



France

Setting the Standard for Automation™



# Deep-Learning pour la Robotique Collaborative

Pr. Fabien Moutarde  
Centre de Robotique, MINES ParisTech, PSL Université

Fabien.Moutarde@mines-paristech.fr  
<http://people.mines-paristech.fr/fabien.moutarde>

En partenariat avec



Intelligence artificielle et industrie du futur  
Grenoble – 5 et 6 février 2019

## Le centre de Robotique de MINES ParisTech

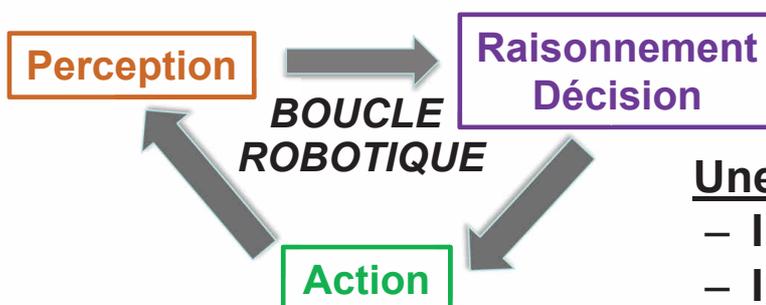


- ~ 40 personnes (dont ~15 doctorants)
- 1.8 M€/an de contrats de recherche
- Véhicules Intelligents et Robotique mobile :
  - analyse temps-reel de scène (par Machine-Learning) pour aides à la conduite et Véhicules Autonomes, Intersections Automatisées et manoeuvres cooperatives, etc...
  - Chaire DriveForAll (Safran-Valeo-PSA), VeDeCom, SoftBank\_Robotics*
- Reconnaissance Geste/actions & Interactions Humains-Machine :
  - robotique collaborative (H2020 project COLLABORATE, PSA Chair Robotics&VR)
  - Commande gestuelle dans habitacle automobile (PSA)
  - reconnaissance actions/intentions de piétons (Footovision, VedeCoM)
- Réalité Virtuelle, numérisation 3D & localisation-cartographie
  - immersion and réponse réaliste
  - 3D mobile mapping: digitalisation bâtiments/rues/infrastructures
  - SLAM pour la robotique mobile
- Estimation non-lineaire (IKF), commande sans modèle

- **LES Intelligences Artificielles et l'Apprentissage Machine (ML)**
- Les IAs pour la Robotique Collaborative
- Que permet le Deep-Learning sur des images ?
- Reconnaissance de gestes/actions pour Interaction et Collaboration Homme-Robot
- Apprentissage de comportement de robots par Imitation ou Renforcement

## Qu'est-ce que l'intelligence (humaine) ?

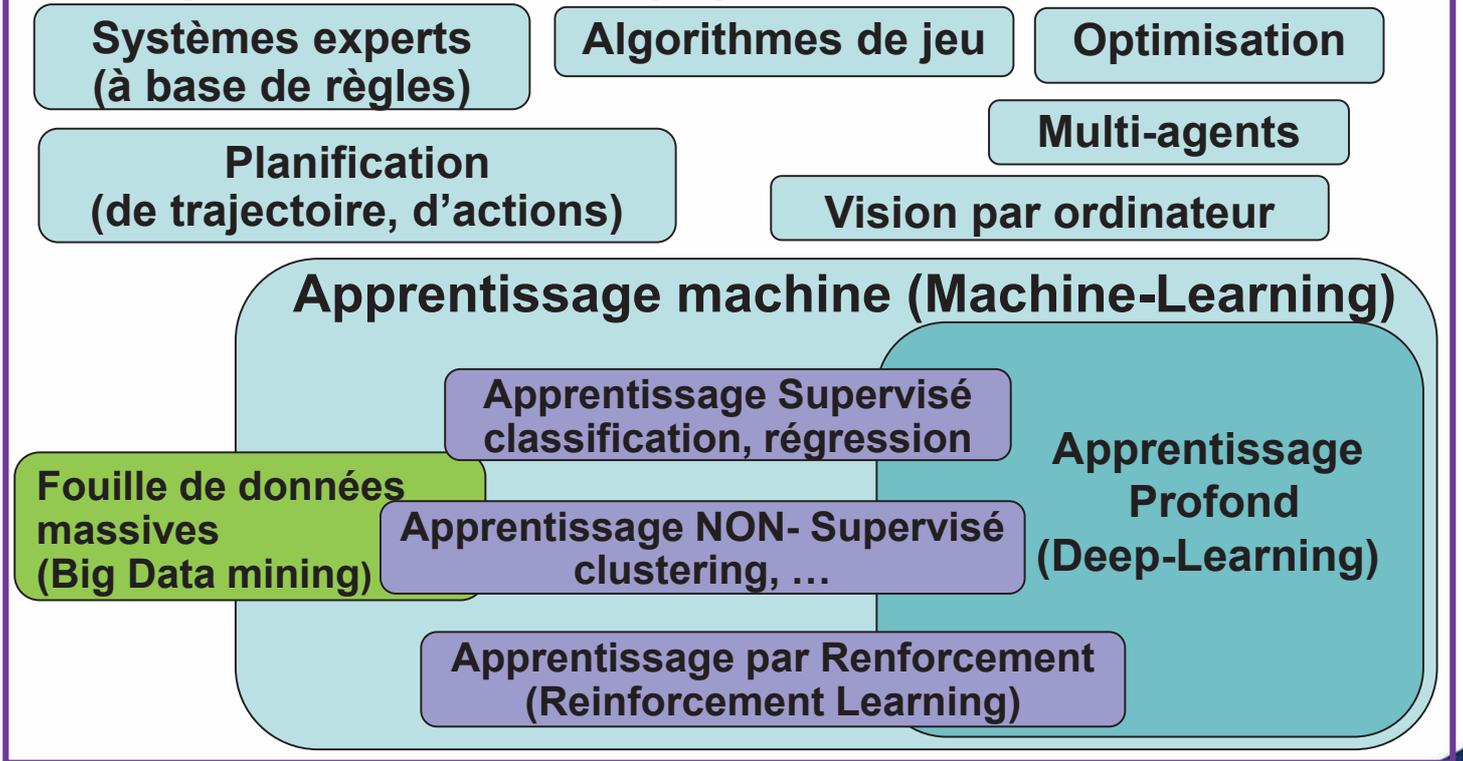
- Intelligence = raisonnement ?  
ou Intelligence = adaptation ?
- En fait, il existe de **NOMBREUX TYPE DIFFÉRENTS D'INTELLIGENCE**



### Une typologie possible :

- Intelligence "de perception"
- Intelligence de prédiction
- Intelligence de raisonnement
- Intelligence de comportement
- Intelligence "d'interaction"
- Curiosité
- Création

## Intelligence Artificielle (IA)

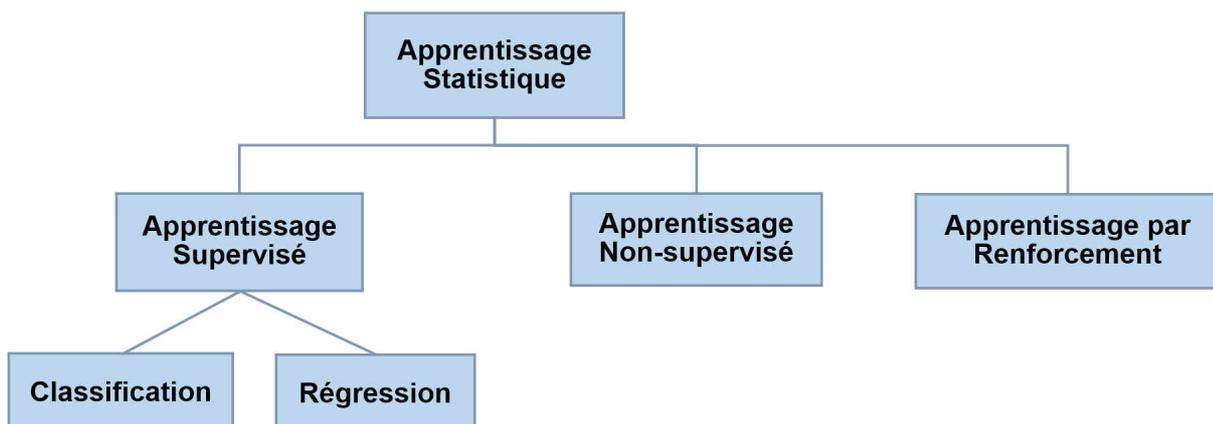


## Principe et typologie du Machine-Learning (ML)

Apprentissage Artificiel/Machine/Statistique (Machine-Learning)

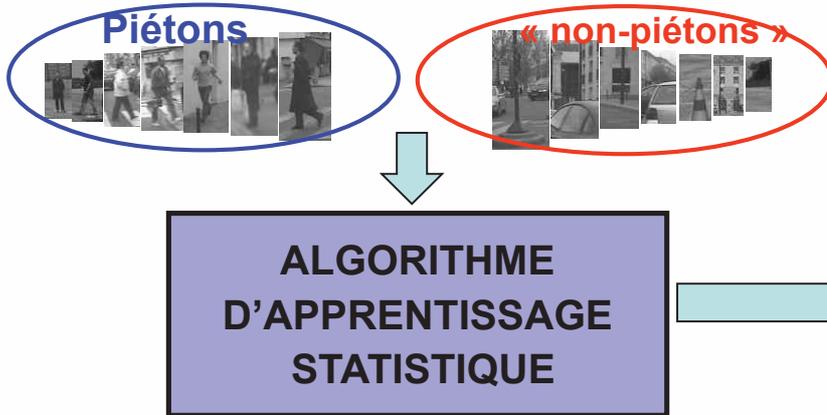
=

**MODELISATION EMPIRIQUE**  
par statistiques + optimisation

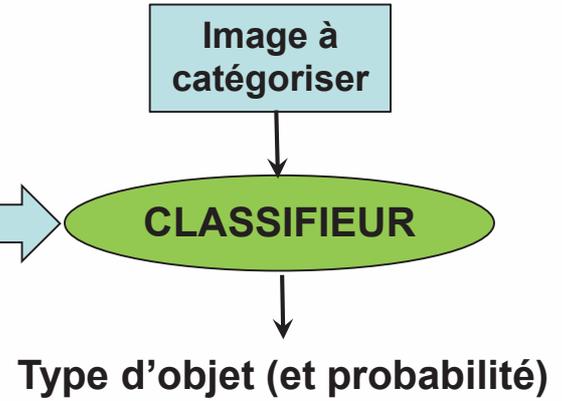


# Principe du Machine-Learning SUPERVISÉ

Hors-ligne dans un centre de R&D



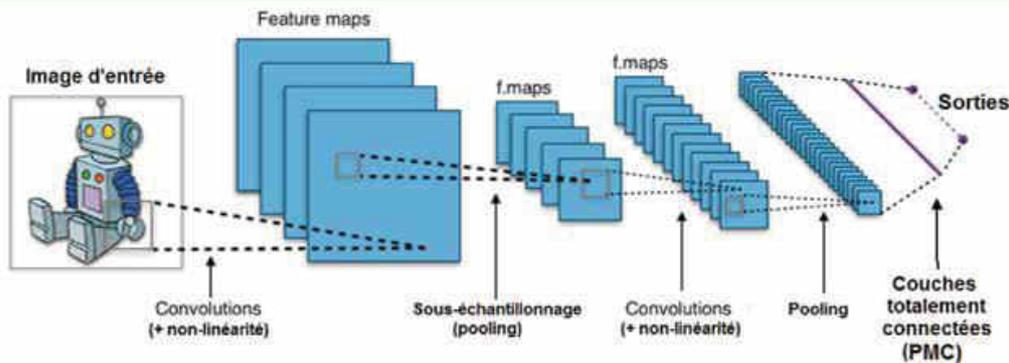
En temps-réel embarqué



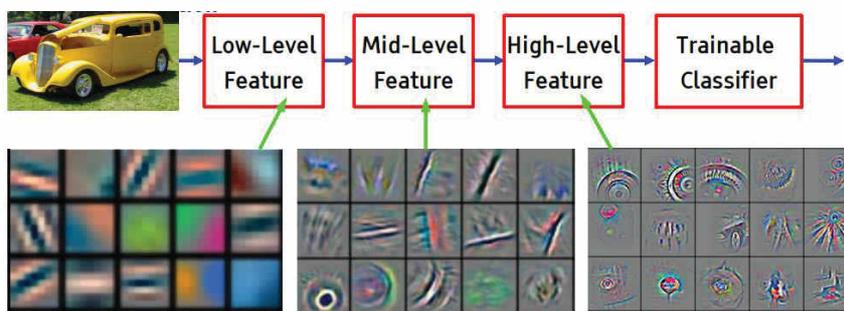
**Le classifieur est FIGÉ !**

Objectif défini par un ingénieur R&D  
Apprentissage SUPERVISÉ  
Résultat selon base d'exemples "étiquetés"  
(et paramètres d'apprentissage)

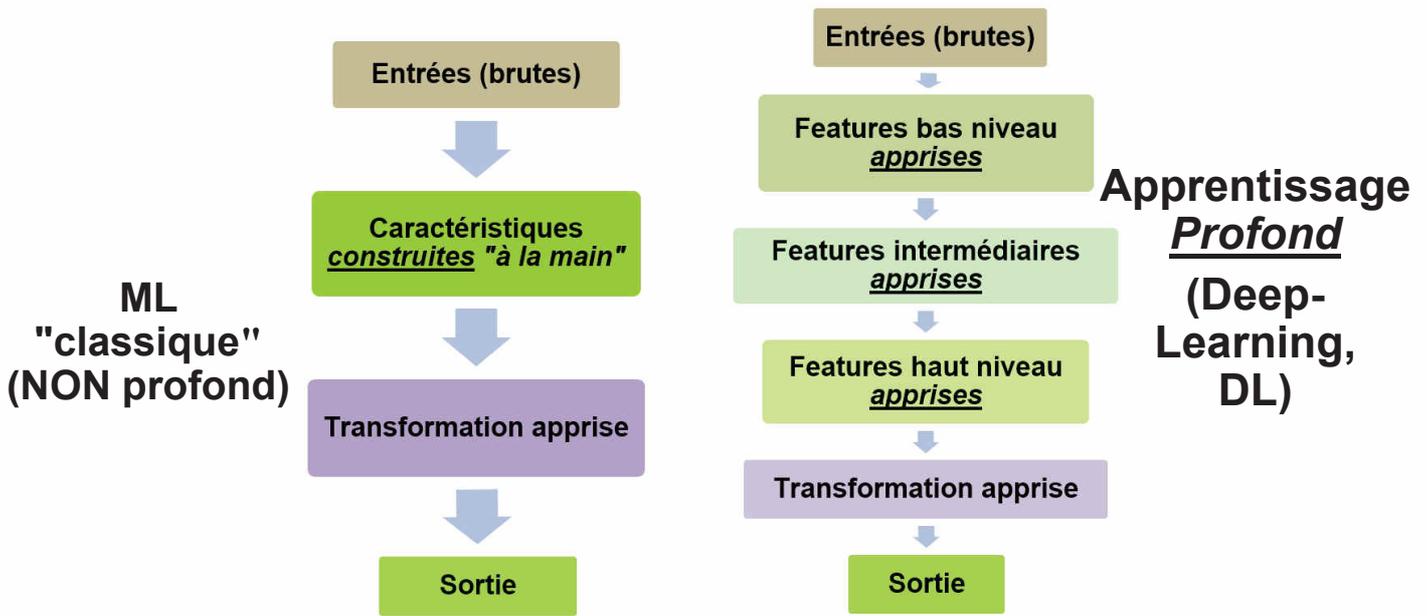
# Apprentissage PROFOND (Deep-Learning)



**Réseau neuronal "convolutif" :**  
nombreuses couches, et des millions de paramètres

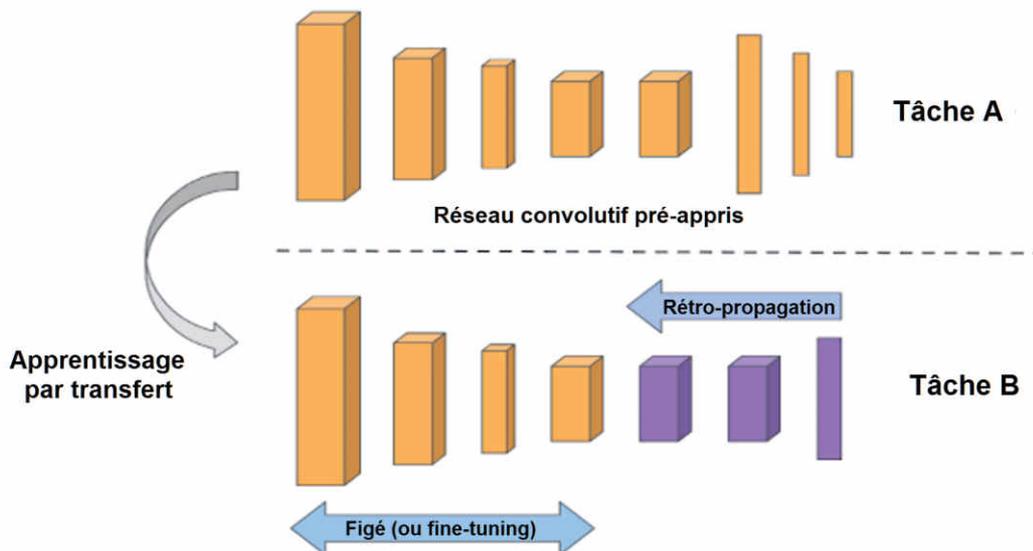


# Spécificité du Deep-Learning : apprentissage de représentation



Le Deep-Learning apprend conjointement les features et la classification

# Transfer Learning : le DL facilite la réutilisation



→ Innombrables applications (sémantisation images, estimation postures, ego-localisation, 3D from 2D, ...)

**convNets ~ « couteau suisse » quasi-incontournable pour tout ce qui est analyse et interprétation d'images**

- LES Intelligences Artificielles et l'Apprentissage Machine (ML)
- **Les IAs pour la Robotique Collaborative**
- Que permet le Deep-Learning sur des images ?
- Reconnaissance de gestes/actions pour Interaction et Collaboration Homme-Robot
- Apprentissage de comportement de robots par Imitation ou Renforcement

## Robots « traditionnels »



**Actions répétitives, rapidité, force importante, ...  
MAIS dangereux et surtout PEU ADAPTATIFS  
(simples « automatismes »)**

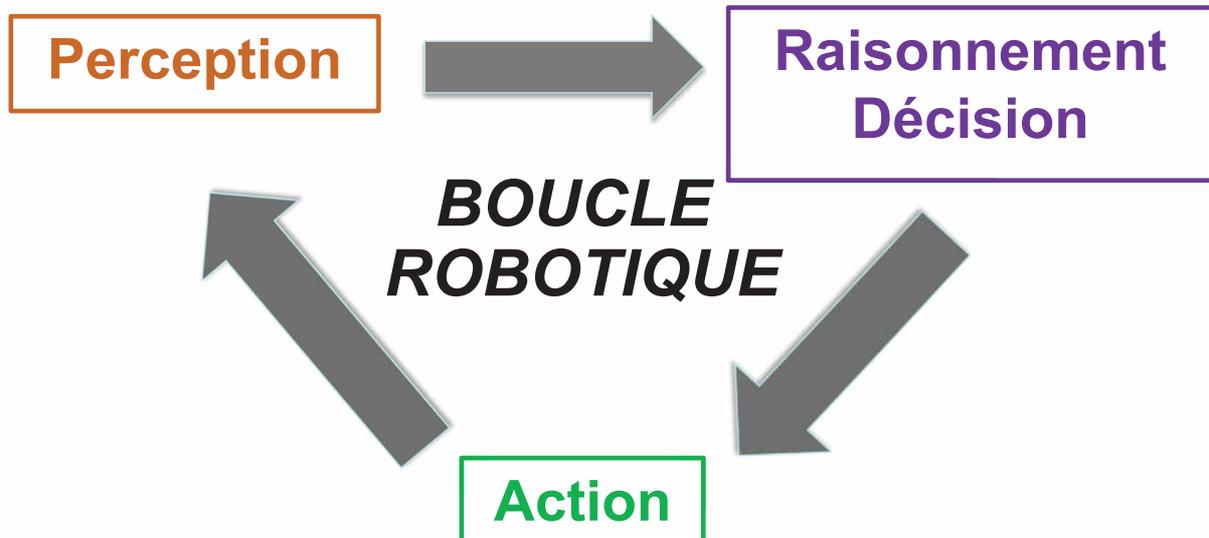
# Robots « intelligents » (≈ adaptatifs et/ou interactifs)



Réagir selon environnement  
et/ou interagir avec humains



# Principe général de fonctionnement des robots



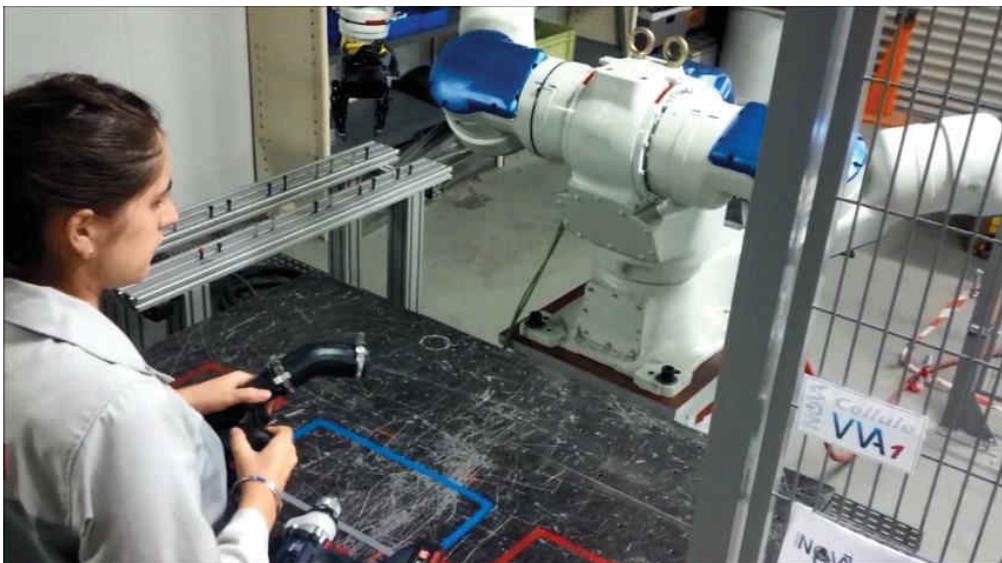
# Quelles types d'Intelligence faut-il aux robots ??

- **Analyser & interpréter un environnement fluctuant**
  - Reconnaître un lieu et s'auto-localiser
  - Détecter/localiser et catégoriser les « objets »
  - Suivre et prédire leurs mouvements
  - Deviner les « intentions »
- **Choisir l'action/mouvement à faire**
  - Logique décisionnelle
- **Adapter/optimiser l'action/mouvement choisi(e)**
  - Avoir un comportement, plutôt qu'une règle rigide
- **Interagir avec des humains ou entre robots**
  - Reconnaissance vocale
  - Interprétation du langage, et savoir « dialoguer »
  - Reconnaissance de gestes/actions, voire d'émotions
  - Coordination/collaboration entre robots

Deep-Learning pour la Robotique Collaborative, Pr. Fabien Moutarde, Centre de Robotique, MINES ParisTech, PSL, 5/2/2019

## Perception Intelligente pour les Robots Collaboratifs

**Besoin primordial :  
analyser et interpréter mouvements, actions & activités  
des Humains autour**



**Reconnaissance d'action pour collaboration Homme-Robot**  
*[centre de Robotique de MINES ParisTech, Chaire PSA "Robotique et Réalité Virtuelle"]*

Deep-Learning pour la Robotique Collaborative, Pr. Fabien Moutarde, Centre de Robotique, MINES ParisTech, PSL, 5/2/2019

# Défis actuels et futurs pour la Robotique Collaborative

- **Compréhension/reconnaissance d'activité**
- **Inférer les INTENTIONS des humains**
- **Apprentissage de COMPORTEMENT adaptatif**
  - Apprentissage par démonstration
  - Apprentissage par renforcement
  - Abstraction de la tâche plutôt qu'enregistrement de trajectoire
  - One/few shot(s) learning
- **Coordination/collaboration**
  - entre robots
  - avec humains (apprentissage de « règles sociales » implicites)

## Plan

- LES Intelligences Artificielles et l'Apprentissage Machine (ML)
- Les IAs pour la Robotique Collaborative
- **Que permet le Deep-Learning sur des images ?**
- Reconnaissance de gestes/actions pour Interaction et Collaboration Homme-Robot
- Apprentissage de comportement de robots par Imitation ou Renforcement

# Le Deep-Learning appliqué aux images

- Classification d'image
- Détection et catégorisation visuelle d'objets
- Segmentation sémantique of d'images
- ...
- Ego-localisation basée images
- Estimation de postures (Humaines)
- Inférence de la 3D (depth) à partir monovision
- ...
- Apprentissage de comportement basé images
  - Programmation de robots par démonstration ou imitation
  - Deep Reinforcement Learning

Deep-Learning pour la Robotique Collaborative, Pr. Fabien Moutarde, Centre de Robotique, MINES ParisTech, PSL, 5/2/2019

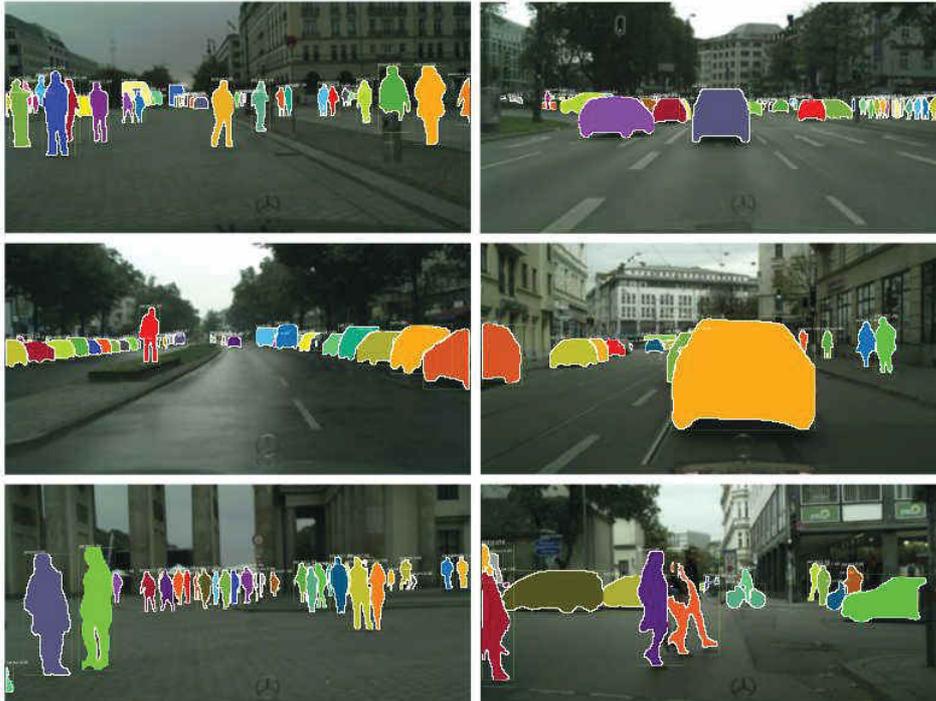
## Détecter et catégoriser visuellement les d'objets



## Détection et catégorisation visuelles avec un Faster\_RCNN

Deep-Learning pour la Robotique Collaborative, Pr. Fabien Moutarde, Centre de Robotique, MINES ParisTech, PSL, 5/2/2019

# Déterminer par DL les contours des objets



**Mask R-CNN** extrait les contours et formes des objets  
(au lieu de juste des boîtes englobantes)

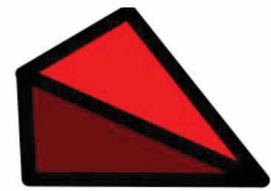
# Exemple de Segmentation sémantique par DL



[C. Farabet, C. Couprie, L. Najman & Yann LeCun: Learning Hierarchical Features for Scene Labeling, IEEE Trans. PAMI, Aug. 2013.]

La segmentation sémantique catégorise aussi les régions étendues comme “route”, “sol”, “herbe”, “bâtiment”, “mur”, etc (et pas uniquement les “objets”)

# PoseNet : ego-localisation (position et orientation) par DL



Input RGB Image

Convolutional Neural Network (GoogLeNet)

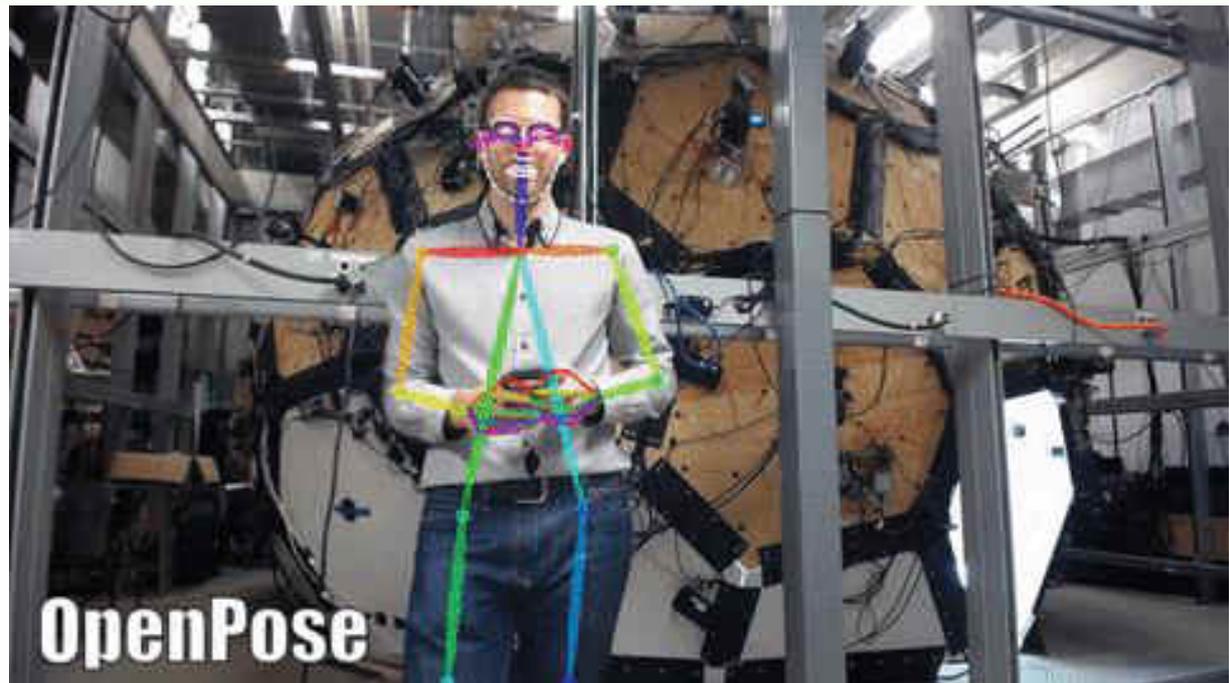
6-DOF Camera Pose



Figure 4: Map of dataset showing training frames (green), testing frames (blue) and their predicted camera pose (red). The testing sequences are distinct trajectories from the training sequences and each scene covers a very large spatial extent.

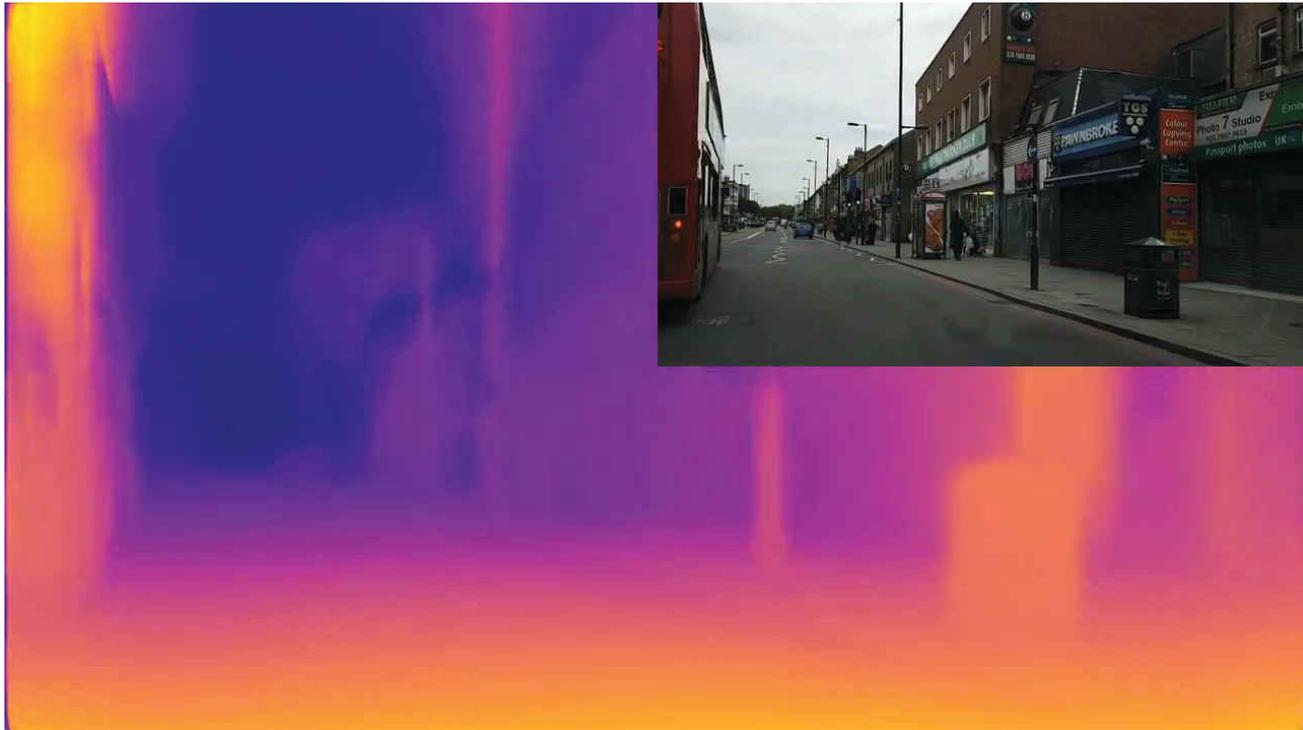
[A. Kendall, M. Grimes & R. Cipolla, "PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization" , ICCV'2015, pp. 2938-2946]

# Estimation de postures par Deep-Learning



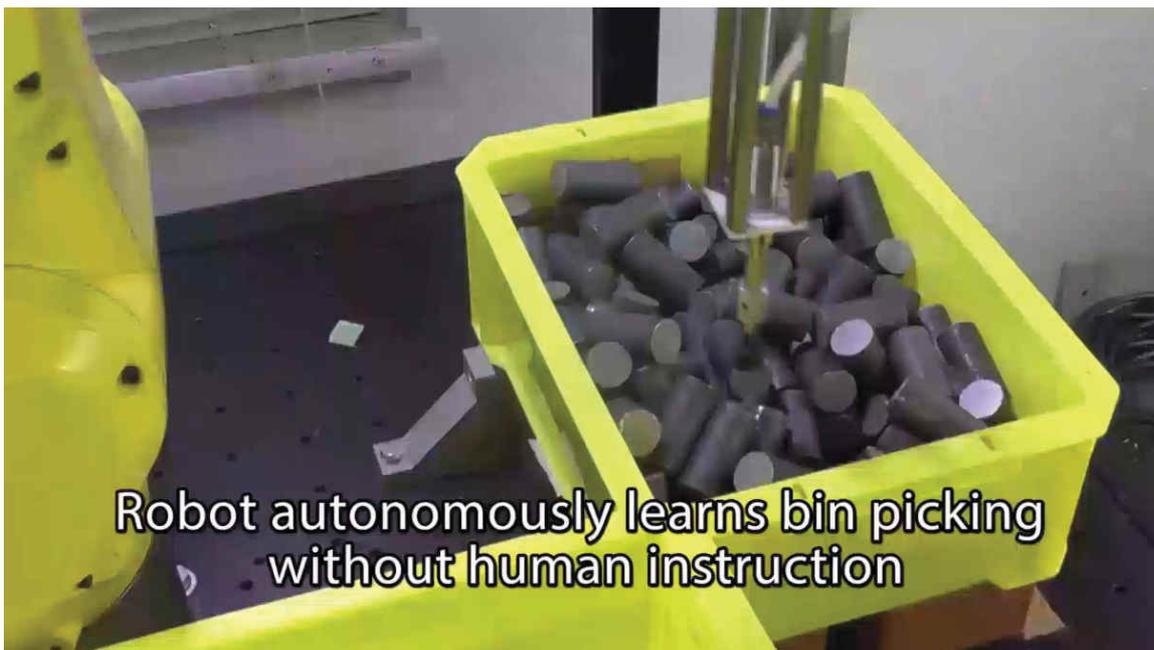
Real-time estimation of Human poses on RGB video  
[Realtime Multi-Person 2D Pose Estimation using Part Affinity Field, Cao et al., CVPR'2017 [CMU]]

# Inférence de la 3D (depth) à partir de monovision RGB



*Unsupervised monocular depth estimation with left-right consistency*  
*C Godard, O Mac Aodha, GJ Brostow - CVPR'2017 [UCL]*

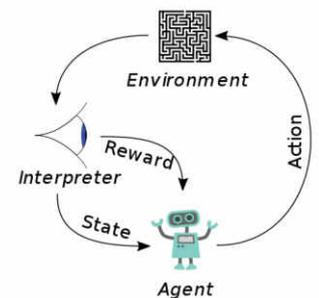
# Apprentissage autonome de tâche par un robot



Robot autonomously learns bin picking  
without human instruction

**Supervised learning, but with success/failure easily  
estimated automatically, for a bin-picking task**

# Apprentissage de comportement par renforcement



Deep-Learning pour la Robotique Collaborative, Pr. Fabien Moutarde, Centre de Robotique, MINES ParisTech, PSL, 5/2/2019

## Plan

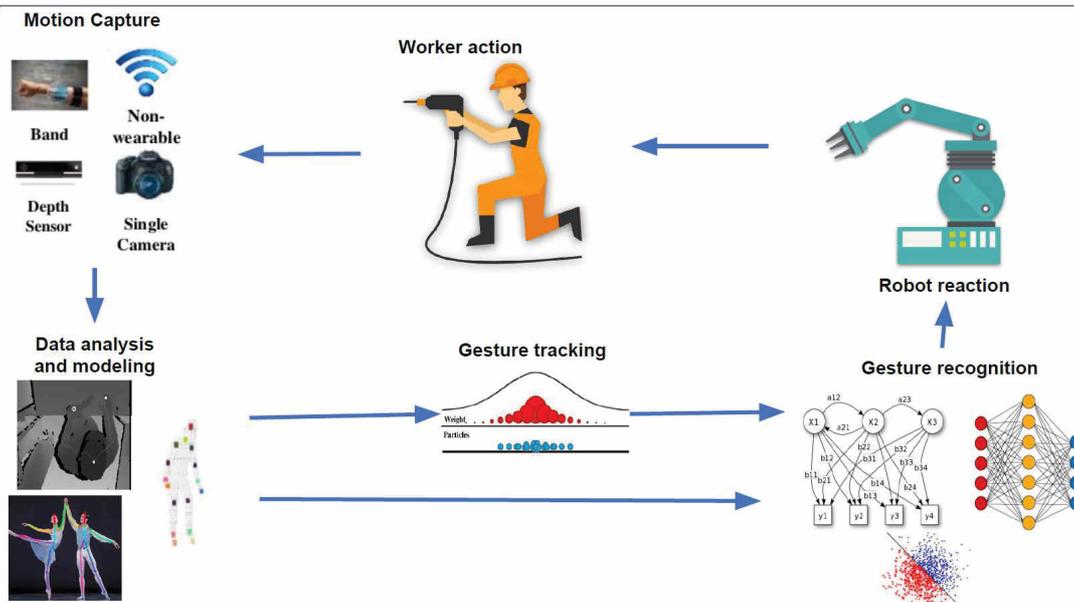
- LES Intelligences Artificielles et l'Apprentissage Machine (ML)
- Les IAs pour la Robotique Collaborative
- Que permet le Deep-Learning sur des images ?
- **Reconnaissance de gestes/actions pour Interaction et Collaboration Homme-Robot**
- Apprentissage de comportement de robots par Imitation ou Renforcement

# Collaboration et Interaction Homme-Robot

Besoin d'analyser et interpréter mouvements, actions & activités :

- Reconnaissance d'actions pour les robots collaboratifs
- Inférence des intentions des Humains (surtout pour robots mobiles)
- Communication Gestuelle

# Cadre général de la Collaboration Homme-Robot

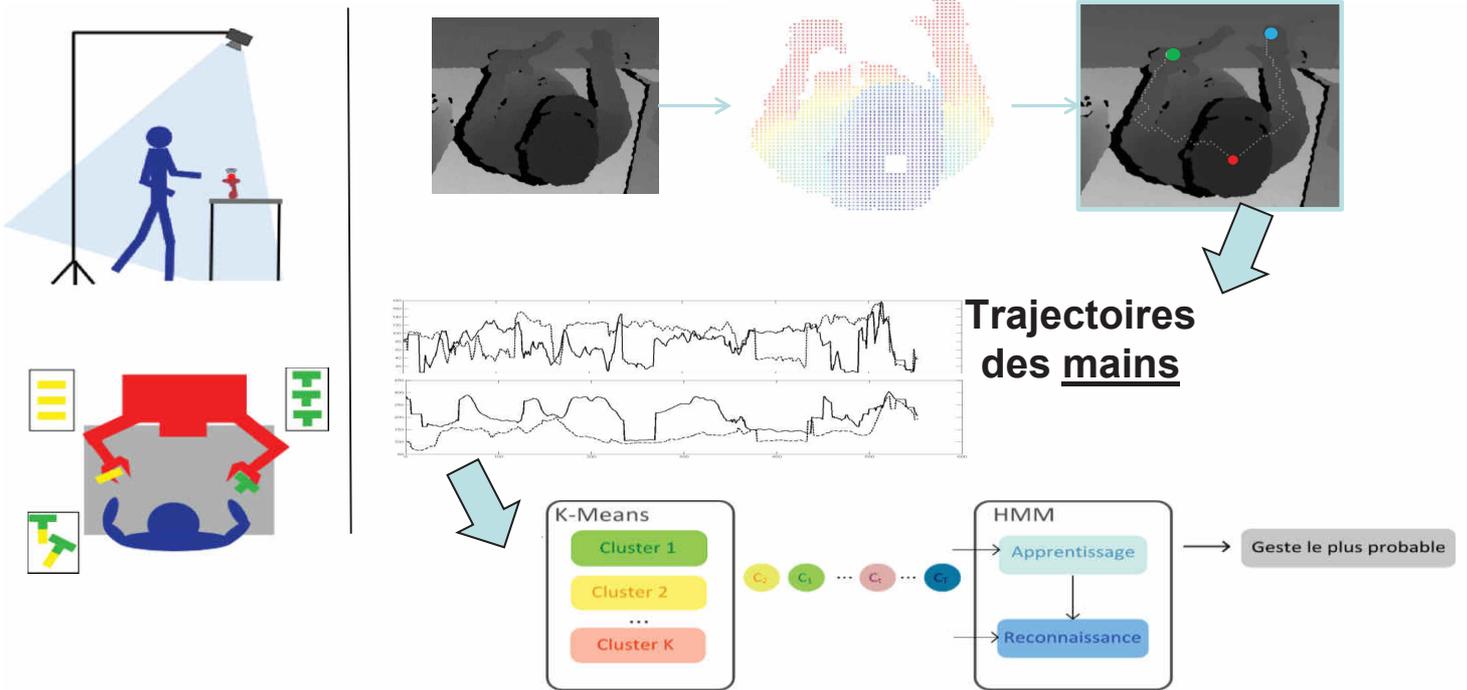


H2020 European project COLLABORATE (oct.2018-2021)



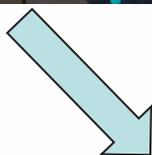
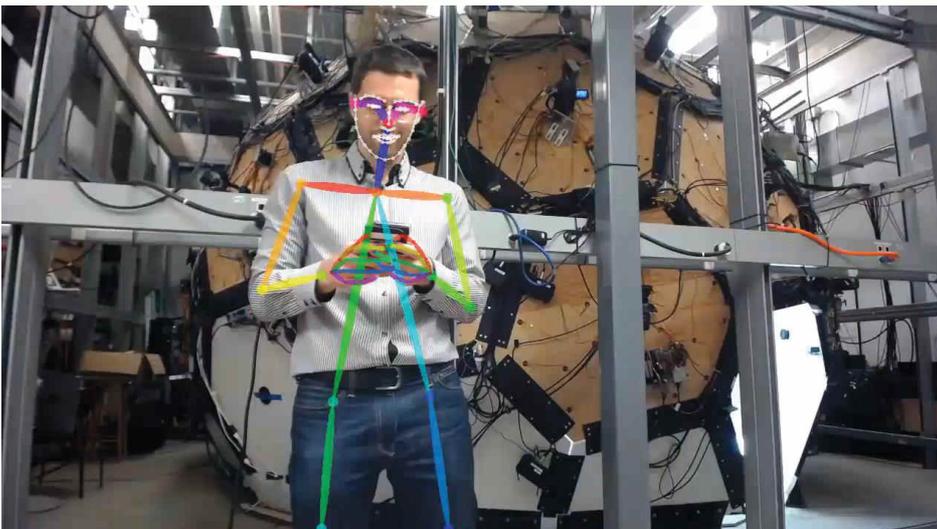
<https://collaborate-project.eu/>

# Exemple de reconnaissance d'actions pour Collaboration Homme-Robot

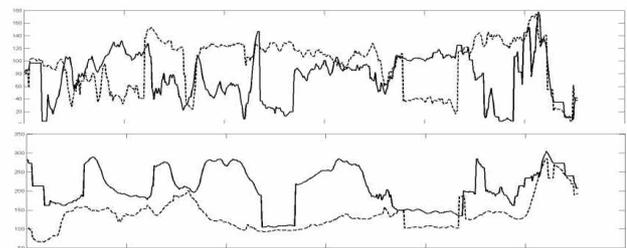


*PhD thesis of Eva Coupeté at MINES\_Paris (defended in 2016), sponsored by Chaire PSA "Robotique et Réalité Virtuelle"*

# Estimation de postures possible avec caméra RGB ! (openPose)

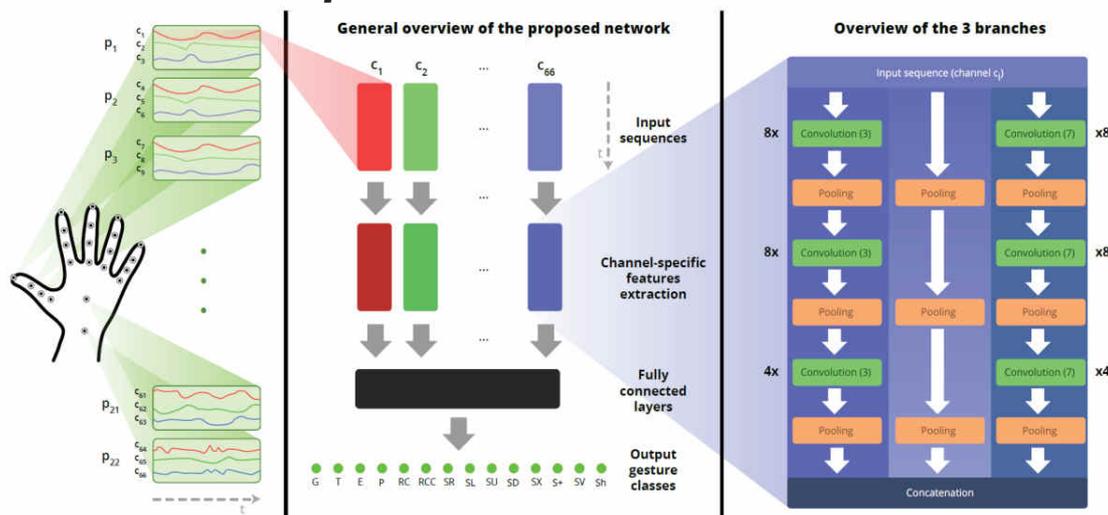


Trajectoires des articulations



## Deux approches principales :

- Deep Recurrent Neural Network (RNN) e.g. LSTM or GRU
- Convolutions *Temporelles*



["Convolutional Neural Networks for Multivariate Time Series Classification using both Inter- and Intra- Channel Parallel Convolutions"](#), G. Devineau, W. Xi, F. Moutarde and J. Yang, RFIAP'2018.

["Deep Learning for Hand Gesture Recognition on Skeletal Data"](#), G. Devineau, W. Xi, F. Moutarde and J. Yang, FG'2018.

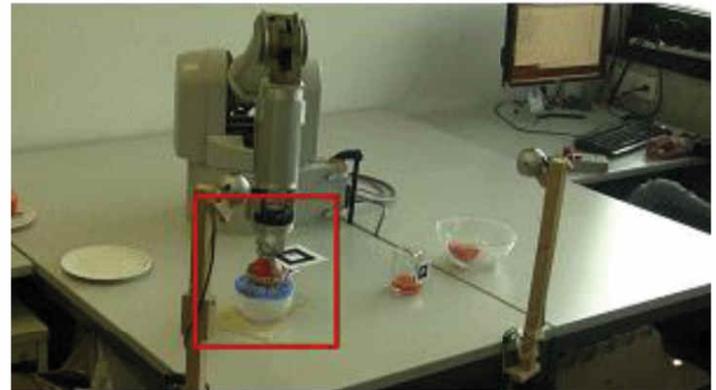
[PhD thesis of Guillaume Devineau @ MINES\_ParisTech, supervised by me]

Deep-Learning pour la Robotique Collaborative, Pr. Fabien Moutarde, Centre de Robotique, MINES ParisTech, PSL, 5/2/2019

## Plan

- LES Intelligences Artificielles et l'Apprentissage Machine (ML)
- Les IAs pour la Robotique Collaborative
- Que permet le Deep-Learning sur des images ?
- Reconnaissance de gestes/actions pour Interaction et Collaboration Homme-Robot
- Apprentissage de comportement de robots par Imitation ou Renforcement

# Enseigner une tâche au robot "par démonstration"



## « Learning from Demonstration » (LfD)

Deep-Learning pour la Robotique Collaborative, Pr. Fabien Moutarde, Centre de Robotique, MINES ParisTech, PSL, 5/2/2019

# Robots collaboratifs incluant LfD



Yumi robot by ABB



Baxter Robot by Rethink Robotics



LWR by KUKA



UR3, Universal Robotics



Welding Robots, Robotik



Carbon Robotics

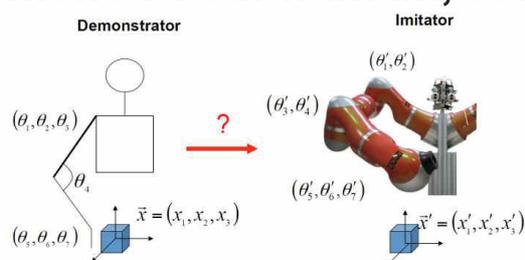
# De l'apprentissage par Démonstration, à l'apprentissage par Imitation ?

## Plusieurs types de modalités possibles pour LfD :

- Kinesthésique : « tenir la main » du robot
- Haptique : donner des feedbacks « physiques » au robot

## Ou encore apprentissage par Imitation par :

- Capteurs de mouvement : enregistrer mouvement réalisé par démonstrateur humain, et *mettre en « correspondance »*

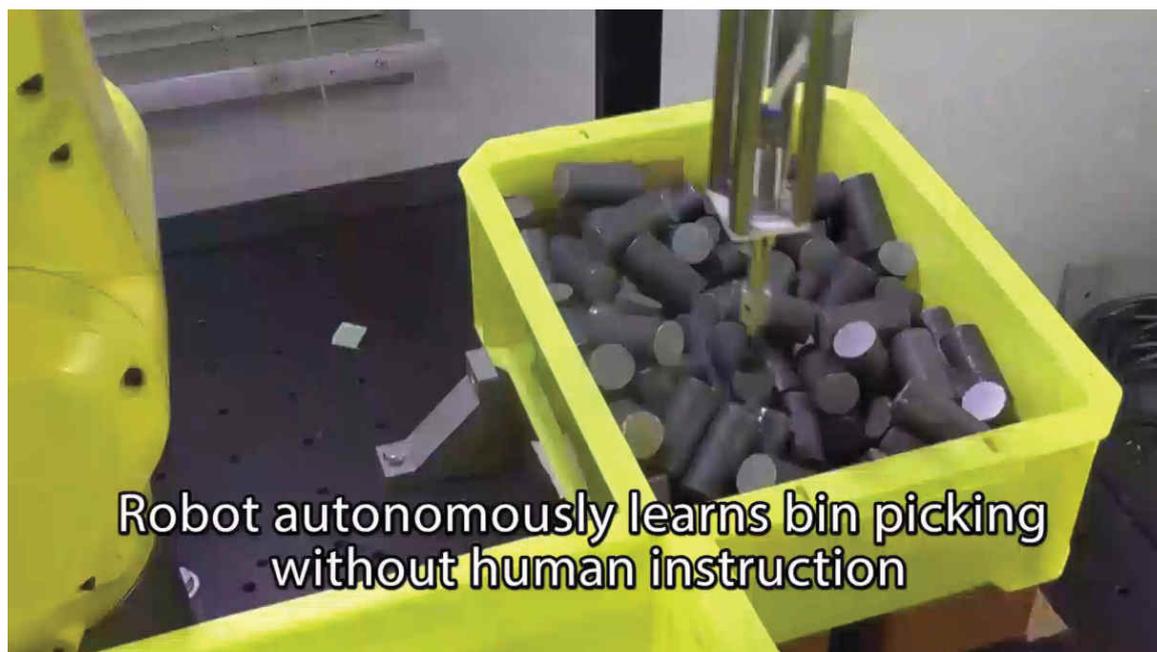


- Vision : analyse par vision de démonstration par un humain

*Pb : pas d'info haptique (force),  
+ erreur/imprécision potentielle  
de la squelettisation*

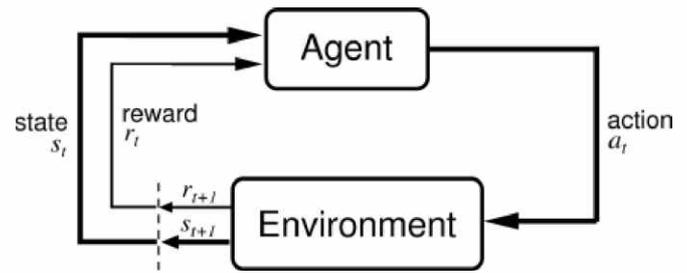
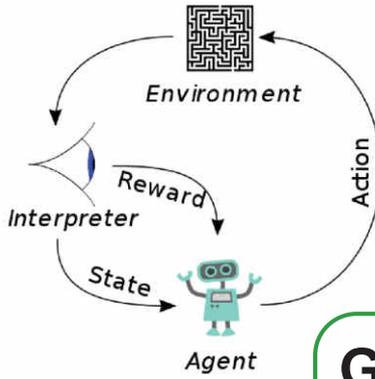


# Apprentissage autonome de tâche par un robot



Supervised learning, but with success/failure easily  
estimated automatically, for a bin-picking task

# Apprentissage par Renforcement (Reinforcement Learning, RL)



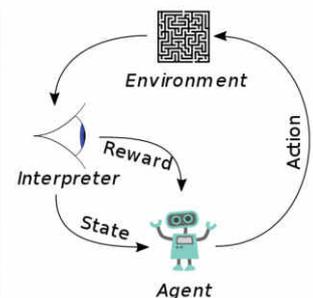
**Goal: find a “policy”  $a_t = \pi(s_t)$  that**

**Maximizes** 
$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}, \gamma \in [0, 1[$$

**Deep Reinforcement Learning (DRL)** if Deep NeuralNet used as model (for policy and/or its “value”): DQN, Actor-Critic A3C

**End-to-end driving: policy  $\pi$  searched as ConvNet(front-image)**

# Apprentissage de comportement par renforcement



# Conclusions

- **Principaux challenges actuels pour la Robotique Collaborative et Intelligente :**  
**Interaction Humain-Robot, reconnaissance ou/et compréhension des actions ou comportements humains, inférence des INTENTIONS Humaines, apprentissage de COMPORTEMENTS ADAPTATIFS complexes**
- **Les Réseaux Convolutifs profonds peuvent faire bien + que classifier des images : segmentation sémantique, ego-localisation par vision, estimation de posture Humaine, inférence de profondeur par monovision, apprentissage de comportements adaptatifs complexes fondés sur image**
- **Pour analyse des mouvements ou intentions Humains, combiner estimation de postures par DL avec Convolution Temporelle profonde de séries temporelles semble prometteur**
- **Pour apprentissage de comportement, le Deep Reinforcement Learning fondé images obtient déjà des résultats intéressants**

# Questions ?

