



# **End-to-End Model-Free Reinforcement Learning for Urban Driving using Implicit Affordances**

Marin Toromanoff, Emilie Wirbel, Fabien Moutarde

Center for Robotics,  
MINES ParisTech, PSL  
16 marzo 2020

# Summary



## **Winner of camera Only track of the CARLA challenge**

much more challenging than the original CARLA benchmark

## **The first RL agent successfully driving from vision**

in urban environment including intersection management and traffic lights detection besides lane keeping, pedestrians and vehicles avoidance

## **implicit affordances**

Allowing training of replay memory based RL with much larger network and input size than most of network used in previous RL works.

## **Increased complexity**

Extensive parameters and ablation studies of implicit affordances and reward shaping

# Challenges in AD



## Highly variable situations

End-to-end system needed



## Imitation learning leads to distribution mismatch

human driver is always in an almost perfect situation




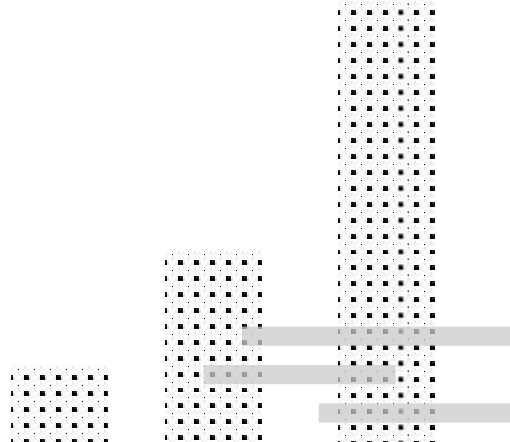
## Data augmentation

currently mostly limited to lane keeping and lateral control



# DRL TO THE RESCUE



- 
- Does not suffer from distribution mismatch
  - RL algorithms rely on a replay buffers allowing to learn from past experiments
- 



# DRL DRAWBACKS

- It can need a magnitude larger amount of data than supervised learning to converge
- difficulties when training large networks with many parameters
- buffers can limit the size of the input used (e.g. the size of the image)
- algorithm appears as a black box from which it is difficult to understand how the decision was taken

# DRL DRAWBACKS

- It can need a magnitude larger amount of data than supervised learning to converge
- difficulties when training large networks with many parameters
- buffers can limit the size of the input used (e.g. the size of the image)
- algorithm appears as a black box from which it is difficult to understand how the decision was taken
- **Use coined affordances = privileged information as auxiliary losses**

# End-to-End Autonomous Driving with RL:

## Related work

- Imitation learning better results than an RL baseline using the A3C algorithm with discrete actions.
- RL with DDPG and continuous actions to fine-tune an imitation agent
- “Learning to Drive in a Day” in which an agent is trained directly on the real car for steering.
- A really recent work also integrates RL on a real car and compares different ways of transferring knowledge learned in CARLA in the real world.

# End-to-End Autonomous Driving with RL:

## Related work

- SplitNet finding features from perception task >> model-free RL agent
- network to predict high-level information such as probability of collision or being off-road in the near futures >> model-based RL scheme

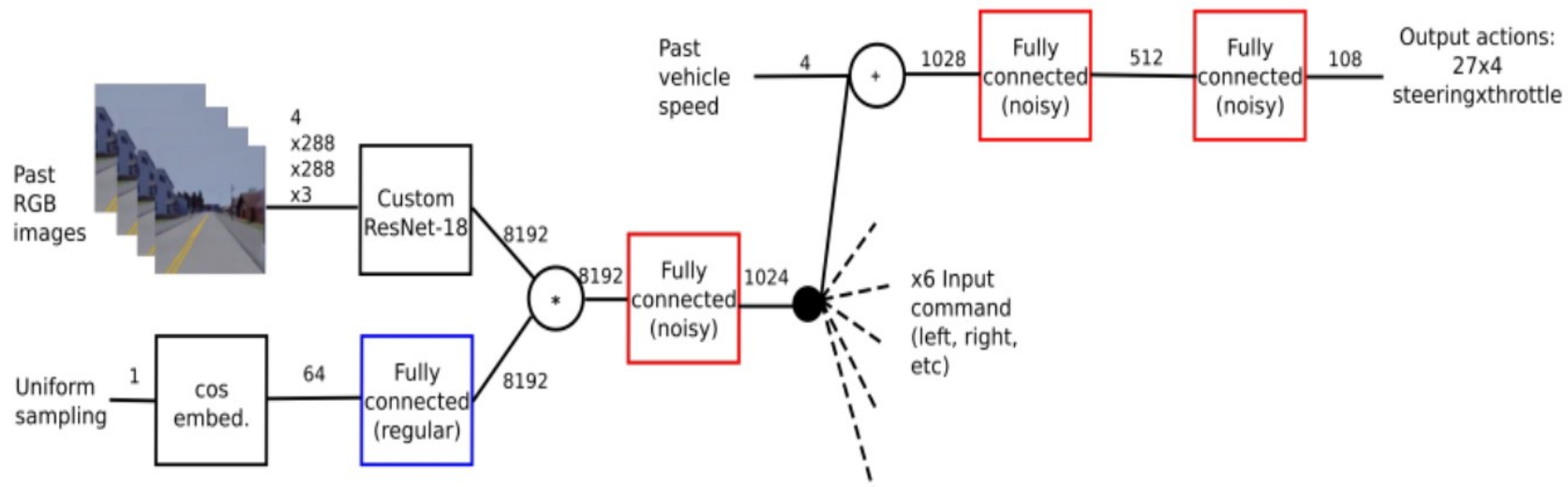


# RL Setup: Rainbow-IQN Ape-X

- Value-based RL as it is the current state-of-the-art on Atari

# RL Setup

- Value-based RL as it is the current state-of-the-art on Atari
- Implementation of Rainbow-IQN Ape-X without the dueling network  
<https://arxiv.org/pdf/1803.00933v1.pdf>



# RL Setup: Rainbow-IQN Ape-X



Unlike Most of networks  
in model-free RL &  
images = small network

**18 convolutional layers  
and 30M parameters**



4×288×288×3 gray scale  
image by concatenating  
4 consecutive frames

**temporality & traffic  
lights detection**



Resnet-18 and a conditional network  
to handle 6 different maneuvers

**most of state-of-the art advances in  
supervised learning such as  
residual connections and  
batchnorm**

# RL Setup

- Value-based RL as it is the current state-of-the-art on Atari
- Implementation of Rainbow-IQN Ape-X without the dueling network  
<https://arxiv.org/pdf/1803.00933v1.pdf>
- Train on multiple maps of CARLA at the same time

# RL Setup

- Value-based RL as it is the current state-of-the-art on Atari
- Implementation of Rainbow-IQN Ape-X without the dueling network  
<https://arxiv.org/pdf/1803.00933v1.pdf>
- Train on multiple maps of CARLA at the same time
- The reward used relies mostly on the waypoint API

# RL Setup

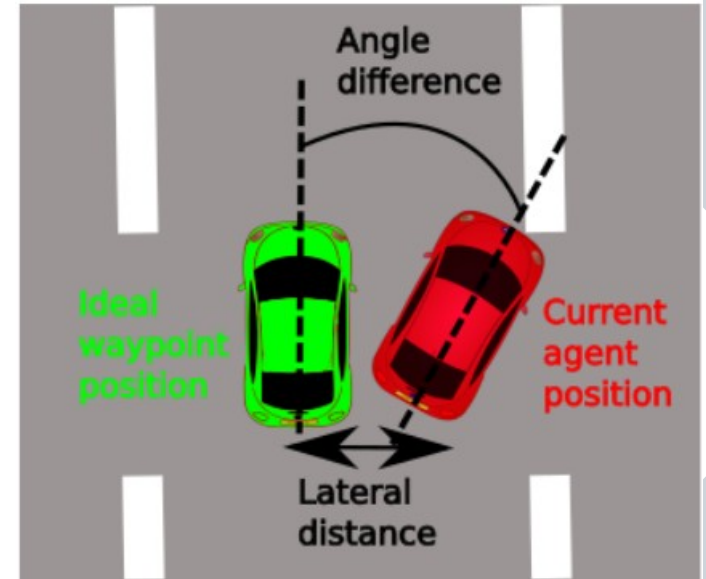
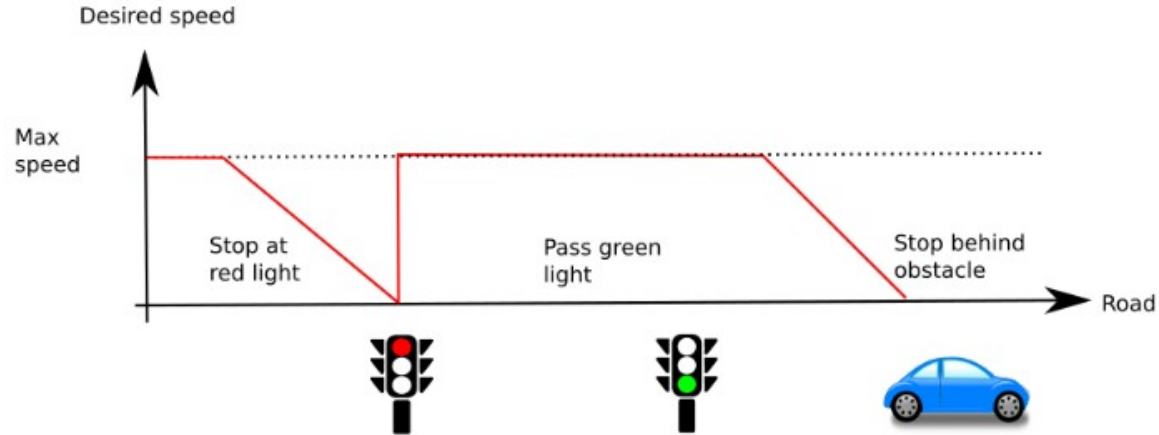
- Value-based RL as it is the current state-of-the-art on Atari
- Implementation of Rainbow-IQN Ape-X without the dueling network  
<https://arxiv.org/pdf/1803.00933v1.pdf>
- Train on multiple maps of CARLA at the same time
- The reward used relies mostly on the waypoint API
- At the beginning of an episode, the agent is initialized on a random waypoint on the city

# RL Setup

- Value-based RL as it is the current state-of-the-art on Atari
- Implementation of Rainbow-IQN Ape-X without the dueling network  
<https://arxiv.org/pdf/1803.00933v1.pdf>
- Train on multiple maps of CARLA at the same time
- The reward used relies mostly on the waypoint API
- At the beginning of an episode, the agent is initialized on a random waypoint on the city
- At an intersection, give a random possible maneuver (Left, Straight or Right) to the agent. The reward relies on three main components: desired speed, desired position and desired rotation.



# RL Setup: Reward function

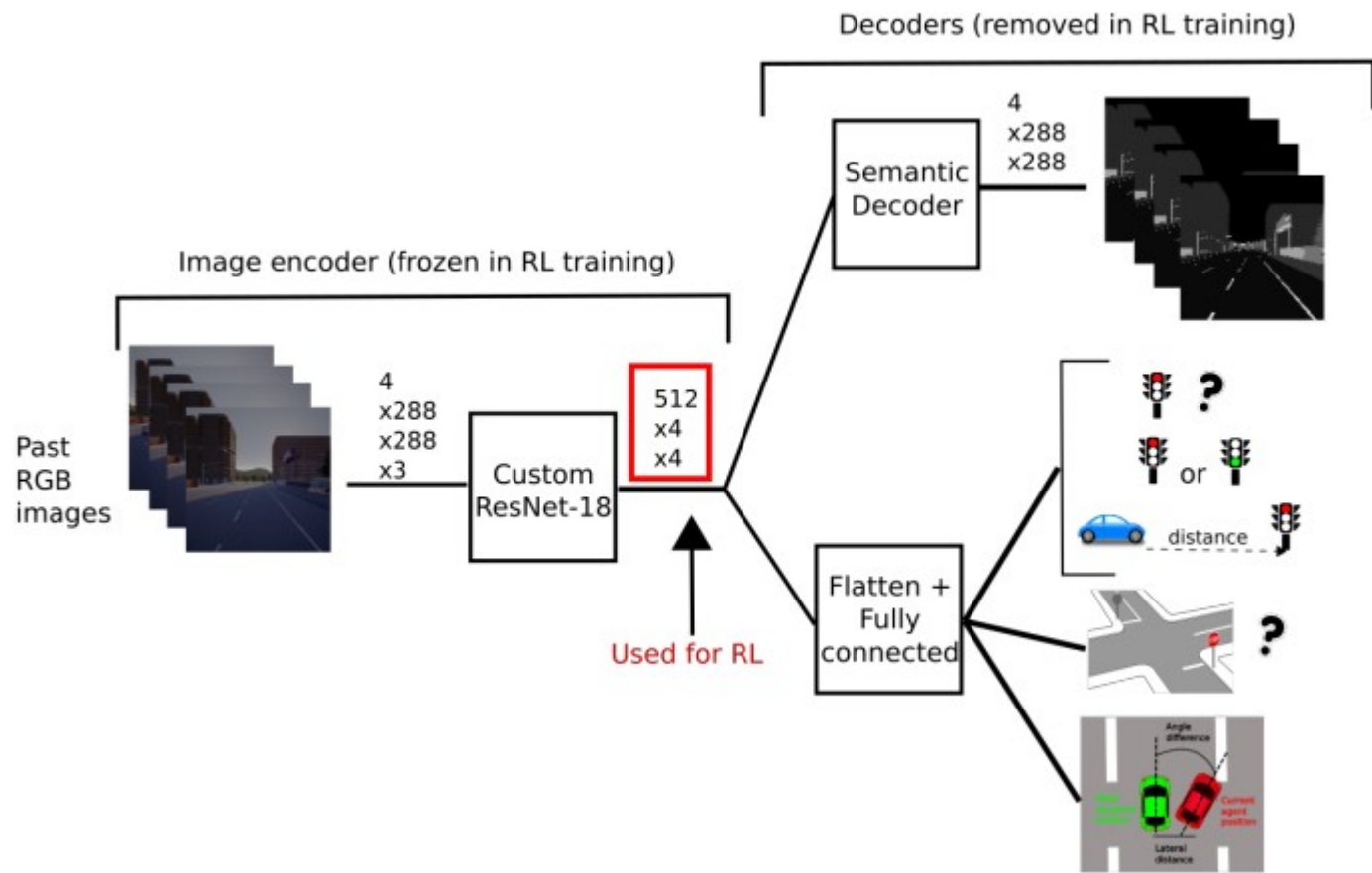


# DRL with large networks challenges

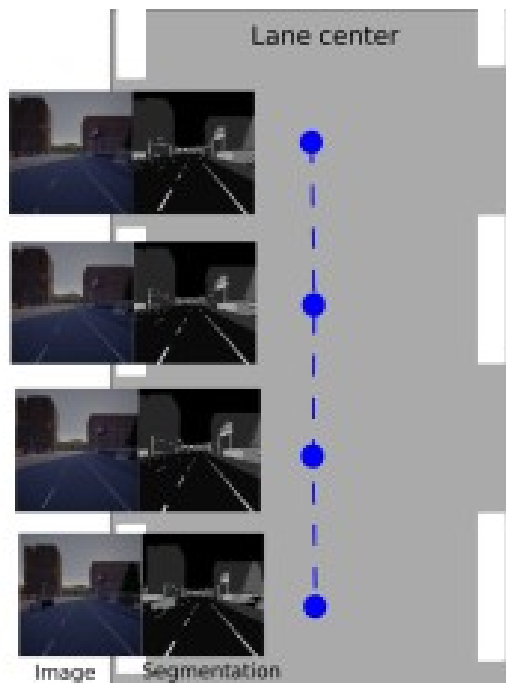
- Much longer and harder to train
- replay memory = memory usage constraints



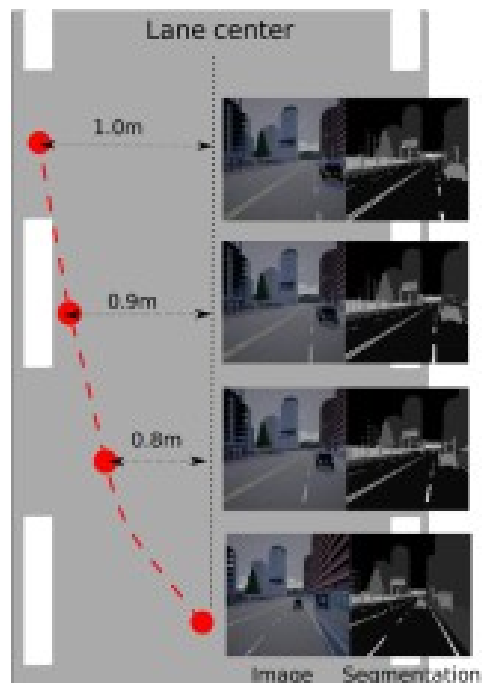
**pre-train the convolutional encoder part of the network to predict some high-level information and then freeze it while training the RL**



# Data augmentation



Encoder Training (autopilot)



RL driving (online trajectory)

# Actions

$9 < \text{steering values} < 27$

Throttle = 3

Braking = 1

36 ( $9 \times 4$ ) or 108 ( $27 \times 4$ ) actions for our experiments

# Actions



To reach more fine-grained discrete actions

simultaneously use multiple predictions of different trained agents and average them

# Experiments and Ablation Studies: Scenario

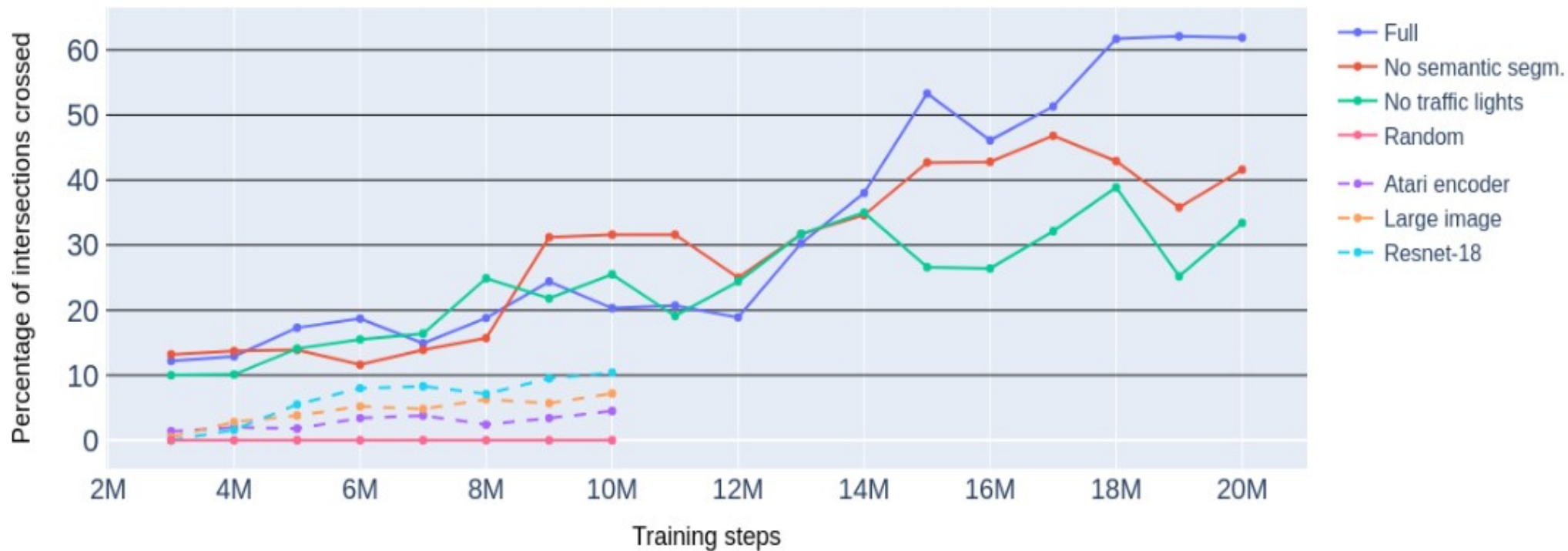
- Hardest environment in the available maps of CARLA
- Randomly spawn pedestrians crossing the road ahead of our agent
- Changing weather
- 20M iterations on CARLA, with 3 actors (6.6M steps for each actor) with framerate of 10 FPS. Equivalent to 20 days of simulated driving
- 10 scenarios of urban situations each one consisting in 10 consecutive intersections

# Experiments and Ablation Studies: metrics

- Average percentage of intersections successfully crossed
- Percentage of traffic lights passed without infraction
- Percentage of pedestrians passed without collision
- the mean absolute rotation between the agent and the road along the episode



# Experiments and Ablation Studies: Results



## Experiments and Ablation Studies: Results

Encoder used	Inters.	TL	Ped.
Random	0%	NA	NA
No TL state	33.4%	80%	82%
No segmentation	41.6%	96.5%	63%
All affordances	61.9%	97.6%	76%

## Experiments and Ablation Studies: Results

Task	RL	CoRL2017 (train town)				Ours	NoCrash (train town)		
		CAL	CILRS	LBC			Task	LBC	Ours
Straight	89	100	96	100		100	Empty	97	100
One turn	34	97	92	100		100	Regular	93	96
Navigation	14	92	95	100		100	Dense	71	70
Nav. dynamic	7	83	92	100		100			

Task	RL	CoRL2017 (test town)				Ours	NoCrash (test town)		
		CAL	CILRS	LBC			Task	LBC	Ours
Straight	74	93	96	100		100	Empty	100	99
One turn	12	82	84	100		100	Regular	94	87
Navigation	3	70	69	98		100	Dense	51	42
Nav. dynamic	2	64	66	99		98			

Table 4. Success rate comparison (in % for each task and scenario, more is better) with baselines [7, 31, 5, 3] on train weathers.

## Experiments and Ablation Studies: Generalization

Training	Unseen EU Town	Unseen US Town
Only Town05	2.4%	42.6%
Multi town	58.4%	36.2%

Table 3. Generalization performance (*Inters.* metric).

## Future work

- apply our implicit affordances scheme for policy-based or actor-critic
- to train our affordance encoder on real images in order to apply this method on a real car.