Deep learning on graph for semantic segmentation of point cloud

Alexandre Cherqui

Signal Processing Laboratory 2 (LTS2), EPFL

Introduction

In order to better identify objects, we can use phtogrammetry to combine different 2D points of view and thus create a 3D representation which contains information such as the structures or the heights of objects, or highlight the hidden ones. Semantic segmentation consists in performing a dense labelling of each point. Because graphs enable an efficient representation of the 2D manifold embedded in the 3D space, they are in essence well suited for this application since they allow efficient computations and to capture the local neighboorhood of each point.

Model

1. To build the graph, we first mesh the point cloud as in figure 1 so that each edge between the tops of the generated triangles is a link between the corresponding nodes of the graph which weight is defined by: $w_{i,j} = \exp\left(-\frac{d_{i,j}^2}{2\sigma^2}\right)$, where $d_{i,j}$ is the euclidian distance between vertices i and j, and σ is the scale hyper-parameter.

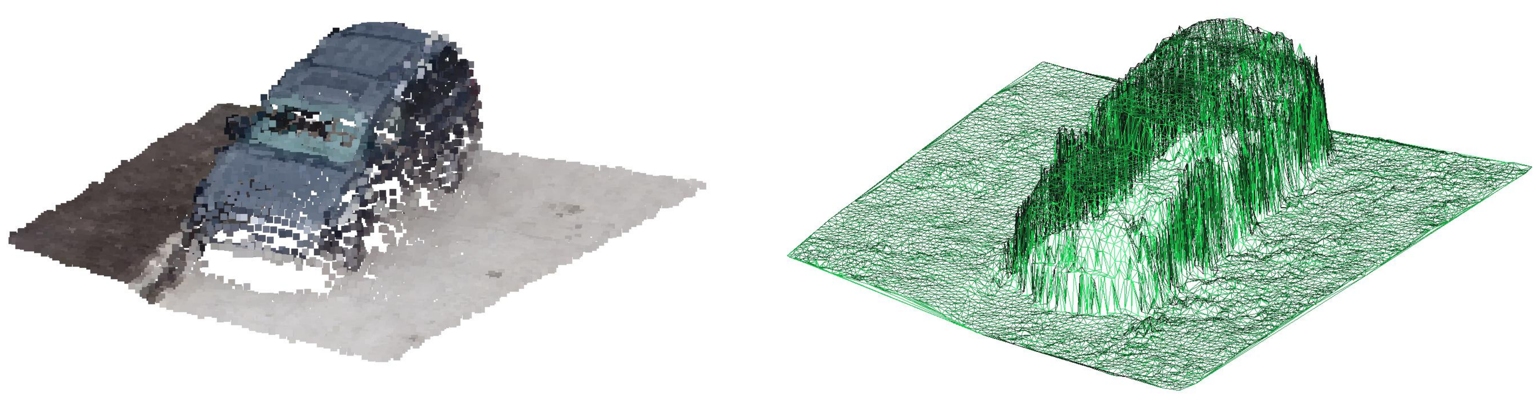
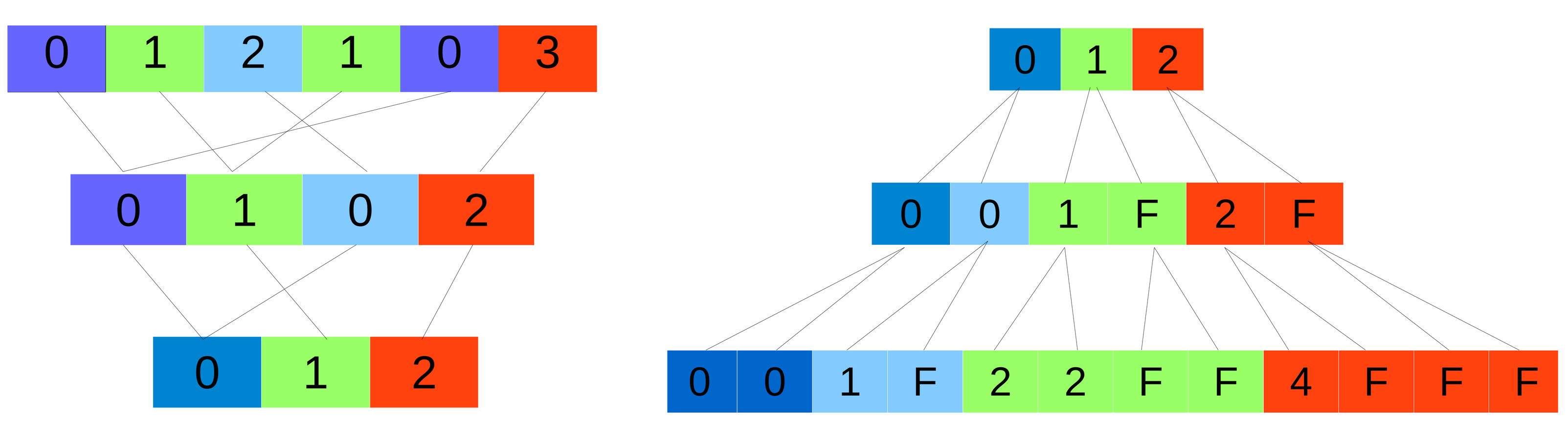


Figure 1 : Mesh generation on a car.

2. Graph convolutions, first defined in the spectral domain [1], can efficiently be done in the spatial domain through the use of Chebychev polynomials T_j [2] which aim at defining the learnable filters. Thus, the graph convolution of a signal $s \in \mathbb{R}^n$ and a filter $x \in \mathbb{R}^n$ can be defined as: $s *_{\mathcal{G}} x = \sum_{j=0}^{K-1} \theta_j T_j(L) s$ where θ_j are learnable parameters and L is the Laplacian of the graph.

3. Then, in order to easily perform the pooling operation, we need to *build a binary tree* such as in figure 2.



(a) Match nodes with respect to their edges weights for the different levels of coarsening

(b) Reorder the nodes so that the union of two matched neighbors from layer to layer forms a binary tree (add fake nodes F if needed)

Figure 2 : Form a binary tree to ease the pooling operation.

4. For our model, we used the architecture depicted in figure 3.

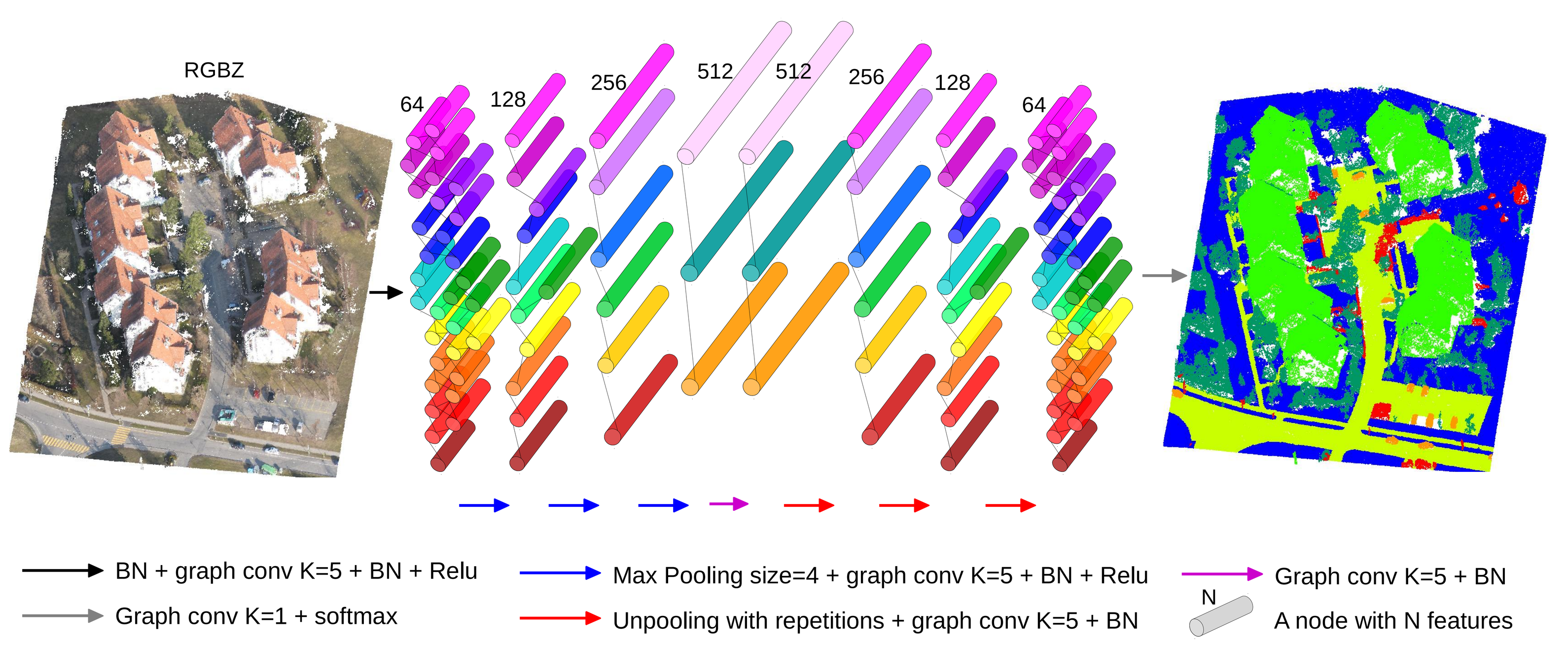


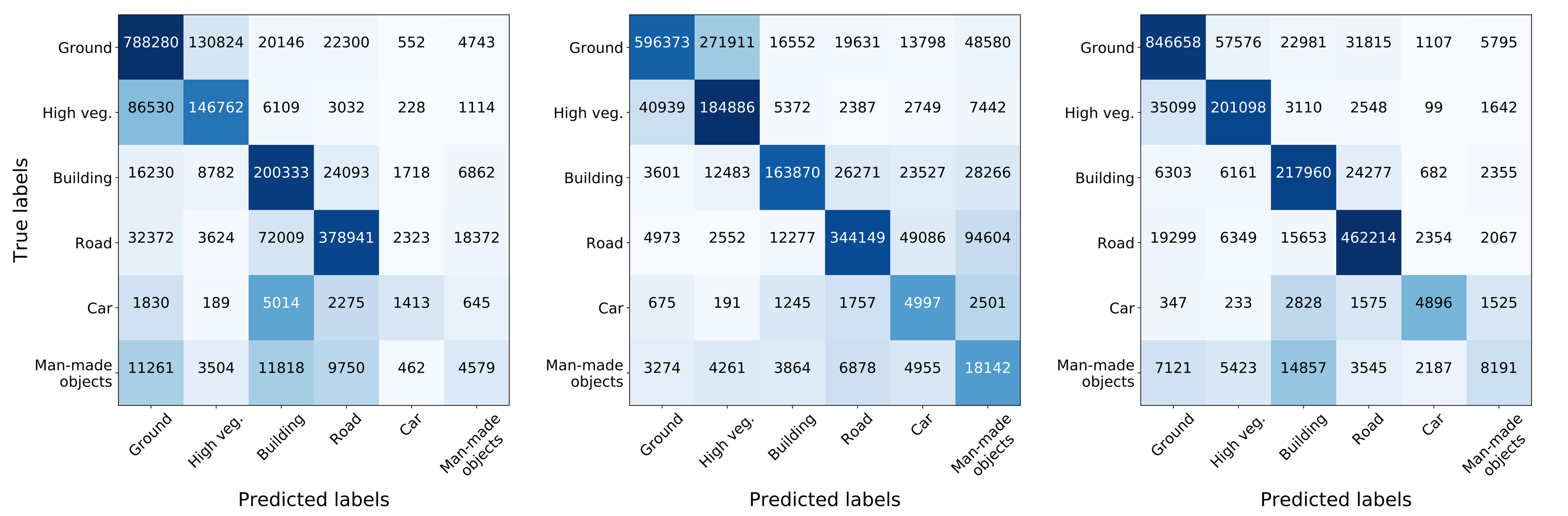
Figure 3 : Model architecture. Spectral distances between colors are related to spatial distances between intra- and inter-layers real nodes.

Results

Table 1 and figure 4 enable to quantitatively compare our model with the used baselines (Random Forest and XGBoost), while figure 5 allows a qualitative comparison.

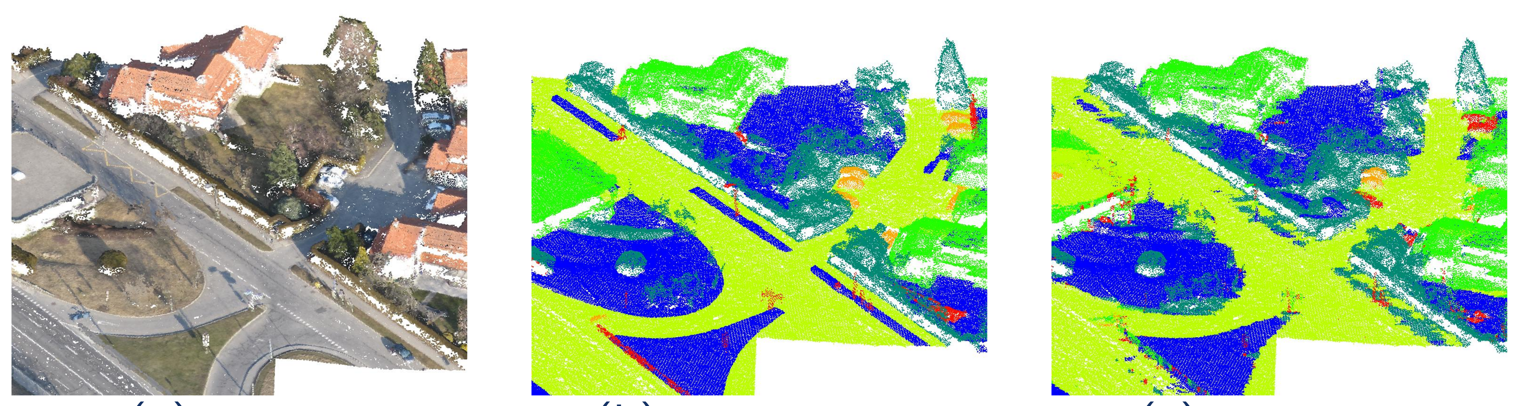
Performances	Overall accuracy (in %)	Mean accuracy (in %)
Random Forest	74.93	52.92
XGBoost	64.68	59.44
Our model	85.85	68.09

Table 1 : Performances on the test set of the cadastre with RGBZ.



(a) Random Forest (b) XGBoost (c) Our model

Figure 4 : Confusion matrices computed on the test set with RGBZ.



(a) Test set (b) Ground truth (c) Predictions

Figure 5 : Qualitative results of our model on the test set.

As we can see, our model outperforms the baselines.

Conclusion

We developed a model for semantic segmentation of aerial photogrammetry points clouds. To do so, we used deep learning on graph to get better results than random forest or XGBoost with a reduced number of features. In the future, we intend to implement dilated convolutions and skip connections. Further, we would like to explore the learning on different graphs and then the learning of the graph itself.

Acknowledgements

I would like to thank my supervisors Michaël Defferrard and Frank De Morsier for their valuable and constructive suggestions. I would also like to thank the staffs of the LTS2 and Picterra for welcoming me and enabling me to observe their daily activities.

References

[1] David I. Shuman, Sunil K. Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. Signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular data domains. *CoRR*, abs/1211.0053, 2012.

[2] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *CoRR*, abs/1606.09375, 2016.