JURSE 2019 :
Urban object classification with 3D deep learning

Younes ZEGAOUI [1,2]

Marc CHAUMONT [2,3]
Gérard SUBSOL [2]
Philippe BORIANNE [4]
Laurent DERUELLE [1]

1 : Berger-Levrault, Toulouse, France
2 : LIRMM, Univ. Montpellier, CNRS, France
3 : Univ. Nîmes, France
4 : AMAP, Univ. Montpellier, CIRAD, CNRS, INRIA, IRD, Montpellier, France

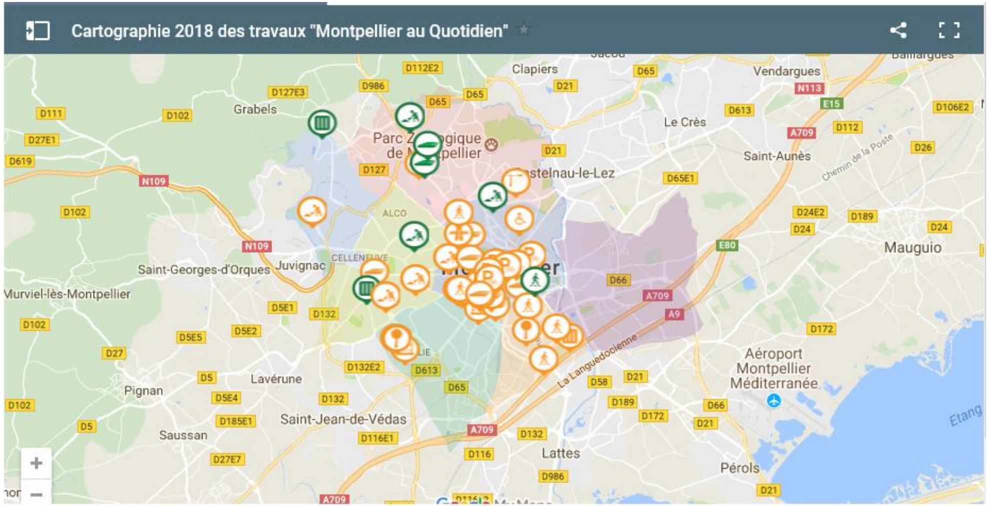
Plan :

- Context
- Deep-Learning in 3D
- The PointNet network
- Experiments
- Discussion and perspectives

Context

- Too many objects for manual approaches

- Ex Montpellier :
  - 282 000 inhabitants (2015)
  - >1000 km roads
  - 2450 streets-> 5 to **8 000 signs**
  - ~**240km de cables** (60 km tramway)
  - **~ 40 000 trees**



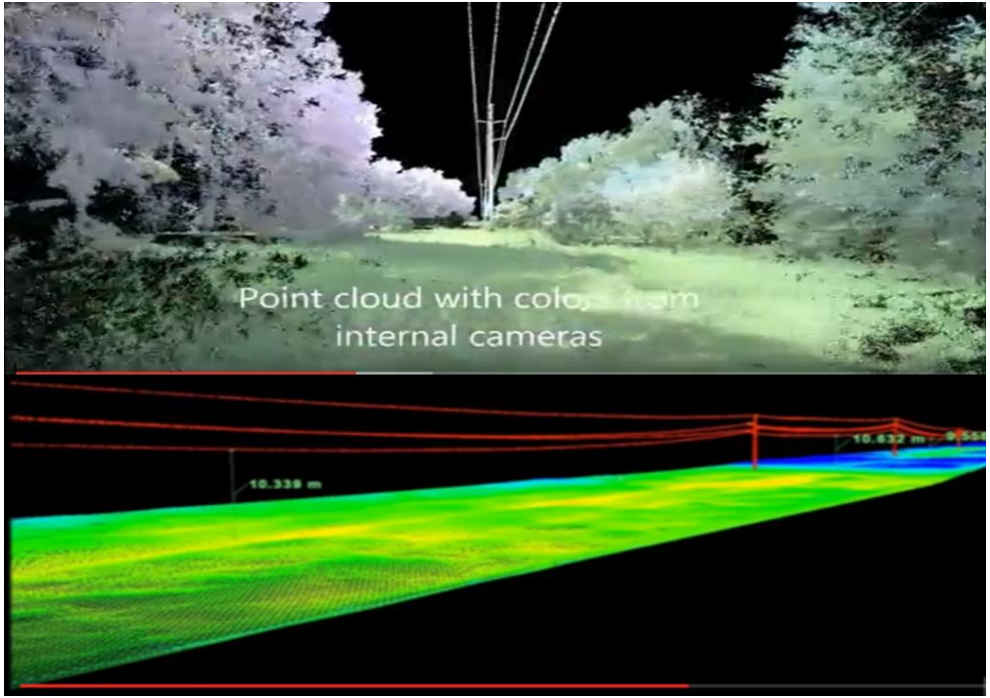
Légende couleurs : en cours de réalisation - Réalisé

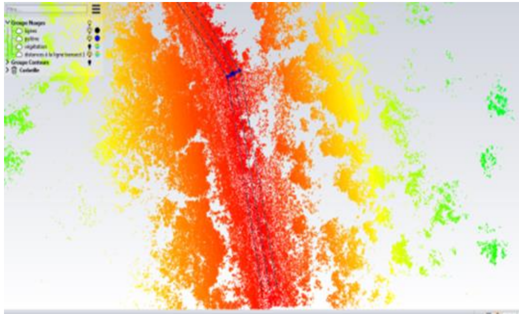
LiDAR : 3D data

Use case by EDF-RTE :

- High voltage line monitoring
- **interaction with the envrionment**

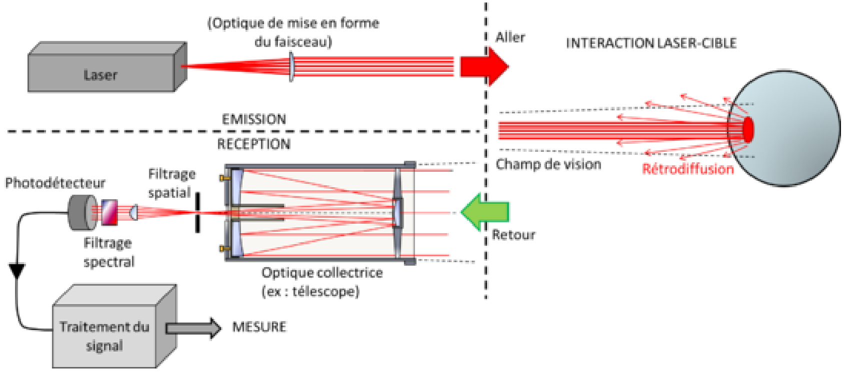
→ actions planning (cutting trees)





LiDAR : how it works

- LIght Detection And Ranging
- Based on **écholocation**
- **Laser** pulse
- Scan an entire scene
- Calculating distances, 3D **cartography**, etc…
- Differents types : dynamic, static, UAV, ...



RIEGL vz 400

Velodyne

3D data

- Dynamic LiDAR acquisition to scan bigger scene such as city district

- Output data :
    - 3D Point cloud (x, y, z)
    - Color images
    - Geolocated

- Constraints : data size (number of points), occlusion.

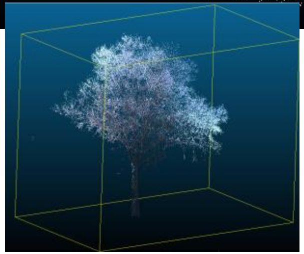
LiDAR à balayage Sick

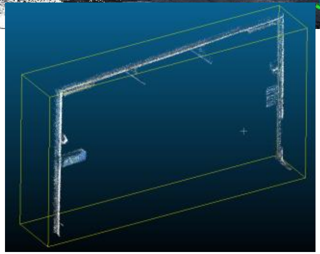
sac à dos Leica

LiDAR : data

Implications :

- Need of **automatic processing**
- **Pattern recognition in 3D** and shape analysis
- **Validating the models**





3D dataset

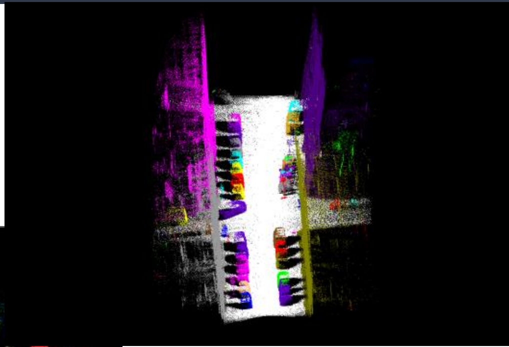
Few public datasets :

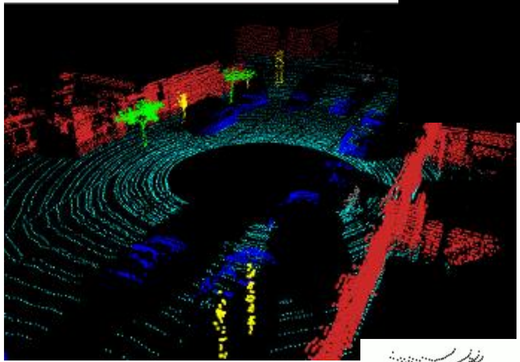
- Paris rue Madame dataset : http://cmm.ensmp.fr/~serna/rueMadameDataset.html

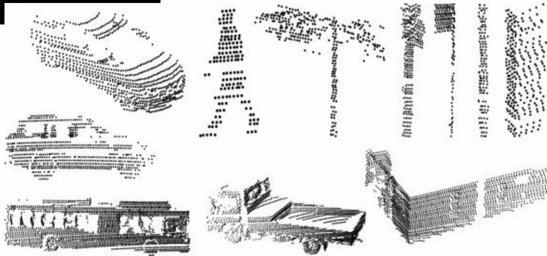
- Kevin Lai dataset : https://sites.google.com/site/kevinlai726/datasets

- Sydney urban dataset : http://www-personal.acfr.usyd.edu.au/a.quadros/objects4.html

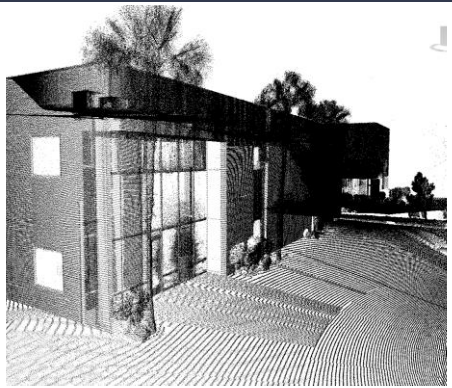
- Our own dataset :







Vocabulary

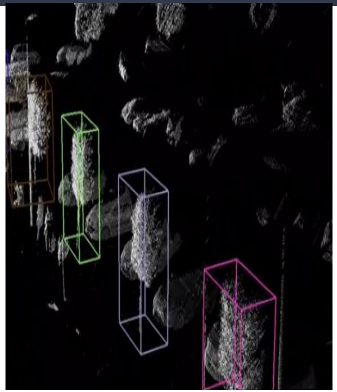


Detection :

Is there a tree?
Yes/No

Classification :

Classe of the object ?
Tree/Car/Pole...



Localisation/instance segmentation :

Where are the objects?
Bounding boxes 3D + Classe

Généralising 2D CNN to 3D data

- A point cloud is nothing more than a vertex list

- Datas are not structured as opposed to 2D images:
    - No direct neighborhood
    - No particular order in the vertex list

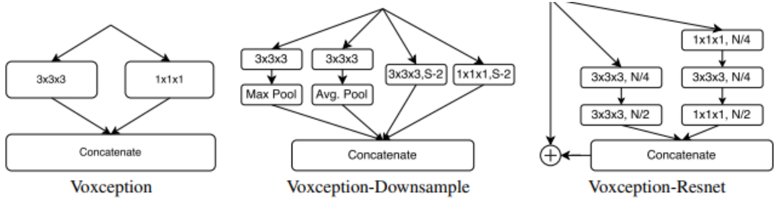
➔ 2D convolutions cannot be applied

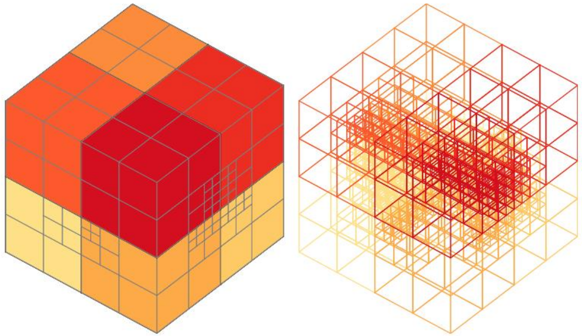
Deep-learning in 3D Voxelisation

- Voxelisation of the point cloud

- Based on classical 2D CNN architecture : ResNet, VGG...

- Memory size (VRAM), sparsity of the data





VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition

D. Maturana, S. Scherer (IROS 2015)

OctNet: Learning Deep 3D Representations at High Resolutions

G. Riegler, A. Osman, A. Geiger (CVPR 2017)

Generative and Discriminative Voxel Modeling with Convolutional Neural Networks

A. Brock, T. Lim, J.M. Ritchie, N. Weston (NIPS-WS 2016)

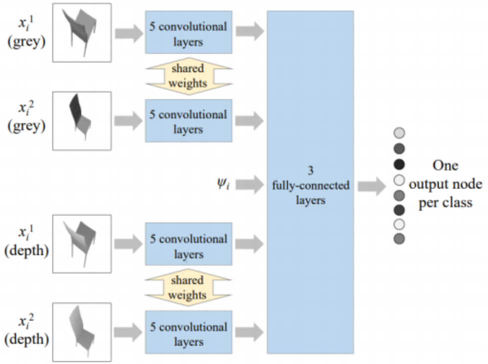
Deep-learning en 3D
Multi-views

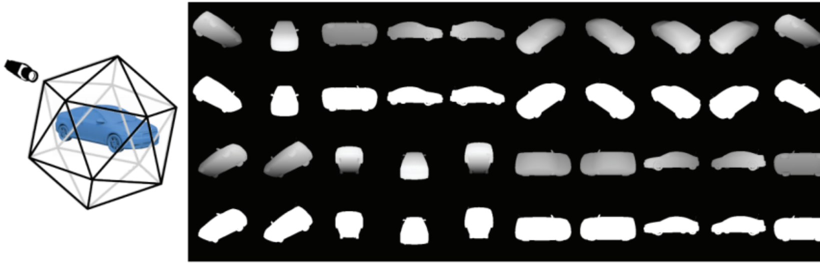
- Virtual camera with different angles.

- For each camera angle synthesized:
  - Color images/greyscale images
  - Depth map
  - Silhouette

- Depend of the choices of the angles

- Works well with CAD models but not really fit for LiDAR data





Volumetric and Multi-View CNNs for Object Classification on 3D Data
C. R. Qi, H. Su, M. Niessner, A. Dai, M. Yan, L. J. Guibas (CVPR 2016)

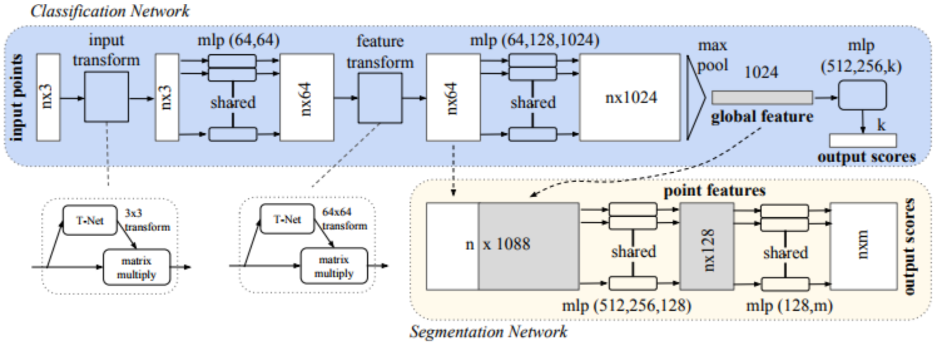
Pairwise Decomposition of Image Sequences for Active Multi-View Recognition
E. Johns, S. Leutenegger, A. J. Davison (CVPR 2016)

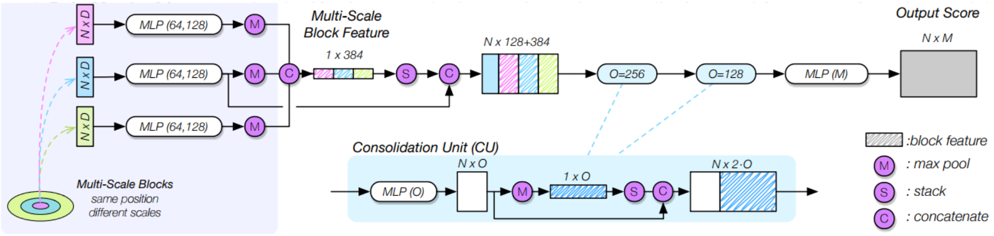
Deep-learning en 3D
Directly on points coordinates

- Training directly on points cloud

- Learning a transformation matrix

- Invariant to the point order:
    - Features extracted for each point
    - Max pooling on the points to extract a global feature vector for each cloud



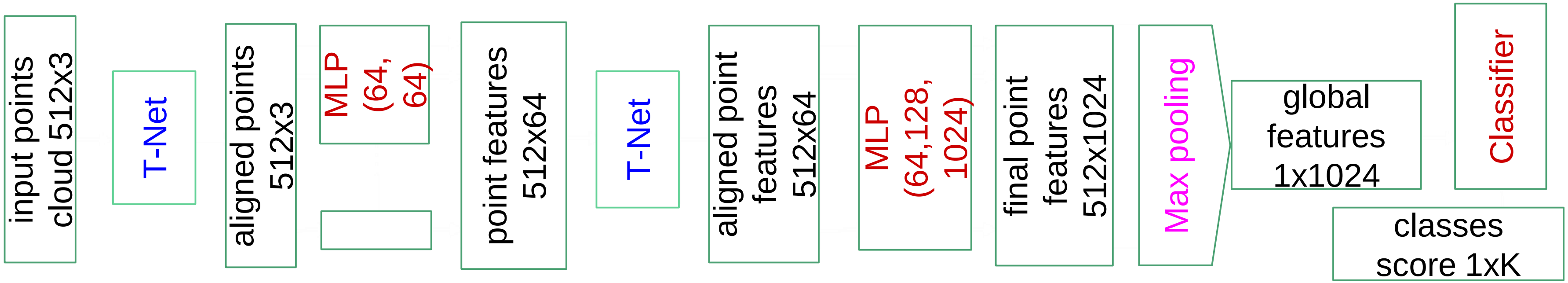


PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation
C. R. Qi, H. Su, K. Mo, L. J. Guibas CVPR 2017

Exploring Spatial Context for 3D Semantic Segmentation of Point Clouds
F. Engelmann, T. Kontogianni, A. Hermans, B. Leibe (ICCV workshop 2017)

PointNet Network

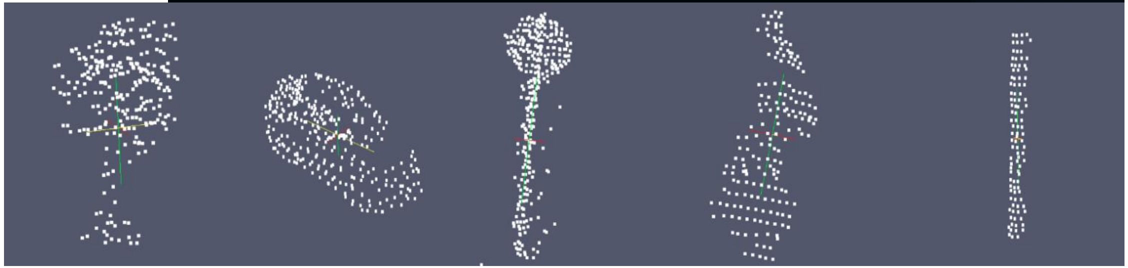
- No intermediary data structure
- Very good performances considering how shallow the network is
- The vertex lists are the input of the network
- T-Net for the transformation matrix
- Tenserflow implementation: https://github.com/charlesq34/pointnet/



Urban ojects classification

- One points cloud = one object
- Using the PointNet network
- Subsampling to 512 points

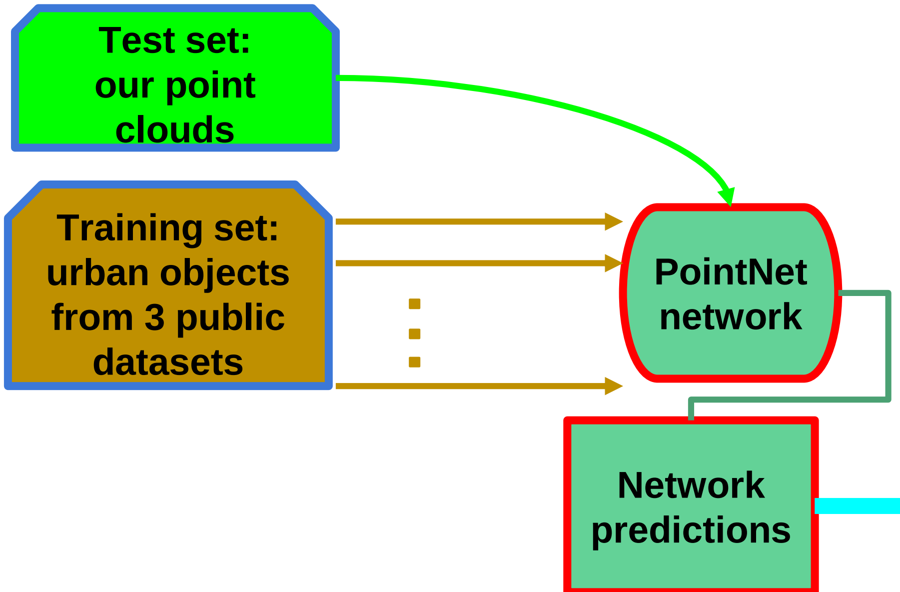
- Urban objects split in 5 classes :
  - Tree
  - Car
  - Traffic sign/light
  - Pole
  - Pedestrian



Tree Car Traffic sign Pedestrian Pole

Classification experiments

- Training on public dataset(727 objects)
- Test on our dataset(174 objects)



		Ground truth				
		tree (75)	*car (39)*	*traffic sign/light (8)*	*pole (23)*	*person (15)*
Classification	*"tree"*	69	2	0	8	0
	"car"	1	33	0	0	0
	"traffic sign/light"	4	0	4	12	2
	"pole"	0	0	3	1	0
	"person"	1	0	1	0	12
	"building"	0	0	0	2	0
	"noise"	0	4	0	0	1
	F measure	0.896	0.904	0.267	0.074	0.828

Classification experiment:
Bigger training set

- Adding another dataset to our training set: http://npm3d.fr/paris-lille-3d
- New training set: 1668 objets in total

		Vérité terrain (annotations)				
		arbre (75)	*voiture (39)*	*signalisation (8)*	*poteau (23)*	*piéton (15)*
Classification	*"arbre"*	70	0	0	2	1
	"voiture"	2	39	0	0	1
	"signalisation"	2	0	7	10	2
	"poteau"	0	0	1	11	0
	"piéton"	0	0	1	0	11
	"bâtiment"	1	0	0	0	0
	"bruit"	0	0	0	0	0
	F mesure	0.946	1.000	0.467	0.629	0.815

Baseline :

		Vérité terrain (annotations)				
		arbre (75)	*voiture (39)*	*signalisation (8)*	*poteau (23)*	*piéton (15)*
Classification	*"arbre"*	69	2	0	8	0
	"voiture"	1	33	0	0	0
	"signalisation"	4	0	4	12	2
	"poteau"	0	0	3	1	0
	"piéton"	1	0	1	0	12
	"bâtiment"	0	0	0	2	0
	"bruit"	0	4	0	0	1
	F mesure	0.896	0.904	0.267	0.074	0.828

Classification experiment:

Class fusion

- Fusing the ‘traffic sign/light’ and the ‘pole’ classes into one TSLP class
- New training set: 6 instead of 7

		Vérité terrain (annotations)			
		arbre (75)	*voiture (39)*	*TSLP(31)*	*piéton (15)*
Classification	*"arbre"*	69	3	3	0
	"voiture"	1	27	0	0
	"TSLP"	2	0	28	5
	"piéton"	0	0	0	9
	"bâtiment"	1	0	0	0
	"bruit"	2	9	0	1
	F mesure	0.920	0.806	0.849	0.750

Baseline :

		Vérité terrain (annotations)				
		arbre (75)	*voiture (39)*	*signalisation (8)*	*poteau (23)*	*piéton (15)*
Classification	*"arbre"*	69	2	0	8	0
	"voiture"	1	33	0	0	0
	"signalisation"	4	0	4	12	2
	"poteau"	0	0	3	1	0
	"piéton"	1	0	1	0	12
	"bâtiment"	0	0	0	2	0
	"bruit"	0	4	0	0	1
	F mesure	0.896	0.904	0.267	0.074	0.828

Classification experiment:
Number of points

- Subsampling the clouds to 2048 points instead of 512
- Same training set but with more points per object

Classification	Vérité terrain (annotations) arbre (75)	voiture (39)	signalisation (8)	poteau (23)	piéton (15)
"arbre"	72	3	0	4	0
"voiture"	0	31	0	0	0
"signalisation"	1	0	5	10	2
"poteau"	0	0	1	1	0
"piéton"	2	0	2	0	13
"bâtiment"	0	4	0	2	0
"bruit"	0	1	0	6	0
F mesure	0.935	0.886	0.385	0.080	0.813

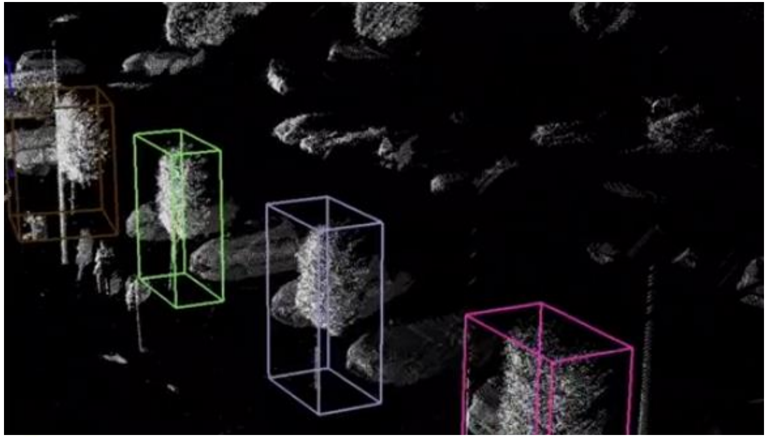
Baseline :

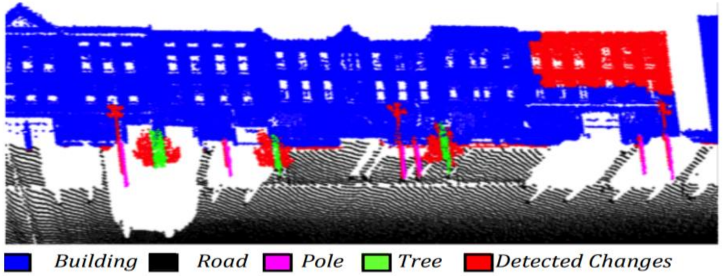
Classification	Vérité terrain (annotations) arbre (75)	voiture (39)	signalisation (8)	poteau (23)	piéton (15)
"arbre"	69	2	0	8	0
"voiture"	1	33	0	0	0
"signalisation"	4	0	4	12	2
"poteau"	0	0	3	1	0
"piéton"	1	0	1	0	12
"bâtiment"	0	0	0	2	0
"bruit"	0	4	0	0	1
F mesure	0.896	0.904	0.267	0.074	0.828

Conclusion and perspectives

- Encouraging results considering small size of dataset
- Further experiments shows us how we can improve the results even when using only vanilla PointNet:

- We want to collect more datas and from more diverse area.
- We consider the performances to be sufficient in order to go from classification to localization.





Kamal Aijazi, Ahmad & Checchin, Paul & Trassoudaine, Laurent. (2013). Automatic change detection and incremental updating for accurate 3D urban cartography. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. 77-84. 10.1109/ITSC.2013.6728214.