

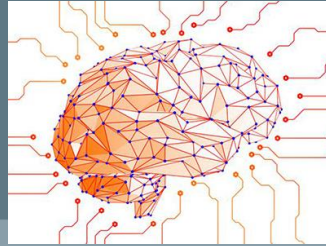


Les réseaux neuronaux

Quelques bonnes idées pour démarrer, comprendre un CNN



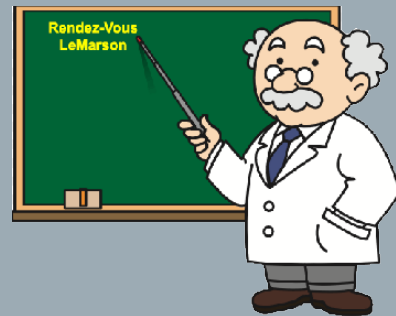
Émission animée
par Claude Marson



Les réseaux neuronaux

Le sommaire aujourd'hui

- ❖ C'est quoi l'Intelligence Artificielle ?
- ❖ Les réseaux neuronaux et la science du tâtonnement
- ❖ Deep Learning
- ❖ L'exemple d'un réseau convolutif



Les réseaux neuronaux

C'est quoi l'Intelligence Artificielle ?

L'informatique est une technique, l'Intelligence Artificielle est une science



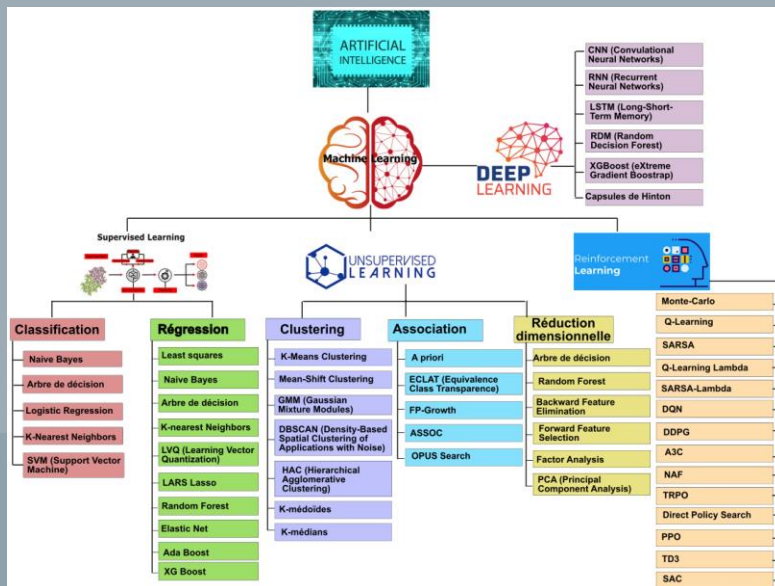
Définition

- Marvin Minsky : "La construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant accomplies de manière plus satisfaisante par des êtres humains, car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique".
- Distinction entre IA forte et IA faible



Les réseaux neuronaux

Positionnement des technologies IA

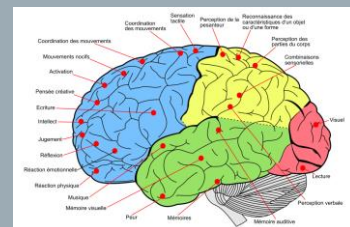
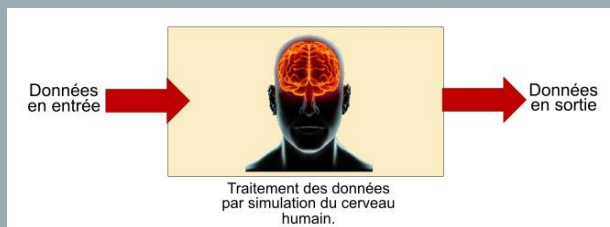


Les réseaux neuronaux

La science du tâtonnement

Les réseaux neuronaux

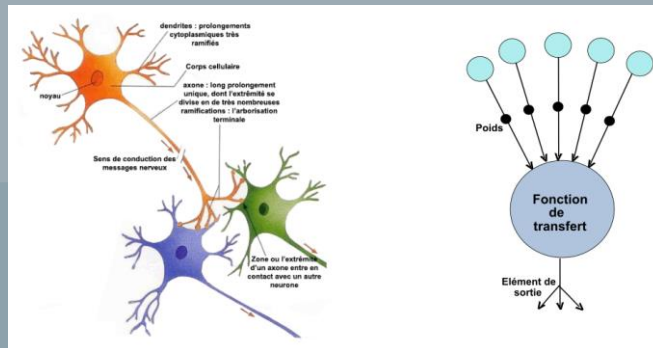
Principes de base d'un réseau neuronal



- Un réseau neuronal est un algorithme qui prend des données en entrée, qu'il traite, pour produire des résultats en sortie
- L'algorithme est fondé sur une simulation (lointaine) du fonctionnement d'un cerveau humain, avec ses neurones, synapses et dendrites
- Il pourra s'agir d'un traitement très simple, avec une seule couche (perceptron) ou plusieurs couches, avec propagation des résultats intermédiaires d'une couche à l'autre, voire de très nombreuses couches (deep learning)

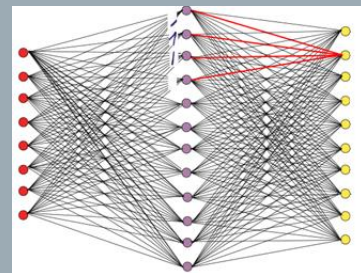
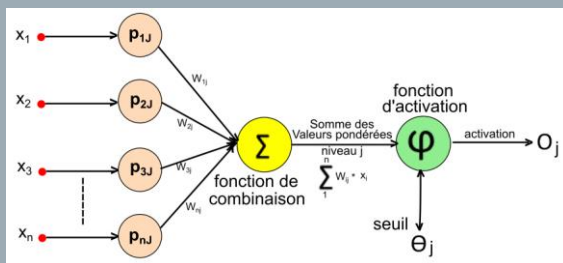
Les réseaux neuronaux

- Les réseaux neuronaux sont des plates-formes matérielles, sur lesquelles sont implémentées des algorithmes, qui simulent le comportement d'un cerveau humain (neurones et synapses), par un système de couches de processeurs (neurones), reliées par des connexions réseaux (synapses et dendrites).
- Un neurone biologique reçoit des signaux transmis par d'autres neurones (interaction dendrites-synapses), les stimuli.
- Le neurone analyse ces signaux en les sommant. Si le résultat obtenu est supérieur au seuil d'activation (excitabilité), il envoie un signal (potentiel d'action) le long de son axone vers d'autres neurones biologiques.
- Ce signal peut renforcer ou diminuer l'activité des neurones qui le reçoivent, selon que les synapses sont excitatrices ou inhibitrices.
- Un neurone biologique est connecté à plusieurs milliers d'autres neurones.
- Un neurone formel ou artificiel simule le fonctionnement d'un neurone biologique :
 - Les synapses sont modélisées par des poids
 - Le corps cellulaire est modélisé par une fonction de transfert, dite fonction d'activation
 - L'axone est l'élément de sortie.



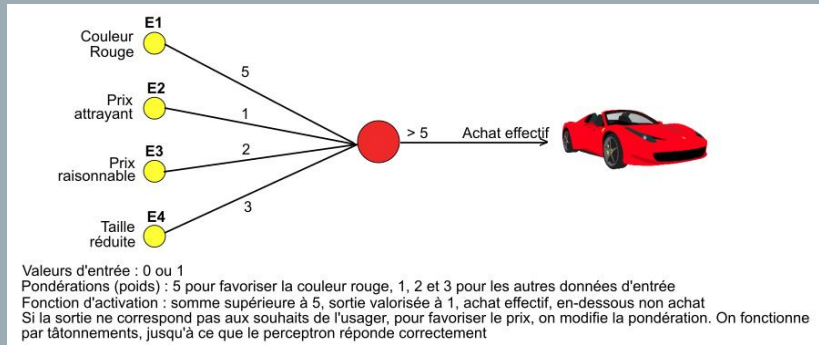
Les réseaux neuronaux

Les principes (de base) d'un réseau neuronal



- Les poids sont notés P_{ij} , qui vont du neurone i au neurone j : on voit bien que chaque poids témoigne de l'importance de l'entrée sur le résultat
- Chaque poids pondère le stimulus émis par le neurone source X_i
- Le neurone j fait la somme des stimuli pondérés : $\sum (P_{ij} * X_i)$
- S'il y a n neurones sources i liés au neurone j , celui-ci va calculer : $\sum (P_{ij} * X_i)$, avec i de 1 à n
- C'est cette valeur que le neurone j va traiter par une fonction de transfert
- La fonction de transfert (activation, seuillage, combinaison) est essentielle : c'est elle qui va déclencher un stimuli vers le niveau $j+1$ si la somme des pondérée niveau j est supérieure au seuil d'excitation
- Il en existe un grand nombre : seuil, linéaire, linéaire saturée, linéaire positive, sigmoïde, tangente hyperbolique...
- Les neurones formels (artificiels) organisés en couches successives deviennent des réseaux de neurones

Le principe de la pondération



- La pondération permet de donner plus ou moins d'importance aux entrées incidentes
- Par tâtonnements ou de manière automatique, en fonction d'un modèle et des informations rétropropagées, on se rapproche de la cible par des itérations successives

Les réseaux neuronaux

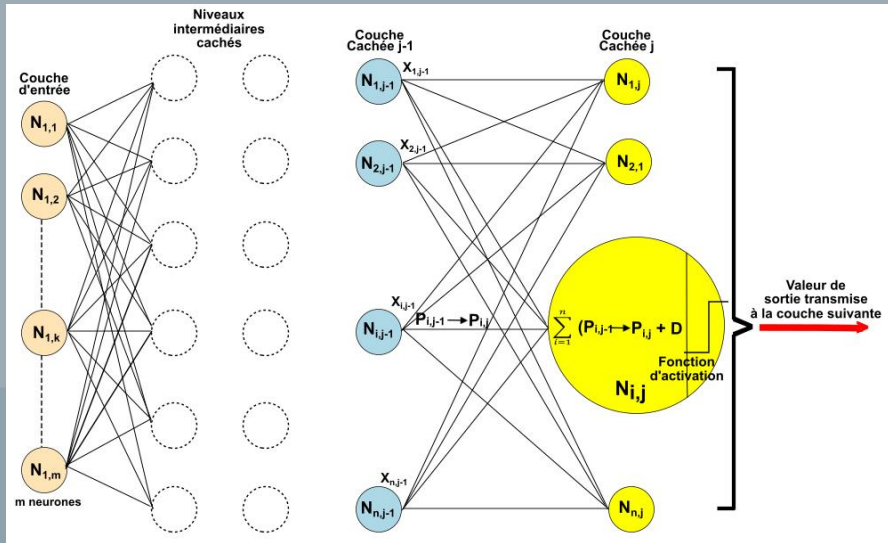
Les fonctions de transfert (activation)

- Chaque neurone applique une fonction de transfert, qui est la même pour tous les neurones appartenant à une même couche.
- Cette fonction mathématique est très importante, qui permet de faire passer l'information transmise par une couche i à une couche $i+1$, de façon discriminée pour chacun des neurones, en appliquant une transformation bien précise, qui dépendra de la nature du traitement que l'on voudra effectuer et de la finalité du réseau de neurones.
- On transforme une valeur en entrée en une valeur de sortie, qui elle-même va devenir la valeur d'entrée des neurones de niveau plus élevé.

<p>$f(x)=x$</p> <p>$f(x)=0$ si $x < 0$ $f(x)=1$ si $x \geq 0$</p> <p>$f(x)=1/(1 + e^{-x})$</p> <p>$f(x)=\tanh(x)$</p> <p>$f(x)=\arctan(x)$</p> <p>$f(x)=x/(1 + x)$</p> <p>$f(x)=0$ si $x < 0$ $f(x)=x$ si $x \geq 0$</p> <p>$f(x)=\alpha x$ si $x < 0$ $f(x)=x$ si $x \geq 0$</p>	<p>Fonction rampe</p> <p>Fonction marche</p> <p>Fonction sigmoïde</p> <p>Tangente hyperbolique</p> <p>Arctangente</p> <p>Fonction sigée</p> <p>Unité de rectification linéaire (ReLU)</p> <p>Unité de Rectification Linéaire Paramétrique (PReLU)</p>	<p>$f(x) = \alpha (e^x - 1)$ si $x < 0$ $f(x) = x$ si $x \geq 0$</p> <p>$f(x) = \log_e(1 + e^x)$</p> <p>$f(x) = ((\sqrt{x^2 + 1}) - 1)/2 + x$</p> <p>$f(\alpha, x) = -(\log_e(1 - \alpha(x + \alpha)))/\alpha$ si $\alpha < 0$ $f(\alpha, x) = x$ si $\alpha = 0$ $f(\alpha, x) = (e^{\alpha x} - 1)/\alpha + \alpha$ si $\alpha > 0$</p> <p>$f(x) = \sin(x)$</p> <p>$f(x) = 1$ si $x = 0$ $f(x) = \sin(x)/x$ si $x \neq 0$</p> <p>$f(x) = e^{-x^2}$</p>	<p>Unité Exponentielle Linéaire (ELU)</p> <p>Unité de Rectification Linéaire Douce (SoftPlus)</p> <p>Identité courbée</p> <p>Exponentielle douce paramétrique</p> <p>Sinus</p> <p>Sinus cardinal</p> <p>Fonction gaussienne</p>
---	---	---	--

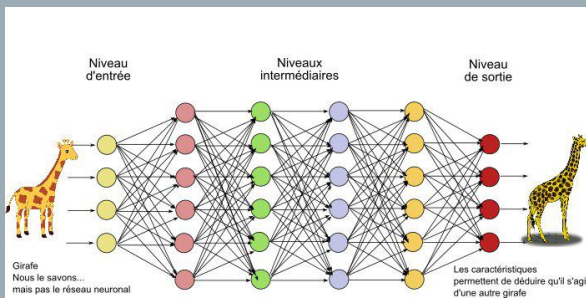
Les réseaux neuronaux

Concrètement...

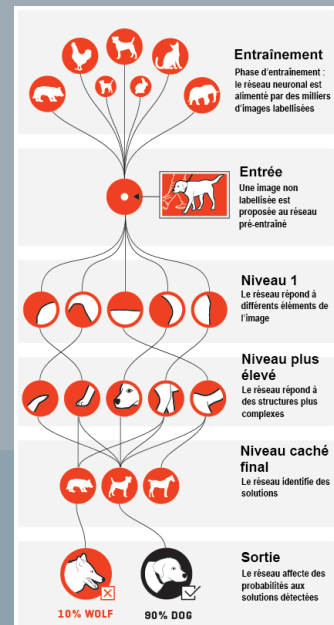


Les réseaux neuronaux

Deep Learning (apprentissage profond)



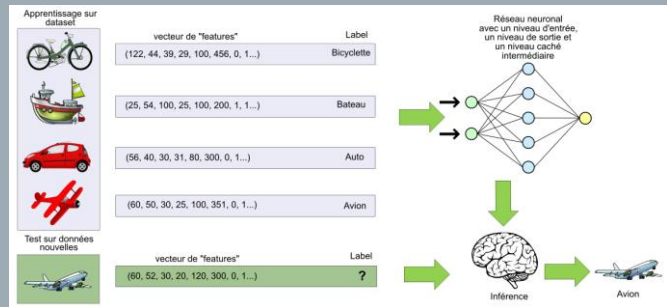
- Le « deep learning » ou apprentissage en profondeur, est un ensemble de méthodes qui bénéficient des progrès des réseaux neuronaux, pour reconnaître des objets, images, images dans les images, sons, phrases dans une communication téléphonique, etc.
- C'est une branche particulière du "Machine Learning", mais avec un grand nombre de couches "cachées", d'où l'idée de "deep" (100 à 200 couches contre quelques couches pour le "Machine Learning")
- Extraction automatique des caractéristiques d'entrée : l'algorithme analyse les données soumises
- Plusieurs différences avec le machine learning
 - Le gros volume de données traitées labellisées, pour l'entraînement
 - Grande puissance de calcul (GPU parallèles, clusters et Cloud...) : un apprentissage de plusieurs semaines peut être réduit à quelques heures ou moins
- L'un des réseaux profonds les plus répandus est le mode CNN ou ConvNet, réseaux de neurones à convolution, des réseaux multi-niveaux (on le verra plus loin)



Les réseaux neuronaux

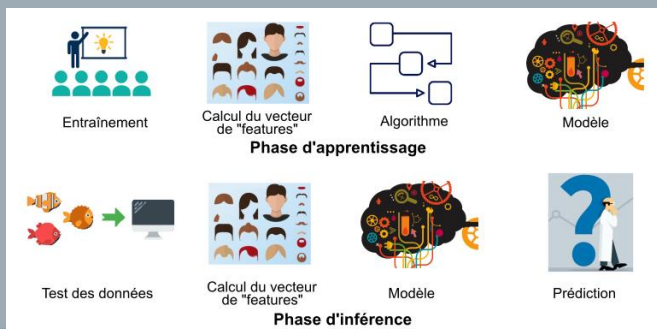
Deep learning

- Technique d'analyse des données qui consiste à apprendre à un algorithme à tirer parti de l'expérience des... humains
- Basée sur deux concepts : l'apprentissage et l'inférence
- Ce sont des algorithmes qui "apprennent" à partir de données qui leur sont soumises, dont ils déterminent les "features" (vecteur) en fonction du résultat souhaité
- Ils s'adaptent et deviennent de plus en plus performants au fur et à mesure que le nombre d'informations d'apprentissage augmente
- On parle d'apprentissage automatique, parce que l'algorithme apprend et se familiarise avec un concept de manière progressive : un objet, un visage, une langue, un comportement, un sentiment, une valorisation... Il établit un modèle : le vecteur.
- Si par la suite, on lui soumet une autre instance de ce concept, il sera capable de le reconnaître, avec ou sans aide. Toute la question étant de lui « apprendre » à effectuer cette reconnaissance.



Les réseaux neuronaux

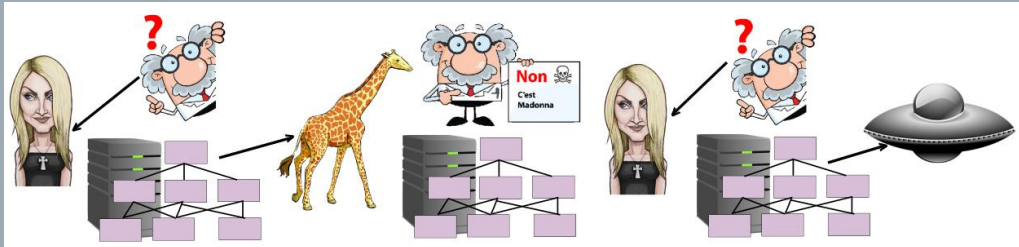
Deep Learning : chronologie des opérations Les deux phases : apprentissage et inférence



Le deep Learning par apprentissage revient à répéter une série d'opérations, jusqu'à ce que le résultat soit bon :

- Définir l'objet à reconnaître (objet au sens large)
- Collecter les données d'apprentissage (vecteur de "features")
- Lancer l'algorithme d'entraînement
- Lancer l'algorithme d'inférence : ce qui revient à tester le modèle sur de nouvelles données et à recalculer les features, puis à faire une prédiction.
- Collecter les données de retour : adéquation du modèle avec l'objet à reconnaître
- Raffiner l'algorithme d'inférence
- Recommencer jusqu'à ce que le résultat soit acceptable
- Utiliser le modèle pour effectuer des prédictions

La mécanique neuronale du deep learning



Toujours deux phases :

- Dans la première phase, le système cherche à appliquer une transformation non linéaire aux valeurs d'entrée et à créer un modèle de sortie, c'est la phase d'**éducation** ou d'**entraînement**, qui s'effectue sur des données labellisées
- La seconde phase, de **découverte**, consiste à améliorer le modèle avec une méthode mathématique (dite derivative) pour "coller" au plus près de la réalité. Elle s'effectue sur des données non labellisées

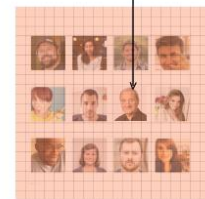
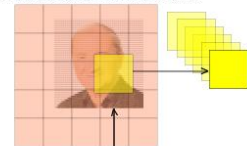


Les réseaux neuronaux

Focus sur les réseaux convolutifs

- CNN (Convolutional Neural Network) peut être considéré comme un empilement de perceptrons multicouches (*), chacun traitant une partie de l'information globale : reconnaissance d'images, vidéos, traitement naturel du langage.
- La convolution est un outil mathématique très utilisé en retouche d'image, car il en fait ressortir les caractéristiques (traits élémentaires...), mais aussi en interprétation climatique ou sismique
- L'image de départ est subdivisée en sous-images dont le traitement est assuré par un neurone artificiel
- Sur chaque sous-image, une fonction de convolution est appliquée à l'aide d'un filtre sur des échantillons de même taille qu'une matrice de filtrage, dont les valeurs évoluent avec l'apprentissage

L'image à reconnaître est divisée en échantillons (16 x 16, 32 x 32...) ou "tuile". Un neurone artificiel est dédié au traitement d'une tuile.



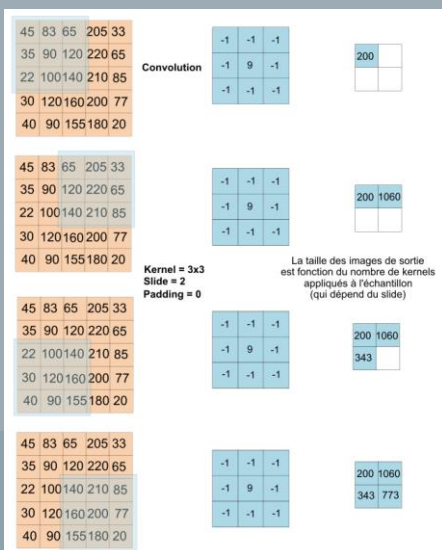
Imager d'origine à analyser

- ❖ L'échantillon est caractérisé par :
 - ❖ Le kernel, la dimension de la matrice de filtrage
 - ❖ Le "slide", le nombre de pixels de décalage entre deux sous-ensembles d'un même échantillon
 - ❖ Le "padding" est l'adjonction de valeurs nulles autour de l'image, pour éviter les dépassements dans le décalage du filtrage
- ❖ L'apprentissage du réseau convolutif, c'est-à-dire l'affinement de son comportement applicatif, se fera ensuite sur trois critères :
 - ❖ Les matrices de filtres qui caractérisent les convolutions, dont les valeurs vont changer pour s'adapter, tout en gardant le même kernel et algorithme
 - ❖ Le biais : valeur ajoutée après la convolution à chaque pixel (le même pour un filtre donné)
 - ❖ Le poids entre les couples de neurones

(*) : perceptron, le Réseau Neuronal Artificiel le plus simple

Les réseaux neuronaux

La fonction de convolution

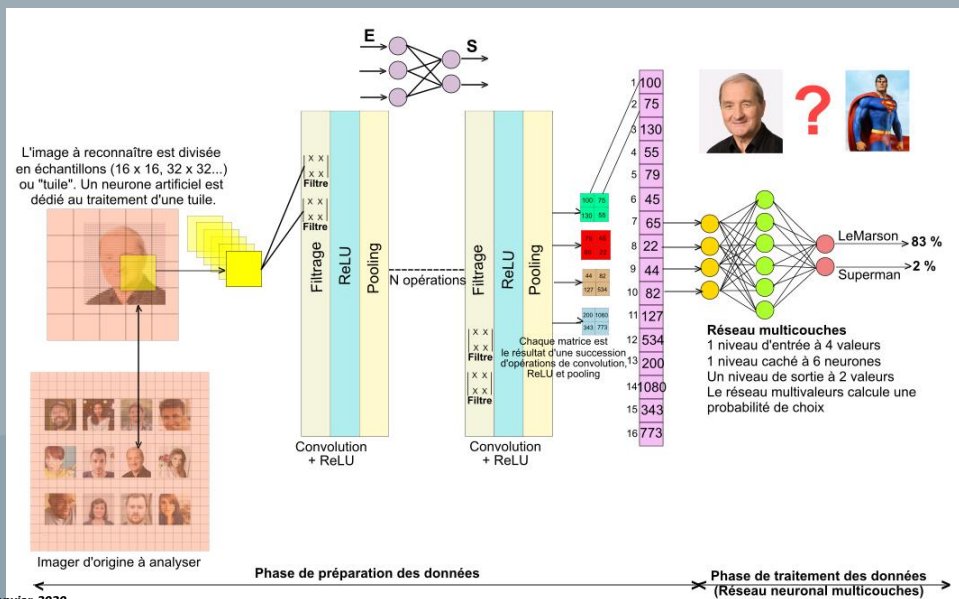


- ❖ La fonction de convolution s'applique aux sous-échantillons d'images
- ❖ Opération mathématique, qui va produire une matrice en sortie, généralement de taille réduite.
- ❖ Il existe plusieurs manières d'effectuer le calcul de convolution :
 - ❖ Convolution classique
 - ❖ La convolution "dilatée"
 - ❖ La convolution transposée
 - ❖ La "separable convolution"
- ❖ La convolution est associée à une fonction d'activation ReLU, qui se justifie pour augmenter les écarts entre deux caractéristiques d'images (suppression des valeurs négatives)
 - ❖ Pourquoi : les opérations de convolution sont linéaires (additions, multiplications...), or les images ne le sont pas, avec de gros écarts entre des pixels proches : il faut donc recréer cette linéarité
- ❖ Le "pooling" est une opération qui revient à remplacer un carré de pixels (2x2, 3x3) par une seule valeur, ce qui diminue la taille de l'image
- ❖ Les types de "pooling" :
 - ❖ "max pooling" : on prend la valeur la plus élevée de la sélection
 - ❖ "mean pooling" : on prend la valeur moyenne de toute la sélection
 - ❖ "sum pooling" : on prend la somme des pixels
- ❖ Le "flattening" (mise à plat) revient à mettre bout à bout les images pour en faire un vecteur

Les réseaux neuronaux

Le processus de convolution complet

EN DIRECT AVEC LEMARSON



LeMarson 31 Janvier 2020

19 / 20

Nos prochains rendez-vous

EN DIRECT AVEC LEMARSON

Lundi 3 février 2020 :
 Vendredi 7 février 2020 :
 Vendredi 14 février 2020 :
 Vendredi 21 février 2020 :
 Vendredi 28 février 2020 :

Actualités du TI
La programmation fonctionnelle
La fin du scandale des certificats payants ?
La justification financière des projets qualité des données
ITIL v4, vous avez du temps à perdre ?



Je vous remercie de votre attention et à bientôt

LeMarson 31 Janvier 2020

20 / 20