

**JEAN PONCE** professeur à l'ENS (Ulm)

VISION ARTIFICIELLE ENTRETIEN AVEC UN CHERCHEUR

Le développement de l'intelligence artificielle suscite de nombreux débats alimentés par les craintes les plus tragiques et les espoirs les plus fous. Écouter un chercheur en contact quotidien avec cette nouvelle discipline permet de mieux discerner les vraies évolutions qui se profilent.

Pouvez-vous nous présenter vos travaux en lien avec l'intelligence artificielle ?

Le but de la vision artificielle est d'interpréter automatiquement une image (que ce soit une photo, une vidéo ou une image de profondeur obtenue grâce à des capteurs spécialisés) en développant des modèles dont l'apprentissage nécessite le moins de supervision possible. En effet, c'est le travail d'étiquetage des données qui est le plus coûteux, et plus la machine est autonome dans sa reconnaissance et son interprétation d'images, meilleur est le modèle. Il s'agit donc (entre autres) d'essayer de comprendre les connexions et les structures implicites entre des images: qu'est-ce qui fait qu'au-delà de la diversité entre les tailles, résolutions, fonds de l'image, etc., on peut regrouper l'ensemble des images de castors ?

Y a-t-il eu des progrès continus en vision artificielle ? Quels sont les espoirs pour cette discipline dans les années à venir et les obstacles sur lesquels vous butez ?

Il y a eu en effet des progrès continus en vision artificielle, contrairement à d'autres champs de l'intelligence artificielle qui ont connu des phases d'accélération puis de ralentissement.

« Il y a eu des progrès continus en vision artificielle, contrairement à d'autres champs de l'IA »

Depuis la fin des années 1990, de gros progrès ont été accomplis en reconnaissance visuelle (notamment par Cordelia Schmid et son équipe à Grenoble): celle-ci ne se limite

plus à des images très simples, mais continue à « marcher » avec des géométries et des fonds complexes par exemple. En parallèle, les appareils photos détectent les visages au moment de la prise de vue, ce qui permet une plus grande netteté de l'image.

REPÈRES

Dans le département informatique de l'ENS que dirigeait Jean Ponce jusqu'en septembre 2017, il y a une dizaine d'équipes dont certaines qui travaillent sur des sujets d'intelligence artificielle: l'équipe de Francis Bach qui fait de l'apprentissage statistique, celle de Stéphane Mallat qui s'occupe d'analyse et de classification des données ainsi que de *deep learning* et la sienne spécialisée dans la vision artificielle, ce qui permet d'échanger sur les différents sous-domaines de l'IA.



Qu'est-ce qui fait qu'au-delà de la diversité entre les tailles, résolutions, fonds de l'image, etc., on peut regrouper l'ensemble des images de castors ?

RENDRE LES PHOTOS PLUS NETTES

On arrive aujourd'hui à considérablement améliorer la qualité d'une image grâce aux méthodes de « débruitage » comme en attestent les images ci-après (Mairal et al., 2009).



Pour moi, nous ne sommes pas plus près aujourd'hui qu'il y a vingt ans d'une « machine intelligente » qui détrônerait l'humanité. Je sais que des personnes très compétentes comme Elon Musk, Bill Gates, Stephen Hawking ont lancé un appel à la prudence vis-à-vis des machines intelligentes qui pourraient menacer à terme l'humanité en l'absence de régulation, mais je considère actuellement cette crainte comme dénuée de fondement.

En revanche, il est vrai que nous allons vers une intelligence artificielle plus intégrée : alors que les différents domaines de l'IA s'étaient séparés dans les années 1980, les progrès effectués dans chacune de ces sous-branches (comme la robotique, le traitement du langage naturel, la vision artificielle, etc.) permettent actuellement d'aller vers plus d'intégration de leurs outils. La robotique notamment se doit d'intégrer plusieurs couches « d'intelligence » : un robot doit maîtriser la perception, apprendre par expérience,

communiquer, etc. Cependant, attention, plus on empile les modules, plus le système est compliqué et plus il y a de risque quant à sa fiabilité. Par exemple, il est difficile d'identifier ce qui

a causé l'accident mortel provoqué par une voiture Tesla. Est-ce le système de

En 2012, un bond de performance a été permis par l'utilisation de réseaux de neurones pour la reconnaissance visuelle (cf. les travaux de Geoffrey Hinton à Toronto). Cette technologie existait depuis les années 1990 mais était surtout utilisée pour la reconnaissance d'écriture manuscrite par exemple et avait été peu exploitée dans le cadre du traitement de photographies. Cette technologie a permis des progrès importants dans ce cadre mais des problèmes demeurent pour la vidéo où elle ne donne pas des résultats franchement meilleurs que les méthodes classiques, notamment à cause de la variabilité temporelle des images (on arrive à bien reconnaître un visage ou un chien sur une image fixe, mais le problème se complexifie quand les images se déforment avec le mouvement). Un autre problème que l'on rencontre actuellement dans notre discipline est le manque de sémantique dans les étiquetages manuels qui sont un point clef des méthodes d'apprentissage. Qu'est-ce que je caractérise comme une action par exemple ? Comment utiliser une sémantique commune d'une base de données à l'autre ?

Quels sont les progrès actuels liés au deep learning ?

Se rapproche-t-on de machines qui s'autonomisent par rapport aux humains ?

Le *deep learning* est très à la mode, mais il faut relativiser ses progrès, en tout cas dans le domaine de la vision artificielle : la technologie n'est pas très différente de celle des années 1990 (alors que celle-ci était davantage en rupture avec la technologie des décennies précédentes).

« Nous ne sommes pas plus près aujourd'hui qu'il y a vingt ans d'une « machine intelligente » qui détrônerait l'humanité »

L'APPRENTISSAGE N'EST PAS LA PANACÉE

L'intelligence artificielle progresse aussi en dehors de l'apprentissage et de larges pans de l'IA ne sont pas concernés par ces évolutions. Par exemple, une méthodologie fondamentale en robotique est ce qu'on appelle « SLAM » (*simultaneous localization and mapping*), où le robot est équipé de capteurs (souvent visuels) lui permettant de construire un modèle tridimensionnel de son environnement et de l'utiliser pour naviguer, sans aucun apport (en général) de méthodes d'apprentissage.



© ÉCOLE POLYTECHNIQUE - J. BARANDE

Beaucoup d'étudiants ont des bourses que ce soit par les grandes écoles comme Polytechnique ou des bourses de thèse.

captage des informations, de leur analyse, de la prise de décision qui était défaillant ? C'est pourquoi l'intégration des différents modules revêt une importance toute particulière, même si nous sommes encore loin du moment où cela sera le seul problème, chacun des modules ayant encore de gros progrès à accomplir.

Vous avez été longtemps chercheur aux États-Unis, comment se situent la France et l'Europe dans la compétition avec les autres pays (États-Unis, Chine notamment) ? Comment s'articulent les liens entre recherche académique et industrielle ?

Les conditions de recherche en France ou en Europe sont très favorables, notamment grâce à l'ERC (European Research Council) qui octroie des fonds généreux pour des programmes de recherche. Lorsque j'étais chercheur aux États-Unis, je passais beaucoup plus de temps à rechercher des contrats que mainte-

nant. De plus, en France, beaucoup d'étudiants ont des bourses, que ce soit par les grandes écoles comme l'ENS ou Polytechnique ou des bourses de thèse, ce qui soulage financièrement les laboratoires qui les emploient.

Pour ce qui est des relations entre laboratoires académiques et industriels, les premiers sont très appréciés de l'industrie qui embauche les étudiants qui en sont issus, donnent des fonds sans contrepartie, etc. Les relations sont donc bonnes, et la recherche académique continue d'avoir un rôle important à jouer (les industries sont moins intéressées maintenant par les brevets ou produits clés en main que par le fait d'avoir accès à des cerveaux bien formés). Les laboratoires académiques ont moins de données que ceux des GAFAPar

exemple, mais les laboratoires de ces derniers se heurtent à la protection de la vie privée, ce qui limite considérablement l'usage qu'ils peuvent faire de ces données. De plus, l'objectif de la recherche en vision artificielle est de travailler avec le moins d'annotation possible donc le

« Pour ce qui est de la vision artificielle, la France compte parmi les meilleurs mondiaux ! »

fait de ne pas avoir accès à tous les « tags » des photos que peuvent collecter les entreprises n'est pas toujours préjudiciable.

Enfin, on a souvent tendance à ne penser qu'aux géants de l'internet, mais

l'IA progresse également grâce à d'autres industries qui ont énormément de données (on peut penser aux banques ou à la médecine par exemple), domaines dans lesquels l'Europe a aussi des champions. Pour ce qui est de la vision artificielle, la France compte parmi les meilleurs mondiaux! ■