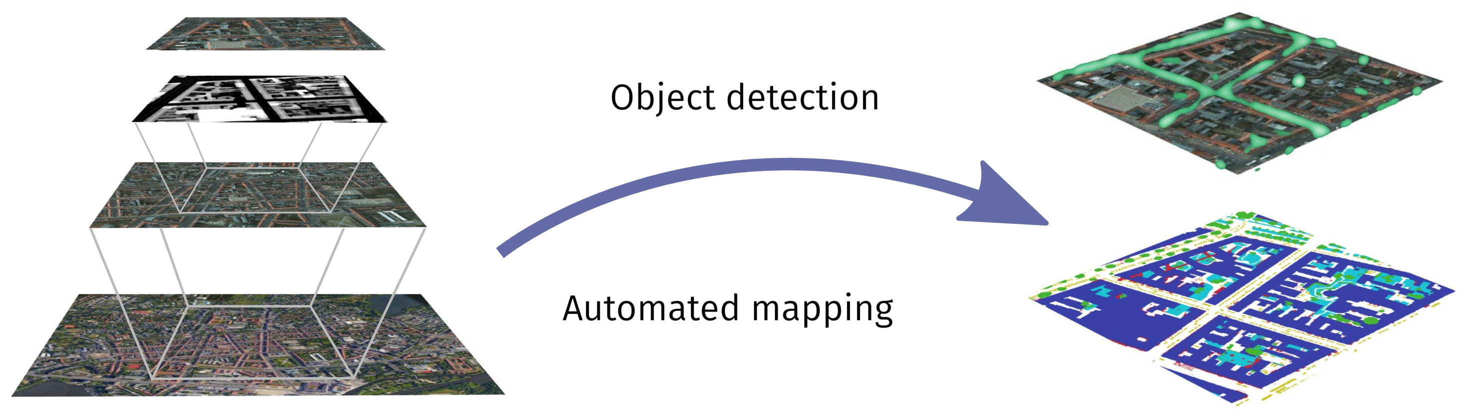
Joint Learning from Earth Observation and OpenStreetMap Data to Get Faster Better Semantic Maps

Nicolas Audebert[1,2], Bertrand Le Saux[1], Sébastien Lefèvre[2]

ONERA - *The French Aerospace Lab*[1], Univ. Bretagne-Sud - IRISA[2]

CVPR 2017, July 21-26, Honolulu

Automated semantic semantic mapping



Using Earth Observation data for :
- Urban mapping: buildings, roads, vehicles, vegetation...
- Large-scale mapping: water bodies, agricultural areas, urban monitoring...
- Object detection: large vehicles, anormal vegetation...

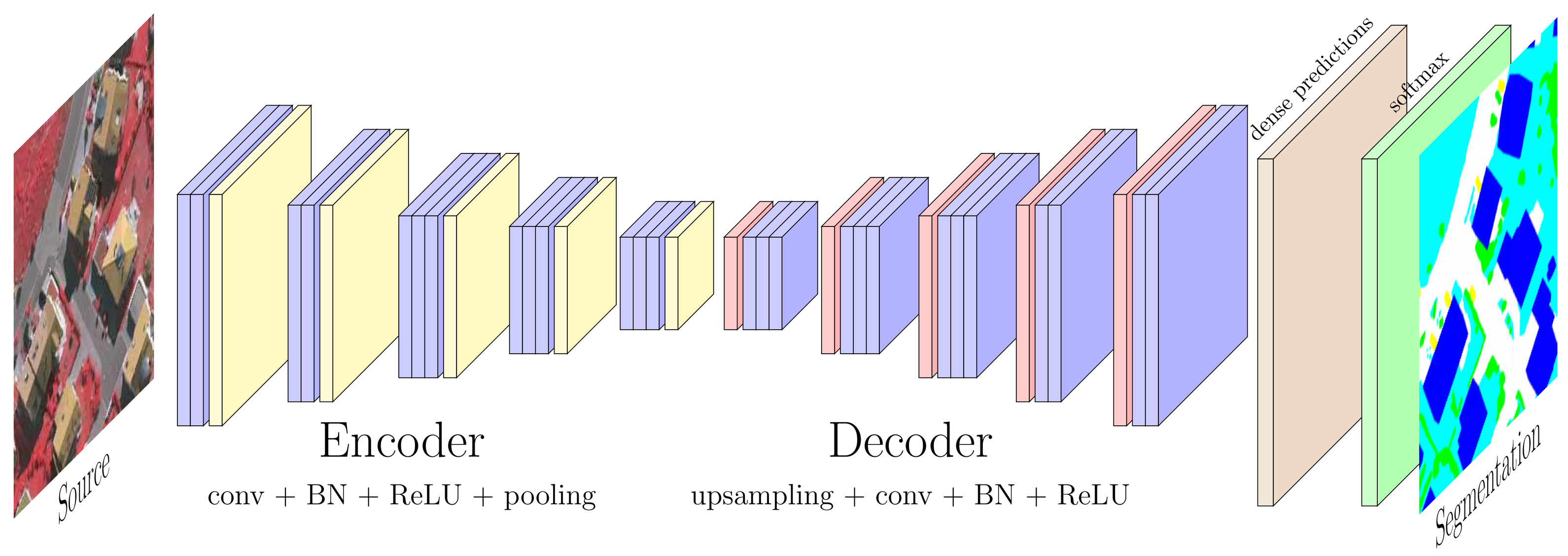
Machine learning has proved itself useful:
- On optical aerial and satellite images
- Using deep convolutional neural networks
- By mixing heterogeneous data, including geographical priors

Deep learning

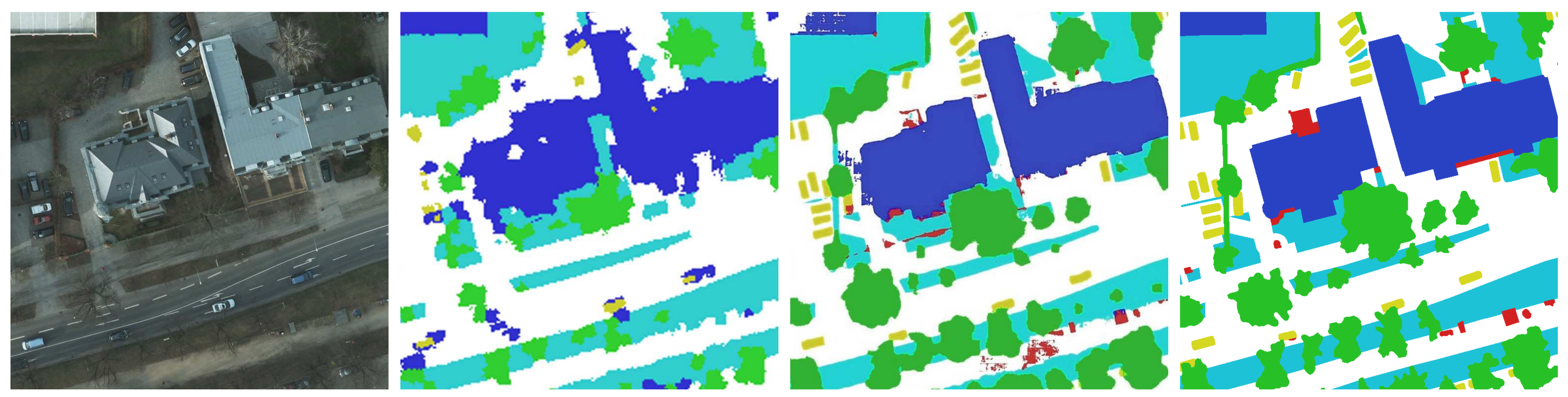
Convolutional Neural Networks (CNN) :
- State-of-the-art for semantic segmentation
- Excellent results for road mapping from aerial images [1]

Fully Convolutional Networks for remote sensing data

- FCN preserve the 2D spatial structure of the data
- Perform a dense pixel-wise classification
- Capture spatial relationships between pixels



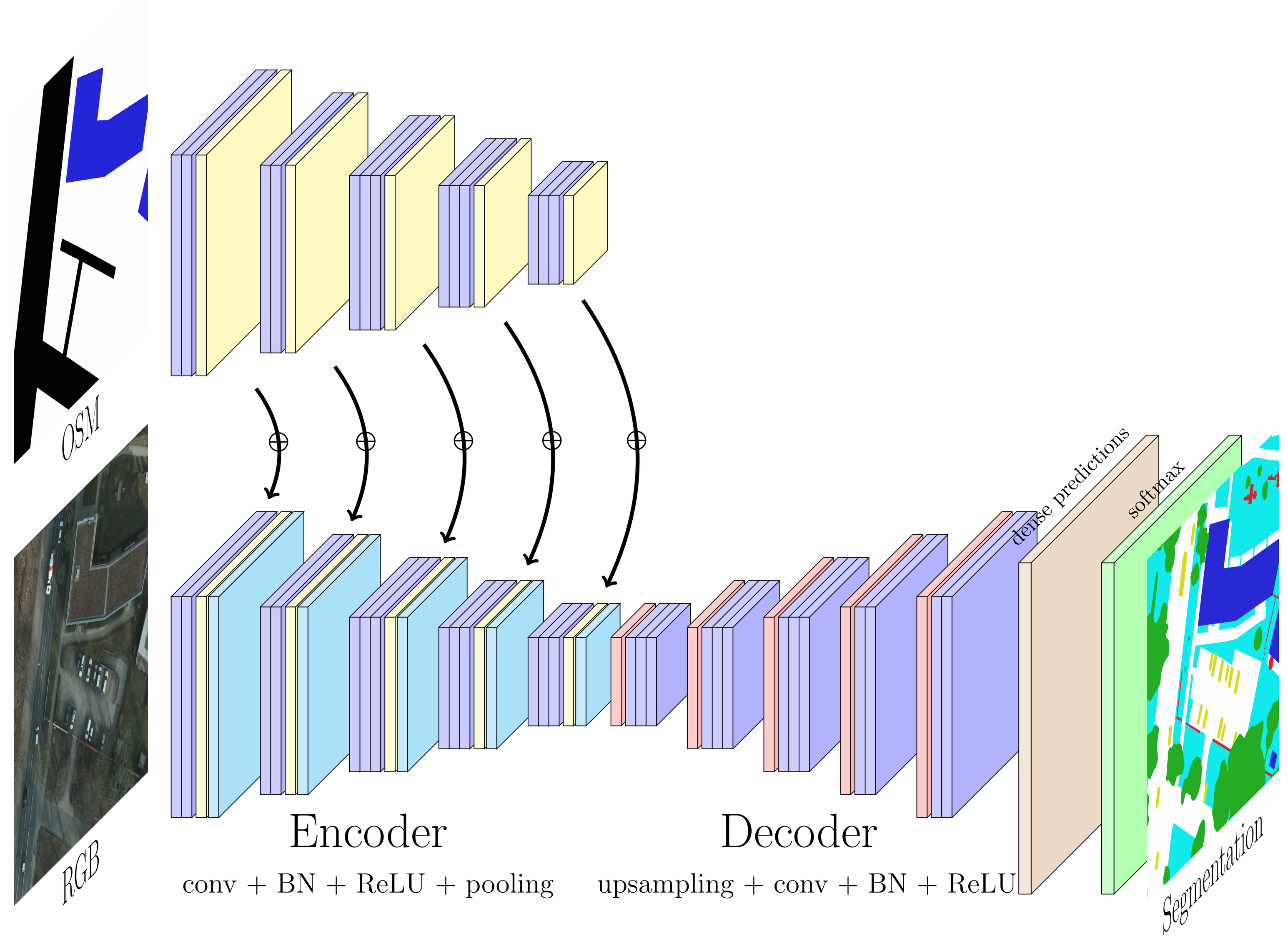
SegNet architecture, designed for autonomous vehicles and adapted to remote sensing [2]



Example of a predicted semantic map on the ISPRS Potsdam dataset.

Fusing optical and geographic data

Heterogeneous data sources can be processed parallely and fused to improve the semantic mapping [3], [4]. The FuseNet model [5] is here used to learn jointly from RGB images and OpenStreetMap crowdsourced annotations.

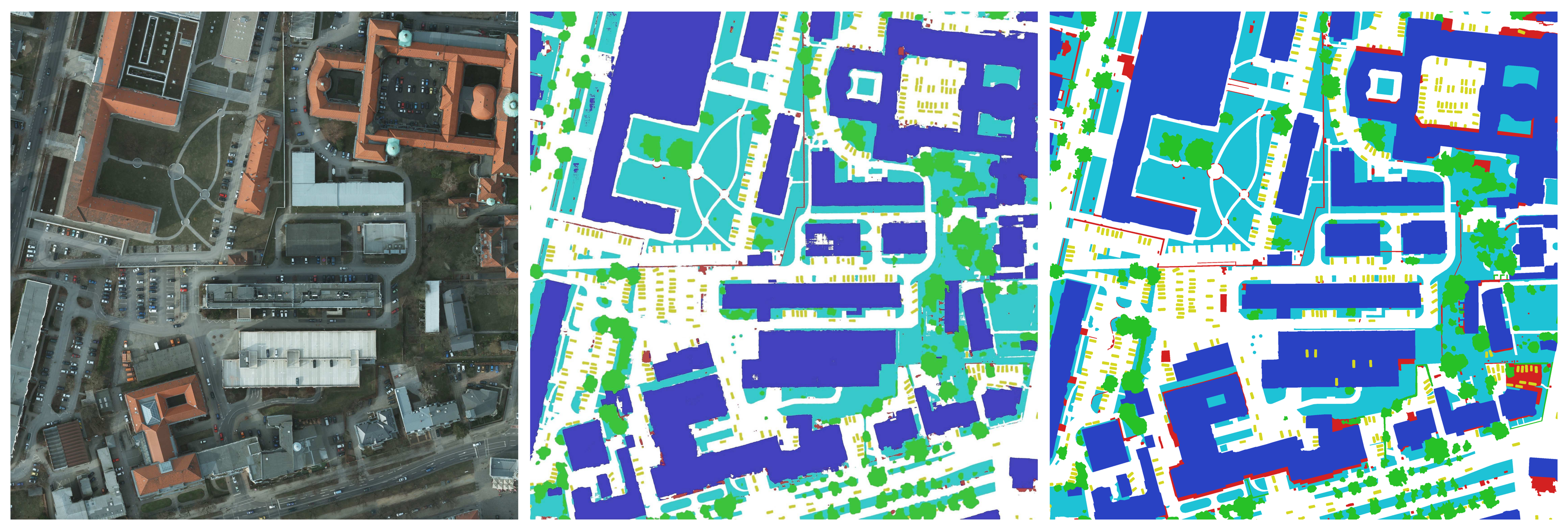


Results

Comparison on the ISPRS Potsdam aerial images dataset

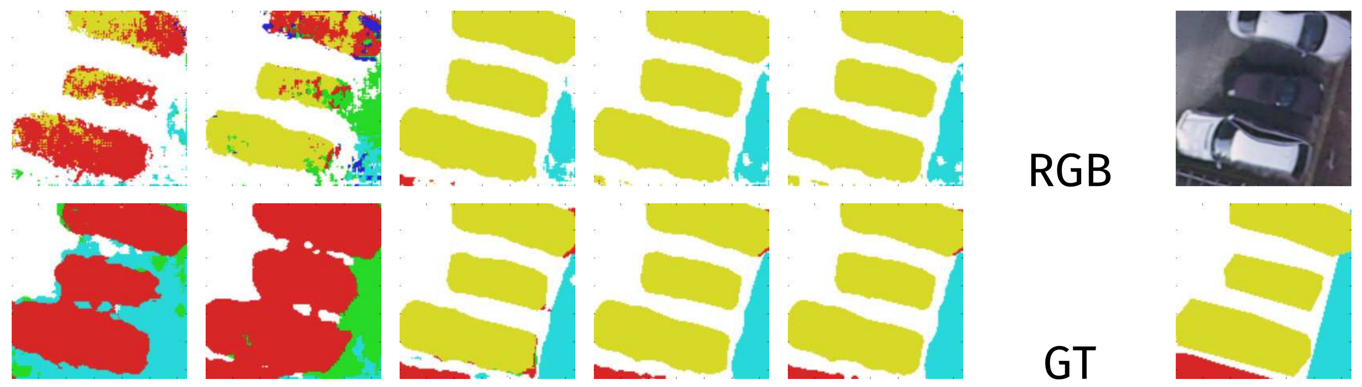
Methode	Roads	Buildings	Low veg.	Trees	Vehicles	Accuracy
RF IRRGB	77.0 %	79.7 %	73.1 %	59.4 %	58.8 %	74.2 %
SegNet RGB	93.0 %	92.9 %	85.0 %	85.1 %	95.1 %	89.7 %
RF IRRGB+OSM	85.6 %	92.4 %	73.8 %	59.5 %	67.6 %	80.9 %
CR RGB+OSM	93.9 %	92.8 %	85.1 %	**85.2** %	95.8 %	90.6 %
FuseNet	**95.3** %	**95.9** %	**86.3** %	85.1 %	**96.8** %	**92.3** %

→ the largest improvements are obtained on the buildings and roads classes which are clearly identified in OSM. Other classes benefit from the additional information on hard patches (vegetation and vehicles).



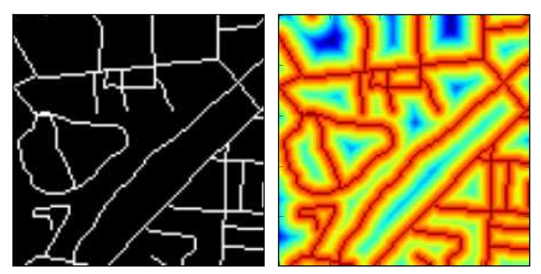
(white: roads, blue: buildings, cyan: low vegetation, green: trees, yellow: vehicles, red: clutter)

Learn faster with better structures



The model converges faster using OpenStreetMap data as it can focus on the harder part of the RGB data. FuseNet reaches the same accuracy as the RGB SegNet with 25% less iterations and a mean loss of 0.39 vs. 0.45 on the test set, which indicates better generalization abilities.

Pre-processing OpenStreetMap data



The geo-information can be rasterized either as sparse binary map or as a dense distance map. Experiments show that binary information is sufficient, as convolutional layers will propagate the information.

Conclusion

- We use deep neural networks to learn semantic mapping of large areas at multiple scales.
- Data fusion allows us to learn jointly from RGB images and structured geo-information from OpenStreetMap.
- Source code and pre-trained models:

`https://github.com/nshaud/DeepNetsForEO`

References

[1] V. Mnih and G. E. Hinton, “Learning to Detect Roads in High-Resolution Aerial Images”, in *Computer Vision – ECCV 2010*, ser. Lecture Notes in Computer Science 6316, Springer Berlin Heidelberg, Sep. 2010, pp. 210–223.

[2] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–1, 2017.

[3] N. Audebert, B. Le Saux, and S. Lefèvre, “Semantic Segmentation of Earth Observation Data Using Multimodal and Multi-scale Deep Networks”, in *Computer Vision – ACCV 2016*, Springer, Cham, Nov. 2016, pp. 180–196.

[4] N. Audebert, B. Le Saux, and S. Lefèvre, “Fusion of heterogeneous data in convolutional networks for urban semantic labeling”, in *2017 Joint Urban Remote Sensing Event (JURSE)*, Mar. 2017, pp. 1–4.

[5] C. Hazirbas, L. Ma, C. Domokos, and D. Cremers, “FuseNet: Incorporating Depth into Semantic Segmentation via Fusion-Based CNN Architecture”, in *Computer Vision – ACCV 2016*, Springer, Cham, Nov. 2016, pp. 213–228.

Nicolas Audebert's work is funded by the research project NAOMI, a Total and ONERA collaboration that aims to develop novel remote sensing techniques for industrial applications.