Franz Rottensteiner
Institute of Photogrammetry and GeoInformation
Leibniz Universität Hannover, Germany

"CLASSIFICATION AND SEMANTIC SEGMENTATION OF POINT CLOUDS: WHAT IS MISSING?"

Point clouds acquired from terrestrial or airborne sensors are one of the major data sources for retrieving semantic information about our environment. In this context, different tasks have to be solved to make the semantic information, contained in a point cloud only in an implicit form, explicit:

- Point cloud classification takes a point cloud as an input and determines which object is represented by that point cloud, assuming that it just represents one such object.
- Point cloud labelling (or semantic segmentation of point clouds) assigns a class label representing an object type to each point of the point cloud.
- Object detection aims at finding instances of specific objects in a point cloud, delivering the objects' positions and outlines. These outlines can be coarse in the form of bounding boxes or fine if object detection is coupled with semantic segmentation.

For the last years, we have seen tremendous progress in all of these tasks, in particular based on supervised classification methods. Researchers have developed expressive handcrafted features that can be extracted from a local neighbourhood of each point, and they have adapted supervised classification techniques from Machine Learning to the processing of point clouds. The consideration of context in the classification process by graphical models, e.g. Conditional Random Fields (CRF), has further improved the accuracy that can be achieved, in particular for small objects.

Thus, the development of classification techniques for point clouds can be seen as a success story. Nevertheless, when comparing it to classification tasks relying on images as a primary data source, it would seem that progress in point cloud processing is lagging behind. The extraction of semantic information from images has been revolutionised by deep learning techniques, in particular by convolutional neural networks (CNN), which have been shown to outperform other classification techniques by a large margin when solving the tasks mentioned earlier on the basis of images. Whereas in the meantime deep learning has been adapted for the classification of point clouds, there has been considerably less research on that topic so far. There are several reasons for that:

- Taking images is much easier than acquiring point clouds, and the sensors are cheaper. Consequently, semantic information extraction from images is also relevant for the consumer market, triggering research by large private companies such as Google. There is less commercial interest in point cloud processing.
- Point clouds have a more complex structure than images, so the task is more difficult. For instance, the concept of a convolution at the core of CNN is not easily transferred to point clouds. When using a grid-based structure for representing a point cloud, information is lost; otherwise, more complex mathematical concepts have to be applied. Until very recently, the best deep learning method in the Semantic.3d benchmark of ETH Zurich was based on the classification of 2D images simulated from a point cloud. Now there are first methods directly operating on point clouds that achieve better results, but there is still a demand for methodological research, even more so for processing airborne point clouds.
- To a large degree, the success of deep learning in image classification has been triggered by the availability of large publicly available benchmark datasets that can be used both for training and for testing. Whereas there are some benchmark datasets for point clouds, e.g. the Semantic.3d benchmark and the ISPRS 3D labelling challenge, the data volume is still rather small compared to the benchmarks in the Computer Vision community.

Our community cannot do much about the first item, but there are quite a few things we can do to solve the other ones:

- We should continue our efforts in benchmarking in order to provide the community with large databases of labelled data so that the prerequisites of deep learning techniques with respect to training data are fulfilled and it becomes easier to train deep learning models for the classification of point clouds.
- This particularly applies to airborne data, where the amount of publicly available training data is simply not sufficient at the moment.
- We also have to do methodological work on deep learning with point clouds. It could focus on some of the following topics:
  - Development and the comparison of different network architectures for the classification of point clouds and related tasks, based on benchmark data mentioned earlier, and again also with additional focus on airborne data;
  - Methods for reducing the requirements w.r.t. the availability of training data, e.g. methods for transfer learning (i.e., transferring a classifier trained on a certain data set to another one where the data might follow a different distribution or where the class structure may be different) or training techniques that can cope with wrong class labels of training samples (label noise);
  - Precise delineation of object boundaries. Whereas considerable efforts have been spent on that topic in Computer Vision, comparable methods are still largely unavailable for point cloud processing.
  - Consideration of context in the classification process. Whereas local context is implicitly considered by deep learning methods, the consideration of long-range dependencies between objects is still an unsolved problem.

These methods mainly aim at transferring developments related to deep learning to point cloud processing. Independently from the classifier that is involved, it would seem that most of the current work aims at labelling points; using these results to generate objects and a 3D representation of these objects seems to be another important direction of research.