

Les réseaux de neurones

PRINCIPE

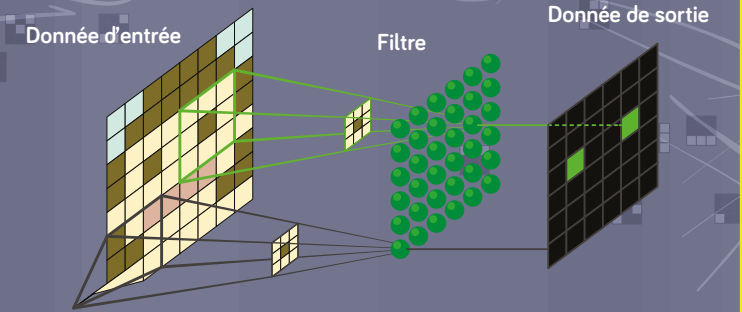


Apprendre à reconnaître un visage, sans savoir ce qu'est un visage ! C'est la prouesse de certains types de réseaux de neurones, grâce à des algorithmes d'apprentissage de plus en plus complexes.

Principe très basique : d'un côté, des neurones qui possèdent des entrées et une pondération associée (les paramètres) ; de l'autre, les données de l'image qui sont les valeurs des pixels. Les neurones effectuent des calculs entre les pondérations et les pixels, et restituent des résultats qui nourrissent les neurones des couches suivantes pour obtenir une donnée de sortie finale. Si celle-ci est correcte par rapport à la donnée d'entrée, le réseau a bien appris à reconnaître la donnée initiale. Si non, l'algorithme modifie les pondérations jusqu'au résultat souhaité.

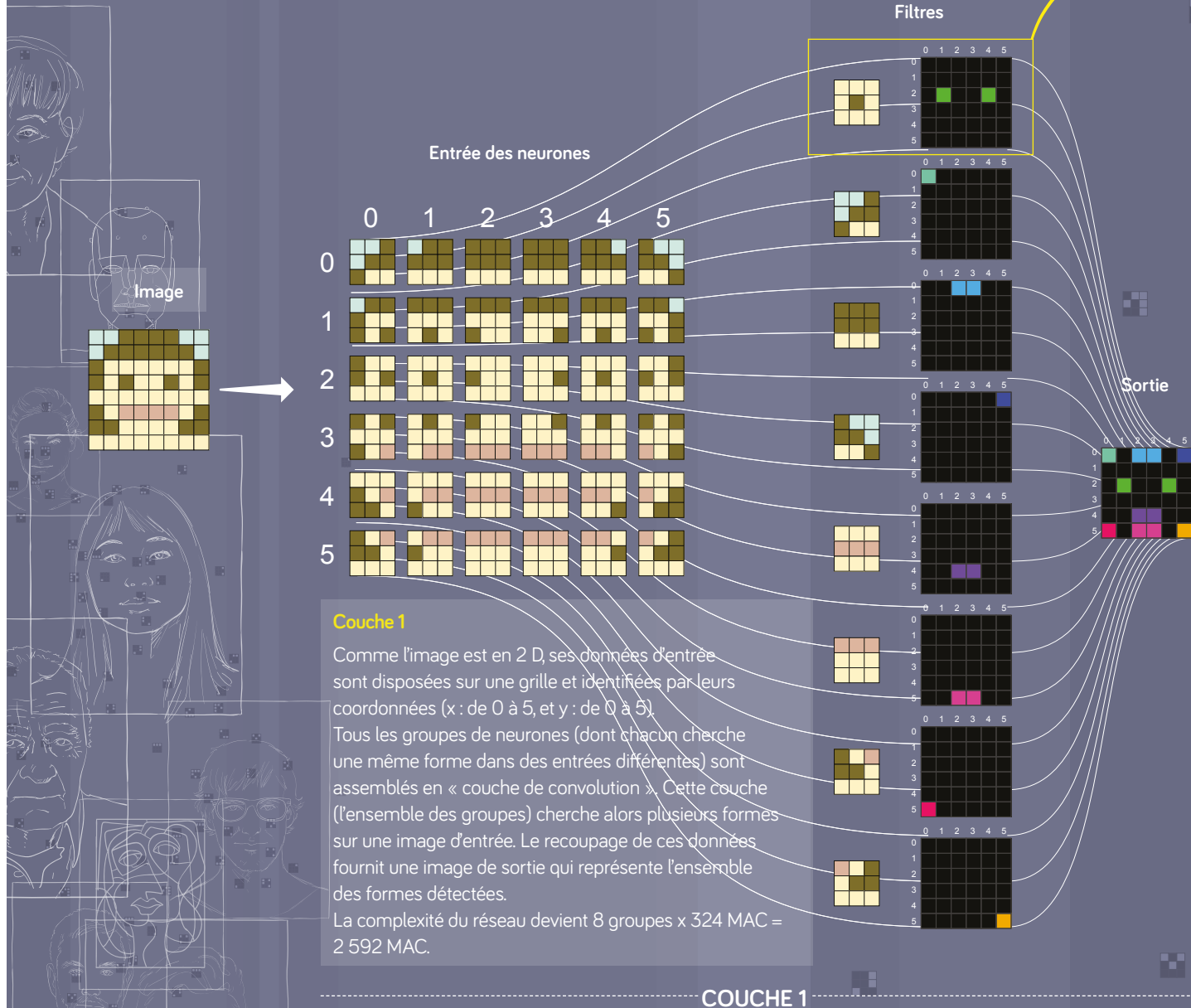
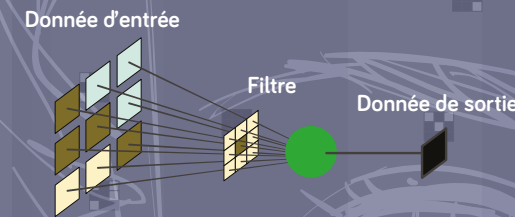
Un neurone

Le neurone (bille verte) connecte ses pondérations aux données d'entrée d'une partie de l'image. Le calcul du neurone constitue un « filtre » dont les paramètres sont fixés par l'algorithme pour que le neurone identifie une forme précise (ici, ce qui serait l'œil du visage) dans une partie précise de l'image. Le résultat de ce calcul est la donnée de sortie : ici, un carré noir qui signifie que le neurone n'a pas identifié l'œil dans la partie de l'image à laquelle il était connecté. Précisément, le calcul consiste en la somme pondérée des données d'entrée : c'est-à-dire que le neurone multiplie ses propres valeurs (pondération de ses entrées) avec celles des pixels et qu'il en fait la somme. Cette opération dite « multiplication et accumulation (MAC) » qualifie la complexité d'un réseau : ici, il y a 9 MAC (en relation avec les 9 pixels de la donnée d'entrée).



Un groupe de neurones

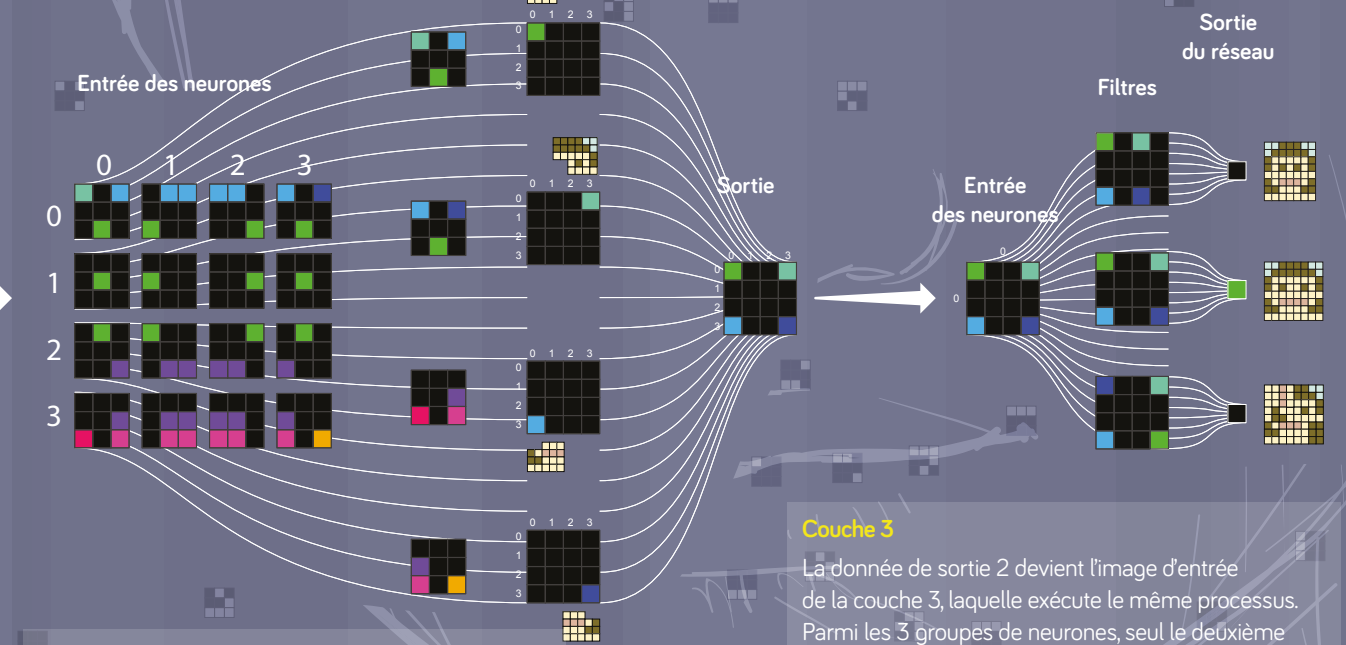
Les neurones sont assemblés dans un groupe. Tous cherchent la même forme (œil) mais dans des zones différentes de l'image d'entrée. Ils ont la même fonctionnalité mais avec des paramètres (filtres) différents. La donnée de sortie est exprimée avec différentes intensités de couleurs : ici, les deux pixels verts à 100 % signifient que 2 neurones ont identifié la forme « œil » à des endroits précis (mais différents) de l'image. Ici, la complexité du réseau est : 6 x 6 neurones effectuant chacun 9 MAC = 324 MAC.



Couche 1

Comme l'image est en 2 D, ses données d'entrée sont disposées sur une grille et identifiées par leurs coordonnées (x : de 0 à 5, et y : de 0 à 5). Tous les groupes de neurones (dont chacun cherche une même forme dans des entrées différentes) sont assemblés en « couche de convolution ». Cette couche (l'ensemble des groupes) cherche alors plusieurs formes sur une image d'entrée. Le recoupage de ces données fournit une image de sortie qui représente l'ensemble des formes détectées. La complexité du réseau devient 8 groupes x 324 MAC = 2 592 MAC.

COUCHE 1



Couche 2

La donnée de sortie de la couche 1 devient l'image d'entrée de la couche 2. D'autres neurones prennent le relais, avec de nouveaux filtres et fournissent de nouvelles données de sorties (différentes couleurs et positions, selon les différents filtres) qui sont envoyées dans une troisième couche.

COUCHE 2

Couche 3

La donnée de sortie 2 devient l'image d'entrée de la couche 3, laquelle exécute le même processus. Parmi les 3 groupes de neurones, seul le deuxième a correctement appris à reconnaître le visage présenté en entrée, en témoigne le carré vert. L'algorithme d'apprentissage va alors changer les pondérations (valeurs intrinsèques des neurones) de toutes les couches afin que les deux autres groupes de neurones parviennent également à reconnaître le visage.

COUCHE 3

TOUT
S'EXPLIQUE



Plateforme N2D2 du CEA-List, qui permet de concevoir, entraîner et évaluer des réseaux de neurones profonds.

Le défi des réseaux de neurones convolutionnels (CNN)

S'il est naturel pour l'homme de savoir ce qu'est un visage ou combien de personnes il y a sur une photo ; un ordinateur n'y verra que des pixels et des nombres ! Dans les premières techniques d'analyses d'images, l'homme décrivait à l'ordinateur un modèle de visage qui était relatif à différents critères comme des variations de couleur. L'ordinateur devait alors trouver, deux zones sombres (les yeux) et une zone claire au milieu (le nez), ce qu'il parvenait à faire... plus ou moins bien. Quand les visages étaient trop éclairés ou légèrement déformés, il ne pouvait pas les reconnaître. Les modèles devaient se complexifier à l'extrême pour bien expliquer à l'ordinateur ce qu'était un visage.

L'architecture d'un CNN

Une nouvelle technique fut développée : les réseaux de neurones, dont ceux dit « convolutionnels » (CNN). Là, il ne s'agit plus de donner des modèles à l'ordinateur ; mais de lui faire apprendre ce qu'est un visage à partir d'une base de données de milliers d'images annotées (les annotations étant les réponses attendues, par exemple : « visage » ou « autre »), base de données dans laquelle il trouve lui-même les caractéristiques importantes d'un visage. Pour ce faire, le réseau CNN est assemblé en groupes de neurones, dont les calculs (filtres) cherchent une même forme dans l'image. Et ces groupes sont organisés en couches de convolution : les premières couches servent à identifier des formes simples (lignes verticales, horizontales, obliques, etc.) ; les suivantes mettent ces informations en commun pour reconnaître des formes plus complexes (un oeil) ; et les dernières, appelées « classifieur », sont capables de classer des informations et de prendre une « décision » (par exemple, « je reconnais un visage »).

Ce sont ainsi des millions de filtres à paramétrer, pour chercher autant de formes.

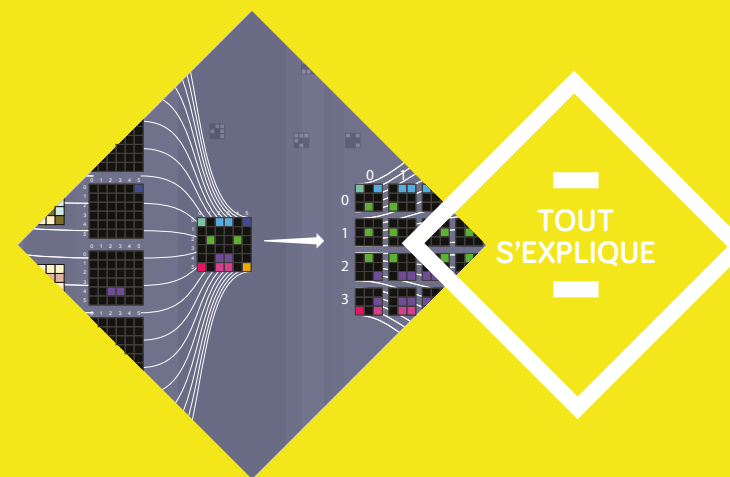
L'algorithme d'apprentissage

Le paramétrage est une tâche déterminante. Elle est confiée à des algorithmes d'apprentissage qui permettent de trouver les bons filtres de toutes les couches. L'algorithme fournit à l'ordinateur des données d'entrée (image) et les réponses (annotations) qui sont attendues en sortie. Si l'ordinateur se trompe, l'algorithme modifie les filtres. En répétant cette opération, l'ordinateur apprend progressivement. Mais il faut s'assurer qu'il n'apprend pas par cœur ! C'est pourquoi des images non présentes dans la base de données d'apprentissage sont soumises au réseau. S'il reconnaît bien ces nouvelles images, il aura bien « généralisé », c'est-à-dire qu'il est capable de reconnaître quelque chose qu'il ne connaît pas. Si non, la base d'apprentissage doit être modifiée.

Complexité des réseaux

L'exemple présenté dans l'infographie n'est pas comparable à la réalité. Pour rappel, le réseau de neurones convolutionnel précurseur développé en 2012 (il y a déjà 7 ans) par l'équipe de Geoffrey Hinton (Google Brain) comptait 650 000 neurones et 60 millions de connections sur 8 couches ! Les perspectives de ces réseaux pour l'intelligence artificielle sont immenses tout comme les compétences clés à maîtriser : identifier les architectures de réseaux efficaces pour résoudre un problème donné ; concevoir les algorithmes d'apprentissages idoines ; expertiser les bases de données ; repenser les plateformes de calcul, leur parallélisation, leur simplification...

les défis 235
du cea



TOUT
S'EXPLIQUE

Les réseaux de neurones

Sous-ensemble de technologies de l'intelligence artificielle, les réseaux de neurones sont des programmes informatiques dont l'architecture s'inspire de celle des neurones biologiques. Ces neurones artificiels, reliés entre eux et organisés en couches successives, apprennent à reconnaître des données à partir d'exemples et peuvent généraliser cet apprentissage afin d'aider à la prise de décision.

ENJEUX



Les réseaux de neurones sont au cœur de l'intelligence artificielle (IA), laquelle est désormais un enjeu majeur pour de nombreux secteurs (santé, énergie & environnement, mobilité, industries...) tel que le confirme un récent rapport du ministère de l'Économie et des Finances (étude Pipame). Parmi les technologies de l'IA, le *deep learning* (méthodes d'apprentissage profond qui servent notamment pour les réseaux de neurones convolutionnels (CNN) dédiés à l'analyse d'images) est en plein essor. Les plus grosses entreprises, telles les Gafa, l'utilisent massivement pour des applications commerciales comme la diffusion personnalisée de publicité. Le *deep learning* est également au cœur de nombreuses recherches en vue d'adapter ses technologies aux problématiques industrielles. La Direction de la recherche technologique du CEA, principalement au sein de l'institut List, y consacre

une part importante de son activité pour l'intégrer dans de nombreuses applications. Par exemple : détection d'incidents par des capteurs en usine ; identification et différenciation de véhicules dans une scène par le logiciel Deep Manta pour le véhicule autonome ; analyse de défauts dans des matériaux à partir de données de contrôle non destructif ; vision par ordinateur développée avec Thales dans le cadre du laboratoire commun VisionLab ; extraction d'information textuelle LIMA intégrée dans certains processus de Bureau Veritas... Pour cela, le CEA-List a notamment développé N2D2, une plateforme logicielle de *deep learning* en *open source* qui permet de concevoir, entraîner et évaluer des réseaux de neurones profonds, et de tester automatiquement une version optimisée de ces réseaux sur différentes solutions matérielles d'exécution.