

Intelligence Artificielle & Véhicule Autonome : Quelle IA pour quelles fonctions ? Avancées récentes et perspectives

**Pr. Fabien Moutarde
Centre de Robotique (CAOR)
Mines ParisTech
PSL Research University**

`Fabien.Moutarde@mines-paristech.fr`

`http://people.mines-paristech.fr/fabien.moutarde`

Faut-il de l'intelligence pour conduire ?

Démo de conduite automatisée par le centre de Robotique de MinesParis durant... IV'2002 !

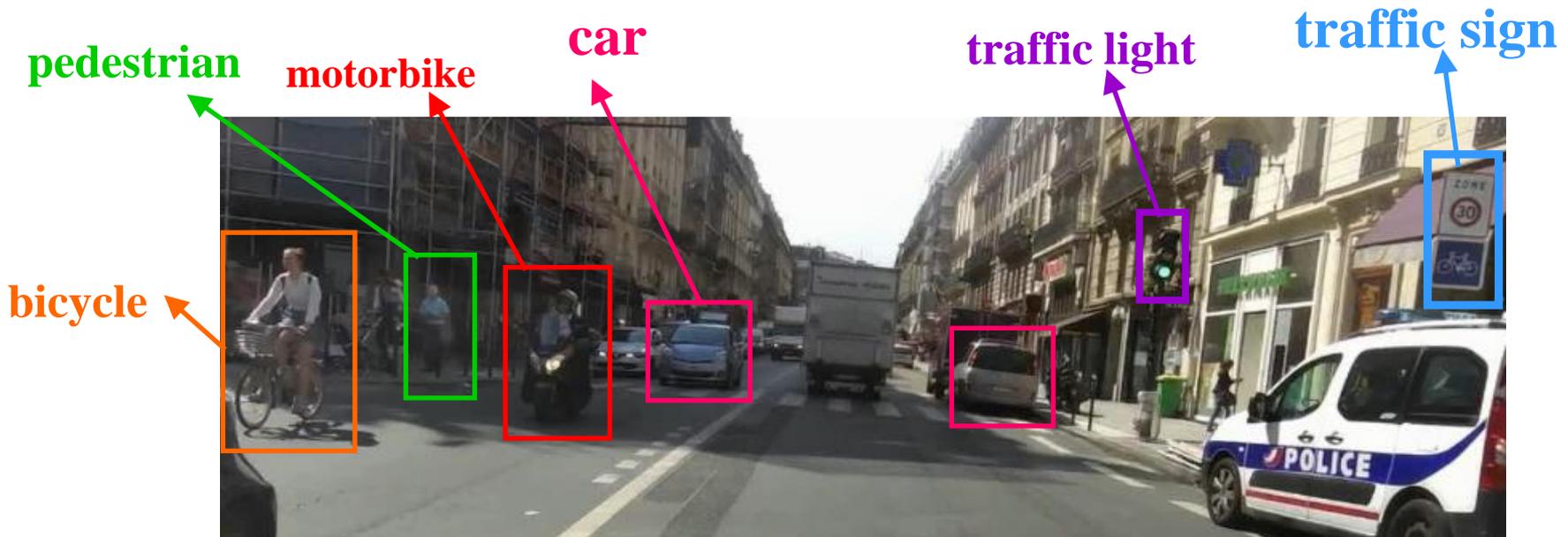


Boucle robotique classique et assez simple : asservissement commande du volant selon estimation position par rapport à route

En circulation « ouverte », c'est + compliqué (surtout en urbain)...



Besoin avant tout de « ***compréhension*** » ***temps-réel de scène***



**Composant-clef pour Aides à conduite (ADAS)
et la conduite automatisée**

- **Forte contrainte temps-réel : traiter au moins ~10 frames/second**

Intelligence Artificielle ? Ou surtout Machine-Learning ?

- **Intelligence Artificielle, un domaine très vaste et hétérogène :**
 - Algorithmes de jeu (échecs, Go, etc..)
 - Raisonnement à base de règles, systèmes experts
 - Multi-agents, émergence comportement collectif
 - ...
 - Optimisation, recherche opérationnelles, programmation dynamique,
 - **Planification (de trajectoire, de tâches, etc...)**
 - **Apprentissage machine (statistical Machine-Learning)**
= modélisation fondée sur des données
(*optimisation, via exemples, de modèle paramétrique*)

**Principalement les 2 derniers sous-ensembles de l'IA
qui sont pertinents pour les véhicules autonomes**

« IA » : pour quels défis du véhicule autonome ?

- **Interprétation « sémantique » environnement véhicule :**
 - Détection et compréhension signalisation routière (panneaux, feux, ...)
 - Détection et catégorisation des « objets » autour du véhicule (voitures, piétons, vélos,...)
 - Prévision des mouvements des autres usagers
 - Inférence intentions conducteurs/piétons (ou policier !) d'après leurs regards/gestes
 - Ego-localisation du véhicule + précise (<1m) que GPS
- **Planification de trajectoire (y compris vitesse)**
dans environnement dynamique incertain
- **Pour ADAS et automatisation niveaux 2-3 :**
 - Analyse et compréhension attention et activités ou gestes du « conducteur-superviseur »
- **A + grande échelle spatio-temporelle :**
 - Prévision évolution trafic pour optimiser itinéraires

Où en était la perception visuelle pour véhicules « intelligents » vers 2012 ?



Algo du CAOR pour TSR vers 2011

→ ~95% OK

[HoG features + SVM pour détection
+ RDF pour reconnaissance]



Algo du CAOR pour ADAS vers 2009

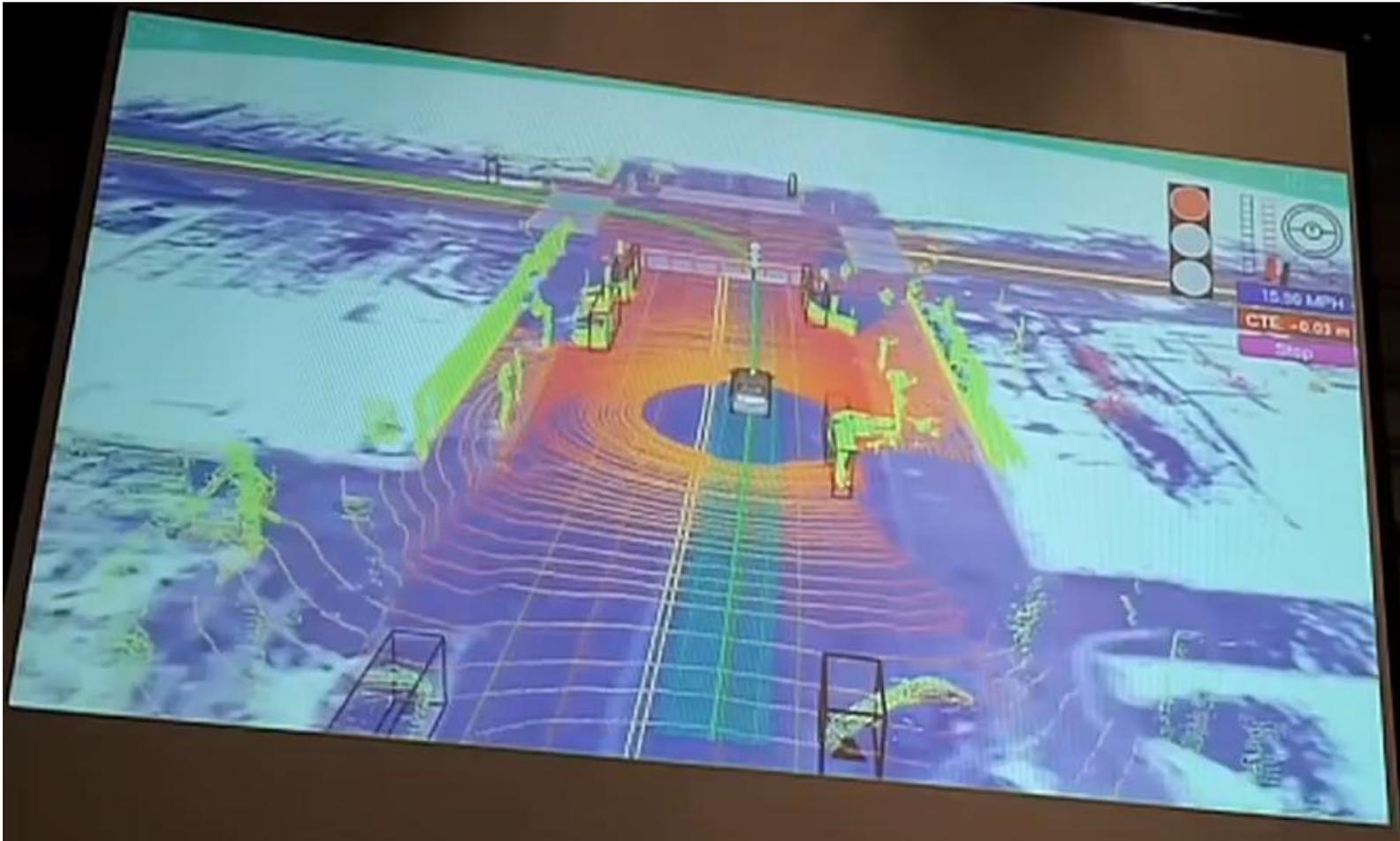
→ ~95% OK (voitures) et ~80% OK (piétons)

[ControlPoints features + adaBoost]



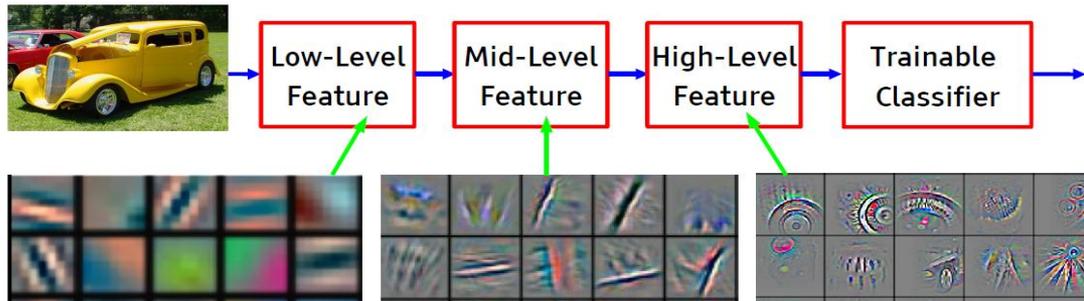
- **Features « sur mesure »**
- **Classif par adaBoost, SVM, RDF**
- **Combinaisons différentes, optimisées pour chaque catégorie (panneaux, véhicules, piétons, ...)**

Et le Google-car ?



Utilisation capteur laser 360° multi-couches Velodyne
avec segmentation 3D + catégorisation 3D

Quels avantages apportés récemment par apprentissage profond (convNet) ?



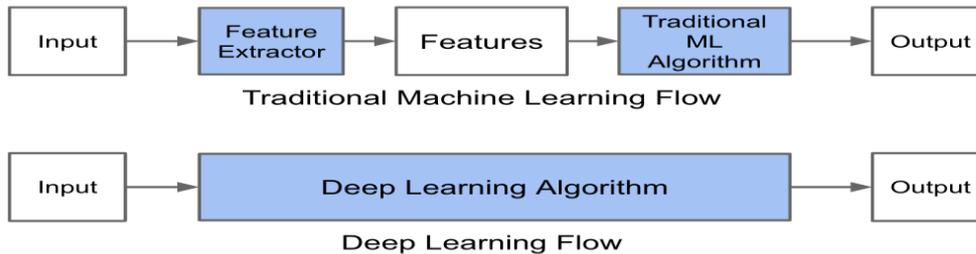
Principe : apprendre FEATURES en + de classification



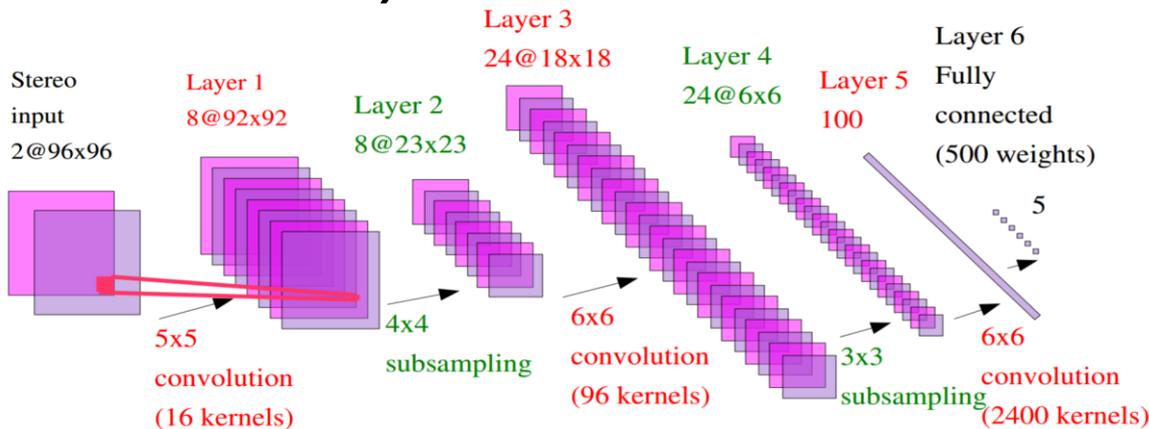
[C. Farabet, C. Couprie, L. Najman & Yann LeCun: Learning Hierarchical Features for Scene Labeling, IEEE Trans. PAMI, Aug.2013.]

- Unification algos détection/reco pour toutes catégories
- Plus grande robustesse car appris sur base ENORME

Que sont les Convolutional Networks (et apprentissage profond) ?



CNN= Succession de couches Convolutions_paramétrées/Pooling, + apprentissage simultané (par descente gradient, avec exemples images+labels-classe) de REPRESENTATION et classification

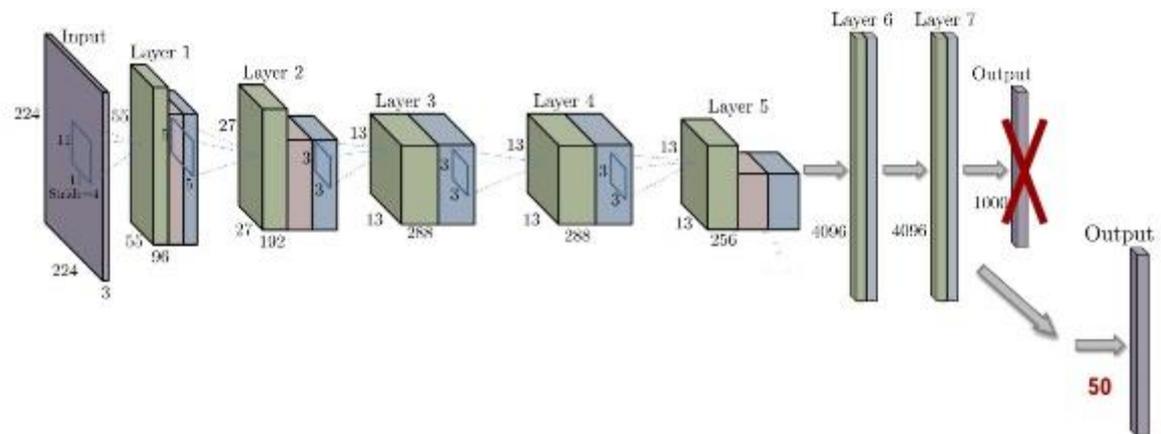


**Pour AlexNet/GoogLeNet, ~ 5-10 couches convolutions+pooling
 → millions de paramètres à optimiser !**

→ besoin base apprentissage ENORME type ImageNet (millions d'exemples et milliers de classes !!) pour apprendre « from scratch »

Généralité de la représentation obtenue par apprentissage profond !

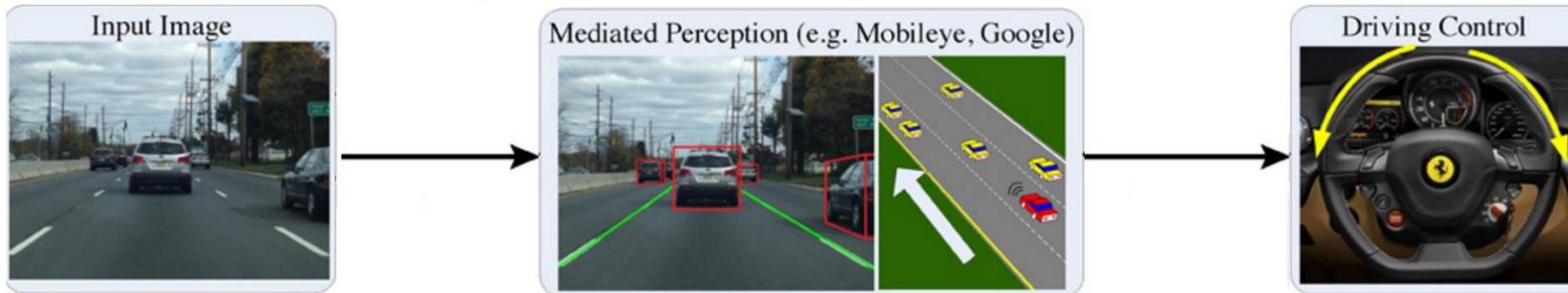
En enlevant dernières couches (celles de classification) de convNet appris sur ImageNet, on obtient une transfo de l'image d'entrée en une représentation « semi-abstraite », qui peut servir de base pour apprendre autre chose (« transfer learning ») :



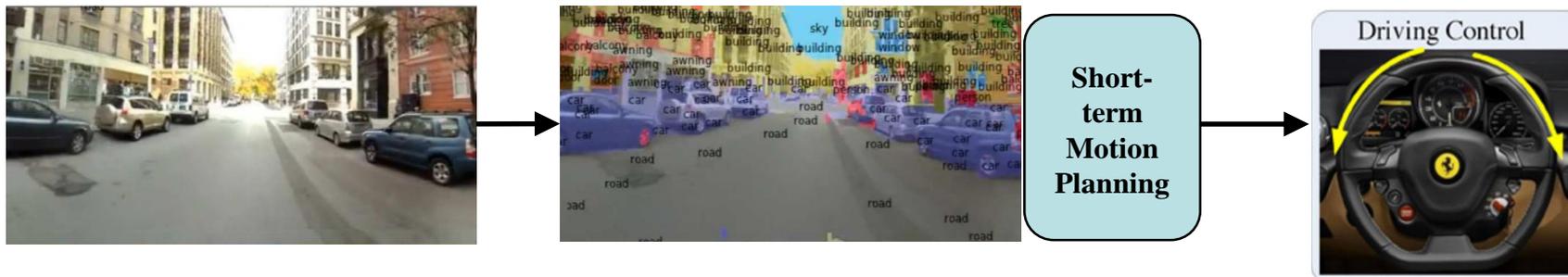
- Reconnaissance AUTRES catégories
- Contrôle direct volant (DeepDrive)
- Localisation précise (position+cap) = PoseNet
- ... et même informations tridimensionnelles

Plusieurs chaînes de traitement possibles pour la conduite automatisée

- Approche « classique » :



- Sémantisation image par ConvNet (Deep-Learning) :



- End-to-end Deep-Learning (nVidia, Berkeley) :



Points encore « durs » pour conduite automatisée :

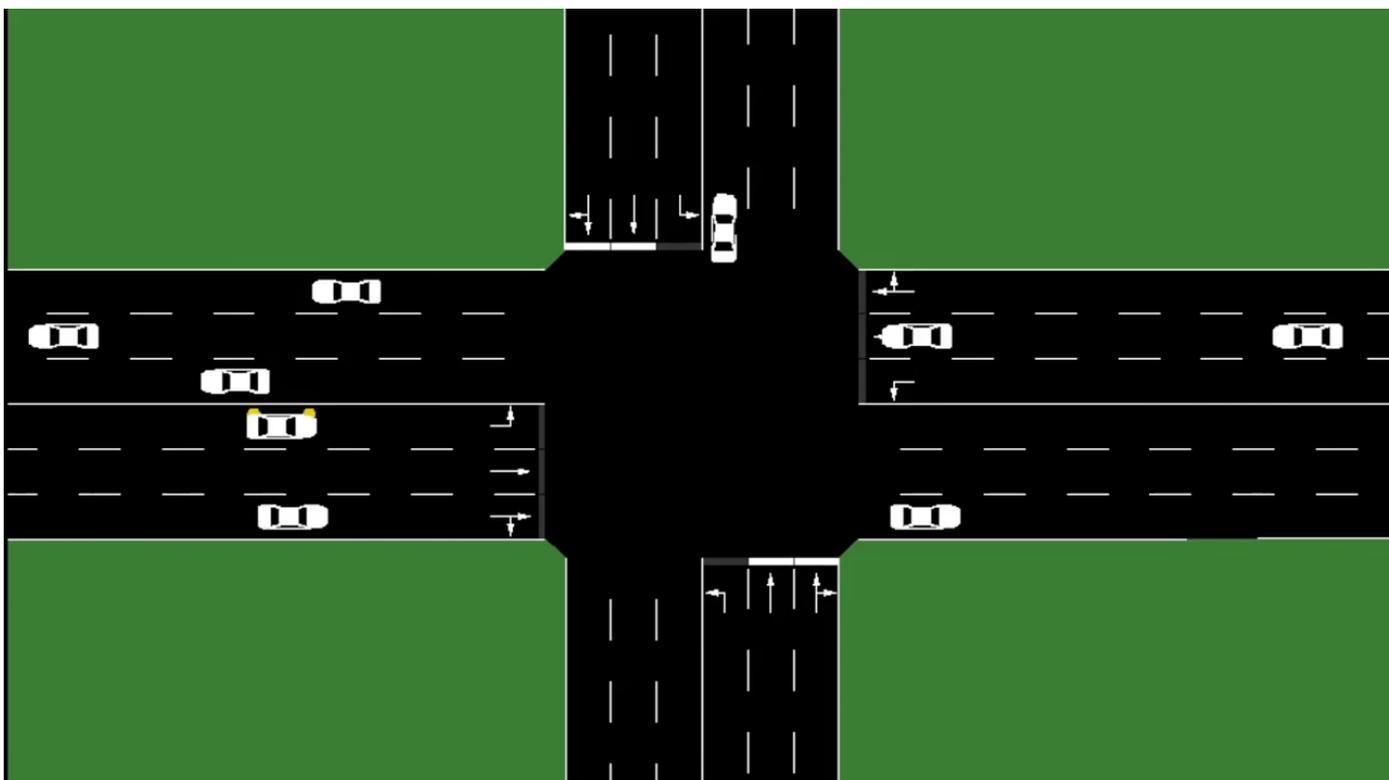
- Collecte **MASSIVE** de données réelles (image/radar/lidar/...)
- Validation quantifiée de la fiabilité des algos !
- **Ego-localisation précise**
- Manœuvres d'urgence
- **Suivi attention/activité conducteur-superviseur**
- Prévision **mouvements/trajectoires** des autres usagers
- Inférence intentions conducteurs/piétons (ou policier !)
d'après leurs **regards/gestes** (et aussi infos VA → autres usagers)

Points « durs » pour conduite automatisée et coopérative :

- Interactions entre véhicules autonomes, et avec autres
- Perception collaborative
- **Intersections automatisées**

*Chaire **DriveForAll**, financée par Safran+Valeo+PSA,
labos = CAOR/Mines_ParisTech (leader)+Berkeley +EPFL+SJTU*

Questions ?



Travail récent CAOR (+INRIA) sur coordination véhicules dans projet FP7 AutoNet2030