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]

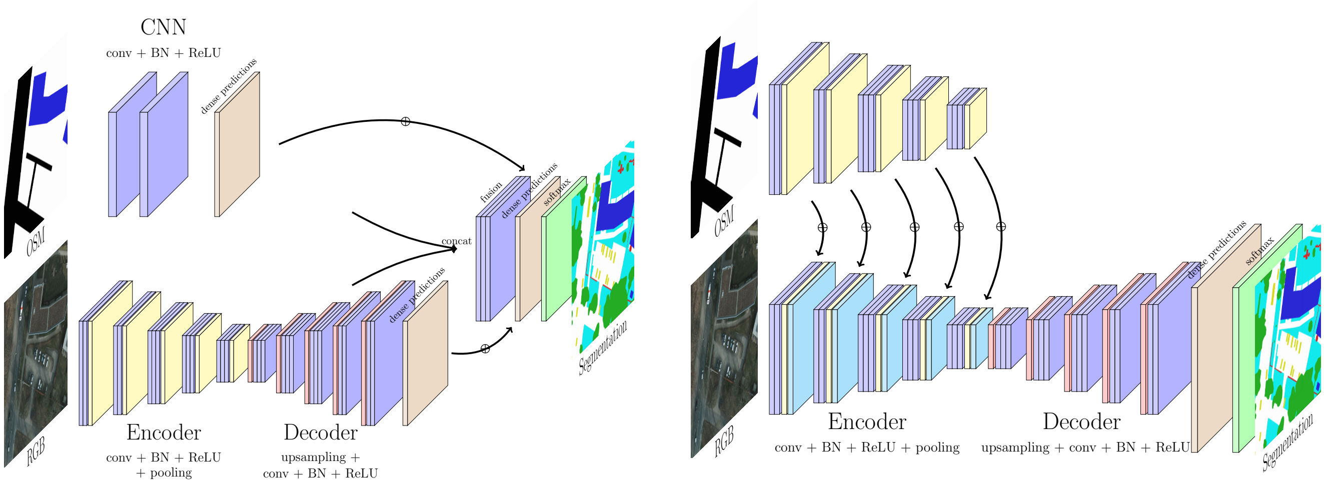
[1] DTIS, ONERA, Université de Paris Saclay, F-91123 Palaiseau - France
[2] Univ. Bretagne-Sud, UMR 6074, IRISA, F-56000 Vannes, France

{nicolas.audebert,bertrand.le_saux}@onera.fr, sebastien.lefevre@irisa.fr

1 Introduction

In this work, we investigate the use of OpenStreetMap data for semantic labeling of Earth Observation images. Deep neural networks have been used in the past for remote sensing data classification from various sensors, including multispectral, hyperspectral, SAR and LiDAR data. While OpenStreetMap has already been used as ground truth data for training such networks, this abundant data source remains rarely exploited as an input information layer. In this paper, we study different use cases and deep network architectures to leverage OpenStreetMap data for semantic labeling of aerial and satellite images. Especially, we look into fusion based architectures and coarse-to-fine segmentation to include the OpenStreetMap layer into multispectral-based deep fully convolutional networks. We illustrate how these methods can be successfully used on the ISPRS Potsdam dataset. We show that OpenStreetMap data can efficiently be integrated into the vision-based deep learning models and that it significantly improves both the accuracy performance and the convergence speed of the networks.

2 Method



(a) Optical and OSM data fusion using residual correction [1]. (b) FuseNet [3] architecture applied to optical and OSM data.

FIGURE 1 – Deep learning architectures for joint processing of optical and OpenStreetMap data.

In this work, we investigate two architectures for optical/OSM data fusion : the residual correction from [1] and the FuseNet architecture [3]. Both models are based on the SegNet semantic segmentation architecture [2] as illustrated in figures 1a et 1b. We validate the data fusion schemes on the ISPRS Potsdam dasat [4], a 5cm/px very high resolution IRRGB semantic labeling dataset. We download and generate the corresponding OpenStreetMap rasters with the footprints for roads, buildings, vegetation and water bodies. We train our models on the ISPRS Potsdam dataset in an end-to-end fashion, following the guidelines from [1]. Our results are shown in tableau 1 with a visual assessment in figure 2.

TABLE 1 – Test results on the ISPRS Potsdam dataset (pixel-wise overall accuracy and F1 score per class).

Method	imp. surfaces	buildings	low veg.	trees	cars	Overall
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
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**

As could be expected, the inclusion of OSM data improves the semantic labeling performance, with significant improvements on "road" and "building" classes. This is not surprising considering that those classes already have a representation in the OSM layers which can help disambiguating predictions coming from the optical source. OSM data already covers the majority of the roads and the buildings, therefore simplifying the inference of the "impervious surface" and "building" classes. OSM data also helps discriminating between buildings and roads that have similar textures.

3 Conclusion

In this work, we showed how to integrate ancillary GIS data in deep learning-based semantic labeling. We presented two methods : one for coarse-to-fine segmentation, using deep learning on RGB data to refine OpenStreetMap semantic maps. We validated our methods on the ISPRS Potsdam 2D Semantic Labeling Challenge on which we increase our semantic labeling overall accuracy by 2.5% by integrating OpenStreetMap data in the learning process. Our findings show that GIS sparse data can be leveraged successfully for semantic labeling on those two use cases, as it improves significantly the classification accuracy of the models. We think that using crowdsourced and open GIS data is an exciting topic of research, and this work provides new insights on how to use this data to improve and accelerate learning based on traditional sensors.

Références

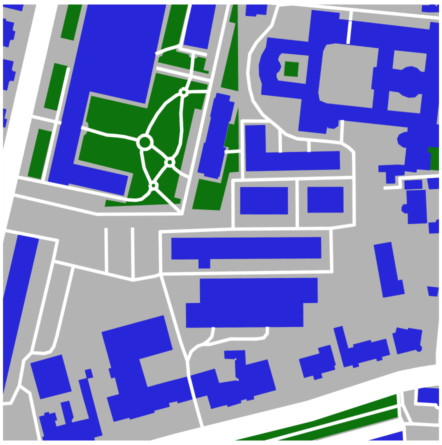
[1] N. Audebert, B. Le Saux, and S. Lefèvre. Semantic Segmentation of Earth Observation Data Using Multimodal and Multi-scale Deep Networks. In *Computer Vision – ACCV 2016*, pages 180–196. Springer, Cham, Nov. 2016.

[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*, PP(99) :1–1, 2017.

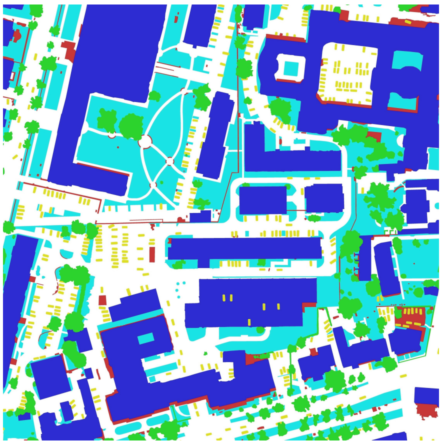
[3] C. Hazirbas, L. Ma, C. Domokos, and D. Cremers. FuseNet : Incorporating Depth into Semantic Segmentation via Fusion-Based CNN Architecture. In *Computer Vision – ACCV 2016*, pages 213–228. Springer, Cham, Nov. 2016.

[4] F. Rottensteiner, G. Sohn, J. Jung, M. Gerke, C. Baillard, S. Benitez, and U. Breitkopf. The ISPRS benchmark on urban object classification and 3d building reconstruction. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1 :3, 2012.

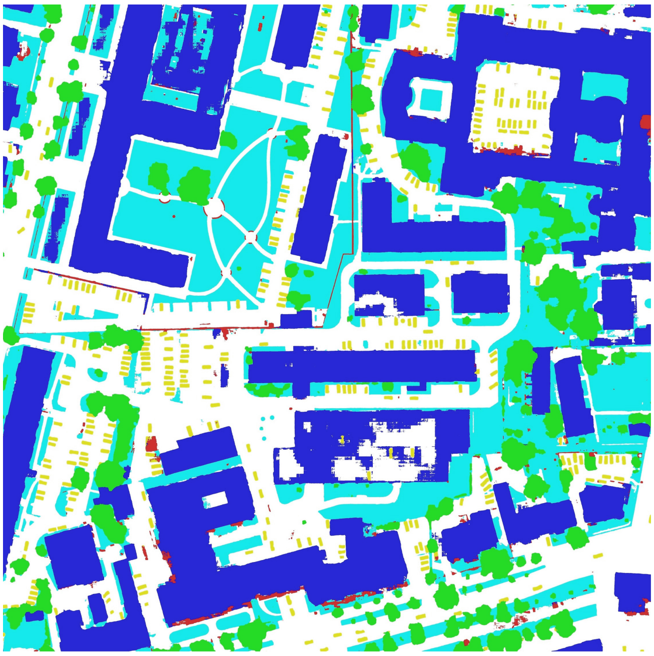
(a) RGB input



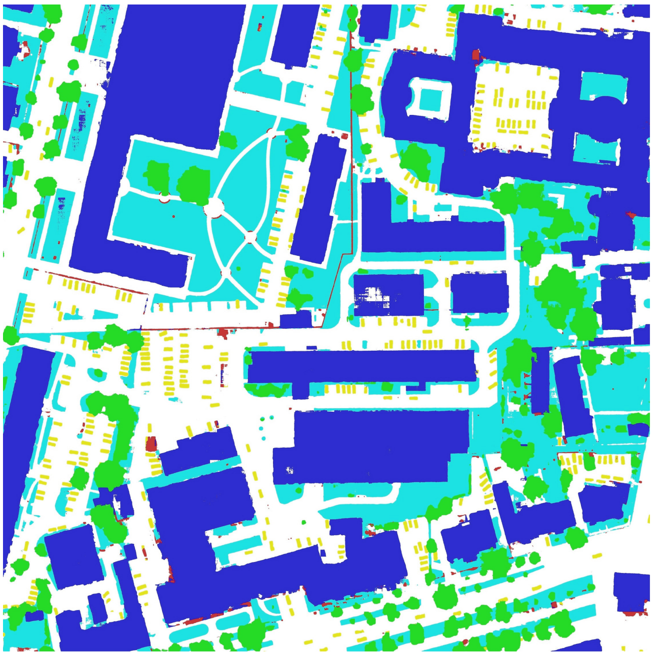
(b) OSM input



(c) Target ground truth



(d) SegNet (RGB)



(e) FuseNet (OSM+RGB)

FIGURE 2 – Excerpt from the classification results on Potsdam