Machine Learning Models for Scene Understanding

EOP-Φ 31/03/2020

Bertrand Le Saux

bertrand.le.saux@esa.int

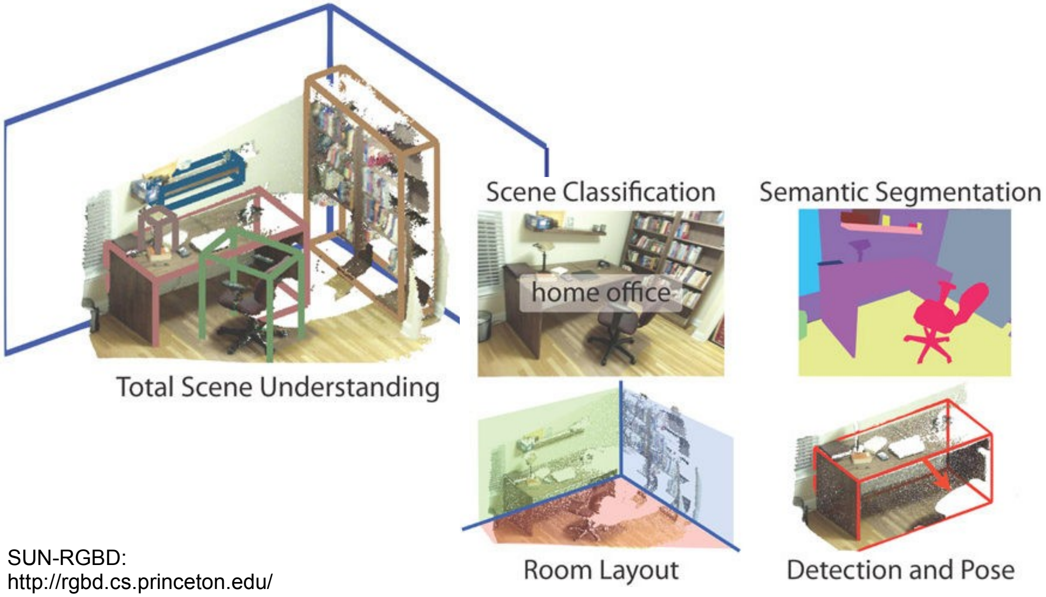
Scene understanding ?

Machine Learning Models for Scene Understanding [T2]

Scene understanding ?

General scene understanding:

object detection, semantic labeling, 3D structure, denoising, motion and action recognition, captioning, etc.



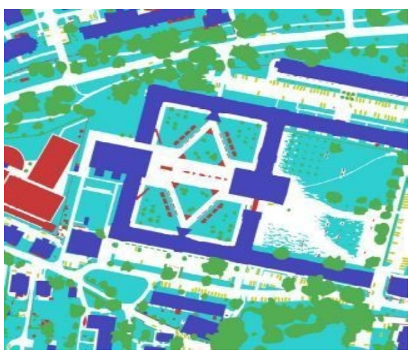
SUN-RGBD:
http://rgbd.cs.princeton.edu/

Varcity project, ETHZ
http://www.varcity.eu/

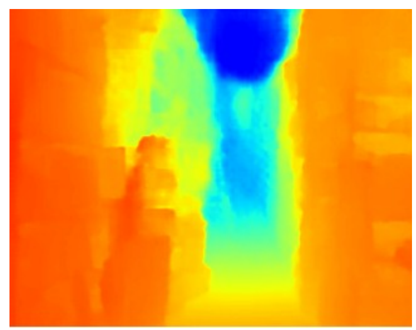
→ *Build functions able to estimate semantics and geometry of a scene*

Scene understanding ?

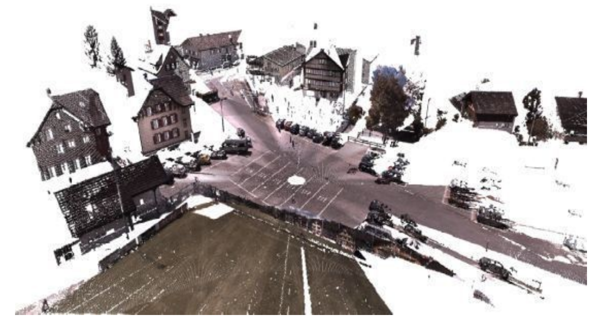
Labeling pixels

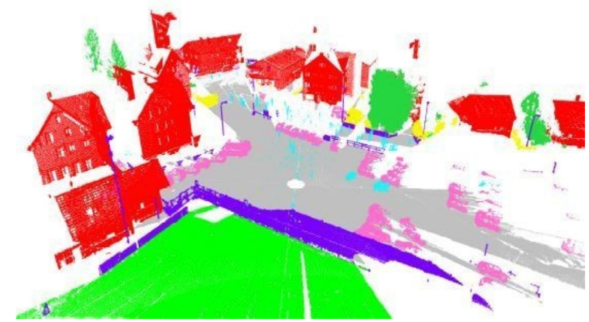


Estimating depth



Labeling 3D points





→ *Build functions able to estimate semantics and geometry of a scene*

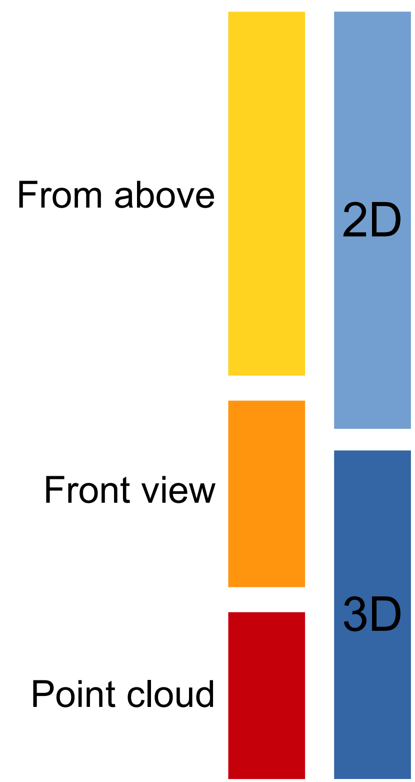
Outline

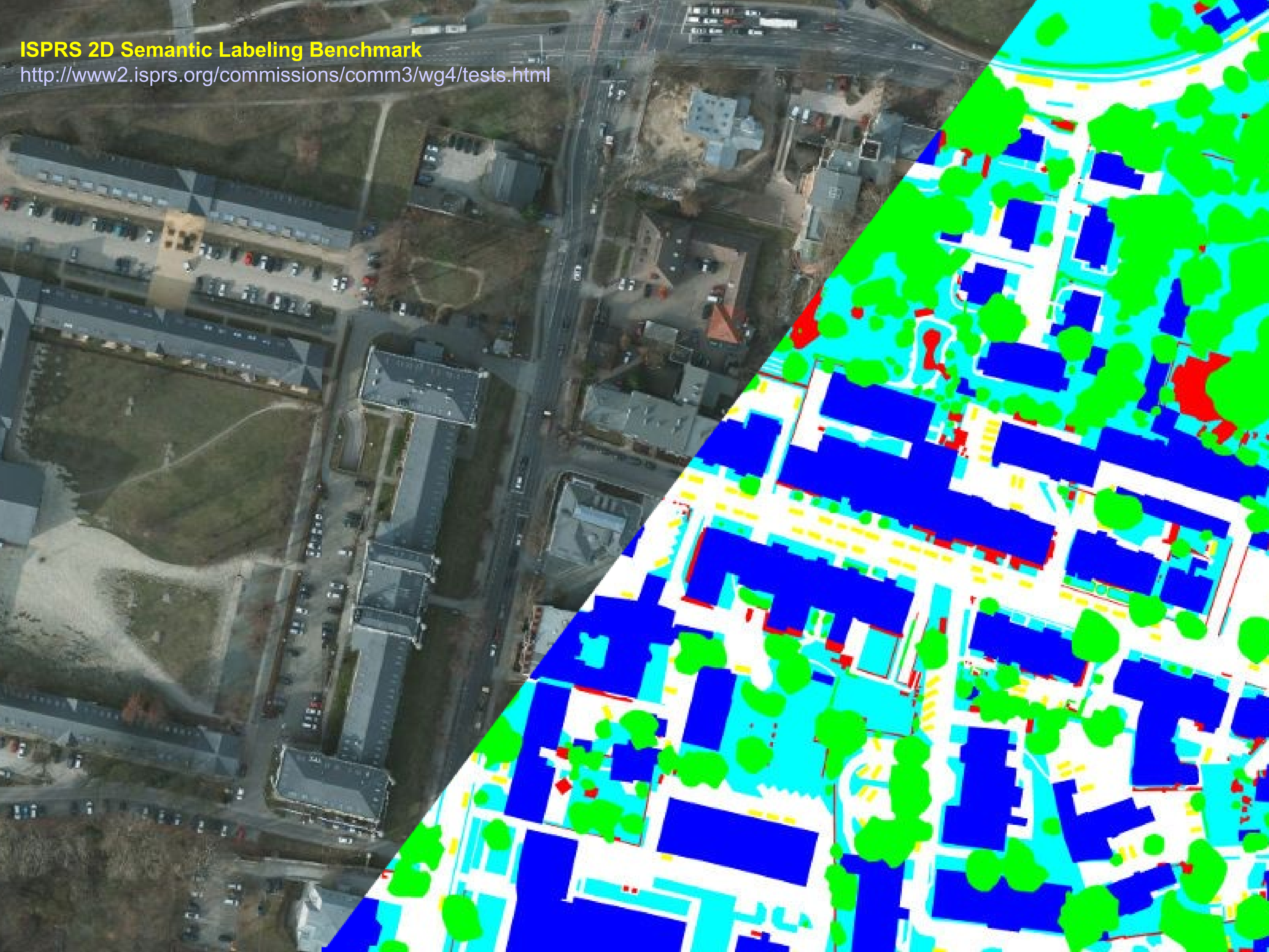
In this talk:

- Why using conv. nets for semantic mapping ?
- Dense conv. nets for semantic segmentation
- Fusion of heterogeneous data
- Joint learning with open-source cartography
- Multi-temporal analysis
- Hyperspectral data classification

- Distance-transform regression for semantic labeling
- Losses for single-image depth prediction
- Robotic exploration

- 3D point-cloud semantic mapping with SnapNet
- Urban mapping

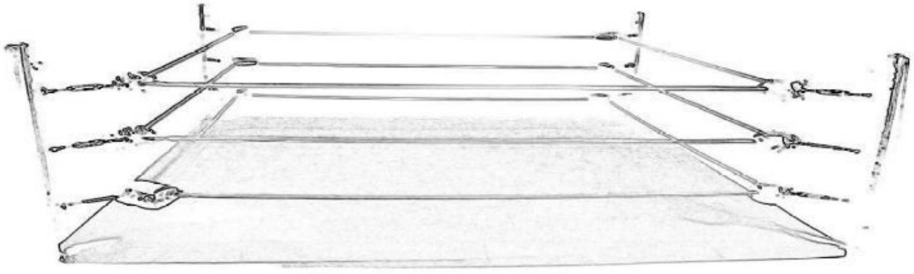



ISPRS 2D Semantic Labeling Benchmark
http://www2.isprs.org/commissions/comm3/wg4/tests.html

Benchmarking classification

(or : why using conv. nets for semantic mapping ?)

(with Adrien Lagrange, Anne Beaupère, Alexandre Boulch, Adrien Chan-Hon-Tong, Stéphane Herbin, Hicham Randrianarivo, Marin Ferecatu)



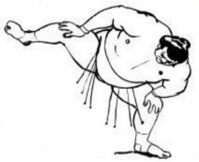
Classification algorithms in competition

Expert rules and indices (NDVI, ...)

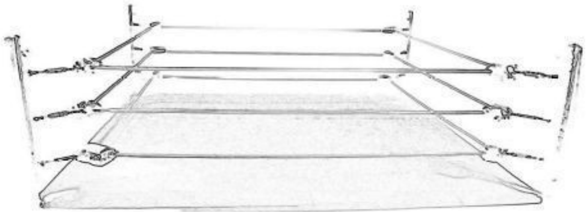
Raw data + SVM

Handcrafted features + SVM (Superpixels, HOGs, normals...)

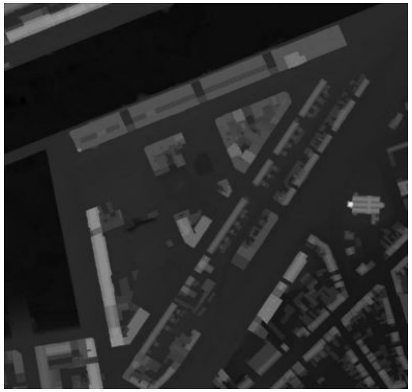
Object-based Detectors (DPMs)

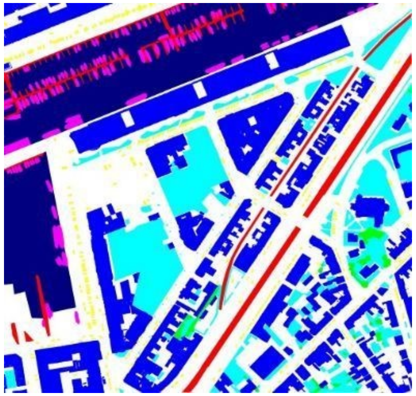


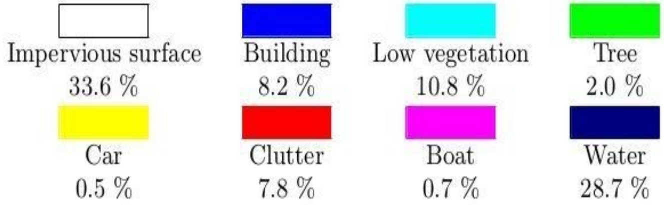
Conv. Neural Nets



Data Fusion Contest 2015 : VHR images, DSM, 8-class semantic reference

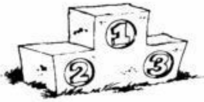






Benchmark today available on: http://dase.grss-ieee.org/

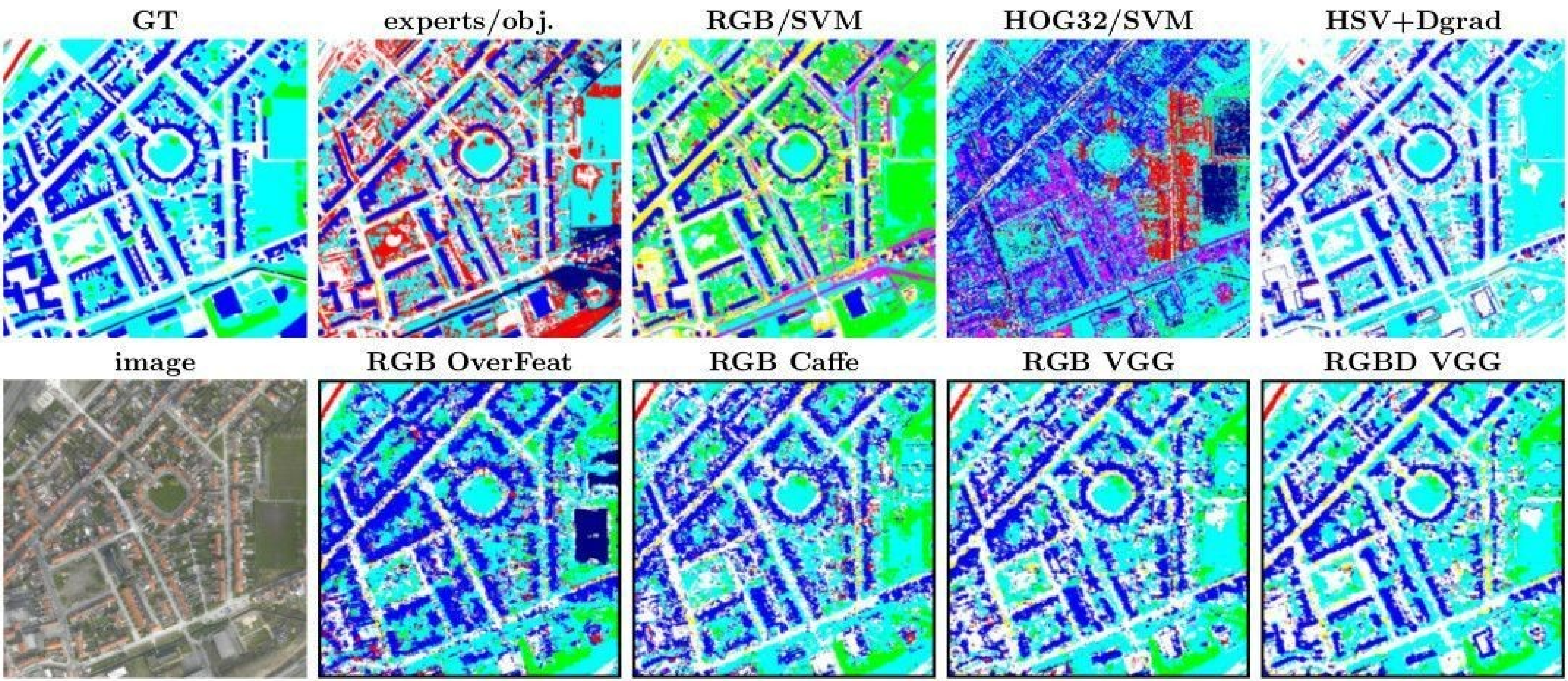
Results : classification measures



3D	*Algorithm*	Imp. surf.	Build.	Low veg.	Tree	Car	Clutter	Boat	Water	Overall acc. %	Cohen κ
★	Expert	58.97	63.87	74.55					92.39	∅	∅
	RGB/SVM	53.89	53.53	50.32	32.97	24.02	13.75	12.12	98.52	60.77	0.52
★	RGBd/SVM	14.51	67.79	38.03	27.43	7.15	1.12	14.58	98.45	50.76	0.41
★	RGBdI/SVM	60.86	69.01	57.12	38.12	11.59	20.49	15.04	94.42	63.83	0.56
	HOG32/SVM	28.94	43.17	48.77	27.32	30.24	17.39	12.61	88.02	52.45	0.41
	HOG16/SVM	39.52	38.45	35.65	29.99	21.93	16.13	13.52	80.02	49.4	0.36
	HSV/SVM	71.60	46.97	68.38	0.12	0.00	13.71	0.00	92.14	70.16	0.60
★	HSVDGr/SVM	73.30	70.85	68.75	0.17	0.00	17.11	0.00	92.37	73.60	0.65
	SOM							51.45		∅	∅
	DtMM					48.46				∅	∅
	RGB OverFeat/SVM	55.86	63.34	59.48	64.44	36.03	28.31	41.51	92.07	67.97	0.59
	RGB Caffe/SVM	62.32	62.66	63.23	60.84	31.34	32.49	46.57	95.61	71.06	0.63
	RGB VGG/SVM	63.18	64.66	63.60	66.98	31.46	43.68	51.92	95.93	72.36	0.64
★	RGBd VGG/SVM	66.02	74.26	65.04	66.94	32.04	44.96	50.61	96.31	74.77	0.67
★	RGBd^{+} VGG/SVM	67.66	72.70	68.38	78.77	33.92	45.6	56.10	96.50	76.56	0.70
★	RGBd^{+} trained AlexNet	79.10	75.60	78.00	79.50	50.80	63.40	44.80	98.20	83.32	0.78

Processing of Extremely High-Resolution LiDAR and RGB Data: Outcome of the 2015 IEEE GRSS Data Fusion Contest–Part A: 2-D Contest, Campos-Taberner et al., **JSTARS'2016**

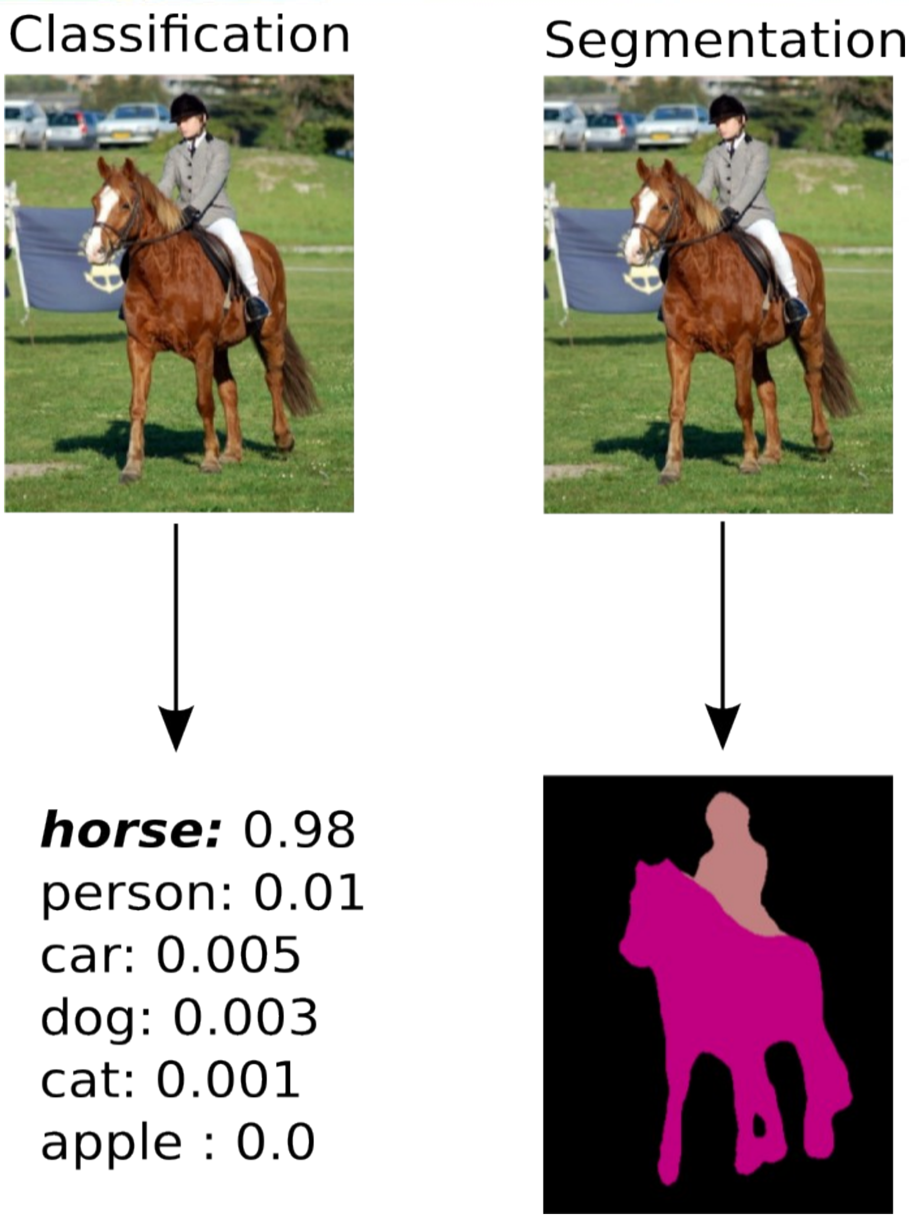
Results : classification map #6



Dense conv. nets for semantic segmentation

(with Nicolas Audebert and Sébastien Lefèvre)

Semantic Segmentation



Classification: 1 image → 1 label

Segmentation: 1 pixel → 1 label

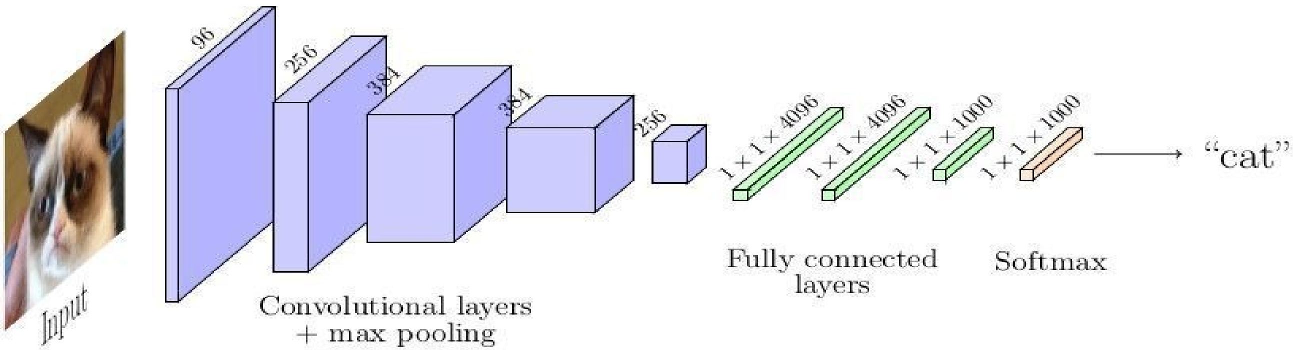
Image = structured pixel ensemble

Network architecture :

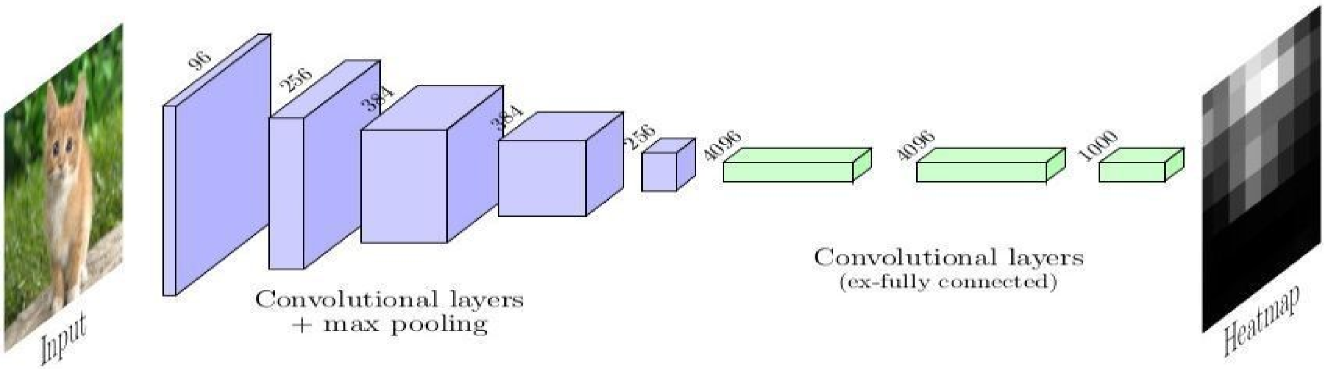
Fully Convolutional Network

[Long et al. 2015]

Fully convolutional networks

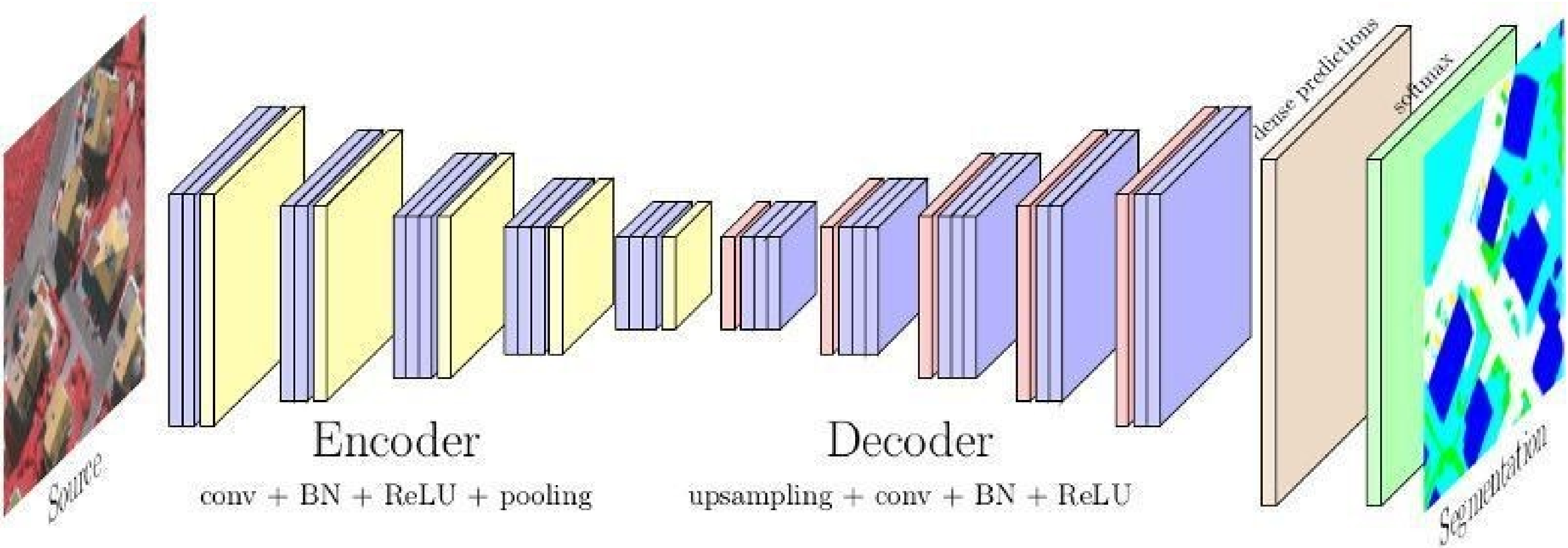


Standard AlexNet



Fully-convolutional AlexNet

Semantic Segmentation



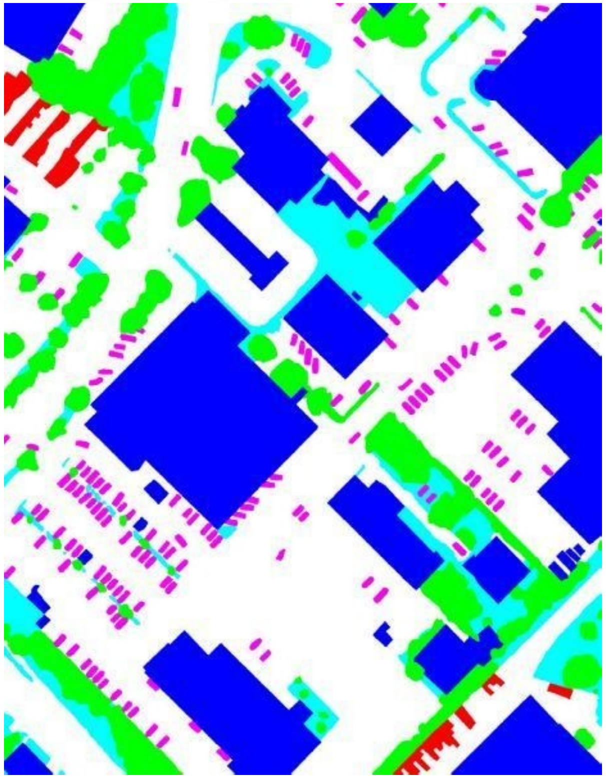
SegNet : A deep convolutional **Encoder-Decoder** architecture for Image Segmentation. Badrinarayanan, V., Kendall, A., Cipolla, R., *TPAMI 2016*

And today : U-net [Ronneberger et al., 2015], Hourglass [Newell et al., 2016]...

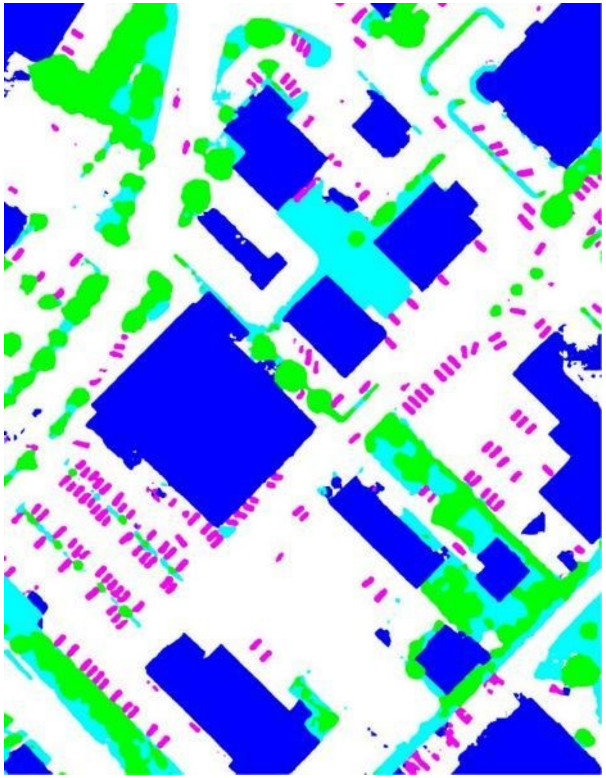
SegNet for semantic segmentation of EO data

ISPRS / Vaihingen
IR/R/G 10cm/pixel

Ground-truth



SegNet



F1 road 93%

F1 building 95%

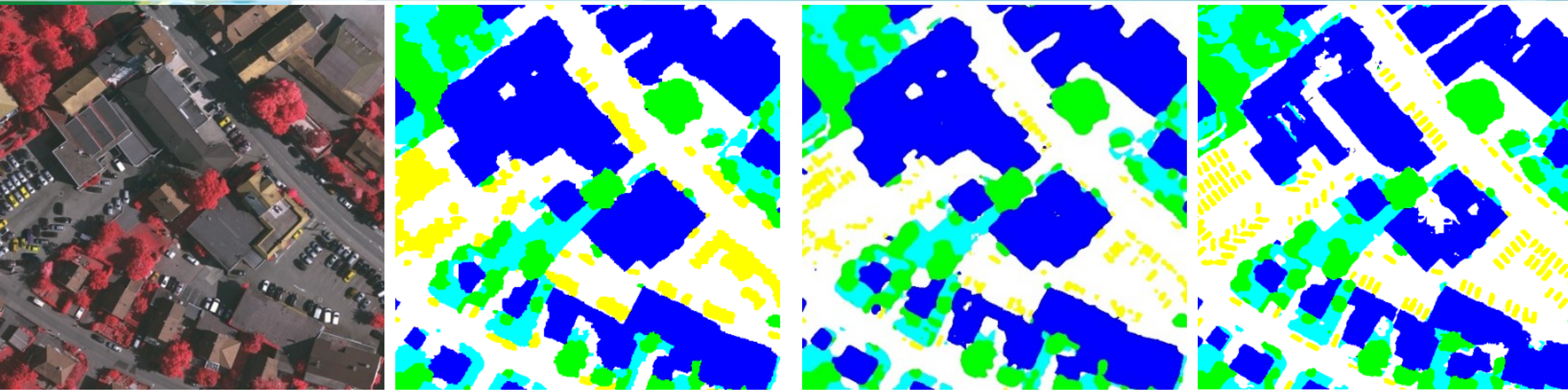
F1 low. Veg. 84%

F1 trees 82%

F1 cars 81%

Overall acc. 89.1%

SegNet compared



ISPRS / Vaihingen
IR/R/G 10cm/pixel

CNN+RF+CRF

FCN

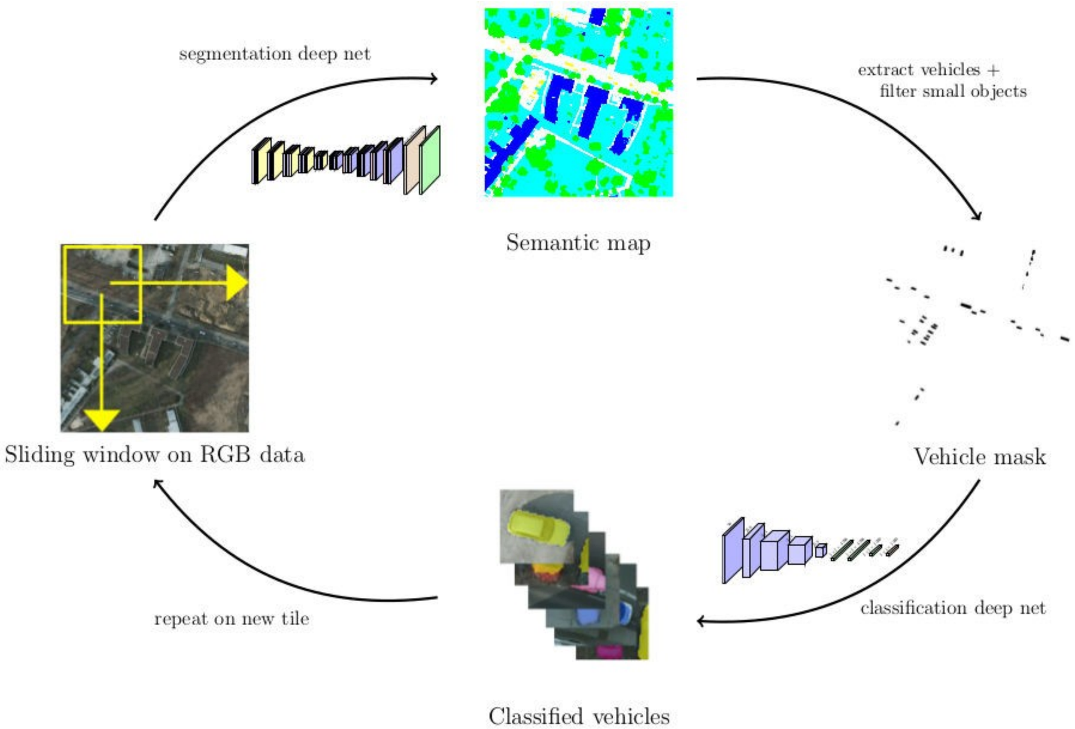
SegNet

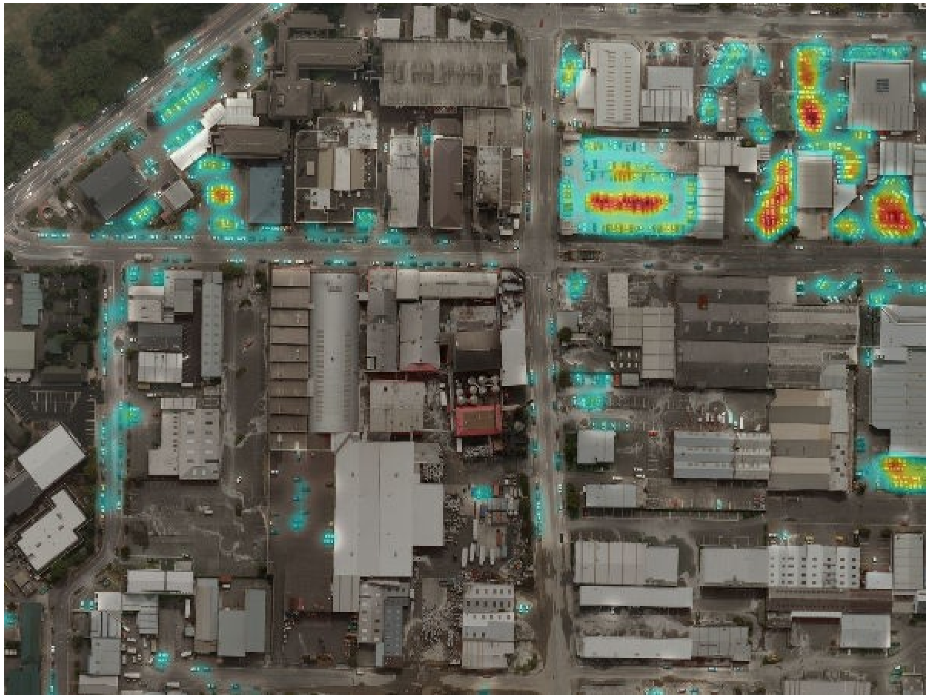
Summary:

- Encoder-decoder frameworks result in precise maps
- Very good overall accuracy
- Precise segmentation of small objects (vehicles…)
- Pre-trained models available in the Caffe Model Zoo / in pytorch
- ➢ Check out: https://github.com/nshaud/DeepNetsForEO

Semantic segmentation of Earth-observation data using multimodal and multi-scale deep networks, *Nicolas Audebert, Bertrand Le Saux, Sébastien Lefèvre*, **ACCV'2016**

Segment-before-detect





- Segmentation is precise enough to detect vehicles by simple connected component extraction
- Allows study of vehicle repartition and density in cities

Segment-before-detect: Vehicle Detection and Classification through Semantic Segmentation of Aerial Images, *Nicolas Audebert, Bertrand Le Saux*, *Sébastien Lefèvre*, **Remote Sensing'2017**

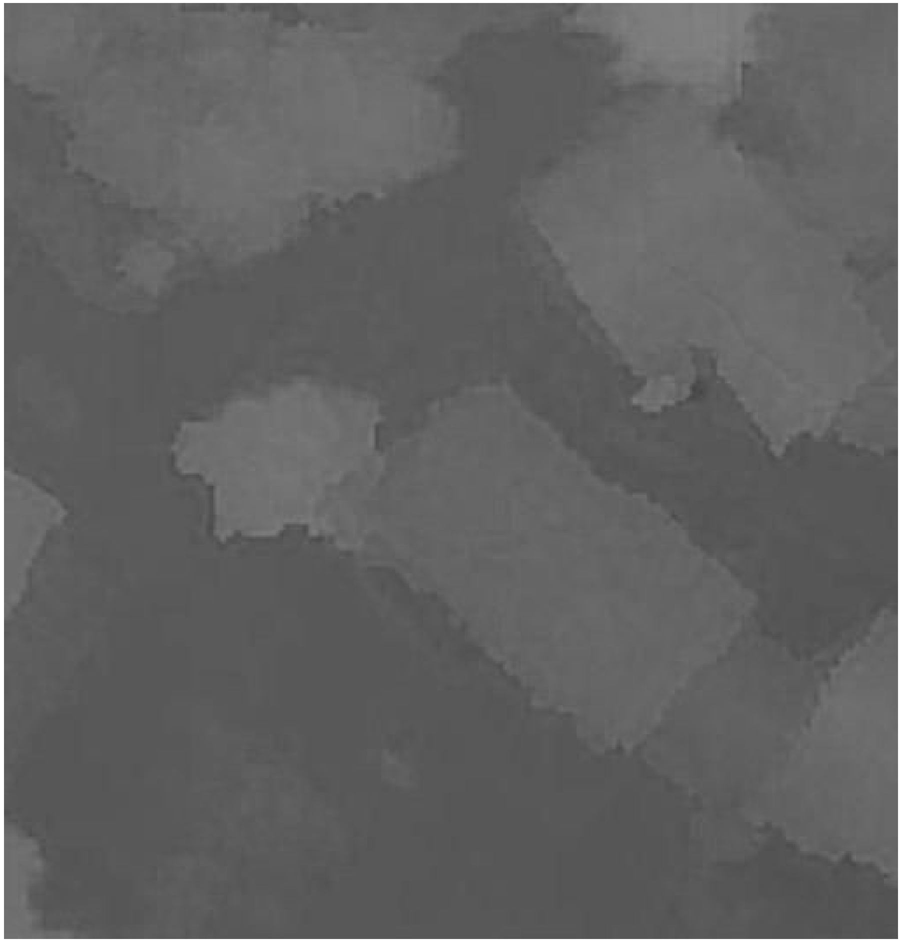
Fusion of heterogeneous data:

residual correction

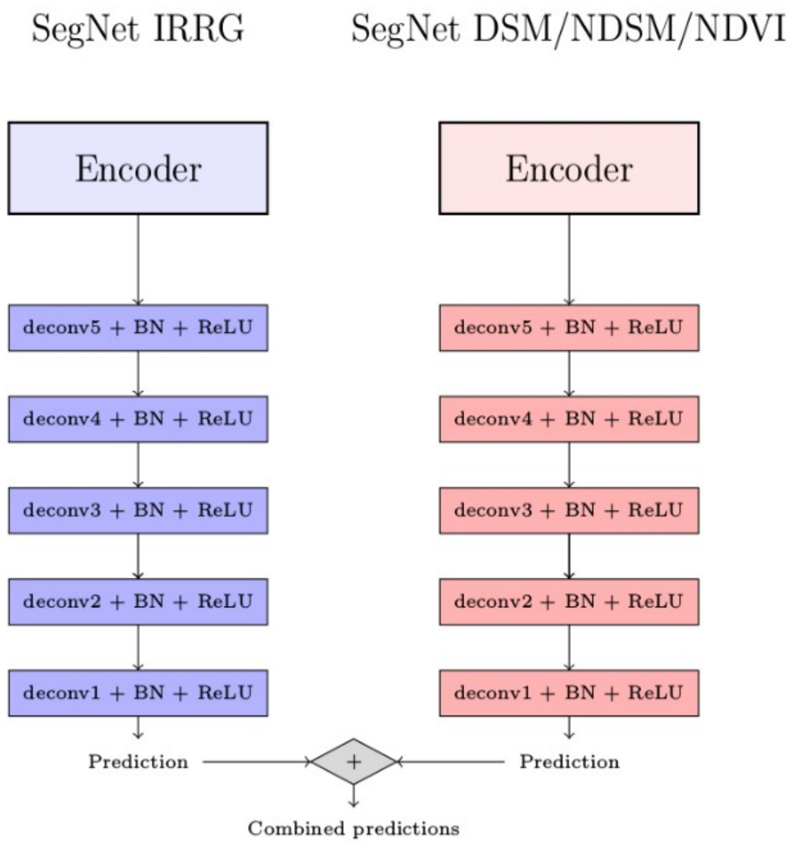
(with Nicolas Audebert and Sébastien Lefèvre)

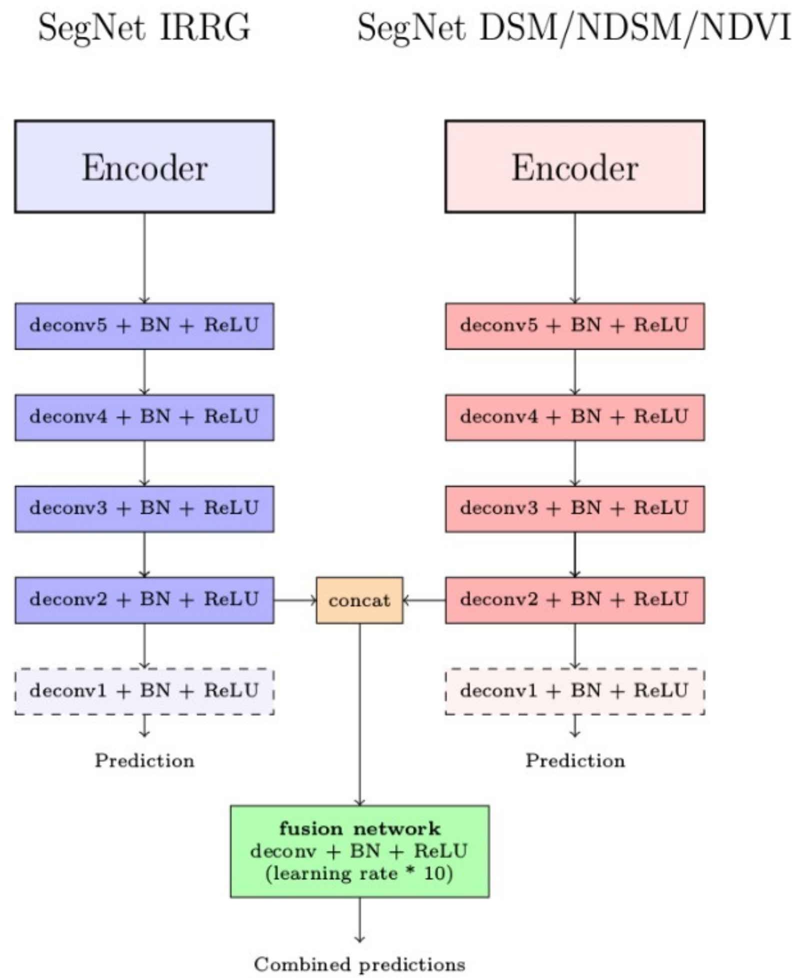
Fusion with residual correction

How can we use complementary data such as optical IR/R/G and LiDAR (DSM / nDSM) together ?



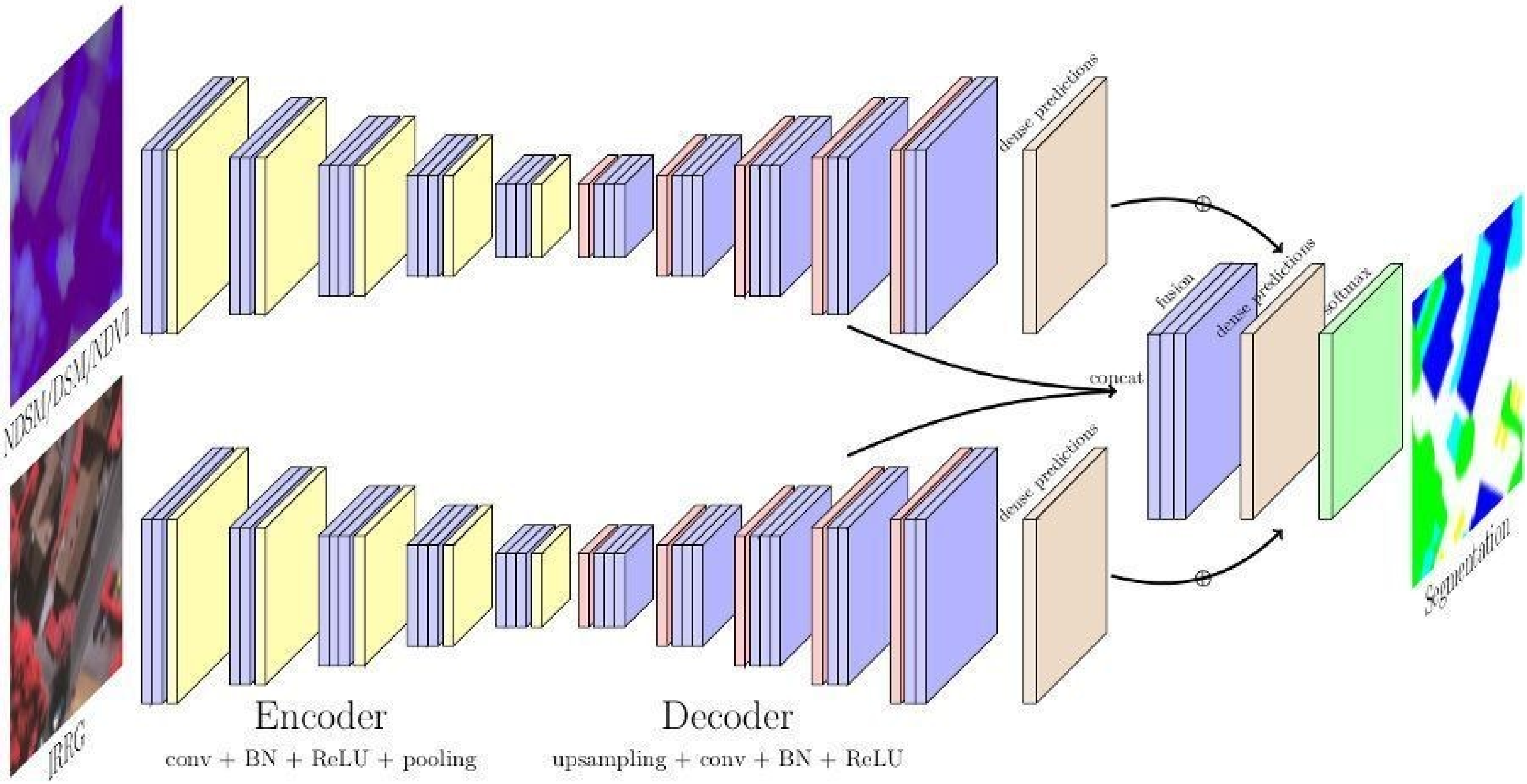
Fusion with residual correction





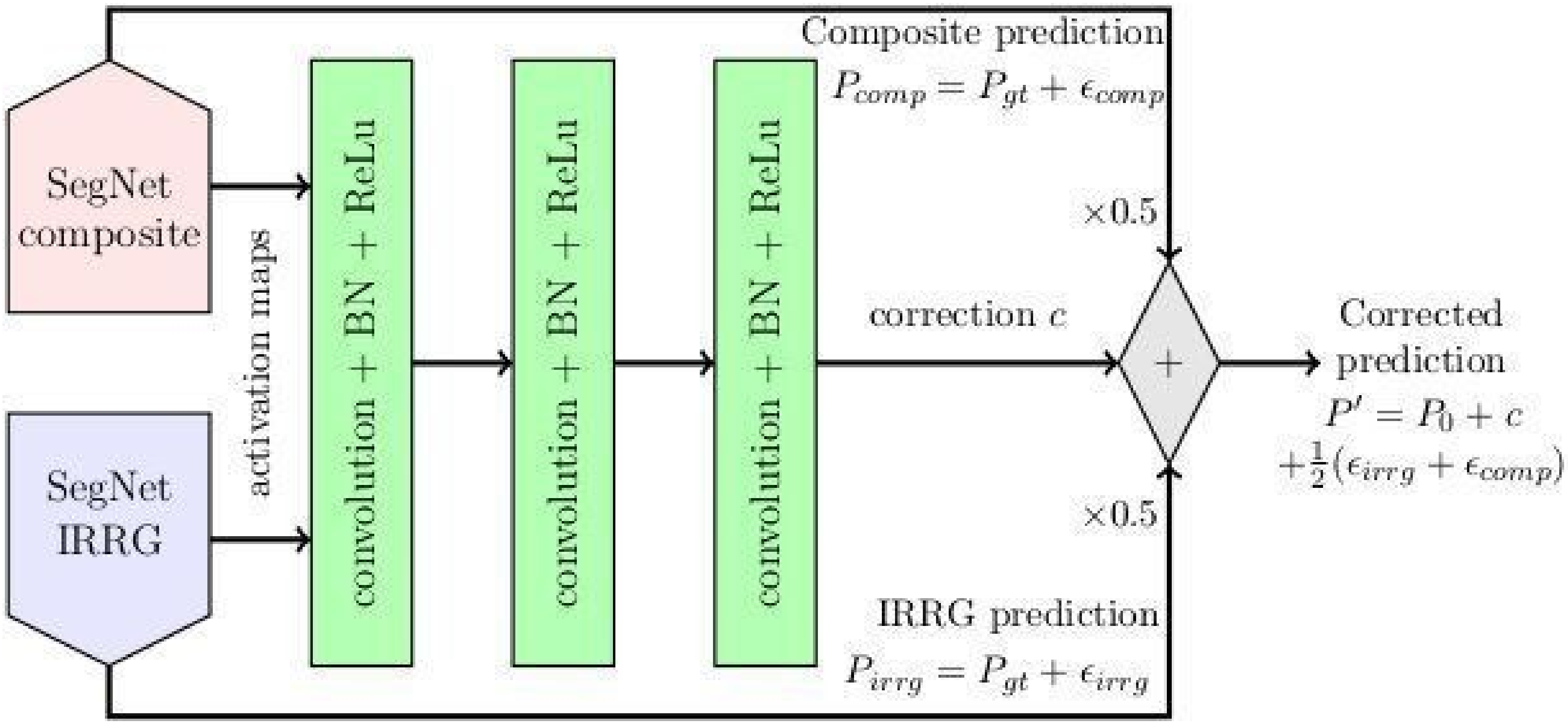
Dual stream with
naive fusion (averaging the 2 predictions) vs. Learning-based fusion

Fusion with residual correction



- Dual-stream: RGB and Composite (DSM, NDSM, NDVI)
- Learning-based fusion based on ***residual correction***

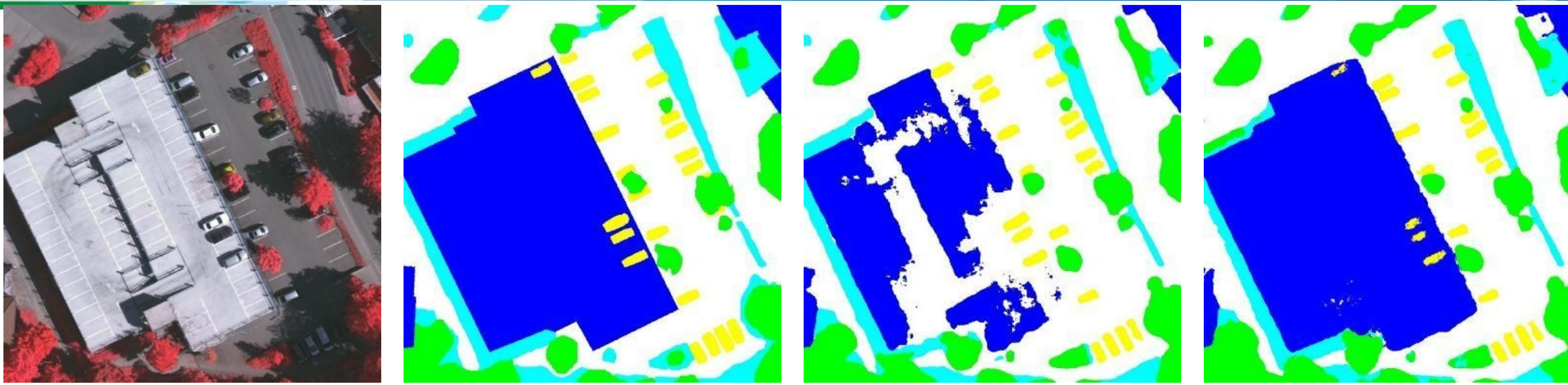
Fusion with residual correction



- Inspired by residual learning [He et al., 2015]
- Learn to correct 2nd-order prediction error

Semantic Segmentation of Earth Observation Data Using Multimodal and Multi-scale Deep Networks, *Nicolas Audebert, Bertrand Le Saux*, *Sébastien Lefèvre*, **ACCV'2016**; **ISPRS J. PHOTO'2018.**

Residual correction results



IR/R/G | Groundtruth | IR/R/G prediction | Fused predictions

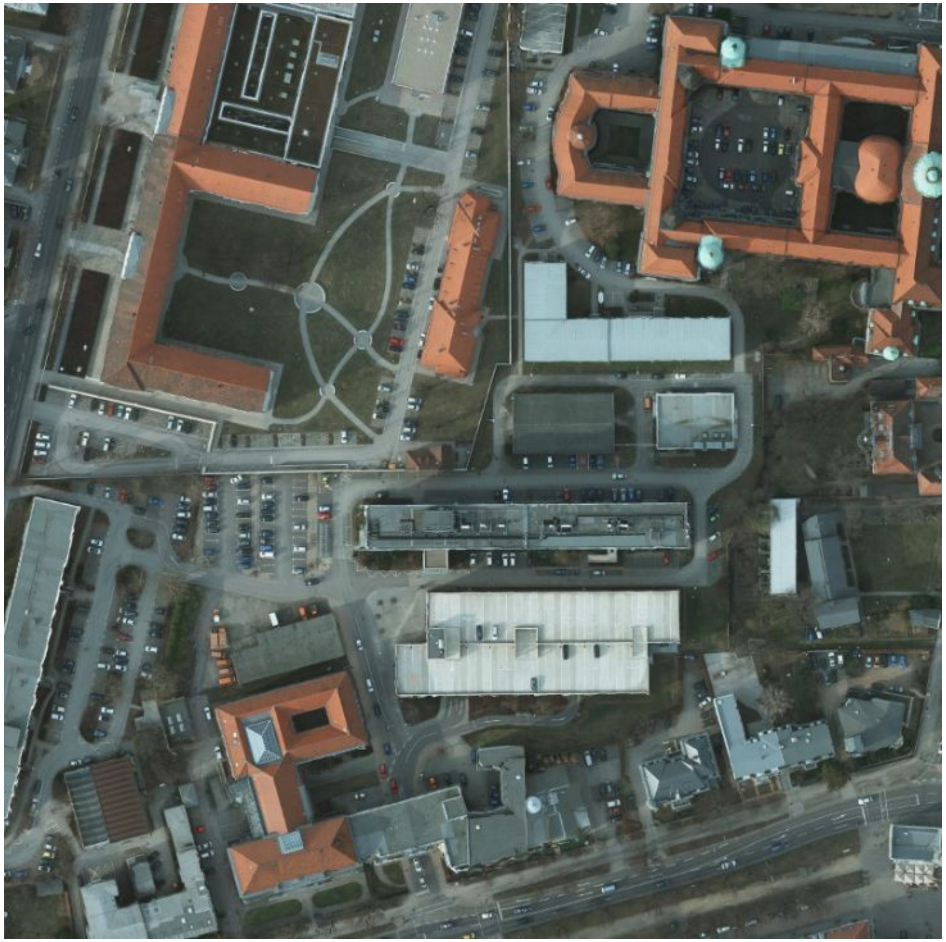
Method	imp surf	building	low veg	tree	car	Accuracy
RF + CRF (“HUST”)	86.9%	92.0%	78.3%	86.9%	29.0%	85.9%
CNN ensemble (“ONE_5”)	87.8%	92.0%	77.8%	86.2%	50.7%	85.9%
FCN (“DLR_2”)	90.3%	92.3%	82.5%	89.5%	76.3%	88.5%
FCN + RF + CRF (“DST_2”)	90.5%	93.7%	83.4%	89.2%	72.6%	89.1%
SegNet++	**91.5%**	94.3%	82.7%	89.3%	**85.7%**	89.4%
SegNet++ + fusion	91.0%	**94.5%**	**84.4%**	**89.9%**	77.8%	**89.8%**

Joint learning with additional cartography

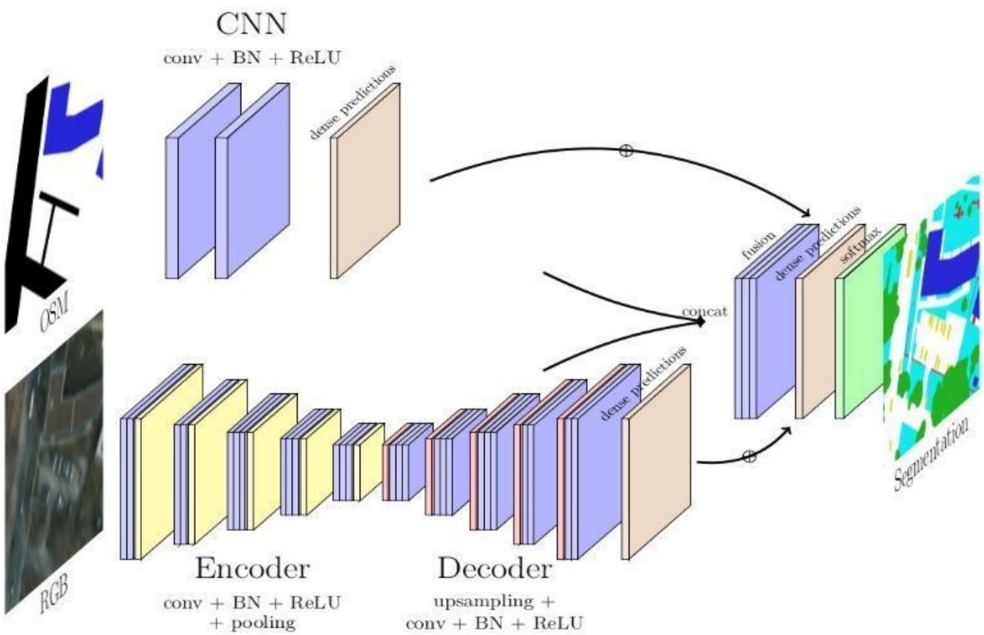
(with Nicolas Audebert and Sébastien Lefèvre)

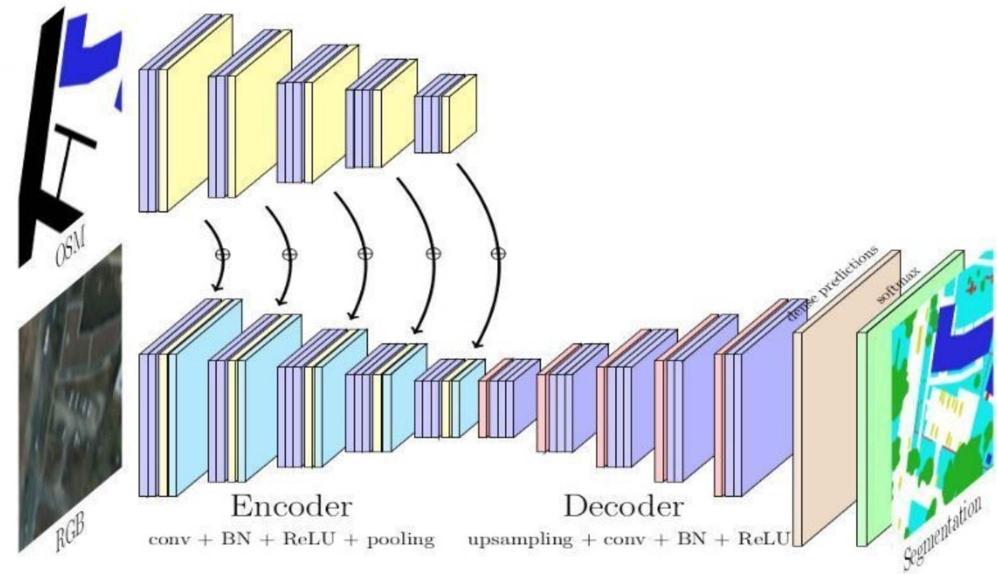
Joint-learning with additional cartography

How can we use collaborative, open source cartography to help us ?



Joint-learning with additional cartography





Optical and OSM data fusion using residual correction

Fusenet architecture applied to optical and OSM

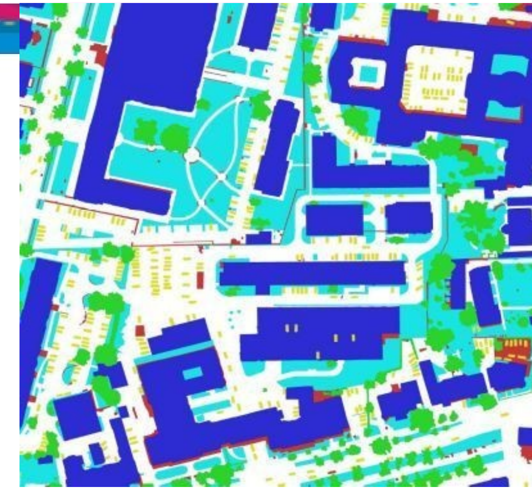
Fusenet : Hazirbas et al., ““FuseNet: Incorporating Depth into Semantic Segmentation via Fusion-Based CNN Architecture”, ACCV 2016

Joint-learning results

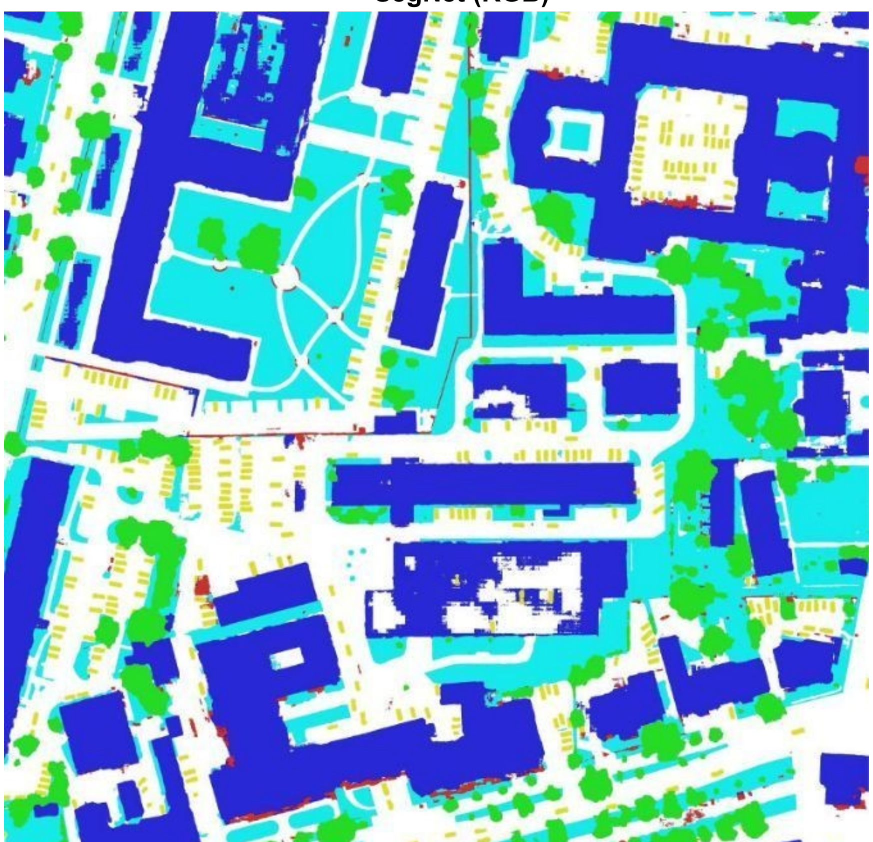
RGB

OSM

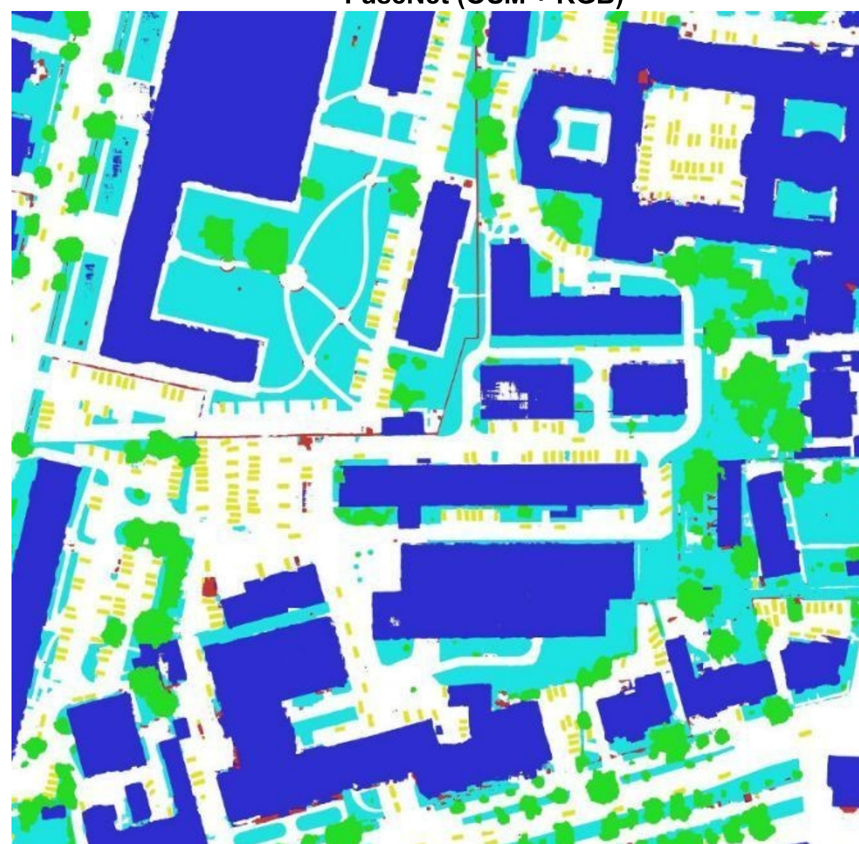
Ground truth



SegNet (RGB)



FuseNet (OSM + RGB)

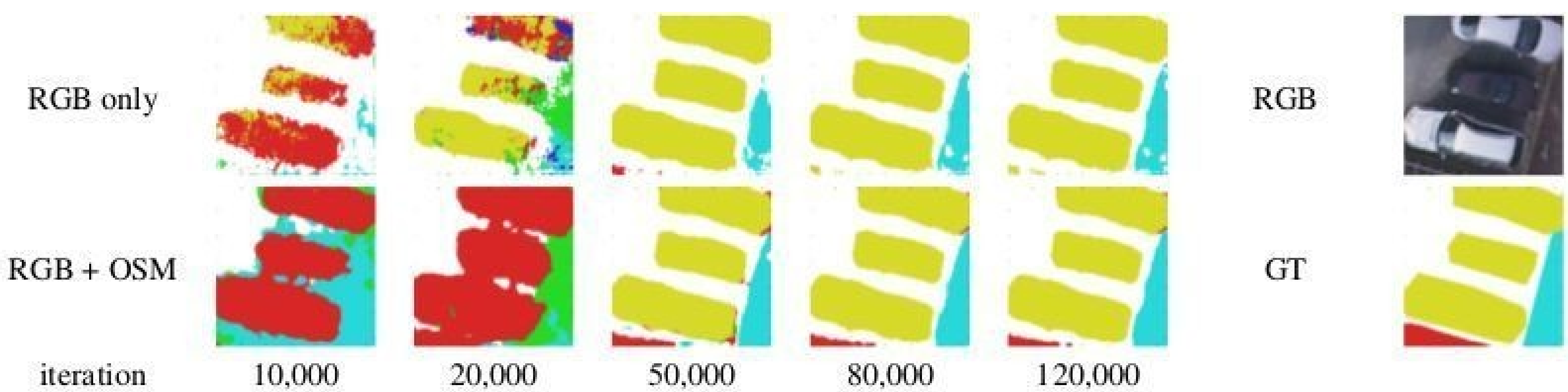


Joint-learning with additional cartography

Classification results

OSM	Method	imp. surfaces	buildings	low veg.	trees	cars	Overall
Binary	OSMNet	54.8	90.0	51.5	0.0	0.0	60.3
∅	SegNet RGB	93.0	92.9	85.0	85.1	95.1	89.7
Binary	Residual Correction RGB+OSM	93.9	92.8	85.1	**85.2**	95.8	90.6
	FuseNet RGB+OSM	**95.3**	**95.9**	86.3	85.1	**96.8**	**92.3**

Evolution during training



→ Converges faster and yields in better-defined structures

Joint Learning from Earth Observation and OpenStreetMap Data to Get Faster Better Semantic Maps, N. Audebert, B. Le Saux, S. Lefèvre, **CVPR/Earth Vision'2017**

Multi-temporal activity analysis

(with Rodrigo Daudt, Alexandre Boulch and Yann Gousseau)

Multi-temporal activity analysis

Rio (Brazil) - Original Copernicus Sentinel Data 2018 available from the European Space Agency (https://sentinel.esa.int).

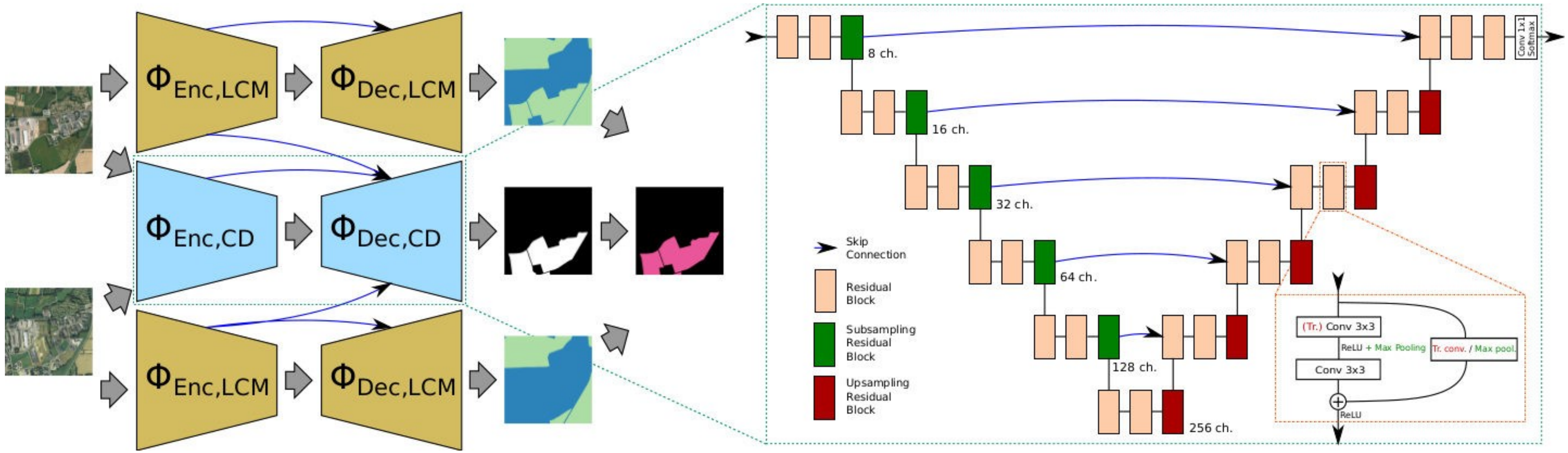
How to extend semantic analysis to multitemporal data ?

→ detect changes ;

→ monitor activity in high-revisit rate acquisitions ;

→ focus on specific changes (urban, agriculture, vehicles, industrial activity...)

Multi-temporal activity analysis

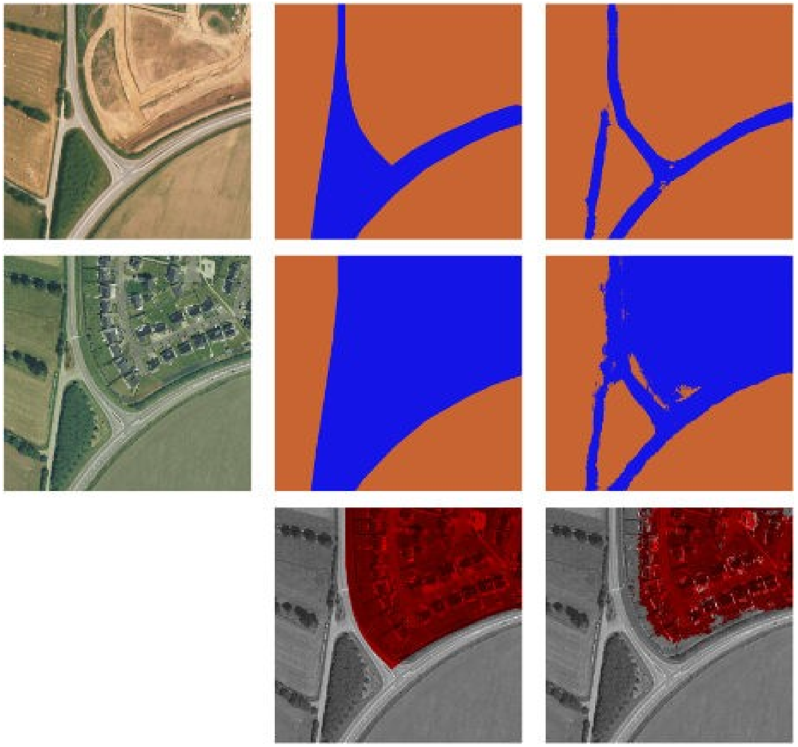


Semantic Change Detection:

- Fully convolutional networks for change detection
- Joint **multi-task learning** of land cover and change maps
- Creation of the first large scale dataset for semantic change detection:
  HRSCD – High Resolution Semantic Change Detection Dataset
  https://ieee-dataport.org/open-access/hrscd-high-resolution-semantic-change-detection-dataset

Multi-Task Learning for High Resolution Semantic Change Detection,
R. Caye Daudt, B. Le Saux, A. Boulch, Y. Gousseau, **CVIU 2019**

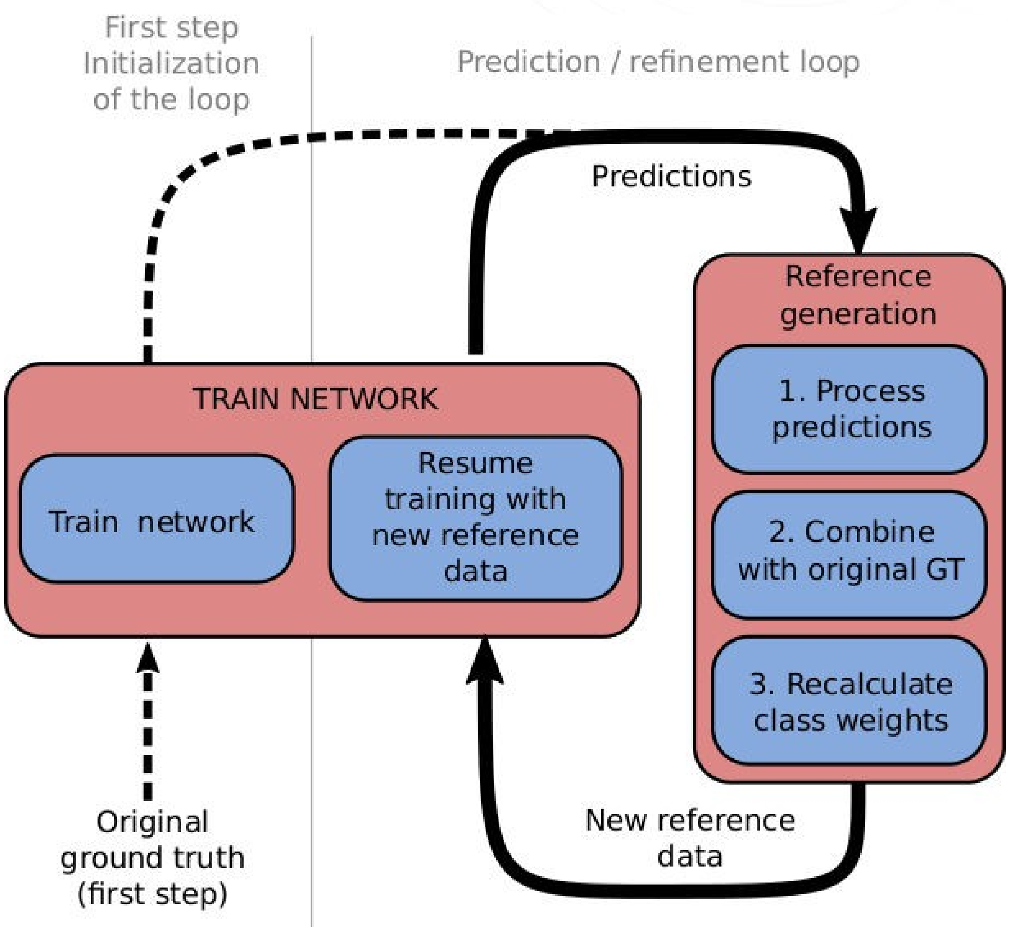
Multi-temporal activity analysis



- End-to-end, fully convolutional networks for change detection
- Prediction of land covers and change maps

➔ Dense prediction of urban evolution in open data

Multi-temporal activity analysis



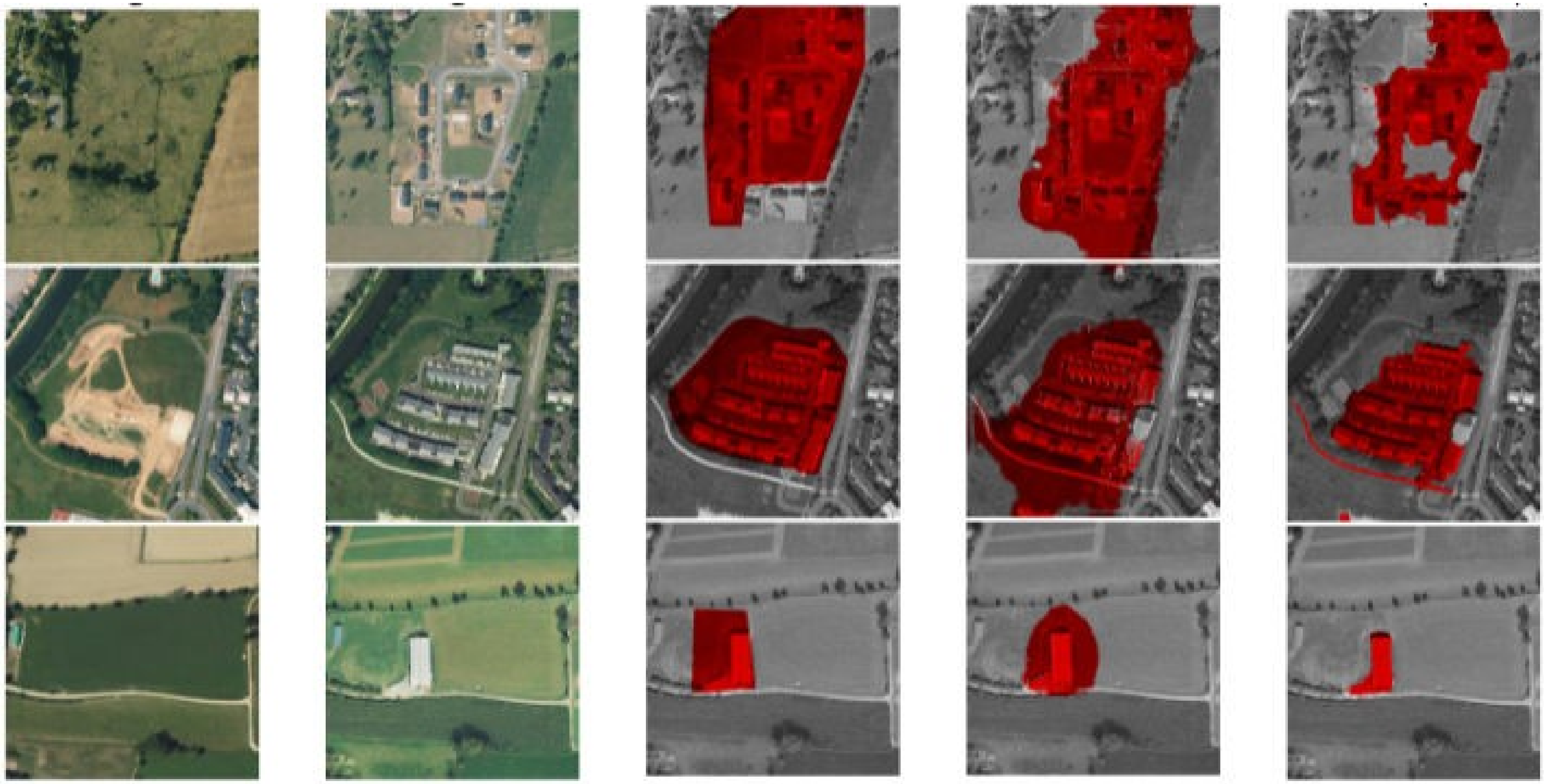
… but reference data might be unreliable !

➔ Weak-learning

- Iterative training with data cleansing
- Process predictions with **Guided Anisotropic Diffusion** to fit the images

Guided Anisotropic Diffusion and Iterative Learning for Supervised Change Detection, *R. Caye Daudt, B. Le Saux, A. Boulch, Y. Gousseau*, **CVPR/EarthVision 2019**

Multi-temporal activity analysis



- (Cautious) iterative model training / reference cleansing method
- Prediction of “true change” maps

➔ Better trained networks, reducing the effect of approximate labels

Hyperspectral data classification

(with Nicolas Audebert and Sébastien Lefèvre)

Hyperspectral data classification

Houston (Texas, USA) – IEEE GRSS IADF TC's Data Fusion Contest 2018
(http://www.grss-ieee.org/community/technical-committees/data-fusion/data-fusion-contest/).

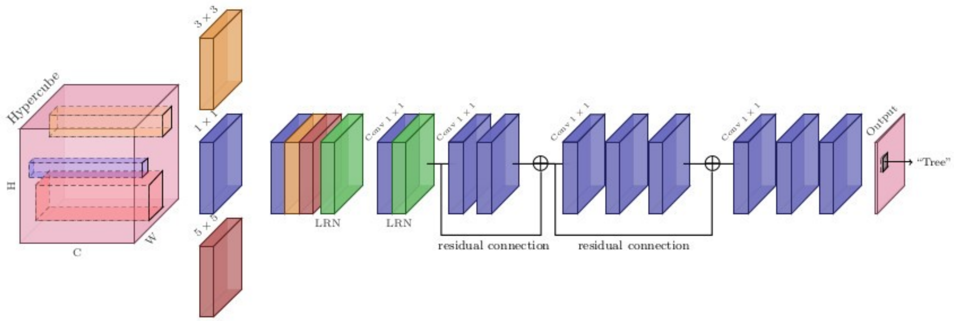
How to extend semantic analysis to hyperspectral data ?

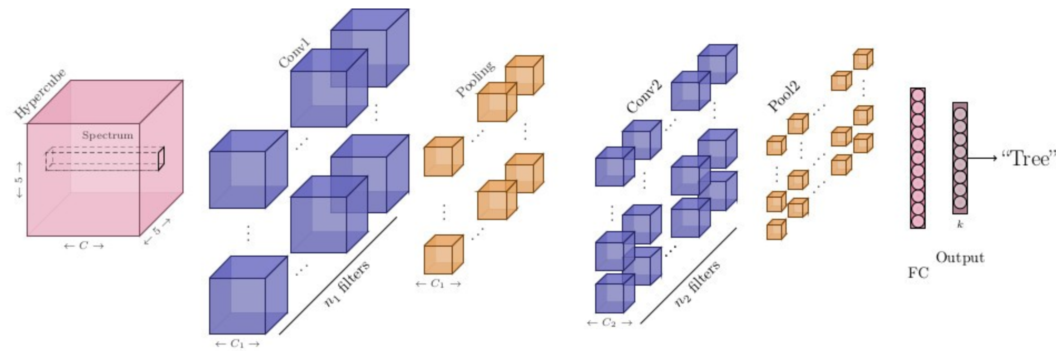
→ RGB to 100+ bands, image to data cube ;

→ finer spectral description, out-of-visible ;

→ lower resolution but finer class discrimination (materials, stressed or healthy vegetation...)

Hyperspectral data classification





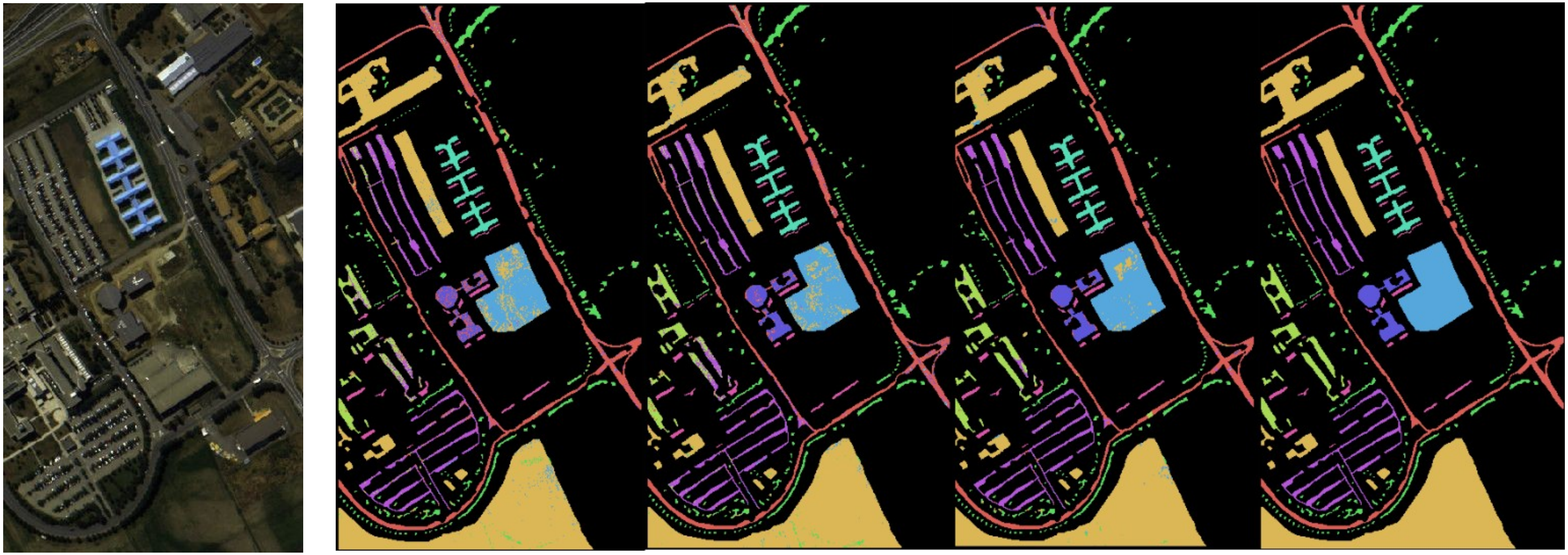
Several conv. net architectures adapted to HSI classification :

- Spectrum-based (1D), spatial-spectral
- 3D-convolution CNNs

➢Open-source toolbox DeepHyperX: https://github.com/nshaud/DeepHyperX

Audebert et al.. *Deep learning for Hyperspectral classification: A comparative review*, GRSM'19

Hyperspectral data classification



Composite SVM CNN 1D CNN 3D Ground-Truth

Pavia Univ. dataset :

- 1D conv. nets slighly better than standard SVM
- 3D conv. nets offer better spatial regularization (retrieve local 3D spatial-spectral patterns)

Deep Learning Classification of Hyperspectral Data: A Comparative Review, N. Audebert, B. Le Saux, S. Lefèvre, **GRSM 2019**

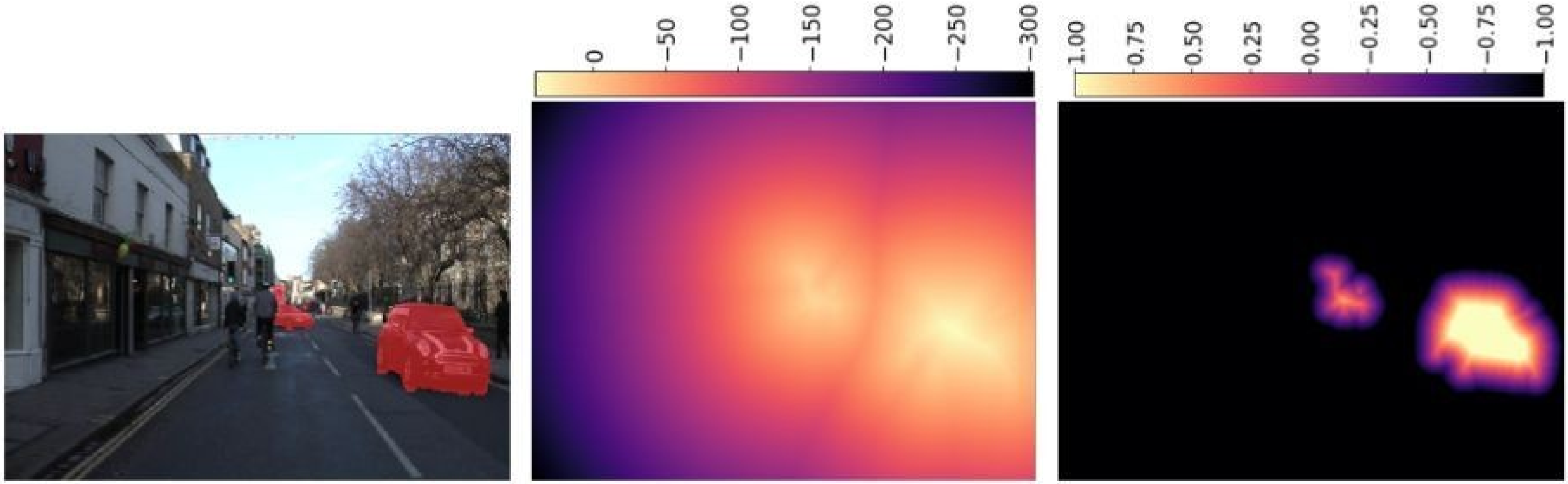
RGB to depth NYUv2 dataset

https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.htm
l

Distance Transform Regression for Semantic Segmentation

(with Nicolas Audebert, Alexandre Boulch and Sébastien Lefèvre)

Distance Transform Regression for Semantic Segmentation

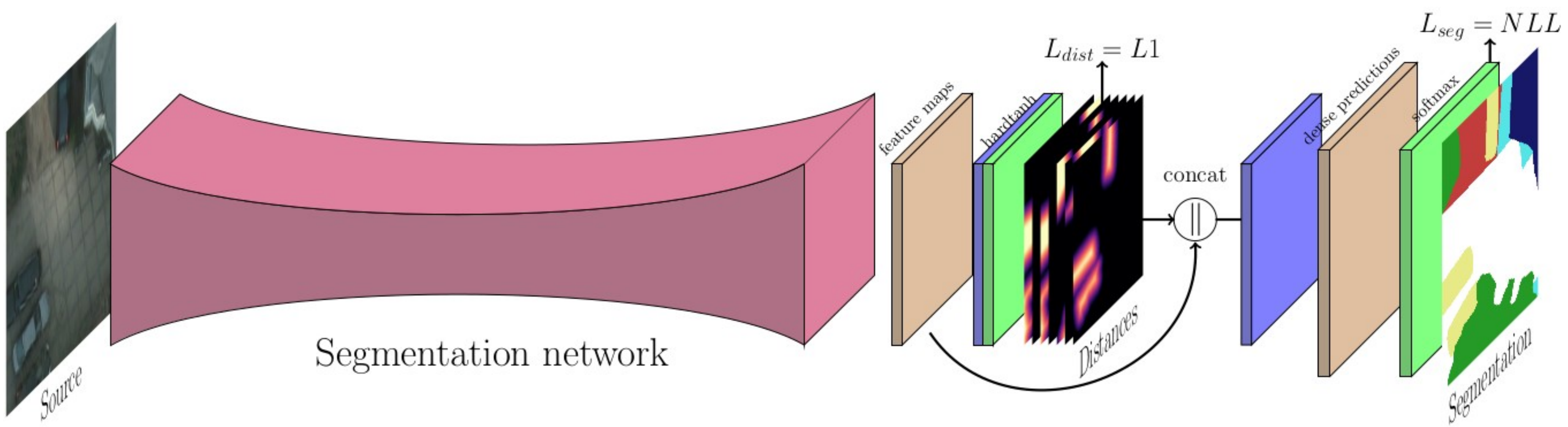


(a) Car segmentation binary mask. (b) Signed distance transform (SDT). (c) Truncated and normalized SDT.

Play with losses to change the objective :

- Classification borders are often imprecise, even in ground-truth !
- Add more information to drive the optimization, e.g. distance to the boundary

Distance Transform Regression for Semantic Segmentation

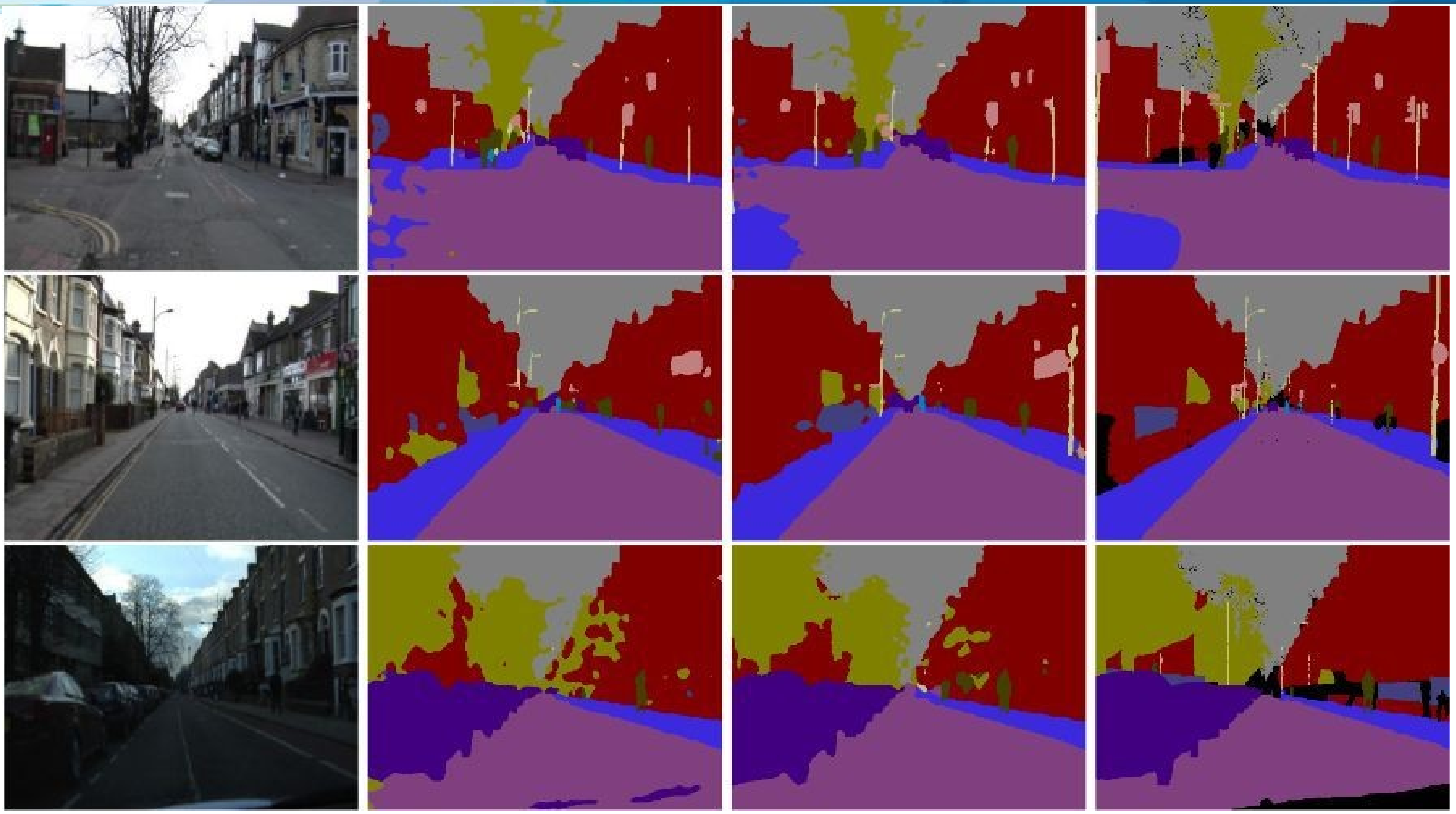


Multi-task learning :

- L1-Regression on the truncated distance maps, and
- Cross-entropy classification on the class label masks.

→ *Regularization of the classification*

Distance Transform Regression for Semantic Segmentation



CamVid RGB PSPnet (classif) PSPnet (classif+DTR) GT

→ Improves consistency / smoothness for sidewalks, trees, poles and trafic signs

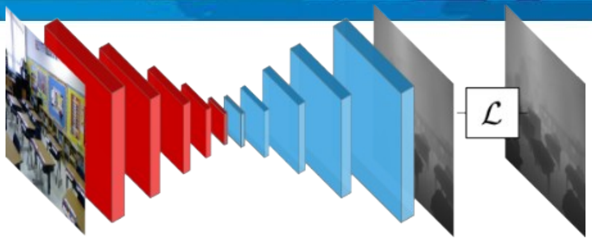
Distance transform regression for spatially-aware deep semantic segmentation, N. Audebert, A. Boulch, B. Le Saux, S. Lefèvre, **CVIU 2019**

Regression Losses for Single-Image Depth Estimation

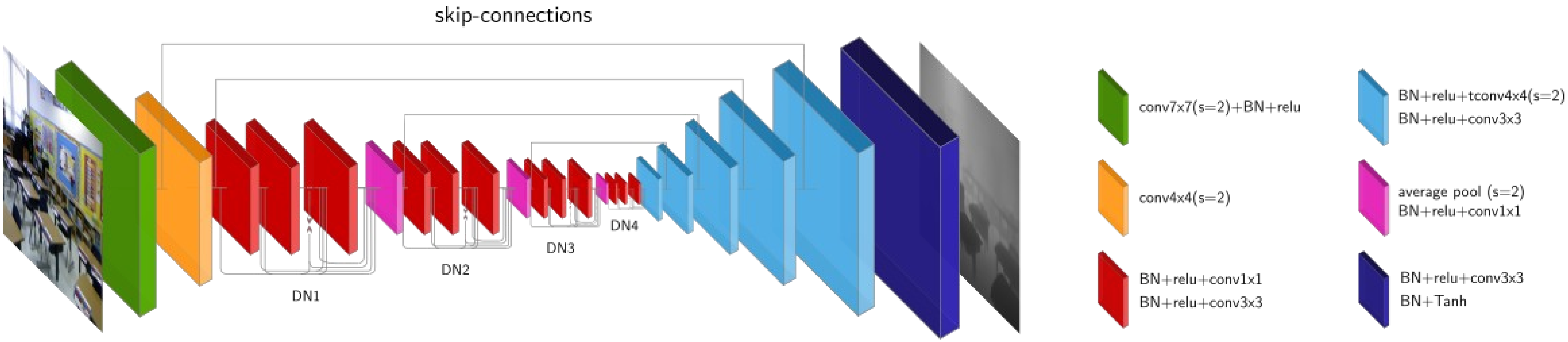
(with Marcela Carvalho, Pauline Trouvé-Peloux, Frédéric Champagnat and Andrès Almansa)

Regression for Depth Estimation

→ *Objective : regression on a depth map*



D3Net :

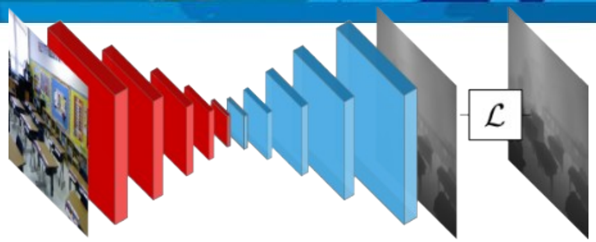


Encoder-decoder network with :

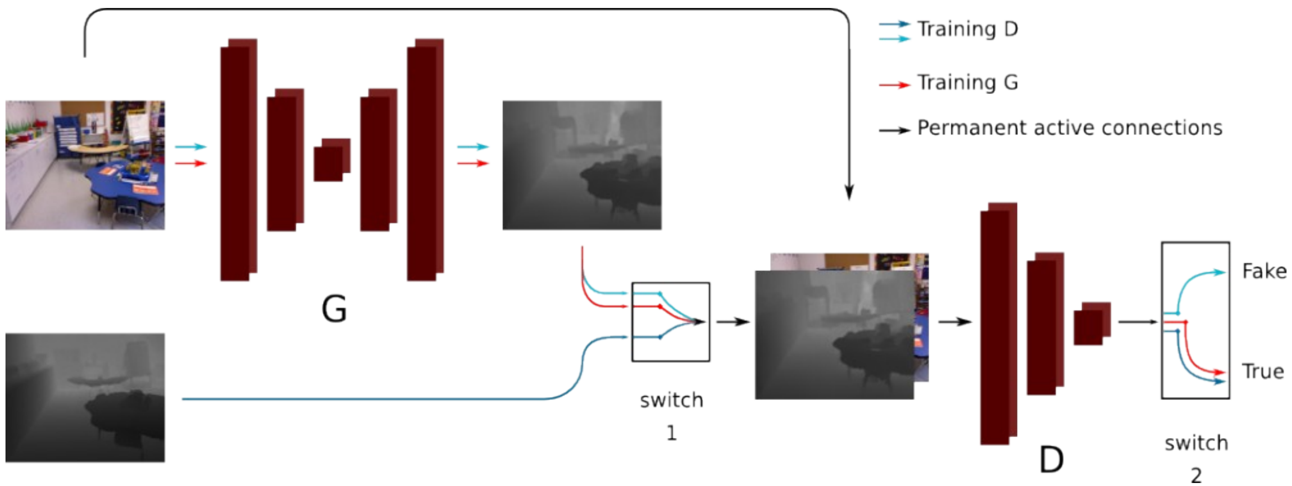
- Dense blocks in the encoder,
- Skipping connections between encoder and decoder for context-awareness...

Regression for Depth Estimation

→ *Objective : regression on a depth map*



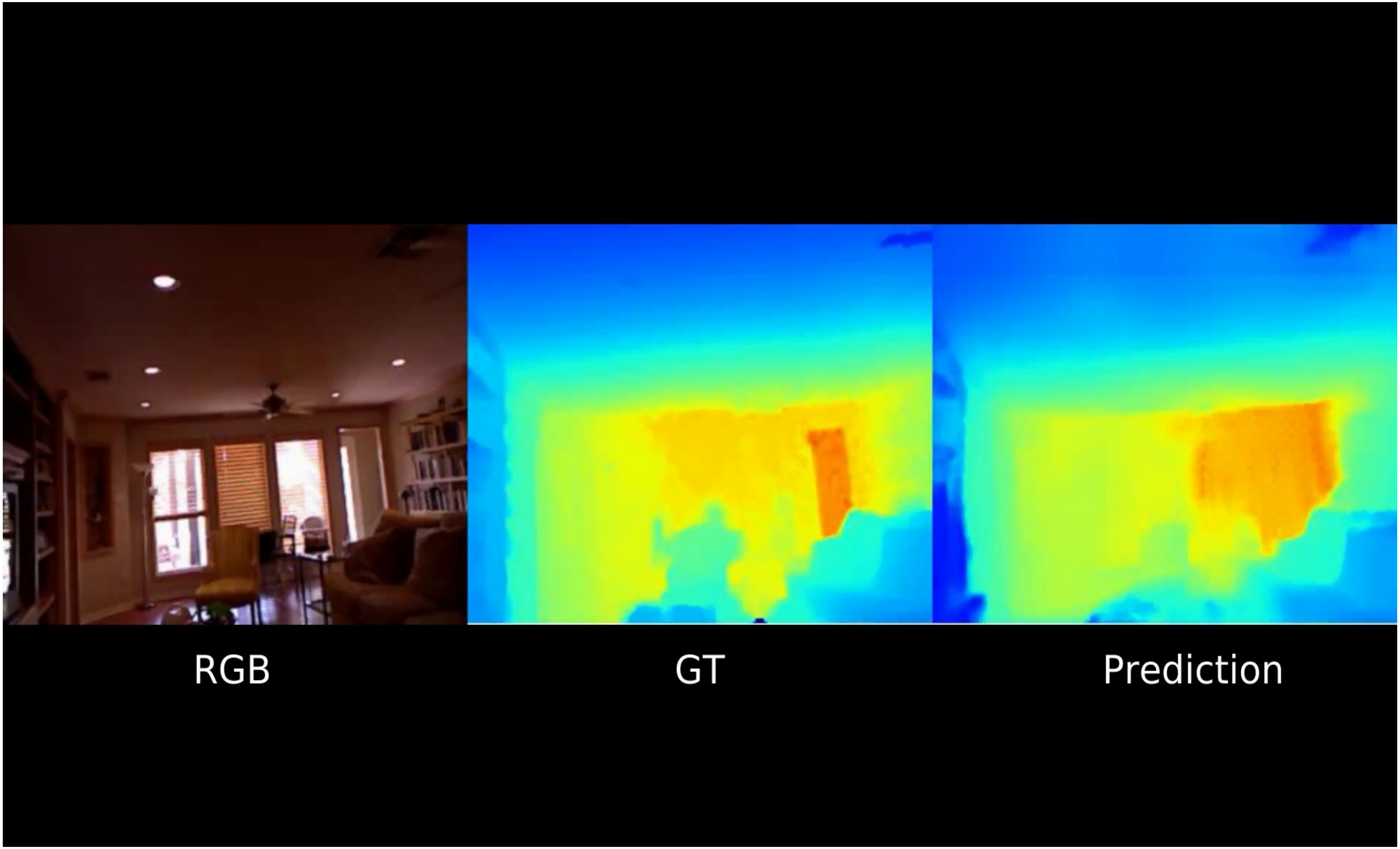
D3Net :



Regression loss with :

- *L1* for global estimation, and
- *Adversarial loss (LS-GAN)* for details (if enough samples!).

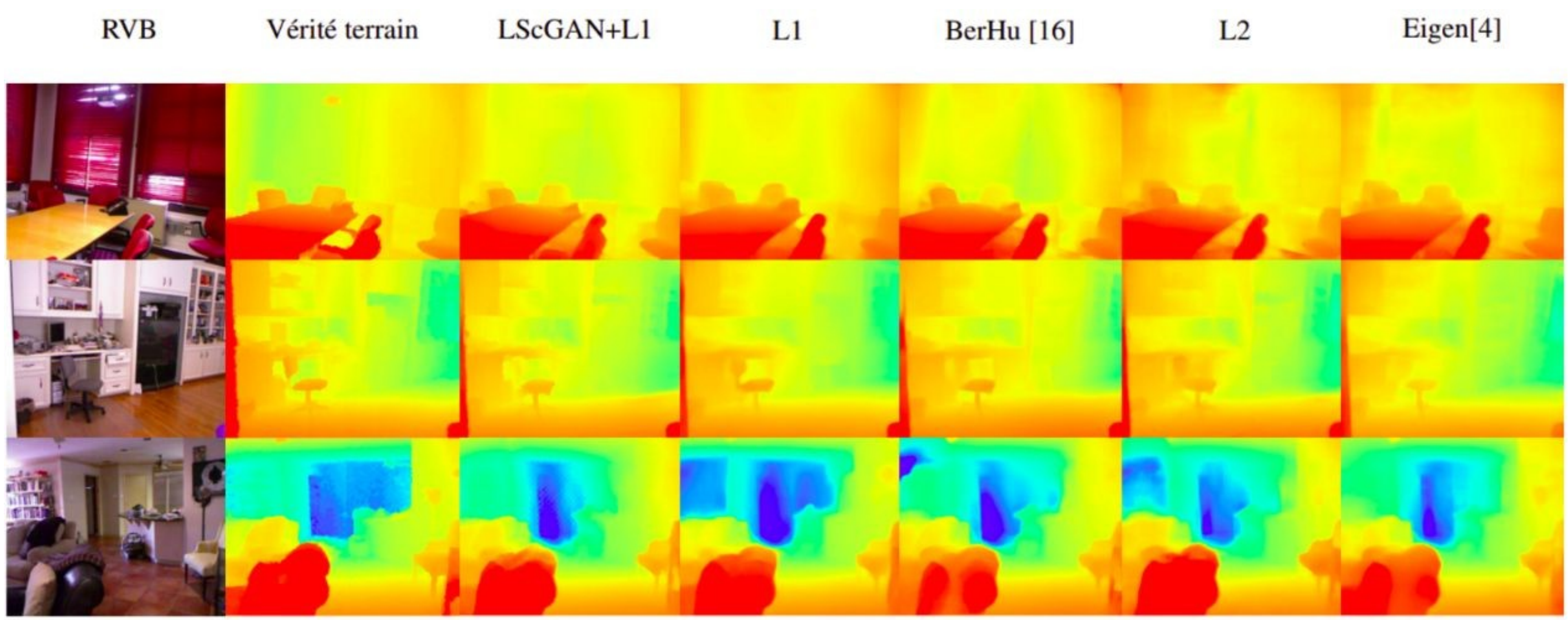
Regression for Depth Estimation [D3Net.mp4]



Regression for Depth Estimation

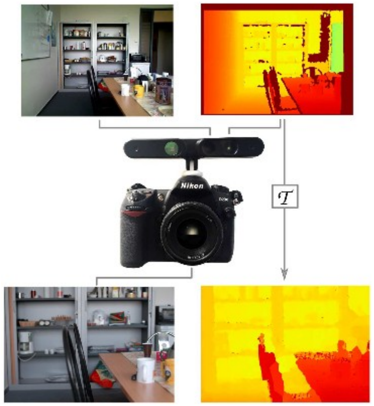
Results :

Methods	Error↓ rel	log10	rms	rmslog	Accuracy↑ $\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Eigen & Fergus 2015	0.158	-	0.641	0.214	76.9%	95.0%	98.8%
Laina et al. 2016	0.127	0.055	0.573	**0.195**	81.1%	95.3%	98.8%
D. Xu et al. 2017	0.121	*0.052*	0.586	-	81.1%	95.4%	98.7%
Cao et al. 2017	0.141	0.060	0.540	-	81.9%	96.5%	99.2%
Jung et al. 2017	*0.134*	-	0.527	-	**82.2%**	*97.1%*	*99.3%*
Kendall & Gal 2017	**0.110**	**0.045**	0.506	-	81.7%	95.9%	98.9%
D3-Net	0.136	-	**0.504**	*0.199*	*82.1%*	95.5%	98.7%

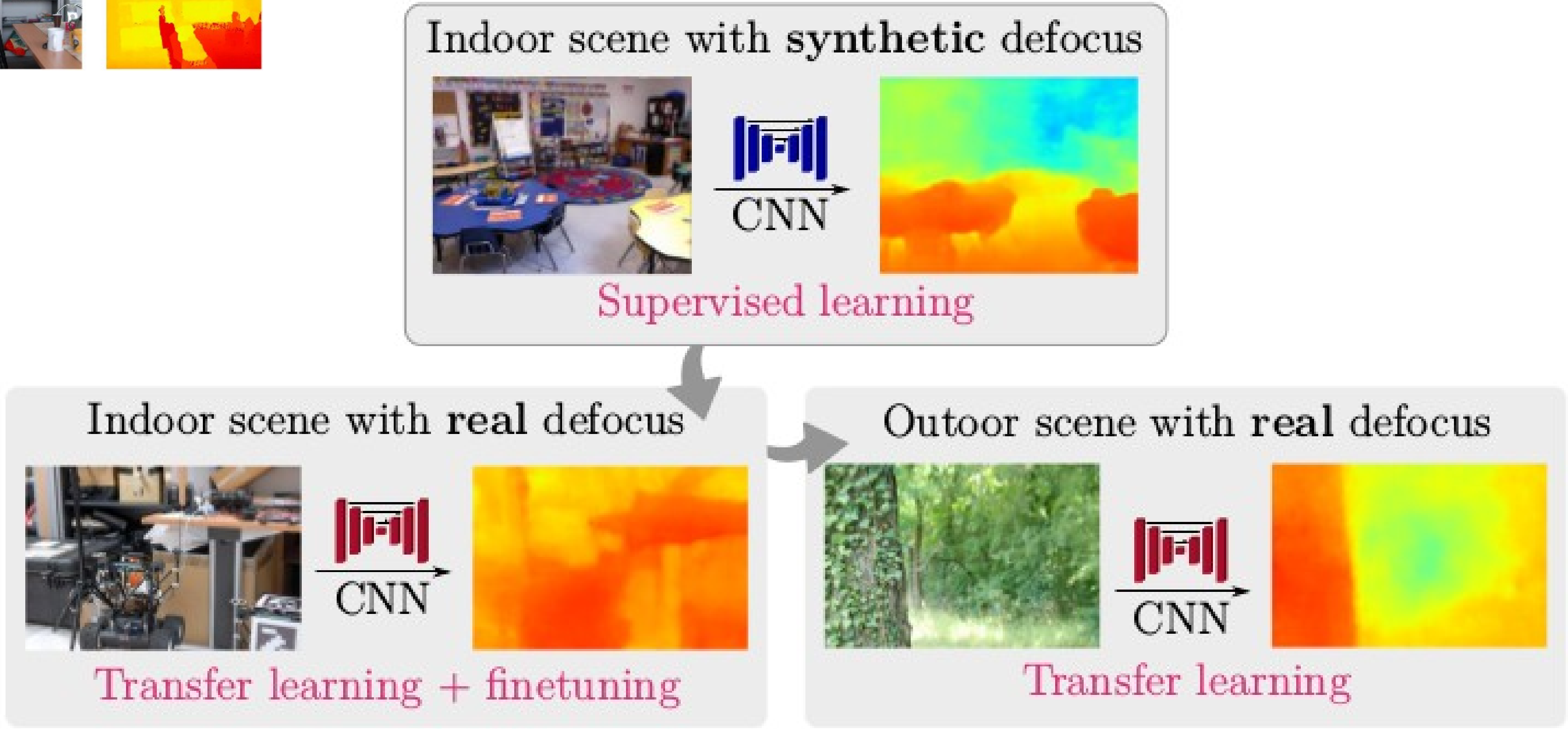


On regression functions for depth estimation, Marcela Carvalho, Bertrand Le Saux, Pauline Trouvé, Andrés Almansa, Fréderic Champagnat, **ICIP'18**

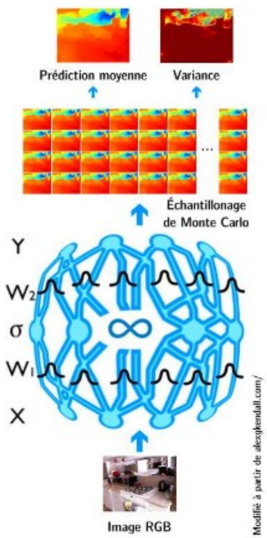
Regression for Depth Estimation



Deep from Defocus “in the wild” :
→ using lens with small depth of field



Regression for Depth Estimation

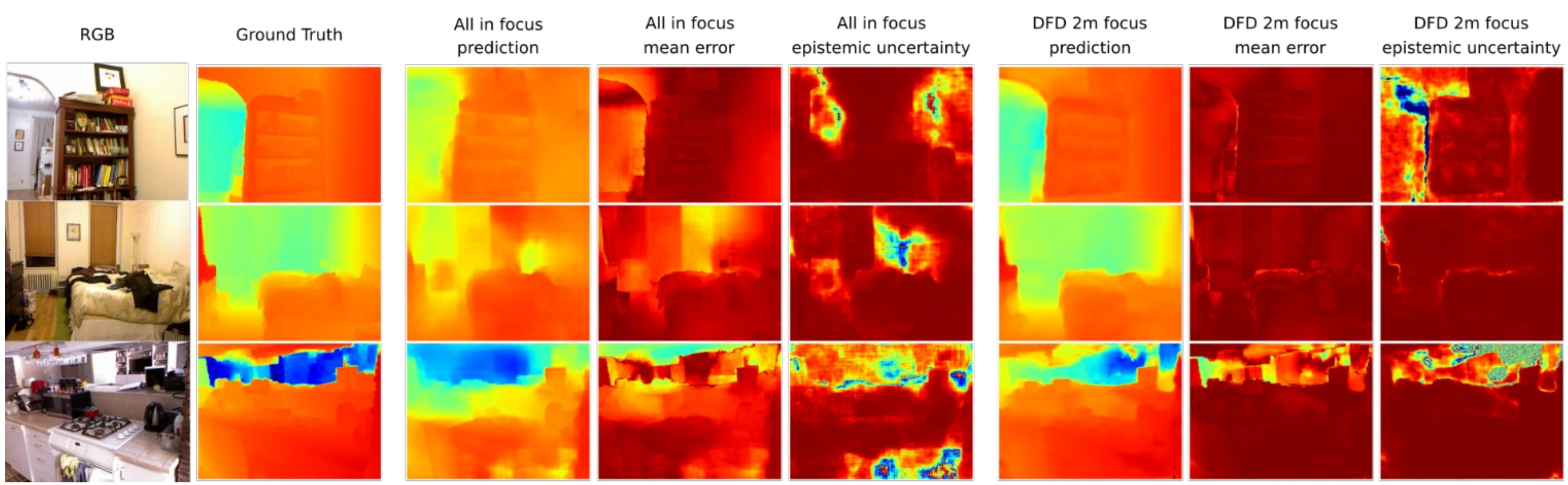


Measuring epistemic uncertainty of the network

- Bayesian net
- Monte-Carlo dropout

Uncertainty with and without depth-from-defocus :

- Uncertainty on low-textured areas
- Defocus reduces errors and increases confidence



On regression functions for depth estimation, M. Carvalho, B. Le Saux, P. Trouvé-Peloux, A. Almansa, F. Champagnat, **ECCV/3DRiW'18**

3D Robotic Exploration

(with Joris Guerry, Alexandre Boulch and David Filliat)

3D robotic exploration

Point-cloud from a single-view: RGB-D data

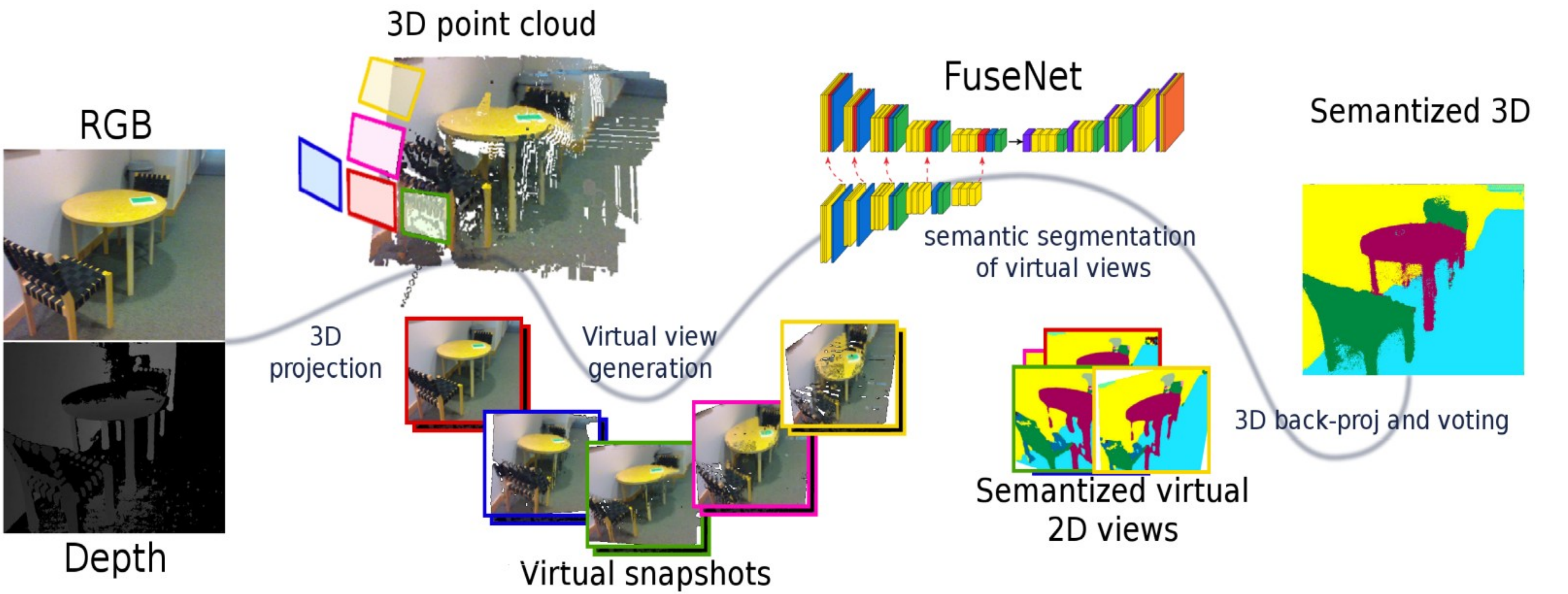
http://rgbd.cs.princeton.edu/

- Even with a single low-resolution, cheap RGB-D acquisition → rich 3D information
- But scene understanding depends on the point of view !

3D robotic exploration

Point-cloud from a single-view: RGB-D data

- Sampling strategy : around the original point of view
- Then quite standard SnapNet pipeline

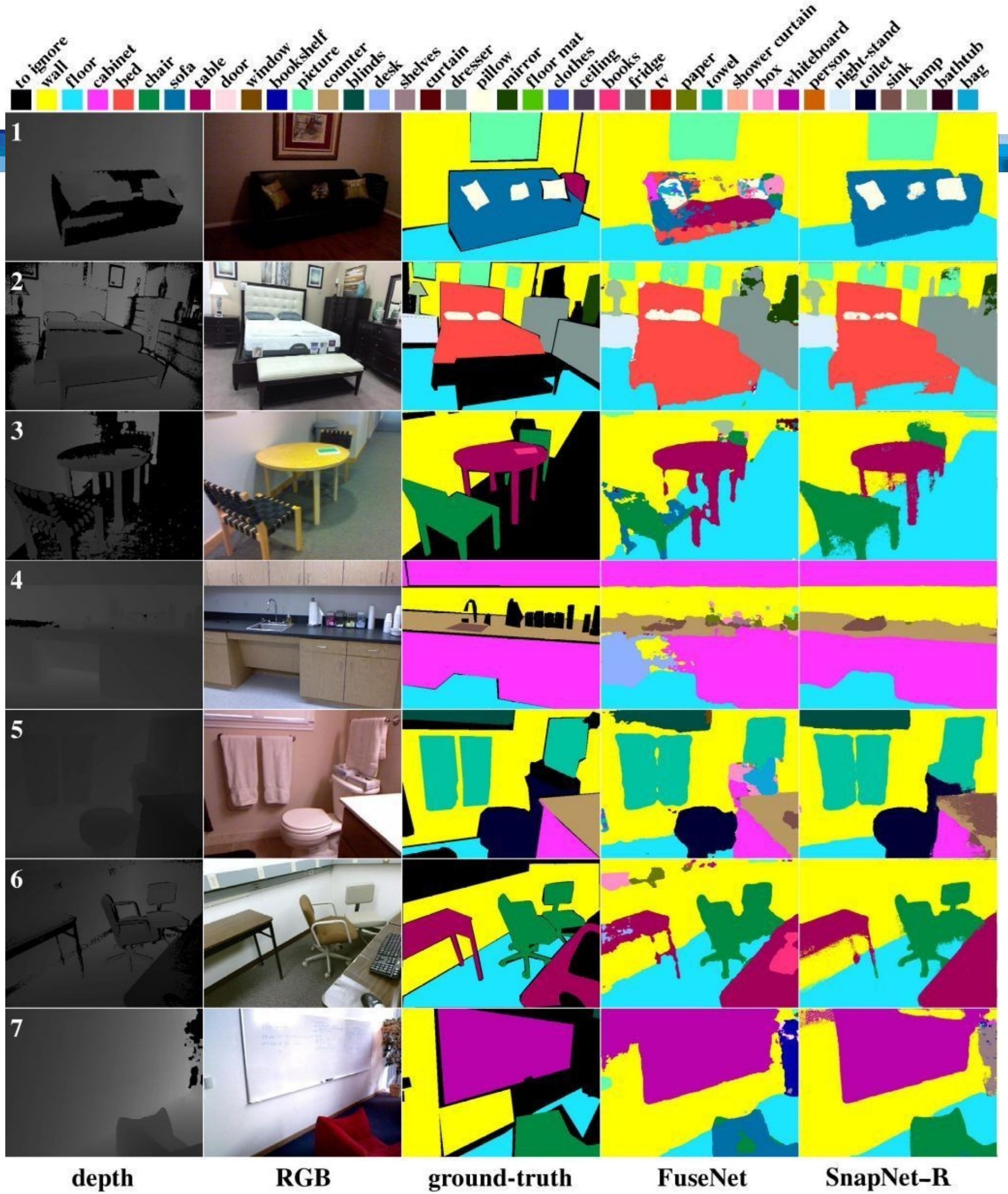


→ Works as *3D-consistent data augmentation*

SnapNet-R: Consistent 3D Multi-View Semantic Labeling for Robotics, Guerry et al., **ICCV/3DRMS'2017**

Point-cloud from a single-view: RGB-D data

SunRGBD



3D robot exploration

Point-cloud from a single-view: RGB-D data

SunRGBD

	Training		Testing				
experiment	preproc.	augm.	preproc.	augm.	OA	MA	IoU
LSTM-CF [30] (RGB)	✗	✗	✗	✗	–	48.1	–
FCN 8s [32] (RGB)	✗	✗	✗	✗	68.2	38.4	27.4
Bayesian SegNet [27] (RGB)	✗	✗	✗	✗	71.2	45.9	30.7
Context-CRF [31] (RGBD)	✗	✗	✗	✗	**78.4**	53.4	**42.3**
*FuseNet SF5[23] (RGBD)	✗	✗	✗	✗	76.3	48.3	37.3
DFCN-DCRF [26] (RGBD)	✗	✗	✗	✗	76.6	50.6	39.3
*1 FuseNet SF5	✗	✗	✗	✗	76.88	52.61	39.17
1 FuseNet SF5	✗	✗	✗	✗	77.21	54.81	39.11
2	✗	✗	✓	✗	74.87	52.47	36.68
3	✗	✗	✓	✓	72.52	53.27	33.89
4	✓	✗	✗	✗	72.81	52.02	34.32
5	✓	✗	✓	✗	77.20	55.03	39.33
6	✓	✗	✓	✓	70.25	***56.87***	30.32
7	✓	✓	✗	✗	75.51	53.71	36.65
8	✓	✓	✓	✗	77.57	56.70	38.83
9 SnapNet-R	✓	✓	✓	✓	***78.04***	**58.13**	***39.61***
10** FusetNet SF5 (HD)	✗	✗	✗	✗	71.44	45.97	29.74
11** SnapNet-R(HD)	✓	✓	✓	✓	73.55	50.07	33.46

SnapNet-R: Consistent 3D Multi-View Semantic Labeling for Robotics, Guerry et al., **ICCV/3DRMS'2017**

3D robot exploration

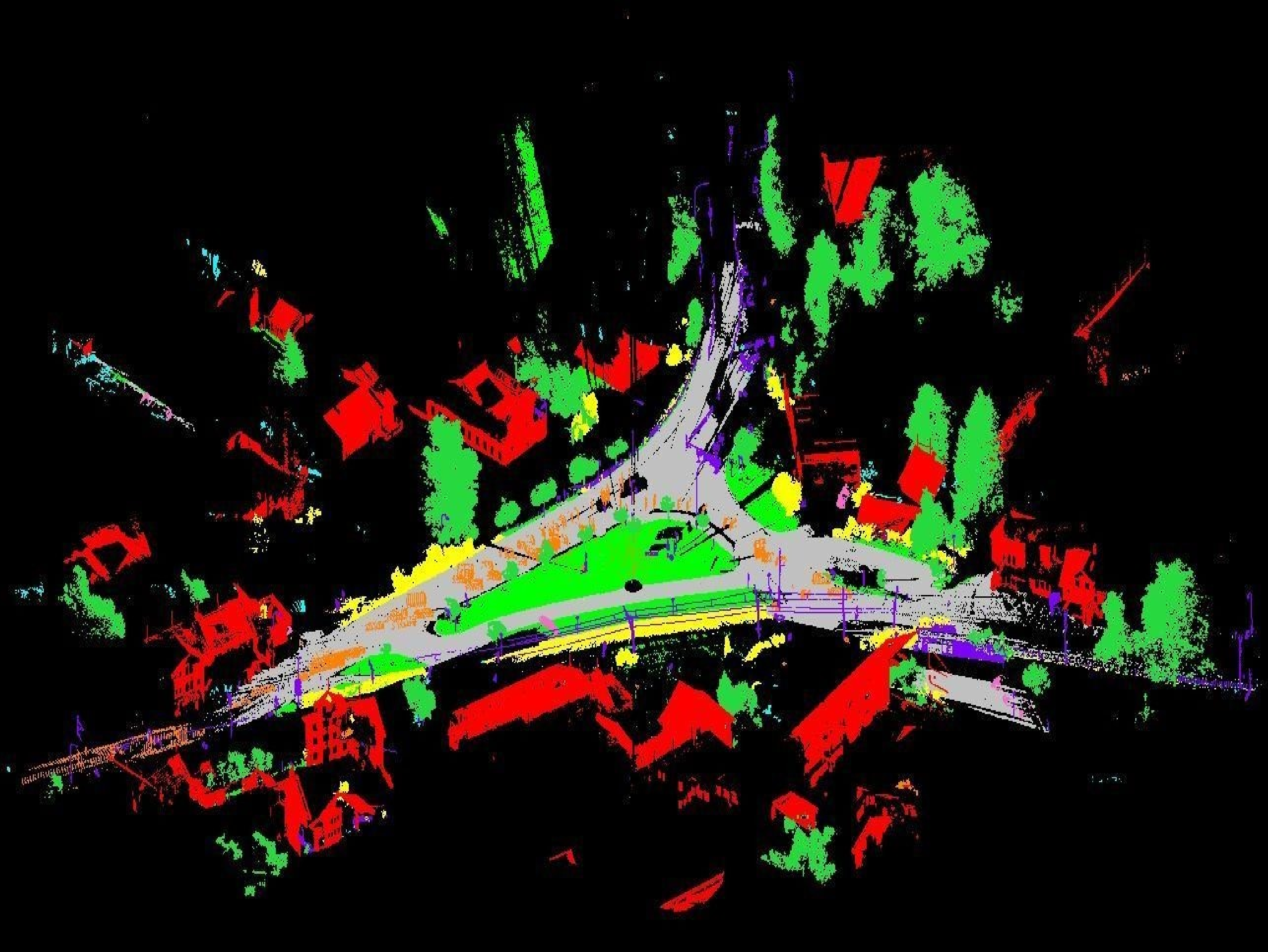
Point-cloud from a single-view: RGB-D data

NYUv2

experiment	OA	MA	IoU
40 classes			
RCNN [17] (RGB-HHA)	60.3	35.1	28.6
FCN 16s [32] (RGB-HHA)	65.4	46.1	34.0
Eigen et al.[12](RGB-D-N)	65.6	45.1	34.1
Context-CRF [31] (RGB-D)	67.6	49.6	37.1
*FuseNet SF3[33] (RGB-D)	66.4	44.2	34.0
*MVCNet-MP [33](RGB-D)	**70.66**	***51.78***	**40.07**
FuseNet SF5 (RGB-D)	62.19	48.28	31.01
SnapNet-R (RGB-D)	***69.20***	**60.55**	***38.33***

13 classes			
Couprie et al.[10] (RGB-D)	52.4	36.2	–
Hermans et al.[24] (RGB-D)	54.2	48.0	–
SceneNet (DHA)[21] (DHA)	67.2	52.5	–
Eigen et al.[12] (RGB-D-N)	75.4	66.9	52.6
*FuseNet SF3 [33] (RGB-D)	75.8	66.2	54.2
*MVCNet-MP [33] (RGB-D)	***79.13***	70.59	***59.07***
Eigen-SF-CRF [35] (RGB-D)	63.6	66.9	–
FuseNet SF5 (RGB-D)	78.41	***72.07***	56.33
SnapNet-R (RGB-D)	**81.95**	**77.51**	**61.78**

SnapNet-R: Consistent 3D Multi-View Semantic Labeling for Robotics, Guerry et al., **ICCV/3DRMS'2017**



Large-Scale Point-Cloud Classif Benchmark / ETHZ

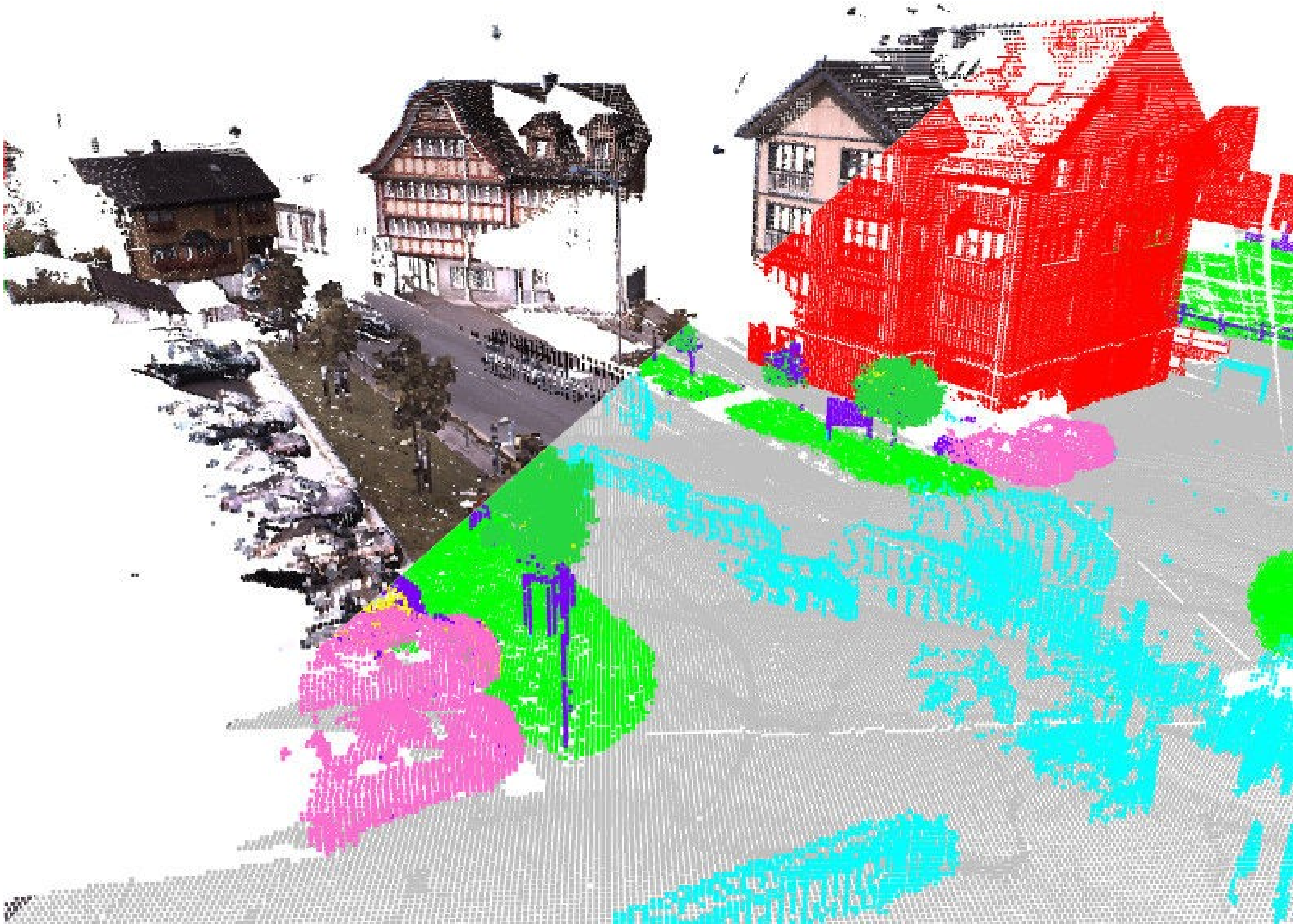
http://semantic3d.net

3D Point-Cloud Semantic Labeling with SnapNet

(with Alexandre Boulch, Joris Guerry and Nicolas Audebert)

3D semantic labeling

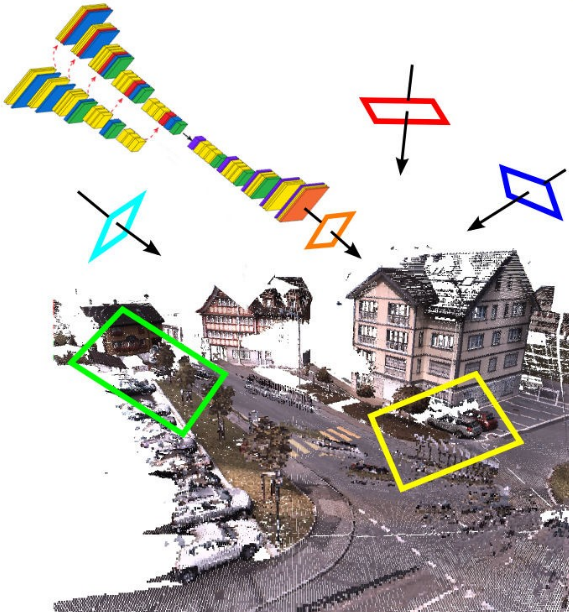
How to understand and classify an environment captured in 3D?
(by LiDAR or photogrammetry)

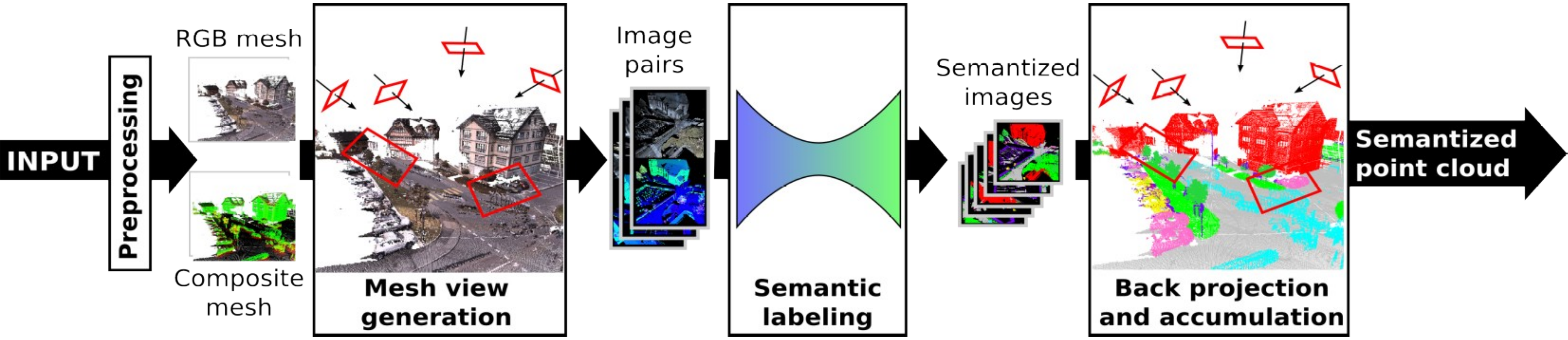


SnapNet for 3D semantic labeling

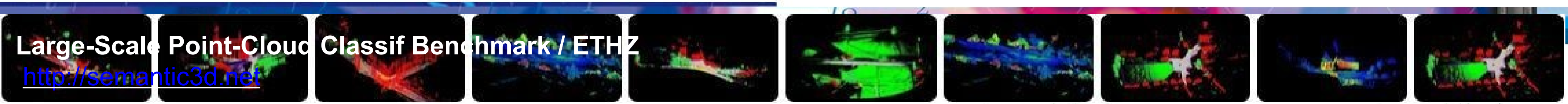
Objective : Label each 3D point with class label

Key-idea : Take snapshots all-over the point cloud, and classify them !

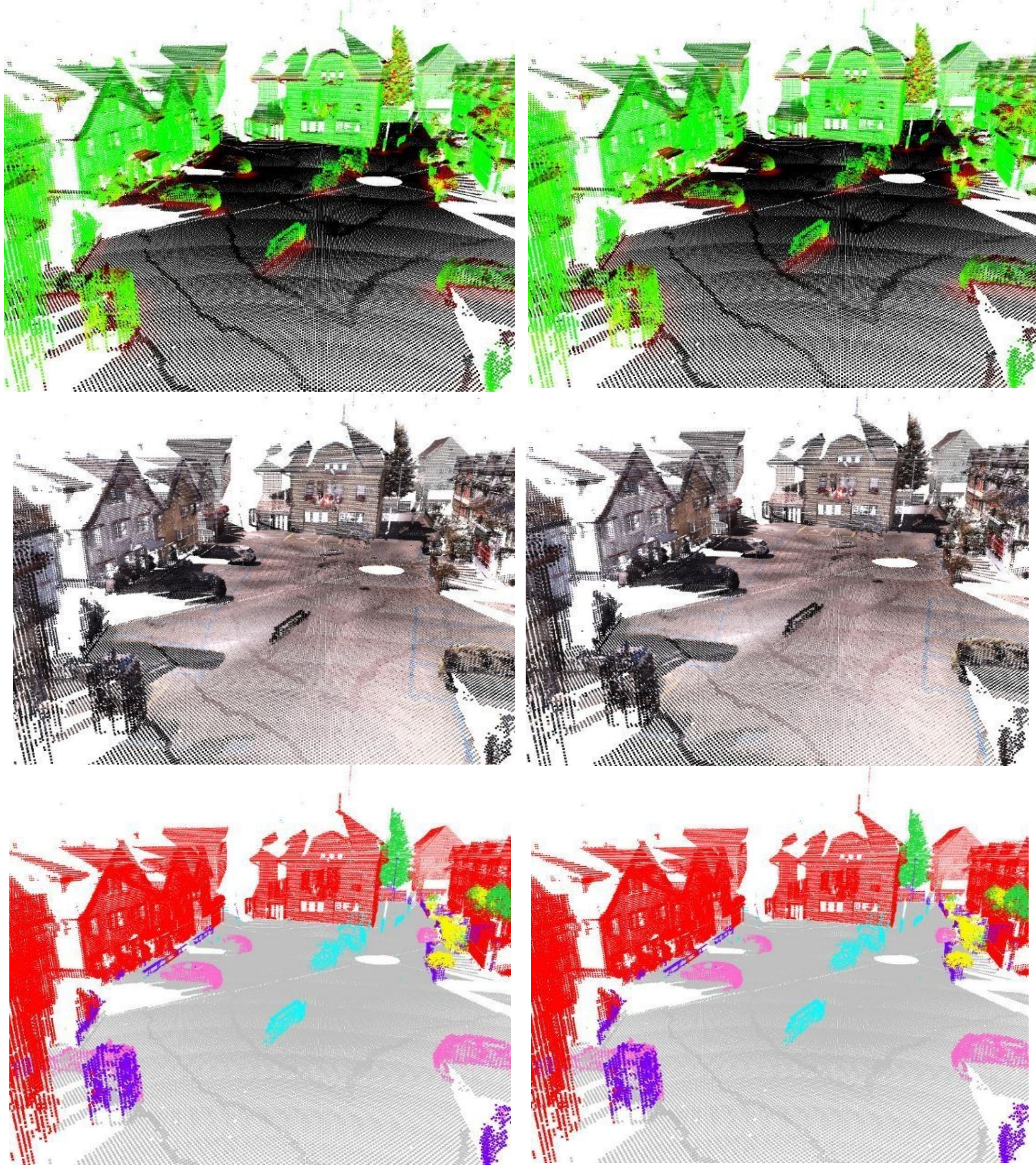


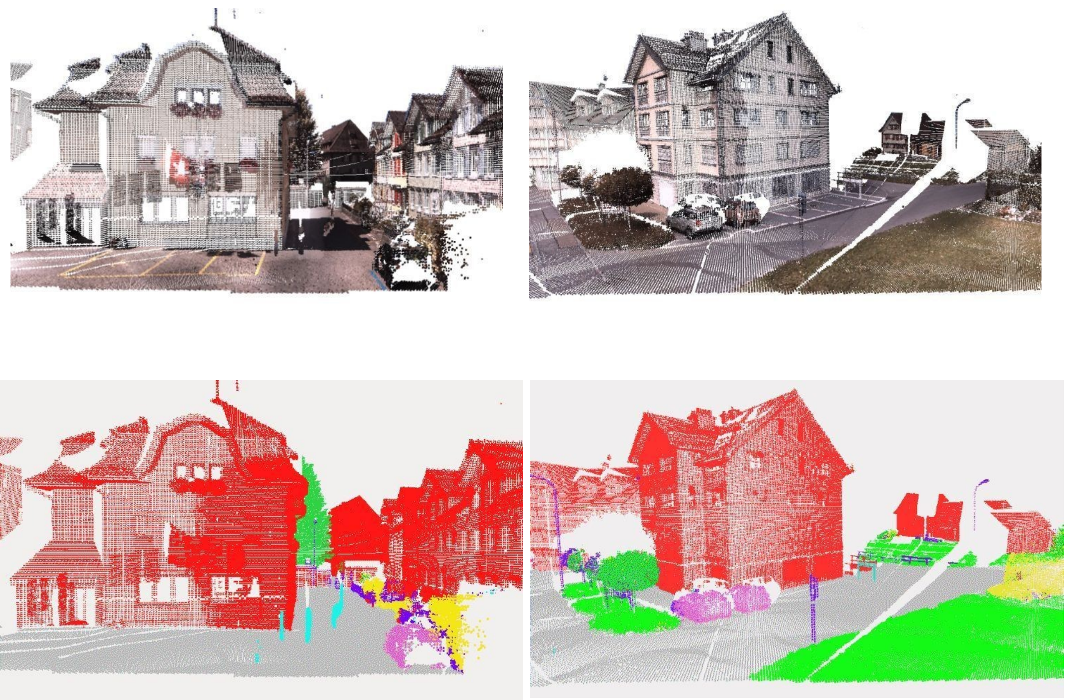


SnapNet : urban classification

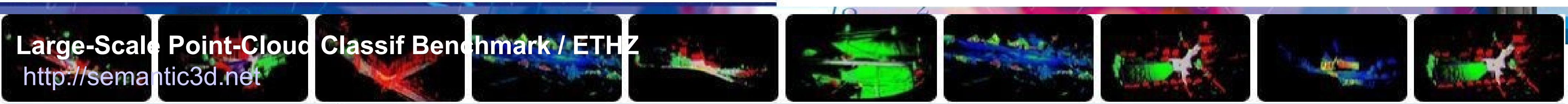


3D urban mapping from LiDAR





SnapNet : urban classification



	Name	A_IoU	OA	[s]	IoU 1	IoU 2	IoU 3	IoU 4	IoU 5	IoU 6	IoU 7	IoU 8
1	SEGCloud	0.613	0.881	1881.00	0.839	0.660	0.860	0.405	0.911	0.309	0.275	0.643
	L. P. Tchapmi, C. B.Choy, I. Armeni, J. Gwak, S. Savarese, SEGCloud: Semantic Segmentation of 3D Point Clouds, International Conference on 3D Vision (3DV), 2017											
2	SnapNet_	0.591	0.886	3600.00	0.820	0.773	0.797	0.229	0.911	0.184	0.373	0.644
	Unstructured point cloud semantic labeling using deep segmentation networks. A. Boulch, B. Le Saux and N. Audebert, Eurographics 3DOR 2017											
3	DeePr3SS	0.585	0.889	0.00	0.856	0.832	0.742	0.324	0.897	0.185	0.251	0.592
	F. Lawin, M. Danelljan, P. Tosteberg, G. Bhat, F. Khan, M. Felsberg. Deep Projective 3D Semantic Segmentation. In , 2017.											
4	3D-FCNN-TI	0.582	0.875	774.00	0.840	0.711	0.770	0.318	0.899	0.277	0.252	0.590
	L. P. Tchapmi, C. B.Choy, I. Armeni, J. Gwak, S. Savarese, SEGCloud: Semantic Segmentation of 3D Point Clouds, International Conference on 3D Vision (3DV), 2017											
5	DLUT_SR	0.563	0.860	1.00	0.953	0.849	0.548	0.296	0.832	0.192	0.320	0.518
	Anonymous submission											
6	TMLC-MSR	0.542	0.862	1800.00	0.898	0.745	0.537	0.268	0.888	0.189	0.364	0.447
	Timo Hackel, Jan D. Wegner, Konrad Schindler: Fast semantic segmentation of 3d point clouds with strongly varying density. ISPRS Annals - ISPRS Congress, Prague, 2016											
7	DeepNet	0.437	0.772	64800.00	0.838	0.385	0.548	0.085	0.841	0.151	0.223	0.423
	Anonymous submission											
8	TML-PCR	0.384	0.740	0.00	0.726	0.730	0.485	0.224	0.707	0.050	0.000	0.150
	Mind the gap: modeling local and global context in (road) networks: Javier Montoya, Jan D. Wegner, Lubor Ladicky, Konrad Schindler. In: German Conference on Pattern Recognition (GCPR), Münster, Germany, 2014											

1: man-made terrain; 2: natural terrain; 3: high vegetation; 4 low-vegetation; 5: buildings; 6: hardscape; 7: scanning artefacts; 8: cars

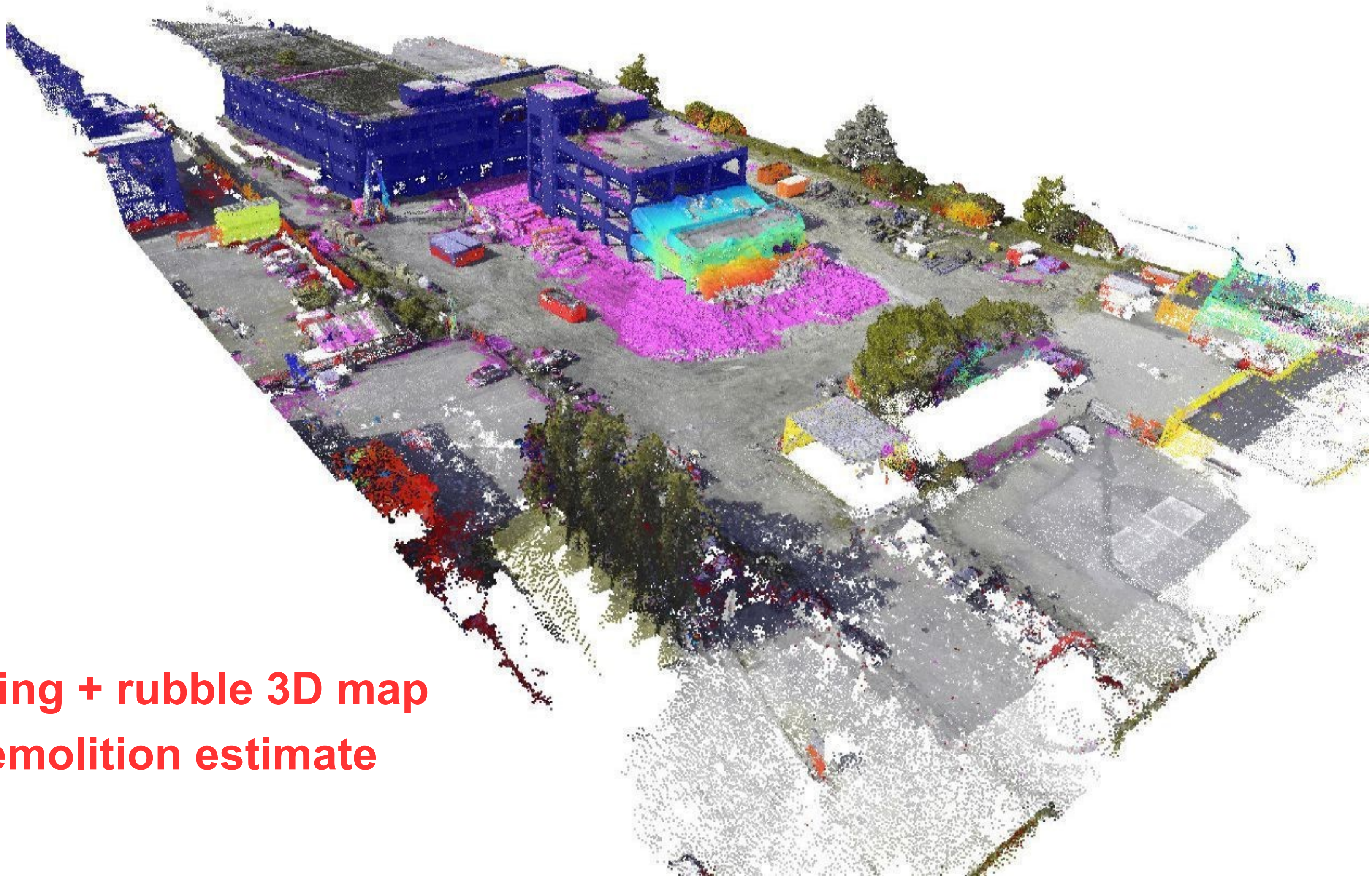
IoU: Intersection over Union; A_IoU: Average IoU; OA: Overall per-pixel Accuracy

Point-cloud semantic labeling using deep segmentation networks, *Alexandre Boulch, Bertrand Le Saux*, *Nicolas Audebert*, **Eurographics/3DOR'2017**

SnapNet : Search-and-rescue classification

Lyon (Fr.) : FP7 Inachus Pilot Test #2 in May 2017

- Point-clouds from micro-UAVs and photogrammetry
- Urban semantizer → buildings, terrain, vegetation…
- Rubble predictor

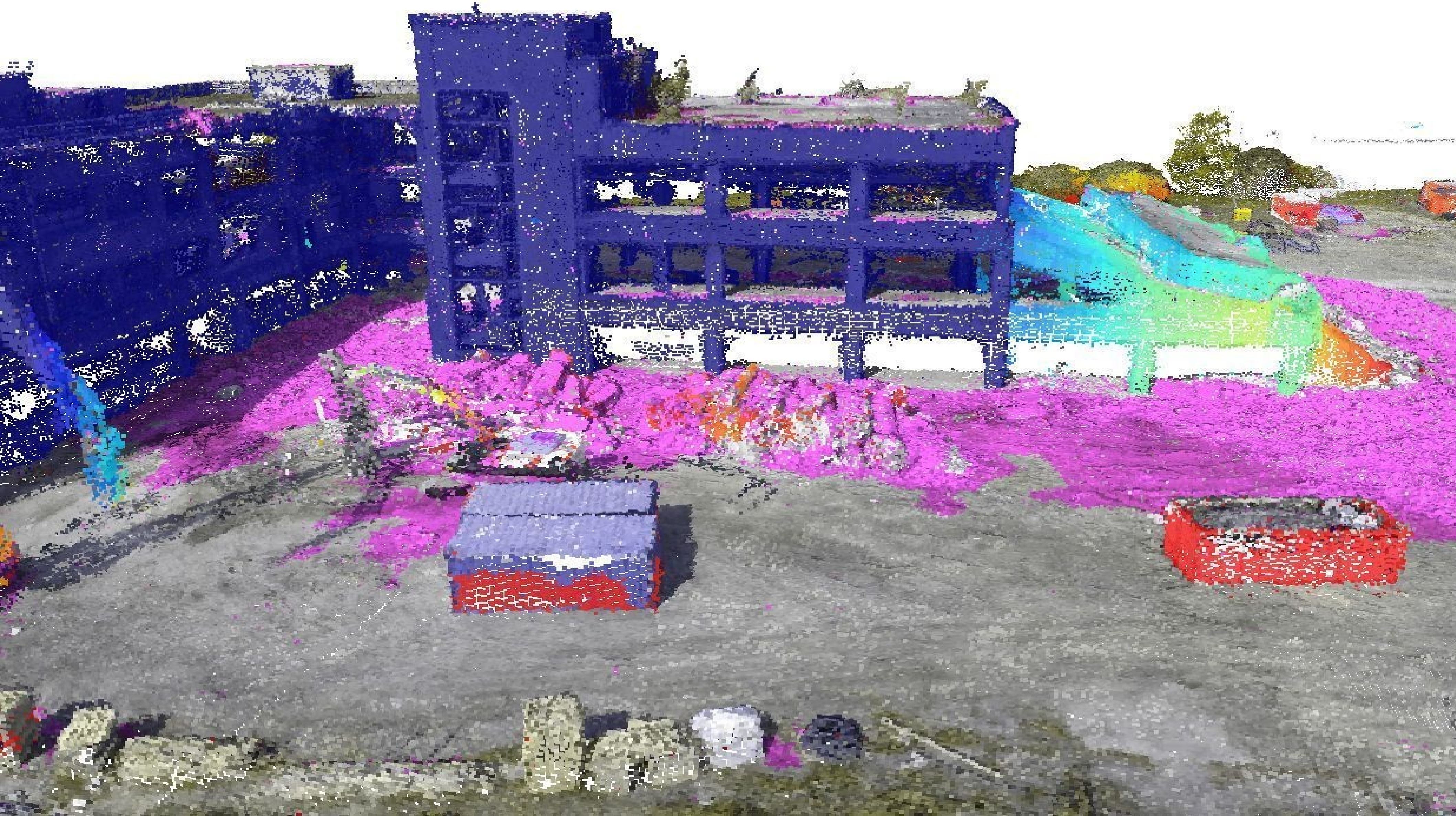


→ **Building + rubble 3D map with demolition estimate**

SnapNet : Search-and-rescue classification

Lyon (Fr., Inachus Pilot Test #2 in May 2017) :

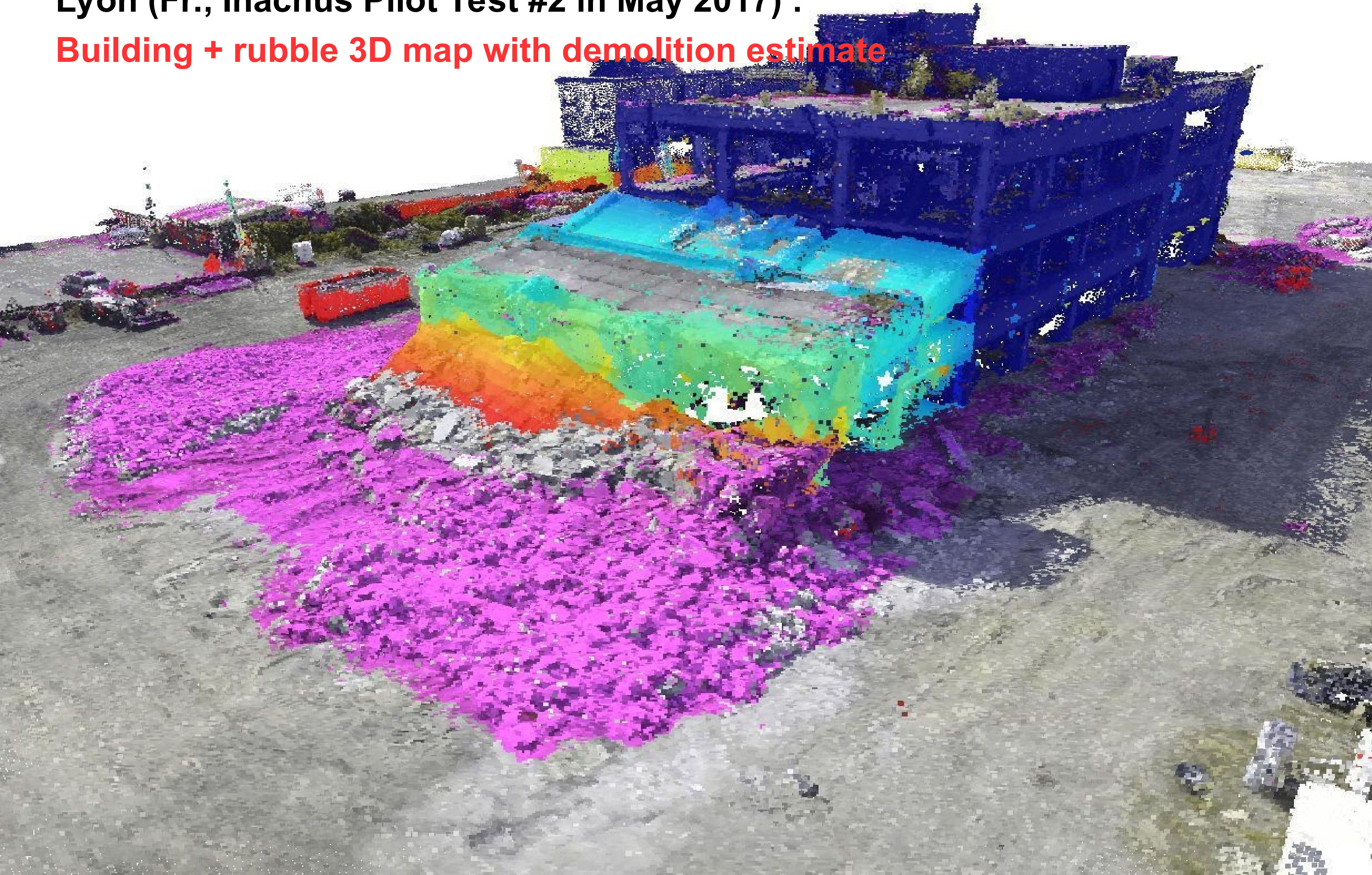
Building + rubble 3D map with demolition estimate



SnapNet : Search-and-rescue classification

Lyon (Fr., Inachus Pilot Test #2 in May 2017) :

Building + rubble 3D map with demolition estimate



Concluding remarks

Concluding remarks

Overall objective : Understanding the environment.

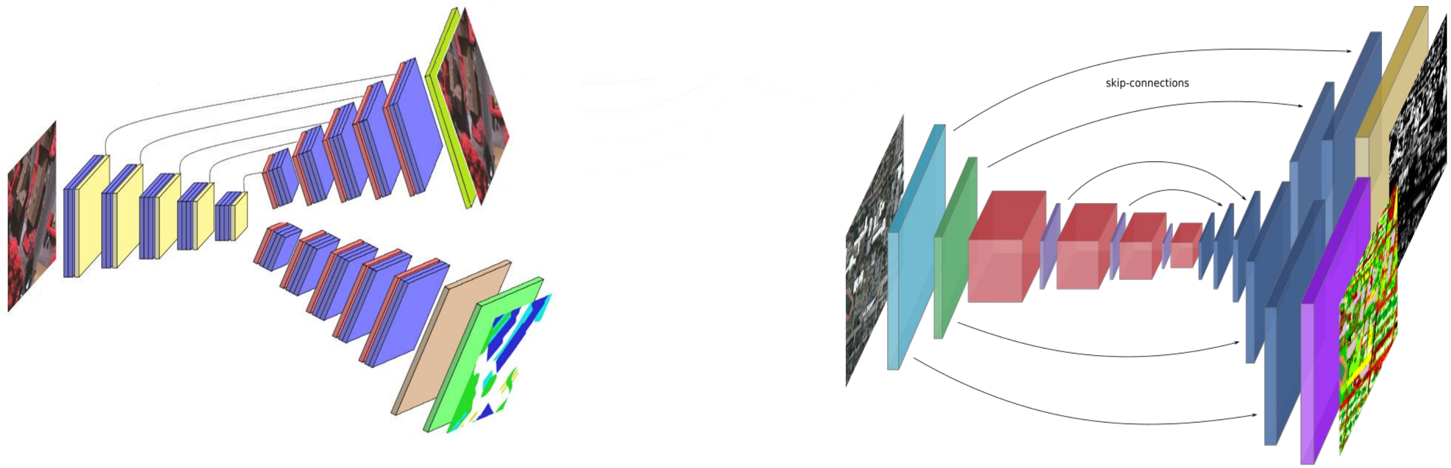
A few common threads :

- Mostly **discriminative models**, chosen for efficiency, using strong *a priori* information to cope with the scarcity of data

- Use of **multiple viewpoints** on the scene (more and more, randomized) to recover 3D structure

- Leveraging **multimodal** information and data to get better analysis, and in particular combining appearance and 3D information

Challenge #2 : large scale scene understanding



Short-term: Building better models

- Multi-task learning for self supervision[1]
- Weak-learning from imprecise or wrong reference (not human-generated)
- Interactive and active learning[2] for making more robust models and predictions
- Multi-temporal analysis to monitor Earth activity

➢ Mapping + DSM generation: https://github.com/marcelampc/aerial_mtl/

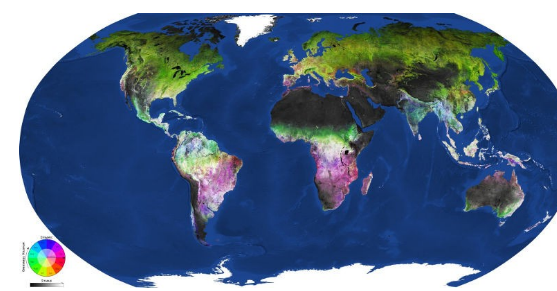
[1] with M. Carvalho et al., J. Castillo-Navarro et al.
[2] with G. Lenczner et al.

Multitask Learning of Height and Semantics from Aerial Images,
M. Carvalho, B. Le Saux, P. Trouvé, A. Almansa, F. Champagnat, **GRSL'19**

Challenge #2 : large scale scene understanding

Middle-term:
Improving the generalization of Earth observation models

- Semi-supervised and self-supervised learning to leverage unlabeled data[1]
- Learning from synthetic data / synthesize data for training



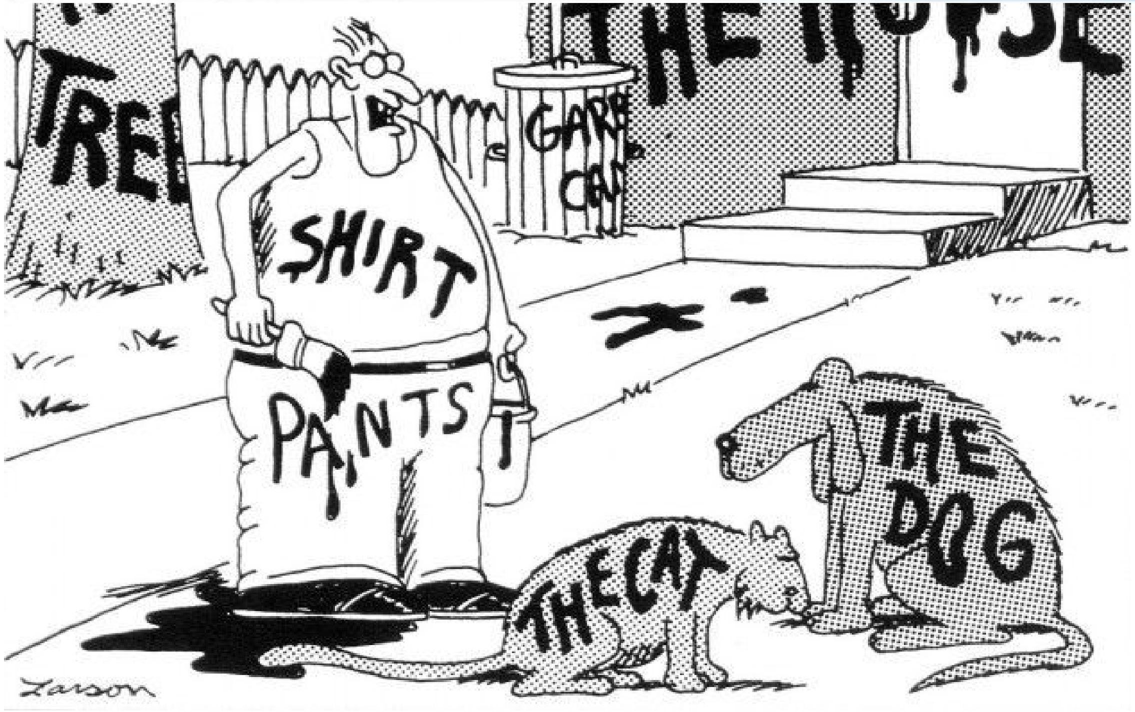
(c) ESA/Copernicus

Long term: large scale highly-multimodal and 3D Earth observation → **Digital Twin Earth**

- Geo-spatial analysis, by leveraging geo-referenced multi-source data
- Large-scale 3D from space, including multi-temporal 3D analysis

[1] with J. Castillo-Navarro, A. Boulch & S. Lefèvre

Questions ?



(c) Gary Larsson / The Far Side

Mail : bertrand.le.saux@esa.int

Web : http://blesaux.github.io