

———— CHAPITRE 2 ————

APPRENTISSAGE PROFOND

[MCours.com](https://www.m-cours.com)

2.1 INTRODUCTION

Dans ce chapitre, nous présenterons l'apprentissage profond et les réseaux de neurones profonds, c'est-à-dire des réseaux de neurones avec plusieurs couches cachées. Ils sont capables d'apprendre des abstractions de caractéristiques des exemples d'entrée, de comprendre les caractéristiques de base des exemples et de faire des prédictions basées sur ces caractéristiques.

2.2 APPRENTISSAGE PROFOND

Selon les fondateurs Yann LeCun, Yoshua Bengio Geoffrey Hinton : "L'apprentissage profond permet aux modèles informatiques, composés de plusieurs couches de traitement, d'apprendre des représentations de données avec plusieurs niveaux d'abstraction."[\[11\]](#)

Une autre définition des auteurs : "L'apprentissage profond est une classe de techniques d'apprentissage automatique, où l'information est traitée en couches hiérarchiques pour comprendre les représentations et les caractéristiques des données dans des niveaux de complexité croissants."[\[12\]](#)

En d'autres termes, Deep Learning est un sous-ensemble des méthodologies et techniques de Machine Learning (ML) qui utilisent le réseau neuronal artificiel. C'est l'adaptation des réseaux neuronaux qui imite la structure du cerveau humain [\[12\]](#).

La figure (2.1) illustre le contexte du DL par rapport à l'IA.

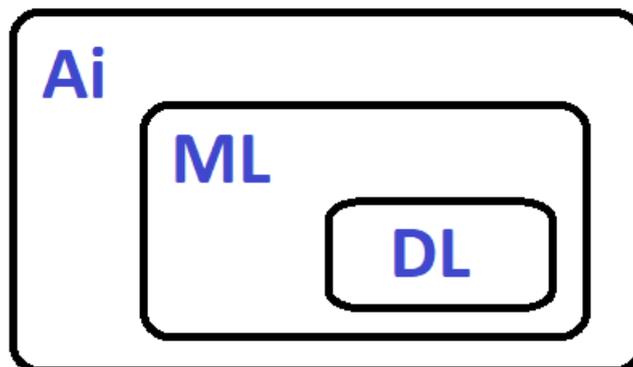


FIGURE 2.1: Contexte de l'apprentissage profond

2.2.1 Comparaison entre l'apprentissage profond et l'apprentissage automatique

Dans cette section nous allons présenter la différence majeure entre le DL et ML, qui est l'intervention humaine. Les systèmes d'apprentissage automatique (Machine Learning) ont besoin de l'intervention humaine pour identifier et coder manuellement les fonctionnalités et les caractéristiques, en fonction du type de données. Par contre, un système de l'apprentissage profond (Deep Learning) tente d'apprendre ces fonctionnalités, sans une intervention humaine supplémentaire. Les quantités de données impliquées dans ce processus sont énormes. Au fur et à mesure que l'entraînement s'enchaîne, la probabilité des réponses correctes augmente [13]. La figure (2.2) illustre cette différence.

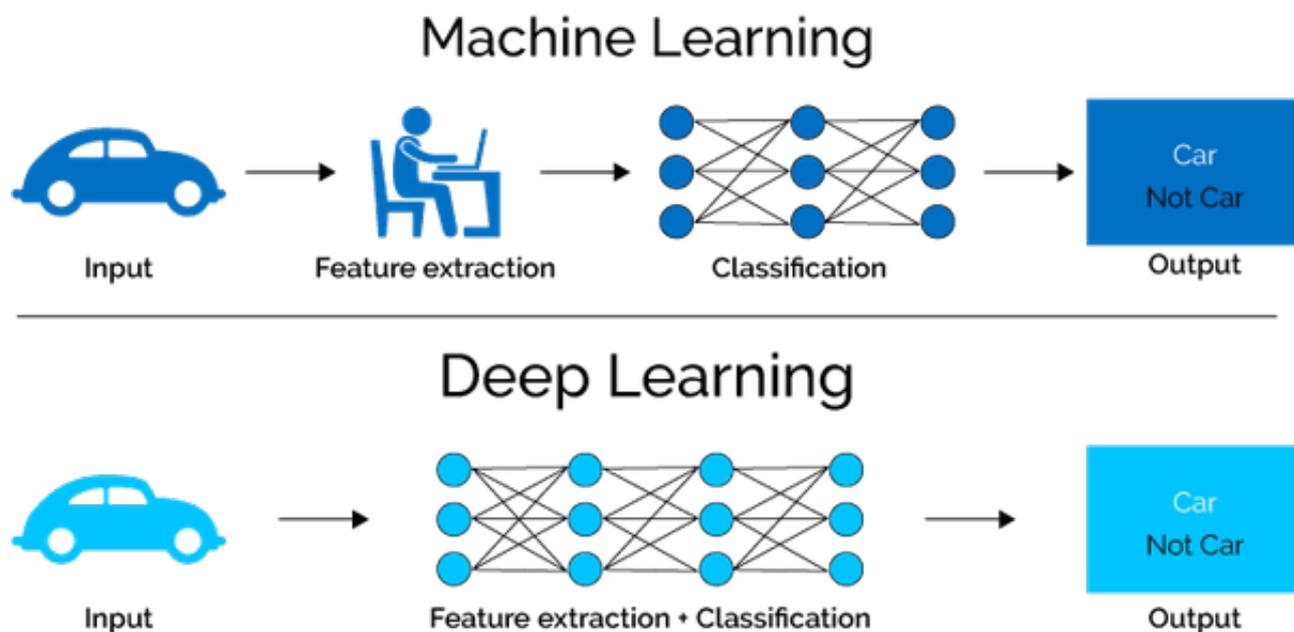


FIGURE 2.2: Comparaison entre l'apprentissage profond et l'apprentissage automatique [14]

2.2.2 Historique de l'apprentissage profond

Dans [15], se trouve un aperçu du contexte historique de l'apprentissage profond, il présente les étapes majeures qui ont mené à ce que nous avons maintenant. Ces étapes sont résumées dans le tableau (2.1) :

Année	Contributeur	Contribution
300 AC	Aristotle	Introduction de l'associationnisme, qui essayent de comprendre le cerveau
1873	Alexander Bain	Introduction des premiers modèles de réseaux de neurones
1943	McCulloch and Pitts	Proposition d'un modèle considéré comme L'ancêtre des réseaux de neurones
1949	Donald Hebb	il introduit la règle d'apprentissage de Hebb
1958	Frank Rosenblatt	Introduction du premier perceptron
1974	Paul Werbos	Introduction de la retro-propagation
1980	Teuvo Kohonen	Introduction des cartes auto organisatrices
1980	Kunihiko Fukushima	Introduction du Neocognitron,
1982	John Hopfield	Introduction des réseaux de Hopfield
1985	Hilton and Sejnowski	Introduction des machines de Boltzmann
1986	Paul Smolensky	Introduction de Harmonium
1986	Michael I. Jordan	Définition et introduction des réseaux de neurones récurrents
1990	Yann LeCun	Introduction de LeNet
1997	Schuster and Paliwal	Introduction des réseaux de neurones récurrents bidirectionnels
1997	Hochreiter and Schmidhuber	Introduction de LSTM,
2006	Geoffrey Hinton	Introduction des Deep belief Network
2009	Salakhutdinov and Hinton	Introduction des Deep Boltzmann Machines
2012	Alex Krizhevsky	Introduction d'AlexNet qui remporta le challenge ImageNet

TABLE 2.1: Historique de l'apprentissage profond [15]

2.3 UTILITÉ DE L'APPRENTISSAGE PROFOND

Dans cette section, nous allons présenter les principaux points qui montrent l'importance du Deep Learning et l'utilité par rapport aux autres techniques de l'apprentissage automatique.[16]

- Les fonctionnalités conçues manuellement, sont souvent sur-spécifiées, incomplètes et prennent beaucoup de temps à concevoir et à valider.
- Les fonctionnalités apprises sont faciles à adapter, rapides à apprendre.
- L'apprentissage profond fournit un cadre très souple, pour représenter des informations.
- Peut apprendre à la fois sans surveillance et sous surveillance.

2.4 APPLICATION DE L'APPRENTISSAGE PROFOND

Parmi les nombreuses applications de l'apprentissage profond, nous pouvons citer [16] :

- **La reconnaissance faciale** : un algorithme de DL va apprendre à détecter les caractéristiques d'un visage sur une photo. Il s'agit en premier lieu, de donner un certain nombre d'images à l'algorithme, puis à force d'entraînement, l'algorithme va être en mesure de détecter un visage sur une image.
- **Voitures autonomes** : les entreprises qui construisent de tels types de services d'aide à la conduite, ainsi que des voitures autonomes telles que Google, doivent apprendre à un ordinateur à maîtriser certaines parties essentielles de la conduite, à l'aide de systèmes de capteurs numériques au lieu de l'esprit humain.
- **Reconnaissance d'image** : Un autre domaine populaire en matière de DL est la reconnaissance d'images. Son objectif, est de reconnaître et d'identifier les personnes et les objets dans les images, ainsi que de comprendre le contenu et le contexte.
- **Génération automatique de texte** : C'est une tâche intéressante, où un corpus de texte est appris et à partir de ce modèle, un nouveau texte est généré, mot par mot ou caractère par caractère.

2.5 TYPES DES RÉSEAUX DE NEURONES PROFONDS

Les types des réseaux de neurones profonds les plus utilisés, dans le procédé de l'apprentissage profond, sont les CNN :

2.5.1 Réseaux de neurones Convolutionnels CNN

Les réseaux de neurones convolutionnels sont parmi les types les plus utilisés, pour imiter le comportement du cerveau humain. Ils sont inspirés par les travaux de Hubel et Wiesel sur le cortex visuel chez les mammifères.[17]

Pendant les années 1990, ces réseaux se sont popularisés avec les travaux de Y. LeCun et al. sur la reconnaissance de caractères.[18]

2.5.1.1 Les types de couches dans les CNN :

Nous allons nous focaliser sur les couches les plus classiques : la couche de convolution (Conv), la couche entièrement connectée (FC) et la couche d'union (pooling).

La couche de convolution (Conv)

Elle est utilisée pour la détection de motifs avec une invariance en translation, ensemble de filtres qui sont appris au cours de l'entraînement. La taille et le nombre de ces filtres sont définis à priori, ils parcourent la totalité de l'image, le plus souvent avec une fonction d'activation Relu. Elle représente la réponse de ce filtre à chaque position spatiale. Ces cartes d'activations sont concaténées le long de la dimension de l'image, pour former un nouveau tenseur. La figure (2.3) montre le traitement d'une couche de convolution ¹.

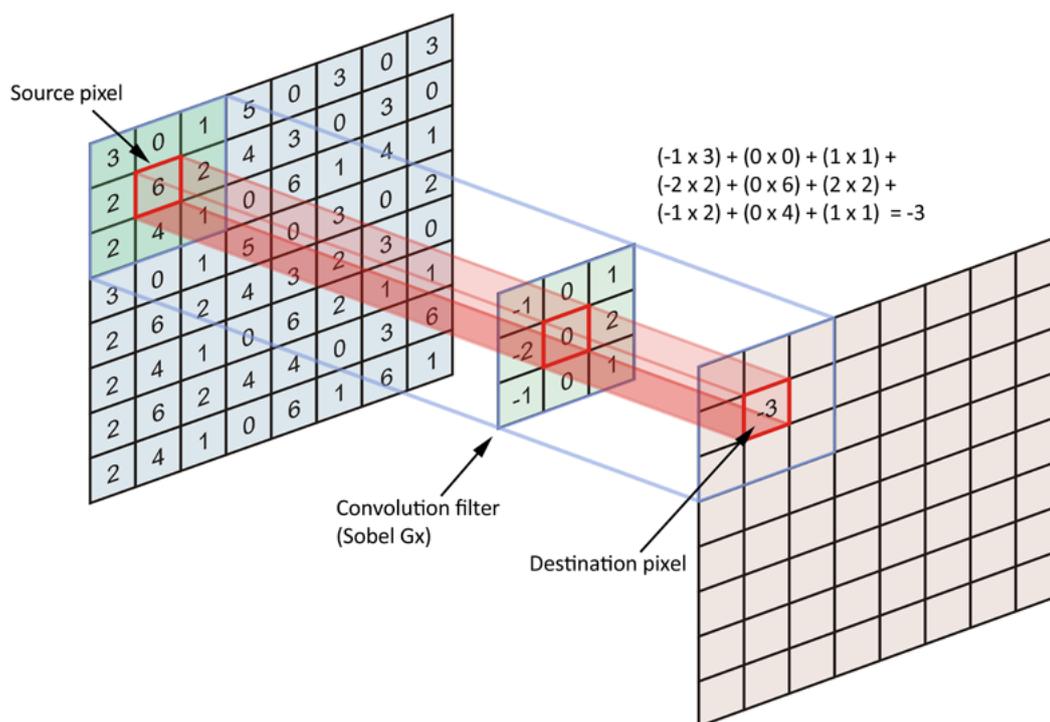


FIGURE 2.3: Couche de convolution

La couche d'union (Pooling)

Elle est utilisée pour réduire la détection des patterns ambigus, après une couche de convolution.

1. <https://thigiacmaytinh.com/wp-content/uploads/2018/05/kernel.png>

Ces fonctions sont prédéfinies et réduisent, pour les couches ultérieures, le nombre de paramètres à apprendre, tout en élargissant le champ de réception (receptive field) [19].

Elles opèrent indépendamment, des profondeurs du réseau et ne nécessitent pas de poids à entraîner.

L'idée derrière l'opération de "pooling" est de retirer une certaine information dans un même voisinage pour l'entraînement [20]. La figure (2.4) illustre ce procédé.

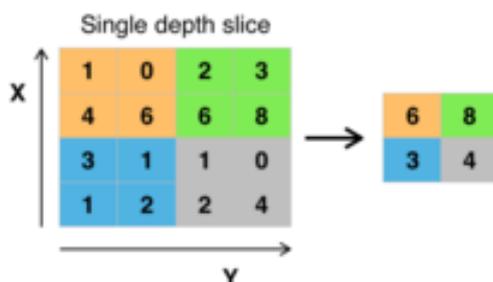


FIGURE 2.4: Couche de pooling [19]

La couche entièrement connectée (FC)

Elle est utilisée pour associer les différents motifs afin d'en déduire la classe. Ce type de couches reçoit un vecteur en entrée et produit un nouveau vecteur en sortie.

Pour cela, cette couche applique une combinaison linéaire puis, éventuellement, une fonction d'activation aux valeurs reçues en entrée.

2.5.1.2 Schéma Général d'un CNN :

Dans la figure (2.5) nous illustrons le schéma général d'un CNN, qui est constitué généralement d'un nombre de couches de convolution, certaines d'entre elles sont suivies par des couches de pooling, et finalement, par des couches entièrement connectées.

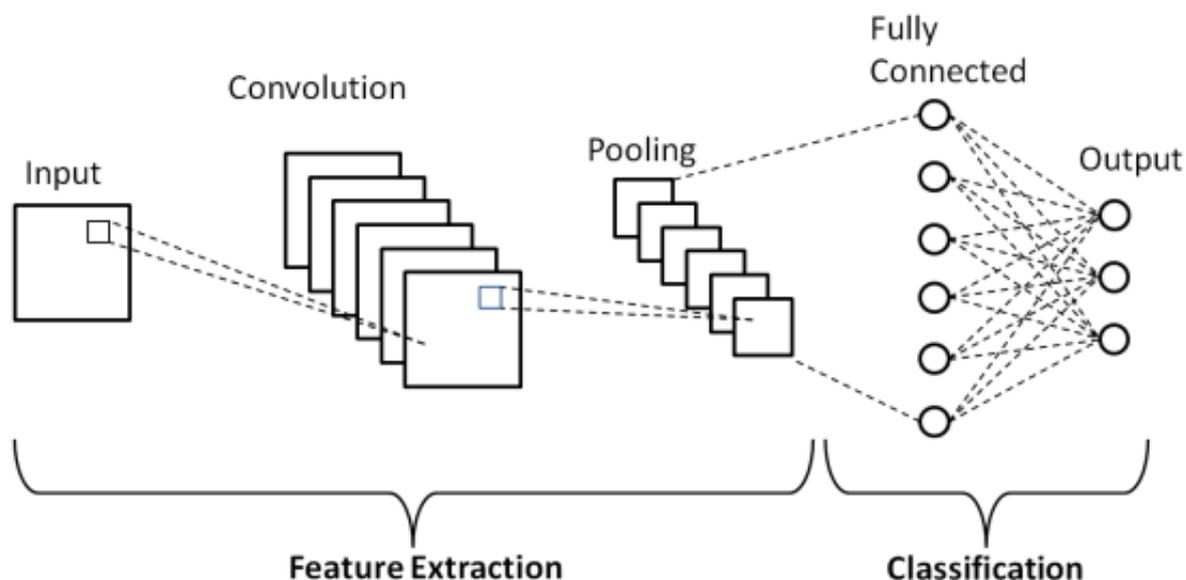


FIGURE 2.5: Schéma général d'un CNN [21]

2.6 L'ARCHITECTURE DES RÉSEAUX DE NEURONES CONVOLUTIONNELS (CNN)

Dans cette section, nous allons montrer les principales architectures des réseaux de neurones convolutionnels CNN :

2.6.1 LeNet

LeNet est une architecture de CNN classique proposée par Yann LeCun en 1998. Cette architecture a été conçue pour reconnaître les patterns visuels, directement à partir de l'image.

Elle permet de reconnaître les patterns tels que les caractères manuscrits avec une robustesse aux distorsions et aux transformations géométriques [22]. Le schéma de l'architecture LeNet est illustrée dans les figures (2.6) et (2.7).

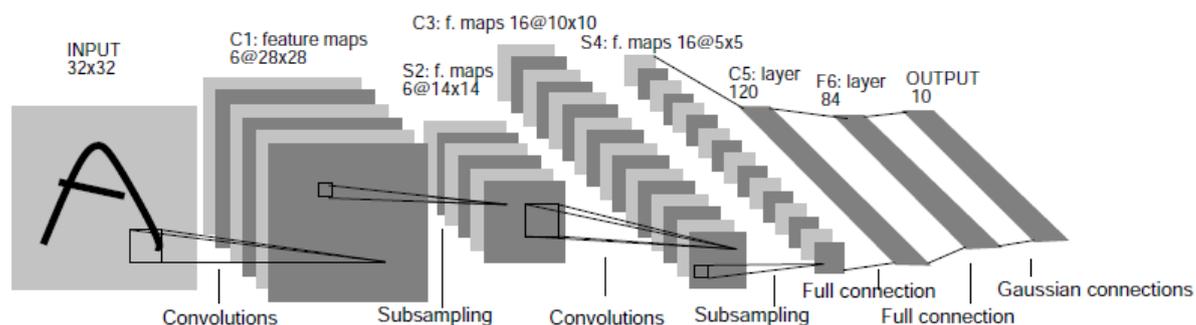


FIGURE 2.6: Schéma de l'architecture LeNet [22]

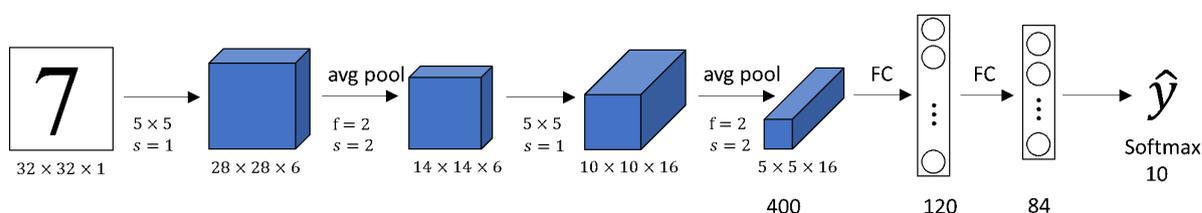


FIGURE 2.7: Schéma de l'architecture détaillée de LeNet [22]

2.6.2 AlexNet

AlexNet est un CNN conçu par Alex Krizhevsky publié avec Ilya Sutskever et son chef de doctorat Geoffrey en 2012.

Il contient huit couches : les cinq premières étaient des couches de convolution, certaines d'entre elles sont suivies par des couches de pooling, les trois dernières étaient des couches entièrement connectées (FC). Ces couches utilisent une fonction d'activation ReLu [23]. La figure (2.8) illustre le schéma d'AlexNet.

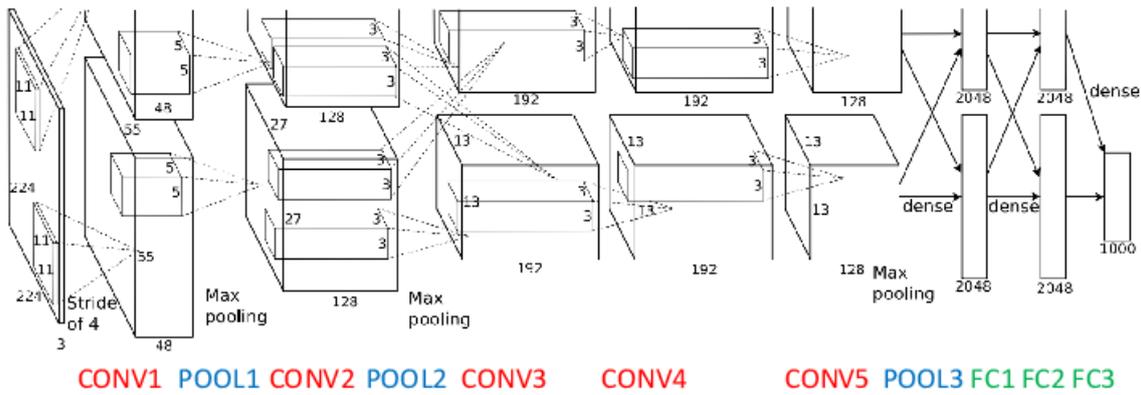


FIGURE 2.8: Schéma de l'architecture AlexNet [23]

AlexNet est considéré comme l'une des architectures les plus influentes publiées dans le domaine de la vision par ordinateur et de l'apprentissage profond.

2.6.3 VGGNet

VGGNet est un réseau neuronal convolutionnel de profondeur de 16 couches, présenté par les chercheurs du Visual Graphics Goup à l'université d'Oxford en 2014, il se caractérise par sa forme pyramidale [24]. La figure (2.9) montre cette architecture.

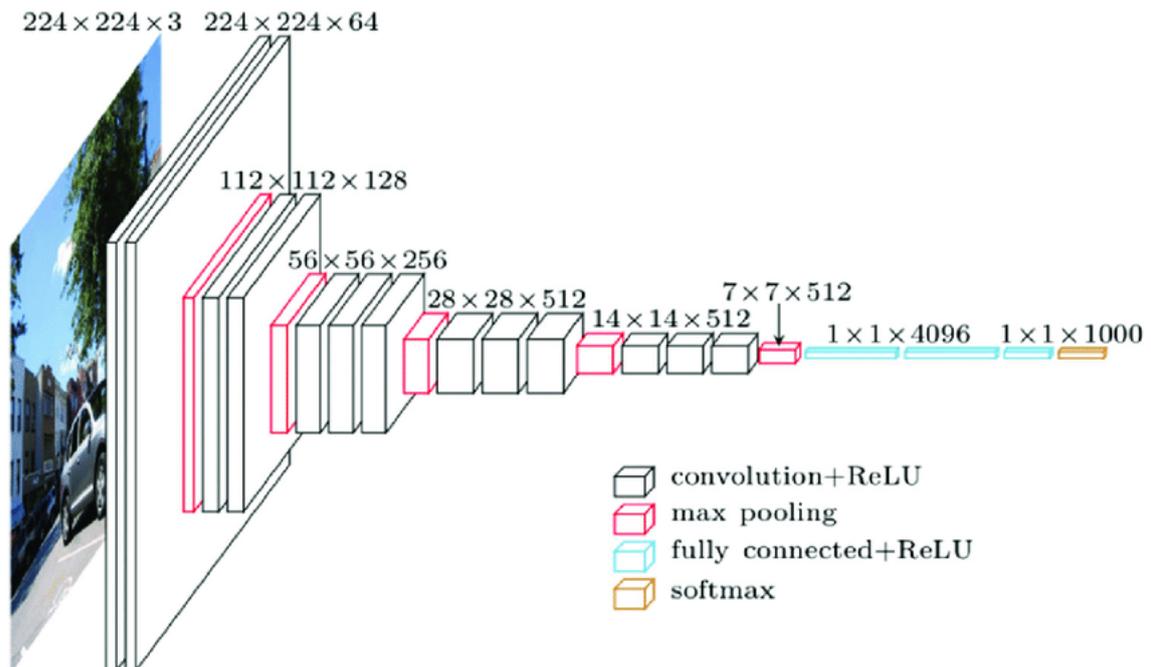


FIGURE 2.9: Schéma de l'architecture VGG [24]

2.6.4 ResNet

Le réseau de neurone Résiduel ResNet est un réseau qui utilise une architecture nouvelle avec « connexion par saut ». Ces connexions sont connues sous le nom d'unités récurrentes fermées [25].

Elles ressemblent fortement aux RNN, ces unités sont appelées « Residual Block » [26]. Le schéma de ces unités est présenté dans la figures (2.10)

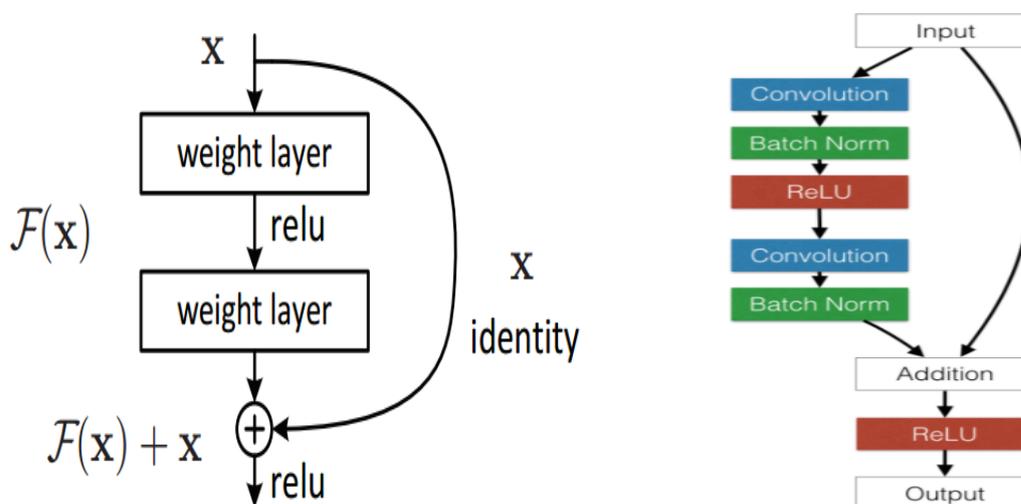


FIGURE 2.10: Schéma du "Residual Block" [25]

2.6.5 GoogleNet

GoogLeNet est le nom de la nouvelle architecture proposée par les chercheurs de Google, pour participer à la compétition ILSVRC en 2014. La contribution apportée par cette architecture, est une unité de base appelée « Inception Module » [27].

Ce modèle utilise « Average Pooling » au lieu des couches entièrement connectées. Cela réduit un grand nombre de paramètres comparé à AlexNet [28]. Les figures (2.11) et (2.12) montrent l'architecture du modèle "GoogLeNet" et de "Inception Module".

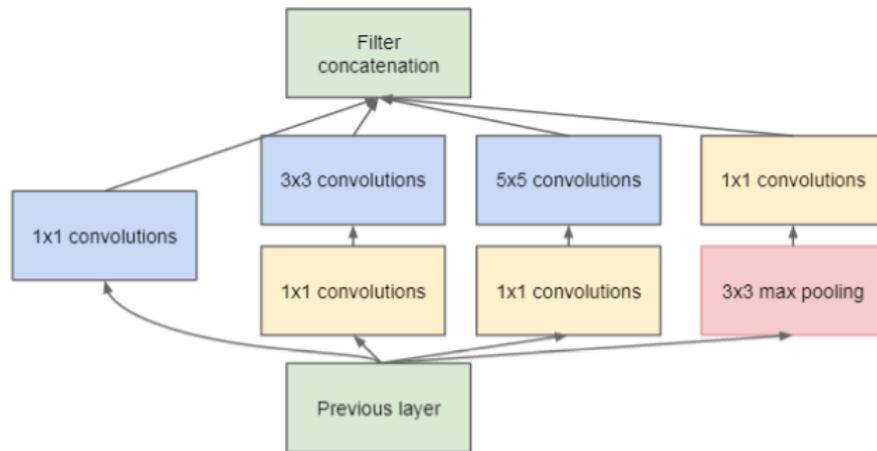


FIGURE 2.11: Schéma de "Inception Module" [27]

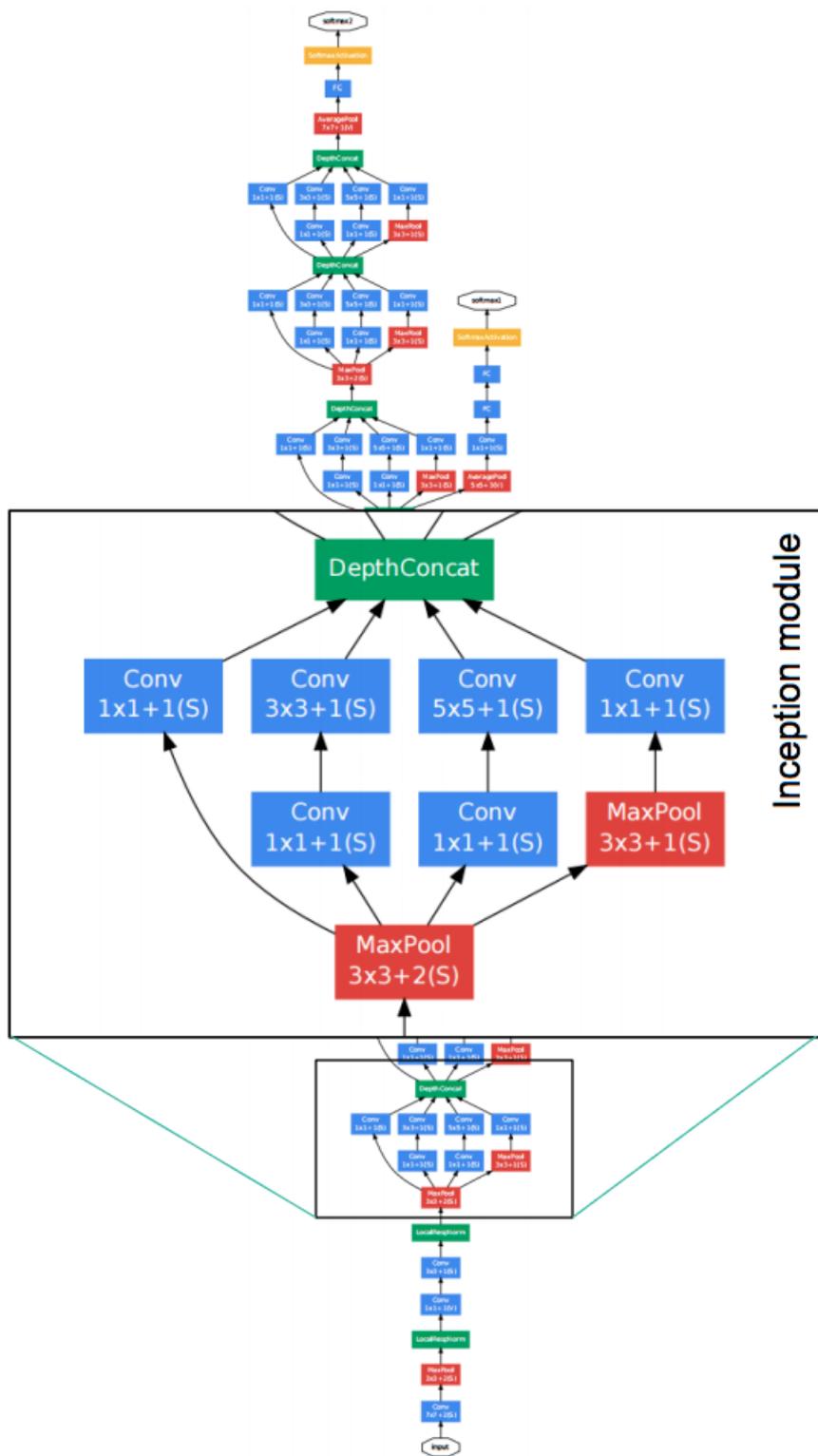


FIGURE 2.12: Schéma de l'architecture GoogleNet [28]

2.7 CONCLUSION

Dans ce chapitre nous avons présenté le domaine de l'apprentissage profond. Nous avons introduit puis expliqué les réseaux de neurones profonds (particulièrement les réseaux convolutionnels), qui seront essentiels pour la reconnaissance des chiffres nécessaire à notre problématique. Finalement, nous clôturons ce chapitre par l'illustration des différentes architectures des réseaux de neurones convolutionnels. Dans le chapitre qui suit, nous allons nous intéresser à l'utilisation des réseaux de neurones profonds dans le contexte d'une reconnaissance de caractères. Le cas des plaques minéralogiques a été choisi pour sa proximité avec notre problématique.

MCours.com