> culture matérielle, un discours plus riche, sensible aux dimensions rituelle et culturelle, qui dépasse la simple séduction esthétique.

Le musée voit aussi converger vers lui des réseaux dont la coexistence pourrait sembler insolite si on oubliait que ses directeurs font preuve d'un grand pragmatisme pour consolider leur entreprise de modernisation et de popularisation du Trocadéro : des scouts, des missionnaires religieux, des explorateurs et voyageurs, des fonctionnaires coloniaux, des journalistes, des artistes et écrivains, des banquiers et financiers, des galeristes et collectionneurs, des nobles, des étudiants et des savants…

ATTIRER DADAS, SURRÉALISTES ET AUTRES ARTISTES D'AVANT-GARDE

De fait, les directeurs du musée d'ethnographie font flèche de tout bois pour mettre en valeur leur ambitieuse entreprise de rénovation du lieu, tribune et vitrine de cette science sur la scène parisienne : brouillant les frontières, ils profitent tout à la fois du goût du grand public pour l'aventure et l'exotisme, abondamment relayé dans la presse à gros tirage et les récits de voyage pittoresques, mais aussi de la ferveur des élites mondaines pour les arts dits primitifs, notamment des cercles artistiques d'avant-garde, en particulier les dissidents surréalistes et les artistes dadas. C'est toute la gageure de l'idéologie défendue par Rivet et Rivière telle qu'elle se manifeste dans le discours, la muséographie et l'animation éducative, scientifique et culturelle du musée : ils veulent en faire un lieu à la fois populaire, c'est-à-dire un instrument d'éducation et de propagande tourné vers le peuple, et à l'avant-garde de la culture. Pour eux, l'ethnologie, en tant que nouvelle science qui bouscule les hiérarchies épistémologiques, culturelles et raciales établies, doit être associée de plein droit à la culture.

À partir de 1932, ils parviennent ainsi à jouer sur les deux registres : celui du musée scientifique en mode majeur et celui du musée des beaux-arts en mode mineur. Pour lutter contre les préjugés raciaux et culturels, ils montrent au public tout ce qu'il peut avoir en commun avec cette humanité exotique : le geste et la parole, la technique et l'art. Conservatoire de la civilisation matérielle qui ouvre elle-même sur l'univers mental propre à chaque société, le musée d'ethnographie du Trocadéro veut faire la preuve par l'objet de l'indéfectible solidarité unissant tous les hommes, et montrer leurs communes aptitudes techniques qui sont autant de marches franchies dans l'ascension vers le progrès.

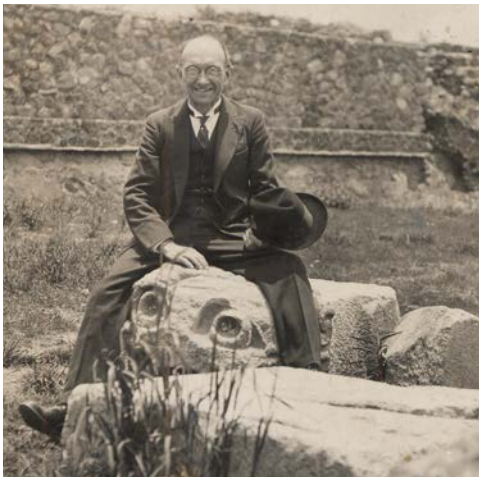
La mise en œuvre de cette volonté passe par un investissement tous azimuts dans les places fortes de la culture en pleine expansion,

LA CROISADE DES PREMIERS ETHNOLOGUES FRANÇAIS

En France, l'ethnologie entre officiellement à l'université en 1925, avec la fondation de l'Institut d'ethnologie, en 1925, sous l'égide du ministère des Colonies. On doit sa création au philosophe Lucien Lévy-Bruhl, au sociologue Marcel Mauss et à l'ethnologue Paul Rivet qui le dirige. L'Institut d'ethnologie forme la première génération d'ethnologues, hommes et femmes qui ont décidé de faire de leur vocation un métier. Dans la professionnalisation de ce métier, le terrain ethnographique devient « le rite d'initiation pratique », « l'épreuve du feu » (selon les mots de Rivet) qui permet d'observer par soi-même des cultures différentes, d'en comprendre les logiques sociales et rituelles, les valeurs.

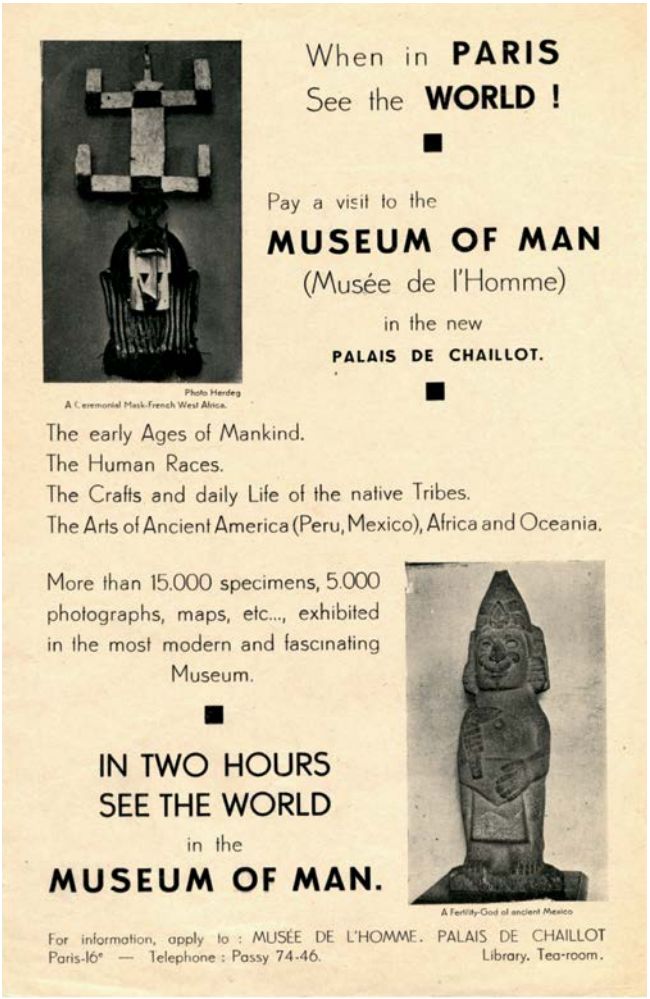
Plus de cent missions ethnographiques sont organisées entre 1925 et 1940, preuve de la vitalité de l'ethnologie. Elles obéissent à deux impératifs : la collecte d'objets de la culture matérielle afin de constituer les archives des sociétés traditionnelles ; l'ethnographie de sauvetage qui se hâte de documenter des modes de vie qui seraient menacés de disparition à cause de la colonisation et de la modernisation occidentales. Les possessions coloniales en Afrique et en Asie sont des terres de mission privilégiées, car faciles d'accès pour les ethnographes. Paul Rivet prône une alliance étroite entre le terrain ethnographique, la collecte d'objets, la recherche ethnologique et le musée d'ethnographie du Trocadéro qu'il dirige depuis 1928, véritable musée-laboratoire.

L'ambition est de quadriller la planète d'ethnographes et de collecteurs travaillant en coordination avec le musée. Si les terrains sous domination coloniale sont plus nombreux, ils sont loin d'être exclusifs. Jacques et Georgette Soustelle partent deux ans au Mexique et étudient les Otomis et les Lacandons ; Claude et Dina Lévi-Strauss mènent plusieurs enquêtes au Brésil, en particulier sur les Bororos et les Nambikwaras ; Alfred Métraux réalise des missions en Argentine, en Bolivie, à l'île de Pâques. André Leroi-Gourhan séjourne deux ans au Japon, Germaine Tillion et Thérèse Rivière s'installent dans l'Aurès algérien ; la grande mission Dakar-Djibouti, dirigée par Marcel Griaule, traverse l'Afrique entre 1931 et 1933, et popularise la culture dogon ; Paul-Émile Victor effectue plusieurs hivernages avec les Inuits de la côte orientale du Groenland ; Denise Paulme et Deborah Lifchitz mènent une enquête intensive à Sanga, au Mali ; Georges Devereux séjourne chez les Sedangs d'Indochine ; etc. Les ambitions sont multiples : faire connaître les sociétés lointaines de manière scientifique et didactique, témoigner de leur richesse culturelle, matérielle et artistique, restituer la diversité de façons d'être au monde, loin des épithètes péjoratives et dépréciatives.
C. L.



PAUL RIVET (1876-1958)

Ici photographié à Teotihuacán, au Mexique, en août 1929, Paul Rivet s'intéressa à l'ethnologie en accompagnant une mission géodésique en Amérique au début du XXe siècle. Il en devint un pionnier.



Dans la continuité des réformes du musée d'ethnographie du Trocadéro, le musée de l'Homme multiplia les événements et déploya de vastes campagnes de communication.

quelle qu'elle soit, et dans les médias de masse comme la radio, la presse écrite et la photographie. Des causeries ethnologiques sont ainsi diffusées chaque semaine sur les ondes de Radio-P.T.T., et dans les magazines et les quotidiens, on lit régulièrement des récits des ethnographes sur leur terrain. Rivet et Rivière n'hésitent pas à s'affranchir de la retenue et de la modération attendue des savants pour communiquer haut et fort sur les missions de leur musée en recourant parfois à des moyens inédits pour faire leur publicité.

Le musée se départit de son image d'austérité savante et se montre davantage accueillant et pédagogue vis-à-vis du public, des publics. Grâce à Rivière, issu du monde des arts, il est l'un des tout premiers musées à entrer dans l'ère de la communication, de la publicité, de l'événementiel, avec l'organisation de coups médiatiques, de fêtes, de bals, d'inaugurations d'expositions très régulières qui en font l'un des endroits les plus courus de Paris. Pour les journaux, le Trocadéro devient un lieu à la mode où il se passe toujours quelque chose, où règne un esprit moderne, chic, cultivé, jeune, joyeux, aventureux, savant sans être ennuyeux.

Rivet et Rivière montent des opérations de prestige qui n'excluent pas des événements grand public, populaires comme un gala de matchs de boxe au Cirque d'Hiver en avril 1931 ou la venue de Joséphine Baker au musée en juin 1933. Rivière, qui connaît déjà l'artiste américaine, l'a invitée pour une visite des nouvelles salles présentant les collections recueillies par la mission Dakar-Djibouti, qui traversa l'Afrique d'ouest en est entre 1931 et 1933. Il a imaginé une séance de prises de vue particulière en sa compagnie à l'occasion de l'ouverture prochaine de l'exposition consacrée à cette mission.

LES GRANDS BALS ETHNIQUES DU MUSÉE DU TROCADÉRO

Les inaugurations d'expositions et de salles rénovées (une bonne cinquantaine en cinq ans) suivent un rythme trépidant et sont immanquablement l'occasion d'événements festifs et médiatiques. De grands bals sont ainsi donnés par le musée, comme celui organisé à l'occasion de l'exposition coloniale de 1931 sur le thème des danses et chants africains, ou celui organisé à l'hôtel George V en mars 1934. Ce bal aux couleurs du Hoggar, une chaîne de montagnes du sud de l'Algérie, exploite habilement l'imaginaire orientaliste pour susciter la curiosité en attendant l'inauguration de la grande exposition sur le Sahara, qui rencontrera un énorme succès avec plus de 70 000 entrées.

Les initiatives de Rivet et Rivière ont ainsi propulsé le musée d'ethnographie du Trocadéro hors de son orbite scientifique naturelle et l'ont ouvert à une diversité de publics rarement atteinte depuis. Jamais le champ d'attraction de l'ethnologie ne fut plus vaste. Aujourd'hui, le musée du quai Branly et le musée de l'Homme s'inscrivent tous deux dans ce même esprit. Héritier des riches collections d'ethnographie de l'ancien musée de l'Homme, le musée du quai Branly, inauguré en juin 2006, continue à porter la volonté de mise en valeur de la diversité culturelle et artistique. Quant au nouveau musée de l'Homme inauguré en octobre 2015, il restitue non seulement l'histoire naturelle de l'humanité depuis la Préhistoire jusqu'à aujourd'hui dans toute la diversité physique de l'espèce humaine, mais aussi l'histoire culturelle de ses relations à son environnement, de son adaptation et de son utilisation de la nature. Plus que jamais, l'anthropologie est au cœur de la culture. ■

BIBLIOGRAPHIE

A. Delpuech, C. Laurière et C. Peltier-Caroff, **Les Années folles de l'ethnographie. Trocadéro 28-37**, MNHN, 2017.

C. Laurière, **Paul Rivet, le savant et le politique**, MNHN, 2008.

A. Conklin, **Exposer l'humanité. Race, ethnologie et empire en France (1850-1950)**, MNHN, 2015.

C. Blanckaert (dir.), **Le Musée de l'Homme. Histoire d'un musée-laboratoire**, MNHN/Artlys, 2015.

RENDEZ-VOUS

INTELLIGENCES ARTIFICIELLES : UN APPRENTISSAGE PAS SI PROFOND !

Les systèmes de reconnaissance automatique ont d'étonnantes faiblesses. Exploiter ces failles permet de s'amuser, mais aussi d'améliorer les procédures d'apprentissage… ou de concevoir de nouvelles attaques malveillantes.

L'AUTEUR

JEAN-PAUL DELAHAYE
professeur émérite à l'université de Lille et chercheur au Centre de recherche en informatique, signal et automatique de Lille (Cristal)

L'apprentissage automatique à l'aide de réseaux neuronaux «profonds» est à la mode. Comme souvent quand il s'agit d'intelligence artificielle, il y a ceux qui perçoivent raisonnablement les capacités des nouvelles idées et il y a ceux qui imaginent bien plus. Cet enthousiasme excessif s'est produit avec les premiers succès des méthodes d'apprentissage automatique dans la décennie 1950; cela a recommencé avec les systèmes experts dans les années 1980; puis quelque temps après avec les réseaux de neurones; et maintenant, c'est une forme de ces réseaux qui fait croire que nous sommes sur le point de mettre dans nos ordinateurs une intelligence générale susceptible de nous surpasser.

Le présent article n'a pas pour objectif de dénigrer une remarquable technique qui a récemment triomphé au jeu de go, qui aide à concevoir des véhicules autonomes, qui améliore la traduction automatique et qu'on maîtrise de mieux en mieux; les livres d'Aurélien Géron (*voir la bibliographie*) vous initieront à cette science nouvelle. Nous cherchons ici à remettre les pieds sur terre à ceux qui l'imaginent comme une panacée informatique.

RÉSEAUX NEUROMIMÉTIQUES

Les neurones formels proposés en 1959 sont des modèles informatiques et simplifiés des neurones de notre cerveau. On conçoit des programmes qui en simulent un grand nombre en les regroupant en couches successives, comme les neurones du cerveau humain. Les neurones formels de la couche d'entrée du réseau reçoivent des informations, par exemple sous la forme d'images décomposées en pixels: chaque pixel est une entrée. En fonction de leurs paramètres internes, les neurones de cette première couche envoient des signaux aux neurones de la seconde couche, qui eux-mêmes, en fonction de leurs paramètres internes, envoient des signaux à la troisième couche, etc.

Les influx de sortie, c'est-à-dire de la dernière couche, désignent les réponses possibles. Ce sont par exemple des lettres A, B, C …: si les images données en entrée sont des lettres manuscrites et imprécises, le but recherché est que le réseau, une fois instruit, sache correctement reconnaître les lettres qui lui sont proposées. La phase d'apprentissage consiste à ajuster les paramètres internes de chaque neurone pour que le réseau réponde correctement. Souvent, plusieurs milliers de paramètres sont à calculer. Tout l'art de la programmation des réseaux de neurones consiste, à l'aide d'un grand nombre d'images avec leurs classifications (par exemple des formes manuscrites dont on indique au réseau à quelles lettres elles correspondent), à faire converger les jeux de paramètres des neurones vers un état à peu près stabilisé et qui sera tel que lorsqu'on proposera des images, par exemple d'autres lettres manuscrites non utilisées, elles soient correctement identifiées.

On parle d'apprentissage supervisé: lors d'une phase initiale, on indique au réseau les réponses voulues; il apprend et, si tout s'est bien passé, il devient alors capable de trouver, seul, les bonnes réponses à des entrées

Jean-Paul Delahaye a récemment publié : **Les Mathématiciens se plient au jeu**, une sélection de ses chroniques parues dans *Pour la Science* (Belin, 2017).

PERFORMANCES : L'EXEMPLE DU SYSTÈME CAPTIONBOT

1

L'identification automatique d'objets sur une photo a fait récemment de grands progrès et les systèmes d'aujourd'hui sont capables, avec un certain succès, de reconnaître le contenu d'images délicates. Vous vous en rendrez compte vous-même en testant le système en ligne CaptionBot conçu par Microsoft (https://www.captionbot.ai).

Le système combine deux réseaux neuronaux. Le premier s'occupe de reconnaissance d'images, le second du traitement du langage naturel pour composer les réponses. En prenant en compte un grand nombre d'images accompagnées de leur description, le logiciel apprend à associer les caractéristiques de l'image à des descriptions écrites de ce qu'elles montrent. Il tente alors de faire de même avec de nouvelles images. Le système continue de progresser. En février 2018, on obtenait par exemple les résultats ci-dessous (les réponses ont été traduites).

Je pense que c'est un arbre avec en fond un coucher de soleil.

Je pense que c'est une personne tenant une guitare.

Je pense que c'est une femme assise sur un canapé.

Je pense que c'est un homme sur une plage.

Je n'en suis pas très sûr, mais je pense que c'est un ours brun assis sur une table.

Je pense que c'est un grand étalage de maïs.

Je pense que c'est une paire de chaussures bleues.

Je n'en suis pas très sûr, mais je pense que c'est un avion dans le ciel.

nouvelles. Un autre type de méthodes pour instruire un réseau de neurones formels se fonde sur l'idée de le récompenser quand il propose une bonne réponse et de le pénaliser sinon; ce sont les méthodes d'« apprentissage par renforcement ».

La puissance de calcul disponible aujourd'hui, les bases de données volumineuses collectées pour les exemples utilisés en phase d'apprentissage, ainsi que la multiplication des couches internes d'un réseau (on en utilise parfois plusieurs dizaines) ont conduit récemment à de spectaculaires réussites, comme la victoire des machines sur les humains au jeu de go.

Les chercheurs du domaine étaient persuadés qu'un réseau ayant bien appris était nécessairement robuste et doté d'une propriété de continuité : des traits dessinant de manière vague et approchée une lettre seront reconnus correctement, et une image proche d'une image reconnue sera elle aussi reconnue.

Or ce n'est pas le cas : les chercheurs ont mis au point des images pièges, conçues pour qu'un réseau ayant appris à reconnaître une certaine catégorie d'images propose une réponse incongrue et fausse quand on lui soumet une image proche, qu'un humain classe aisément. >

> Insistons sur la prudence nécessaire. Yann LeCun, pionnier de l'« apprentissage profond » ayant dirigé la recherche en intelligence artificielle chez Facebook, est clair : « C'est un abus de langage que de parler de neurones ! De la même façon qu'on parle d'aile pour un avion, mais aussi pour un oiseau, le neurone artificiel est un modèle extrêmement simplifié de la réalité biologique. » Il ajoute : « Les intelligences artificielles les plus abouties ont aujourd'hui moins de sens commun qu'un rat ! »

Nous disposons de méthodes qui surpassent les humains pour certains types de problèmes et c'est un remarquable succès. Cependant, dans de nombreux domaines, l'informatique a depuis longtemps créé des systèmes qui dépassent les humains. Pouvez-vous mémoriser 300 numéros de téléphone ? Savez-vous calculer 3^{10} en une seconde ? Non. Votre téléphone portable, lui, fait cela facilement. Les réussites des réseaux de neurones, des méthodes d'apprentissage automatique et des méthodes de fouille de données massives (les *big data*) sont de nouveaux exploits des machines, mais ces succès ne signifient pas que les machines sont devenues véritablement intelligentes.

L'ÉTRANGE EXPÉRIENCE DE 2014

Revenons à nos exemples pièges. Une équipe de sept chercheurs affiliés aux universités de New York et Montréal, à Google et à Facebook, réunie autour de Christian Szegedy, a publié en 2014 une méthode permettant de tromper les systèmes de reconnaissance visuelle.

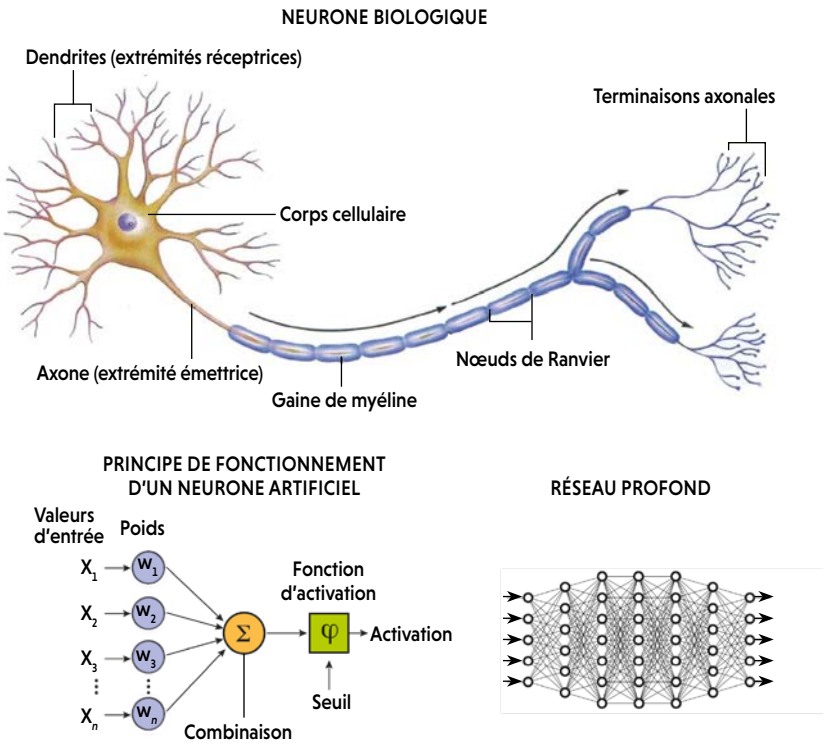
Les chercheurs en réseaux neuronaux affirment volontiers que les réseaux « généralisent ». Si un réseau apprend à reconnaître un cheval à partir d'un ensemble de photos de chevaux, il aura la capacité de reconnaître un cheval sur des photos qu'il n'a jamais vues. Il semblait alors aller de soi qu'un réseau qui classe correctement une photo de cheval donnée classera correctement la même photo où quelques pixels seulement ont été modifiés.

Ce n'est pas le cas : le réseau indiquera parfois qu'il voit une vache à la place du cheval, que tous les humains identifient pourtant. Si l'hypothèse de continuité « Ce qui est proche d'une image correctement traitée l'est aussi » jusque-là admise était correcte, alors aucune méthode piège de ce type ne pourrait fonctionner. Il est troublant que la technique du groupe de Christian Szegedy fonctionne presque toujours et puisse fabriquer des exemples trompeurs pour une vaste catégorie de réseaux de neurones et de données d'apprentissage. Ces chercheurs écrivent : « Pour tous les réseaux que nous avons étudiés et pour toutes les bases d'images utilisées pour l'apprentissage, nous réussissons à engendrer des exemples très proches, visuellement indiscernables d'images correctement reconnues, qui sont mal classées par le réseau. »

Pour construire les images pièges, on utilise une fonction qui estime le risque d'erreur pour chaque image possible, puis on utilise un algorithme d'optimisation qui, de petites modifications en petites modifications, déforme l'image >

2 NEURONES ET RÉSEAUX PROFONDS

Les neurones artificiels sont des versions simplifiées (matérielles ou logicielles) des neurones biologiques. Le plus souvent, on les conçoit de façon que les signaux d'entrée (des influx plus ou moins forts) sont mélangés : on calcule par exemple une combinaison linéaire de ses influx d'entrée en utilisant des paramètres w_1, w_2, ..., w_n propres au neurone. Si cette valeur calculée dépasse un certain seuil (encore un paramètre du neurone), alors un influx de sortie est émis. Un réseau dit profond comporte de nombreuses couches de neurones formels qui interagissent ainsi (schéma du bas, à droite). En réponse aux signaux donnés en entrée (par exemple les pixels d'une image), le réseau produit des signaux en sortie (qui codent par exemple l'objet reconnu dans l'image).



3

DES IMAGES PIÈGES

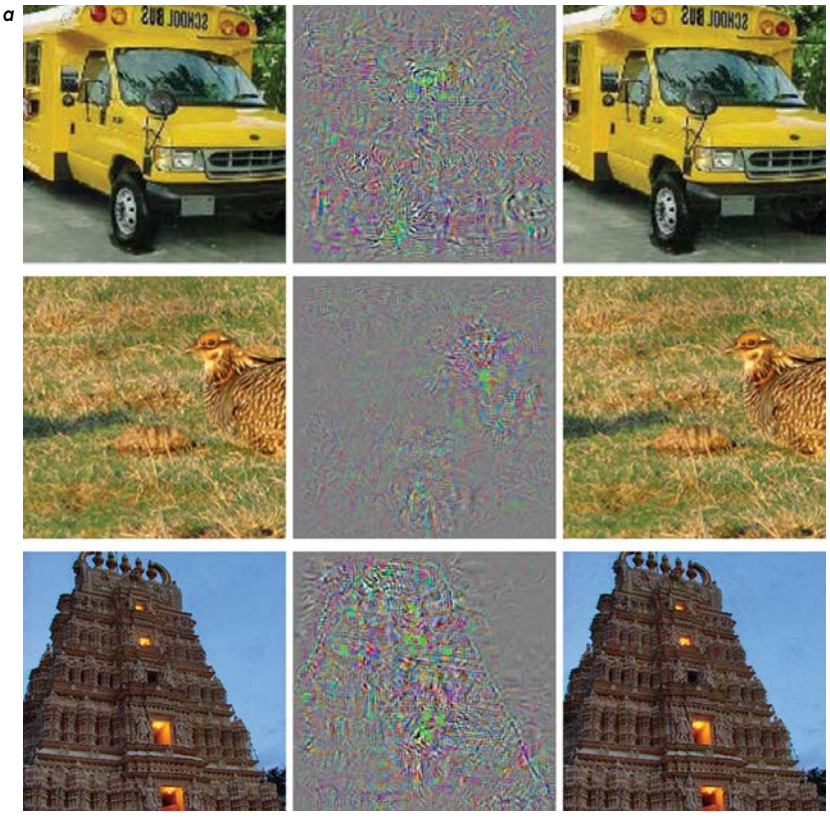
Un procédé de modification d'images permet de piéger les algorithmes neuronaux de reconnaissance d'images et les conduit à produire des réponses absurdes.

Les images proposées dans la colonne de gauche de la figure *a*, après la phase d'apprentissage du réseau de neurones, sont toutes correctement reconnues. Christian Szegedy et son équipe calculent alors des perturbations *(colonne centrale)* qui, appliquées aux images de la colonne de gauche, donnent les images de la colonne de droite. Ces dernières sont alors, dans chaque cas, reconnues par le système comme figurant des autruches !

Dans le premier tableau de la figure *b*, les chiffres manuscrits des colonnes 1, 3, et 5 sont identifiés correctement par le réseau de neurones (après apprentissage). En revanche, le réseau identifie mal les mêmes images qui ont subi de légères modifications (colonnes 2, 4 et 6), alors que l'œil humain les identifie aussi bien et ne voit presque pas de différences avec les premières images.

Dans le second tableau de la figure *b*, on a placé en colonnes 2, 4 et 6 les images des colonnes 1, 3 et 5 soumises à un assez fort bruit aléatoire. Le réseau continue pourtant de fournir les bonnes identifications dans plus de la moitié des cas, alors que les humains en sont incapables.

Ces deux tableaux montrent donc que ce qu'un réseau de neurones apprend dans sa phase d'apprentissage est d'une nature assez différente de ce que nous-mêmes apprenons.



> en augmentant le plus possible ce risque d'erreur, tout en changeant le moins possible l'image manipulée. De proche en proche, sans que l'œil humain ne perçoive de changements notables, l'algorithme construit une image qui fera trébucher le réseau.

UN NOUVEAU DOMAINE DE RECHERCHE

Comme souvent en recherche, quand une bonne idée est découverte, une multitude de travaux la précisent et la prolongent. Depuis l'article de 2014, des dizaines d'équipes ont trouvé et perfectionné toutes sortes d'astuces produisant des images pièges, et cela quelle que soit la méthode d'apprentissage mise en œuvre pour instruire le réseau. Dans certains cas, un seul pixel modifié induit en erreur le réseau (*voir l'encadré 4*). Assez souvent, une image qui piège un réseau trompera aussi d'autres réseaux ayant appris selon une méthode différente. Christian Szegedy explique:

«Nos exemples pièges sont relativement robustes, et trompent aussi des réseaux de neurones ayant un nombre différent de couches que celui qui les a engendrés ou provenant de phases d'apprentissage n'utilisant que des sous-ensembles des données.»

En apprentissage automatique, l'erreur que commettent les systèmes est parfois due à ce qu'on nomme le surajustement (*overfitting*). La méthode apprend au système à réagir sur un nombre limité d'exemples, mais dès qu'on en sort, le système se trompe, un peu comme une courbe que l'on force à passer par tous les points de mesure disponibles et qui est incapable de prédire correctement une nouvelle valeur.

Cependant, on a du mal à y voir la bonne explication des exemples pièges, puisque la classification automatique par le réseau fonctionne en général. Ce qui se produit est plus subtil qu'un surajustement: le réseau a généralisé les exemples utilisés dans la phase d'instruction, mais cette généralisation a laissé des failles que détectent les méthodes d'optimisation employées pour le piéger, en faisant apparaître des cas d'erreur inattendus.

De surcroît, on a su fabriquer des objets 3D qui, quand on les photographie, produisent des images pièges: ce n'est pas une image très spécifique qui trompe le réseau, mais toutes celles qu'engendre un objet réel. C'est inquiétant: on pourrait par exemple fabriquer un faux panneau «Stop» qu'un véhicule autonome interpréterait comme un panneau lui donnant la priorité! Ces images pièges rendent possibles de nouveaux types d'attaques contre des systèmes informatiques intégrant des réseaux de neurones. Parmi les travaux menés, mentionnons ceux ayant établi qu'on peut piéger plusieurs images à la fois: on cherche à optimiser le pixel qui augmente le plus la mesure du risque d'erreur non pas sur une seule image, mais sur un ensemble d'images. On construit ainsi de proche en proche une perturbation unique, une matrice de valeurs, qui, appliquée à chaque image d'un vaste ensemble d'images, produira une image piège.

Ce que l'on a réussi à faire pour les images a été adapté aux sons. Les méthodes décrites par Moustafa Alzantot, de l'université de Californie à Los Angeles, et ses collègues permettent de modifier insensiblement le son d'une voix qui prononce «yes» de façon que le système neuronal l'interprétera comme «no» (https://nesl.github.io/adversarial_audio/). On imagine les graves conséquences qu'auraient de telles méthodes si on les utilisait à des fins malveillantes...

Pour évaluer l'effet du bruit ajouté aux sons initiaux pour fabriquer les exemples

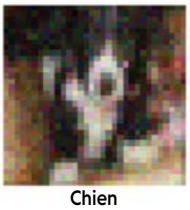
4 TROMPER AVEC UN SEUL PIXEL

Les images comportant un petit nombre de pixels sont plus délicates à déchiffrer. Pourtant, nous savons le faire assez bien, comme l'illustrent ces six images en basse résolution, assez facilement identifiables.

Dans l'article « One pixel attack for fooling deep neural networks » (prépublication arXiv:1710.08864, 2017), les chercheurs Jiawei Su, Danilo Vasconcellos Vargas et Sakurai Kouichi expliquent comment, après une phase d'apprentissage conduisant à de bonnes identifications (respectivement un cerf, un oiseau, un chien, un cheval, un bateau, un chat), ils ont réussi, en ne modifiant qu'un seul pixel de chaque image (qu'on voit assez nettement), à berner le réseau qui a alors reconnu autre chose, respectivement un avion, une grenouille, un chat, un chien, un avion, un chien.


Cerf
Avion (49,8 %)


Oiseau
Grenouille (88,8 %)


Chien
Chat (75,5 %)


Cheval
Chien (88 %)


Bateau
Avion (62,7 %)


Chat
Chien (78,2 %)

pièges, les chercheurs ont mené une comparaison avec des sujets humains. Chacune des personnes d'un groupe de 23 participants a écouté 1500 séquences sonores pièges sans être informée de ce qu'il fallait trouver. Dans 89% des cas, la reconnaissance par les sujets humains n'a pas été affectée, alors que la machine s'est trompée dans 100% des cas.

CORRIGER LES MÉTHODES

Ce qu'on a découvert donne des pistes pour créer de nouvelles méthodes d'apprentissage plus résistantes. L'idée la plus simple consiste, après la première période d'apprentissage, à construire de nombreuses images pièges et à les faire apprendre en complément au réseau en lui indiquant les bonnes réponses, tout comme un professeur de langue indique les faux amis à l'étudiant qui, averti, ne commettra pas l'erreur associée à ces expressions trompeuses.

Une autre idée simple fournit facilement une première protection : la compression de données. Gintare Dziugaite, de l'université de Cambridge, et ses collègues ont noté dès 2016 l'intérêt et les effets des méthodes de compression informatique pour se protéger des images pièges. Avant de faire apprendre une image au réseau, on lui appliquera un algorithme de compression d'images d'usage courant. Cela modifie un peu certains pixels, mais la compression, en simplifiant l'image, la débarrasse des caractéristiques inutiles que le réseau pourrait prendre en compte alors qu'elles ne sont qu'accidentelles. Ce lissage ramène à ce qui est essentiel dans l'image.

On se retrouve une fois encore dans une situation où le perfectionnement des armes conduit aux perfectionnements des protections. Quelle sera l'issue finale de cette coévolution du glaive et du bouclier? Nul ne le sait, mais cela fera globalement progresser la science de l'apprentissage automatique.

LE POSSIBLE ET L'IMPOSSIBLE

Les réseaux de neurones artificiels sont formidables pour une catégorie de tâches particulières que Andrew Ng, chercheur principal chez Baidu, entreprise chinoise, décrit comme suit: «Si un individu moyen est capable de réaliser une certaine tâche en moins de une seconde, alors il est probable que vous pourrez l'automatiser.» En revanche, leur mode de fonctionnement et leur programmation par apprentissage à base d'exemples entraînent des caractéristiques dont plusieurs sont des limitations.

Gary Marcus, professeur de psychologie à l'université de New York, énumère dans une prépublication (arXiv:1801.00631, 2018) dix problèmes liés aux réseaux de neurones profonds qui nous aident à percevoir ce qu'il faudrait faire pour aller au-delà.

1. L'apprentissage profond exige énormément de données: lors de la phase d'apprentissage, il faut donner au système plus d'exemples que n'en aurait besoin un humain.

2. L'apprentissage à l'aide de réseaux de neurones profonds est en réalité superficiel; l'utilisation de l'adjectif *profond* concerne la structure des réseaux utilisés, mais pas la nature de ce qui est appris et maîtrisé.

3. L'apprentissage profond ne sait pas, jusqu'à présent, traiter de manière directe les structures hiérarchiques ou impliquant une combinatoire complexe entre éléments (par exemple le fonctionnement d'un moteur de voiture).

4. L'apprentissage profond est inefficace pour mener des raisonnements logiques enchaînés, comme déduire de «Lucie est la mère de Jacques et la fille de Jean» que «Jacques est le petit-fils de Jean». On sait construire des méthodes symboliques (non neuronales) qui réussissent ces tâches de raisonnement qu'il faudra probablement associer aux méthodes neuromimétiques.

5. L'apprentissage profond est opaque. Lorsque le réseau a appris à mener une certaine tâche, des milliers de paramètres ont été fixés et, sauf dans les cas les plus simples, il est impossible de les interpréter et de comprendre pourquoi ils ont les valeurs retenues: ils vous donnent une réponse, contentez-vous en!

6. L'apprentissage profond ne permet pas, jusqu'à présent, d'intégrer des connaissances *a priori* ou formulées abstraitement, par exemple du type: «Si sur le dos de l'animal il y a une selle, il est assez probable que c'est un cheval.» Les humains ne comprennent pas tout à l'aide d'exemples, ils intègrent aussi des connaissances élaborées par d'autres auxquels ils font confiance, leurs enseignants par exemple.

7. L'apprentissage profond ne distingue pas une corrélation d'une cause. Il repère des liens, mais n'est pas en mesure de les analyser.

8. L'apprentissage profond suppose un monde stable, ce qui est ennuyeux dans un univers en mouvement ou en évolution rapide.

9. L'apprentissage profond ne garantit pas les résultats. La découverte de méthodes insoupçonnées qui engendrent des exemples pièges en a été une preuve spectaculaire.

10. L'apprentissage profond n'est pas facile à mettre en œuvre, car on ne sait pas à l'avance ce qui marchera, et à chaque nouvelle application, il faut ajuster les paramètres généraux (profondeur du réseau, nombre de neurones, technique d'apprentissage, etc.) sans savoir *a priori* ce qui fonctionnera le mieux.

Vive l'intelligence artificielle, et bravo pour ses succès récents grâce à l'apprentissage profond... Mais restons lucides sur ce qui est fait et reste à faire avant de parler de machines en tout point meilleures que les humains. ■

BIBLIOGRAPHIE

N. Akhtar et A. Mian, **Threat of adversarial attacks on deep learning in computer vision : A survey**, prépublication arXiv:1801.00553, 2018.

M. Alzantot *et al.*, **Did you hear that ? Adversarial examples against automatic speech recognition**, prépublication arXiv:1801.00554, 2018.

Y. Bengio, **La révolution de l'apprentissage profond**, *Pour la Science Hors-Série* n° 98, pp. 43-48, février-mars 2018.

A. Géron, **Machine Learning avec Scikit-Learn** et **Deep learning avec TensorFlow**, Dunod, 2017.

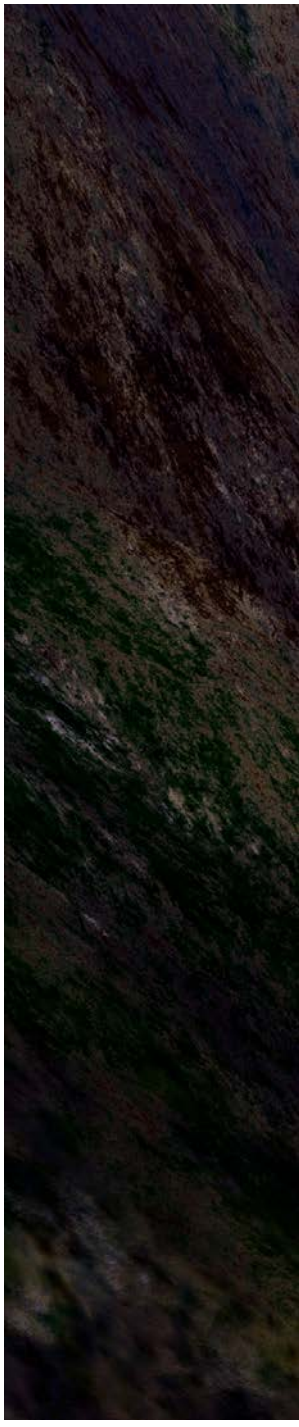
G. K. Dziugaite *et al.*, **A study of the effect of jpg compression on adversarial images**, prépublication arXiv:1608.00853, 2016.

A. Ng, **What artificial intelligence can and can't do right now**, *Harvard Business Review*, vol. 9, 2016.

C. Szegedy *et al.*, **Intriguing properties of neural networks**, International Conference on Learning Representations, 2014, prépublication arXiv:1312.6199, 2014.

L'AUTEUR

LOÏC MANGIN
rédacteur en chef adjoint à *Pour la Science*



UN PAYSAGE PLUS VRAI QUE NATURE

Un artiste espagnol détourne un logiciel de création de paysages tridimensionnels pour composer des mondes virtuels réalistes à partir de toiles de peintres classiques. Invitation au voyage dans l'irréel.

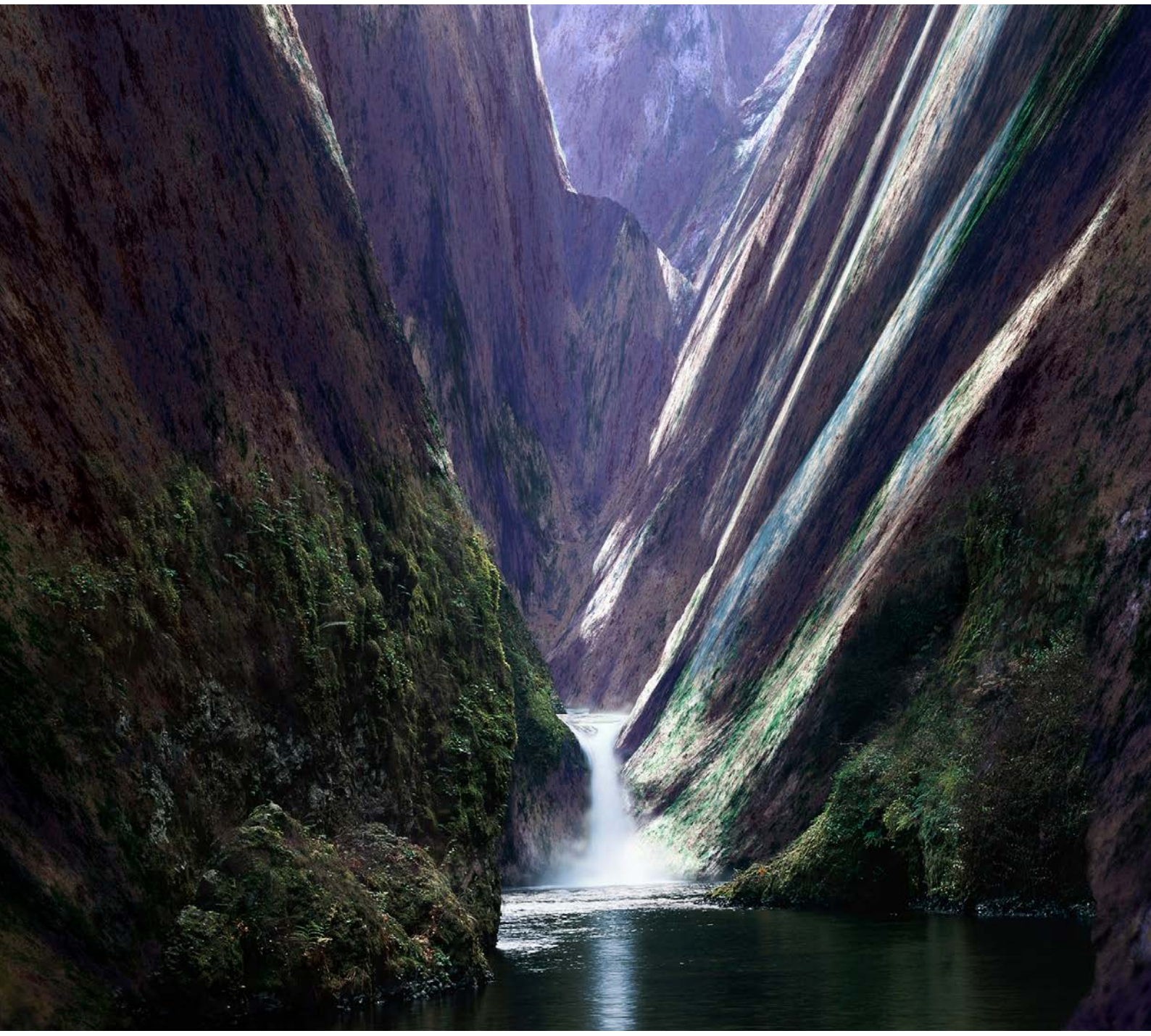
Le *Bosquet* d'André Derain interprété en un paysage 3D par un logiciel informatique.

André Derain, pionnier du fauvisme au début du XX^e^ siècle avec Henri Matisse, Maurice de Vlaminck, Georges Braque… est une figure incontournable de l'art moderne. *Bosquet*, l'une de ses toiles, peinte en 1912 et conservée au musée de l'Ermitage, à Saint-Pétersbourg, en Russie, est visible au Grand Palais, à Paris, jusqu'à cet été (*voir page ci-contre*). Pas convaincu par ce que vous voyez? Rien d'étonnant: l'œuvre (*voir ci-dessus*) est passée entre les mains de l'artiste espagnol Joan Fontcuberta. Plus déconcertant encore, elle est présentée dans le cadre d'une exposition intitulée *Artistes & Robots*! Le mystère s'éclaircira.

La manifestation propose une quarantaine d'œuvres pour lesquelles les artistes se sont adjoint l'aide de machines, de robots, de programmes informatiques, d'intelligence artificielle. Des machines à peindre bricolées par le Suisse Jean Tinguely dans les années 1960 jusqu'aux œuvres les plus récentes conçues grâce à l'«apprentissage profond», l'autonomie du créateur humain s'estompe à mesure que l'on progresse dans l'exposition, laissant une place toujours plus grande à la technologie. De même, l'interactivité avec le public va croissant. L'œuvre s'échappe et vit sa propre vie.

Chemin faisant, les questions surgissent, nombreuses. Une intelligence artificielle peut-elle avoir de l'imagination? Que peut faire un robot qui soit inaccessible à un humain? Une œuvre d'art est-elle nécessairement d'origine humaine? Jusqu'où une machine peut-elle revendiquer le statut d'artiste? Et bien d'autres encore.

Et l'on découvre le travail de Joan Fontcuberta. Dans son projet *Orogenesis* (qui concerne la naissance des montagnes), il utilise des logiciels, notamment Terragen, qui traduisent des informations bidimensionnelles en paysages tridimensionnels virtuels et réalistes. Ces programmes ont d'abord été mis au point pour des applications



J. Fontcuberta, Orogenesis : Derain, 2004, Tirages C-print sur Dibond, 120 x 160 cm. © Adagp, Paris 2018

militaires et scientifiques, par exemple pour reconstruire des reliefs à partir de données satellitaires ou cartographiques. Mais l'artiste les détourne.

Il leur soumet des toiles de Paul Cézanne, William Turner, John Constable, Caspar David Friedrich… et donc d'André Derain, et contraint les logiciels à les interpréter afin qu'ils créent des mondes parallèles. Les contours et les tons des peintures deviennent des montagnes, des rivières, des vallées, des cascades, des nuages. Les paysages qui en résultent, splendides, extrêmement réalistes et plausibles, mais de pure fiction, sont ensuite photographiés, et ce sont ces clichés qui sont à voir au Grand Palais.

Selon Joan Fontcuberta, il s'agit d'« hallucinations contemporaines définies par la technologie ». Les machines acquerraient ainsi une capacité d'imagination, et de fait, l'artiste n'est ici que l'instigateur, le travail de production étant délégué à l'ordinateur.

Le choix du médium n'est pas anodin. En effet, la photographie a longtemps été considérée comme le reflet de la réalité, un témoin digne de foi. Or là, Joan Fontcuberta brouille les frontières entre réel et imaginaire, entre vérité et mensonge. Il le revendique : il souhaite semer le doute, détruire les certitudes, annihiler les convictions… ■

« Artistes & Robots »,
au Grand Palais, à Paris, jusqu'au 9 juillet 2018.

Retrouvez la rubrique Art & science sur www.pourlascience.fr

LES AUTEURS

JEAN-MICHEL COURTY et ÉDOUARD KIERLIK
professeurs de physique à Sorbonne Université, à Paris

LE NEC PLUS ULTRA DE LA CHUTE LIBRE

La situation où un objet tombe sous la seule influence de la gravitation, sans perturbations externes, est impossible à réaliser, même en orbite autour de la Terre. Mais la mission spatiale *Microscope* s'en approche en compensant les forces résiduelles.

Dans l'air, deux objets différents chutent rarement de la même façon : la traînée aérodynamique a une grande influence.

Le 2 avril 2018, la station spatiale chinoise *Tiangong-1* se désintégrait dans l'atmosphère après une interminable chute. Depuis mi-2016 et la coupure des communications avec les opérateurs au sol, il n'était plus possible de maintenir son altitude en actionnant à distance ses moteurs. Moralité : même à 400 kilomètres d'altitude, la station n'était pas en parfaite chute libre le long de son orbite, mais subissait, outre la gravité, des forces résiduelles qui la freinaient et qui l'ont finalement fait tomber vers le sol.

Cet événement illustre une question qui se pose aussi dans d'autres contextes : comment s'approcher au mieux d'une chute libre, c'est-à-dire d'un mouvement dû uniquement à la gravitation ? Y répondre est crucial par exemple pour vérifier avec précision un principe fondamental de la physique, le « principe d'équivalence », qui implique que tous les corps chutent de la même façon.

Depuis Galilée, en effet, les physiciens cherchent à tester avec la meilleure précision possible si des corps de compositions différentes tombent de la même façon dans le vide. Comme Newton l'a ensuite compris, cela revient à vérifier l'égalité entre la « masse pesante », qui caractérise la manière dont un corps est attiré gravitationnellement par d'autres corps, et la « masse inerte », qui caractérise la difficulté qu'il y a à modifier sa vitesse en lui appliquant une force.

COMPARER LA CHUTE D'UNE PLUME ET CELLE D'UNE BILLE

À l'air libre, une telle expérience est vouée à l'échec : à cause du frottement de l'air, une plume tombe bien plus lentement qu'une bille de plomb. Ainsi, avec une feuille de papier de format A4 faisant 80 grammes par mètre carré qui tombe à plat, la traînée aérodynamique compense le poids pour une vitesse de chute de l'ordre de 1 mètre par seconde, une valeur atteinte et dépassée en un dixième de seconde par des objets plus compacts.

La force de traînée est proportionnelle au carré de la vitesse de l'objet par rapport à l'air, à la masse volumique de l'air et à la superficie que présente l'objet dans la direction du mouvement. Comment la diminuer ?

Une solution est de réduire la masse volumique de l'air, autrement dit de faire le vide. Le gain est alors considérable : sans atteindre les valeurs de l'ultravide (10^{-7} pascal) réservées généralement à des cavités