Deep-Learning for Autonomous Driving

Pr. Fabien MOUTARDE
Director of the Center for Robotics
Mines Paris, PSL

Fabien.Moutarde@minesparis.psl.eu
http://people.minesparis.psl.eu/fabien.moutarde

Outline

- **Challenges of Autonomous Driving (AD)**
- Deep-Learning for PERCEPTION in AD
- Deep-Learning for PREDICTION of road users
- Deep-Learning for AD short-term PLANNING

What is challenging in Autonomous Driving?

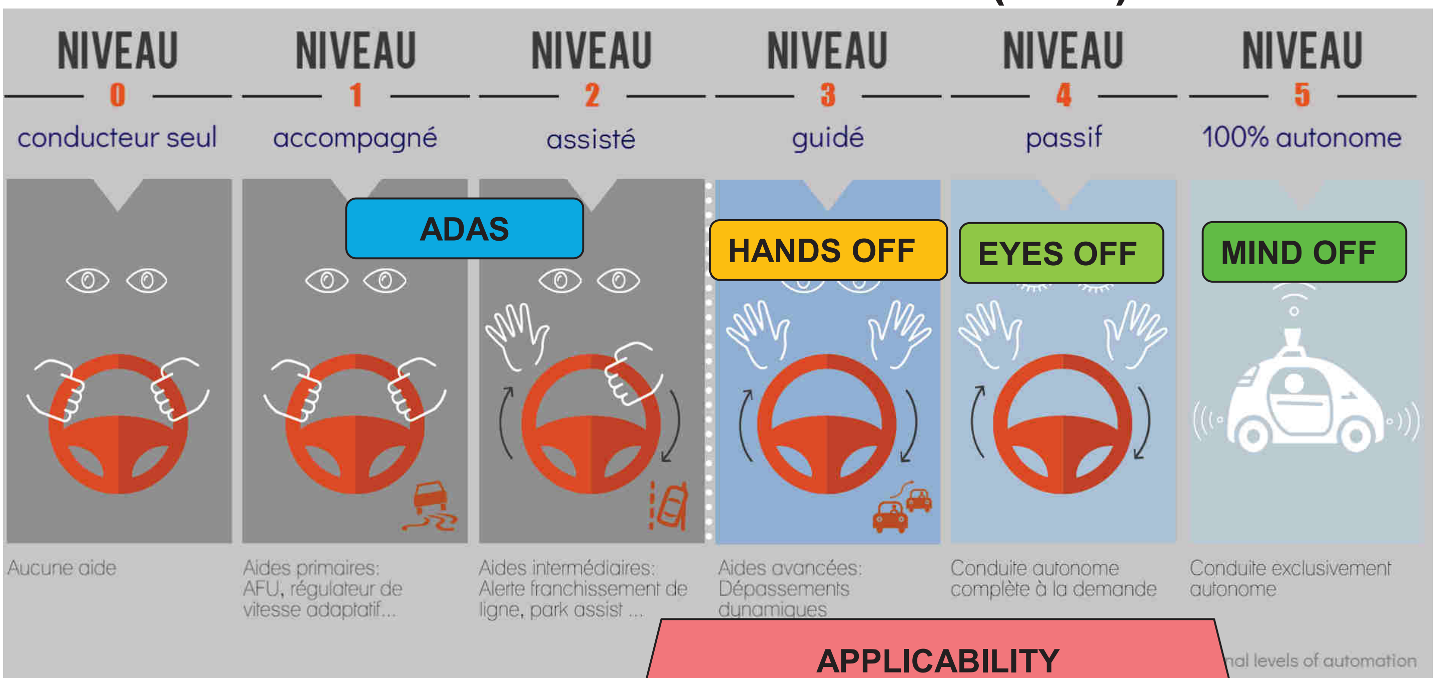
EASY on simple road with good lane markings and no other road users…

Automated Driving experiment (on closed track) by the Center for Robotics of Mines_Paris in 2002

…but much more difficult on open roads, especially urban areas

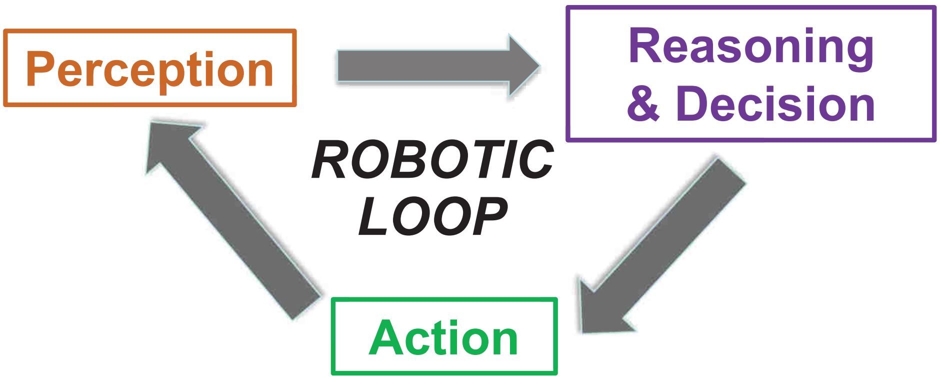
Autonomous?? Or rather *Automated* Driving?

The 5 levels of automation for vehicles (SAE)



APPLICABILITY CAN BE RESTRICTED TO SPECIFIC CONDITIONS (eg HIGHWAYS, …) = "Operational Design Domain" (ODD)

An Automated Vehicle is a mobile robot!



What types of Artificial Intelligences needed for Automated Vehicles (AV) ?

- "Semantic" interpretation of the vehicle's environment:
  - Detect and categorize/recognize objects (cars, pedestrians, bicycles, traffic signs, traffic lights, …)
  - Ego-localization
  - *Predict movements of other road users*
  - *Infer intentions of other drivers and pedestrians (or policeman!) from their movements/gestures/gazes*
- Planning of trajectories (including speed)
  In a dynamic and uncertain environment
- Coordinating/communicating/negociating with other road users
- Cooperative planning of multiple AVs
- For partial automated driving (level 3-4):
  - Analyze and understand attention and activities or gestures of the "driver-supervisor"

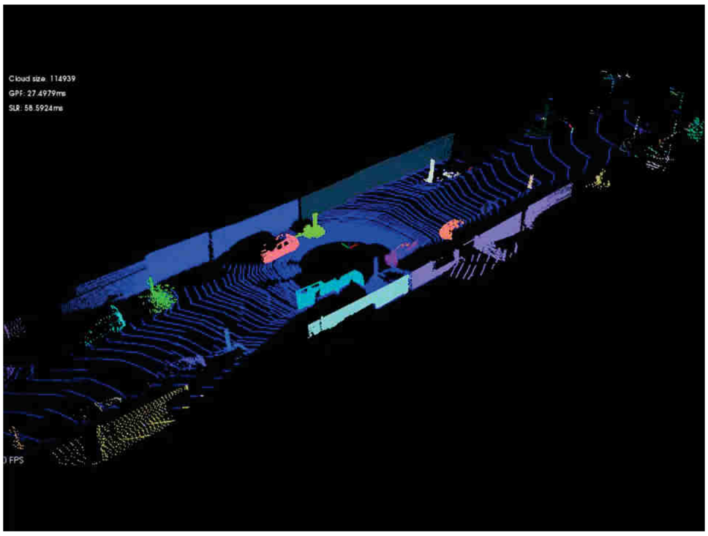
Outline

- Challenges of Autonomous Driving (AD)
- **Deep-Learning for PERCEPTION in AD**
- Deep-Learning for PREDICTION of road users
- Deep-Learning for AD short-term PLANNING

Intelligent Perception for Automated Vehicles

Essential function: SEMANTIC scene understanding
Now mostly done using Deep-Learning

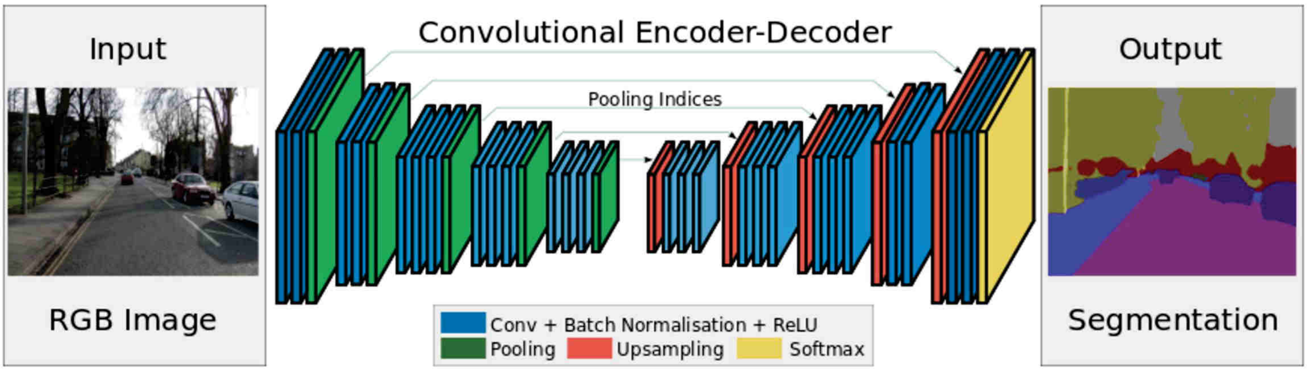
From camera
(Visual object detection
by Faster_RCNN)

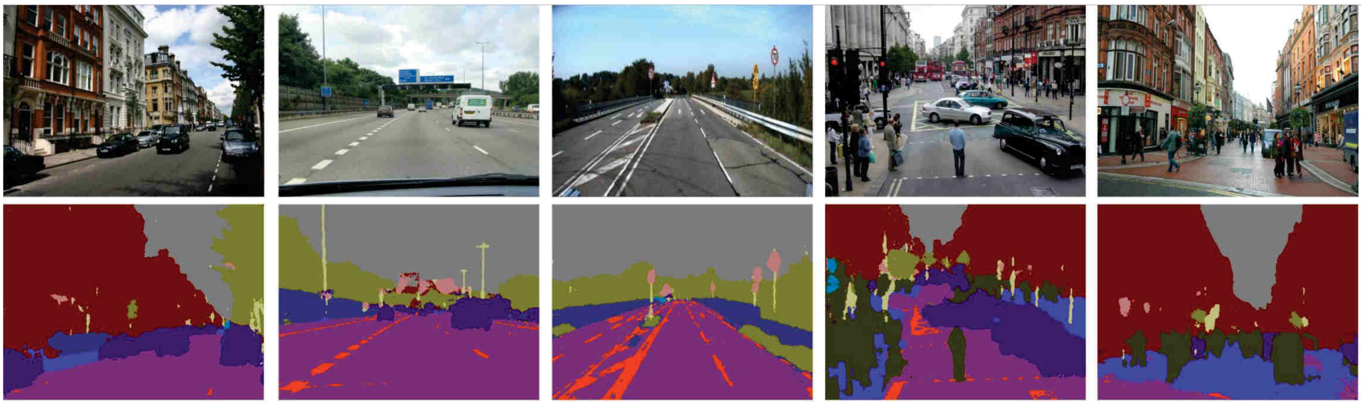


From LIDAR

Strong real-time constraint: process ≥ 15 frames/second

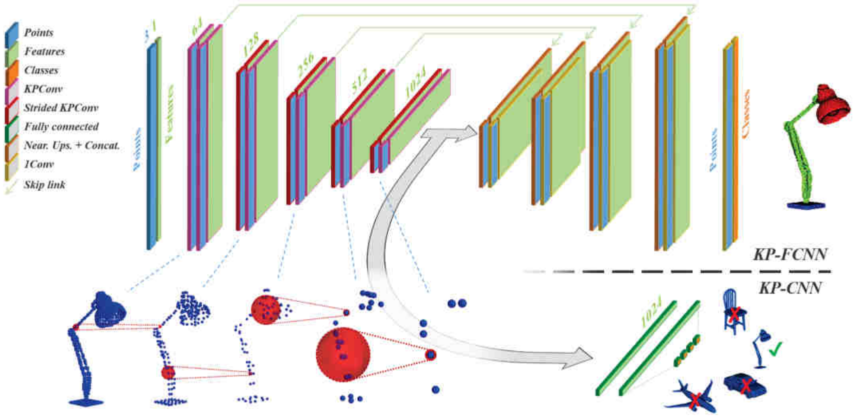
Convolutional Encoder-Decoder for SEMANTIC segmentation



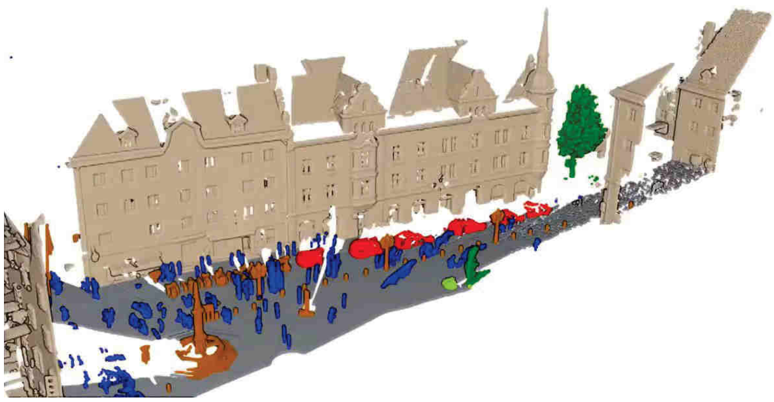


"SegNet: A Deep Convolutional Encoder-Decoder Architecture for ImageSegmentation", Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla [Cambridge (UK), 2015]

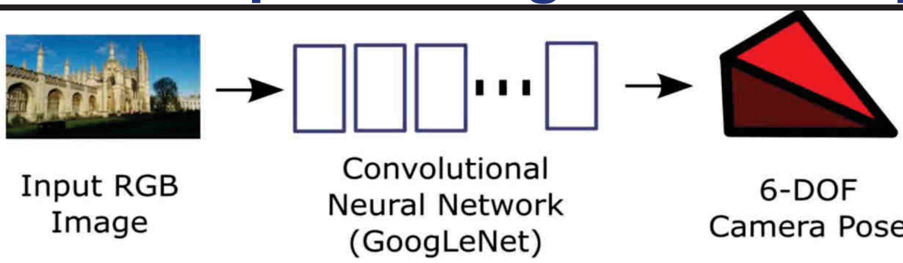
Deep-Learning for semantization of 3D points clouds



KPConv: Flexible and Deformable Convolution for Point Clouds, H. Thomas, C.R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, L.J. Guibas, IEEE International Conference on Computer Vision (ICCV), Oct 2019.



Visual ego-localization by Deep-Learning: PoseNet approach



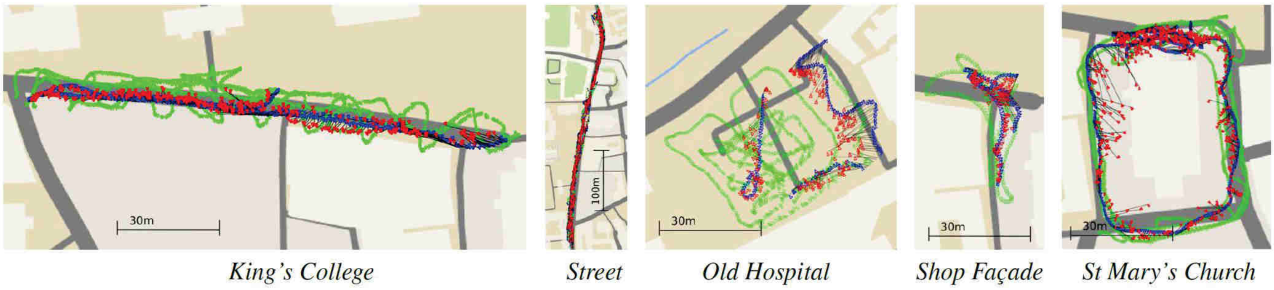


Figure 4: **Map of dataset** showing training frames (green), testing frames (blue) and their predicted camera pose (red). The testing sequences are distinct trajectories from the training sequences and each scene covers a very large spatial extent.

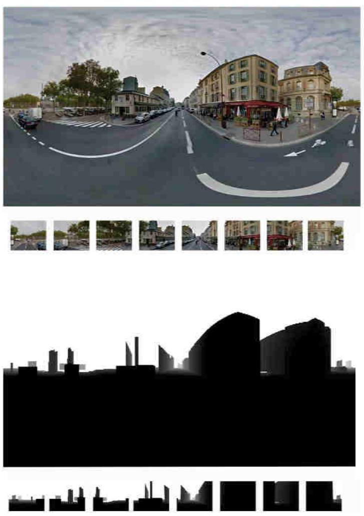
Localization by deep-learning is ~2 times less precise than by classic computer-vision, but is ~20 times faster at inference, and much more robust to blur, occlusions and appearance change

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

Visual ego-localization based on pre-existing geo-referenced images

Outdoor ego-localization using Google-StreetView images (BoW+RANSAC // Deep-Learning)

Synthetized views from Google StreetView panoramas



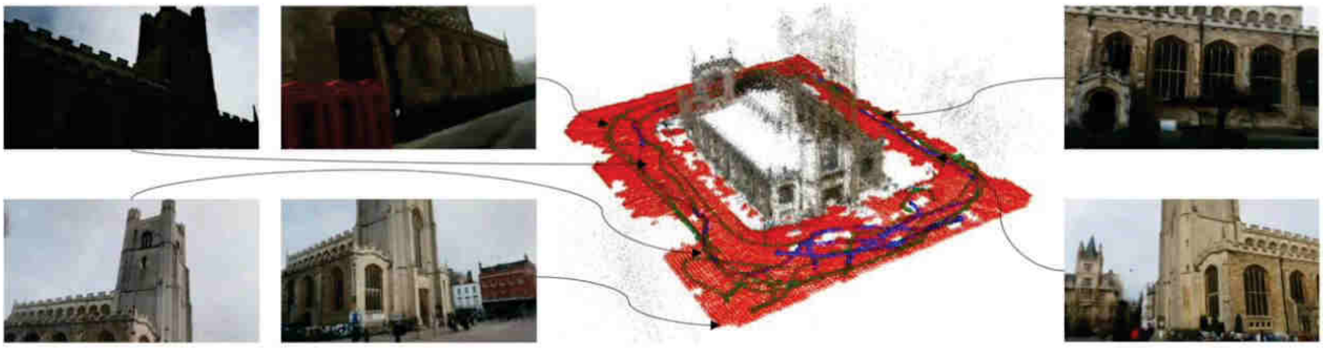
Query image matching

Urban Localization with Street Views using a Convolutional Neural Network for End-to-End Camera Pose Regression, Guillaume Bresson, Yu Li, Cyril Joly, Fabien Moutarde. IEEE Intelligent Vehicles Symposium (IV '2019), June 2019.

Visual ego-localization with DL using data augmentation with NeRF

LENS : Localization enhanced by NeRF synthesis

CoRL 2021



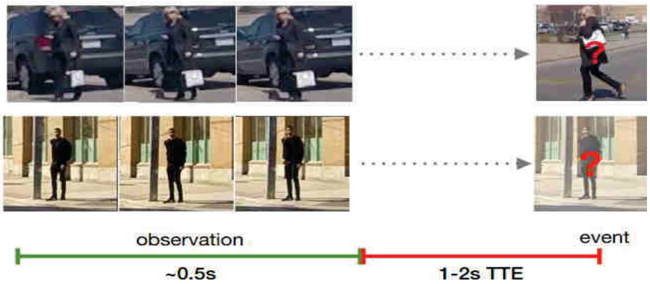
LENS: Localization enhanced by NeRF synthesis, Arthur Moreau, Nathan Piasco, Dzmitry Tsishkou, Bogdan Stanciulescu, Arnaud de La Fortelle, 5th Annual Conference on Robot Learning (CoRL'2021), London (United Kingdom), Nov.2021

Outline

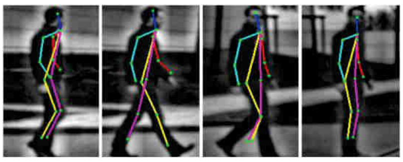
- Challenges of Autonomous Driving (AD)
- Deep-Learning for PERCEPTION in AD
- **Deep-Learning for PREDICTION of road users**
- Deep-Learning for AD short-term PLANNING

Deep-Learning for prediction of pedestrian INTENTIONS

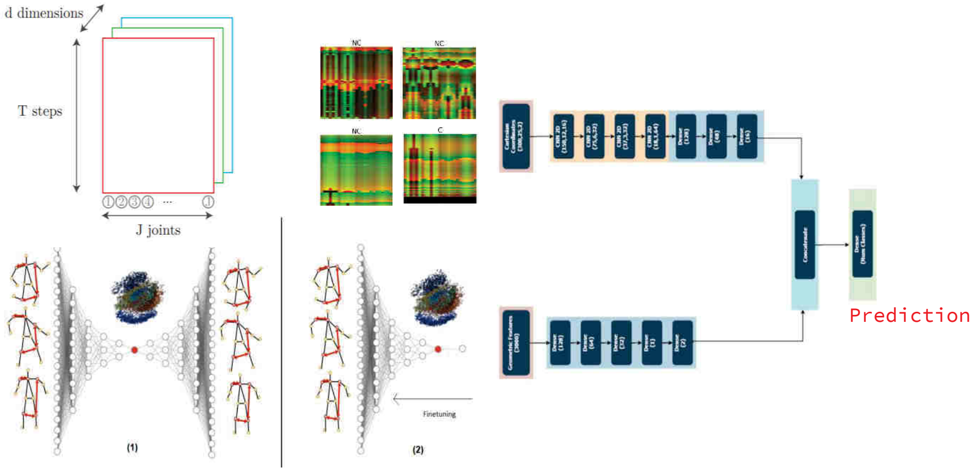
Goal



Our approach: *Skeletal* Pedestrian Intention NETwork



PhD thesis of Joseph GESNOUIN

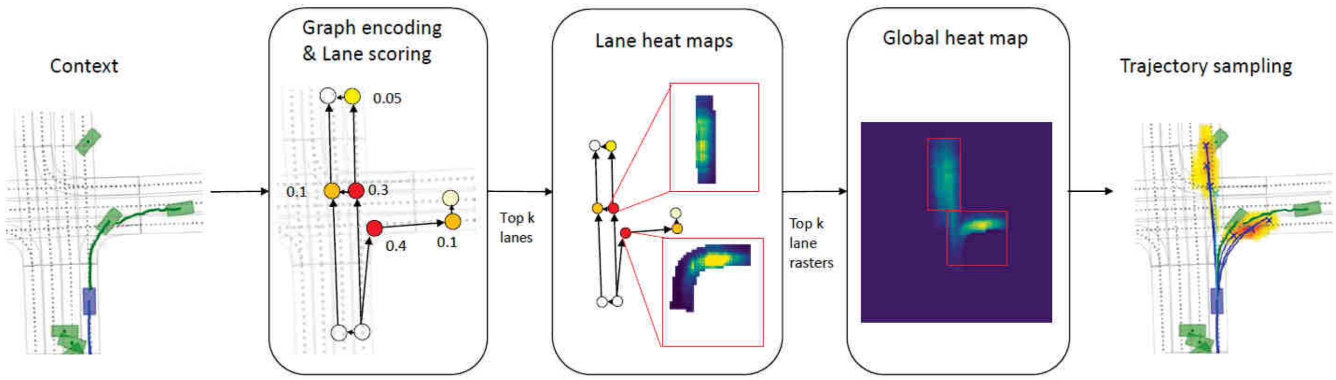


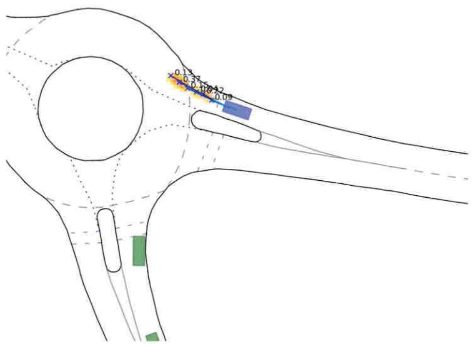
Results of our Skeletal Pedestrian Intention network

Model Name	Model Variants	Model Params (Additional Costs)	PIE ACC	PIE AUC	PIE F1
Static	VGG16 [43]	14.7M	0.71	0.60	0.41
	Resnet50 [15]	23.6M	0.70	0.59	0.38
ATGC [34]	AlexNet	58.3M	0.59	0.55	0.39
ConvLSTM [40]	VGG16	0.001M (VGG)	0.58	0.55	0.39
	ResNet50	0.001M (Resnet)	0.54	0.46	0.26
SPI-Net [12]	CNN MLP	0.1M (OpenPose)	0.66	0.54	0.30
SingleRNN [18]	LSTM	1.4M (2*VGG,OpenPose)	0.83	0.77	0.67
	GRU	1.0M (2*VGG,OpenPose)	0.81	0.75	0.64
MultiRNN [2]	GRU	1.8M (2*VGG,OpenPose)	0.83	0.80	0.71
StackedRNN [53]	GRU	2.6M (2*VGG,OpenPose)	0.82	0.78	0.67
HierarchicalRNN [52]	GRU	3M (2*VGG,OpenPose)	0.82	0.77	0.67
SFRNN [35]	GRU	2.6M (2*VGG,OpenPose)	0.82	0.79	0.69
C3D [44]	RGB	78M	0.77	0.67	0.52
I3D [6]	RGB	12.3M	0.80	0.73	0.62
	Optical flow	12.3M (FlowNet2)	0.81	0.83	0.72
TwoStream [42]	VGG16	134.3M (FlowNet2)	0.64	0.54	0.32
PCPA [19]	Temp. +mod. attention	31.2M (C3D,OpenPose)	0.87	0.86	0.77
TrouSPI-Net (ours)	CBAM attention block	1.5M (OpenPose)	**0.88**	**0.88**	**0.80**
	SE attention block	1.5M (OpenPose)	**0.88**	0.87	**0.80**

***TrouSPI-Net: Spatio-temporal attention on parallel atrous convolutions and U-GRUs for skeletal pedestrian crossing prediction**, J. Gesnouin, S. Pechberti, B. Stanciulescu, and F. Moutarde, 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG'2021).*

Deep-Learning for prediction of VEHICLE trajectories





GOHOME: Graph-Oriented Heatmap Output for future Motion Estimation, *Thomas Gilles, Stefano Sabatini, Dzmitry Tsishkou, Bogdan Stanciulescu, Fabien Moutarde, IEEE International Conference on Robotics and Automation (ICRA'2022), Philadelphie (USA), May 2022*

Outline

- Challenges of Autonomous Driving (AD)
- Deep-Learning for PERCEPTION in AD
- Deep-Learning for PREDICTION of road users
- **Deep-Learning for AD short-term PLANNING**

Principle of "end-to-end" driving

Sensors (cameras, radar, LIDAR, etc…) → Perception → Trajectory Planning → Control → Driving wheel, Acceleration or braking

Current architecture for automated driving

vs. **HUMAN driving: turn/brake** ***<u>by just looking in front!</u>***

≈ "intelligent" visual servoing

Sensors (cameras, radar, LIDAR, etc…) → **Imitation Learning or Reinforcement Learning** → Driving wheel, Acceleration or braking

Center camera → CNN → Network computed steering command → Drive by wire interface

<u>Imitation</u> Learning for end-to-end driving

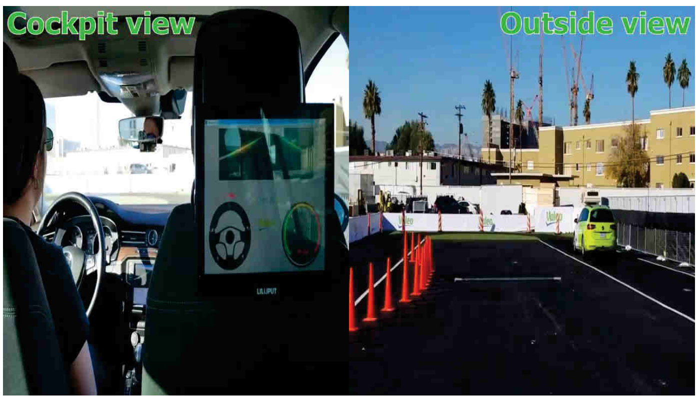
"Copying"human driver

Recorded steering wheel angle → Adjust for shift and rotation → Desired steering command

Fisheye camera → Random shift and rotation → CNN → Network computed steering command → –

Error → Back propagation weight adjustment → CNN

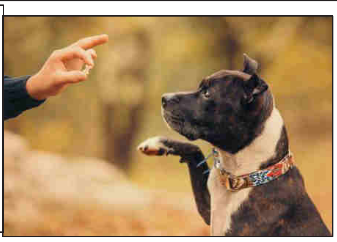
<u>"End to End Vehicle Lateral Control Using a Single Fisheye Camera"</u>*, *Marin Toromanoff, Emilie Wirbel, Frédéric Wilhelm, Camilo Vejarano, Xavier Perrotton et Fabien Moutarde, 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018), Madrid, Spain, 1-5 oct. 2018.

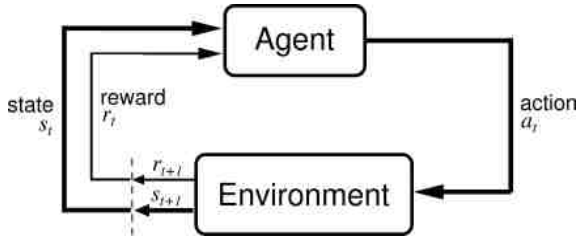


Valeo demo @CES'2018

Short-term trajectory planning with Deep Reinforcement Learning

Reinforcement Learning = training by REWARD



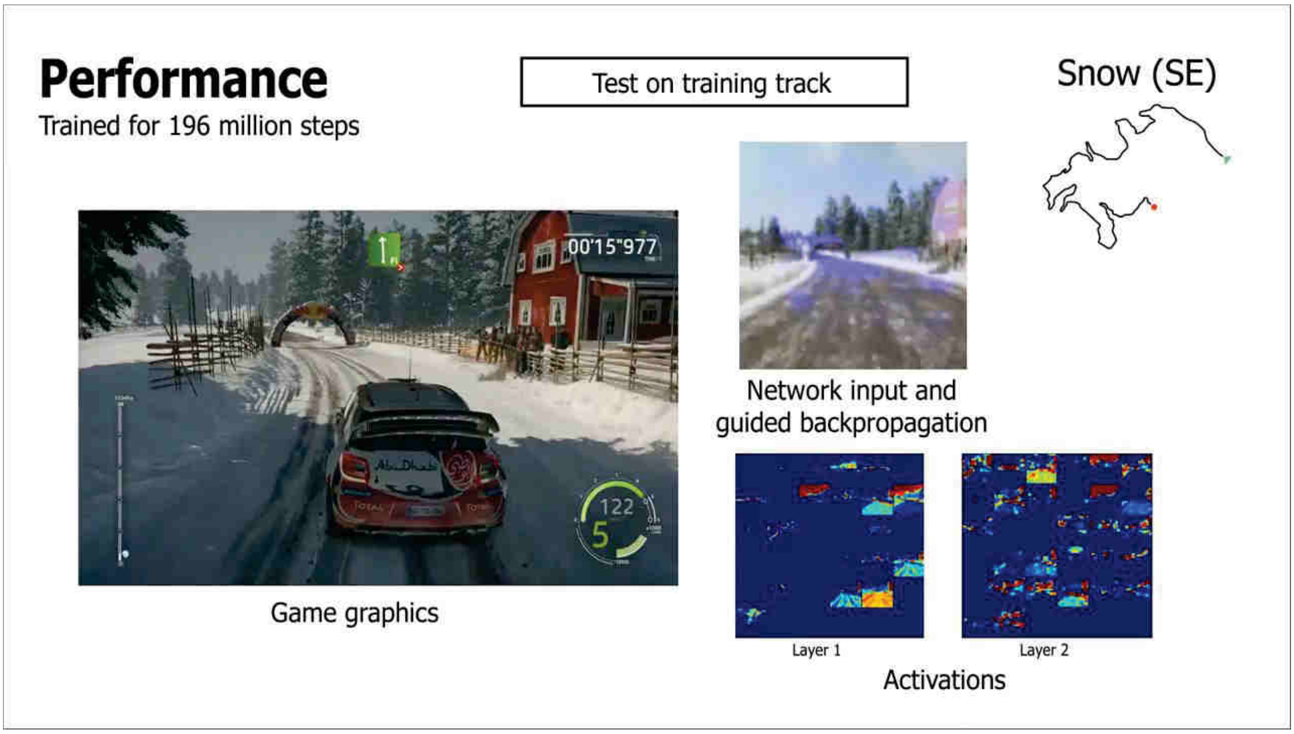


Find a "policy" $a_t=\pi(s_t)$ that maximizes

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

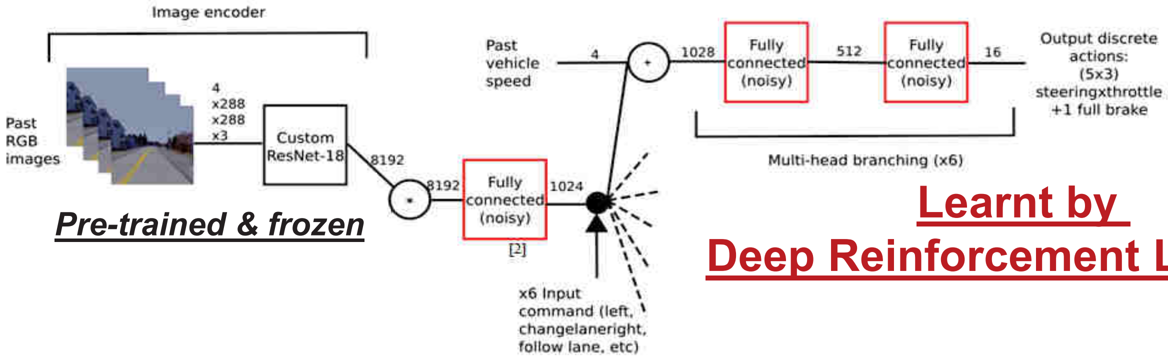
End-to-end driving by Deep Reinforcement Learning

End-to-end driving learnt by RL in a racing-car simulator



[End-to-End Race Driving with Deep Reinforcement Learning](), Maximilian Jaritz, Raoul De Charette, Marin Toromanoff, Etienne Perot, Fawzi Nashashibi, ICRA 2018 - IEEE International Conference on Robotics and Automation, Brisbane, Australia, May 2018.

Deep *Reinforcement* Learning for Automated Driving



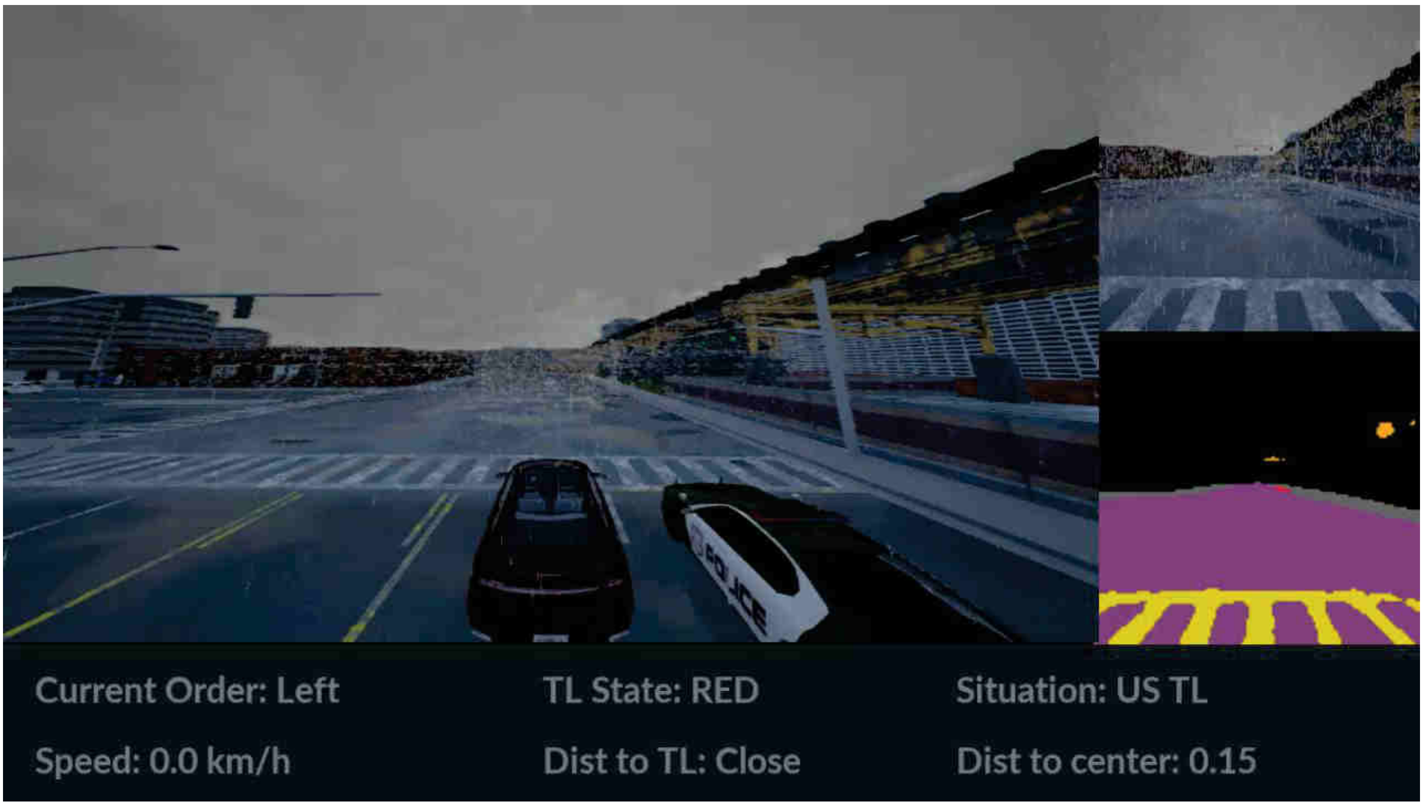
Learnt by Deep Reinforcement Learning

Work by Valeo/MinesParis industrial (CIFRE) PhD

1st prize at « CARLA Autonomous Driving challenge » in 2019&2020

https://leaderboard.carla.org/challenge

Examples of Autonomous Driving obtained with our DRL



"End-to-End Model-Free Reinforcement Learning for Urban Driving using Implicit Affordances", M.Toromanoff, E.Wirbel & F.Moutarde, CVPR'2020

Conclusion & perspectives

Major current AI challenges for Automated Driving:

- Quantified safety validation / HOMOLOGATION??
- Forecasting of road users' movements/trajectories
- Inference of HUMAN INTENTIONS (pedestrians&drivers)
- Coordination/collaboration with other road users
  - between AVs (cooperative planning, etc…)
  - with Humans:
    - Non-verbal communication (gestures, movement, gaze)
    - Learning of implicit "social rules"
- Learning of adaptive&complex BEHAVIOR

  Intelligent and Dynamic short-term planning of trajectories

Questions?