



Mission critical solutions

# Journée Scientifique Annuelle<sup>1</sup>

CRIANN

19 Novembre 2020

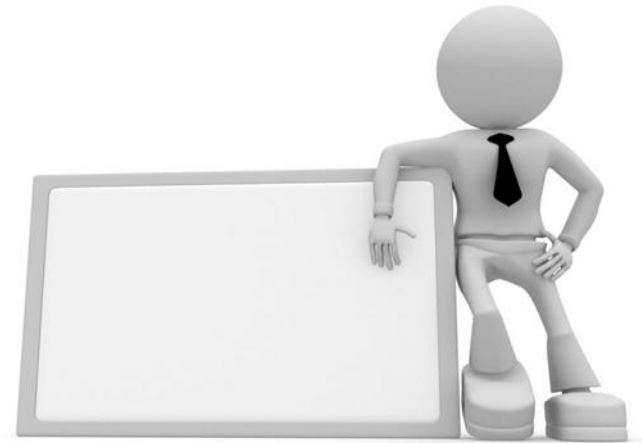


**Contribution à la Perception  
d'Environnement pour la Smart Mobilité.**

Contribution to Environment Perception for Smart  
Mobilité.

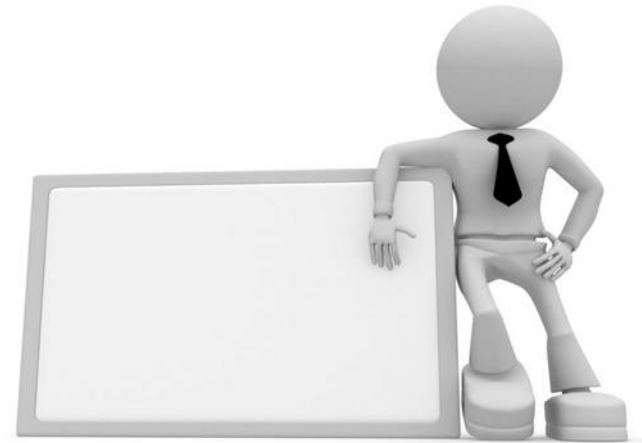


- Contexte & Objectifs
- Perception d'Environnement pour Smart Mobilité
- Projets de Recherche
- Conclusion & Perspectives

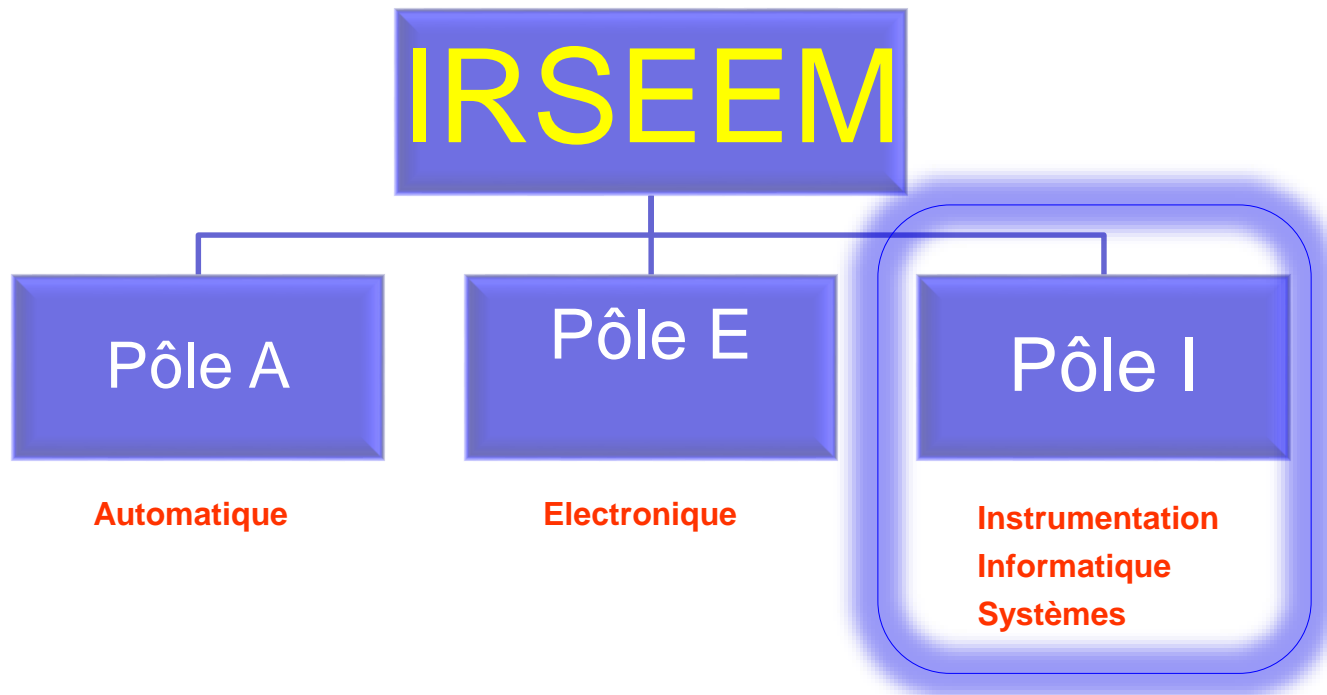




- **Contexte & Objectifs**
- Perception d'Environnement pour Smart Mobilité
- Projets Recherche



## Contexte & Objectifs :

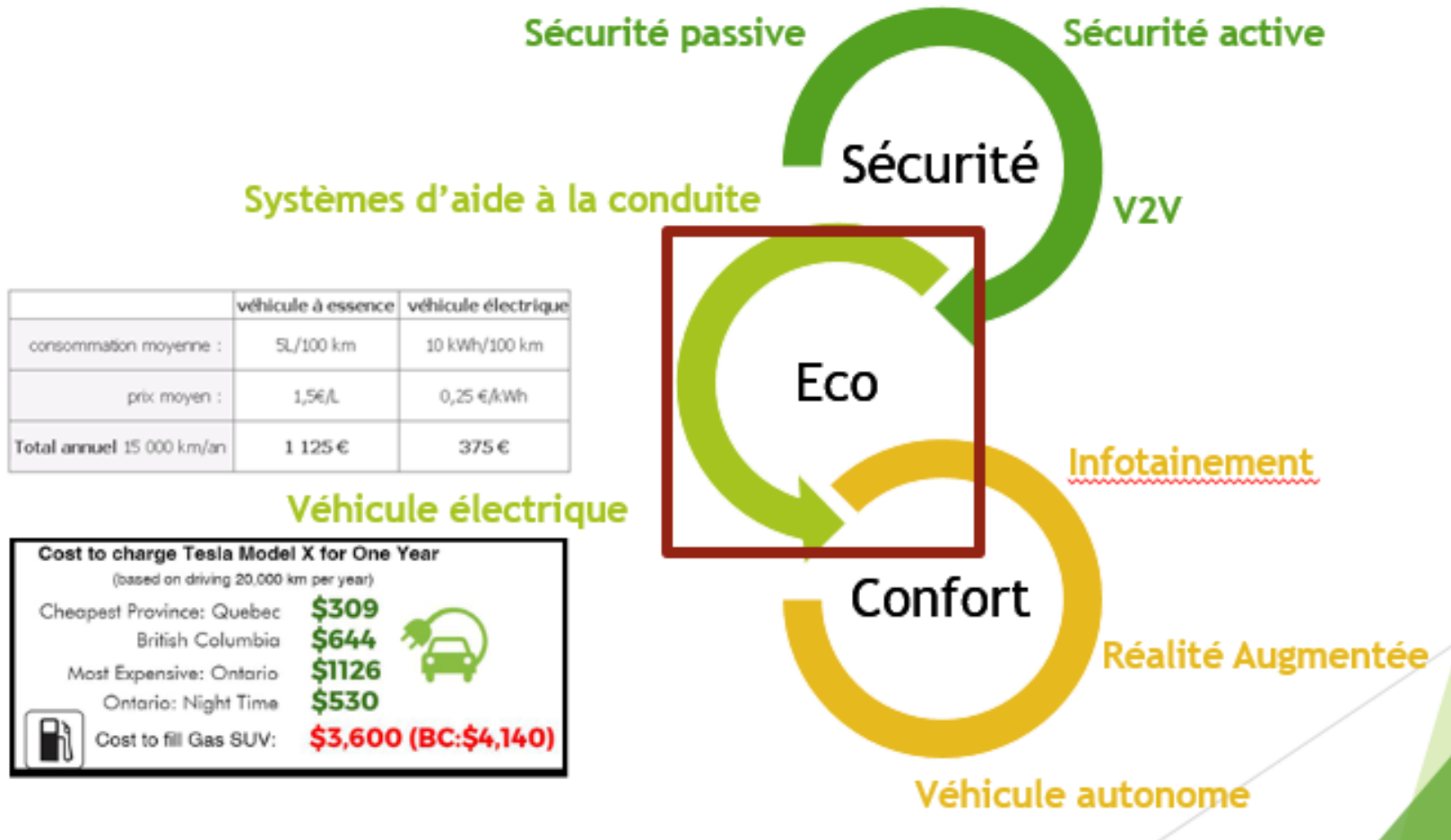




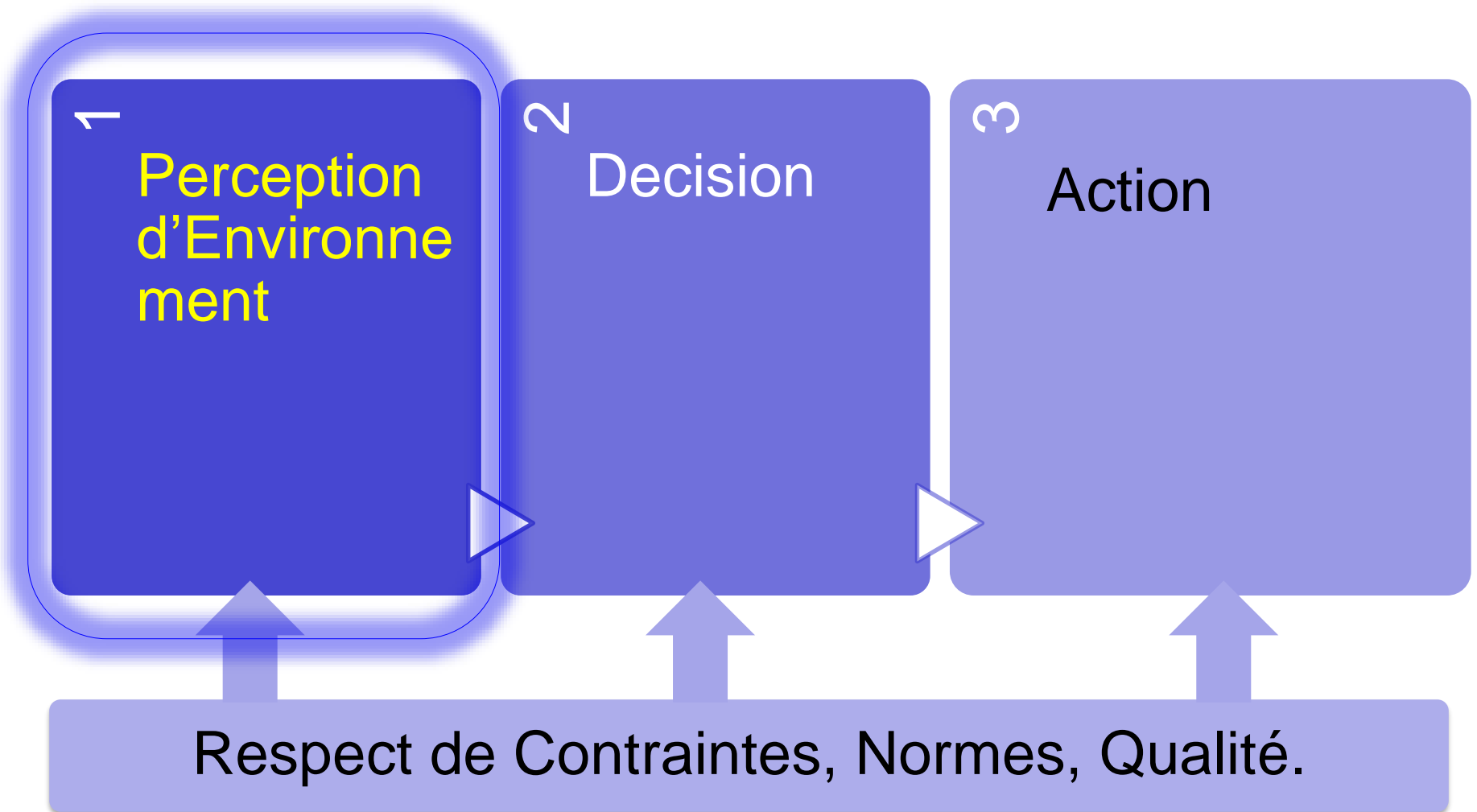
## Contexte & Objectifs :

- Défi de la Smart Mobilité :
  - Développer de nouvelles solutions de mobilité bas-Carbone **efficaces** et **sécurisées**
  - Renforcement de la **sécurité routière** et réduction des **accidents** (objectif **Zéro Accidents**)
  - Réduction de la consommation **d'énergie** et de **l'empreinte** carbone
  - Développer la mobilité **douce** : piétons, vélos, etc.

## Contexte & Objectifs :

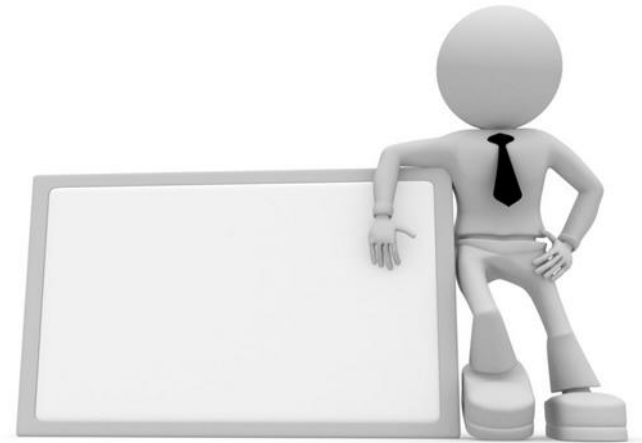


# Chaine d'Acquisition et de Traitement de Données





- Contexte & Objectifs
- Perception d'Environnement pour Smart Mobilité
- Projets Recherche : M2NUM Phase I





## Contexte & Objectifs :

### ■ Projet M2NUM :

#### ■ WP1.5 (Aide à la Conduite et Détection de Piétons) :

- Détection de piétons et reconnaissance d'objets basées Deep Learning
- Estimation de distances des objets
- Analyse du comportement du piéton
- Validation de la plateforme : véhicule autonome



## **Plateforme (TRL5/6)**

Véhicule Autonome : Perception de l'Environnement Proche  
(détection de piétons, reconnaissance d'objets, tracking)



## Plateforme de Détection d'Objets :

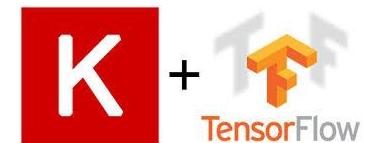
### ■ PC :

- CPU Intel i7 6700hq
- GPU Nvidia GTX960m 2G
- RAM 8G DDR3 et Disque dur SSD



### ■ MYRIA (CRIANN) :

- Supercalculateur
- Cluster de calcul de type x86-64 constitué de plusieurs unités de calculs CPU Intel et GPU Nvidia :



- 9 nœuds de 2 x **K80**
- 9 nœuds de 2 x **P100** : Un peu plus puissant que K80
- Millier CPU Intel



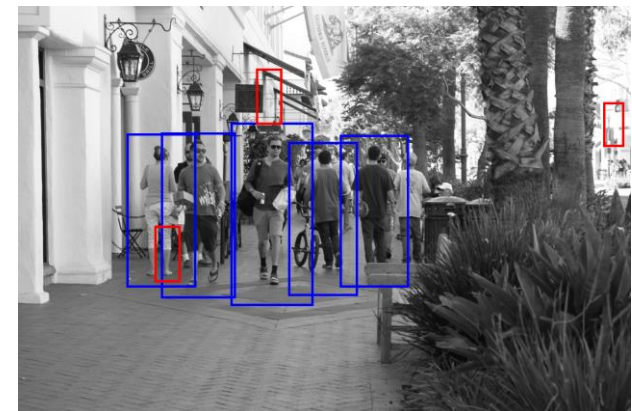
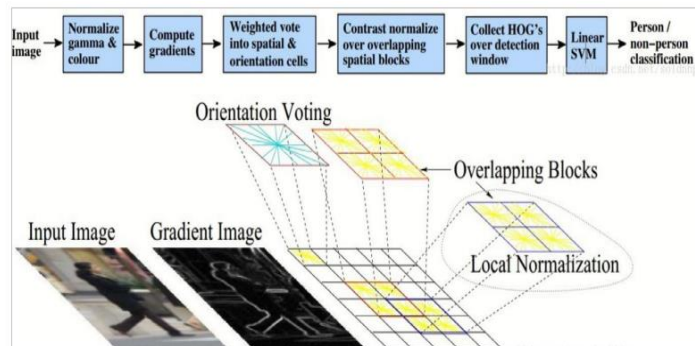
## Détection d'Objets basée HOG/DPM :

### ■ Approches de Reconnaissance de Formes

- Extraction des caractéristiques d'images
- Classification basée sur une base de données d'images positives (**avec des personnes**) et négatives (**sans personnes**)

### ■ Avantages & Inconvénients :

- Très précises et donnent relativement de bons résultats
- Coûteux en temps de calcul



## Détection d'Objets basée HOG/DPM :

- Bases de Données utilisées

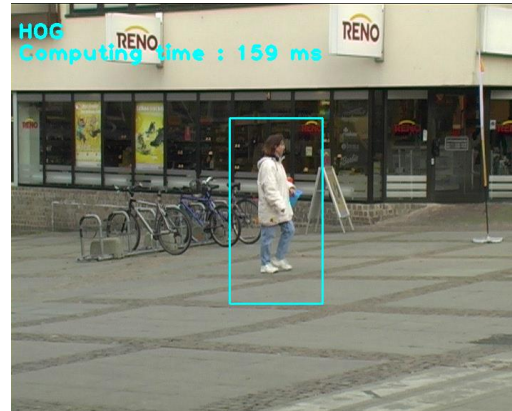


## Détection d'Objets basée HOG/DPM :

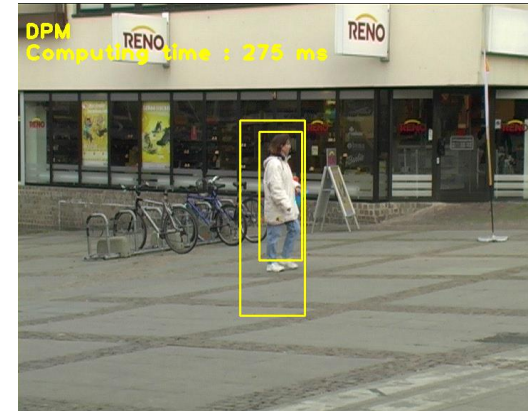
### ■ Comparaison : HAAR, HOG et DPM (TUD-Brussels & ETH)



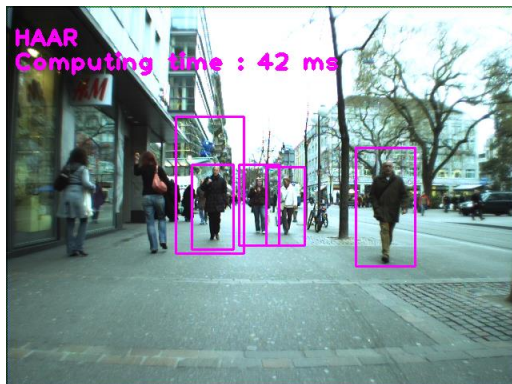
Temps de calcul : 55 ms  
Taille images : 640x480



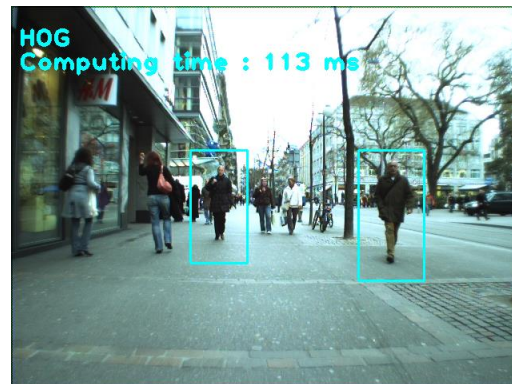
Temps de calcul : 159 ms



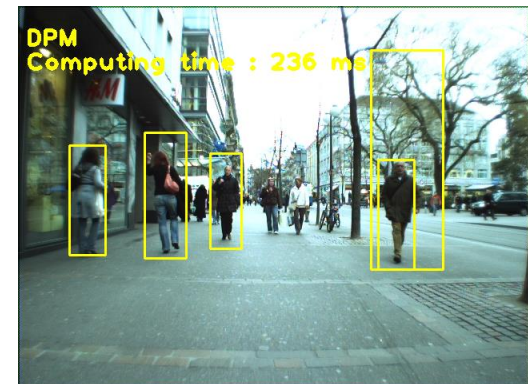
Temps de calcul : 275 ms



Temps de calcul : 42 ms



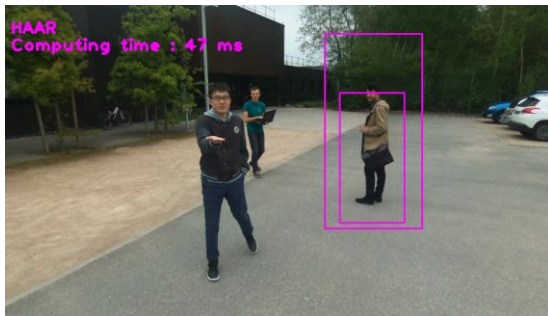
Temps de calcul : 113 ms



Temps de calcul : 236 ms

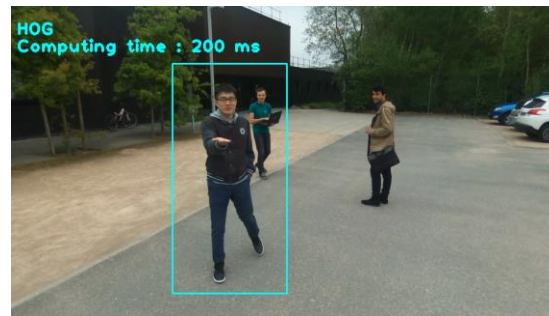
## Détection d'Objets basée HOG/DPM :

### ■ Comparaison : HAAR, HOG et DPM (ESIGELEC)

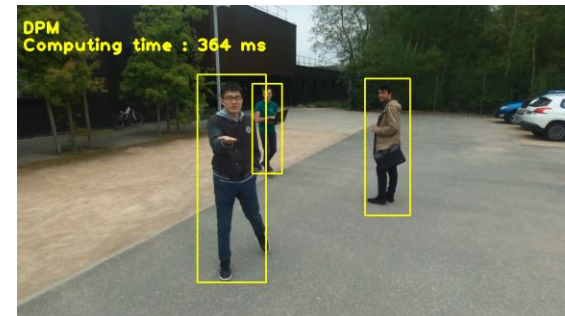


Temps de calcul : 47 ms

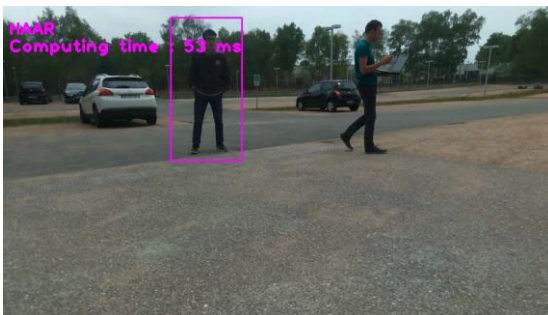
Taille images : 960x540



Temps de calcul : 200 ms



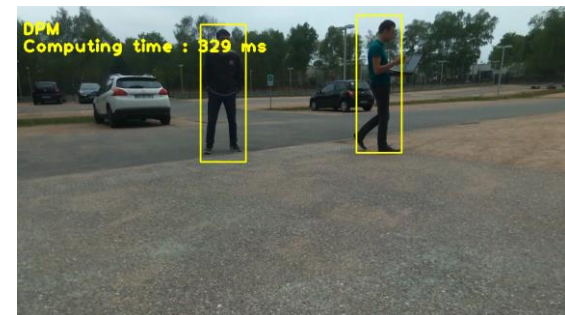
Temps de calcul : 364 ms



Temps de calcul : 53 ms



Temps de calcul : 208 ms



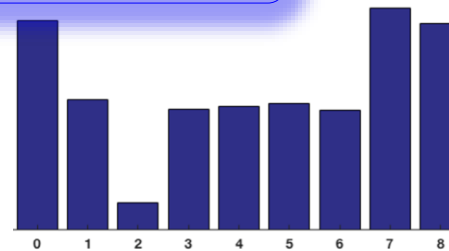
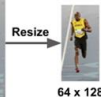
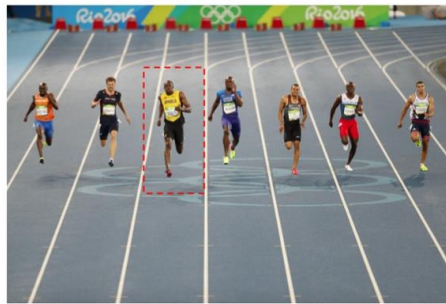
Temps de calcul : 329 ms

## Détection d'Objets basée HOG/DPM :

### ■ HOG/DPM Amélioré : **Faster-DPM**

- Préfiltrage d'images (correction gamma)
- Redimensionnement des images
- Gradients négatifs (18 niveaux, 0 à 360° )
- Normalisation (par blocs de 16x16, 32x32)
- Détourage de l'objet du fond de l'image :

■ Temps de calcul divisé par 4

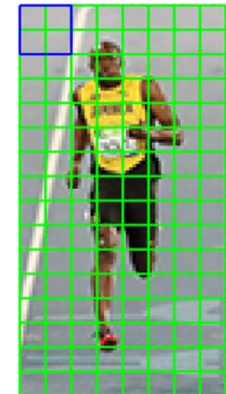


9 niveaux : 0 à 180°

Prise en compte des gradients négatifs

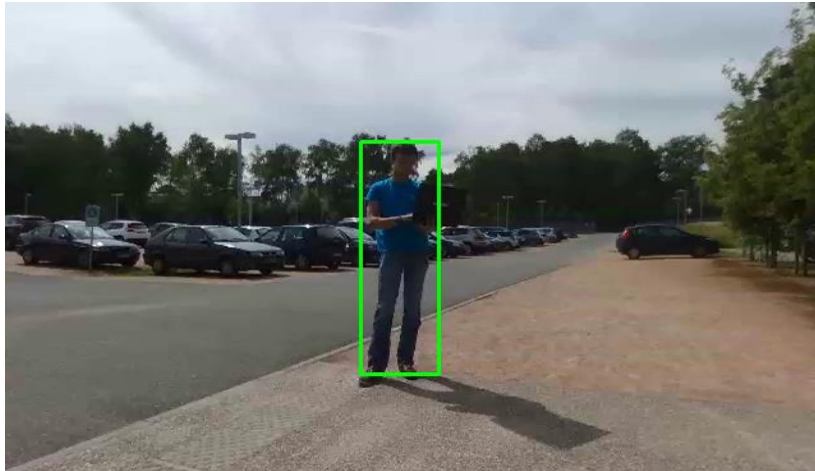
18 niveaux : 0 à 360°

Normalisation par blocs de 16x16 avec décalage de 8x8

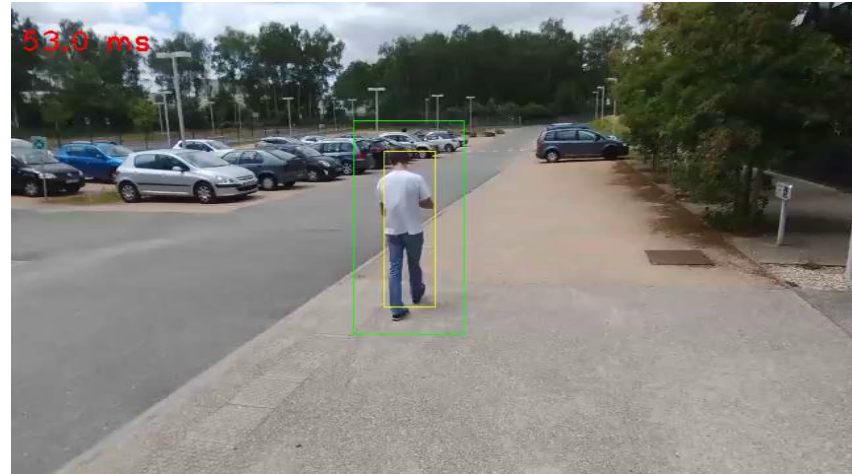


## Détection d'Objets basée HOG/DPM :

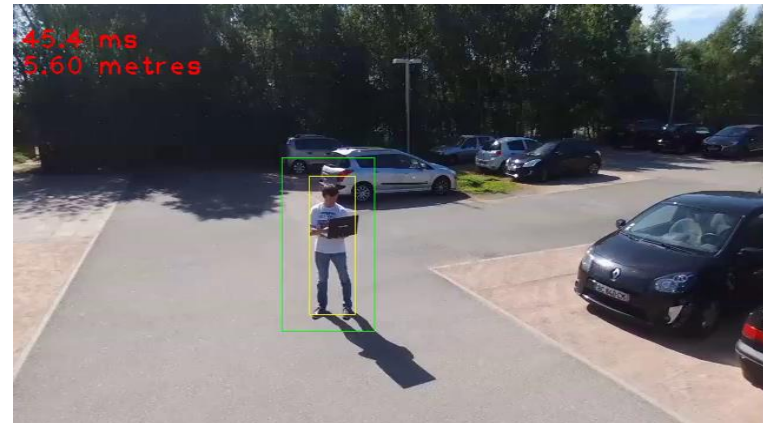
### ■ HOG/DPM Amélioré (Faster-DPM) vs HOG/DPM classique



HOG/DPM Classique



SDPM



$$\text{PID} : V_{\text{rotation}} = K_p * \text{erreur} + K_i * \int_0^t \text{erreur} * dt + K_d * \frac{d}{dt}(\text{erreur})$$

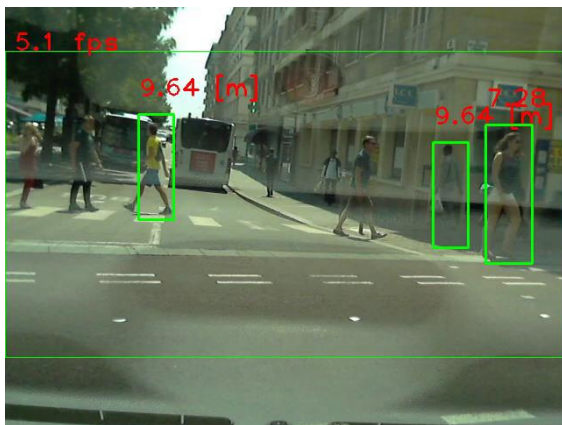
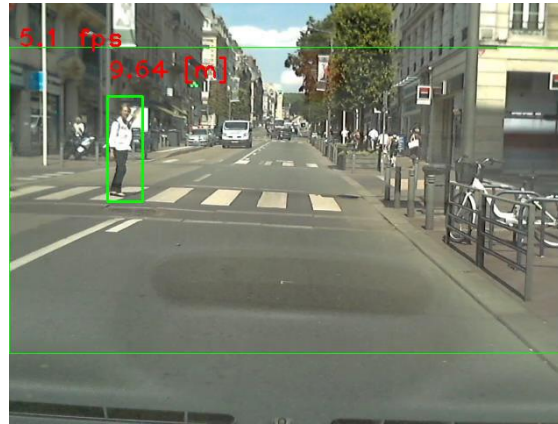
$$\text{STA}^1 : V_{\text{rotation}} = K_p * \left( \frac{\text{erreur}}{|\text{erreur}|+A} \right) * \sqrt{|\text{erreur}|} + K_i * \int_0^t \frac{\text{erreur}}{|\text{erreur}|+A} dt$$





## Détection d'Objets basée HOG/DPM :

### ■ Détection de Piétons basée HOG/DPM Amélioré (SDPM)

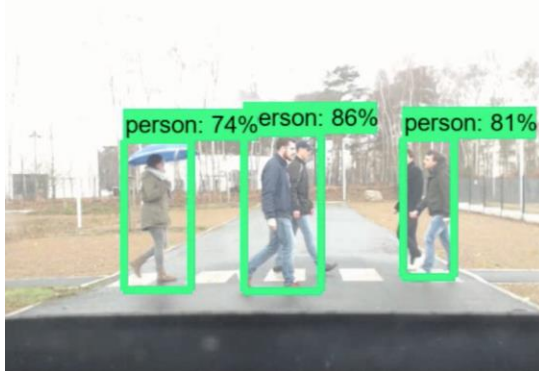
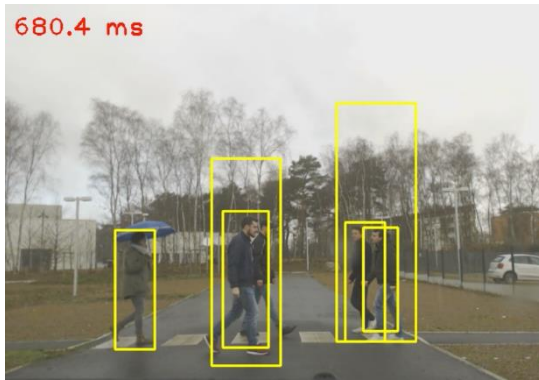


Essais Environnement Réel - Rouen

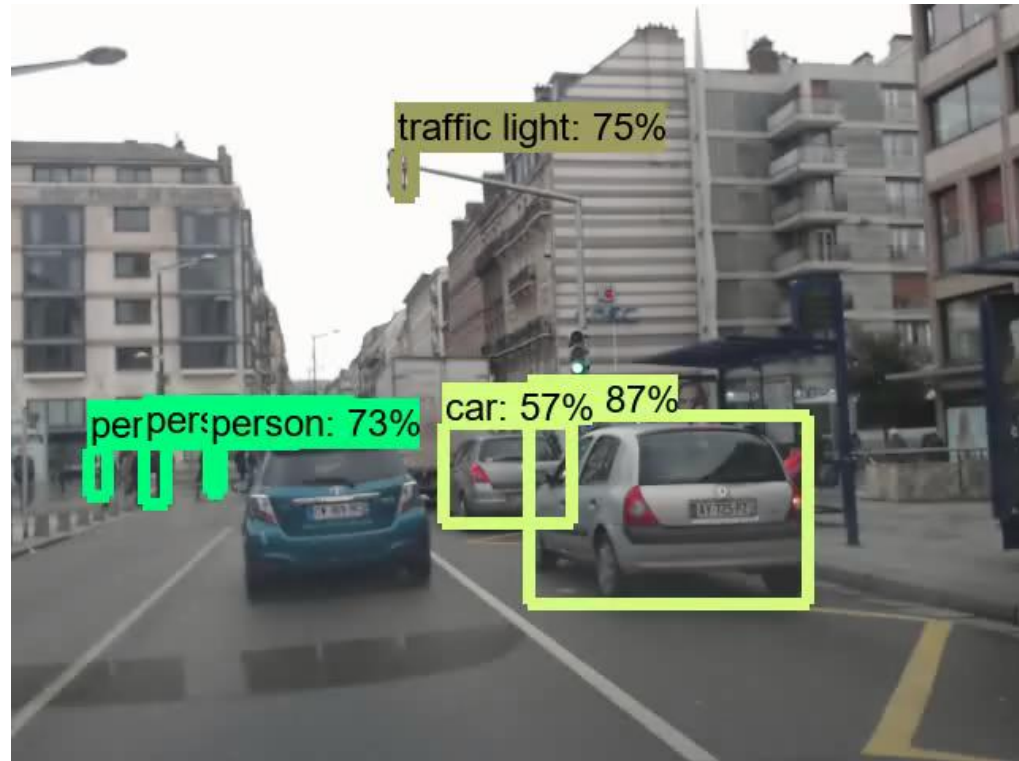


## Contexte & Objectifs

### ■ Deep Learning basé SSD sous TensorFlow



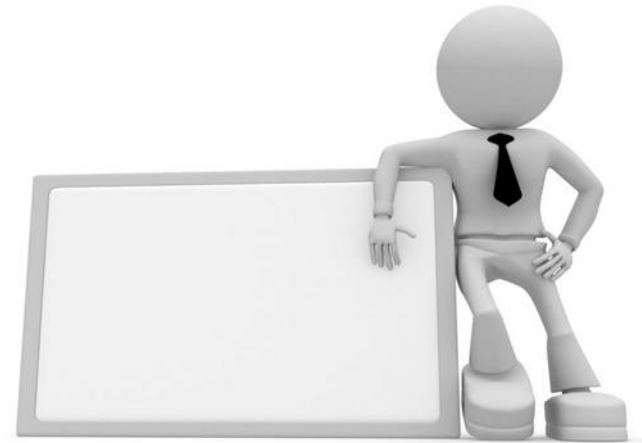
Parking CISE - ESIGELEC



Centre Ville - Rouen



- Contexte & Objectifs
- Perception d'Environnement pour Smart Mobilité
- **Projet Recherche : M2NUM Phase II**



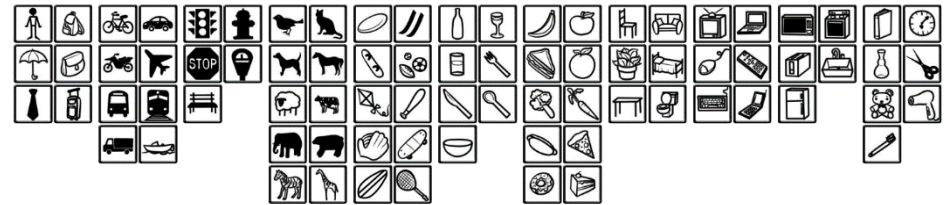
## Plateforme de Détection d'Objets :

### ■ Base de données Pascal VOC :

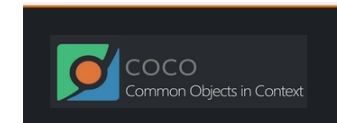


Université de Leeds, ETHZ, Zurich, Université de  
Edinburgh, Université d'Oxford

- 20 Classes (11 530 images)
- Dont environ 2000 images de piétons (pour apprentissage)
- Variété moyenne



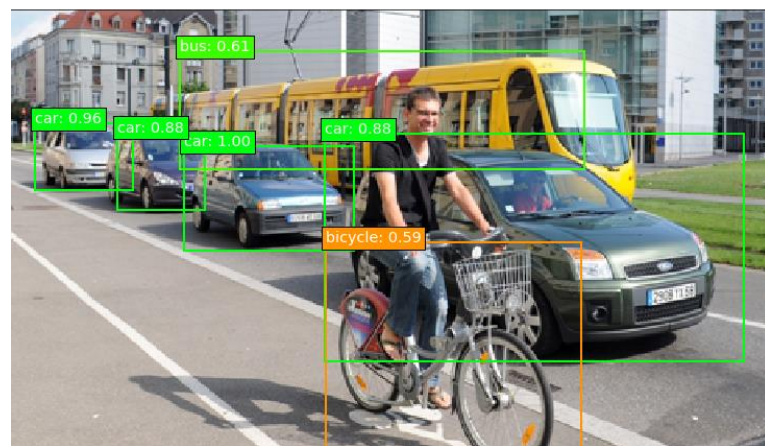
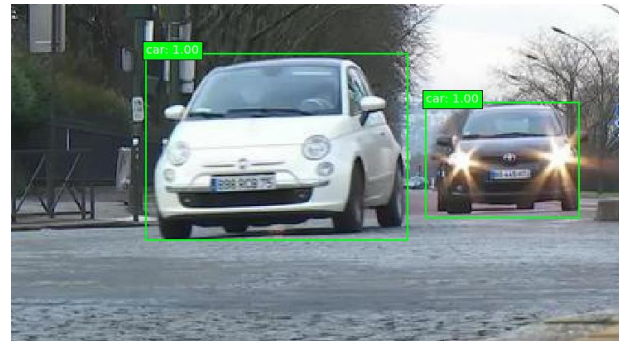
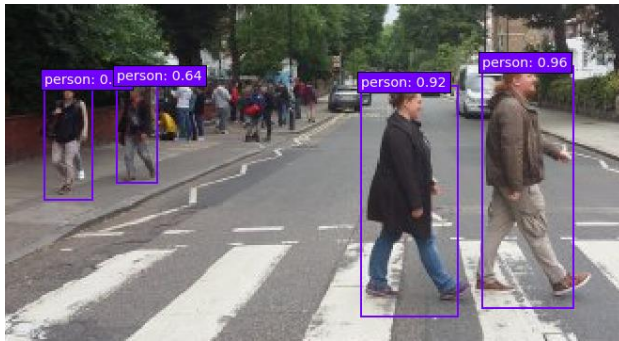
### ■ Base de données COCO (Google) :



- 80 Classes (330 000 images)
- 123.287 images, 886.284 instances (dans le dataset train/val 2017)
- 64115 images de piétons dans le train\_2017 (pour apprentissage)
- Variété très élevée :
  - Taille, comportement, âge, disposition, etc.

## Détection d'Objets :

### ■ Approche SSD : Résultats d'inférence





## Estimation de Distance basée Deep Learning

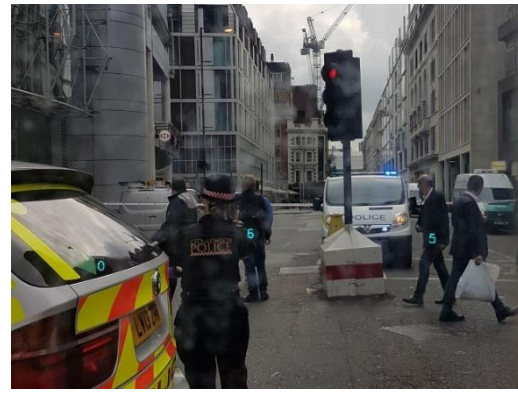
### ■ Monodepth : Résultats (1.3s par image)

### ■ Détection via SSD puis estimation de distance par Monodepth

Bonne estimation

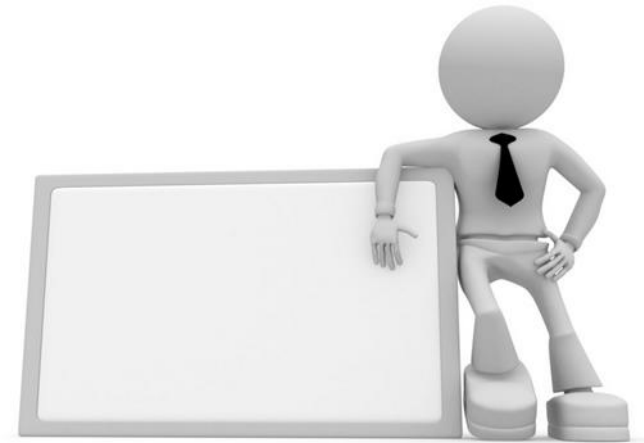


Mauvaise estimation





- Contexte & Objectifs
- Perception d'Environnement pour Smart Mobilité
- Projets Recherche : SEGULA Technologies





## Détection et Tracking d'Objets par Deep Learning :

### ■ Dataset :

	KITTI
Auteurs	Karlsruhe Institute of Technology
Domaine	Routier
Nb classes	8
Nb d'images	15000



## Détection et Tracking d'Objets par Deep Learning :

### ■ Ressources Matériels :

#### ▶ Ordinateur portable personnel:

- Gpu: GTX 1050 4Gb VRAM
- Cpu: i5 8300h
- Ram: 16Gb

#### ▶ LNA Serveur:

- Gpu: 2x GTX 1080Ti 11Gb VRAM
- Utile pour les inférences et entraînements léger

#### ▶ MYRIA (CRIANN):

- Gpu: K80 et P100 11Gb VRAM
- Utile pour entraînements
- Pas de mode débuggage, accès en mode Batch

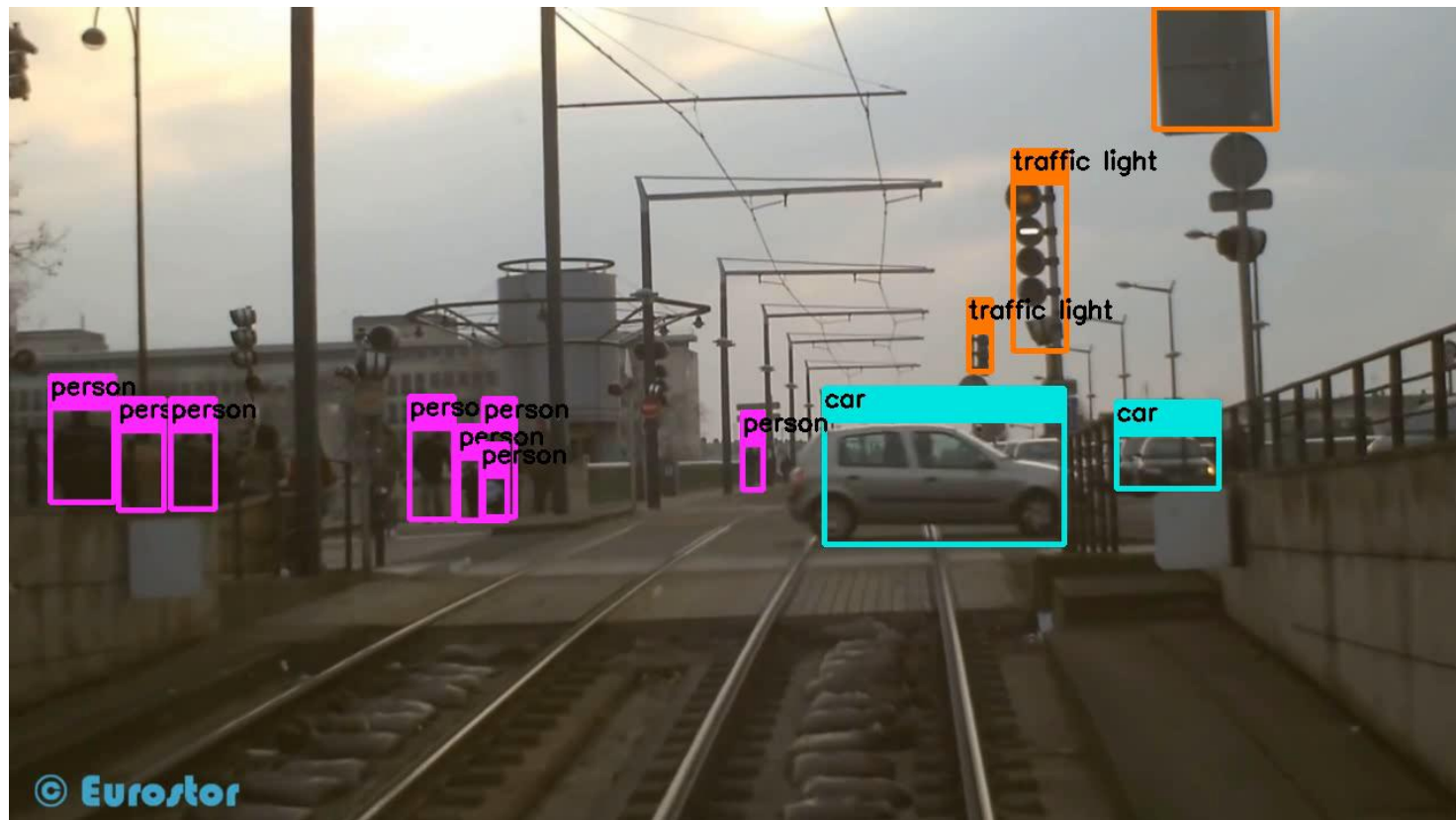
#### ▶ Caméra RGB-D RealSense D435

- 1 Emetteur infrarouge
- 2 Caméras infrarouge
- 1 Caméra RGB
- Peut être utilisée comme caméra stéréoscopique



## Détection d'Objets : YOLO3

### ■ Résultats & Comparaison : SSD vs YOLO V3





## Détection et Tracking d'Objets par Deep Learning :

### ■ YOLO V3 :



### ■ Comparaison des différentes implémentation de SSD :

Implémentation	FPS
Keras	6fps
Tensorflow	11fps
Darknet	13fps
PyTorch	16-18fps

## Détection et Tracking d'Objets par Deep Learning :

### ■ Comparaison des Méthodes :



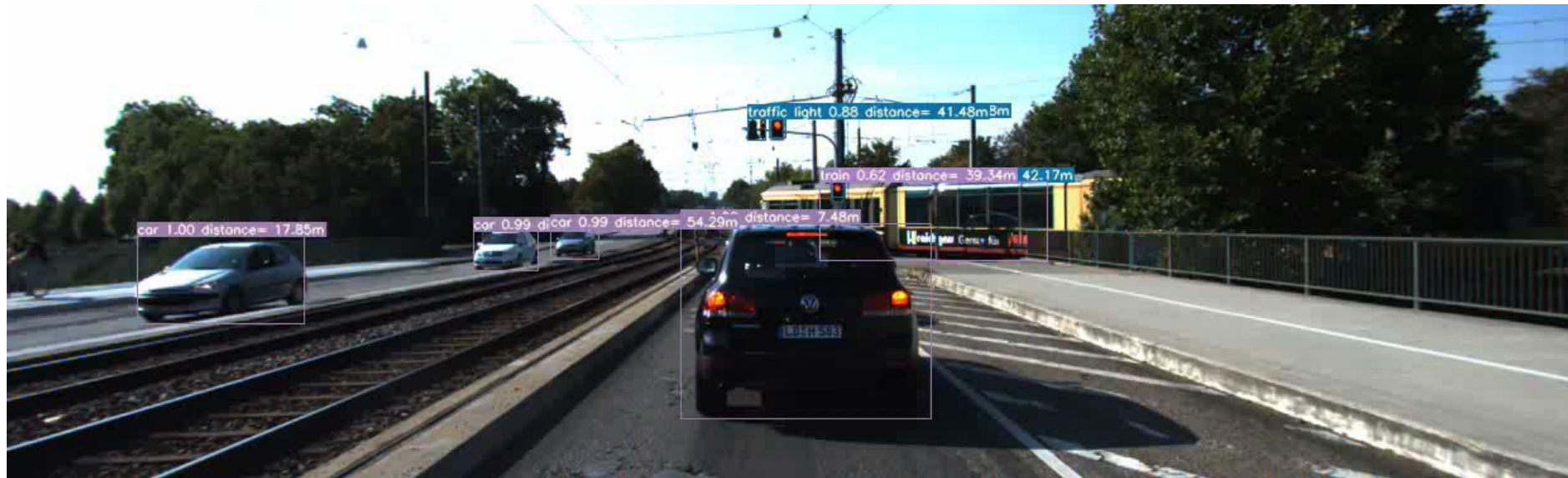
## Détection et Tracking d'Objets par Deep Learning :

### ■ Evaluation des Méthodes d'Estimation de Distances :



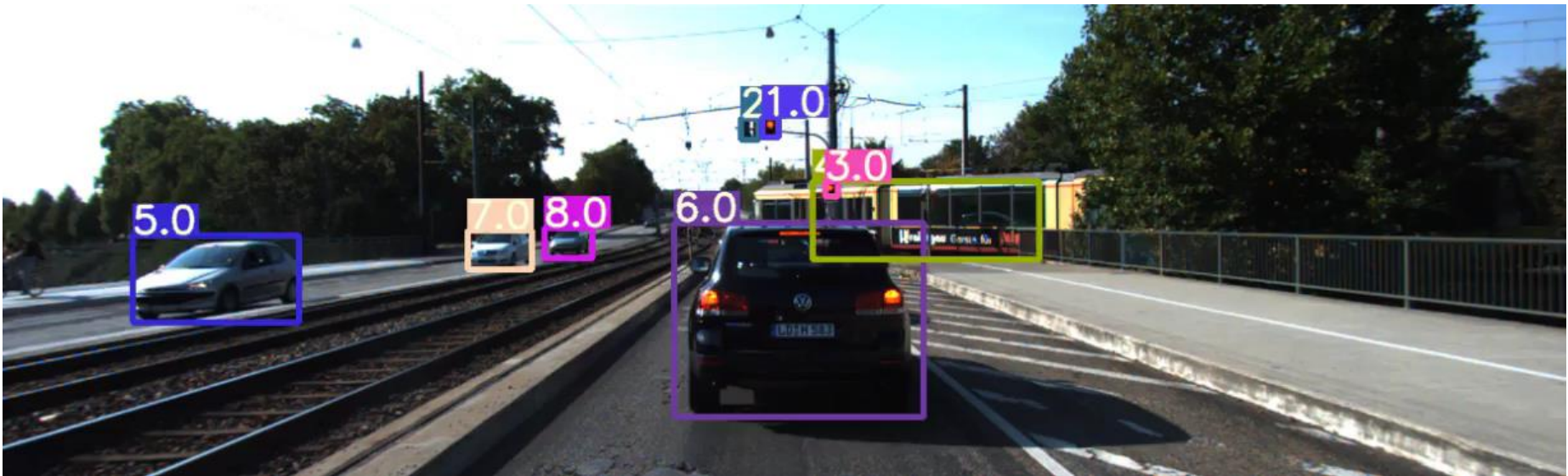
## Détection et Tracking d'Objets par Deep Learning :

- Evaluation de la Distance Réelle des Objets sur une Séquence KITTI :



## Détection et Tracking d'Objets par Deep Learning :

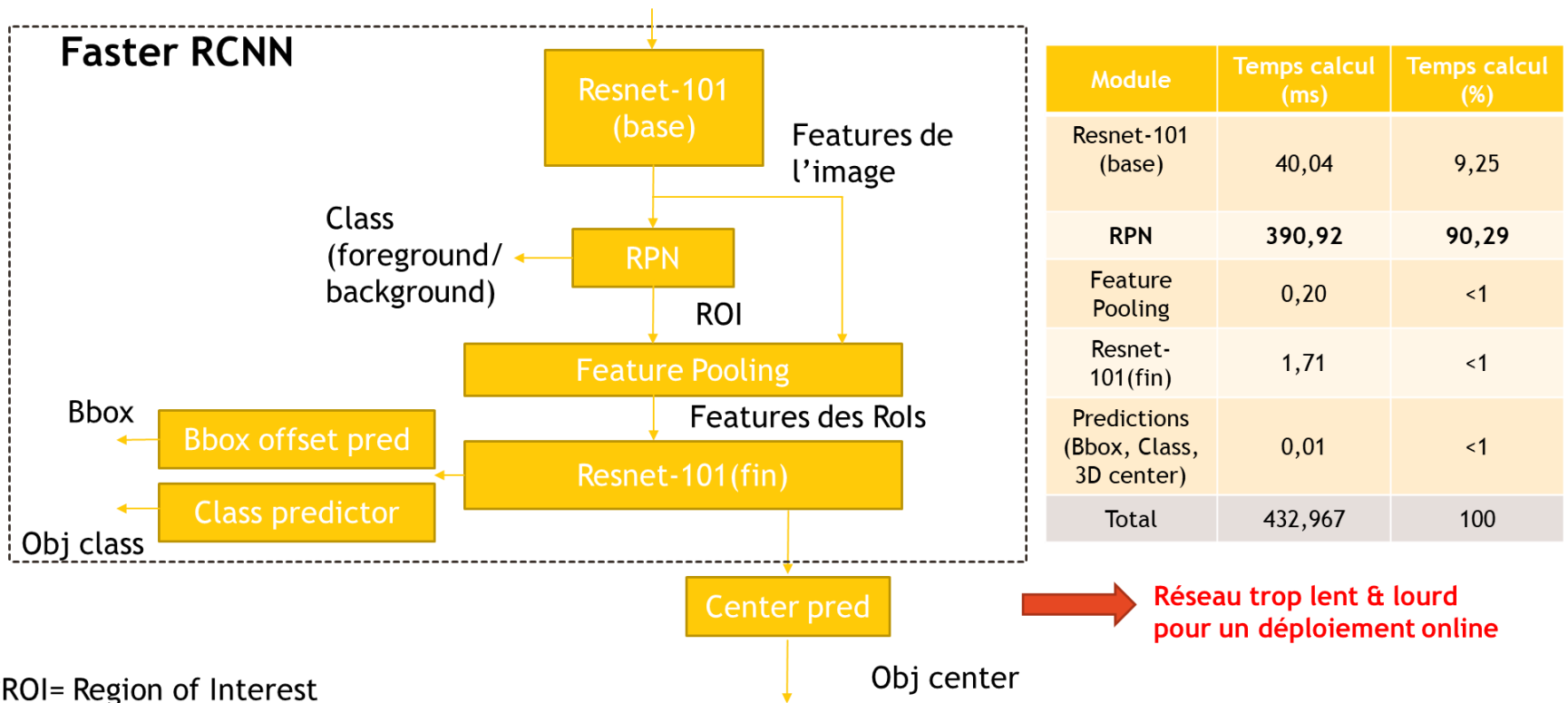
### ■ Tracking :





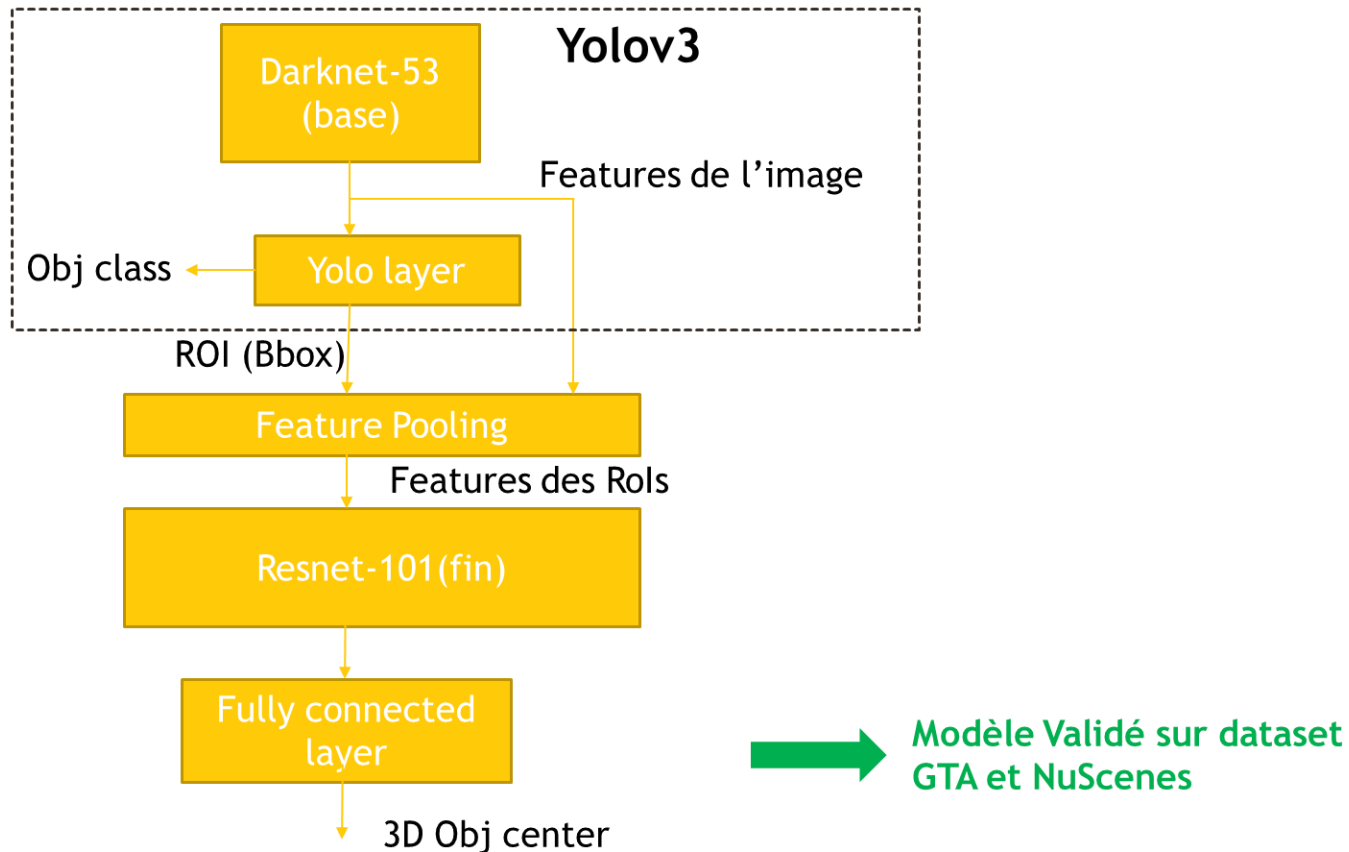
## Détection d'Objets par Deep Learning :

### ■ Détection du Centre 3D :



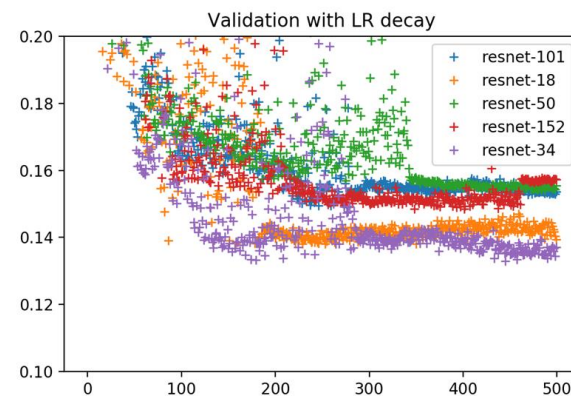
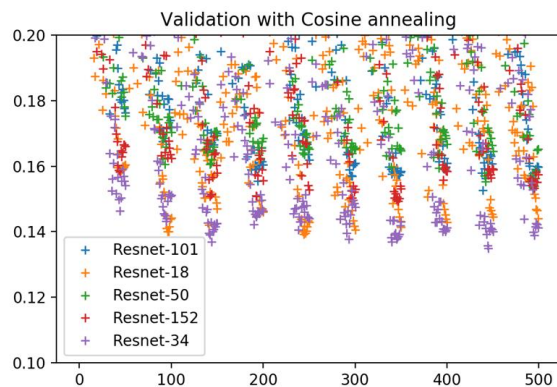
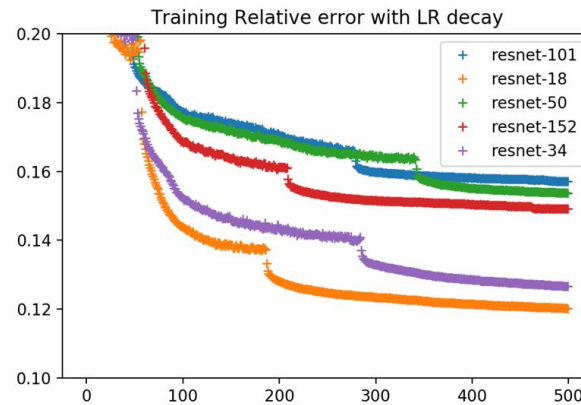
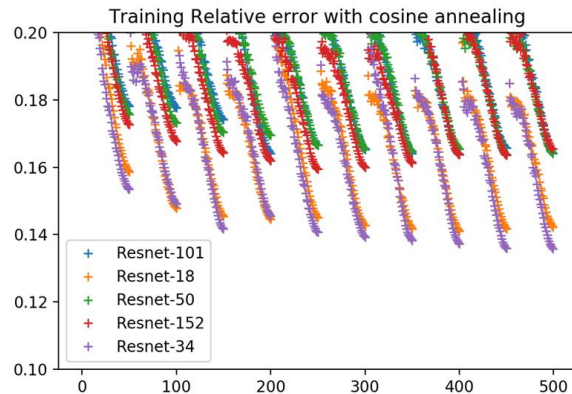
## Détection d'Objets par Deep Learning :

### ■ Détection du Centre 3D : Architecture proposée



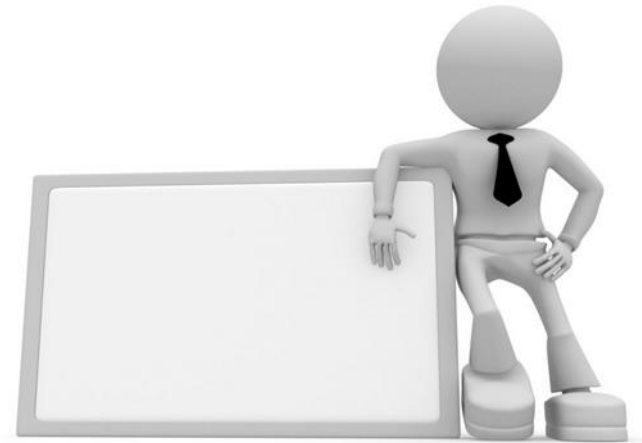
## Détection d'Objets par Deep Learning :

### ■ Détection du Centre 3D : Architecture proposée





- Contexte & Objectifs
- Perception d'Environnement pour Smart Mobilité
- Projets Recherche : ADAPT



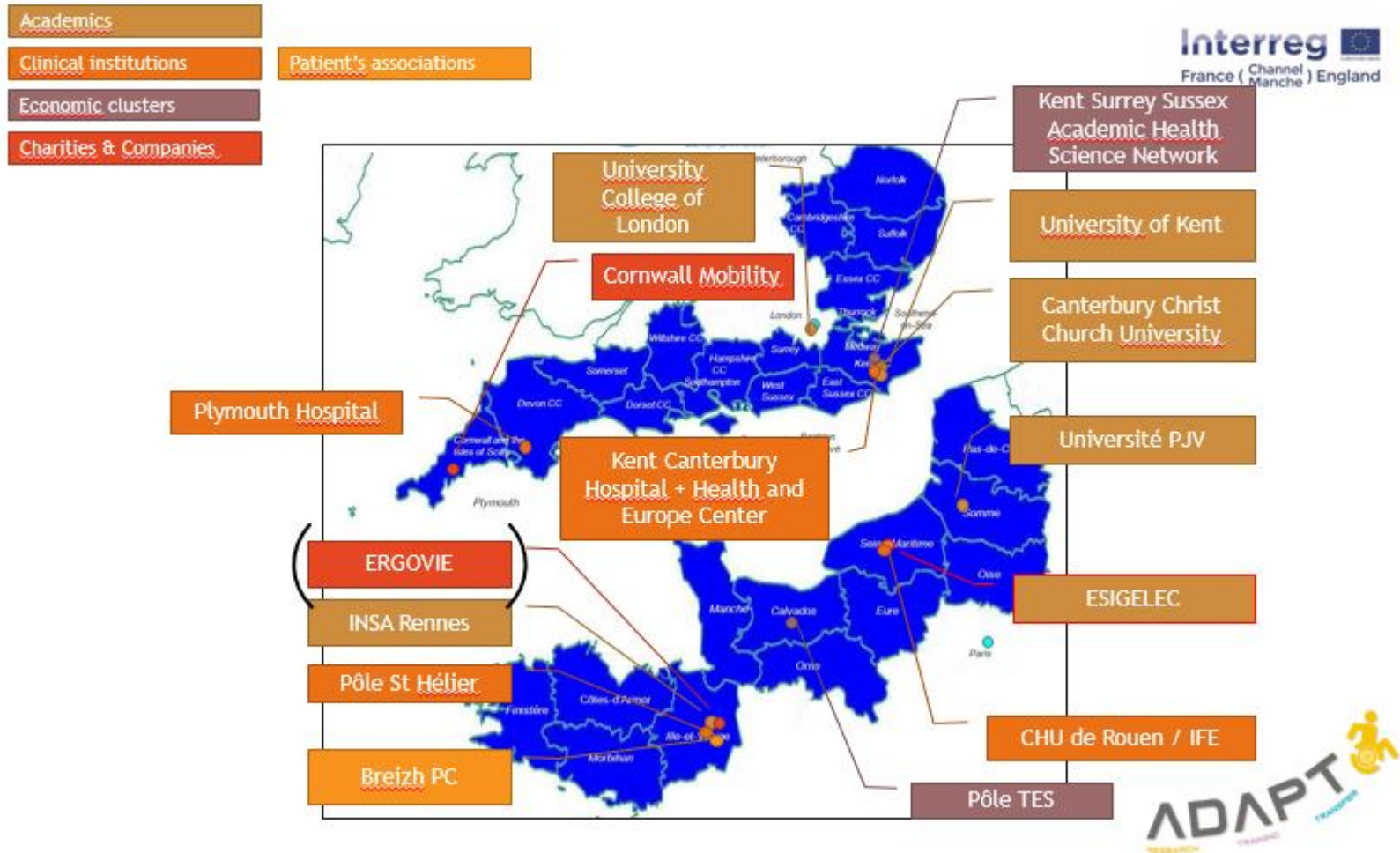
## Détection et Suivi d'Objets par Deep Learning pour Fauteuil Roulant Intelligent : **INTERREG VA FMA (2017-2022)**

### ■ Développement d'un Smart Wheelchair :

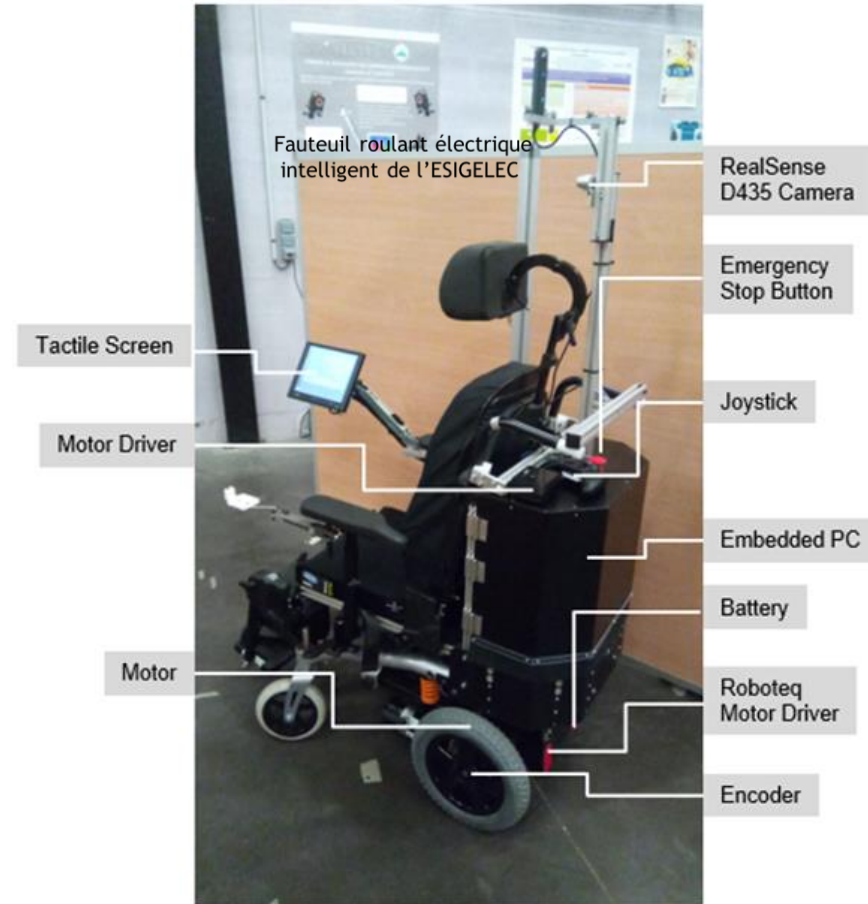
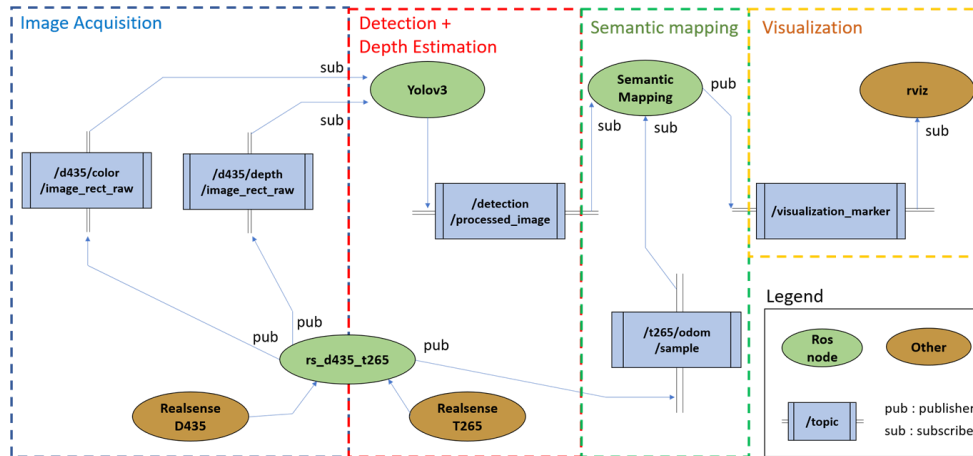




## ADAPT : 16 Partenaires : INTERREG VA FMA



## ■ Développement d'un Smart Wheelchair :





## Développement d'un Smart Wheelchair :

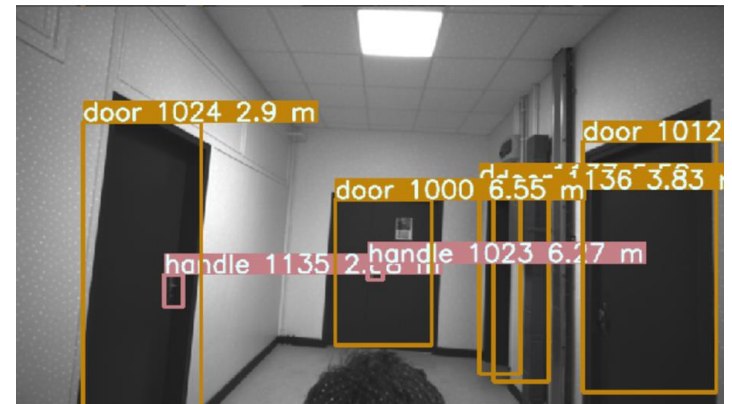
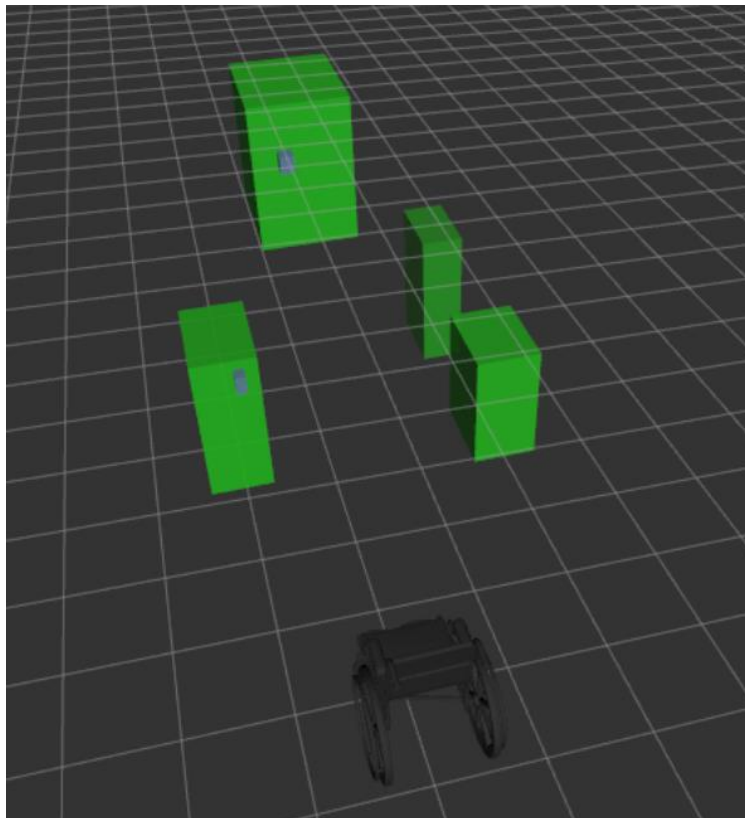






## Détection et Suivi d'Objets par Deep Learning pour un Fauteuil Roulant Intelligent :

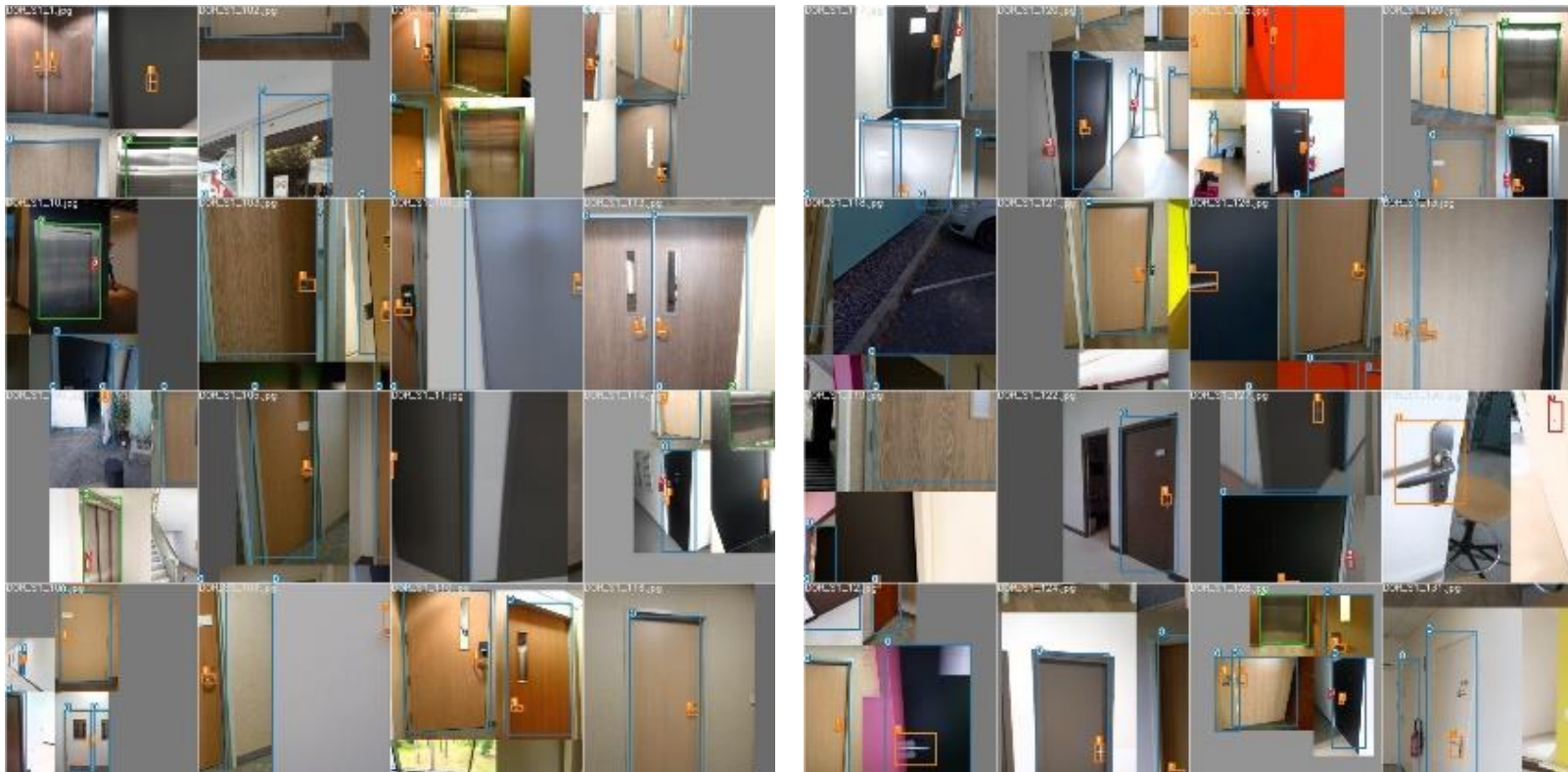
### ■ Détection de Portes et Poignées : Carte sémantique





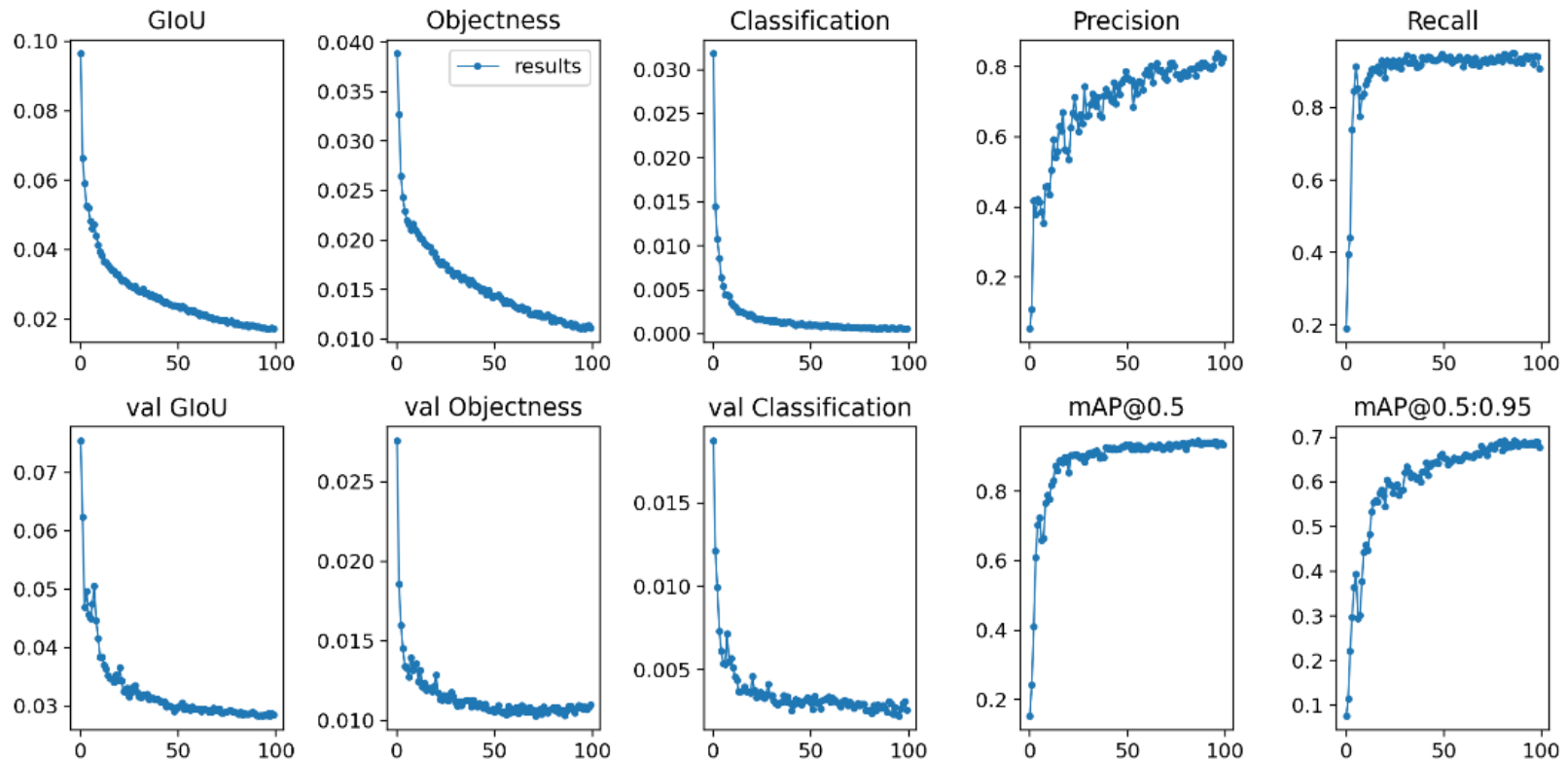
# Détection et Suivi d'Objets par Deep Learning pour un Fauteuil Roulant Intelligent :

## ■ Détection de Portes et Poignées : Carte sémantique



# Détection et Suivi d'Objets par Deep Learning pour un Fauteuil Roulant Intelligent :

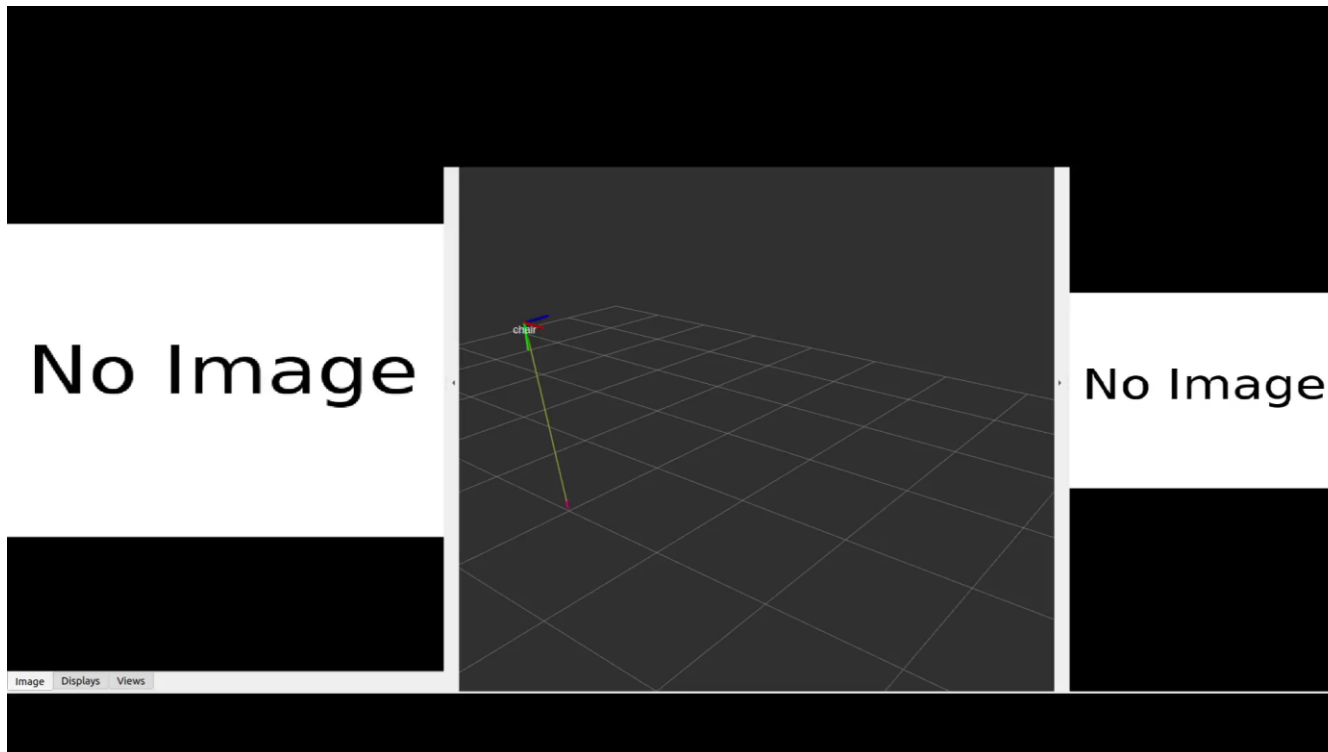
## ■ Détection de Portes et Poignées : Carte sémantique





## Détection et Suivi d'Objets par Deep Learning pour un Fauteuil Roulant Intelligent :

- Détection de Portes et Poignées : Carte sémantique





**ADAPT**  
RESEARCH TRAINING TRANSFER

**Interreg**   
France ( Channel ) England  
European Regional Development Fund

**ESIGELEC**   
**irseem**

**INSA** INSTITUT NATIONAL DES SCIENCES APPLIQUÉES RENNES

UNIVERSITÉ de Poitiers  
*Jules Verne*

The Health and Europe Centre

**UCL**

University of Kent

**CHU**  
ROUEN NORMANDIE

Canterbury Christ Church University

Kent Surrey Sussex Academic Health Science Network

**CM**  
Cornwall Mobility

**pôle TES**  
e-secure transactions

**PÔLE DE MÉDECINE PHYSIQUE ET RÉADAPTION**  
SAINT HÉLÈNE

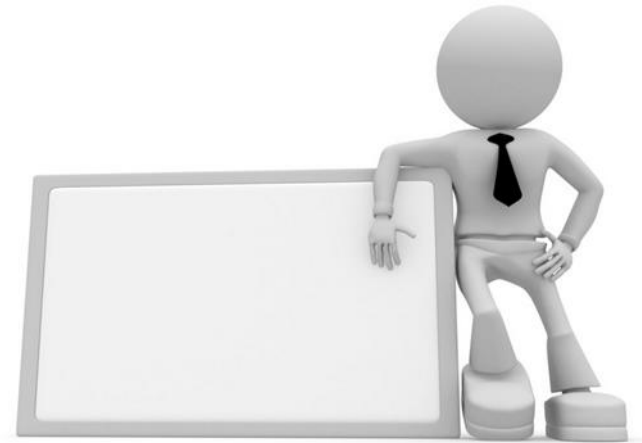
**RESEAU ITC**  
BREIZH

Plymouth Hospitals **NHS**  
NHS Trust

East Kent Hospitals University **NHS**  
NHS Foundation Trust



- Contexte & Objectifs
- Perception d'Environnement pour Smart Mobilité
- Projets Recherche : M2SiNUM



## Projet M2SiNUM (RIN) :

### ■ Projet M2SiNUM : (2018-2021)

■ Modélisation Mathématique avancée et Simulations Numériques pour l'Innovation dans l'Environnement et la Santé

■ Comprend 4 Axes :

■ Innovations pour les Energies

■ Smart City

■ Modélisation Mathématique pour Matériaux Innovants

■ Imagerie



**Plateforme (TRL5)**

Aide à la Conduite et Navigation Autonome



## Projet M2SiNUM (RIN) :

### ■ Partenaires :

LMI, Rouen

LMRS, Rouen

LITIS, Rouen

GREYC, Caen

LMAH, Le Havre

LMNO, Caen

IRSEEM, Rouen

LMN, Rouen

CORIA, Rouen

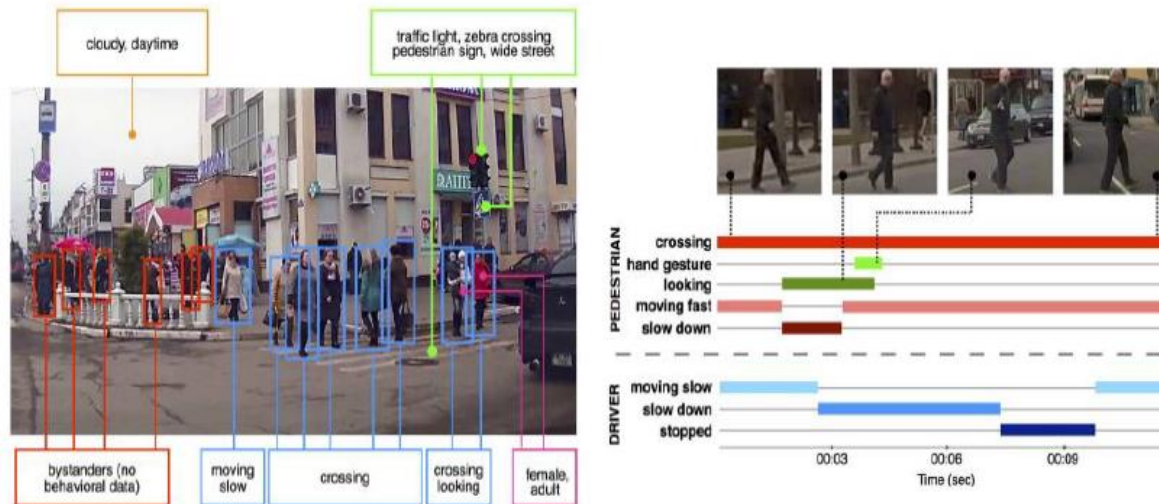
CEREMA, Rouen

CRIANN - Mésocentre de calcul de Normandie



# Analyse de Comportement et Compréhension de Scènes Routières :

## ■ Contexte et Objectif



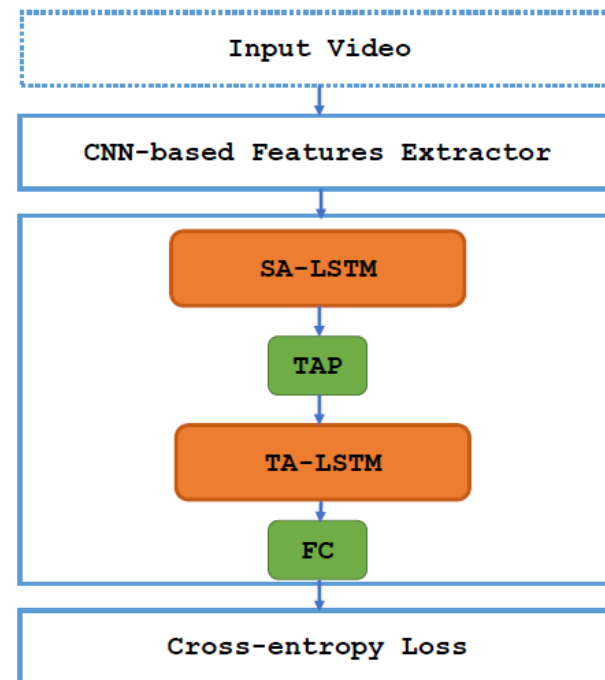
### Towards a comprehensive road scene analysis !

- Tackling the issue of road agents behaviour understanding
- Proposing a spatio-temporal deep network based on attention mechanism which leverages key visual features over time
- Reuse of the multi-head self-attention mechanism with Long-Short Term Memory (LSTM) network.

## Analyse de Comportement et Compréhension de Scènes Routières :

### ■ Approche Proposée :

- CNN features extraction from the input frames
- Backbone of our model is composed of four main layers :
  - SA-LSTM (Spatial Attention-LSTM) layer
  - TAP (Temporal Average Pooling) layer
  - TA-LSTM (Temporal Attention-LSTM) layer
  - FC (Fully Connected) layer
- Cross-entropy loss layer



## Analyse de Comportement et Compréhension de Scènes Routières :

### ■ Approche Proposée :

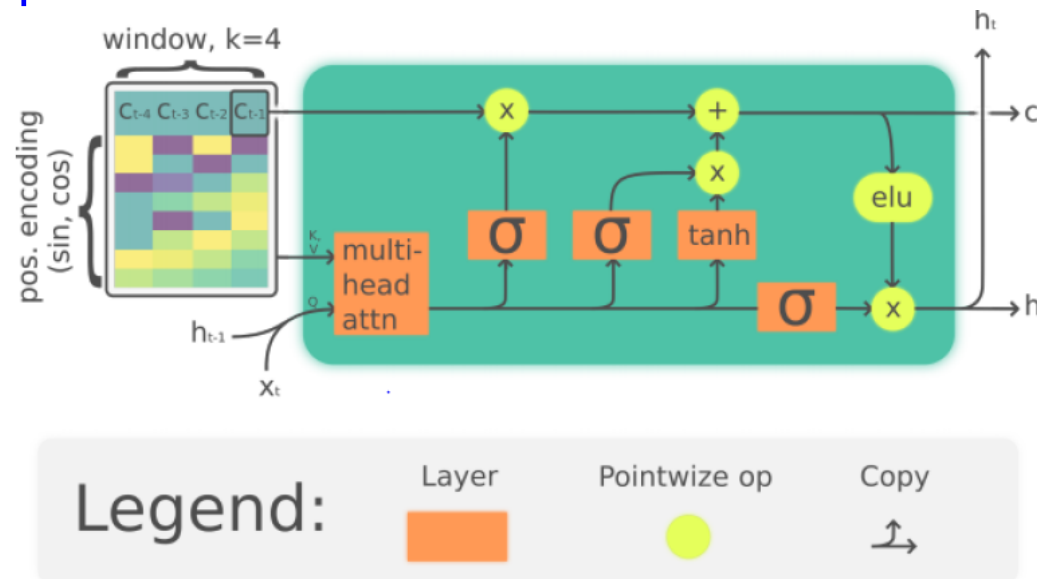


Figure – Multi-head attention mechanism on a windowed queue of its past cell states.  $K$ ,  $V$  and  $Q$  are respectively the attention mechanism's input Keys, Values and only one Query. The most recent states  $c_{[t-1...t-k]}$  are arranged in a "first in, first out" queue of length  $k$ . There is also the not-illustrated fact that the query is formulated from a layer of the concatenation of  $h_{t-1}$  and  $x_t$  on the innermost (feature) axis.



# Analyse de Comportement et Compréhension de Scènes Routières :

## ■ Résultats : Datasets

- ① Joint Attention in Autonomous Driving (JAAD) :
  - over than 300 image sequence ranging from 5 to 15 seconds in duration
  - Collected in North America and Europe
  - Goal : determine pedestrian's behaviours in front of the vehicle specifically annotation provided information with regards to their gaits (mainly walking, standing, looking towards the traffic and not looking) and their corresponding attention level like looking, standing at the curb and checking the traffic.
- ② Honda research institute Driving Dataset (HDD) :
  - around 10k of image sequences acquired under different weather conditions
  - two out of four annotation layers are used : Cause and Goal-oriented action which proposed 5 and 11 classes for driver' behaviours, respectively.



## Analyse de Comportement et Compréhension de Scènes

### Routières :

#### ■ Résultats :

	sign	congestion	traffic light	pedestrian	parked car	mAP%
[Ramanishka CVPR 2018]	46.83	39.72	45.31	2.15	7.24	28.25
A-LSTM +MobileNet (ours)	71.97	74.21	<b>63.00</b>	47.54	45.81	60.50
A-LSTM +Inception (ours)	<b>78.44</b>	<b>82.05</b>	61.84	<b>49.50</b>	<b>63.79</b>	<b>67.12</b>

**Table** – Results for Cause layer of HDD dataset. Behaviors from this layer are immediate causes for either stop or deviate actions.

	Without CI	With CI
[Rasouli ICCV 2018]	0.713	0.815
LSTM-baseline	0.514	0.560
A-LSTM + MobileNet (ours)	0.829	0.841
A-LTSM + Inception (ours)	<b>0.850</b>	<b>0.883</b>

**Table** – Performance evaluation over the JAAD dataset and comparison between visual information and its combination with Contextual Information (CI).



- Contexte & Objectifs
- Perception d'Environnement pour Smart Mobilité
- Projets de Recherche
- Conclusion & Perspectives

