

Une introduction à l'apprentissage profond

Houcine Senoussi

14 janvier 2026

- 1 Introduction
- 2 Réseau convolutif (ConvNet)
 - Couche de convolution
 - Pooling
 - Couches full connected
 - Récapitulatif
- 3 Conclusion
- 4 Références

De quoi s'agit-il ?

- Profondeur d'un réseau neurone : nombre de couches.
- Une première définition : un réseau de neurones profond (**deep**) est un réseau ayant plus de couches qu'un réseau de neurones ordinaires (**shallow**).
- Question : à quoi servent ces couches supplémentaires ?
 - À représenter les données selon différents niveaux d'**abstraction**.
- Exemples de niveaux d'abstraction : une image peut être décrite par l'ensemble de ses pixels (niveau le plus bas), les groupes de pixels ayant une forme particulière, les arêtes et les objets (niveau(x) le(s) plus haut(s)).

- Exemple d'application : Une image représentant un visage humain :
 - Le visage se compose des yeux, du nez, etc ...
 - Chaque partie 'simple' du visage se compose de groupe de pixels ayant une forme particulière.
 - Chaque forme se compose de pixels.
- D'où le complément de la définition : Dans une architecture profonde, chaque niveau (couche) permet un traitement non-linéaire de l'information issue des niveaux plus bas (en termes d'abstraction), c'est-à-dire plus proches de la couche d'entrée. Ce traitement consiste à créer de nouveaux objets correspondant à un niveau d'abstraction plus élevé.

■ Motivation/Intérêt :

- Traiter de manière plus efficace de plus **grandes quantités** de données en utilisant les progrès dans la **puissance** des machines.
- Un nombre de couches réduit peut conduire à une explosion du nombre de neurones par couche.
- Les humains apprennent d'abord des concepts simples, et les combinent ensuite pour représenter des concepts plus abstraits.
 - Cela est notamment dû à l'architecture 'profonde' du cerveau. Ceci est particulièrement vrai pour le cortex visuel qui contient une série d'aires, chacune contenant une représentation différente du champ visuel.

Caractéristiques des réseaux profonds

- Si la première différence entre les réseaux profonds et les réseaux standards est le nombre de couches, cette différence n'est pas la seule car elle entraîne d'autres nouveautés qui sont résumées dans le tableau suivant :

Caractéristique	Réseaux standards	Réseaux profonds
Nombre de couches	Faible	Important
Types de couches	Unique	Plusieurs
Circulation de données	Feedforward	Possibilité de boucles
Connection des couches	Tot. connectées	différents types
Contenu des couches	1 tab. de neurones	1 ou plusieurs tab.
Fonctions d'activation		Nouvelle(s) fonction(s)
Apprentissage	Desc. du gradient	DdG + autres outils

- Plusieurs types de réseaux profonds existent. Parmi lesquels il y a notamment :
 - les réseaux **récurrents** : ce sont des réseaux dans lesquels la sortie d'un neurone peut être connectée à l'entrée d'un neurone appartenant à un niveau inférieur. Autrement dit, ce sont des réseaux contenant des **cycles**.
 - Remarque : on oppose réseaux récurrents aux réseaux 'acycliques' ou 'feed-forward' :
 - les réseaux **convolutifs** (ConvNet, pour Convolutionnal Neural Network) : ainsi appelés car l'une des opérations principales pour passer d'une couche à l'autre est la **convolution** (en l'occurrence entre un tableau à deux dimensions représentant une couche du réseau et un tableau plus petit appelé **filtre** (filter) ou **noyau** (kernel)).

- L'idée de départ des ConvNet est de prendre en compte la **structure spatiale** des données.
 - Le principal domaine d'application de ce type de réseau étant le traitement d'images, les neurones de la couche d'entrée sont disposés sous la forme d'un **tableau à deux dimensions**.
- L'opération principale qui relie deux couches successives est la **convolution**.
 - Cela donne lieu aux **couches de convolution** dont chaque neurone est connecté aux neurones d'un carré $t \times t$ de la couche précédente.
- Une autre opération importante est le **pooling**.
 - Une couche de pooling peut être considérée comme une copie sous-échantillonnée de la couche qui la précède.

- Le réseau se termine par une ou plusieurs couches entièrement connectées, telle qu'on en trouve dans les réseaux de neurones standards.
- D'autres types de couches peuvent exister dans le réseau (voir plus bas).

- Étant donné un tableau T de taille $n \times n$ et un tableau W $m \times m$, avec $m \leq n$, la convolution de T par W est le tableau $Conv$ défini comme suit :
 - $Conv$ est de taille $n - m + 1 \times n - m + 1$.
 - $Conv(i, j) = \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} W(k, l) * T(i + k, j + l)$.
- De manière plus visuelle, pour calculer $Conv(i, j)$:
 - Poser W au dessus de T de manière à faire coïncider $W(0, 0)$ avec $T(i, j)$.
 - Multiplier chaque élément de W par celui de T avec lequel il coïncide.
 - Additionner tous les produits.
- Le tableau W est appelé **filtre** ou **noyau** ou encore template.
- L'objectif de cette opération est en général détecter des parties de l'image T semblable au template W .

- Une couche de convolution est défini à l'aide de la convolution de la couche qui la précède par un filtre contenant des poids $w_{k,l}$.
- Plus exactement, en notant y la sortie des neurones d'une couche de convolution et x celle de la couche précédente, nous avons :
 - $y(i,j) = \sigma(b + \sum_{k=0}^{t-1} \sum_{l=0}^{t-1} w_{k,l} * x_{i+k,j+l})$
- avec :
 - b est le biais,
 - $w_{k,l}$ est le poids associé au neurone $(i+k, j+l)$
 - σ est une fonction d'activation (par exemple : la fonction sigmoïde).

- Il est important de noter que les coefficients w_{kl} et b sont communs à tous les neurones du tableau.
 - L'idée est donc de détecter le même motif à tous les emplacements du tableau contenant les entrées.
- En revanche, une couche de convolution ne contient pas un seul tableau de neurones mais plusieurs. Chacun de ces tableaux est défini à l'aide d'un filtre différent. Cela permet de détecter des motifs différents dans un même tableau d'entrée.
- Les couches de convolution ont donc besoin de définir les hyperparamètre suivants :
 - le nombre de filtres ainsi que la forme et la taille de chacun d'entre eux.

Pooling

- Les couches 'Pooling' arrivent en général à la suite des couches de convolution.
- Leur rôle est de '**simplifier**' / '**résumer**' l'information contenue dans ces dernières.
- Plus précisément, le pooling consiste à :
 - regrouper les neurones d'une couche de convolution par carré $p \times p$. Notons $Carr(i, j)$ le carré dont les neurones occupent les positions $\{(p * i, p * j), \dots, (p * (i + 1) - 1, p * (j + 1) - 1)\}$
 - La couche Pooling contient un tableau dont chaque neurone est connecté à un carré de la couche précédente et dont la valeur est définie par :
 - $P(i, j) = f(Carr(i, j))$.
 - f est la fonction qui sert à 'résumer' les neurones du carré $Carr(i, j)$. On prend $f = \max$. Dans ce cas nous parlerons de Max-Pooling. Les autres fonctions utilisées sont : la moyenne, la racine carrée de la somme des carrés (L2-Pooling).

- Les couches Pooling ont donc besoin de définir les hyperparamètres suivants :
 - taille p de Pooling (en général prise égale à 2) et fonction de pooling.

- La figure suivante montre un exemple de Max-pooling

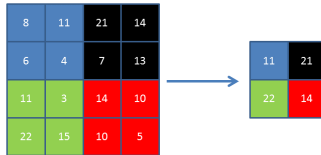


Figure – Pooling

Principaux types de couches-Récapitulatif

- La figure suivante résume la succession entrée-couche convolutionnelle-couche pooling

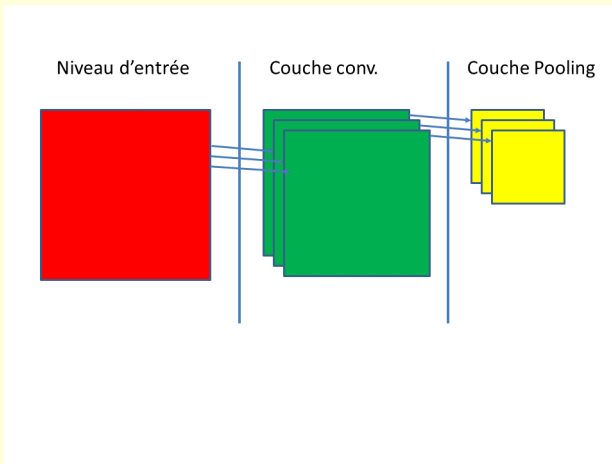


Figure – Succession des couches

- La sortie d'un réseau est assurée par une couche entièrement connectée (FC, fully connected) telle qu'on en trouve dans les réseaux de neurones standards.
 - Un neurone d'une telle couche est connectée à tous les neurones de la couche précédente (par exemple une couche Pooling).
- Comme nouveaux paramètres, nous avons besoin pour définir la couche de sortie des poids w et des biais b associés à chaque neurones de sortie.

- À première vue, un réseau ConvNet est composé d'une couche d'entrée, suivie d'une couche de convolution suivie d'une couche Pooling elle même suivie d'une couche convolution suivie d'une couche pooling ... et se termine par une couche de sortie FC.
- Une couche de convolution et une couche Pooling successives contiennent le même nombre de tableaux.
- Nous pouvons donc noter un réseau CNN de la manière suivante :
 - $Input \rightarrow (Cnv \rightarrow Pool) \times p1 \dots \rightarrow (Cnv \rightarrow Pool) \times pn \rightarrow FC$

- Au fil de leur développement, une nouvelle couche est venue s'ajouter au réseaux de convolution, c'est la couche **dropout**.
- Dans une couche dropout, les sorties de la couche précédente sont mises à 0 avec une probabilité p et copiées telles quelles avec une probabilité $1 - p$.
- Les neurones dont la sortie est mise à 0 ne participent ni à la phase feedforward ni à la rétropropagation.
- La couche dropout a comme objectif de réduire à la fois le **sur-apprentissage** et la **complexité** de l'apprentissage.

- ConvNet reprend les fonctions d'activation utilisées dans les réseaux standards mais y ajoute la fonction **ReLU** (pour Rectified Linear Units) dont l'objectif est de 'corriger' les sorties des neurones par exemple en prenant le max entre 0 et la valeur de la sortie.
- La fonction **ReLU** a été introduite à l'origine pour accélérer l'apprentissage.
- Il existe plusieurs variantes de la fonction **ReLU**.

- Dans les réseaux convolutifs, l'apprentissage se fait par la rétropropagation comme dans les réseaux standards.
- Le coût très élevé de la rétropropagation a conduit à introduire des outils qu'on appelle les optimiseurs (e.g. Adam) dont le rôle est de réduire le coût de cet apprentissage.
 - **Exercice :**
 - Recenser les paramètres qu'un réseau ConvNet doit apprendre.
 - Calculez le nombre de ces paramètres et comparez avec les réseaux standards.

- De la même manière que nous avons des couches de convolution de dimension 2, on peut en avoir de **dimension 1** utilisables, par exemple, en traitement du signal.
- Le cas des images en couleurs (*RGB*) conduit à utiliser des couches d'entrée à trois tableaux de neurones et des filtres de **dimension 3**.

- Prise en compte de la structure spatiale des données.
- Grâce au fait qu'un neurone est en général connecté à un petit nombre de neurones de la couche précédente et que les poids et les biais sont les mêmes dans chaque tableau, les ConvNet réduisent le nombre de paramètres à calculer (par rapport à celui d'un réseau ordinaire totalement connecté).

Définition d'un réseau convolutif

- Récapitulons. Pour définir entièrement un réseau CNN nous devons :
 - Définir la suite de couches avec pour chacune son type et ses caractéristiques.
 - Choisir le nombre de filtres et leurs formes pour chaque couche de convolution. Ces choix dépendent du type d'images à traiter.
 - Exemple : pour la reconnaissance des chiffres manuscrits, on a l'habitude de prendre des filtres de taille 5×5
 - Choisir la taille des carrés ainsi que la fonction à utiliser dans chaque couche Pooling.
 - Choisir le paramètre p de la (ou des) couches dropout éventuelle(s).
 - Appliquer la descente du gradient pour l'apprentissage des différents coefficients et biais, comme pour les réseaux de neurones ordinaires. Il y a souvent besoin, avant de lancer l'apprentissage de choisir un optimiseur.

- Nous avons introduit l'apprentissage profond à travers l'un des réseaux les plus utilisés : Le réseau de neurones convolutifs.
- Pour résumer, on peut retenir les caractéristiques & avantages suivantes des réseaux convolutifs :
 - Existence de plusieurs **types** de couches dont le principal est celui des couches de **convolution**.
 - Connexions **locales** (entre couches successives) et poids **partagés**, et comme conséquence la **réduction** du nombre de paramètres.

- M.A. Nielsen, "Neural networks and deep learning". Determination Press, 2015.
- Cours du CNAM (RCP209)
- Wikipedia.
- Livre : Deep learning methods and applications , Li Dend and Don Yu.
- Divers articles, notamment ceux de Yann Lecun.