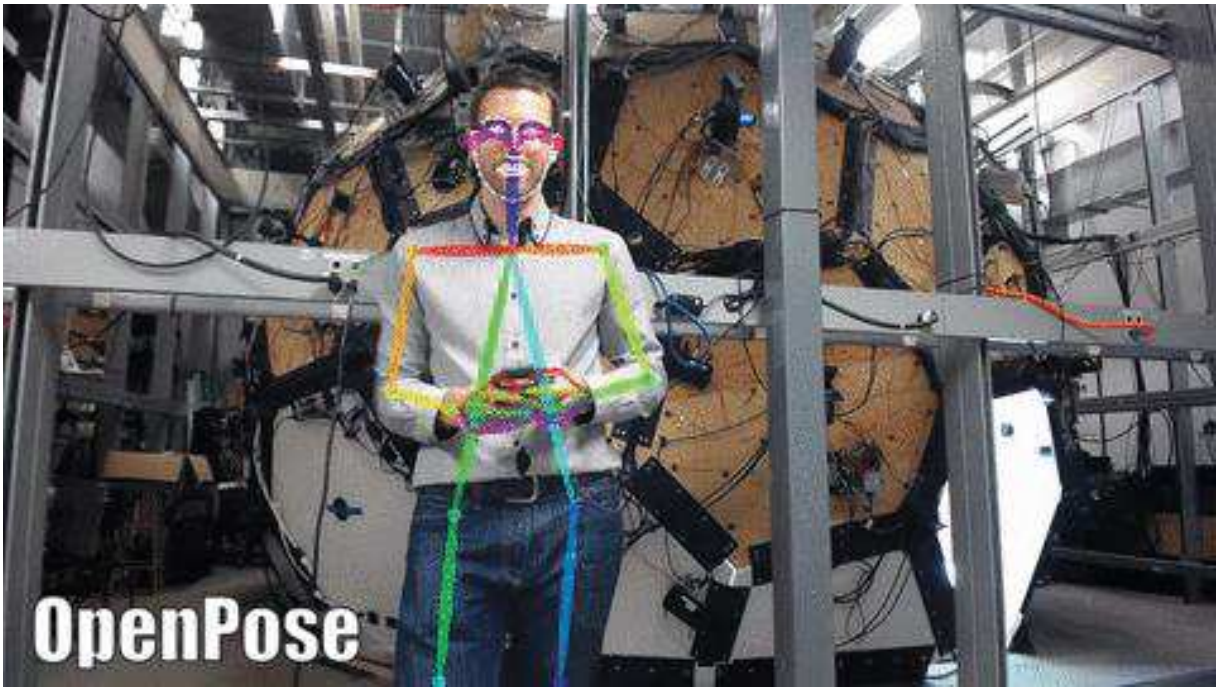
Outline

- Recalls on Convolutional Neural Networks (CNN or ConvNets) and Deep-Learning
- Transfer Learning
- Beyond Image Classification: DETECTION OF OBJECTS
- Instance segmentation with DeepLearning
- DL for Human pose inference and depth estimation
- Semantic segmentation with DeepLearning
- Interest and use of simulations / synthetic videos

Now possible to estimate Human poses from RGB images!



Real-time estimation of Human poses on *RGB* video

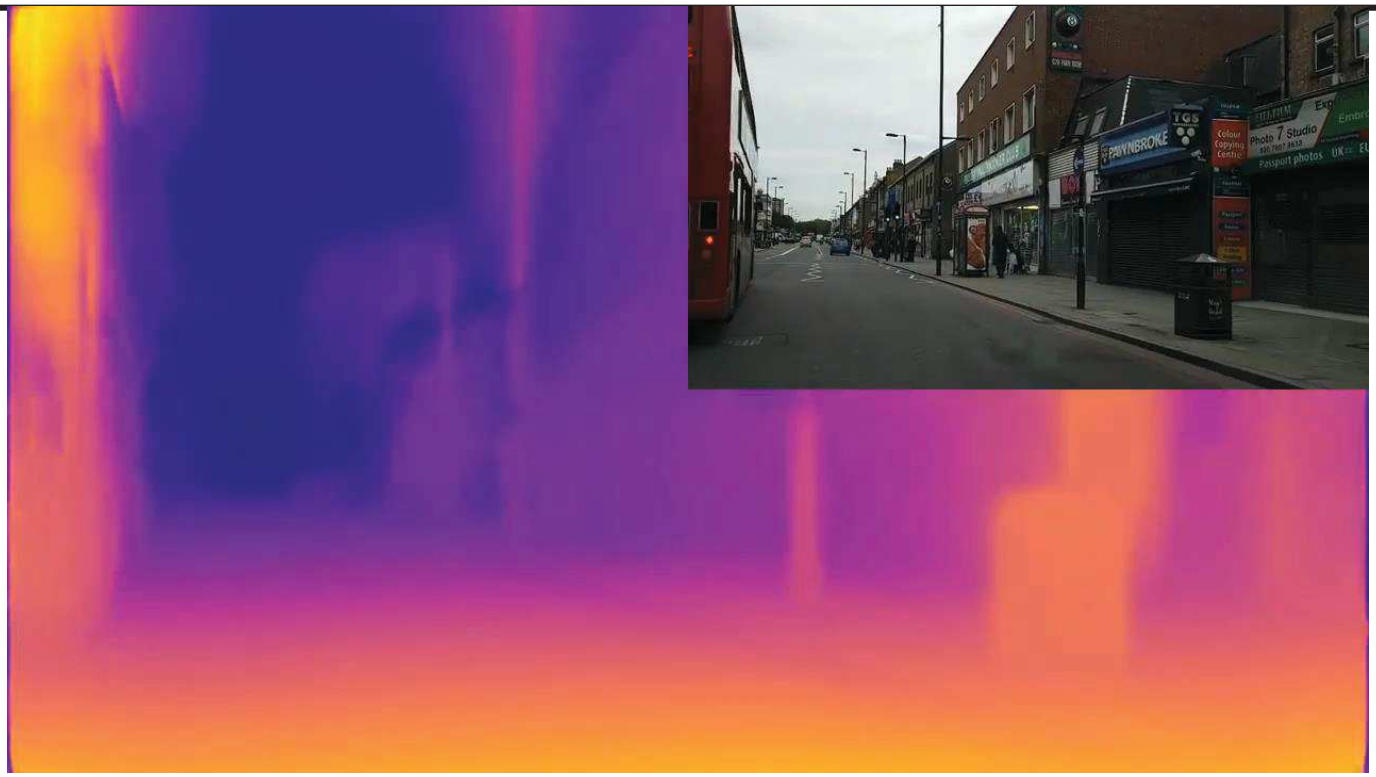
[Realtime Multi-Person 2D Pose Estimation using Part Affinity Field, Cao et al., CVPR'2017 [CMU]

OpenPose on streets

Human Pose estimation by DL methods

- **OpenPose = 2D pose, bottom-up (localize joints, then assemble them into skeletons)**
- **AlphaPose = 2D pose, top-down, slower and less robuts**
- **HMR (Human Mesh Recovery) = 3D pose + estimate body SURFACE as a mesh**

Inference of 3D (depth) from monocular vision

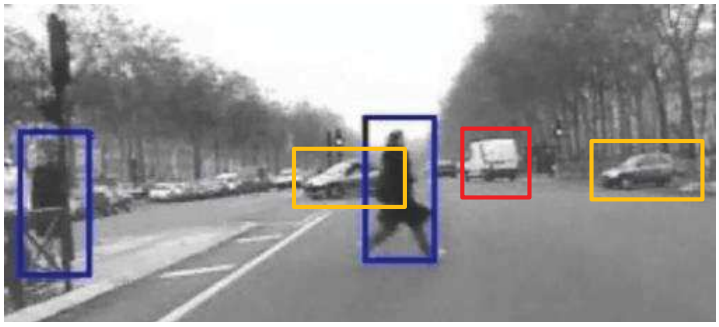


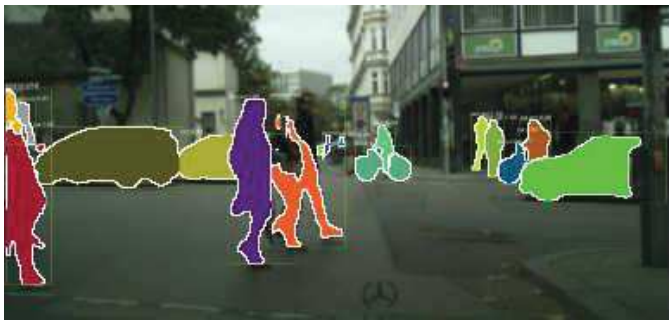
Unsupervised monocular depth estimation with left-right consistency
C Godard, O Mac Aodha, GJ Brostow - CVPR'2017 [UCL]

Outline

- Recalls on Convolutional Neural Networks (CNN or ConvNets) and Deep-Learning
- Transfer Learning
- Beyond Image Classification: DETECTION OF OBJECTS
- Instance segmentation with DeepLearning
- DL for Human pose inference and depth estimation
- Semantic segmentation with DeepLearning
- Interest and use of simulations / synthetic videos

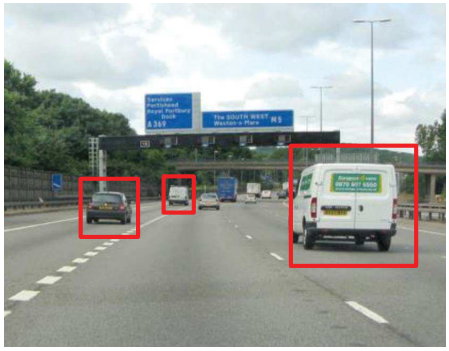
Drawbacks of object detections approach

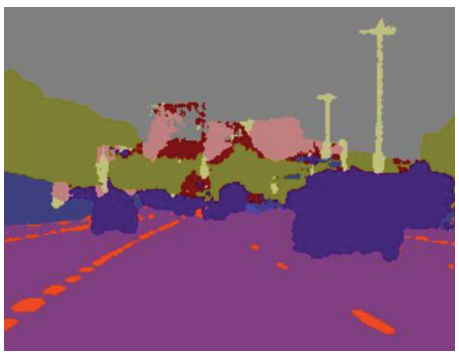




- Problem for objects without sharp boundaries (trees, …) or very dense group of objects (crowd of pedestrians, …)
- Only « compact » objects are categorized
  (what about « road », « sidewalk », « building », …?)

Advantage of Semantic (full) segmentation





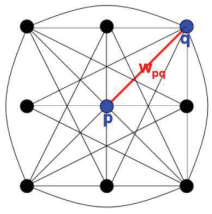
- One single semantic segmenter → all interesting object categories (cars, pedestrians, signs, etc…) and categorization of whole image
- Can also categorize non-compact areas (road, sky, buildings, trees, traffic lanes…)

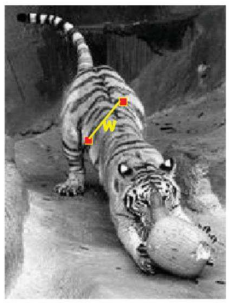
What is image SEGMENTATION?

Identify groups of *contiguous* pixels (connex sets) that « go together »

Many ≠ approaches for image segmentation

- **Clustering (K-means, GMM, MeanShift, …)**
- **Graph-based (graph-cuts)**

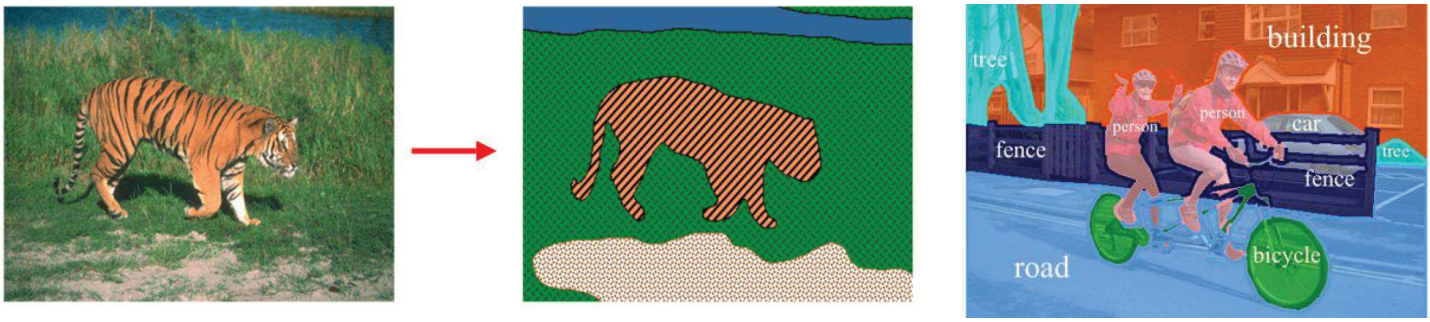




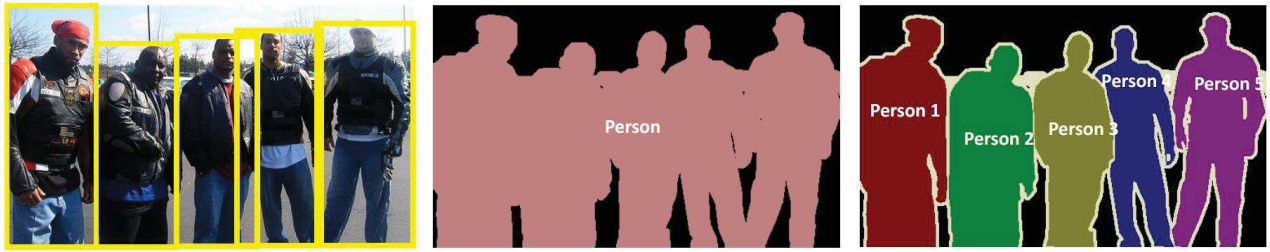
- Node (vertex) for every pixel
- Edge between pairs of pixels, (p,q)
- Affinity weight w_{pq} for each edge
  - w_{pq} measures similarity
  - Similarity is inversely proportional to difference (in color and position…)

- **Mathematical Morphology (watershed, etc…)**
- **Energy minimization (Conditional Random Fields)**
- **Deep-Learning**

What is SEMANTIC Image Segmentation?

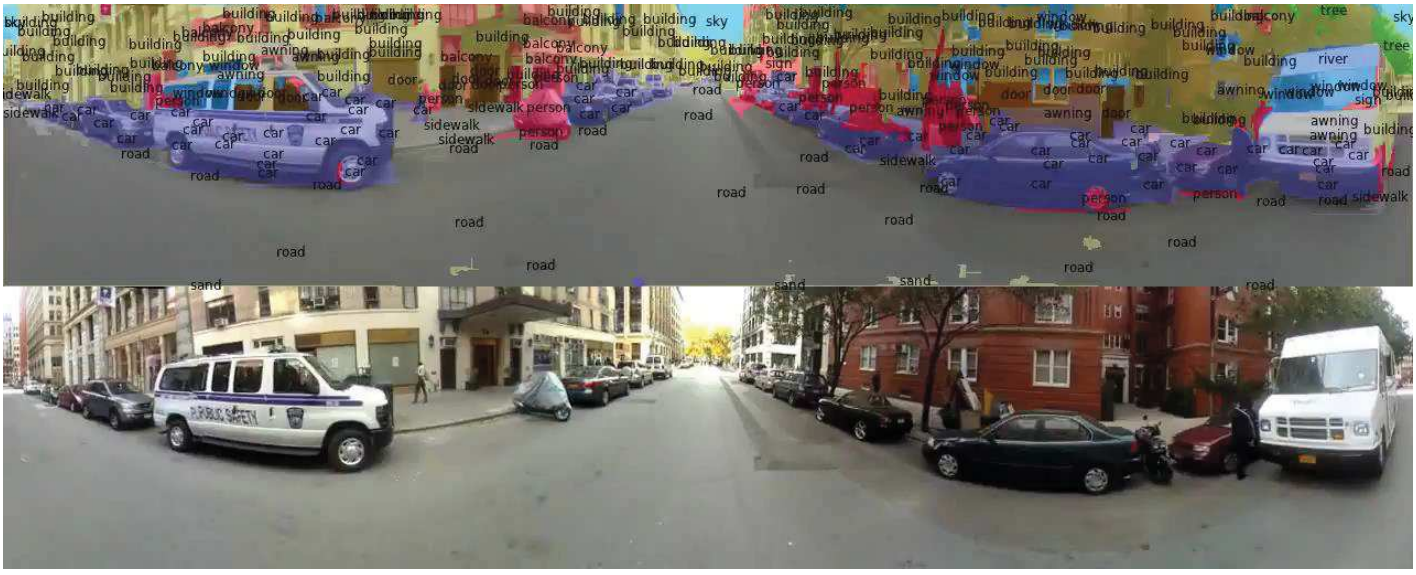


SEMANTIC segmentation:
« go together » = same « type of object »
≠ from just grouping pixels with similar colors or texture



Objects detection | **Semantic Segmentation** | Instance Segmentation

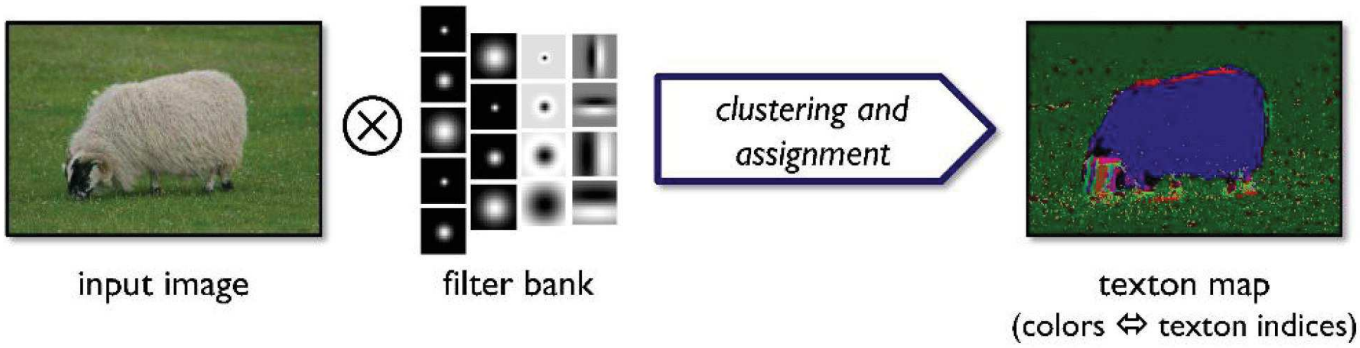
Video example of semantic segmentation *with category labels*



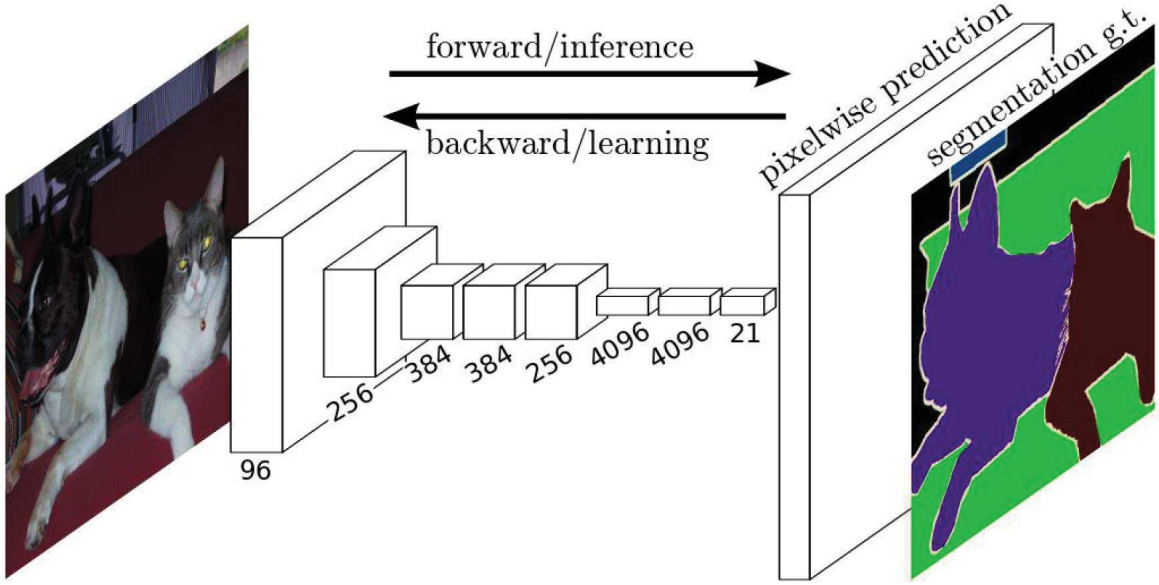
[C. Farabet, C. Couprie, L. Najman & Yann LeCun:
Learning Hierarchical Features for Scene Labeling,
IEEE Trans. PAMI, Aug.2013.

Semantic segmentation BEFORE Deep-Learning

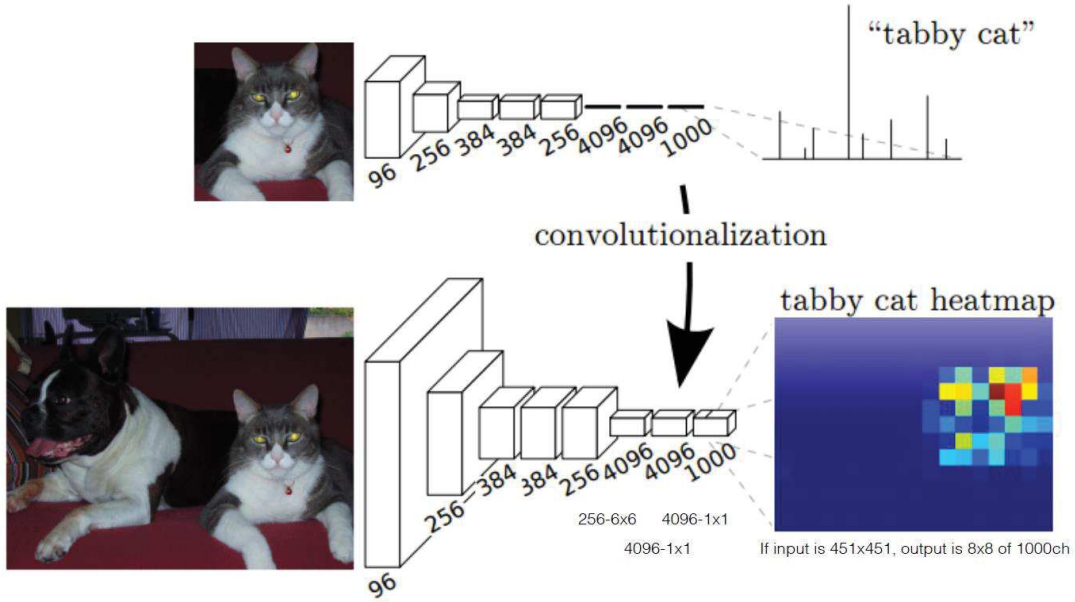
- Relying on Conditional Random Field (CRF)
- Operating on pixels or superpixels
- Interactions between label assignments



Deep-Learning approach for semantic segmentation

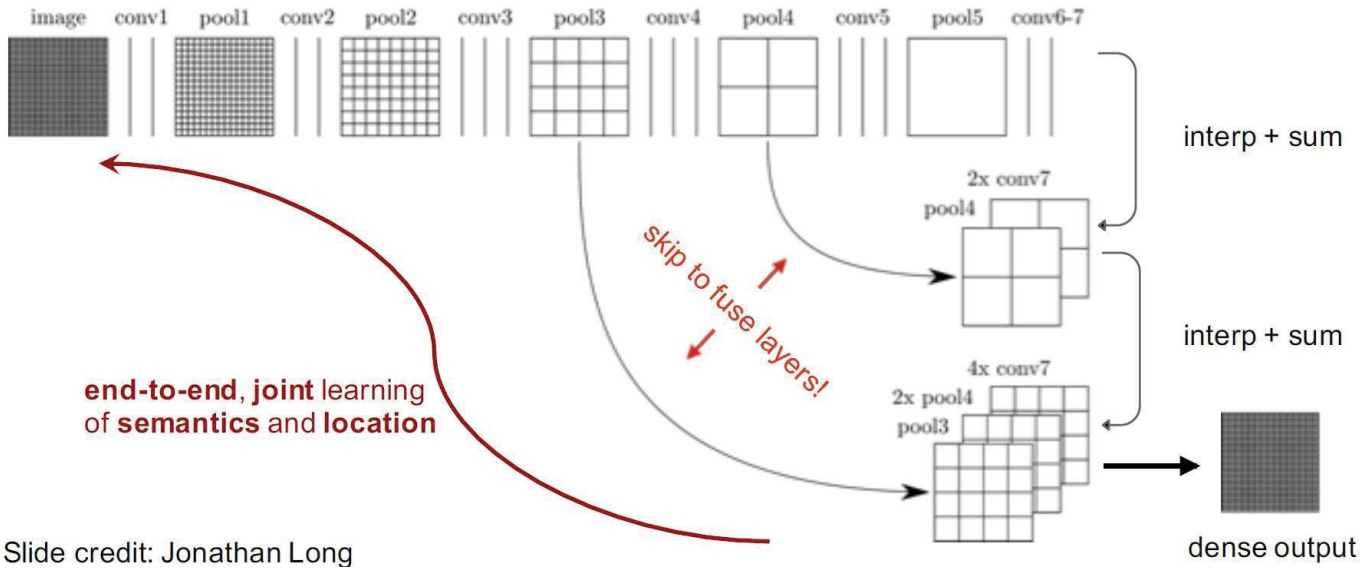


Fully Convolutional Network (FCN)



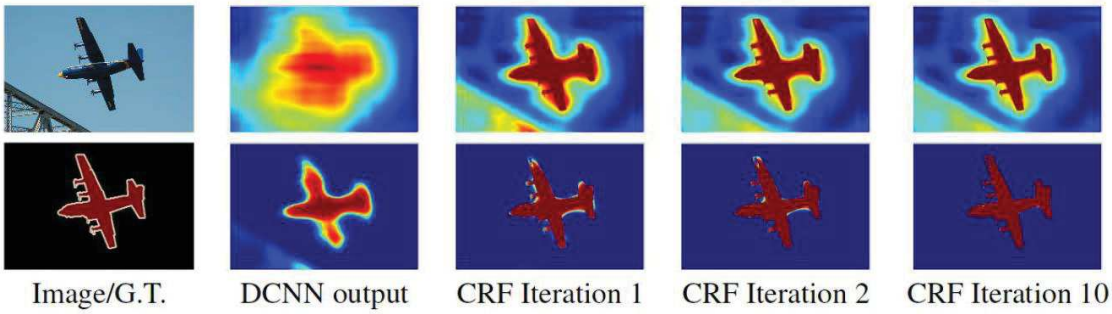
« Fully Convolutional Networks for Semantic Segmentation », Evan Shelhamer, Jonathan Long, and Trevor Darrell, *[Berkeley, 2015]*

FCN principle

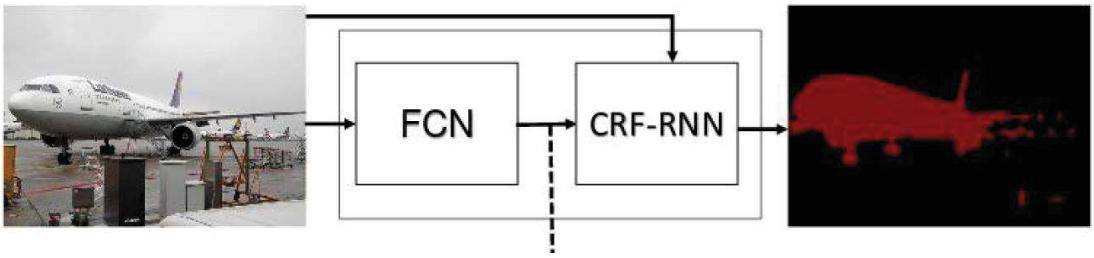


Trick = some connections skipping directly to « fuse layers »

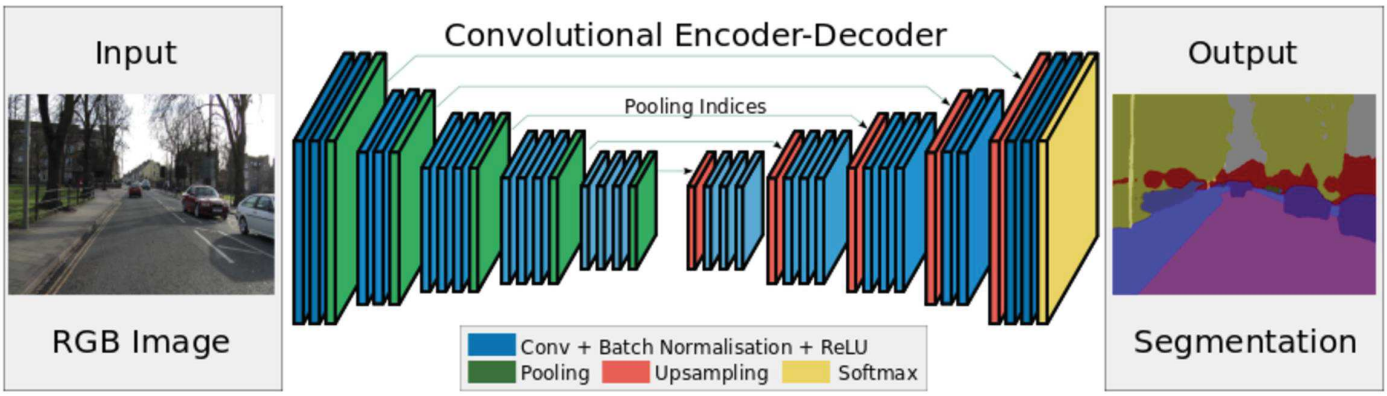
FCN + CRF

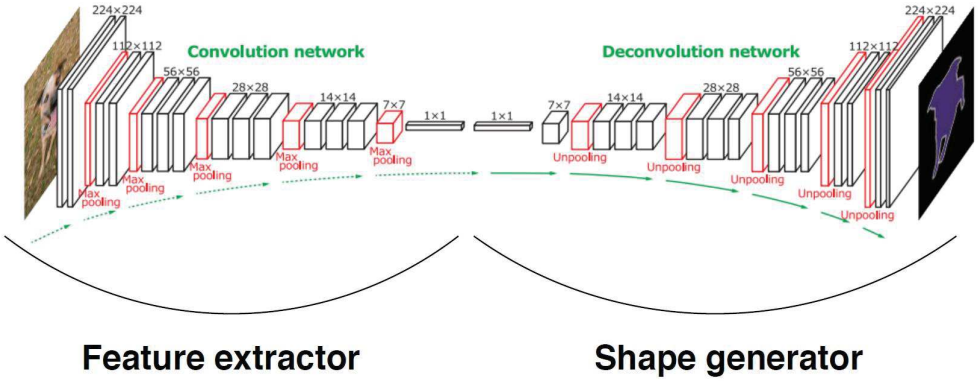


Output from FCN rather blurry and inaccurate, but can be improved by CRF post-processing

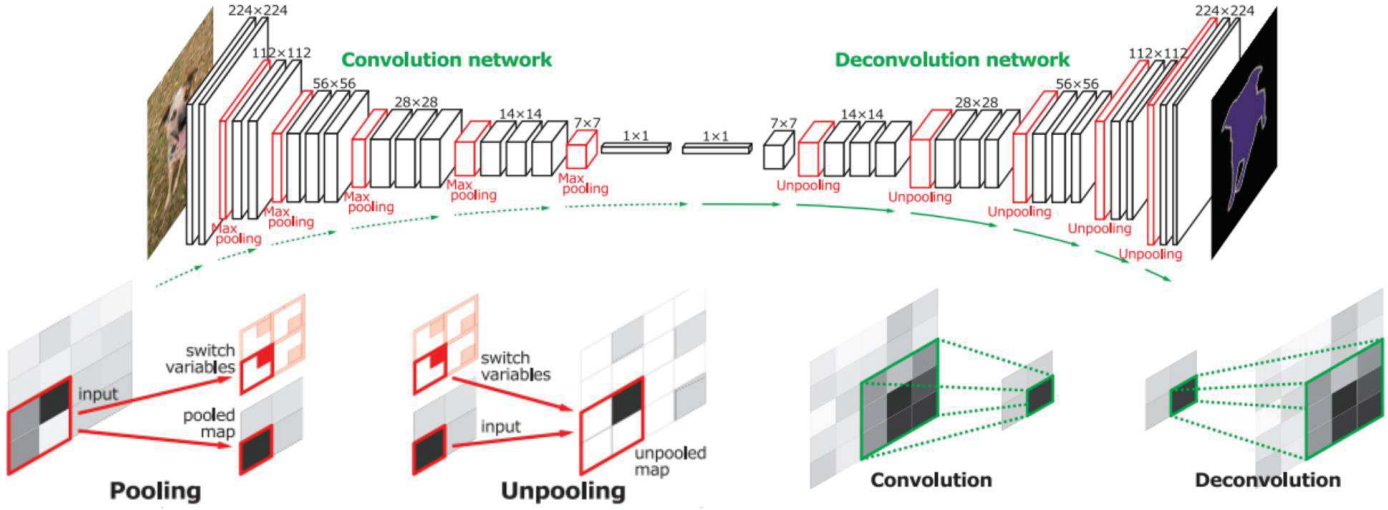


Convolutional Encoder-Decoder

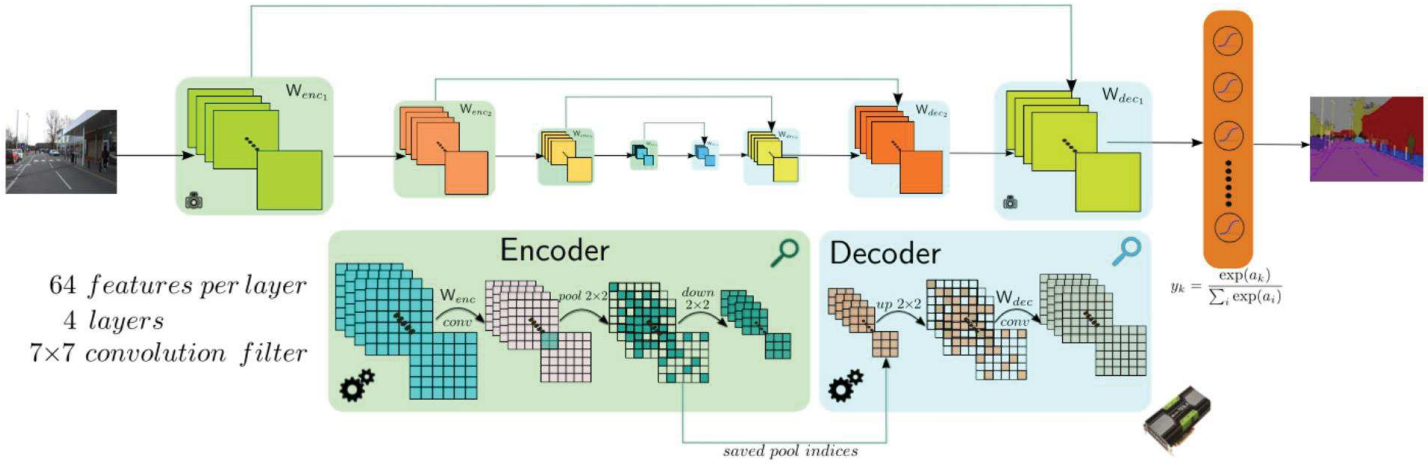




Deconvolution??

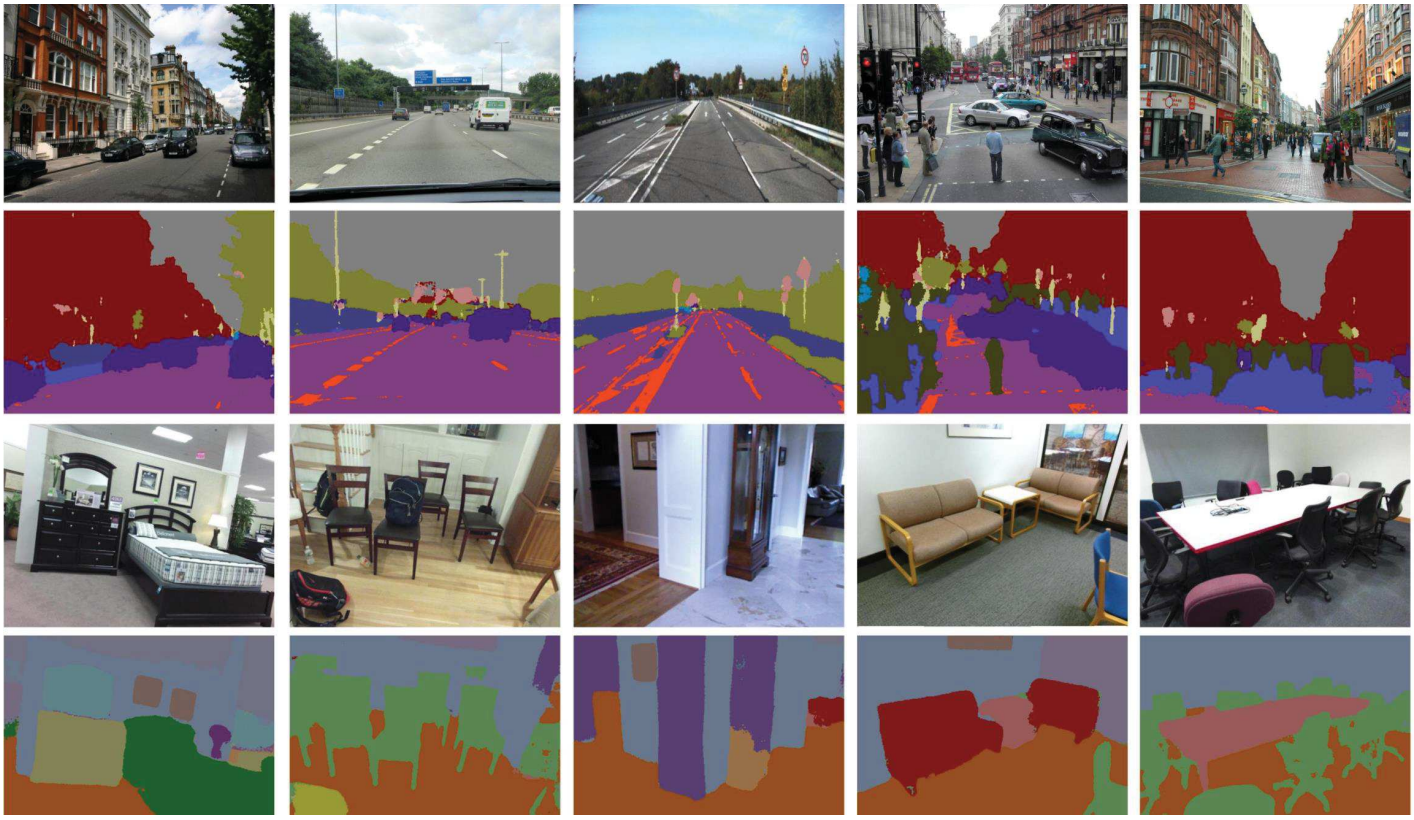


SegNet

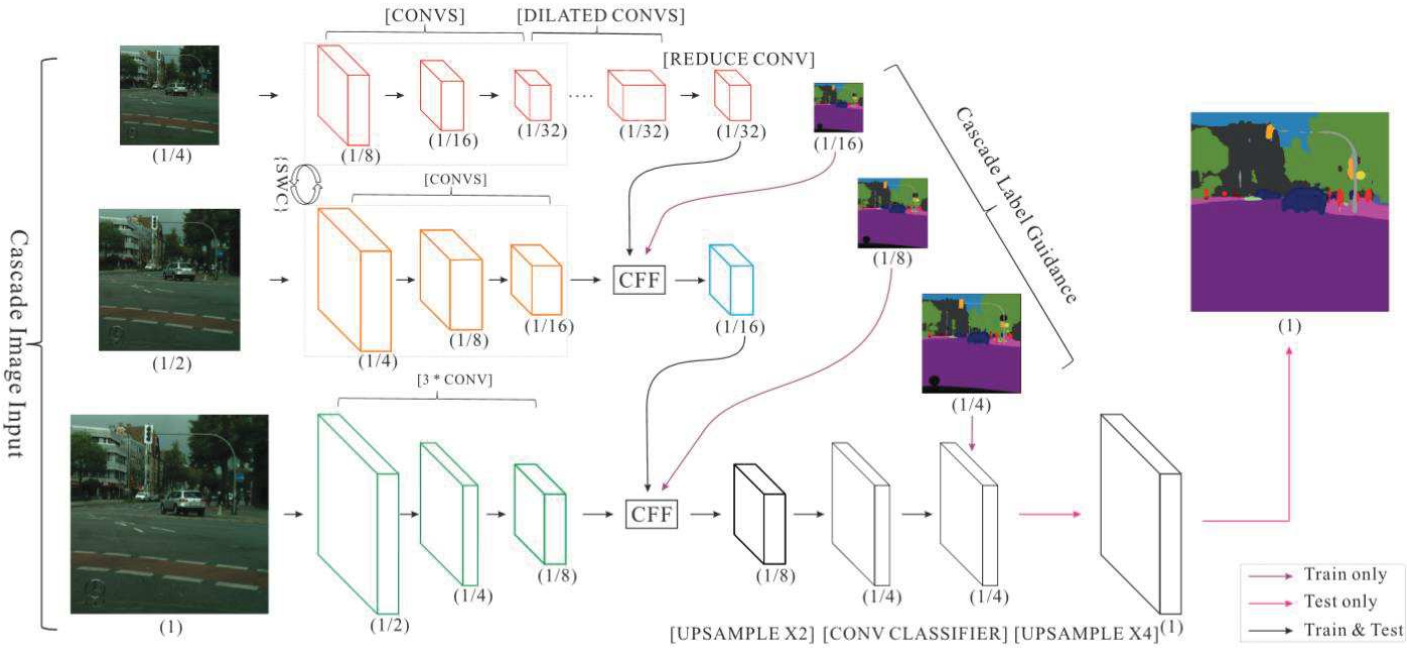


"SegNet: A Deep Convolutional Encoder-Decoder Architecture for ImageSegmentation", Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla *[Cambridge (UK), 2015]*

SegNet example results



ICnet



« ICNet for Real-Time Semantic Segmentation on High-Resolution Images »,
Zhao, Hengshuang & Qi, Xiaojuan & Shen, Xiaoyong & Shi, Jianping & Jia, Jiaya.
Chinese University of Hong-Kong (2017).

And many other competitors!

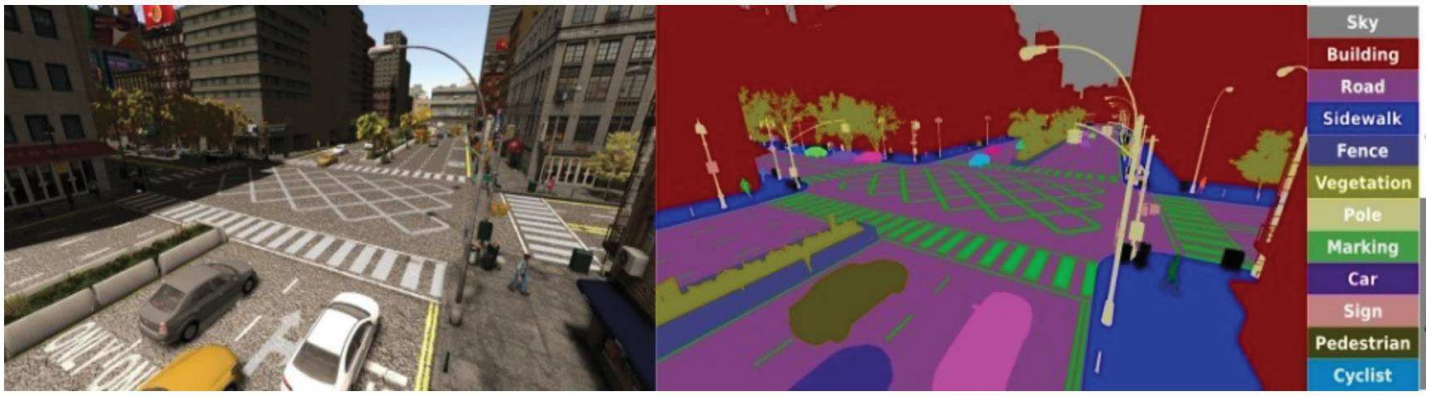
- 2015: U-Net (Keras) - https://github.com/zhixuhao/unet
- RefineNet (2016)
- DeepLab (Caffe) - https://github.com/Robotertechnik/Deep-Lab
- DeepLabv3 (Tensorflow) - https://github.com/NanqingD/DeepLabV3-Tensorflow

Outline

- Recalls on Convolutional Neural Networks (CNN or ConvNets) and Deep-Learning
- Transfer Learning
- Beyond Image Classification: DETECTION OF OBJECTS
- Instance segmentation with DeepLearning
- DL for Human pose inference and depth estimation
- Semantic segmentation with DeepLearning
- Interest and use of simulations / synthetic videos

Synthetic images

More and more realistic



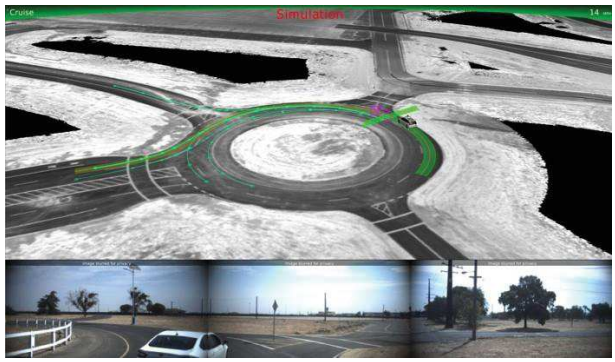
Example from SYNTHIA
(http://synthia-dataset.net)

Interest of synthetic images for Machine-Learning in IV applications

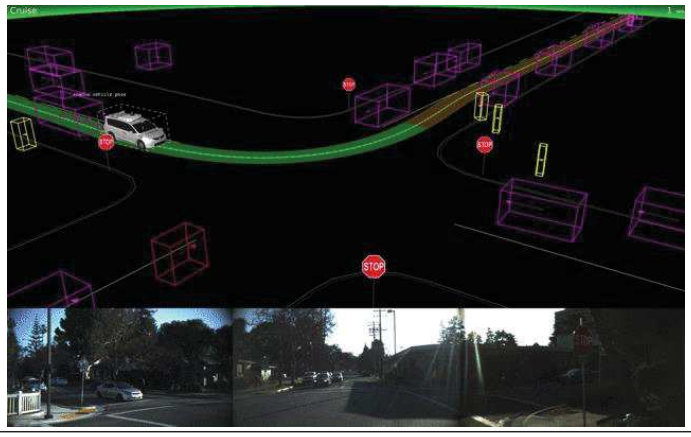
- **Possible to generate as many as needed at nearly no cost (in particular compared to recording while driving)**
- **Easy to generate controlled variability in environment, luminosity conditions, scenarii, etc + also images « dangerous situations »**
- **NO NEED FOR MANUAL LABELLING: ground truth (ie target value) for classifiers, localizers, and semantic segmentation provided automatically**

Simulators dedicated to Autonomous Vehicles

Scenario-buiding with CarCraft by Google/Waymo



Simulation of a virtual scenario in XView by Google/Waymo



CARLA open-source urban driving simulator

- **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 éxecution	− −	+	+	+
Multi-agent	− −	−	−	++

→ Choice of CARLA

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

CARLA simulator

Synthetic images use in ML/DL for IV

- Initial training of a classifier / segmenter / controller only on simulated images / videos / scenarios
- Possible to then adaptation to real-world by fine-tuning on REAL images/video datasets
- Cheaper / more extensive testing than on real-world videos

- REINFORCEMENT LEARNING in simulation !

Examples of autonomous driving obtained by DRL in CARLA

Town02: Single Lane, EU

Weather: Heavy rain

Traffic Light: Red

Network input

Current Order: Left

Current Speed: 1.8 km/h

Work by my PhD student Marin Toromanoff (Valeo/MINES).

Ranked 1st (vision-only track) on CARLA "*Autonomous Driving challenge*" !!