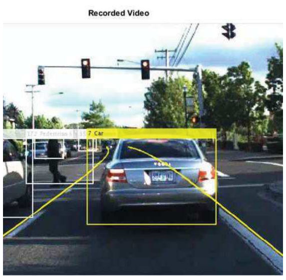
End-to-end driving from vision with Deep-Learning: from Imitation-Learning to Reinforcement-Learning

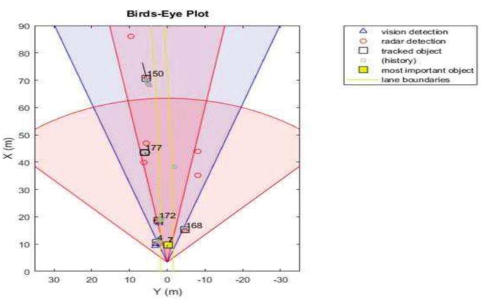
Pr. Fabien MOUTARDE
Center for Robotics
MINES ParisTech
PSL Université
Fabien.Moutarde@mines-paristech.fr
http://people.mines-paristech.fr/fabien.moutarde

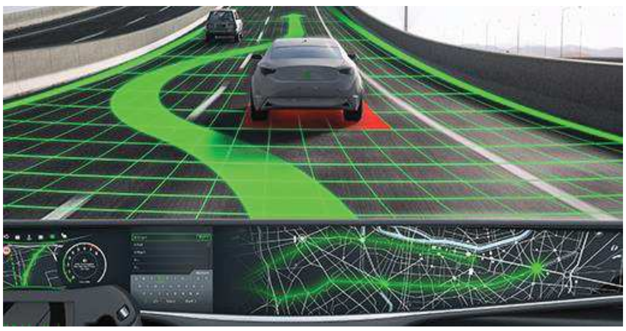
Idea of end-to-end driving

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

Current architecture for automated driving







vs. HUMAN driving =
turn/accelerate-brake by just looking in front
≈ *"intelligent" visual servoing*

Principle of end-to-end driving

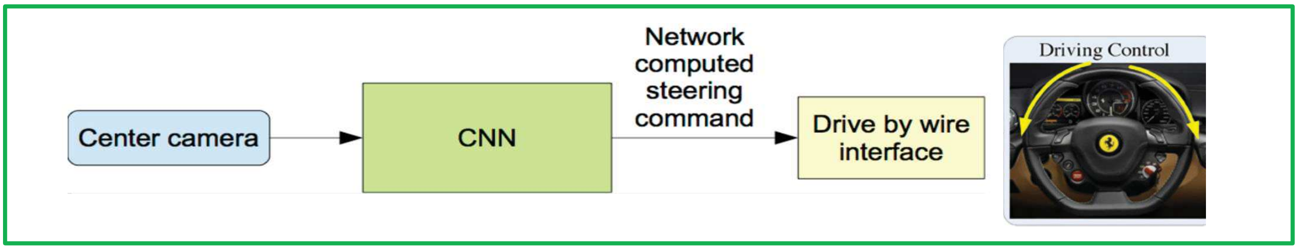
Sensors (cameras, radar, LIDAR, etc…) → Perception → Trajectory Planning → Control → Driving wheel, Acceleration or braking

Current architecture for automated driving

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

≈ "intelligent" visual servoing

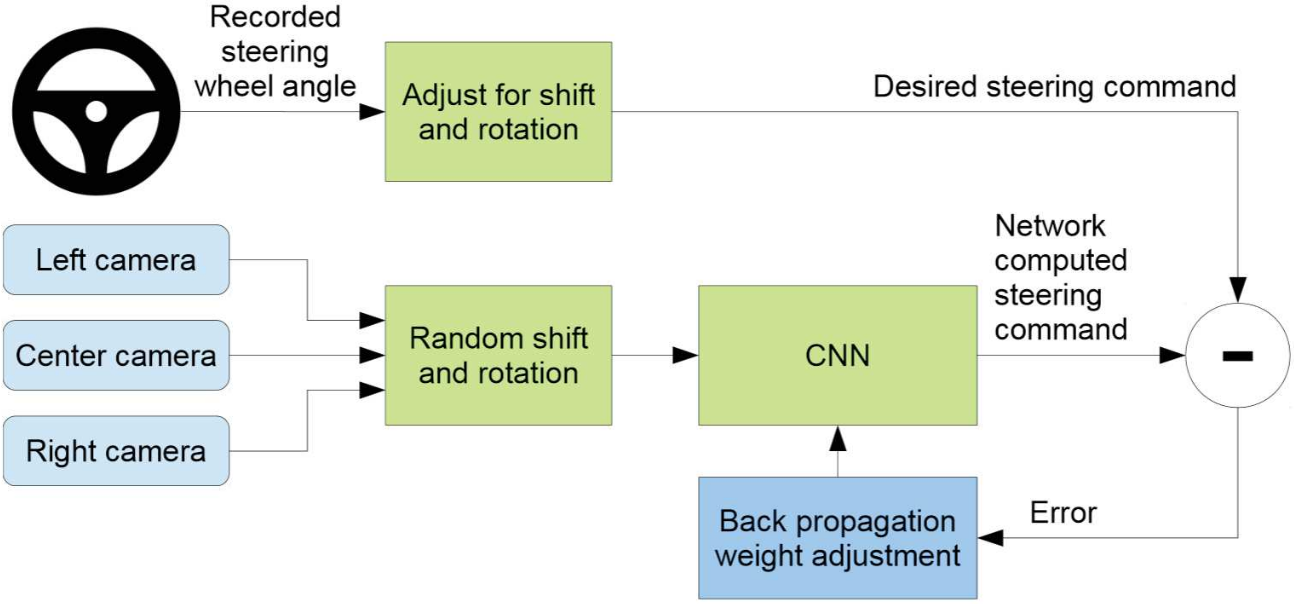
Sensors (cameras, radar, LIDAR, etc…) → Imitation Learning or Reinforcement Learning → Driving wheel, Acceleration or braking



Outline

- **Imitation Learning**
- Reinforcement Learning

Imitation Learning: "copying" human driver

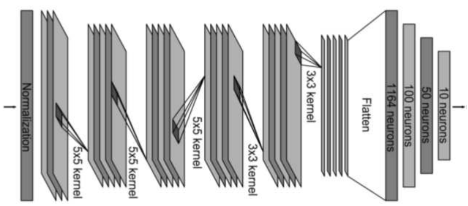


Nvidia approach

[Bojarski et al., End to end learning for self-driving cars. arXiv preprint arXiv:1604.07316 (2016)]

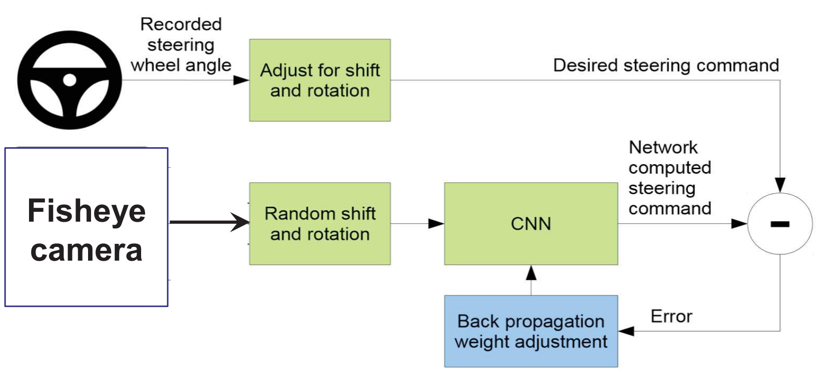
<u>*Imitation*</u> Learning approach proposed by Valeo

ConvNet input: Cylindrical projection of fisheye camera



ConvNet output: steering angle

"Copying" human driver



"End to End Vehicle Lateral Control Using a Single Fisheye Camera", **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.**

Real data for training and testing

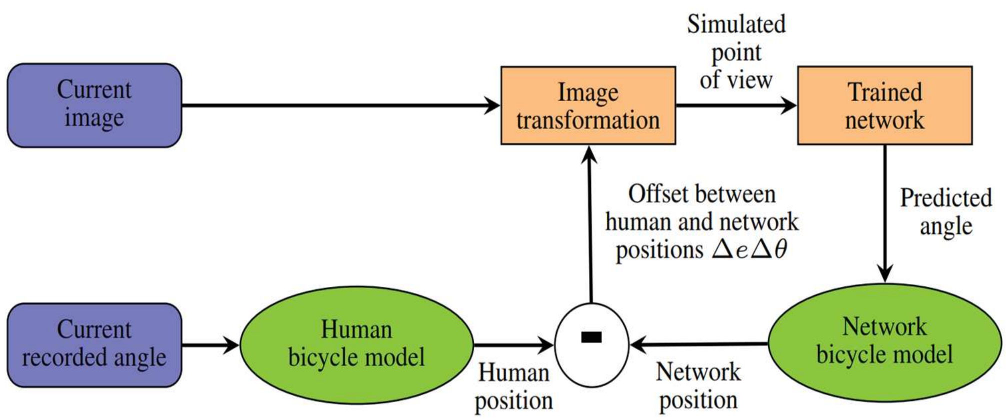
Training+testing dataset = 10000 km and 200 hours of human driving in openroad (highways, urban streets, country roads, etc…) under various weather conditions

TrainSet = 10 million images, TestSet = 3 million images.

TABLE IV: Description of test scenarios

Scenario	Urban	Highways	Sharp turns
Image count	100000	70000	15000
Duration (min)	56	39	8

"Simulator" with real images



"ConvNet in-the-loop" simulator with real images

[Work by my Valeo CIFRE PhD student Marin Toromanoff]

End-to-end driving: closed loop evaluation

Fig. 5: Screenshot of the simulator, blue trajectory is human, red is the car driven by the network. At current time, the network is translated of 7cm on the right. Here the network failed to take the turn, the red trajectory goes off-road, and a recovery is done.

TABLE V: Autonomy (%) and mean absolute distance (MAD, in cm) according to data distribution and validation scenario, the baseline is just going straight.

Scenario	Urban		Highways		Sharp turns	
Metric	Aut. (%)	MAD (cm)	Aut. (%)	MAD (cm)	Aut. (%)	MAD (cm)
Original	99.3	16	**98.7**	**19**	**73.7**	**30**
Sel. #1	98.9	**15**	97.7	25	83.7	**27**
Sel. #2	**99.5**	16	97.2	24	**87.5**	28
Oversamp.	**98**	**18**	**91.8**	**29**	82.5	29
Baseline	8	36	14	41	0	35

TABLE VI: Comparison of performance between individual networks and bagging

Scenario	Urban		Highways		Sharp turn	
Metric	Aut. (%)	MAD (cm)	Aut. (%)	MAD (cm)	Aut. (%)	MAD (cm)
Weights #1	**99.5**	16	97.2	24	**87.5**	28
Weights #2	98.9	15	97.7	25	83.7	**27**
Weights #3	99.3	16	**98.7**	**19**	**73.7**	**30**
Weights #4	98.6	**18**	**92**	**26**	85	29
Weights #5	**98.4**	15	96.4	21	83.7	28
Bagging	**99.5**	**13**	**98.7**	**19**	**87.5**	**27**

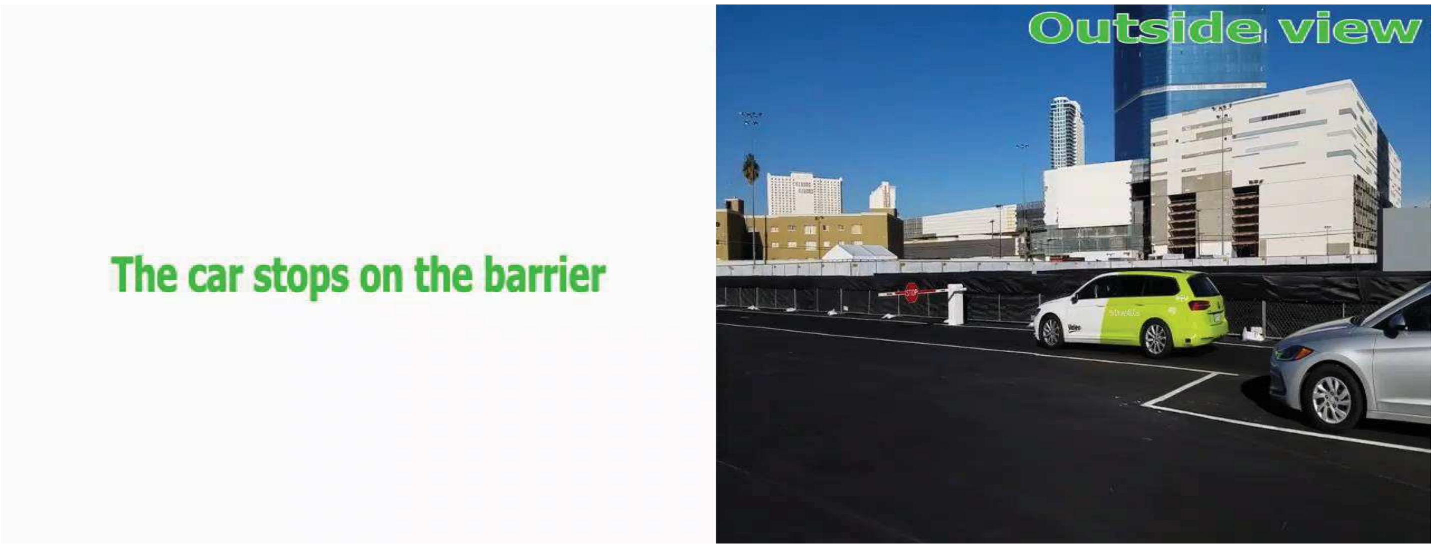
End-to-end driving (learnt purely on real-world data) tested in GTA simulator

[Work by my Valeo CIFRE PhD student Marin Toromanoff]

Real vehicle end-to-end driving (learnt by imitation)

[Work by Valeo using ConvNet trained by my CIFRE PhD student Marin Toromanoff]

Demonstration at CES'2018



[Work by Valeo using ConvNet trained by my CIFRE PhD student Marin Toromanoff]

"End to End Vehicle Lateral Control Using a Single Fisheye Camera", 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.

Synthesis on Imitation Learning

- Aims at "copying" behavior of human drivers
- Requires large amount of real-world training data: video + driving-wheel AND brake-throttle data durings many hours and thousands of varied-enough km driven by human
- Very promising generalization performances

BUT:

- Rare events under-represented in dataset
- Dangerous events unavailable in dataset !!

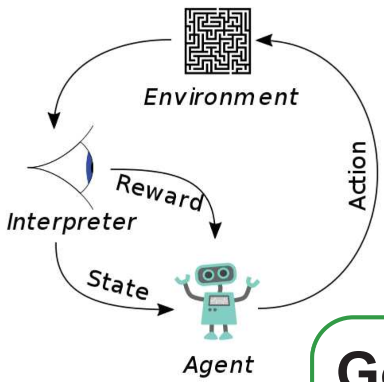
Outline

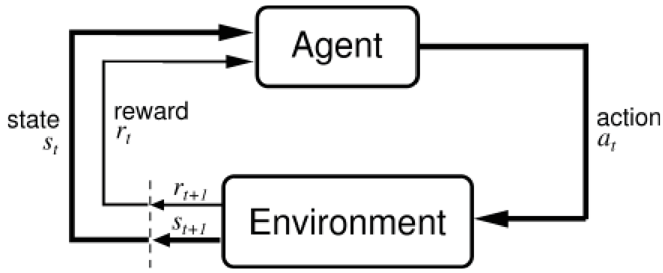
- Imitation Learning
- Reinforcement Learning

Preliminary RL experiment for end-to-end driving

[Work by my Valeo CIFRE PhD student Marin Toromanoff]

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 π searched as ConvNet(front-image)

Value and Q-function

- **Value of a state (for a given policy)**

$$V_\pi(s) = \mathbb{E}_\pi[R_t|s_t = s] = \mathbb{E}_\pi[\sum_{k=0}^{T} \gamma^t r_{t+k}|s_t = s]$$

THERE ALWAYS EXISTS A *DETERMINISTIC* OPTIMAL POLICY π*

$$\forall \pi, \forall s \in S, V_{\pi^*}(s) \geqslant V_\pi(s)$$

- **Q-function**

$$Q_\pi(s, a) = \mathbb{E}_\pi[R_t|s_t = s, a_t = a] = \mathbb{E}_\pi[\sum_{k=0}^{T} \gamma^t r_{t+k}|s_t = s, a_t = a]$$

Richard S. Sutton and Andrew G. Barto, Reinforcement Learning: An Introduction (2017)

3 families of RL algorithms

- **Policy-based** $\pi_\theta \approx \pi^*$
  optimize a *parameterized policy*
- **Value-based** $Q(s, a, \theta) \simeq Q^{\pi^*}(s, a)$
  find optimal (parameterized) Q-value
  → then π(s) = argMax Q*(s,a)

Model-free

- **Model-based** $m(s_t, a_t, \theta') \approx s_{t+1}, r_{t+1}$

Deep Reinforcement Learning with Deep Q Network (DQN)

- **Q-learning:** $Q^{new}(s_t, a_t) \leftarrow (1-\alpha)\cdot \underbrace{Q(s_t,a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \overbrace{\Big(\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1}, a)}_{\text{estimate of optimal future value}} \Big)}^{\text{learned value}}$

- **Optimal policy deduced from optimal Q-value**

$$\pi^*(s) = arg\max_a Q_{\pi^*}(s,a)$$

- **DQN** [1]**: if too many possible states, approximate Q as a neural network, and learn Q* using SGD with loss from Bellman equation**

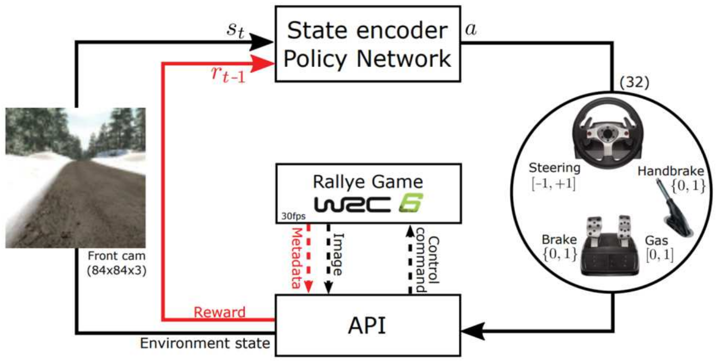
$$L(s_t, a_t, r_{t+1}, s_{t+1}, \theta) = (\underbrace{r_{t+1} + \gamma \max_a Q(s_{t+1}, a, \theta)}_{target} - Q(s_t, a_t, \theta))^2$$

[1] V. Mnih et al : Human-level control through deep reinforcement learning (2015)

Typology of RL algorithms

Famille	Algorithme	On/Off policy	Domaine Action continu/discret
Basés sur la politique	REINFORCE	On policy	Les 2
	Acteur Critique (A3C)	On policy	Les 2
Basés sur la fonction de valeur	SARSA	On policy	Discret
	Q-Learning (DQN)	Off-policy	Discret
Basée sur un modèle	MCTS	Off-policy	Discret
	iLQG	Off-policy	Continu

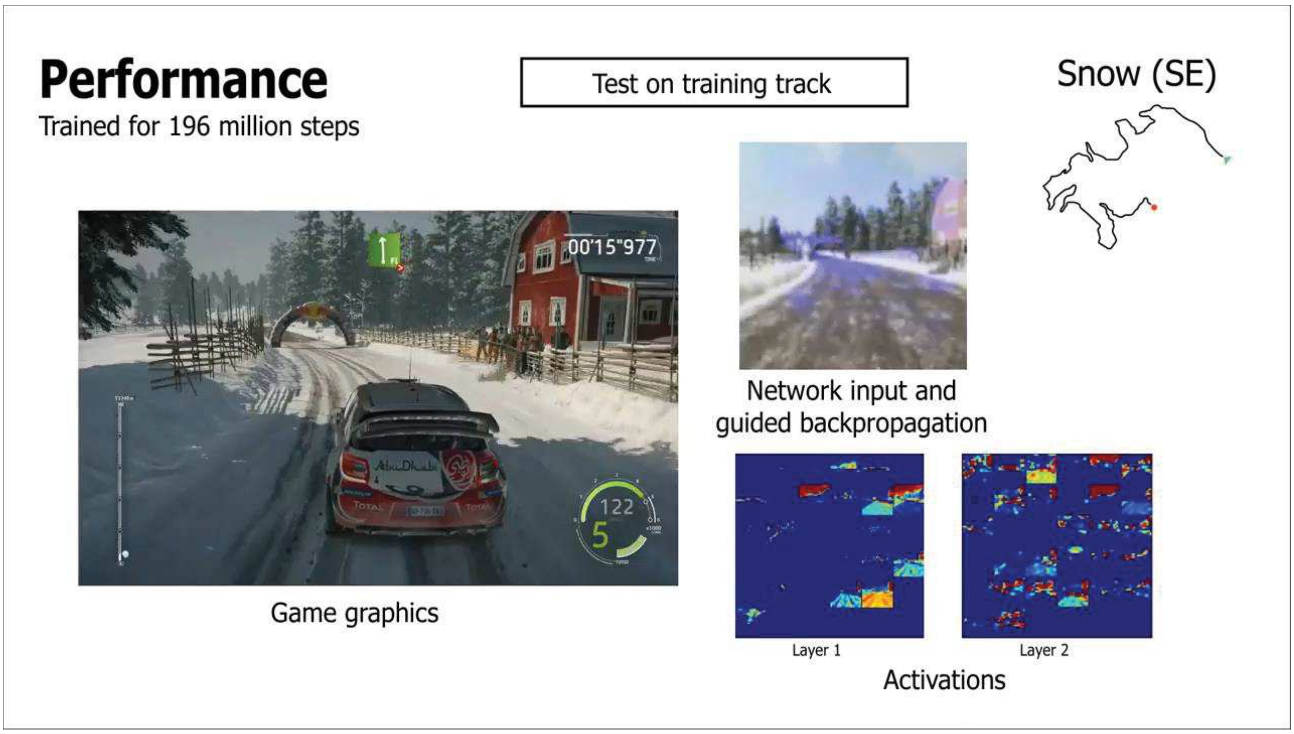
End-to-end driving learning by RL in racing-car simulator



- **DRL used: Actor-Critic (A3C)**
- **Reward: faster=better (race!) + penalize offroad**

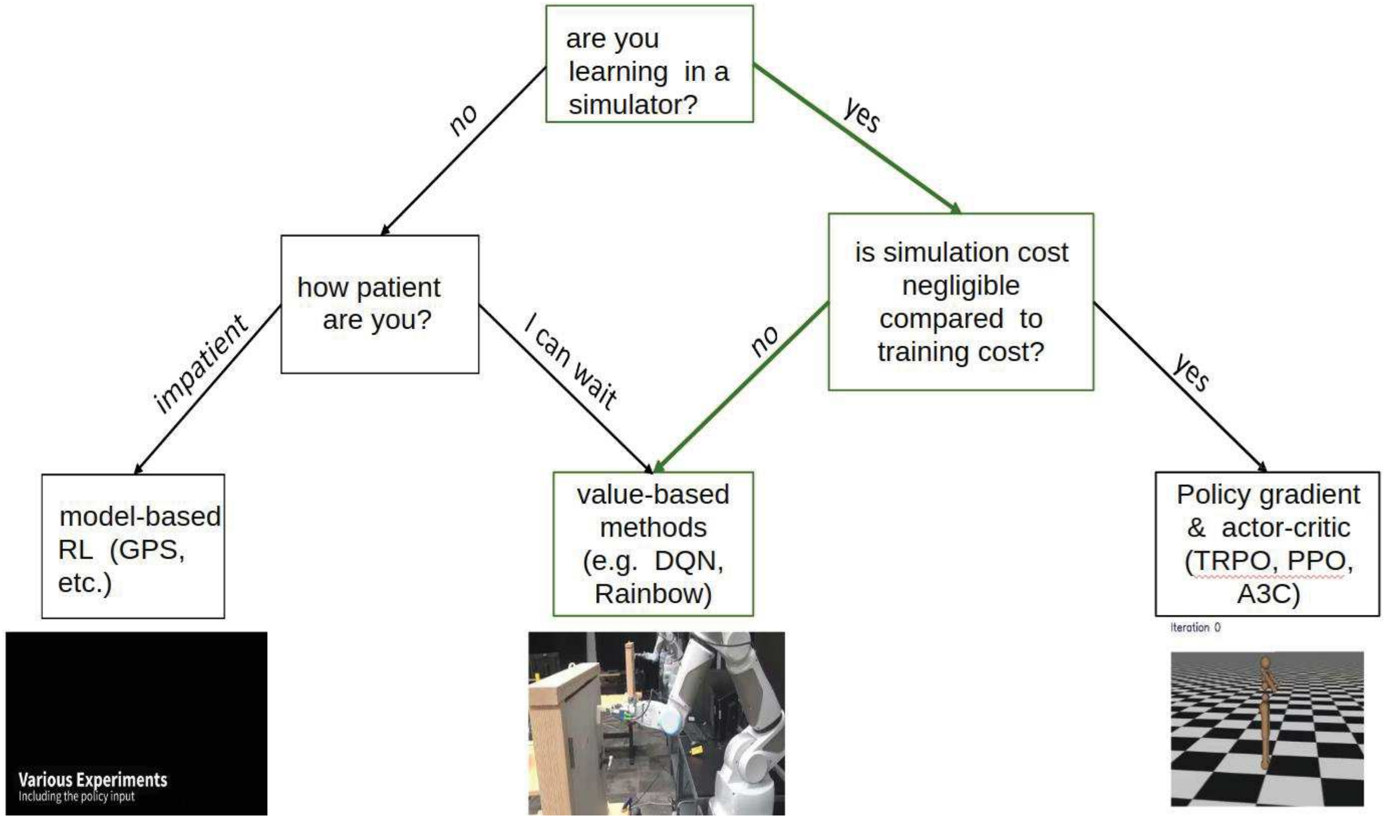
Etienne Perot, Maximilian Jaritz, Marin Toromanoff, Raoul De Charette. End-to-End Driving in a Realistic Racing Game with Deep Reinforcement Learning, International conference on Computer Vision and Pattern Recognition - Workshop, Honolulu, United States, Jul. 2017.

End-to-end driving learnt by RL in 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.

Choice of RL algo?



[1] S. Levine: CS294 *Deep Reinforcement Learning* (2017)

RL for Automated Driving: learn in a simulator!

- **RL requires huge amount of trials AND ERRORS ⇒ Simulation REQUIRED (safety + speed)**
- **Still few driving simulators adapted for DL and RL, and best ones not totally mature**

Simulateur	GTA	DeepDrive.io	AirSim	CARLA [1]
Flexibilité	−−	++	++	++
Variété	++	−−	−	+
Complexité/Réalisme	++	−−	−	−
Objets mobiles	++	−−	−−	+
Vitesse exécution	−−	+	+	+
Multi-agent	−−	−	−	++

[1] A. Dosovitskiy: CARLA: An Open Urban Driving Simulator (2017)

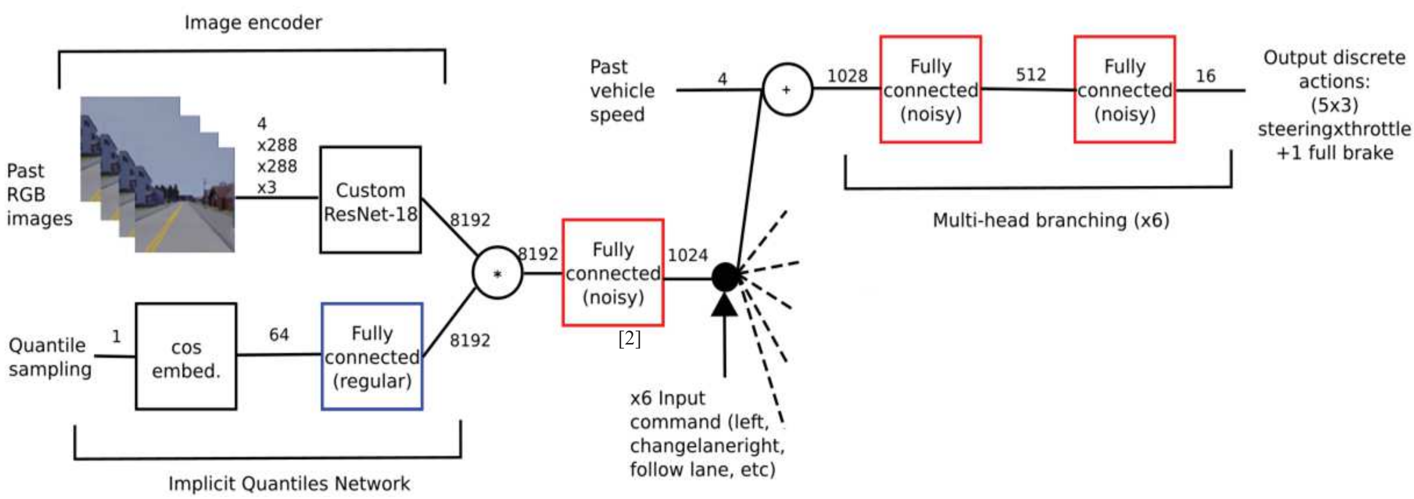
CARLA Autonomous Driving challenge

- **Itinerary to be followed in a city (given by 4 possible orders at intersections: Left, Straight, Right, Follow_Lane)**
  BUT must <u>stay on the road</u>, in the lane, respecting Traffic Lights, and no collision with pedestrians and other cars!
- **Evaluation metrics = Task completion & Distance between infractions, in an UNSEEN CITY**
- **Results (and 10.000$ for winners!) on 2019 July 1st**

Network architecture

- **U.S. Traffic lights ➔ Need to use *COLOR and high-enough resolution* ⇒ big network, hard and slow to train**
- **Use a resnet-18 (10 times more weight than previously used in *DQN-like* network)**
- **Handle turn-orders (at intersections) with multi-head branching [1]**

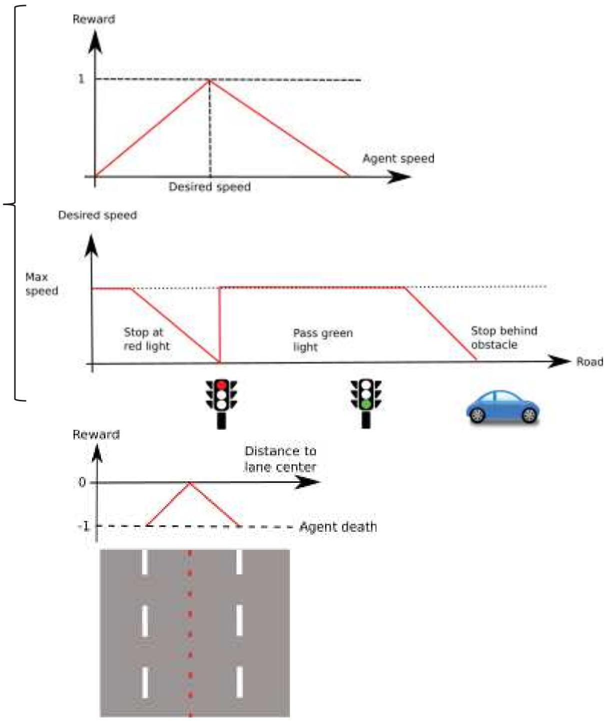
[1] Codevilla et al., *End-to-end driving via Conditional Imitation Learning*, 2017
[2] M. Fortunato et al.. *Noisy Networks for Exploration*, 2017



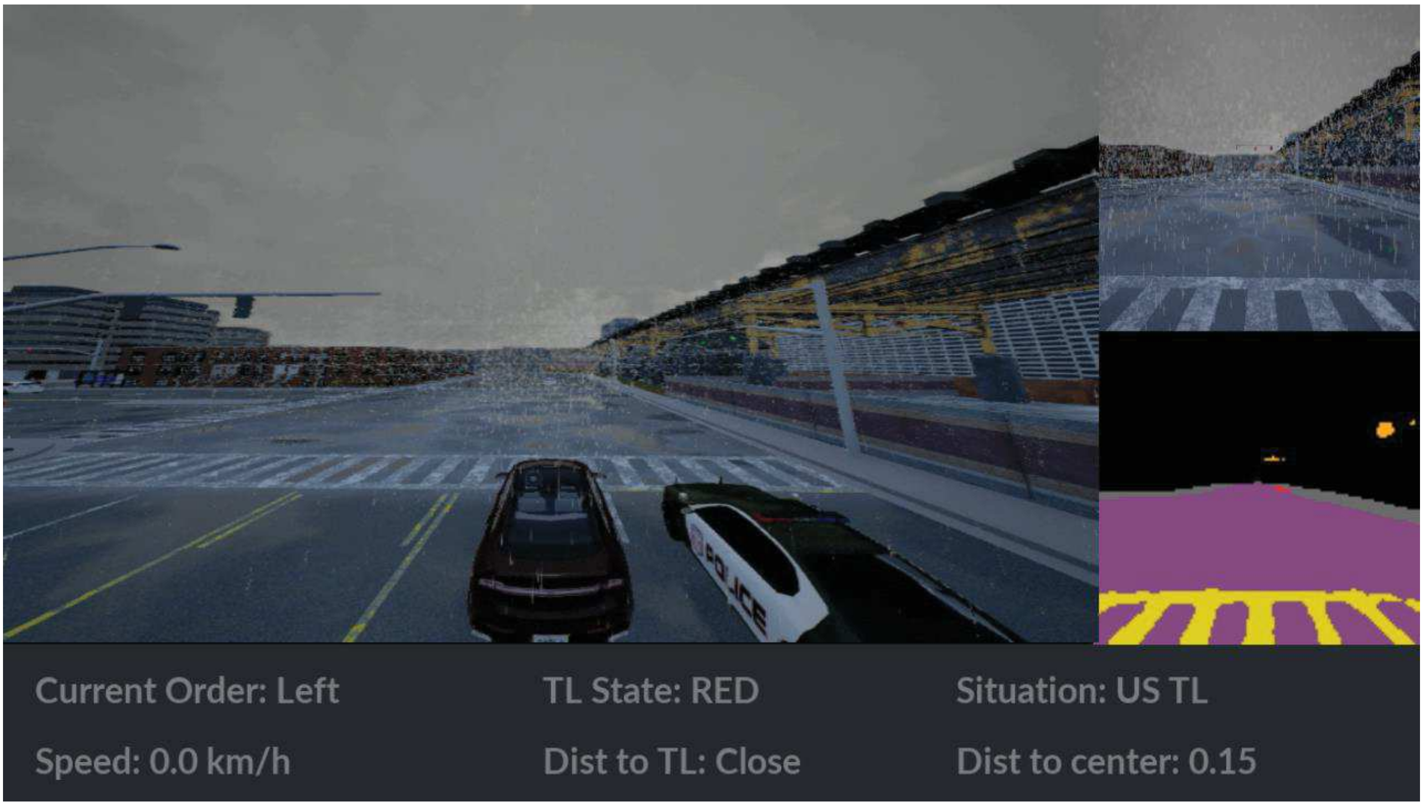
Reward shaping

Rewards scaled in [-1, 1]:

- **Speed: positive reward to follow speed, depends on obstacles & traffic light**
- **Lateral position: negative reward depending on distance to lane center**
- **Episode terminates on collision, running red traffic light, too far from lane center or stuck (if no reason to stop)**



Examples of autonomous driving obtained with our DRL



Conclusions & perspectives on DRL for Automated driving

- DRL allows to learn behavior *without any example provided by human*
- Only the REWARD needed to define objectives

- Very encouraging first results in simulation
- Winner of "vision-only" track at CARLA « Autonomous Driving challenge » !!

- Future work:
  - transferrability to real-world videos
  - Combination of Imitation-Learning and RL?