IA
pour le multimédia et au-delà

Recherches de pointe et perspectives de collaborations

Stéphane DUPONT

Head of Machine Intelligence Research, Numediart Institute

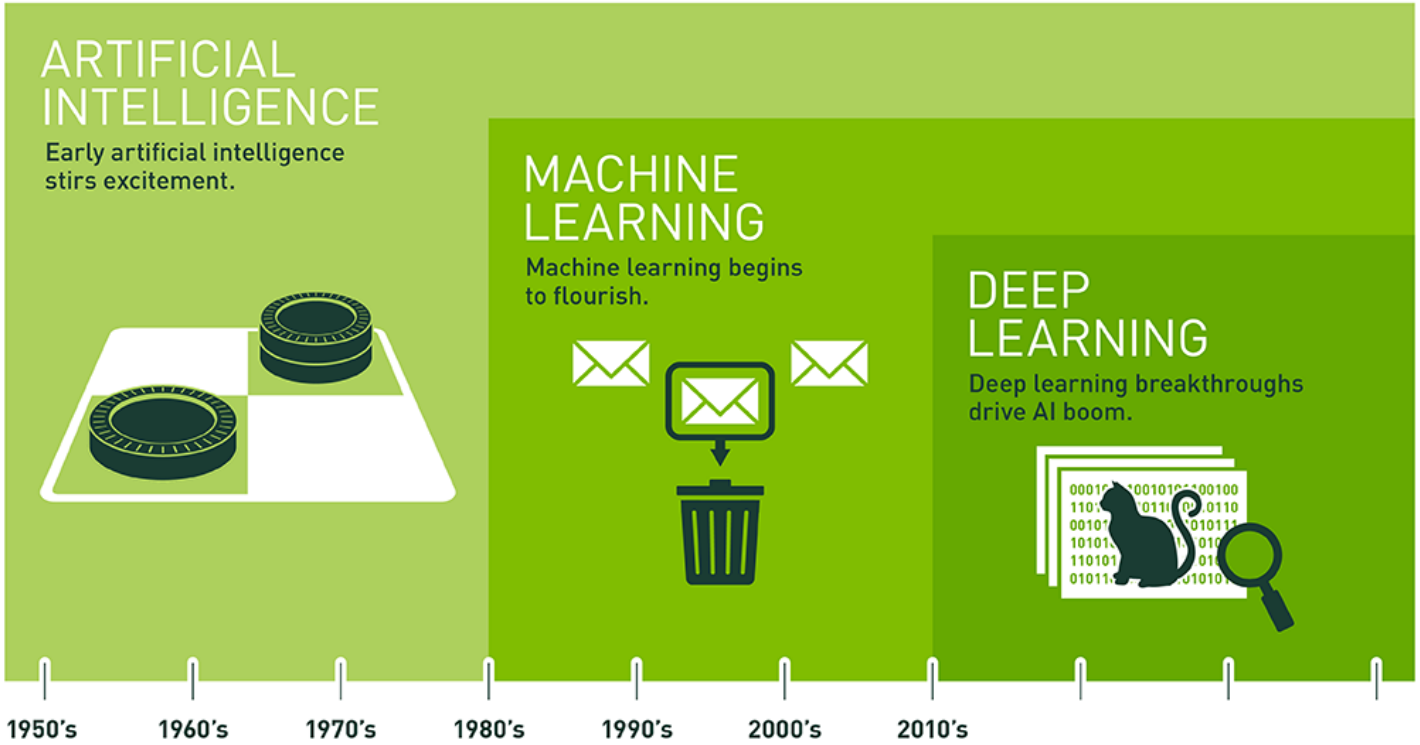
~~IA~~ Deep Learning
pour le multimédia et au-delà

Recherches de pointe et perspectives de collaborations

Stéphane DUPONT

Head of Machine Intelligence Research, Numediart Institute

Le Deep Learning, c'est quoi?



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Sources: https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

Le Deep Learning, c'est quoi?

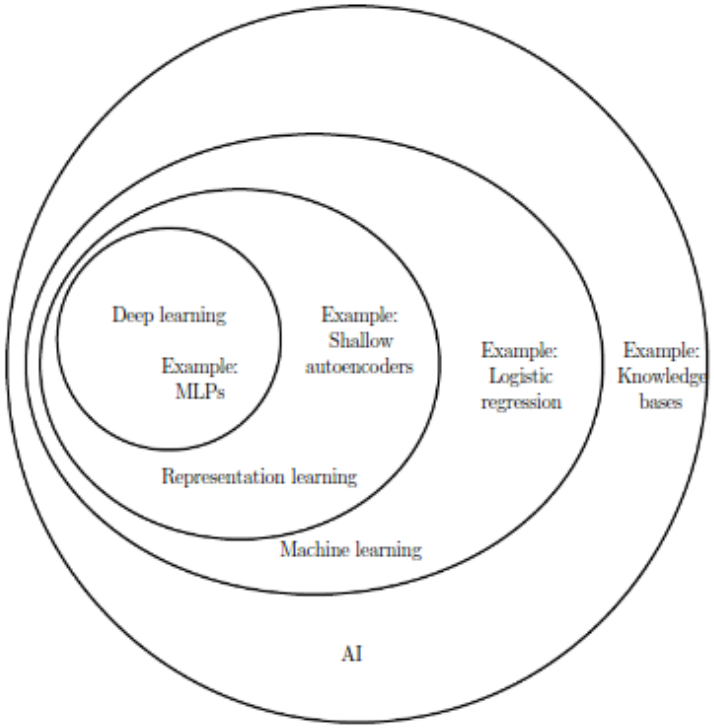
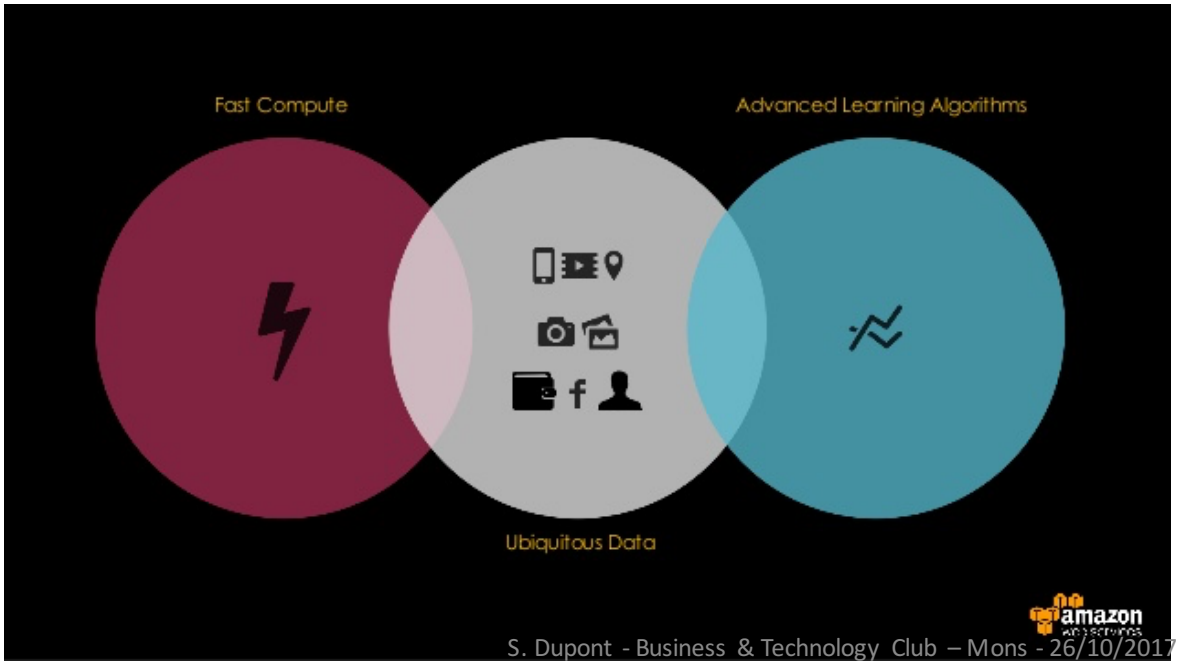


Figure 1.4: A Venn diagram showing how deep learning is a kind of representation learning, which is in turn a kind of machine learning, which is used for many but not all approaches to AI. Each section of the Venn diagram includes an example of an AI technology.

Sources: Ian Goodfellow and Yoshua Bengio and Aaron Courville*, "Deep Learning", MIT Press, 2016*

Le Deep Learning, comment ça marche?

- Estimer les (100aines de millions de) paramètres d'un "modèle" (ou même un "algorithme")
  - sur base de gros volumes d'exemples de données représentatives… (big data and big **dimensionality**)
  - en utilisant des algorithmes d'apprentissage avancés
  - nécessitant des capacités de calcul importantes.



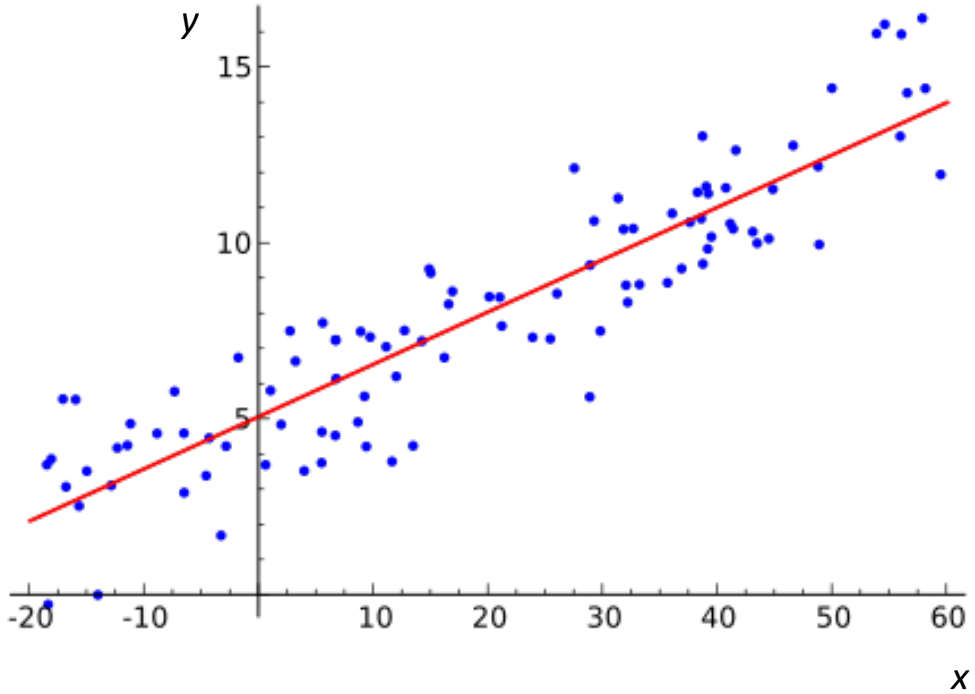
Data: *avantage aux industries centrées sur les données (actuellement surtout marketing)*

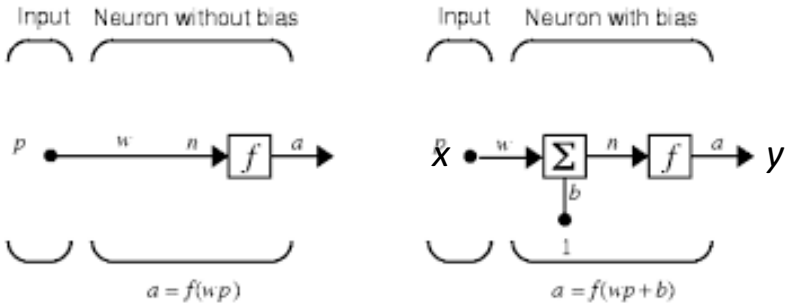
Algorithmes: *le sujet de recherche à faire progresser encore, même si tous les experts utilisent à peu près les mêmes algorithmes*

Substrat de Calcul: *redéploiement de l'industrie micro-électronique: GPU, neuro-morphique, cloud*

Le Deep Learning, comment ça marche?

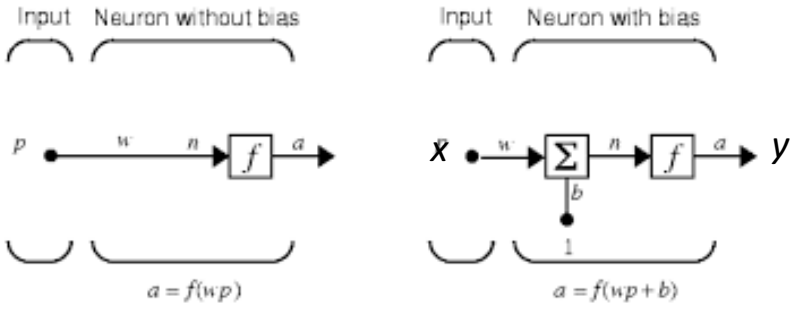
- Ca veut dire quoi: estimer les (millions de) paramètres d'un "modèle/algorithme"
  - $y = f(x)$ x = inputs & y = outputs
  - Un cas simplissime, la régression linéaire

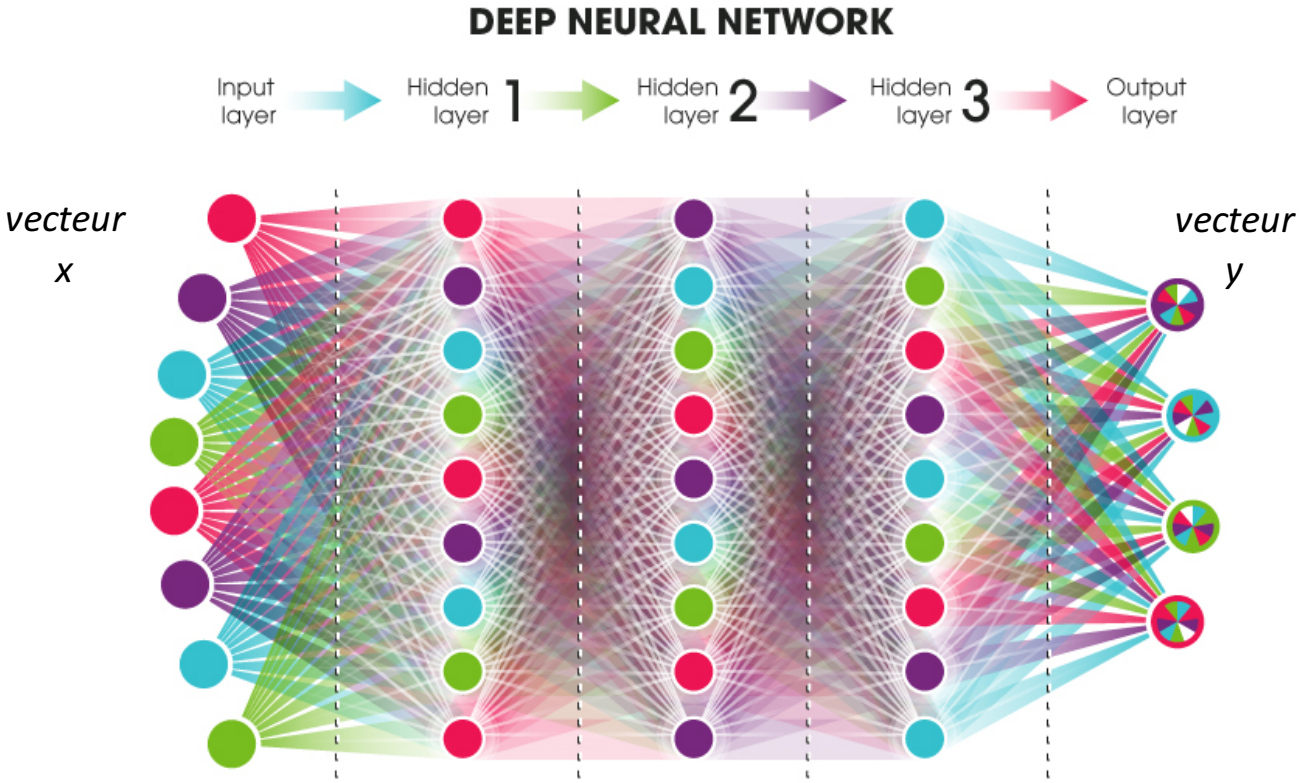




Le Deep Learning, comment ça marche?

- C'est simple et compliqué à la fois: des 10ne de millions de neurones





neuralnetworksanddeeplearning.com - Michael Nielsen, Yoshua Bengio, Ian Goodfellow, and Aaron Courville, 2016.

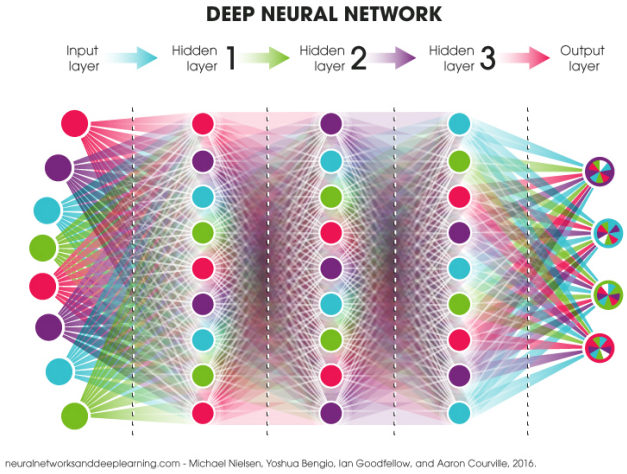
Le Deep Learning, comment ça marche?

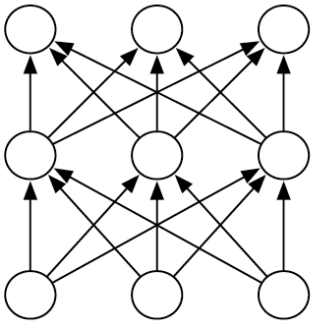
- $y = f(x)$ x = inputs & y = outputs
- Cas d'application réels
  - (**reco parole**) *x* représente un enregistrement vocal, et *y* la transcription de ce qui est dit
  - (**reco image**) *x* représente une image, et *y* une phrase décrivant cette image
  - (**segmentation image**) *x* représente une image, et *y* l'objet sous chaque pixel
  - (**édition image**) *x* représente une image, et *y* une image dans un style pictural
  - (**langage/traduction**) *x* représente une phase en FR, et *y* sa traduction en EN
  - (**immobilier**) *x* représente la surface et localisation d'un bien, *y* son prix
  - (**véhicule autonome**) *x* représente les signaux des senseurs, *y* le contrôle de la colonne de direction
  - (**médical**) *x* représente les signaux vitaux en soins intensifs, *y* la probabilité de mortalité
  - (**logistique**) …
  - (**finance**) …
  - (**énergie**) …
  - (**audit**) …
  - (**chatbots**) …

- **Dimensionalités** très importantes des vecteurs d'entrée/sortie *x* et *y*
- Fonction f fortement **non-linéaire**
- Implique besoin de:
  - Grand nombre d'exemples, bcp. de données (10^{ne} de millions)
  - Fonction f avec nombreux paramètres (idem)
  - Du coup, volume de calcul très important (mais fortement parallélisable)

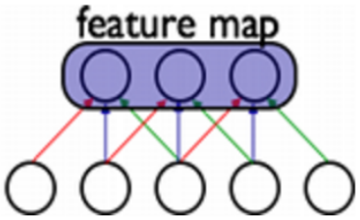
Le Deep Learning, oui mais….

- Architectures/algorithmes de réseaux de neurones profond:
  - Réduire le nombre de paramètres
  - Partager les paramètres entre différents neurones
  - Ajouter du bruit ou imaginer de nouvelles données durant l'apprentissage

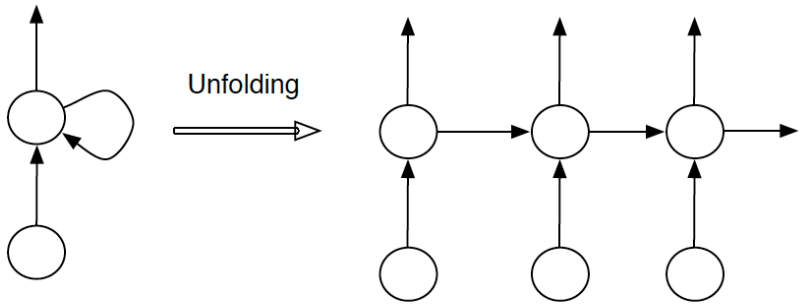




Feed-forward neural network

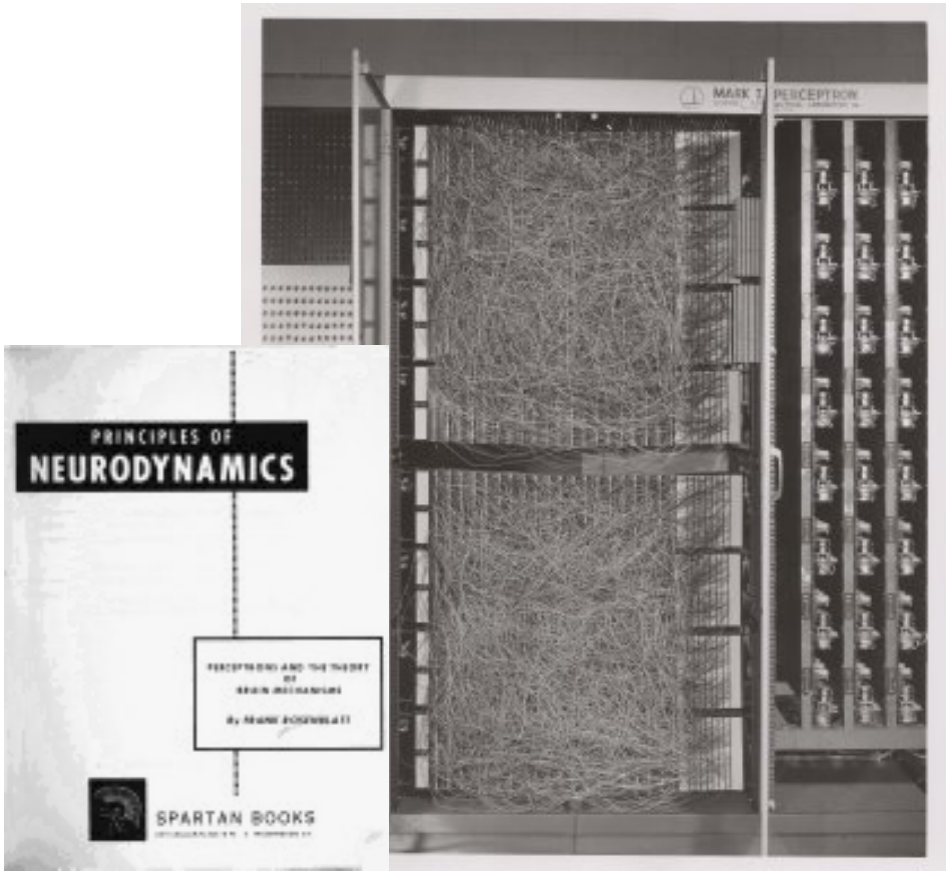


Convolutional neural network



Recurrent neural network

60ème anniversaire...



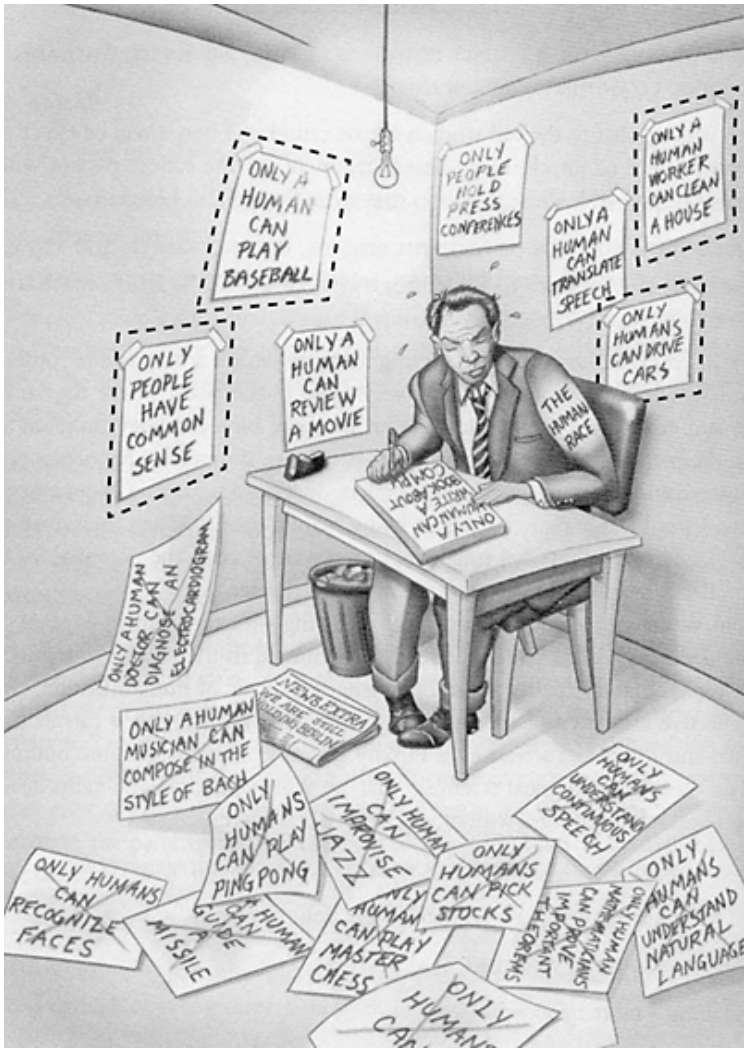
Frank Rosenblatt

400 photocells connected to neurons by wires and potentiometers moved by motors to adjust the weight of the connections and biases of the perceptrons (1957)

Nowadays, you can have 10 of millions of these interconnected running real time in a single chip (GPU/FPGA/…) (2017)

Le Deep Learning, ça marche vraiment?

- Mieux que l'être humain.

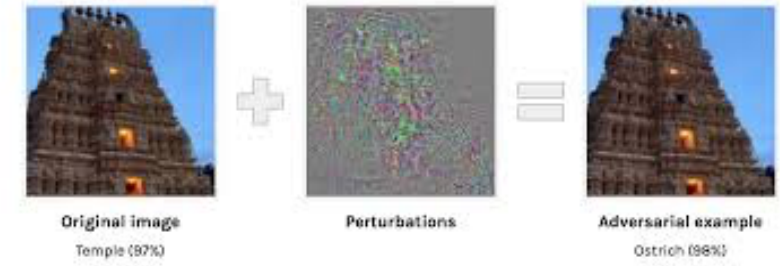


Source: Kurzweil, 1999

Le Deep Learning, ça marche vraiment?

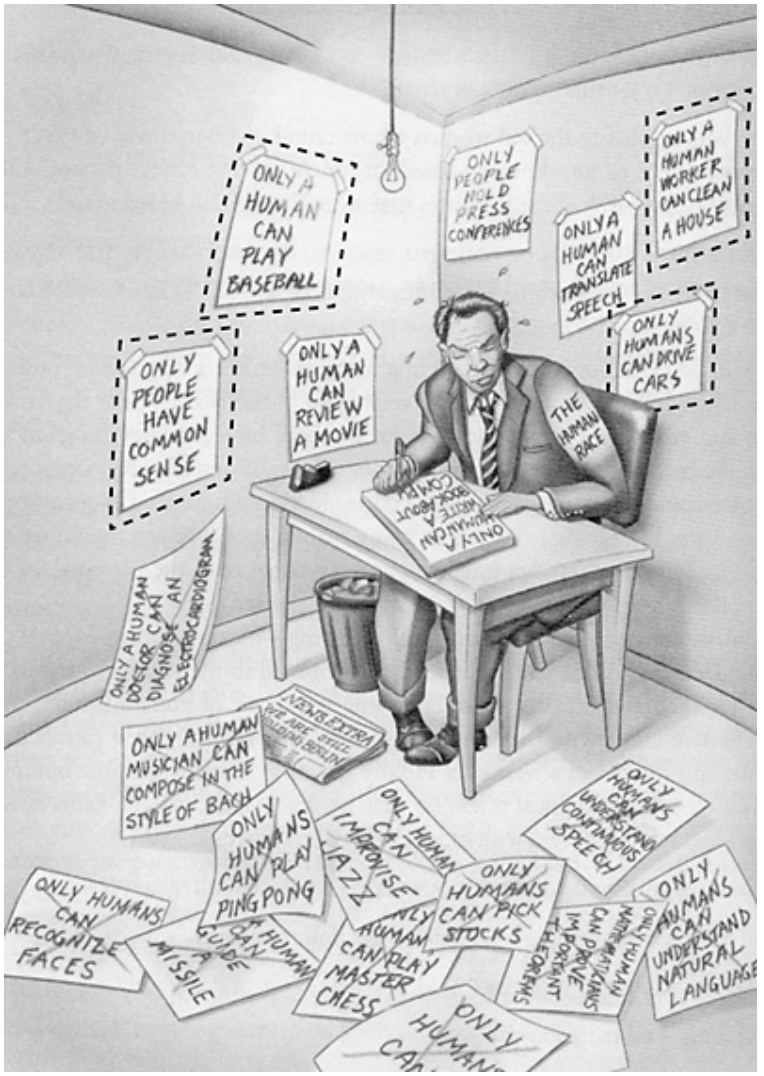
Le Deep Learning, ça marche vraiment?

- Idiots savants – “Narrow AI”
- Failles de “sécurité”: exemples contradictoires (« adversarial »)



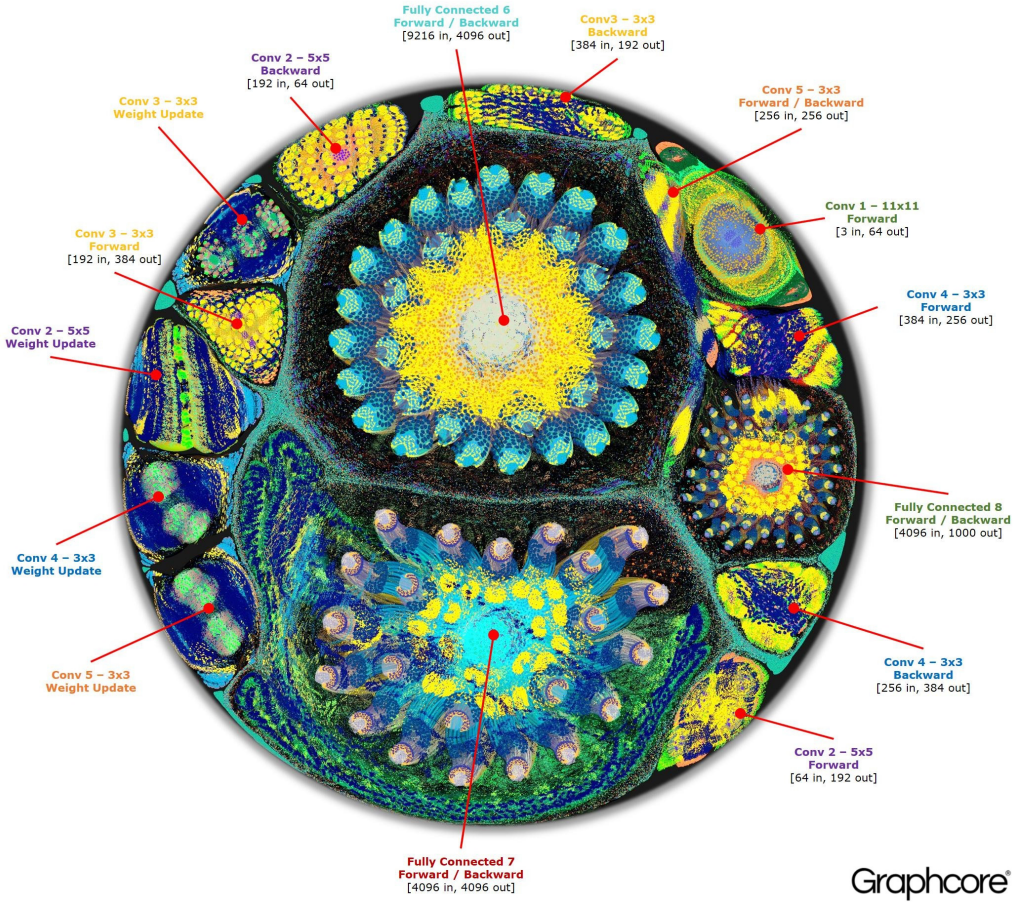
- Préjugés: IA sexiste, raciste, etc.
- Que faire pour les problèmes où les données sont incomplètes?

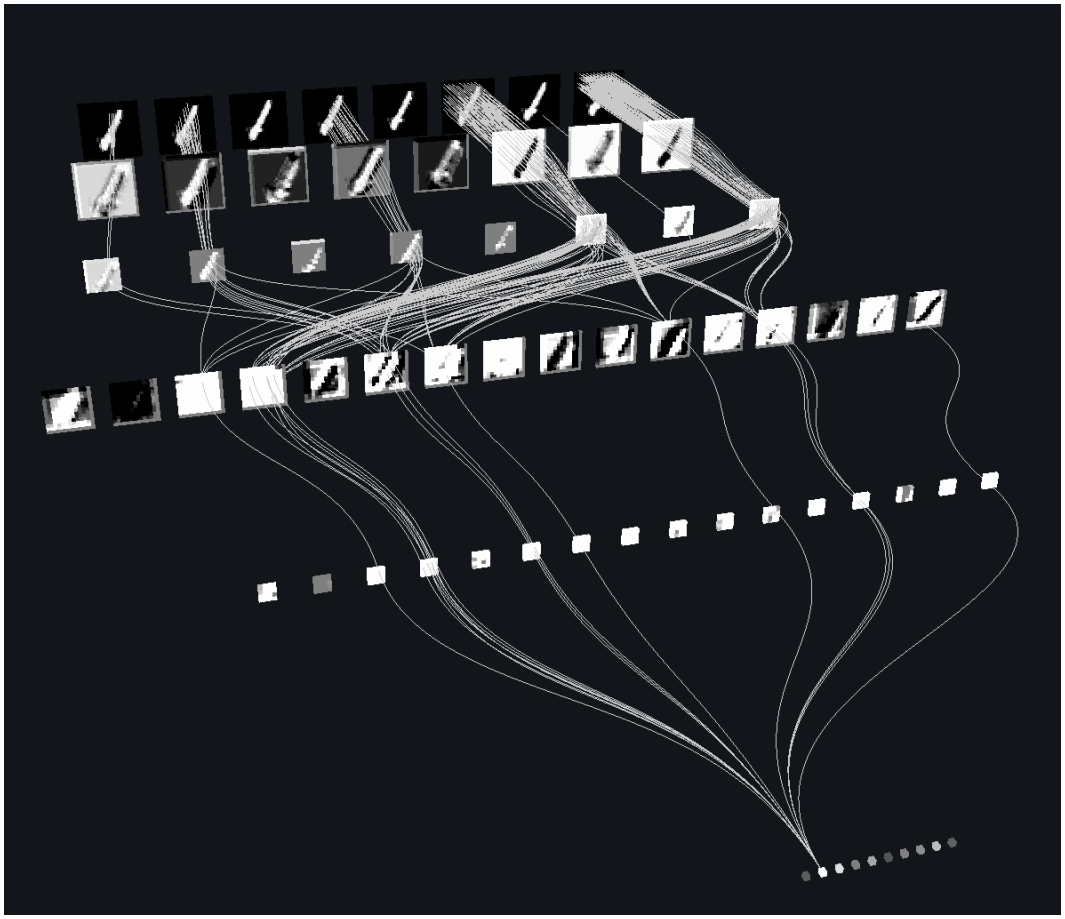
- Boîtes noires



Source: Kurzweil, 1999

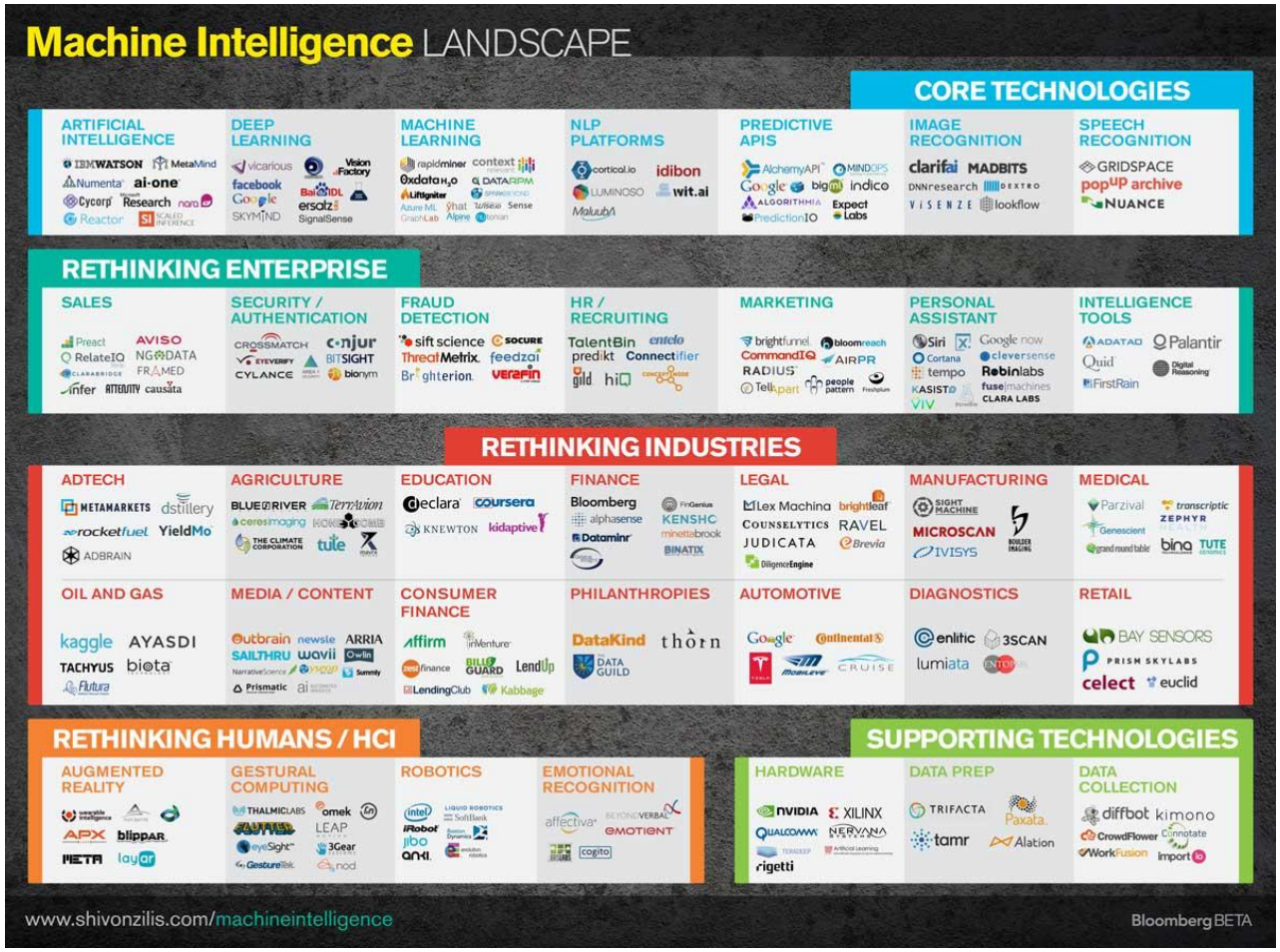
Le Deep Learning, des boites noires?





Sources: http://terencebroad.com/convnetvis/vis.html

Le Deep Learning, une industrie?



Secteurs porteurs

Domaines de l'intelligence artificielle les plus porteurs de croissance d'ici à 2025

1. Détection et identification d'objets
   (Voitures autonomes)
2. Identification, tag et classement d'images
   (Reconnaissance des personnes ou des paysages)
3. Analyse de données médicales
   (Aide au diagnostic pour les maladies graves)
4. Commerce et trading algorithmique
   (Ciblage de clients, trading financier)
5. Localisation et cartographie
   (Aide à la navigation)
6. Maintenance prédictive
   (Anticipation des réparations nécessaires aux outils industriels)

Sources: LE MONDE, IBM, MCKINSEY, DELOITTE, TRACTICA, STOCKROW

Deep Learning à l'UMONS - Historique

- Précurseurs en RNA (pour la parole, et lecture labiale) dès 1995: projets EU Wernicke, Sprach, Respite, Divines, …
- Précurseurs en DNNs dès 2000: architecture multi-bande (2000), adaptation rapide des poids (2003), auto-encodeur multilingue (2005), ….
- Renforcement des équipes depuis 2013: 12 personnes (8 thèses)
- Collaborations:
  - avec laboratoires reconnus internationalement: Montreal MILA, Sherbrooke Necotis, plusieurs équipes INRIA.
  - avec industrie via financements Wallinov.

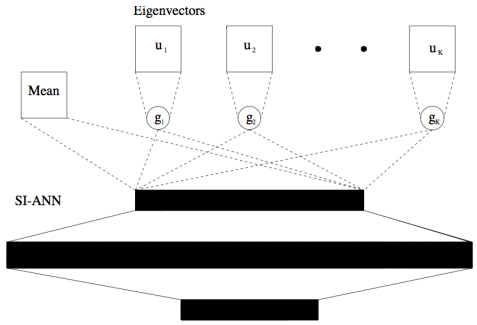
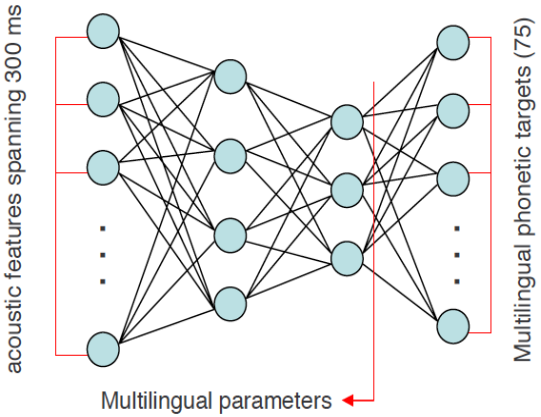


Figure 1: Fast Adapted ANNs



Deep Learning à l'UMONS – Activités en Cours

- Parole, Langage, Affect
  - Reconnaissance vocale
  - Réduction de bruit
  - Synthèse vocale expressive
  - Reconnaissance des expressions et émotions
  - Reconnaissance de gestes
  - Analyse de scènes complexes (« situation awareness ») pour agents interactifs
  - Traduction automatique et réponse à des questions
- Multimédia
  - Identification, étiquetage et classement de sons, de musique, d'images, de croquis, etc.
  - Hachage sémantique profond pour indexation/recherche rapide
  - Recherche d'images par similarité ou par croquis (perf. surhumaines)
  - Attention visuelle
  - …
- Nouveaux résultats de recherche concrets

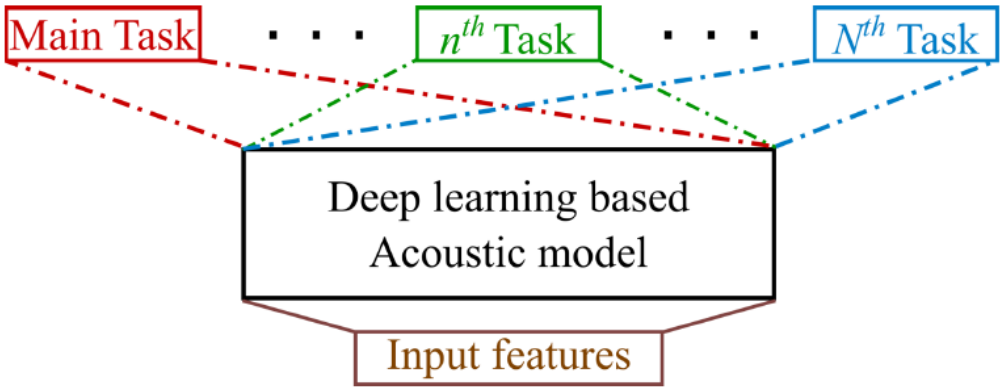
  → protocoles d'apprentissage (ex: multi-tâche, quadruples, ranking)

  → combinaisons de modules DNN (ex: attention, séquence-vers-séquence, modélisation conditionnelle).

  → fusion vision et langage (« grounding », incarnation des connaissances) comme nouvelle façon d'aborder l'intelligence artificielle située.

Compréhension/Synthèse de Signaux Humains

- Reconnaissance vocale
- Débruitage vocal “intelligent”
- Synthèse vocale expressive



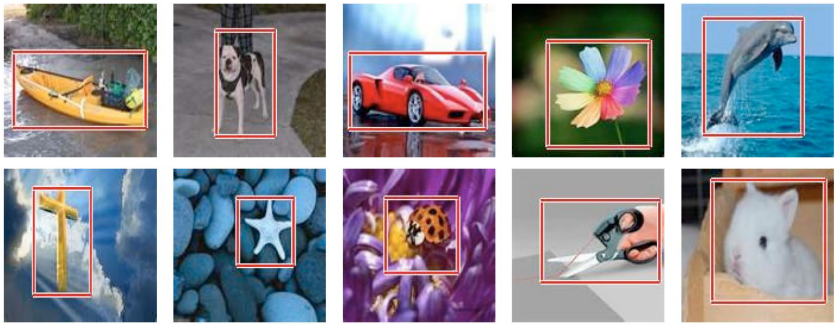
- Au-delà du langage verbal:
  - Reconnaissance et Synthèse de signaux para-linguistiques et extra-linguistiques: amusement, rire, etc…
  - Reconnaissance multimodale d’opinions: personnes exprimant leur avis sur un produit / un sujet: approche multimodale combinant les mots choisis, l’intonation vocale, et les expressions faciales

Delbrouck, J. B., El Haddad, K.& Dupont, S. (2017). An RNN-Based Multimodal Sentiment Analysis: Focusing on Facial Expressions and Multimodal Dynamic Representations. Submitted to ICMI 2017.

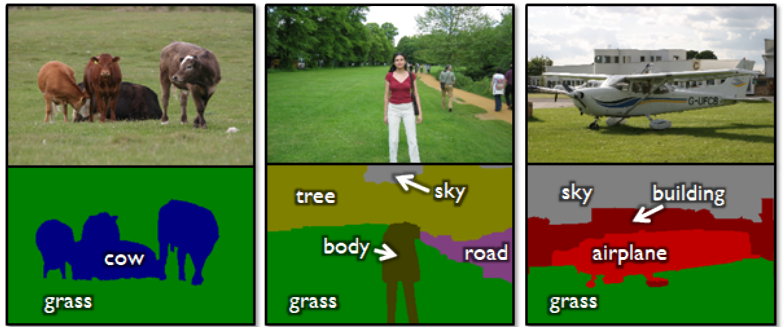
Compréhension visuelle – vision par ordinateur (captions, sous-titres)

- Problèmes abordés par les chercheurs sont de plus en plus complexes: reco. d'objet (un seul objet sur l'image), segmentation des différents objets, descriptions riches (objets, attributs, relations...), réponses à des questions sur le contenu visuel, etc…

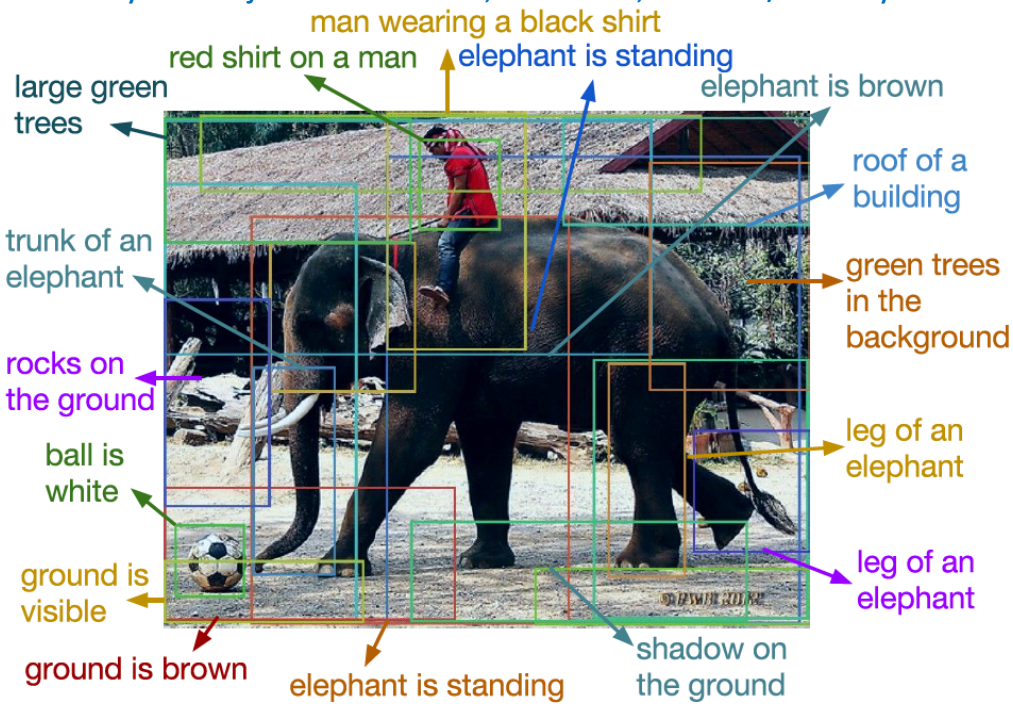
Object detection/localization



Object segmentation



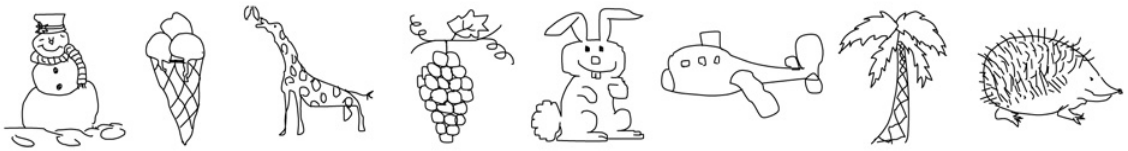
Beyond Objects: Attributes, Relations, Context/Activity



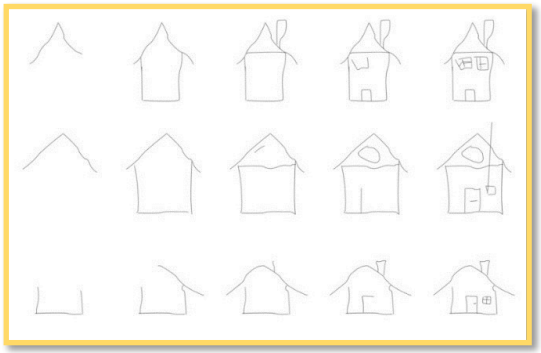
(Stanford Uni. 2016)

Compréhension visuelle – reconnaissance de croquis

- Une nouvelle référence internationale:
  - Précision > **80%** sur 250 catégories: utilisation du « deep learning » pour tenir compte de l'information spatiale, mais aussi temporelle, en cours de tracé.
  - **Premier labo (niveau mondial)** à avoir **dépassé les performances humaines** sur cette problématique.
  - Reconnaissance temps-réel en ligne: progressivement en cours de tracé.

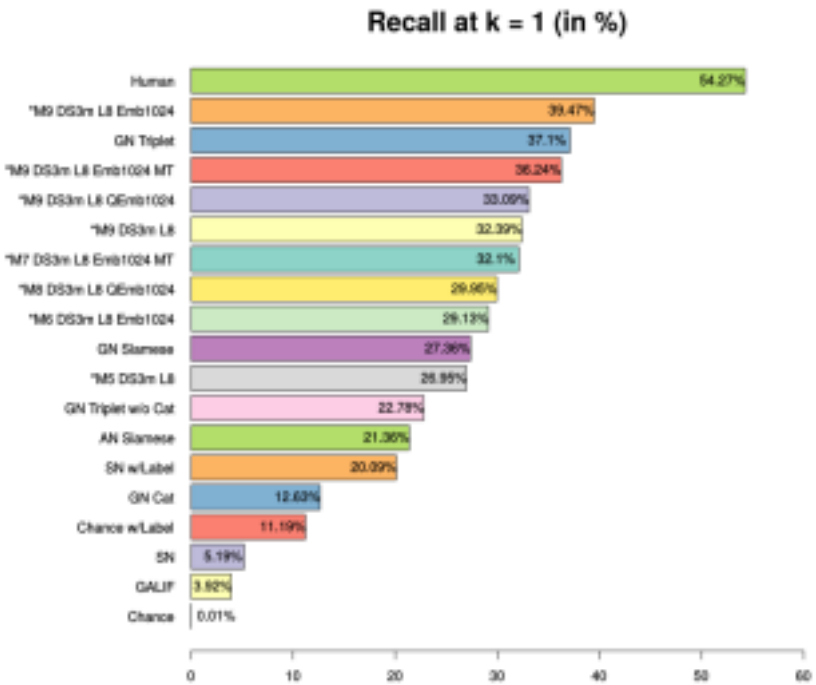


Reconnaissance automatique
de croquis partiels en temps-réel



Recherche d'images en la croquant

- On se rapproche des performances humaines (42% de résultats pertinents en top-1 vs. 54% pour l'humain)



Human

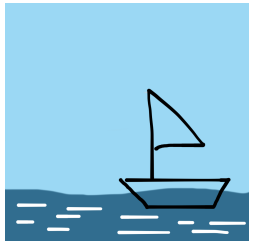
UMONS

Seddati, O., Dupont, S., & Mahmoudi, S. (2017). DeepSketch 3. Multimedia Tools and Applications, 1-27.

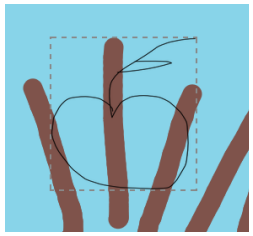
Intégration - Moteur de recherche vidéo
Partenariat projet EU iMOTION

- Lauréats du benchmark VBS 2017

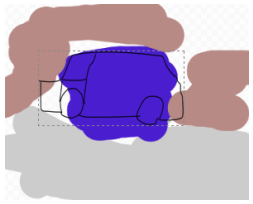
sailboat



apple

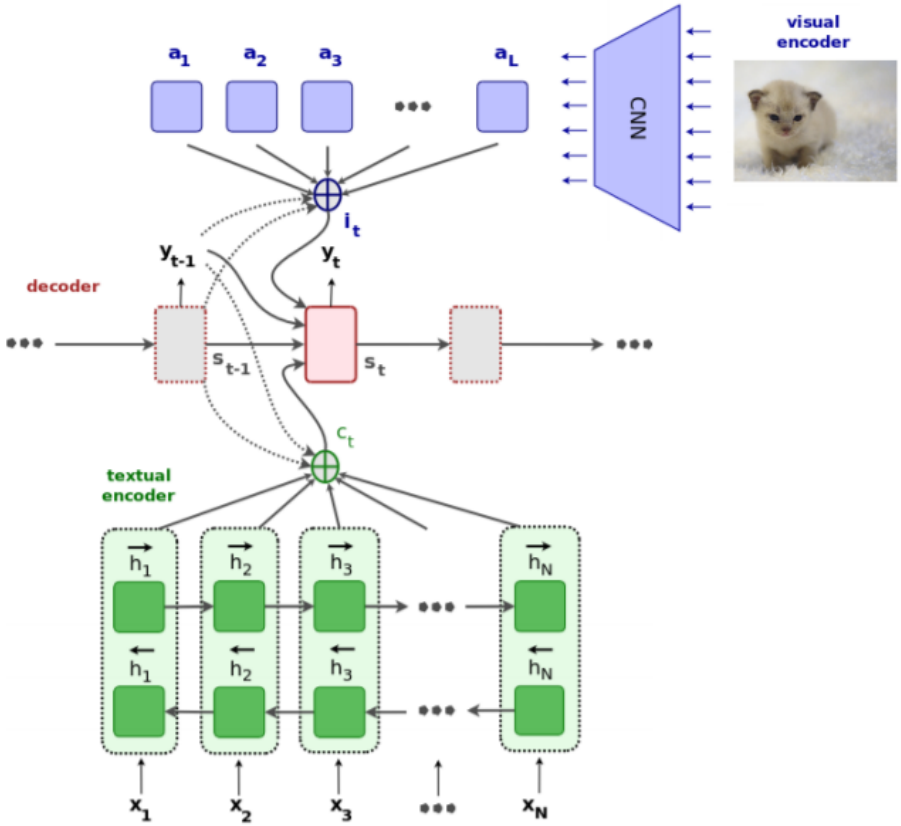


car



Langage (et vision) – traduction automatique

- Multimodalité
- Mécanismes d'attention émergents profonds
- Nous disposons d'un modèle qui constitue un **nouvel état de l'art**: métrique d'évaluation BLEU **28.01%**. Donc, près de 30% des séquences de 4 mots dans la phrase traduite sont identiques à une traduction humaine. En pratique cela produit des phrases tout à fait convaincantes.



Delbrouck, J. B., & Dupont, S. (2017). Multimodal Compact Bilinear Pooling for Multimodal Neural Machine Translation. Submitted to ICLR 2017
Seddati, O., Dupont, S., & Mahmoudi, S. (2017).Triplet Networks Feature Masking for Sketch-Based Image Retrieval 14th International Conference, ICIAR 2017, Montreal, QC, Canada, July 5–7, 2017, Proceedings
Delbrouck, J. B., & Dupont, S. (2017). An empirical study on the effectiveness of images in Multimodal Neural Machine Translation. Submitted to EMNLP 2017.

Langage (et vision) – traduction automatique

- Multimodalité
- Mécanismes d'attention émergents profonds
- Nous disposons d'un modèle qui constitue un **nouvel état de l'art**: métrique d'évaluation BLEU **28.01%**. Donc, près de 30% des séquences de 4 mots dans la phrase traduite sont identiques à une traduction humaine. En pratique cela produit des phrases tout à fait convaincantes.

Die beiden Kinder spielen auf dem Spielplatz .

Ein Mann sitzt neben einem Computerbildschirm .

Delbrouck, J. B., & Dupont, S. (2017). Multimodal Compact Bilinear Pooling for Multimodal Neural Machine Translation. Submitted to ICLR 2017

Seddati, O., Dupont, S., & Mahmoudi, S. (2017).Triplet Networks Feature Masking for Sketch-Based Image Retrieval 14th International Conference, ICIAR 2017, Montreal, QC, Canada, July 5–7, 2017, Proceedings

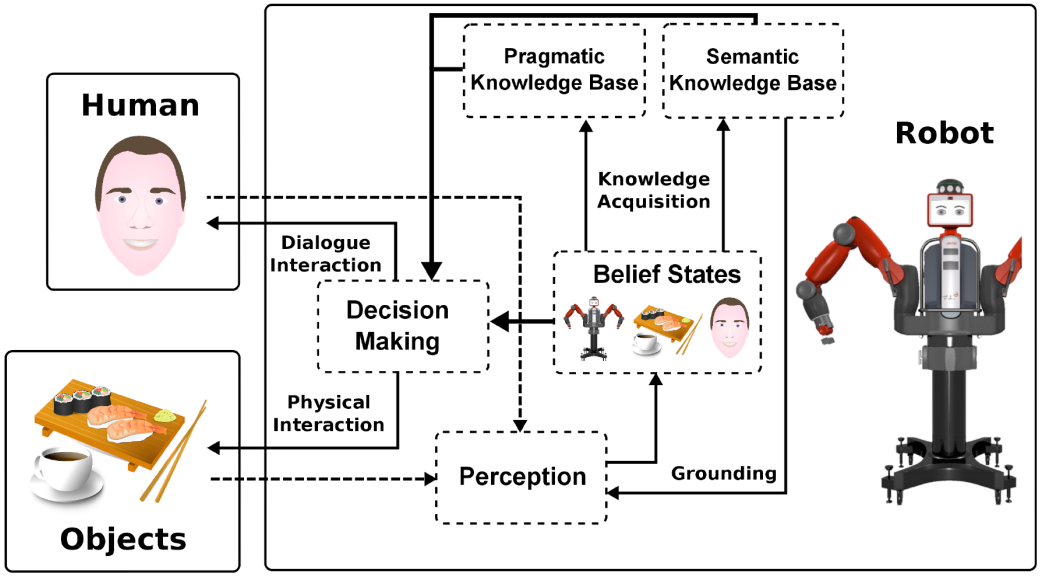
Delbrouck, J. B., & Dupont, S. (2017). An empirical study on the effectiveness of images in Multimodal Neural Machine Translation. Submitted to EMNLP 2017.

Grounding sémantique profond

- En très bref: il est maintenant possible d'améliorer des méthodes de traitement du langage naturel en permettant à la machine de comprendre (ou d'imaginer) la scène visuelle décrite.

- Compréhension réelle du langage

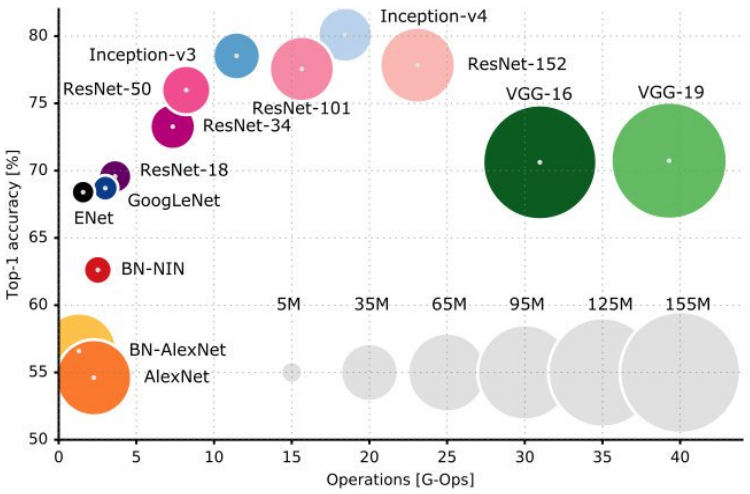
- Dans le cadre d'interactions:
  - Perception
  - Action
  - Modèle Cognitif



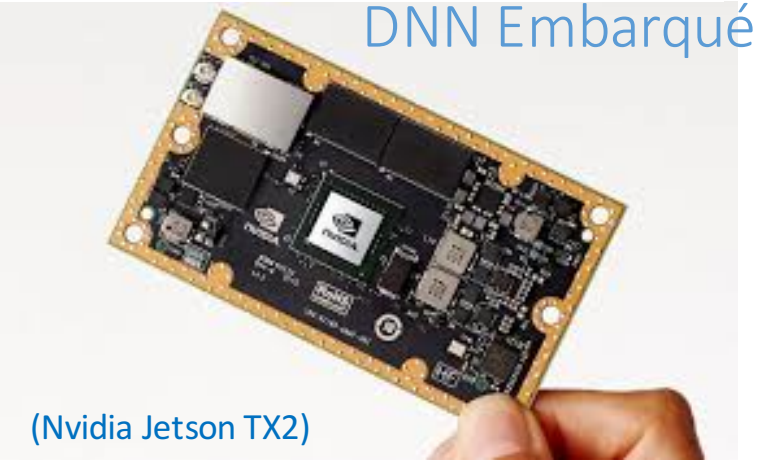
Delbrouck, J. B., & Dupont, S. (2017). Visually Grounded Word Embeddings and Richer Visual Features for Improving Multimodal Neural Machine Translation. Submitted to GLU 2017.

Autres thèmes et applications

- Intelligence embarquée: systèmes de détection d'événements spécifiques.
- Compréhension et raisonnement spatial par DNNs.
- Interaction avec des robots humanoïdes "affectifs".
- Combinaison de mécanismes d'apprentissage (deep learning) et de bases de connaissance (web sémantique)



DNN Embarqué



(Nvidia Jetson TX2)

Vers le Smart IoT

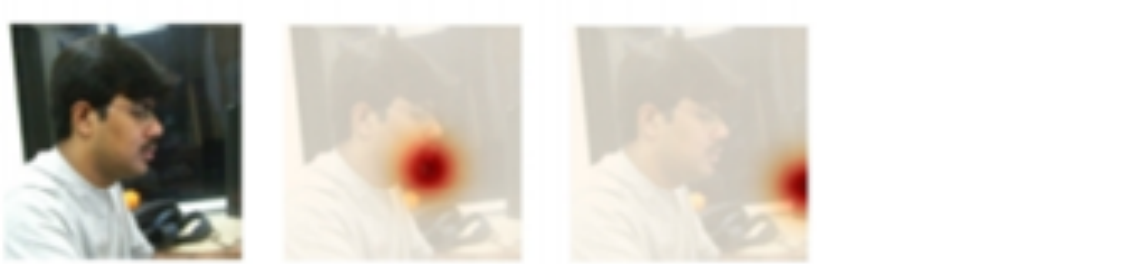
- Grand nombre de capteurs en continu:
  - Dimension des vecteur de données d'entrée importante
  - Temps-réel: volume, vélocité
  - Environnements non-stationnaires

- L'image, c'est facile! Il n'y a que deux dimensions… mais bcp. de pixels….

Intelligence artificielle - Demain

- Des machines qui ont du bon sens
- Des machines qui savent s'expliquer
- Auto-apprentissage

- Neuroscience/pédagogie/informatique
- Symboles & data
- "General AI"



Intelligence Artificielle Créative?

Sources: Gatys 2015

Une masse critique en R&D deep nets…

UMONS Université de Mons

- Living Lab CLICK via FEDER DigiSTORM (Industries Créatives)
- Co-Innovation via FEDER IDEES (Internet de Demain)
- AI meetups Mons
- Travaux de fin d'étude
- Thèses de doctorat cofinancés
- Financements Régionaux Wallinov
- Financements Européens H2020
- Collaboration ad hoc

Dites nous quels sont vos *x* et vos *y*…

UMONS
Université de Mons

… nous pourrions vous aider à trouver votre *f*

Let's make AI great again!

https://www.numediart.org
http://tcts.fpms.ac.be/~dupont/mir.html
stephane.dupont@umons.ac.be

Merci!