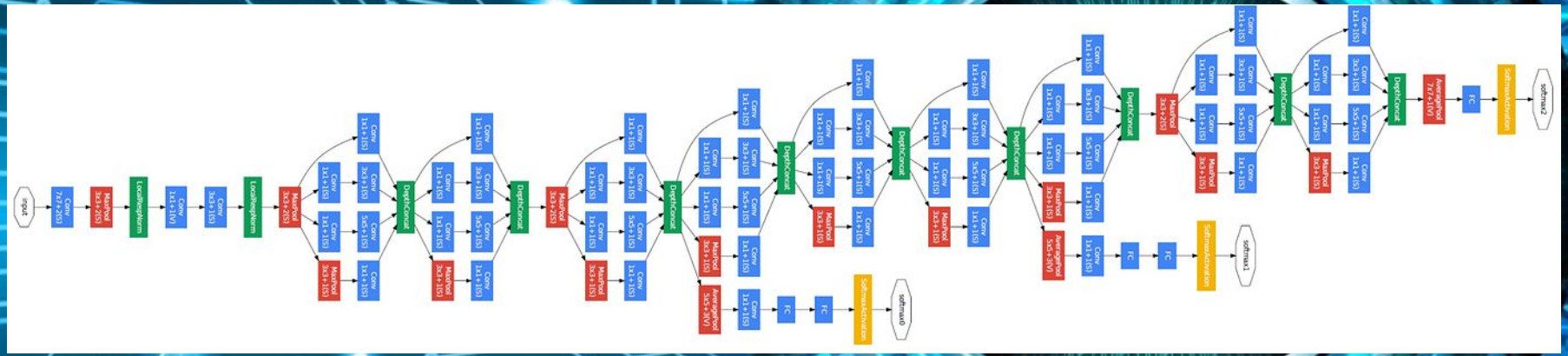
GoogleNet

2014 : GoogleNet est classé 1er au challenge ImageNet (ILSVRC)

Going Deeper with Convolutions (2014)



GoogleNet

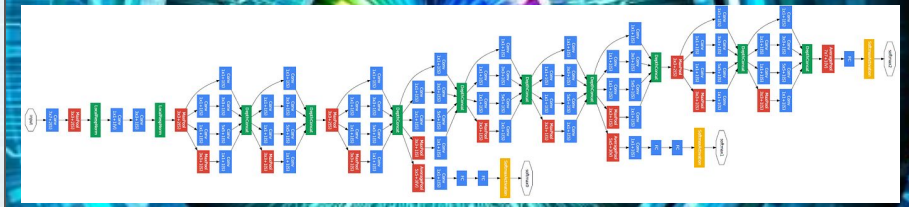

input
Conv 7x7+2(S)
MaxPool 3x3+2(S)
LocalResponseNorm
Conv 1x1+1(V)
Conv 3x3+1(S)
LocalResponseNorm
MaxPool 3x3+2(S)
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
MaxPool 3x3+1(S)
Conv 1x1+1(S)
DepthConcat
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
MaxPool 3x3+1(S)
Conv 1x1+1(S)
DepthConcat
MaxPool 3x3+2(S)
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
MaxPool 3x3+1(S)
Conv 1x1+1(S)
DepthConcat
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
MaxPool 3x3+1(S)
Conv 1x1+1(S)
DepthConcat
AveragePool 5x5+3(V)
Conv 1x1+1(S)
FC
FC
SoftmaxActivation
softmax0
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
MaxPool 3x3+1(S)
Conv 1x1+1(S)
DepthConcat
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
MaxPool 3x3+1(S)
Conv 1x1+1(S)
DepthConcat
AveragePool 5x5+3(V)
Conv 1x1+1(S)
FC
FC
SoftmaxActivation
softmax1
MaxPool 3x3+2(S)
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
MaxPool 3x3+1(S)
Conv 1x1+1(S)
DepthConcat
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
MaxPool 3x3+1(S)
Conv 1x1+1(S)
DepthConcat
AveragePool 7x7+1(V)
FC
SoftmaxActivation
softmax2

GoogleNet

Modules inception :

couches convolutives parallèles avec des filtres de différentes tailles dans chaque module

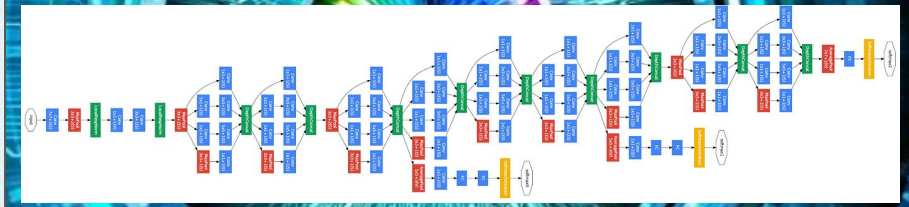
Going Deeper with Convolutions (2014)



GoogleNet

Global average pooling :

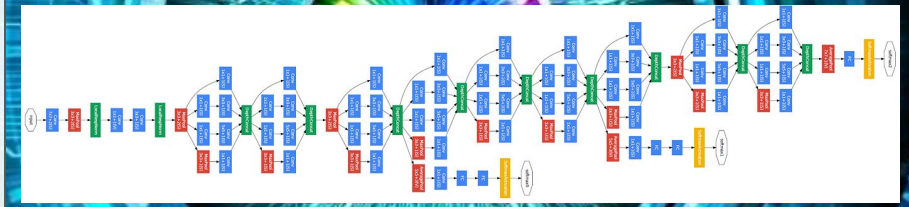
réduit l’overfitting
améliore l’invariance spatiale



GoogleNet

Classificateurs auxiliaires :

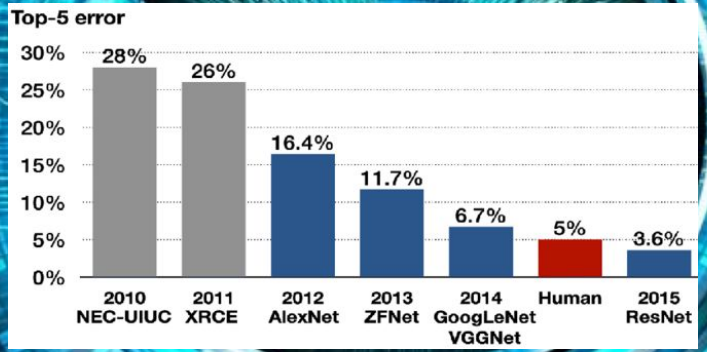
Classificateurs auxiliaires dans les couches intermédiaires réduit le vanishing gradient



ResNet

2015 : ResNet est classé 1er au challenge ImageNet (ILSVRC)

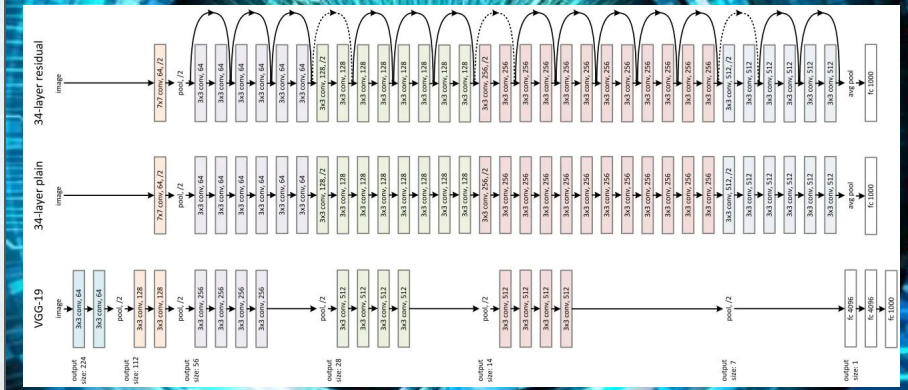
Deep Residual Learning for Image Recognition (2016)



ResNet

Blocs résiduels :

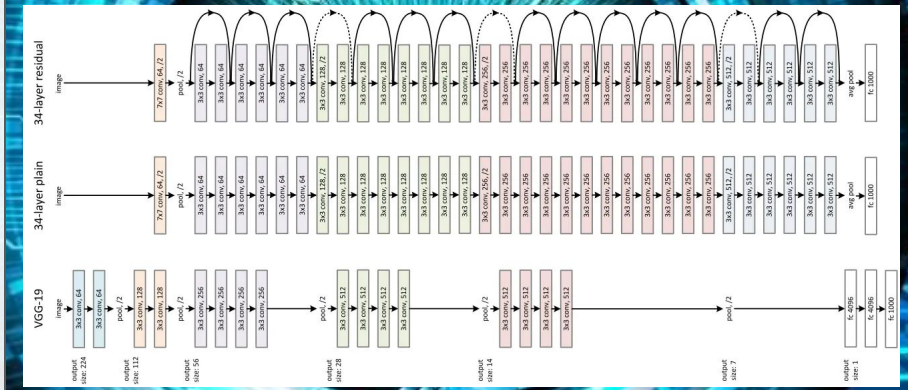
Skip connexions pour améliorer l'apprentissage des fonctions résiduelles



ResNet

Solution au gradient vanishing :

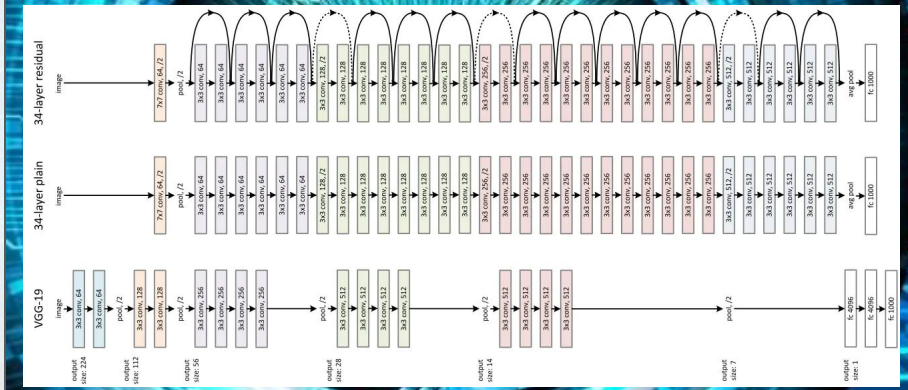
Les skip connections facilitent la conservation du gradient lors de la backpropagation



ResNet

Global average pooling :

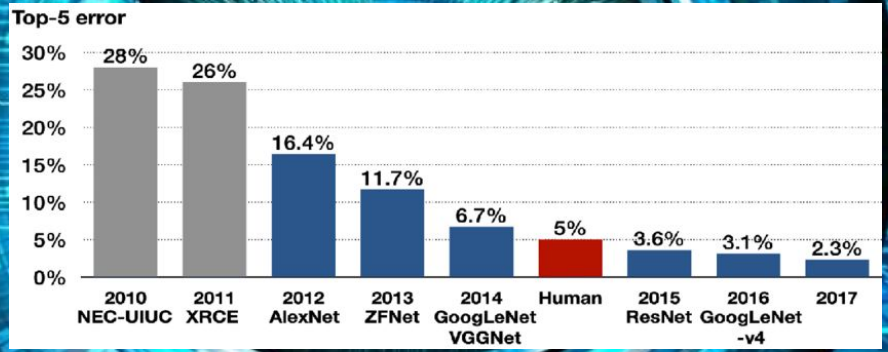
régularise contre l’overfitting
améliore l'invariance spatiale



Future evolution

Apprentissage de l'architecture, portabilité, …

Attention (2017)



Merci !