

Il serait facile pour un système d'apprendre par cœur les réponses correctes sur les exemples connus (il suffit d'une mémoire suffisante). Mais cela n'a pas grand intérêt. Même si on accumule des millions d'images de coucher de soleil, le nombre d'images possibles d'une telle scène est infini. L'objectif de l'apprentissage est d'être capable de donner de bonnes réponses sur de nouveaux exemples, inconnus du système, en généralisant les exemples connus. Le bon niveau de généralisation dépend du contexte. Ainsi, on peut vouloir reconnaître la notion d'arbre en général, mais on peut aussi vouloir distinguer les chênes des hêtres, ou vouloir reconnaître le hêtre d'un jardin donné...

Un tel algorithme doit aussi s'appuyer sur certaines hypothèses relatives aux données et à ce que pourrait être une solution possible à un problème donné. Ainsi, le logiciel doit intégrer comme principe que si des données d'entrée d'une fonction particulière sont semblables, leur sortie ne devrait pas être radicalement différente: modifier quelques pixels sur une image de chat ne devrait pas transformer l'animal en chien.

De telles hypothèses faites sur des images sont utilisées par les réseaux de neurones dits convolutifs. Ces programmes sont à l'origine du renouveau de l'intelligence artificielle. Les réseaux neuronaux convolutifs utilisés dans l'apprentissage profond comprennent de nombreuses couches de neurones organisées de sorte que le logiciel sera robuste vis-à-vis des changements dans l'objet qu'il tente d'analyser. Il reconnaîtra l'élément même s'il a un peu bougé. Ainsi, un réseau bien entraîné identifiera un visage quel que soit l'angle de vue.

UNE STRUCTURE MULTICOUCHE

La configuration d'un réseau convolutif s'inspire de la structure en couches multiples du cortex visuel, c'est-à-dire la partie de notre cerveau qui reçoit les signaux des yeux. Ces nombreuses couches de neurones virtuels justifient le qualificatif de «profond» donné à ces réseaux et confèrent à ces derniers une meilleure capacité à appréhender le monde environnant.

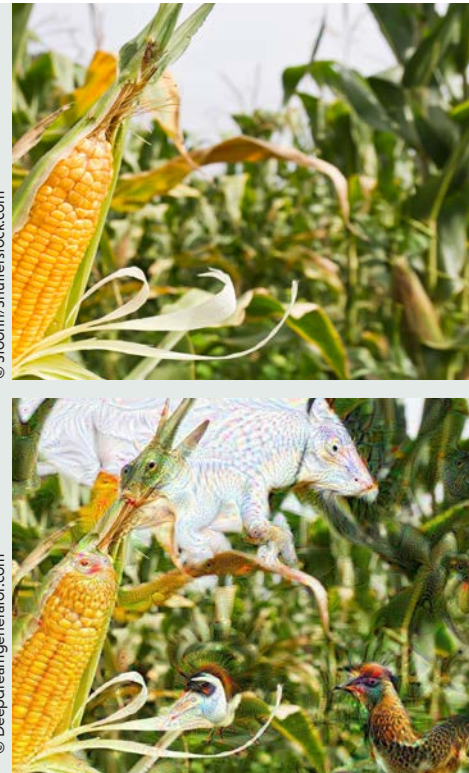
L'apprentissage profond est devenu envisageable il y a une dizaine d'années grâce à certaines innovations, alors que l'intérêt pour l'intelligence artificielle était au plus bas. Un organisme canadien financé par le gouvernement et par des fonds privés, l'Icra (Institut canadien de recherches avancées), a contribué à ranimer la flamme en soutenant un programme dirigé par Geoffrey Hinton, de l'université de Toronto, et dont faisaient partie Yann LeCun, de l'université de New York et du centre de recherche de Facebook à Paris, Andrew Ng, de l'université de Stanford, Bruno Olshausen, de l'université de Californie à Berkeley, moi-même et

APPRENTISSAGE PROFOND ET ART PSYCHÉDÉLIQUE

L'apprentissage profond est performant pour la reconnaissance d'objets dans une image, mais il est en réalité assez difficile de comprendre ce qui se passe au cours de l'analyse à chaque niveau du réseau de neurones. La société Google a développé le logiciel DeepDream qui utilise l'apprentissage profond. Les chercheurs peuvent analyser le résultat à différentes étapes du processus en demandant au logiciel de faire ressortir ce qu'il «voit» dans l'image. Ainsi, telles des paréidolies – l'illusion de voir, par exemple, un chien dans un nuage qui en a vaguement la forme –, le logiciel, qui s'est, par exemple, entraîné à reconnaître des images d'animaux, verra des oiseaux et des rongeurs dans une photographie d'un champ de maïs. Chacun peut créer ses propres images psychédéliques sur le site Deepdreamgenerator.com.

© Sroonn/Shutterstock.com

© Deepdreamgenerator.com



plusieurs autres. En raison du scepticisme ambiant, il était difficile à cette époque de publier des articles sur le sujet et même de persuader des étudiants de faire leur thèse dans ce domaine. Mais nous étions convaincus qu'il fallait persévérer dans cette voie.

La réticence vis-à-vis de l'intelligence artificielle découlait initialement de l'idée que la création de réseaux neuronaux était sans espoir à cause de la difficulté à optimiser leur performance pour apprendre efficacement.

L'optimisation est une branche des mathématiques qui vise à trouver la meilleure combinaison de paramètres, ici les poids synaptiques des neurones virtuels, pour atteindre un certain objectif. Lorsque la relation qui lie les paramètres et l'objectif est assez simple (on dit que l'objectif est une «fonction convexe» des paramètres) alors on peut améliorer progressivement les paramètres et l'on parle d'optimisation convexe. La procédure d'entraînement est répétée jusqu'à ce que les paramètres s'approchent aussi près que possible des valeurs qui produisent le meilleur résultat. On cherche en fait un minimum global, à savoir le jeu de paramètres qui permet d'atteindre la valeur la plus basse (et la meilleure) de l'écart à l'optimum.

En général, la situation n'est pas aussi simple, la fonction reliant les paramètres à l'objectif qu'ils atteignent n'étant pas convexe. Or l'optimisation non convexe représente un défi bien plus difficile. De nombreux chercheurs

➤ pensaient qu'il n'était pas possible de le relever. L'algorithme d'apprentissage pourrait en effet se retrouver coincé dans un minimum local, d'où il ne pourrait sortir en ajustant les paramètres par petites touches.

Or avec mes collègues, nous avons montré que si le réseau de neurones est assez grand, le problème des minima locaux est fortement réduit. Dans un tel réseau, la plupart des minima locaux correspondent à un niveau d'apprentissage quasi équivalent à celui du minimum global.

Bien que les problèmes théoriques de l'optimisation soient, en principe, résolubles, la construction de grands réseaux comportant plus de deux ou trois couches avait souvent échoué. À partir de 2005, les recherches financées par l'Icra ont commencé à porter leurs fruits et les obstacles ont été surmontés progressivement. En 2006, nous sommes parvenus à entraîner des réseaux neuronaux plus profonds en utilisant une technique qui opérait couche par couche.

Par la suite, en 2011, nous avons trouvé un moyen d'entraîner des réseaux encore plus profonds (avec davantage de couches de neurones virtuels) en modifiant les opérations effectuées par chacun des neurones, ce qui les rapprochait davantage des neurones biologiques dans leur façon de traiter l'information. Nous avons aussi découvert qu'en injectant du bruit aléatoire dans les signaux transmis entre les neurones au cours de l'apprentissage – comme ce qui a cours dans le cerveau –, les réseaux apprenaient mieux à identifier correctement une image ou un son.

Par ailleurs, deux facteurs essentiels ont contribué au succès des techniques d'apprentissage profond. Le premier est l'augmentation d'un facteur 10 de la puissance de calcul des ordinateurs grâce aux processeurs dédiés au traitement d'image, ce qui a permis d'entraîner des réseaux plus grands en un temps raisonnable. Le second facteur est que l'on avait désormais accès à d'énormes bases de données étiquetées, avec lesquelles les algorithmes d'apprentissage ont pu s'exercer à reconnaître un « chat », par exemple.

Une autre raison aux récents succès de l'apprentissage profond est que le réseau est capable d'apprendre à effectuer un ensemble de calculs qui construisent ou analysent pas à pas une image, un son ou un autre type de donnée. Avec une profondeur suffisante, ces dispositifs excellent dans beaucoup de tâches de reconnaissance visuelle ou auditive. Des travaux théoriques et expérimentaux récents ont d'ailleurs montré qu'il est impossible d'effectuer efficacement certaines de ces opérations mathématiques sans un réseau assez profond.

Mais que se passe-t-il au cœur du réseau de neurones profond ? Chaque couche transforme son entrée et produit une sortie qui est envoyée à la couche suivante (voir l'encadré

page 45). Les premières couches se concentrent sur les détails aux échelles les plus petites, puis les couches suivantes agrandissent les échelles considérées. Plus les couches sont profondes, plus elles vont représenter des concepts abstraits.

RECONNAÎTRE DES VIDÉOS

Jusqu'à récemment, les réseaux de neurones artificiels se distinguaient en grande partie par leur capacité à effectuer des tâches telles que la reconnaissance de motifs dans des images statiques. Mais d'autres types de réseaux neuronaux affichent des résultats intéressants sur des événements dynamiques. Les « réseaux neuronaux récurrents », par exemple, ont démontré leur capacité à effectuer une séquence de calculs pour traiter la parole ou une vidéo. Les données séquentielles sont constituées d'unités (phonèmes ou mots dans le cas de la parole) qui se suivent. La façon dont les réseaux neuronaux récurrents traitent leurs entrées ressemble un peu au fonctionnement du cerveau. Les signaux qui se propagent parmi les neurones changent constamment à mesure que les entrées sensorielles sont traitées.

Les réseaux récurrents parviennent à prédire quel sera le mot suivant dans une phrase et peuvent ainsi produire une nouvelle séquence de mots, l'un après l'autre. Ils peuvent aussi s'atteler à des tâches plus complexes. Après avoir « lu » tous les mots d'une phrase, le réseau est à même de deviner le sens de la phrase entière. Un réseau récurrent distinct peut alors utiliser le traitement sémantique du premier réseau pour traduire la phrase dans une autre langue.

La recherche sur les réseaux neuronaux récurrents a connu sa propre période de stagnation à la fin des années 1990 et au début des années 2000. Mes propres travaux théoriques suggéraient qu'ils rencontreraient des difficultés à apprendre à récupérer de l'information du passé lointain, à savoir les premiers éléments de la séquence en cours de traitement. Mais certains de ces problèmes ont été résolus grâce à des techniques consistant à stocker l'information pour qu'elle persiste durablement. Ces réseaux neuronaux utilisent la mémoire temporaire (mémoire cache) d'un ordinateur pour traiter des informations multiples et dispersées, comme des idées contenues dans différentes phrases réparties au fil d'un document.

Le retour en force des réseaux neuronaux à apprentissage profond après le long sommeil de l'intelligence artificielle n'est pas uniquement un triomphe technique. C'est aussi une leçon de sociologie des sciences. En particulier, il souligne la nécessité de soutenir des idées qui font fi du *statu quo* technologique et d'encourager la tenue d'un portefeuille de recherches diversifié, laissant de la place à des champs provisoirement passés de mode. ■

BIBLIOGRAPHIE

D. SILVER ET AL., *Mastering the game of Go without human knowledge*, *Nature*, vol. 550, pp. 354-359, 2017.

Y. LECUN ET AL., *Deep learning*, *Nature*, vol. 521, pp. 436-444, 2015.

Y. BENGIO ET AL., *Representation learning: A review and new perspectives*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35(8), pp. 1798-1828, 2013.

ImageNet classification with deep convolutional neural networks, 26th Annual Conference on Neural Information Processing Systems (NIPS 2012), 2012.

SUR LE WEB

Cours de Yann LeCun au Collège de France, 2015-2016 : <https://www.college-de-france.fr/site/yann-lecun/index.htm>