

Vers des systèmes perceptifs autonomes

Stéphane HERBIN

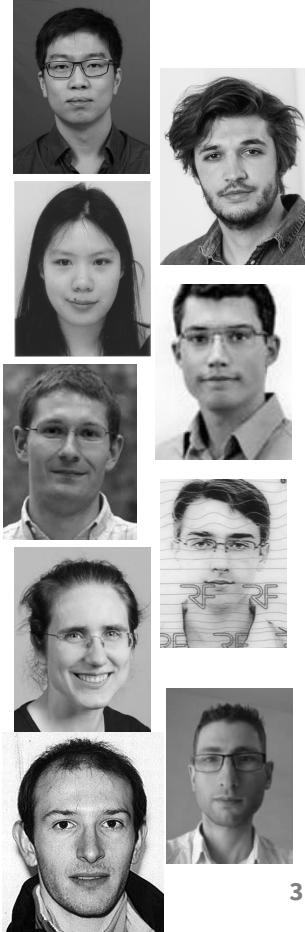
Soutenance HDR - 6/7/20

Parcours

1987-1990	Supélec	<i>Diplôme d'Ingénieur</i>
1990-1992	University of Illinois at Urbana-Champaign	<i>Master of Science</i>
1992-1997	Ecole Normale Supérieure de Cachan	<i>Service National + Doctorat</i>
1997-1998	Société SUDIMAGE	<i>Ingénieur de recherche</i>
1997-1998	Université de Marne La Vallée	<i>Enseignant</i>
1998-2000	MBDA (ex Aérospatiale Matra Missiles)	<i>Ingénieur de recherche</i>
2000-2020	ONERA	<i>Maître de recherche</i>

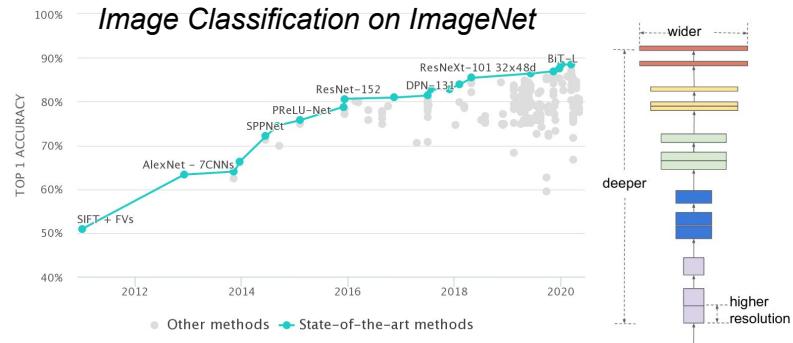
Doctorants

1. **Alexis Lechat (2018-2021)** : « Apprentissage incrémental des systèmes de question réponse visuels »
 - Co-encadrement avec Frédéric Jurie (Univ. De Caen/GREYC).
2. **Maxime Bucher (2015-2018)** : « Apprentissage et exploitation de représentations sémantiques pour la classification et la recherche d'images»
 - Co-encadrement avec Frédéric Jurie (Univ. De Caen/GREYC).
3. **Isabelle Leang (2013-2016)** : « Fusion en ligne d'algorithmes de suivi visuel d'objet »
 - Co-encadrement avec Jacques Droulez (CNRS/UPMC/ISIR) et Benoît Girard (CNRS/UPMC/ISIR).
4. **Cédric Le Barz (2012-2015)** : « Navigation visuelle pour les missions autonomes des petits drones »,
 - Co-encadrement avec M. Sanfourche (ONERA) et M. Cord (UPMC/LIP6)
5. **Joseph Defretin (2008-2011)** : « Stratégies de vision active pour la reconnaissance et d'objets »
 - Co-encadrement avec G. Le Besnerais (ONERA) et N. Vayatis (ENS de Cachan / CMLA)
6. **Christophe Guilmart (2008-2011)** : « Filtrage de segments informatifs dans des vidéos »
 - Co-encadrement avec P. Pérez (Technicolor)
7. **Anne-Marie Tousch (2006-2010)** : « Hiérarchies sémantiques pour l'annotation multifacette d'images»
 - Co-encadrement avec J.-Y. Audibert (ENPC / Certis)
8. **Jonathan Guinet (2004-2008)** : « Acquisition de cible sur séquences aériennes, étude de l'apport d'informations spatio-temporelles»
 - Co-encadrement avec G. Le Besnerais (ONERA) et S. Philipp-Foliguet (Univ. Cergy Pontoise)
9. **Benjamin francesconi (2003-2007)** : « Détection d'objets sur image satellitaire »
 - Co-encadrement avec B. Chalmond (ENS Cachan/ Univ. Cergy Pontoise)



Contexte

- Gain de performance du Deep Learning en traitement de données
- Utilisation de l'IA pour des applications réelles (critiques)? Quelle fiabilité?
- ONERA: mission de transfert de technologie → garantir ou améliorer la maturité
- Projet = Transformer cette question de maturité en une activité de recherche
- Comment l'exprimer?



[Tan & Le, 2019]



Investigators Say Tesla Model 3 Driver Killed In Florida Crash Was Using Autopilot



Alan Ohnsman Forbes Staff
Transportation

f
t
in



The roof of Jeremy Beren Banner's 2018 Tesla Model 3 was sheared off when the car struck a semi-truck trailer in Delray Beach, Florida, on March 1, 2019. [-] NATIONAL TRANSPORTATION SAFETY BOARD

(Updates with Tesla comments)

Safety officials investigating a deadly crash in Florida in which a Tesla Model 3 driver was killed in a collision with a commercial truck said the electric-car maker's semi-automated driving system was being used at the time and that neither it nor the driver took evasive action.

The National Transportation Safety Board issued preliminary findings from its investigation of the March 1 crash in Delray Beach that killed 50-year-old Jeremy Beren Banner, who died after his 2018 Model 3 hit a semi-truck that crossed his path on Florida State Highway 441 in early morning traffic. The Tesla, travelling at 68 miles per hour on a section of highway with a posted speed limit of 55 mph, slammed into the side of the truck's trailer, shearing the Tesla's roof off. The truck driver was unharmed.

Amazon Pauses Police Use of Its Facial Recognition Software

The company said it hoped the moratorium "might give Congress enough time to put in place appropriate rules" for the technology.



Civil liberties advocates began calling for a ban on the use of facial recognition by law enforcement in 2018. Elaine Thompson/Associated Press

By Karen Weise and Natasha Singer

June 10, 2020



APES: “Autonomous PErceptual System”

- Quel sont ses caractéristiques?
- Comment fonctionne-t-il?
- En quoi peut-il être fiable?
⇒ Quelles directions de recherche?

(Pourquoi est-ce un objet de recherche intéressant?)

Définir un APES

Définitions de Perception (wikipedia)

- **Français:** La perception est l'**activité** par laquelle un sujet fait l'**expérience** d'objets ou de propriétés présents dans son **environnement**. Cette activité repose habituellement sur des **informations** délivrées par ses **sens**.
- **Anglais:** La perception (du latin *perceptio*) est l'organisation, l'identification et l'**interprétation** des informations **sensorielles** afin de représenter et de comprendre les **informations** présentées ou l'**environnement**.
- **Espagnol:** La perception est la façon dont le **cerveau** détecte les **sensations** qu'il reçoit à travers les sens pour former une **impression consciente** de la réalité physique de son **environnement** (interprétation).
- **Italien:** La perception est le **processus psychique** qui opère la synthèse des données **sensorielles** sous des formes **significatives**.
- **Allemand:** La perception est le **processus** et le résultat de l'obtention d'**informations** et de traitements de **stimuli** de l'**environnement** et du corps propre.

Perception: deux dimensions

Condition à l'expérience du monde

- Conscience
- Sensation/Réceptivité
- Existence

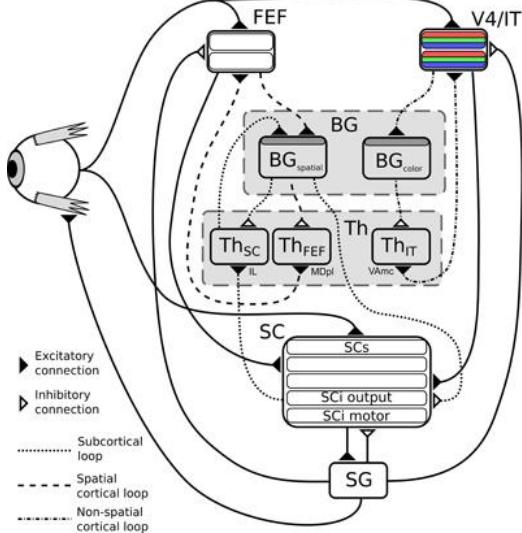
Expression de quelque chose sur le monde

- Processus/Activité
- Représentation/Information/Interprétation
- Vérité

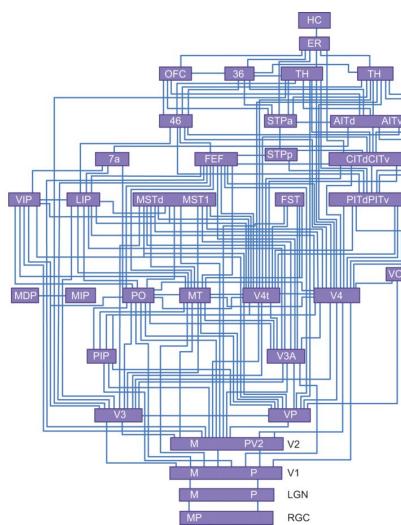
“La perception est une activité informée du monde extérieur par des entrées sensorielles”

“Système” perceptif

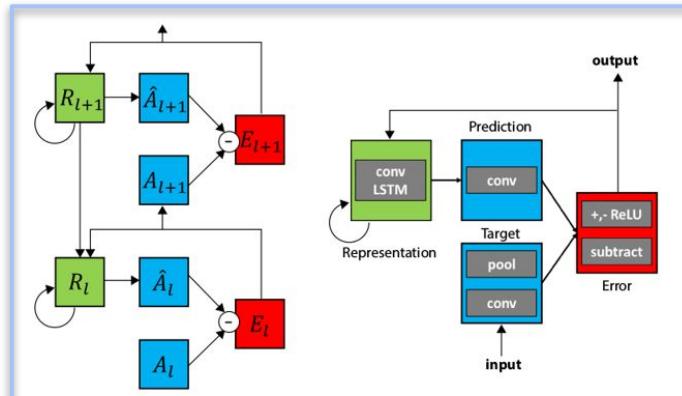
composant/communication/bouclage/dynamique



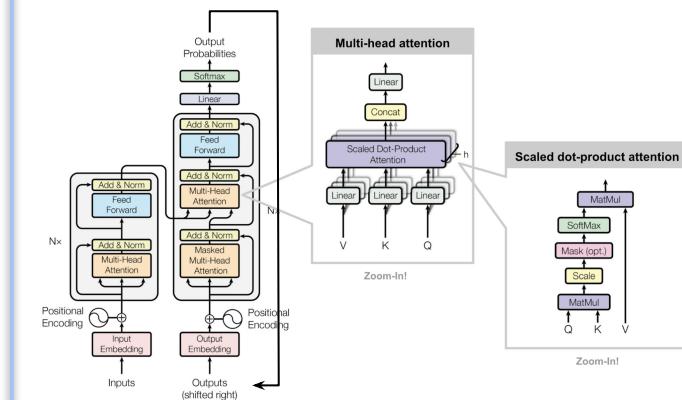
[N'Guyen, Thurat & Girard, 2014]



[Felleman & Van Essen, 1991]



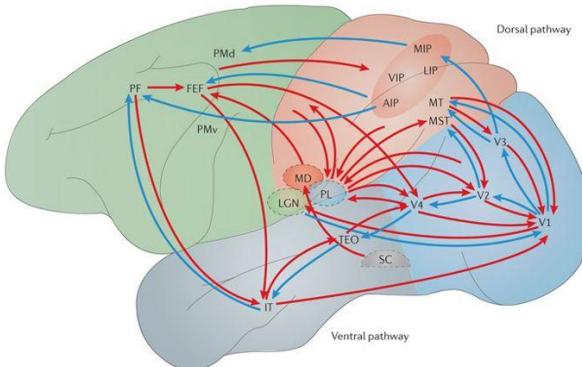
[Lotter, Kreiman & Cox, 2014]



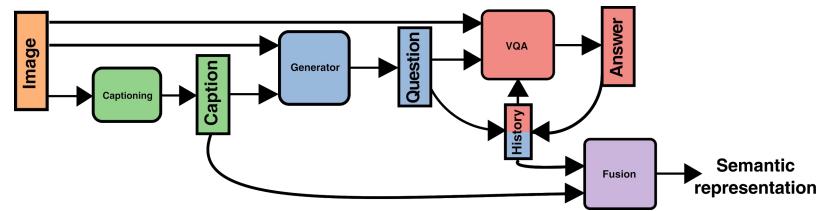
[Vaswani et al., 2017]

Perception & Cognition

- Cerveau
 - Bouclages multiples
 - Haut → Bas niveau (ex: IT → V1)
 - Difficile de hiérarchiser (Top, Bottom?)
- Frontière fixe entre perception pure (“Impénétrable”) et fonctions cognitives introuvable.



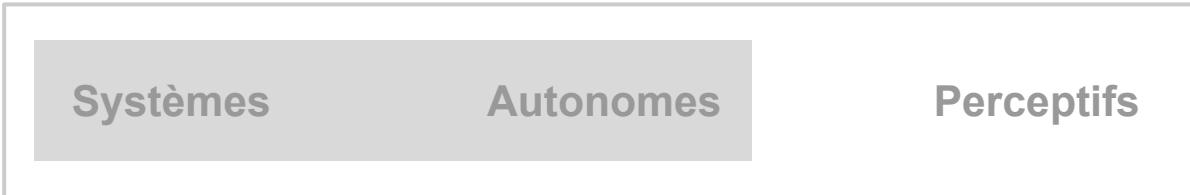
[Gilbert & Li, 2013]



[Bucher, Herbin & Jurie, 2018]

- Réseaux profonds artificiels
 - Architecture complexe
 - Combinaison de modules “cognitifs” et “signal”
- Unité (rigidité?) de représentation

Agent perceptif



Autonomie

De quelle activité ou faculté?

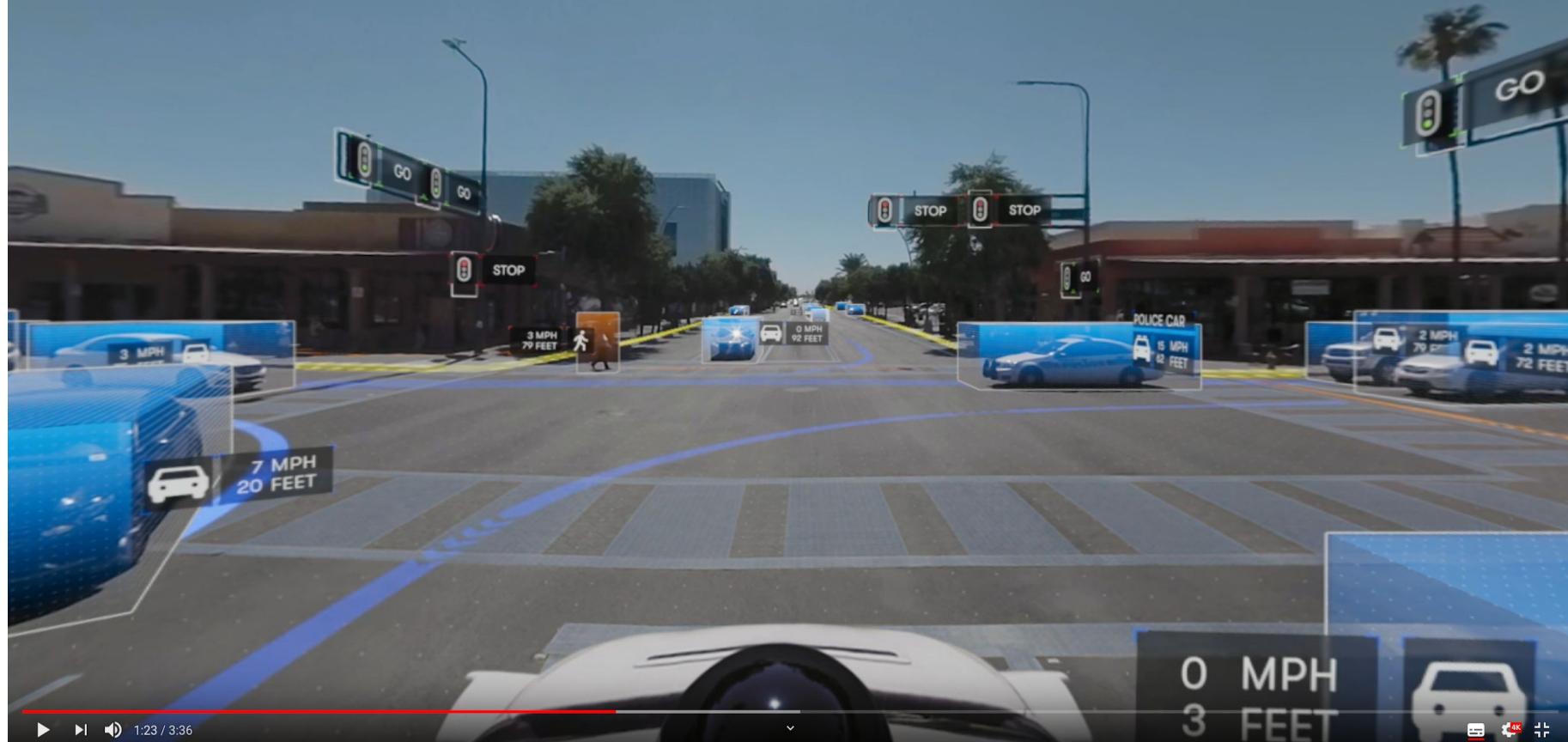
Indépendance par rapport à qui?

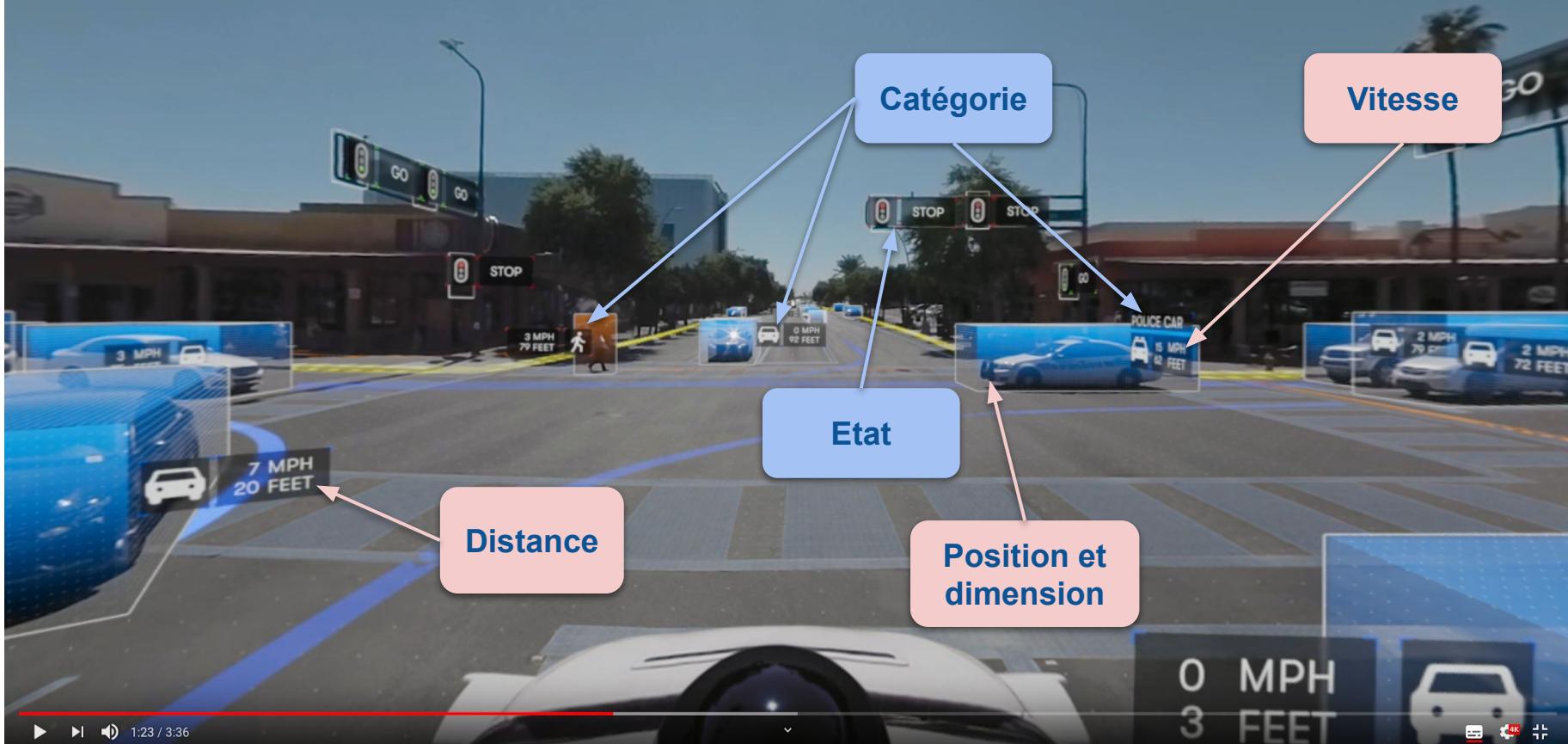
Comment est-elle obtenue?

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/Deceleration	Monitoring of Driving Environment	Fallback Performance of Dynamic Driving Task	System Capability (Driving Modes)
Human driver monitors the driving environment						
0	No Automation	the full-time performance by the <i>human driver</i> of all aspects of the <i>dynamic driving task</i> , even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	Driver Assistance	the <i>driving mode-specific</i> execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the <i>human driver</i> perform all remaining aspects of the <i>dynamic driving task</i>	Human driver and system	Human driver	Human driver	Some driving modes
2	Partial Automation	the <i>driving mode-specific</i> execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the <i>human driver</i> perform all remaining aspects of the <i>dynamic driving task</i>	System	Human driver	Human driver	Some driving modes
Automated driving system ("system") monitors the driving environment						
3	Conditional Automation	the <i>driving mode-specific</i> performance by an <i>automated driving system</i> of all aspects of the <i>dynamic driving task</i> with the expectation that the <i>human driver</i> will respond appropriately to a <i>request to intervene</i>	System	System	Human driver	Some driving modes
4	High Automation	the <i>driving mode-specific</i> performance by an <i>automated driving system</i> of all aspects of the <i>dynamic driving task</i> , even if a <i>human driver</i> does not respond appropriately to a <i>request to intervene</i>	System	System	System	Some driving modes
5	Full Automation	the full-time performance by an <i>automated driving system</i> of all aspects of the <i>dynamic driving task</i> under all roadway and environmental conditions that can be managed by a <i>human driver</i>	System	System	System	All driving modes

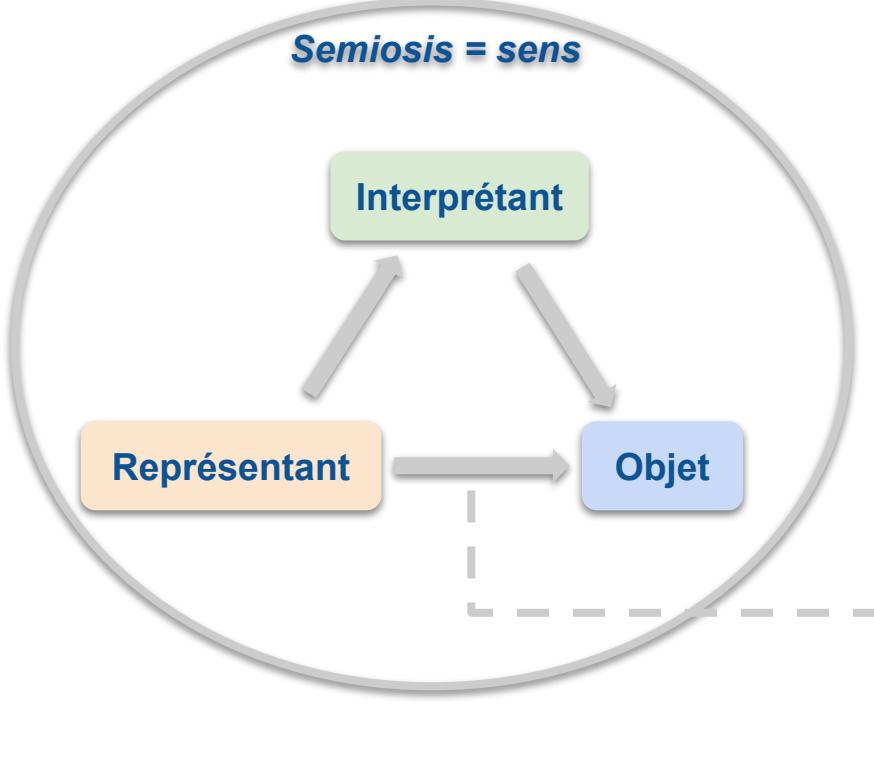
Copyright © 2014 SAE International. The summary table may be freely copied and distributed provided SAE International and J3016 are acknowledged as the source and must be reproduced AS-IS.

Système perceptif: “servile” → “autonome”?

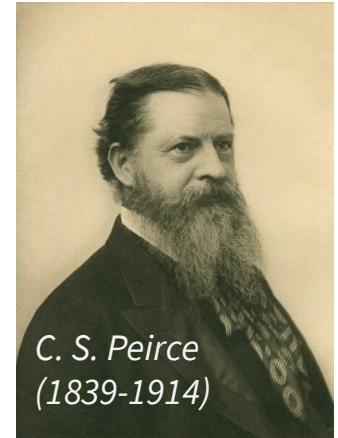




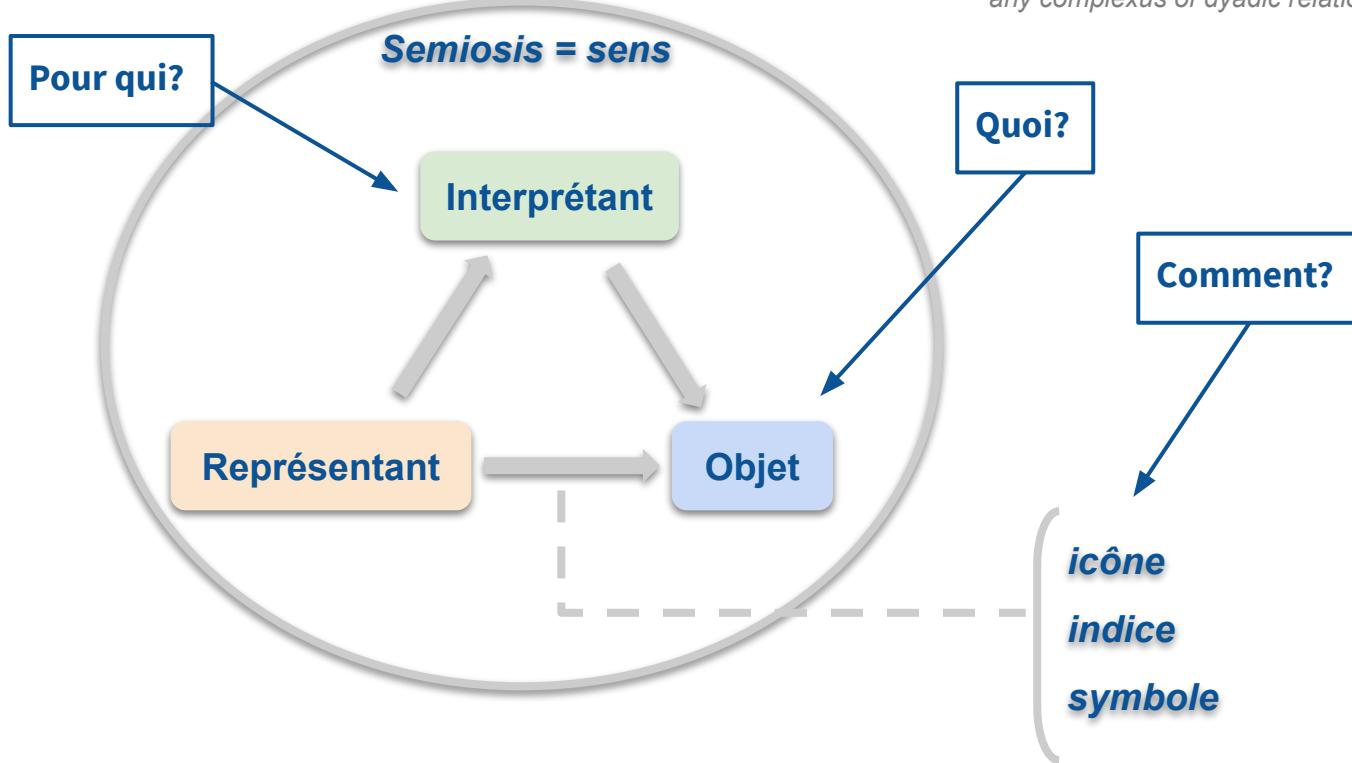
Sémiotique



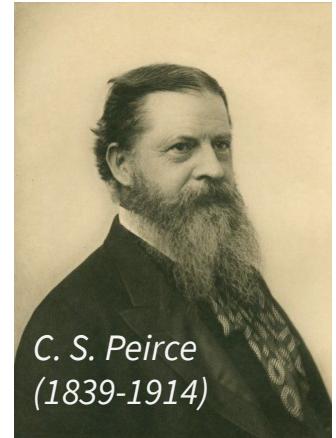
"A Sign, or Representamen, is a First which stands in such a genuine triadic relation to a Second, called its Object, as to be capable of determining a Third, called its Interpretant, to assume the same triadic relation to its Object in which it stands itself to the same Object. The triadic relation is genuine, that is its three members are bound together by it in a way that does not consist in any complexus of dyadic relations"



Sémiotique



"A Sign, or Representamen, is a First which stands in such a genuine triadic relation to a Second, called its Object, as to be capable of determining a Third, called its Interpretant, to assume the same triadic relation to its Object in which it stands itself to the same Object. The triadic relation is genuine, that is its three members are bound together by it in a way that does not consist in any complexus of dyadic relations"



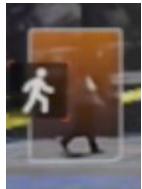
Diversité des références et interprétants

- le conducteur du véhicule arrêté
- l'auto-pilote du véhicule arrivant au croisement
- la police



- icône de l'état d'un feu de signalisation
- indice d'un croisement
- symbole de demande d'arrêt

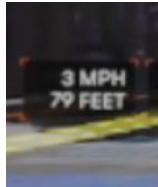
Autres signes



- icône de présence d'un piéton
- indice d'un danger éventuel



- indice de présence future de véhicule
- icône de forme de trajectoire
- icône de sens de déplacement



- indice d'obstacle éventuel

mesure = signe particulier !

Perception et signes

Deux parties:

- Perception = exprime des signes (= “représentants”) ou des mesures pour un destinataire (= “interprétant”)
- Interprétation = le destinataire (= “interprétant”) construit une référence à partir du représentant reçu (“semiosis”)

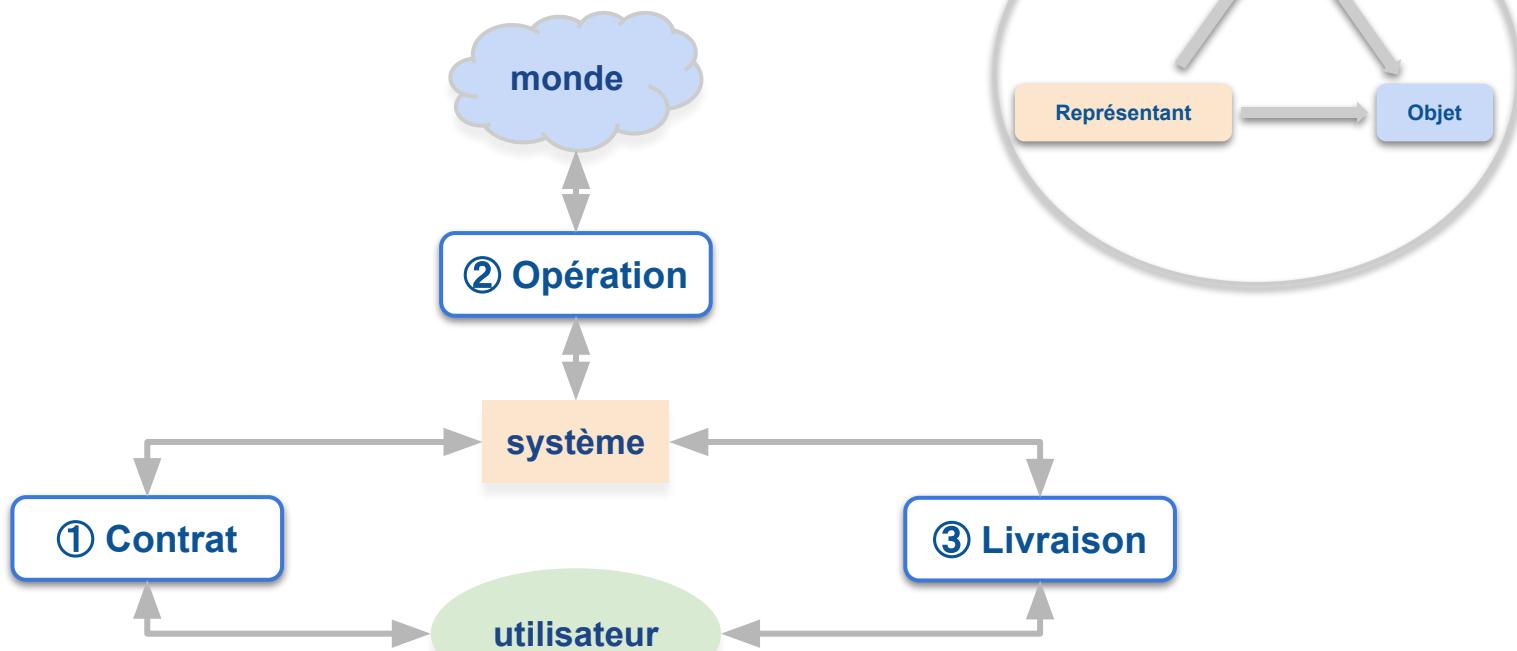
Le destinataire peut être humain ou artificiel

Diversité de références et d’interprétants



Modélisation et évaluation de la perception comme production de signes

Perception comme service



“Service de perception”

Rôle

- Explicite la représentation de la fiabilité du processus de perception = adéquation entre l'attendu et la réalisation

Moyens

- Contractualisation qui définit ce qui est attendu de part et d'autre, et les règles à appliquer en cas d'échec ou de conflit
- Répertoire de tâches perceptives à adapter aux besoins du destinataire/utilisateur/interprétant



Auto-évaluation des capacités perceptives



Modélisation et évaluation d'un “service de perception”

Autonomie perceptive?

Quelle activité ou faculté?

- Production de signes ou mesures

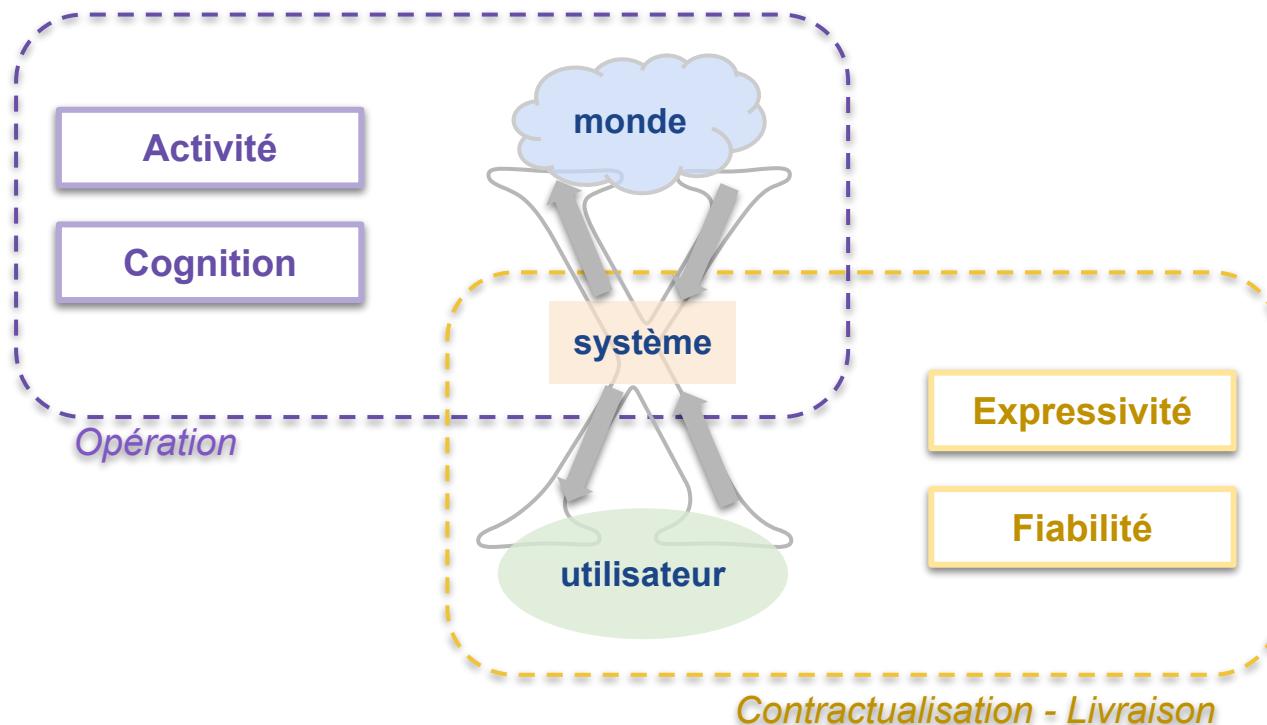
Indépendance par rapport à qui ou quoi?

- Utilisateur interprétant les signes ou mesures

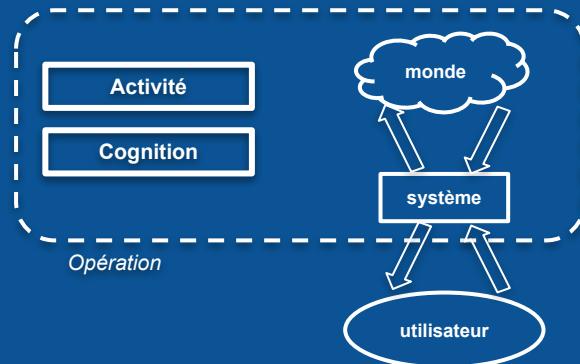
Comment est-elle obtenue?

- En tant que service honorant un contrat négocié

Caractéristiques d'un APES



Comment fonctionnent les APES?



Deux temporalités

Court terme = Opération

→ “perception active”

Long terme = Développement

→ apprentissage

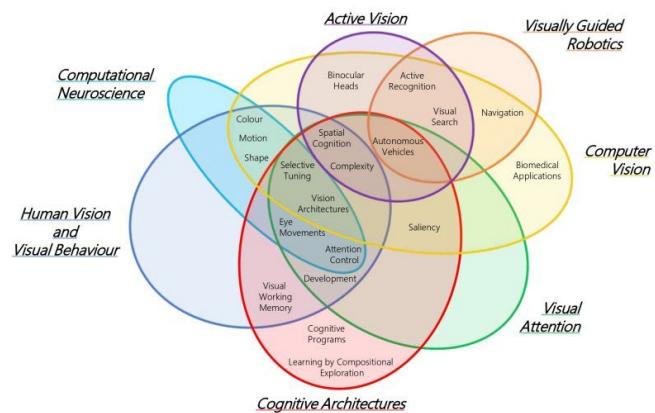
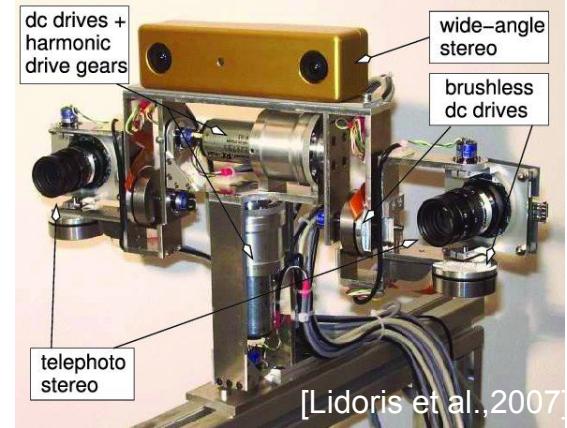
“Perception active”

Entre “robotique”, “vision par ordinateur” et “neurosciences”

Tâches canoniques

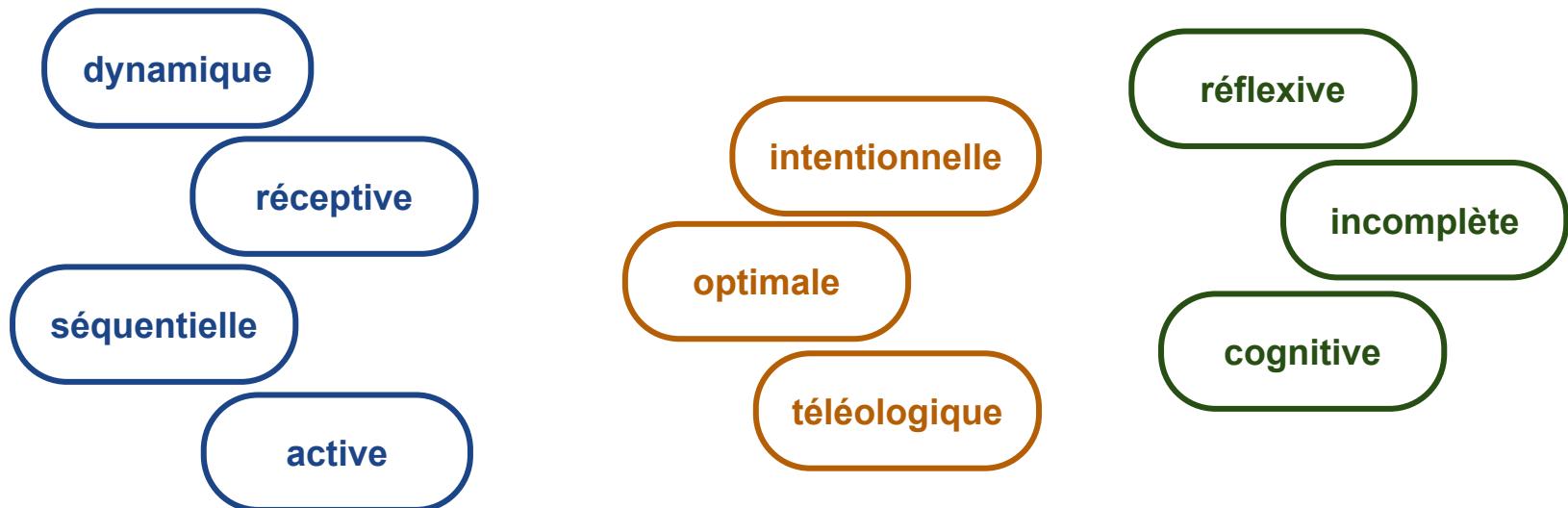
- reconstruction 3D
- recherche d'objet
- reconnaissance d'objet multi-vues
- interprétation de scènes dynamiques et capteurs commandables

Inspiré par perception naturelle



[Tsotsos Lab]

Caractéristiques de la perception active



La perception est active par nature

Formalisation de la perception active

Cadre formel: Décision séquentielle

- Comment manipuler des espaces de grandes dimensions?
- Comment modéliser l'incertitude?
La mettre à jour?
- Quelles actions? (“next best view”, sélection de caractéristiques ou de ressources)
- Quand et comment prédire?

[Herbin, 1996, 1997, 1998, 2003, 2004, 2014]

[Defretin et al., 2010, 2011]

$$\Phi_{t-1}, a_t, x_t \rightarrow \Phi_t$$

$$b_t(y) \propto \Pr[x_t | a_t, \Phi_{t-1}, y] \cdot b_{t-1}(y)$$

$$a_t = \arg \max_a \mathcal{U}_t(\Phi_{t-1}, a)$$

$$\hat{y}_t = \arg \max_y b_t(y)$$



Dynamique fonctionnelle et contextualisation, Espaces de tâches structurés

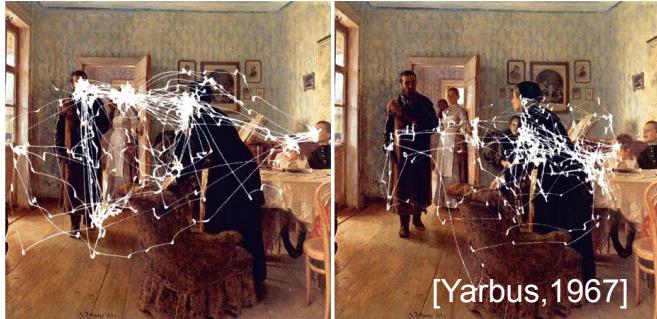
L'Attention: ingrédient essentiel

Rôle

- Optimise les ressources sous contraintes
 - Révèle une activité cognitive (“intentionnalité”)

Caractéristiques

- Sélection (“hard”) vs. Modulation (“soft”)
 - “Overt” vs. “Covert”
 - Interne vs. Externe



Attention généralisée: au delà du contrôle des entrées sensorielles

Développement

Problème

Comment exprimer de “bons” signes ou mesures?

Solution (contemporaine)

Apprentissage = profiter des expériences passées pour préparer un (bon) comportement futur

Quel apprentissage pour un Système Perceptif Autonome?

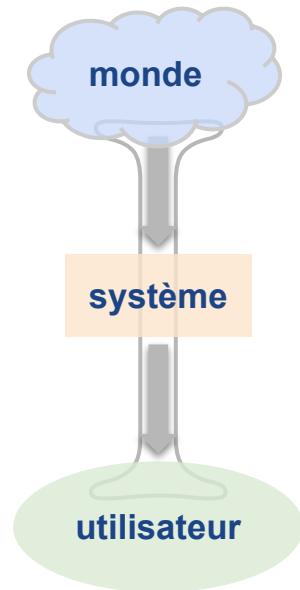
Apprentissage supervisé: la référence

Intérêt

- Prise en compte simultanée du “besoin utilisateur” et du domaine des expériences
- Simplicité fonctionnelle → résultats mathématiques
- Architectures “end-to-end”
- Approche “performante”

Mais

- Hypothèse de stationnarité du domaine
- Dépendance aux données annotées (qualité, quantité, densité)
- Univocité de la sortie
- Ni adaptatif, ni “actif”

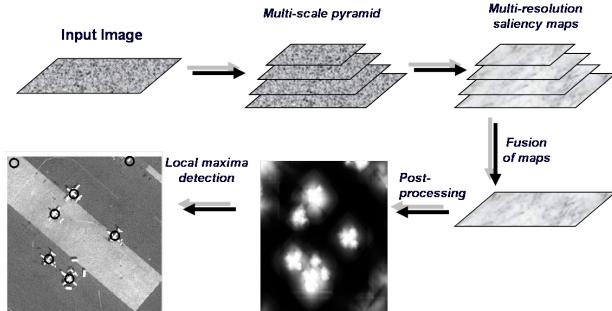


Sortir du supervisé - I

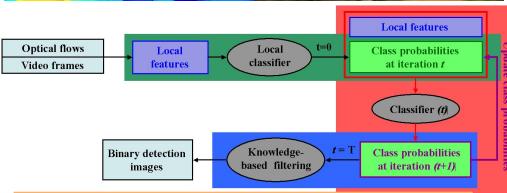
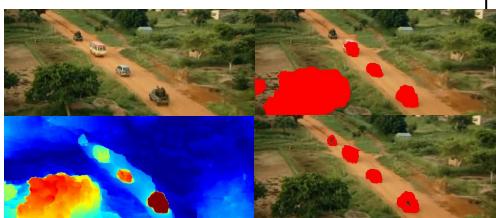
Hybrider: “Model based” + “Data driven”.
 (C'est ce que l'on faisait avant le DL!)

Pb: performances (très) inférieures au supervisé!

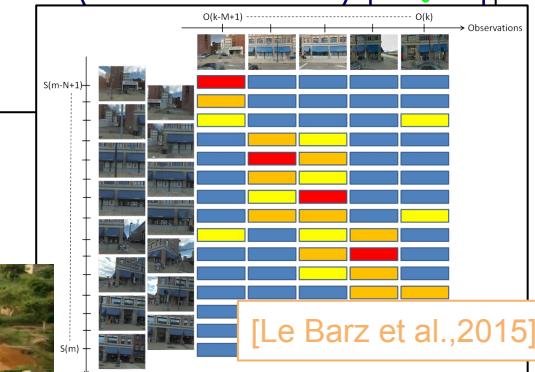
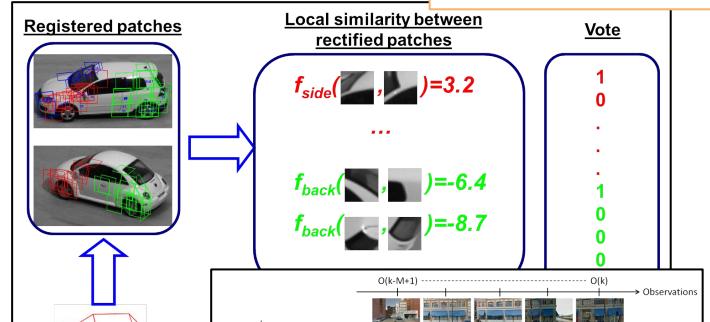
Mais chaînes plus interprétables



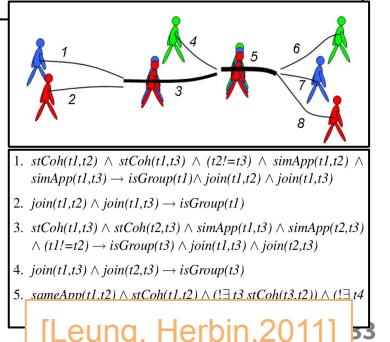
[Francesconi, Chalmond, Herbin, 2006]



[Guilmart, Herbin, Perez, 2011]



[Le Barz et al., 2015]



[Leung, Herbin, 2011]

Sortir du supervisé - II

Autres modalités d'apprentissage:

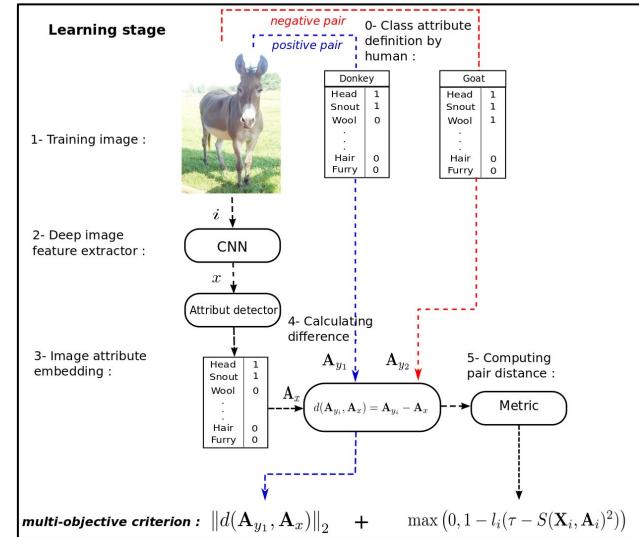
- semi-supervisé, auto-supervisé, adaptation de domaine, méta-apprentissage...

Introduire des “connaissances”:

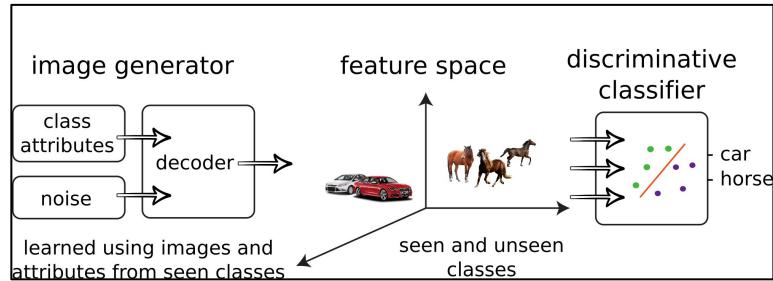
- Description sémantique de problème (“zero-shot” learning)
- Génération conditionnelle de données



Apprentissage hétérogène



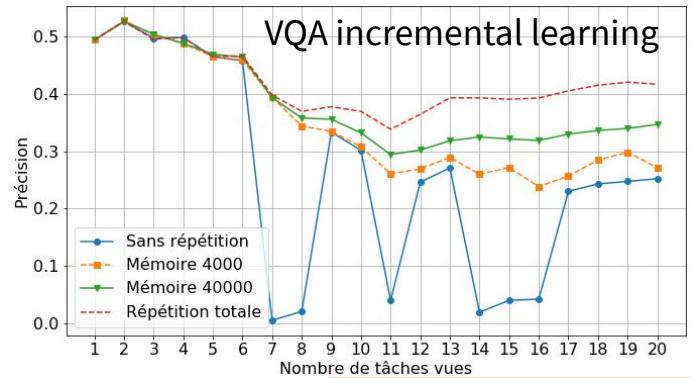
[Bucher, Herbin, Jurie, 2016, 2017]



Sortir du supervisé - III

Dynamiques d'apprentissage: Dépasser le “learn & test”

- Augmenter les compétences et la polyvalence
 - Continual/incremental learning
- Contrôler les demandes d'information
 - Active Learning
- Profiter de toutes les expériences
 - Never-ending/lifelong learning
- Gérer la complexité des tâches
 - Curriculum/developmental Learning



[Lechat, Herbin, Jurie, 2019]



Unifier opération et développement

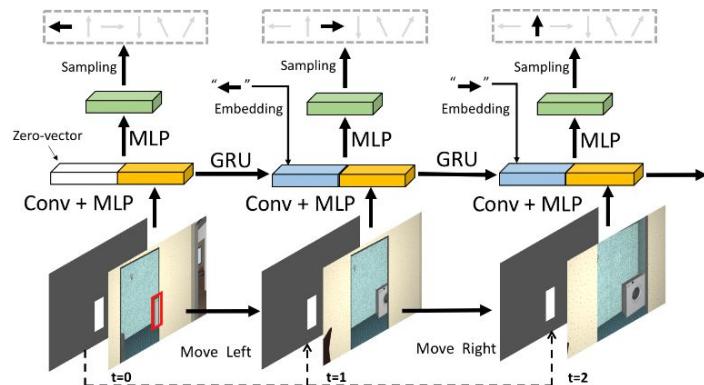
Bilan: Perception active à l'ère du “deep learning”

Nouveaux moyens

- “Embedding” multi-modal (données sensorielles et actions) et “end-to-end”
- Réseaux récurrents (LSTM, GRU,): modélisation empirique des transitions
- Attention (“soft”): “gating”, “self-attention”
- Apprentissage par renforcement profond

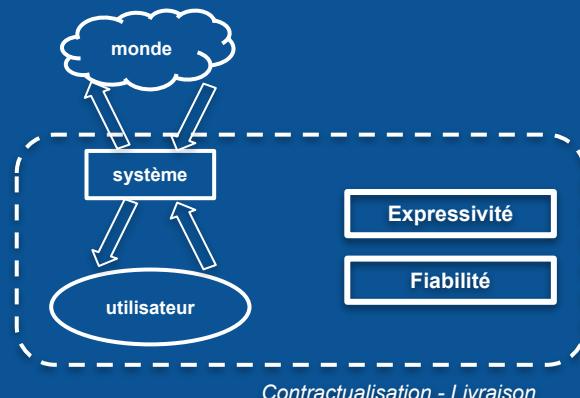
Mais

- Adaptivité limitée: architectures sophistiquées mais fermées, peu évolutives ou polyvalentes
- Abandon du séquentiel?



“Embodied Visual Recognition”
[Yang et al., 2019]

Faire confiance aux APES?



Des systèmes perceptifs fiables?

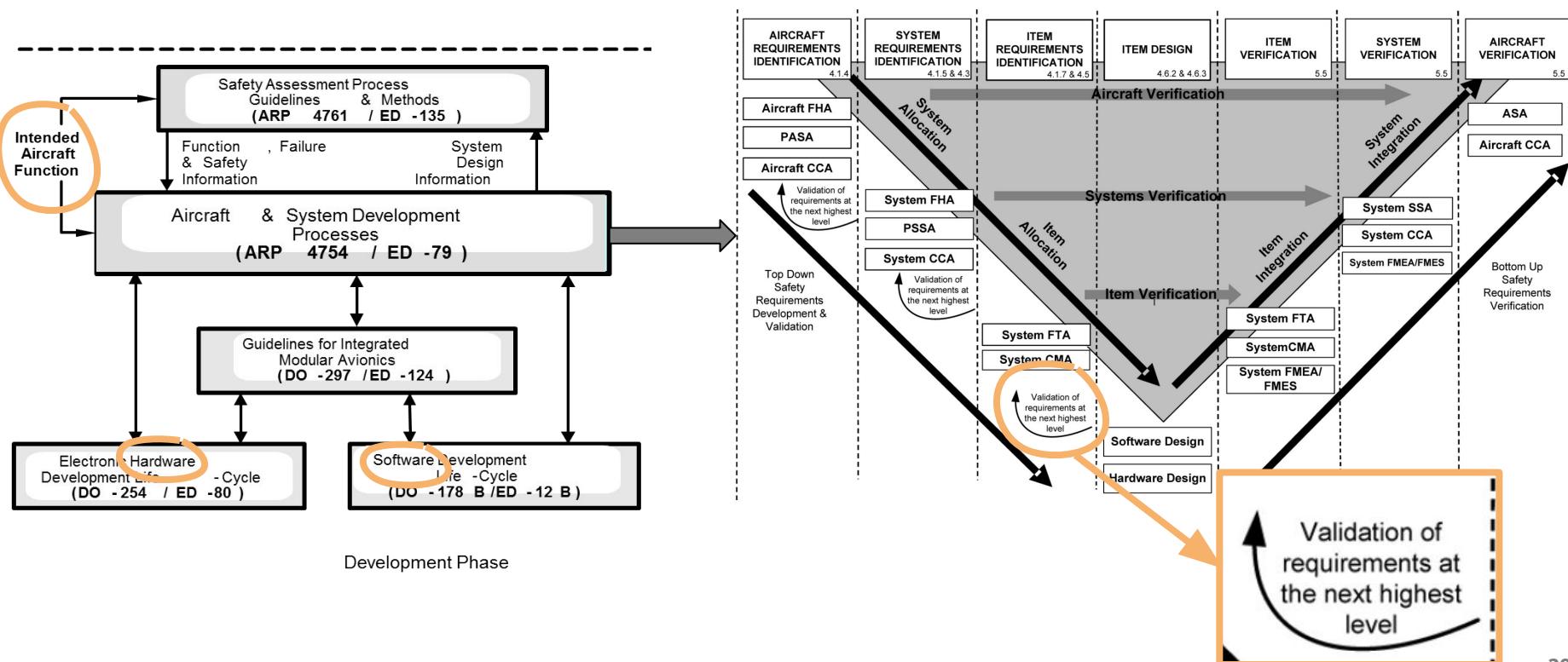
Peut-on s'inspirer des formalismes existants?

Quelles spécificités des systèmes perceptifs?

Quelles améliorations/évolutions proposer?

Rôle de l'autonomie?

Certification en aéronautique



Niveaux d'analyse de Marr & Poggio (Vision)

Computational	What is the goal of the computation? Why is it appropriate, and what is the logic of the strategy by which it can be carried out?
Representation and algorithm	<i>How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?</i>
Software	How is the algorithm formally described and coded?
Hardware implementation	How can the representation and algorithm be realized physically?



“Validation of requirements”

Comment assurer la cohérence entre ces deux niveaux?

Fiabiliser les systèmes perceptifs?

Comment s'assurer que l'algorithme réalise bien la fonction voulue?

- Valider les exigences et anticiper les risques
 - évaluation, robustesse, vérification formelle

Comment empêcher l'algorithme de générer des comportements inattendus?

- Cerner le domaine d'usage
 - détection d'anomalie ou d'attaques, auto-diagnostic

Comment démontrer ou révéler le bon comportement de l'algorithme?

- “Open the black-box”
 - explicabilité, transparence

La “nouvelle” difficulté : impact de l'apprentissage

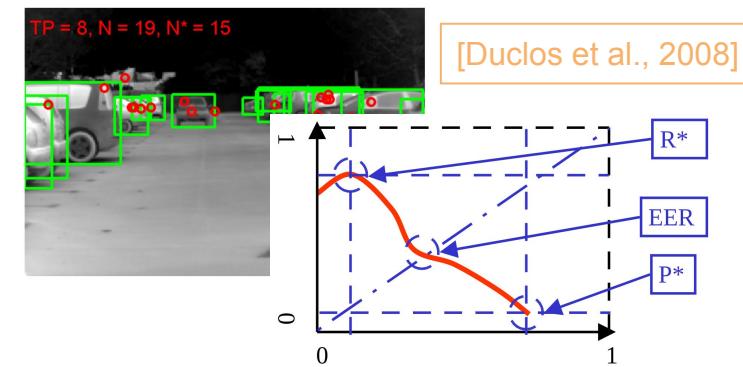
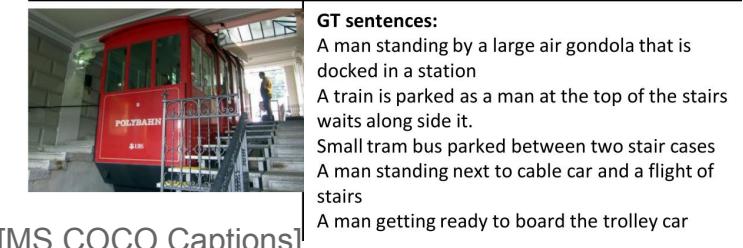
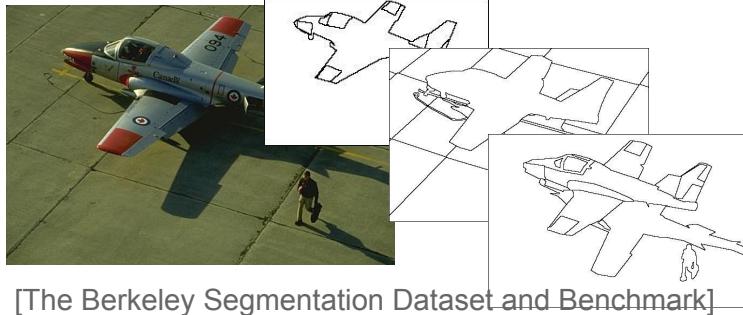
Evaluation

Pratique actuelle: “Benchmark”

- collecter des données représentatives
- exprimer la sortie souhaitée (“groundtruth”)
- définir des métriques = mesures d'écart avec GT
- calculer des stats (moyennes) sur
 - données
 - points de fonctionnement
 - natures de prédictions ...

Toutes ces étapes ont des faiblesses

Que mesure vraiment un benchmark?

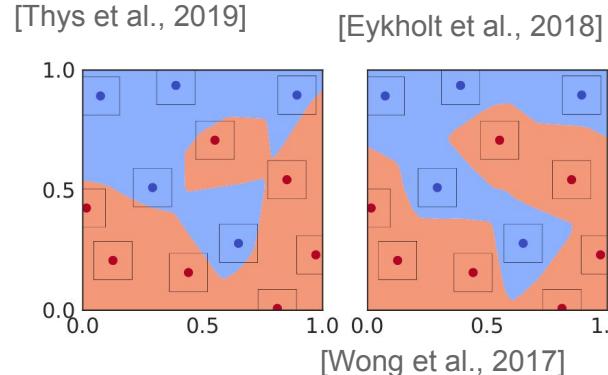
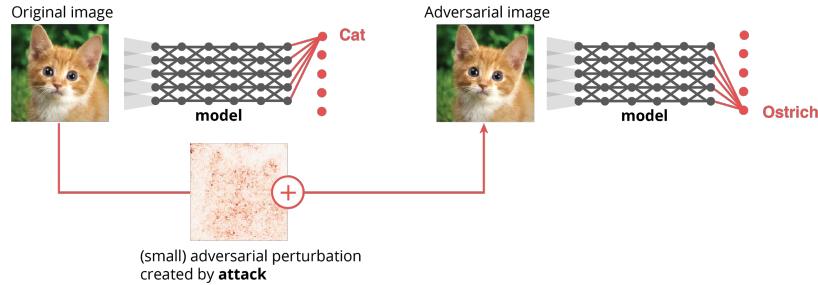


Robustesse (du DL)

Deux stratégies actuelles

- Adversarial Training
 - Inclut les adversaires dans une fonction de coût
 - Conflit avec coût de prédiction et généralisation?
 - Origine des exemples adversaires mal cernée
- Vérification formelle
 - Garanties “exactes” (SMT, programmation linéaire)
 - Passage à l’échelle difficile
 - Formulation spécifique des exigences
 - Ne dit rien sur les capacités de généralisation

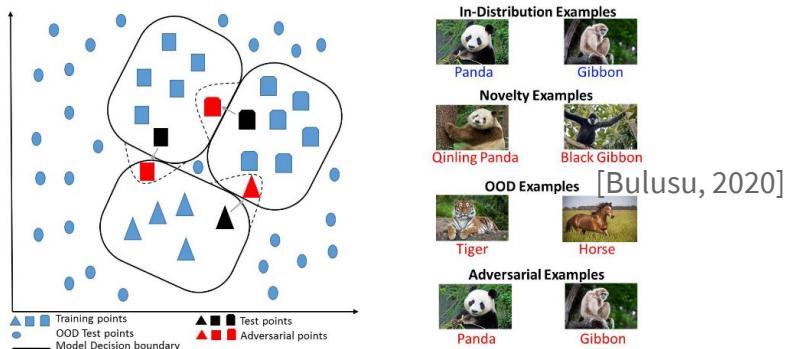
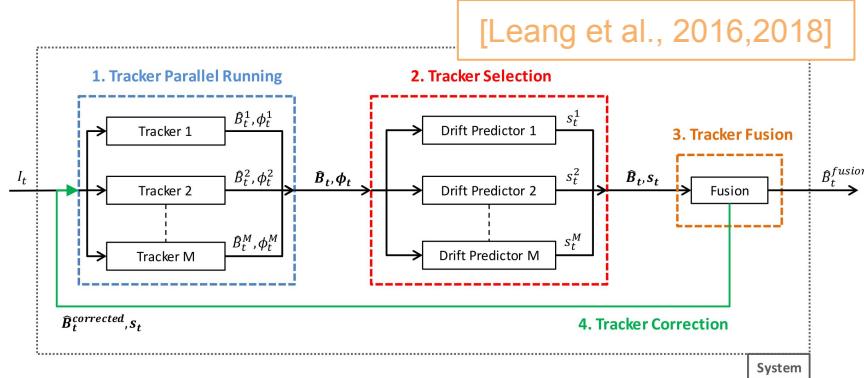
Encore beaucoup d’inconnu!



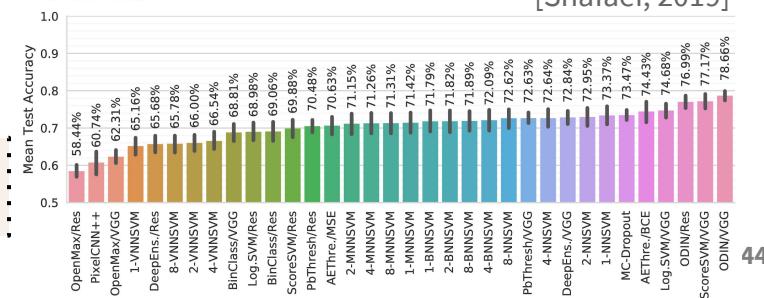
Auto-évaluation

- Détection d'anomalie
 - “Facile” quand anomalie connue (supervisé)
 - “Difficile” sinon (non supervisé)
 - Redondance/fusion plus efficace que sélection
 - “Out Of Distribution detection”
 - Effets “Grandes dimensions” contre-intuitifs
 - Liens avec exemples adversaires?
 - Performances faibles

Argument central pour la fiabilité: est-ce faisable?



[Shafeei, 2019]



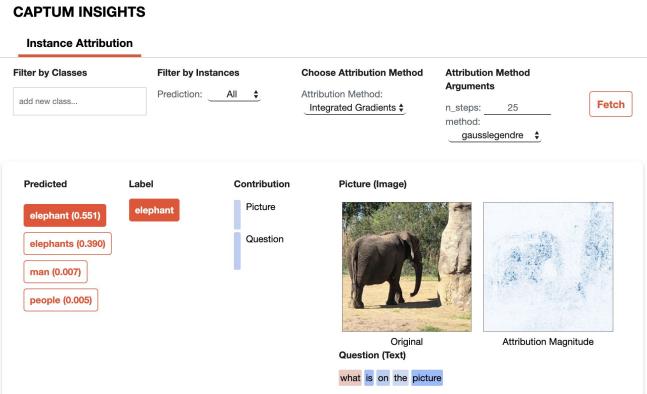
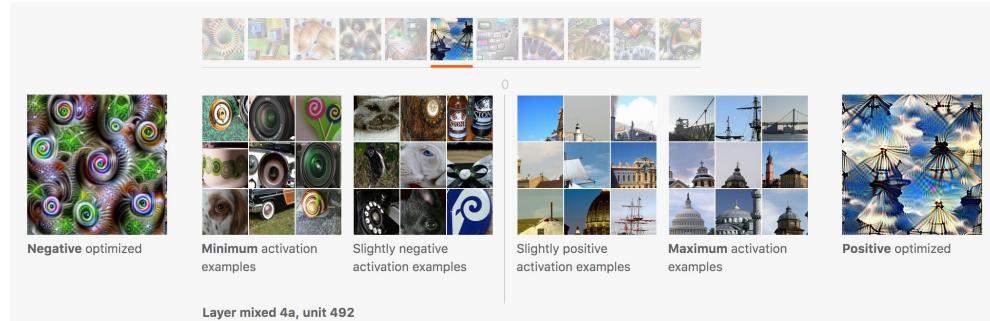
Intégrer détection d'anomalie et récupération

Explicabilité

- De quoi?
 - algorithme ou prédition
- Comment?
 - visualisation de caractéristiques d'entrée ou de variables internes, chemins dans graphes, approximation, génération de données virtuelles, d'exemples proches/opposés...
- Pour qui?
 - scientifique, développeur, expert du domaine, autorité, utilisateur, destinataire des prédictions, étudiant...



Sémiotique?



Clarification des objectifs de l'explicabilité



Evaluation de l'explicabilité

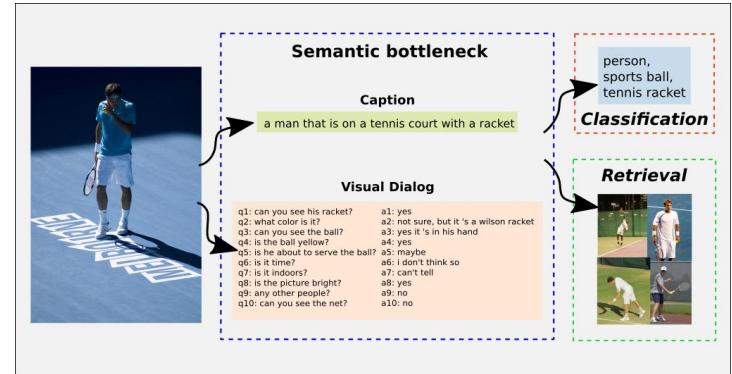
Explicabilité par conception

Faire dépendre la prédiction de représentations ou processus interprétables

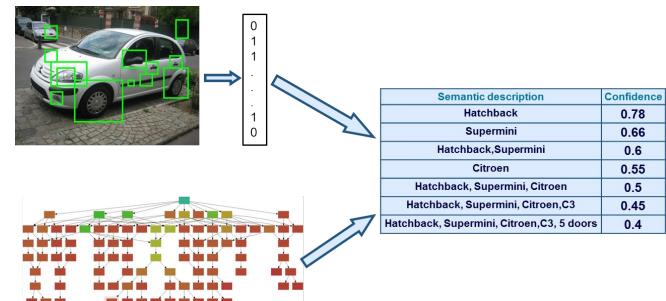
- Sélection des caractéristiques informatives (attention)
- Etat interne interprétable
- Raisonnement
- Incertitude plus expressive



Expliquer pour fiabiliser



[Bucher,Herbin,Jurie,2018]



[Tousch,Herbin,Audibert,2008]

Interactivité interprétable

1. Opération/Contractualisation/Livraison dynamiques
2. Collaborer/Apprendre avec l'utilisateur
3. Dialogue comme support des interactions formelles et justifications interprétables

Solution pour améliorer la fiabilité?



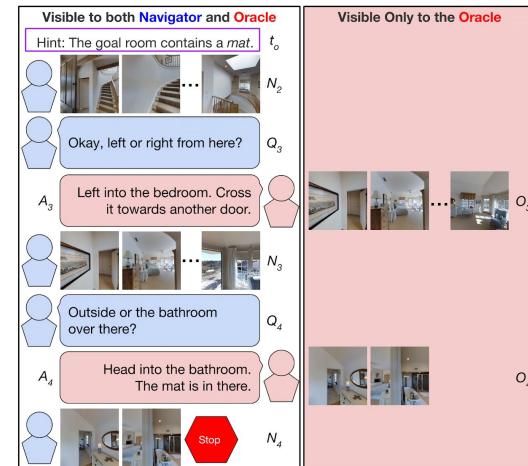
Apprentissage pour l'interaction utilisateur



Langage naturel comme représentation pivot



[Das, 2017]



[Thomason, 2019]

Conclusion

Une perception “adulte”?

Question: comment utiliser la perception artificielle dans des applications exigeantes?

Proposition: munir la perception d'une **autonomie**

- Perception comme **activité continue** associant opération & développement (apprentissage)
- Perception **intelligible** et **responsable** = capable de s'exprimer sur son comportement et de s'auto-évaluer
- Perception autonome = capable de **contractualiser un service** adapté à un utilisateur/destinataire

Thèmes de recherche

- 1. Renforcer les dimensions systémiques et dynamiques de la perception**
 - Polyvalence de la perception et dynamique fonctionnelle
 - Unifier opération et développement (learning)
- 2. Améliorer les échanges avec l'utilisateur/récepteur**
 - Modélisation et évaluation d'un service de perception
 - Formaliser l'interactivité avec l'utilisateur
- 3. “Responsabiliser” la perception**
 - Expliquer pour fiabiliser
 - Auto-évaluation
- 4. Étudier les convergences entre perception artificielle et cognition naturelle**

Questions?

