Bottleneck Identification to Semantic Segmentation of Industrial 3D Point Cloud Scene via Deep Learning

Romain Cazorla[1,2] , **Line Poinel**[2] , **Panagiotis Papadakis**[1] , **Cédric Buche**[3]

[1]IMT Atlantique, Lab-STICC, UMR CNRS 6285, team RAMBO, F-29238 Brest, France
[2]Segula Technologies, France
[3]ENIB, CROSSING, IRL CNRS, Australia
romain.cazorla@segula.fr

Abstract

Point cloud acquisition techniques are an essential tool for the digitization of industrial plants, yet the bulk of a designer's work remains manual. A first step to automatize drawing generation is to extract the semantics of the point cloud. Towards this goal, we investigate the use of deep learning to semantically segment oil and gas industrial scenes. We focus on domain characteristics such as high variation of object size, increased concavity and lack of annotated data, which hampers the use of conventional approaches. To address these issues, we advocate the use of synthetic data, adaptive downsampling and context sharing.

1 Introduction

Point cloud acquisition in the industrial sector allows for the creation of drawings which would be unthinkable through manual measurements. Still, an industrial designer must spend several weeks to generate technical drawings for large installations. To automatize this arduous process, extracting the semantics of the point cloud through segmentation is essential i.e. by assigning a semantic label to each point of the cloud, such as "pipe" or "structural element" (see Figure 1) To alleviate certain challenges associated to the use of point clouds, most earlier works [Guo et al.2020] focus on feature learning and extraction. While this is indeed crucial, alleviating other issues related to the data format is critical for industrial scene segmentation. For this domain, we identify as the most recurring issues the lack of data availability, important size variation between object instances, dense shape articulation that can span very large volumes and investigate ideas that could alleviate these problems. We report performance with current state-of-the-art method in table 1.

Mean	Pipe	Girder	Valve	Barometer	Floor
0.25	0.419	0.135	0.051	0.0	0.926

Table 1: Intersection over Union of some classes obtained via the KPConv method [Guo et al.2020] on a custom industrial dataset.
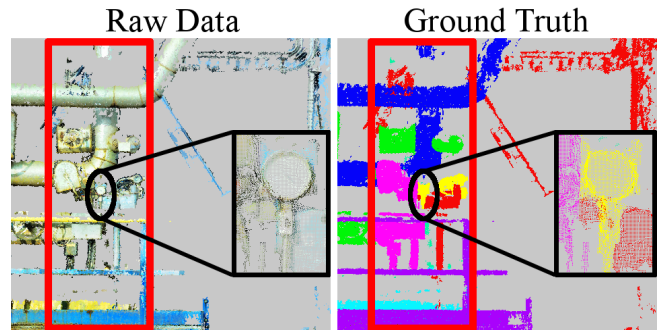

Figure 1: Semantic segmentation of a point cloud. The presence of far smaller objects (barometer highlighted in yellow) than others (red and purple : structural elements) is visible. The area contained in the red square is typical of our domain : 15 objects are present in a 0.6mx1mx1.8m box.

2 Domain-Specific Issues

Relative object size. Industrial equipment varies considerably in size, be it in one or several dimensions. Similarly, scale variation in 2D hinders object recognition as CNN are not scale invariant [Singh and Davis2018]. A multi-resolution approach can be used to reduce the issue of density and size variation [Du et al.2021]. However such method looses information during multi-resolution feature fusion. Also, in a number of cases point density is constant or even superior in the small-sized classes compared to larger-sized classes. Point density is one among several reasons why small size hinders segmentation. In interior environments, like S3DIS[Armeni et al.2016], some classes of objects are significantly smaller than others. We can witness relative loss of segmentation precision due to size. Even with an important number of points, these classes are consistently less well segmented than others. A relevant example is presented in Figure 2. This causes serious concerns in our domain, where some objects are significantly smaller than others but of great importance (such as a flange versus a tank). Finally, classes which are disproportionately large in one dimension, such as piping which often stretches over several dozen meters are difficult to capture as a whole by networks. This is why they are often considered as *stuff* in panoptic segmentation in related domains. [Nivaggioli et al.2019].
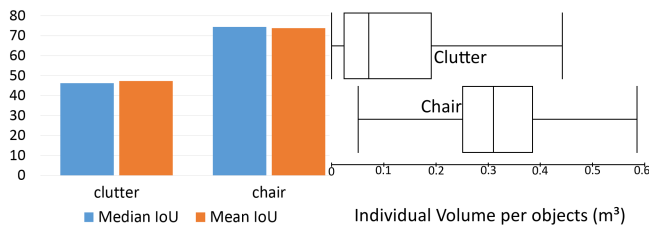
Figure 2: S3DIS Dataset [Armeni *et al.*2016] : Mean and median IoU value for two classes over 13 methods of the literature (left) - Relative object size comparison for two classes (right). The clutter class is 60% less precisely segmented than the chair class. Even if clutter is more complex than a chair, 75% of clutter objects are smaller than 87.5% of the chair objects although they are represented by 2.85 times more instance and 3 times more points than chair.

Scene layout. The specific layout of components may complicate their analysis [Pang2016]. Increased object concavity and size variance may result in dense scenes that are difficult to segment. Unavoidably, this further impedes sensing of surfaces and leads to occlusions. For road or interior scenes, 2D sampling and separate cell processing enables to fix the number of input points per area and hence allows MLP-based feature extraction [Guo *et al.*2020]. On the other hand, for industrial scenery with vertical layout as in the case of distillation towers in refineries, this approach is not well suited as it might lead to subsampling or oversampling, depending on object density. A vertical subdivision of the cloud could be considered, as long as important contextual information related to the floor and higher strata is not lost.

Data availability. Industrial scene data are scarcely available to the public, in contrast to domains such as road scenes or building interiors where the state-of-the-art is compared. Building a custom dataset would involve a tremendous amount of work, in view of similar tentatives such as ScanNet [Dai *et al.*2017]. Due to the object size and scene layout characteristics highlighted earlier, ground-truth generation becomes even harder.

3 Proposed Contributions

Data. In our application domain, few already segmented real-world data exist but a large number of 3D models is accessible. Our approach would then be to exploit synthetically generated data as a means for data augmentation, combined with a domain adaptation approach in order to increase performance [Wu *et al.*2018].

Downsampling. We argue that downsampling is a major cause of the inefficiency with which objects of different sizes are considered, since by its very nature it is a lossy process. As such, selecting the most important pieces of information during this operation is critical. However, information about a small object near a bigger one could be considered as non-critical in operations such as max pooling and as such is entirely lost. Designing a pooling layer able to determine which features are the most important via a learned importance kernel seems a promising idea as it has already shown its strength in 2D images [Gao *et al.*2019].

Scene layout. Context sharing allows [Landrieu and Simonovsky2018] to process large scenes after an initial over-segmentation but strongly depends on this first step. We feel that including more context information about object should alleviate the layout issue. It could allow to differentiate objects with domain specific knowledge (e.g. a flange cannot be on structural elements) but also to globally take into account really large objects such as piping, notably along the Z-Axis.

4 Epilogue

We presented three main issues that hinder the use of deep learning for semantic segmentation of oil and gas industrial scenes, along with ideas for alleviating them. Such improvements could allow to extend our work to Panoptic Segmentation, since adding instance segmentation is shown to improve performance [Wang *et al.*2019], which would be particularly beneficial in our case for the problem of scene layout.

References

[Armeni *et al.*, 2016] I. Armeni, O. Sener, A.R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese. 3D Semantic Parsing of Large-Scale Indoor Spaces. In *CVPR*, 2016.

[Dai *et al.*, 2017] A. Dai, A.X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner. ScanNet: Richly-annotated 3D reconstructions of indoor scenes. In *CVPR*, 2017.

[Du *et al.*, 2021] J. Du, Z. Jiang, S. Huang, Z. Wang, J. Su, S. Su, Y. Wu, and G. Cai. Point Cloud Semantic Segmentation Network Based on Multi-Scale Feature Fusion. *Sensors*, 21(5):1625, 2021.

[Gao *et al.*, 2019] Z. Gao, L. Wang, and G. Wu. LIP: Local Importance-Based Pooling. In *IEEE ICCV*, 2019.

[Guo *et al.*, 2020] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun. Deep Learning for 3D Point Clouds: A Survey. *IEEE TPAMI*, pages 1–1, 2020.

[Landrieu and Simonovsky, 2018] Loic Landrieu and Martin Simonovsky. Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs. In *IEEE CVPR*, 2018.

[Nivaggioli *et al.*, 2019] A. Nivaggioli, J. F. Hullo, and G. Thibault. Using 3D Models to Generate Labels for Panoptic Segmentation of Industrial Scenes. *ISPRS Annals*, IV-2/W5:61–68, 2019.

[Pang, 2016] Guan Pang. *3D Object Detection in Industrial Site Point Clouds*. PhD thesis, USC Libraries, 2016.

[Singh and Davis, 2018] Bharat Singh and Larry S. Davis. An Analysis of Scale Invariance in Object Detection - SNIP. In *CVPR*, 2018.

[Wang *et al.*, 2019] X. Wang, S. Liu, X. Shen, C. Shen, and J. Jia. Associatively Segmenting Instances and Semantics in Point Clouds. In *CVPR*, 2019.

[Wu *et al.*, 2018] Bichen Wu, Alvin Wan, Xiangyu Yue, and Kurt Keutzer. SqueezeSeg: Convolutional Neural Nets with Recurrent CRF for Real-Time Road-Object Segmentation from 3D LiDAR Point Cloud. In *IEEE ICRA*, 2018.