

Article

Object Semantic Segmentation in Point Clouds—Comparison of a Deep Learning and a Knowledge-Based Method

Jean-Jacques Ponciano ^{1,*}, Moritz Roetner ², Alexander Reiterer ^{2,3} and Frank Boochs ¹

¹ i3mainz, Institute for Spatial Information and Surveying Technology, Mainz University of Applied Sciences, 55128 Mainz, Germany; frank.boochs@hs-mainz.de

² Fraunhofer Institute for Physical Measurement Techniques IPM, 79110 Freiburg, Germany; moritz.roetner@ipm.fraunhofer.de (M.R.); alexander.reiterer@ipm.fraunhofer.de (A.R.)

³ Department of Sustainable Systems Engineering INATECH, University of Freiburg, 79110 Freiburg, Germany

* Correspondence: jean-jacques.ponciano@hs-mainz.de

Abstract: Through the power of new sensing technologies, we are increasingly digitizing the real world. However, instruments produce unstructured data, mainly in the form of point clouds for 3D data and images for 2D data. Nevertheless, many applications (such as navigation, survey, infrastructure analysis) need structured data containing objects and their geometry. Various computer vision approaches have thus been developed to structure the data and identify objects therein. They can be separated into model-driven, data-driven, and knowledge-based approaches. Model-driven approaches mainly use the information on the objects contained in the data and are thus limited to objects and context. Among data-driven approaches, we increasingly find deep learning strategies because of their autonomy in detecting objects. They identify reliable patterns in the data and connect these to the object of interest. Deep learning approaches have to learn these patterns in a training stage. Knowledge-based approaches use characteristic knowledge from different domains allowing the detection and classification of objects. The knowledge must be formalized and substitutes the training for deep learning. Semantic web technologies allow the management of such human knowledge. Deep learning and knowledge-based approaches have already shown good results for semantic segmentation in various examples. The common goal but the different strategies of the two approaches engaged our interest in doing a comparison to get an idea of their strengths and weaknesses. To fill this knowledge gap, we applied two implementations of such approaches to a mobile mapping point cloud. The detected object categories are car, bush, tree, ground, streetlight and building. The deep learning approach uses a convolutional neural network, whereas the knowledge-based approach uses standard semantic web technologies such as SPARQL and OWL2 to guide the data processing and the subsequent classification as well. The LiDAR point cloud used was acquired by a mobile mapping system in an urban environment and presents various complex scenes, allowing us to show the advantages and disadvantages of these two types of approaches. The deep learning and knowledge-based approaches produce a semantic segmentation with an average F1 score of 0.66 and 0.78, respectively. Further details are given by analyzing individual object categories allowing us to characterize specific properties of both types of approaches.

Keywords: point cloud; deep learning; knowledge-based; SPARQL; segmentation; 3D; semantic segmentation; classification



Citation: Ponciano, J.-J.; Roetner, M.; Reiterer, A.; Boochs, F. Object Semantic Segmentation in Point Clouds—Comparison of a Deep Learning and a Knowledge-Based Method. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 256. <https://doi.org/10.3390/ijgi10040256>

Academic Editors: Wolfgang Kainz and Greet Deruyter

Received: 2 February 2021

Accepted: 6 April 2021

Published: 10 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Understanding unstructured data is a complex task for computer-based approaches. It requires connecting features in the data with the nature and understanding of the objects of interest. However, the representation of an object in a data set varies strongly, and an automatic process ideally should be able to handle this variety.

For example, the data characteristics depend on the acquisition process (e.g., technology and methodology used) that generates them. The characteristics of the digitized

objects (e.g., material, reflectance, roughness, size), the context of the scene (e.g., urban outdoor, indoor building, ruin excavation), and various other factors external to the acquisition process (e.g., ambient light, light intensity, weather conditions, movement of the measuring instrument or digitized objects) influence the acquisition process. These variations in characteristics generate a diversity of object representations in different data sets, which can differ from our expectations of reality. The diversity of object representation is a real challenge for the approaches of semantic segmentation. The different semantic segmentation approaches can be gathered in three categories: model-driven, data-driven and knowledge-based approaches. Among these three categories, data-driven approaches using deep learning (DL) and knowledge-based (KB) approaches have shown great results. Therefore, this research focuses on these two.

On the one hand, the approaches based on DL (such as deep learning) substitute the lack of understanding of the acquisition process factors to identify reliable patterns. These patterns must be learned in the training stage. To face the challenge of the diversity of object representation, DL requires a vast amount of data to provide enough variety of object representation and to identify reliable patterns. The use of a vast amount of data allows DL to be robust toward different object representations. The DL approaches are dependent on the data set used for the training stage. This dependency defines the strength and the weakness of these approaches. Indeed, DL approaches work as long as the considered data represents the content to be understood. The more variation occurs in object or appearance, the more data are required for training. DL approaches have difficulty detecting the object or the geometry for which they have not been trained. This is the case for unique objects or data sets found in fields such as cultural heritage. In the same way, that DL approaches depend on training data sets, KB approaches depend on the knowledge definition. KB approaches based on semantic technologies integrate human knowledge of the object or sensing process to guide the semantic segmentation process. This knowledge allows the adaptation of the object definition and the semantic segmentation process according to the sensing process, which is composed of the acquired scene context or acquisition process. They have the advantage of being applicable in a large variety of contexts.

Both approaches have a common goal but apply different strategies. It is therefore interesting to compare them to highlight their strengths and weaknesses. For this reason, this research compares two implementations of these approaches, one for each. The comparison focuses on a supervised DL approach developed at the Fraunhofer Institute for Physical Measurement Techniques IPM and a fully knowledge-based approach developed at the Institute of Spatial Information and Surveying Technology at Mainz University of Applied Sciences (i3mainz). The DL approach uses a VGG6 Encoder with an FCN8 Architecture [1]. The fully KB approach uses standard semantic web technologies SPARQL [2] and OWL2 [3] to guide the selection of algorithms and the classification process according to objects or data characteristics.

The comparison uses an outdoor point cloud acquired by cameras and a laser scanner on a mobile mapping system. For that reason, the objects are viewed only from the middle of the road on which the car moved. This affects the representation of the objects in the data set. Further factors (variation in brightness, the vibration of the measuring instrument, meteorological conditions) also influence the acquisition process. This creates a large diversity of object representations within the data (e.g., complete car vs. partially acquired car, different types of cars, etc.). This comparison is based on the detection of a set of “regular” objects (in terms of the geometrical aspect), which include cars, buildings, the ground and streetlights, and a set of “irregular objects” (disparate shapes), which include bushes and trees. This application case has the advantage of providing a variety of objects and of their representation within the data to compare the performance of these two approaches. Their performance is assessed and compared based on these metrics. First, this research presents works related to semantic segmentation. Secondly, the DL approach’s workflow is presented, followed by an explanation of the workflow of the fully knowledge-based approach. This research then presents the results obtained by each approach and

compares them. Finally, the research presents the obtained conclusions based on the observed differences or similarities and corresponding properties of both approaches.

2. Related Work

The segmentation of objects and geometries depends on their characteristics (e.g., size, shape) on the one hand, and the characteristics of the data on the other hand (e.g., density, noise, occlusion, roughness).

Model-driven approaches are based entirely on the characteristics of the objects. The approaches mentioned in [4–6] propose a detection of simple objects based on the mathematical model of their shape. However, they are only able to detect simple geometric objects (such as walls or columns). Approaches [7,8] propose describing objects with different descriptors. These approaches only allow the detection of objects if they have different characteristics and if the data are dense and uniform.

The current state-of-the-art data-driven approaches, which are overwhelmingly dominated by DL techniques based on convolutional neural networks, have existed for some time [9]. More recently, object recognition in 3D point cloud data using deep learning has gained attention not only due to the availability of more raw processing power in recent years, but also because of difficulties with the sparseness of many point clouds, and enormous data sets with many different classes (e.g., an entire city), where the a priori description of knowledge features is usually impossible [1,10]. DL approaches can detect objects after a thorough first training without a priori knowledge. These “bottom-up” approaches are more flexible and usually generalize better than model-based approaches. Semantic segmentation tasks in 3D point clouds can broadly be split into two approaches: direct, via point-based methods, and indirect, via projection-based networks [11].

One of those direct approaches is “PointNet [12]”, which presents an in-depth learning framework that learns point-wise features with several DLP (multilayer perceptron) layers and extracts global shape features with a max pooling layer to allow object detection in the point cloud without simplifying this large and unsorted data set. Using a hierarchical neural network structure, as described in PointNet++ [13], the original “PointNet” approach improved significantly, especially regarding local structural information. Besides these point-wise DLP implementations, there are various other approaches using point convolution methods (DPC [14]), RNN-based methods (DARNet [15]), graph-based methods (DPAM [16]), and lattice representation methods (LatticeNet [17]). The indirect method, however, uses image-based recognition and projects these results into the point cloud.

These projection-based methods can be further split into five categories [11]: multi-view (TangentConv [18]), spherical (RangeNet++ [19]), volumetric (FCPN [20]), permutohedral lattice (LatticeNet [17]) and hybrid representation (MVPNet [21]). In [9] a 3D point cloud is projected onto 2D planes from multiple virtual camera views. Using a Fully Convolutional Network (FCN) to predict pixel-wise scores on these synthetic images and fusing the re-projected scores, a semantic segmentation can be achieved without working directly on the point cloud. In [18] tangent convolutions for dense point cloud segmentation have been introduced, based upon the assumption that point clouds are sampled from locally Euclidean surfaces; the local surface geometry is projected onto a virtual tangent plane to be processed by tangent convolutions.

Although these indirect methods do not fully exploit the underlying geometry and structural information, they perform and scale well for large data sets [11]. Unlike data-driven approaches, KB approaches use knowledge in a “top-down” strategy to detect objects.

The interest in using ontologies for object classification is present in the field of image processing. The work of [22] uses an ontology-based classification to identify different types of buildings in airborne laser scanner data. The ontology, specified in OWL2, formalizes each type of building using appropriate features to describe a qualitative concept identified through the Random Forest Classifier [23]. This approach consists of first delineating buildings’ footprints, then extracting their features and adding them to the ontology. Finally,

reasoning using the Fact++ reasoner [24] classifies the different buildings by determining their type according to their OWL2 specification. This approach has performed well (F-measure = 97.7%) in classifying “residential/small buildings”, but the authors highlight the need for additional information to avoid overlap between the two other classes— “Apartment Buildings” and “Industrial and Factory Buildings”— which obtain F-measures of 60% and 51%, respectively. The work of [25] aims at identifying objects in urban and periurban areas in remote sensing images. This approach of object-oriented image analysis consists of segmentation, extraction of features, and identification using a domain ontology. The domain ontology contains the description of objects according to three types of attributes: spectral, spatial, and contextual. The ontology-based object recognition assigns each region with a feature-based matching score for each concept of the ontology. Next, it defines the concept that obtained the highest matching score as the concept representing the region. This approach obtains an average F-measure of 87% in detecting an orange house, vegetation, road, and water. The approach of the doctoral thesis in [26] aims at allowing semantic image indexing and retrieval. Maillot uses visual concept ontology and reasoning techniques to extract and classify objects from segmented images. The visual concepts of the ontology are associated with low-level features and algorithms. This ontology provides an interface between expert knowledge and the image processing level. Through this condensed review of the most relevant state-of-the-art approaches for detecting objects in point clouds, Two types of the most efficient approaches are highlighted. Deep learning approaches using a “bottom-up” strategy, and KB approaches using a “top-down” scheme.

To study the effectiveness of these two types of approaches on the point clouds resulting from mobile mapping technologies more precisely, we compare an DL approach developed by IPM [9] presented in Section 3 to a KB approach developed by i3mainz [27] presented in Section 4. We assume that the approaches developed by IPM and i3mainz are efficient implementations of DL and KB approaches, respectively.

3. Material and Method

3.1. Data Set Acquisition

Mobile mapping systems are increasingly used to document infrastructure elements (e.g., urban space, road surface, etc.) [28]. The great advantage of these systems is efficient data acquisition at high driving speeds and in flowing traffic. The data quality depends on the corresponding measurement sensor technology and the environment (e.g., surface characteristics).

Typically, a mobile mapping system consists of a positioning system (primary system), a combination of a gyro system, a global positioning system (e.g., global navigation satellite system, GNSS) and an odometer. A Kalman filter calculates a continuous forward solution in real-time. All other sensors are secondary sensors and synchronized with the positioning system by either distance or time. These sensors can be cameras or laser scanners. The results of a measurement run with such a vehicle are georeferenced and time-stamped images, and what is known as a point cloud.

A point cloud represents a scene as an initially unsorted set of points in 3D coordinates (3D point cloud). In addition to the 3D information, the received backscattered laser’s intensity value is available (4D point cloud). State-of-the-art systems generate 3D point clouds with absolute accuracy in the range of a few centimeters by measuring up to two million points per second.

The time signal is the basis for the fusion of different data streams from the sensor systems mentioned (positioning system, laser scanner and cameras). A separate calibration describes the individual sensors’ spatial orientation to each other so that three rotation and three translation parameters are available for each sensor.

The mobile mapping system used to collect data for our experiments was the Mobile Urban Mapper (MUM) developed by IPM [29]. MUM comprised Fraunhofer IPM’s high-end laser scanner CPS [30] with two million measurements per second. The scanning frequency was 200 Hz, while still providing a distance precision of about 3 mm (1 sigma).

Four cameras recorded the vehicle's entire environment and captured images with a resolution of 5 megapixels every 5 meters. MUM was positioned and oriented with an Applanix LV420 (<https://www.applanix.com/products/poslv.htm> accessed on 9 April 2020).

Each measurement technology and context influences the characteristics (e.g., noise, density, regularity) of the generated point cloud. These characteristics have an impact on the semantic segmentation process and its efficiency.

The acquisition of a scene in an urban outdoor context acquired by a mobile mapping system was obtained by sequentially scanning the scene using a laser device mounted on a car moving along the road. Such a sequential scan acquisition method produces data acquisition in the form of "steps" on the one hand and generates occlusion areas (without information) on the other hand. Moreover, the different materials (e.g., metal, glass, and stone) and the distance between the objects and the acquisition technology at the time of acquisition greatly influence the process. Thus, portions of data can be dense and continuous, and others can be discontinuous with low density.

Furthermore, the complexity of the different factors influencing the acquisition process leads to unpredictable situations and problems where the characteristics of the data obtained differ significantly from the expected characteristics.

Indeed, trees and bushes have irregular shapes resulting in a quite rough representation, lowly dense, highly curved, and highly noisy. Additionally, cars are mainly discontinuous and present some unpredictable representation. Buildings and grounds have huge density variations due to their location, which is quite far from the mobile mapping system and causes some building and ground occlusions. Moreover, the shape of the ground is not constantly horizontal and not planar. Finally, there are streetlights, which are small and thin objects compared to the others, and have a low density.

3.2. Deep Learning Workflow

This study's deep learning workflow was based on a multi-view approach originally developed for the classification of urban surface textures [31]. Its key component was a convolutional neural network (CNN) for image segmentation trained in a supervised fashion on a large set of mobile mapping images. The choice of a multi-view approach based on red-green-blue (RGB) images for this problem was motivated by (a) its sensitivity to color as a feature (which is crucial for distinguishing different surface textures) and (b) the need for efficiency in application. Processing the dense point clouds produced for urban mobile mapping was time- and resource-intensive, so shifting the semantic detection to 2D images taken at regular intervals significantly sped up the detection process, the training of the CNN and the optimization of the workflow in comparison to a deep learning-based semantic segmentation of the 3D data. The original implementation segmented the point clouds into more than 30 different object classes, some as fine-grained as Gravel, Paving, Curbstone, and Manhole Covers. For this study, the number of output classes was adapted to the list in Section 4 (car, bush, tree, ground, streetlight and building) by merging all surface objects into the class Ground. The neural component was a standard FCN [1] with a VGG16 encoder. An implementation based on the Caffe framework [32] was used to train the network. Training took several weeks on a data set of 90,000 manually segmented images of urban scenes with data augmentation (color and brightness shifts). The data had been collected during three mapping campaigns throughout Germany over the span of a year. The Adam optimizer was employed with the following parameters: learning rate = 0.001, beta1 = 0.9, beta2 = 0.999, epsilon = 10^{-08} .

Figure 1 details the components of the data processing pipeline. The FCN classified all RGB input images by pixel, resulting in a 2D semantic segmentation of different views of the 3D point cloud from the cameras' positions on the mobile mapping vehicle. Then the intrinsic and extrinsic camera parameters were applied to project the 2D classifications into the georeferenced point cloud.

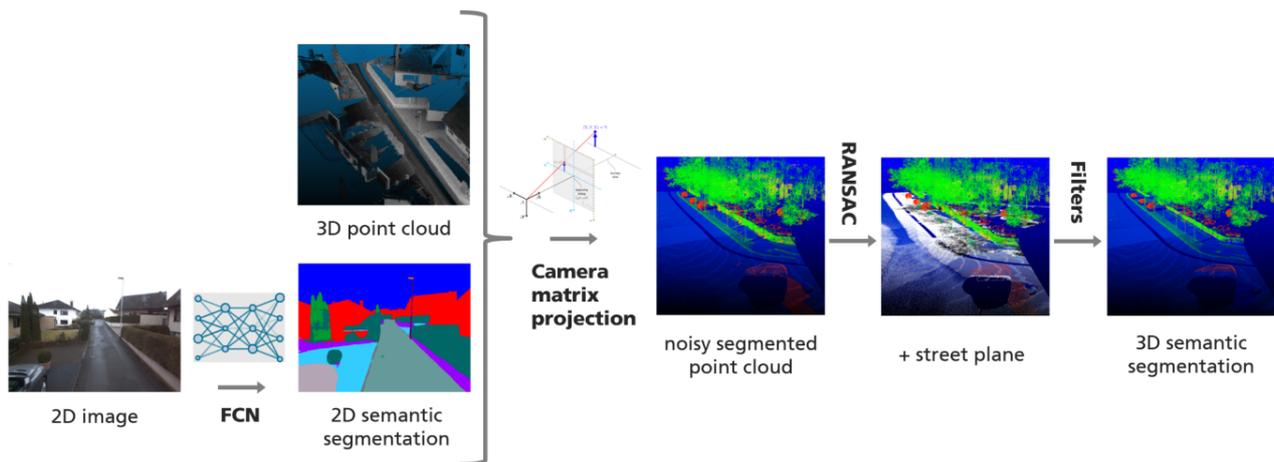


Figure 1. Pipeline and components of the deep learning workflow.

While the semantic segmentation of the images was straightforward and handled entirely by the neural network, complications and ambiguities arose at this stage for several reasons.

- Due to vehicle movement, images and view perspectives overlapped, and points could receive several different classifications from a corresponding pixel; the further away from the camera, the worse the classification result here. Since small errors in calibrations summed up at an exponential rate in relation to the distance from the camera to the point/pixel, we implemented a selection scheme, in which information from an image that had been acquired closer to the vehicle received a higher weight, to determine the final class of a point once the projection had finished.
- The resulting point cloud was very noisy due to projection through objects like windowpanes (which the scanner does not record), partially recorded objects due to the scanner angle, and the generally lower resolution of the point cloud compared to the image.
- Classification results were limited to the fields of view of the cameras. Regions that were not visible in any RGB image could not receive a label during projection, resulting in non-classified areas within the point cloud.
- Finally, errors in calibration summed up at an exponential rate in relation to the distance from the camera to point/pixel.

Therefore the DL approach applied several filter strategies in post-processing to reduce noise.

- Street plane filter: All points that received a Ground label and fit a plane with RANSAC [33] were clustered in the street plane. In the next step, all labels other than Ground on and 0.5 m above the street plane were removed.
- Depth filter: To prevent labels from traveling through objects due to their lower resolution in the point cloud and small calibration errors, a window of depth values of 20 pixels \times 20 pixels around the currently projected pixel was taken into account for large objects, namely Tree and Building. A label was only projected if it was in the majority depth plane (with an error margin of 10 cm).

3.3. Knowledge-Based Workflow

3.3.1. Knowledge Modeling

KB approaches use an explicit representation of knowledge defined through an ontology. These approaches integrate knowledge about data, objects, and algorithms to drive the semantic segmentation process based on data and objects [34–36]. In [27], the approach presents a model of algorithms, data, and objects in OWL2. Figure 2 illustrates the modeling structure of these concepts.

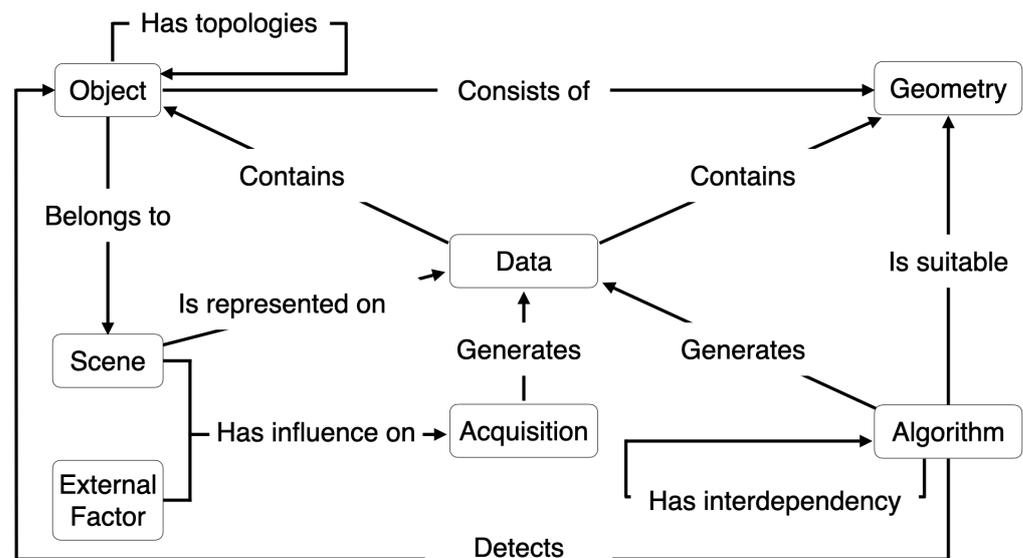


Figure 2. Overview of knowledge modeling.

The concept “Data” representing data is described as containing objects and geometries. It is described as being able to be generated from an acquisition process or an algorithm. This concept is also defined as being able to represent a scene. The acquisition process influences the data through external factors and the scene that it represents. Such knowledge of the acquisition process allows for characterizing the data and adapting them to its characteristics [37].

An object belongs to a scene, consists of geometries, and has a topological relationship with other objects. An algorithm is suitable for specific geometries and detects objects. In accordance with the geometries that constitute the objects, algorithms are selected to adapt the process to objects’ characteristics [38].

Algorithms have interdependencies between them, which means the execution of an algorithm can require another algorithm’s execution. Thus, the semantic segmentation process uses information about algorithms, their interdependence, objects, and data characteristics to select and apply the appropriate algorithms.

This information is described using description logics. Description logics are a family of formal languages for representing knowledge. They are mainly used in artificial intelligence to describe and reason about the relevance of concepts in an application domain. For example, the description logic of a streetlight in Manchester Syntax used for the comparison in Section 5 is shown in Listing 1.

Listing 1: Description of Streetlight in Manchester Syntax.

```

Object that consistsOf some
(Line that hasOrientation only VerticalNormal)
and belongsTo only Outdoor
and hasDistance some
(Distance that assessDistanceOf some StreetLight
and (hasValue exactly 1 xsd:double[>= '30'^^xsd:double])
and hasValue exactly 1 xsd:double[<= '40'^^xsd:double])
and (hasWidth only xsd:double[< '0.5'^^xsd:double])
and (hasLength only xsd:double[< '0.5'^^xsd:double])
and (hasHeight only xsd:double[>= '3'^^xsd:double])
and hasMaterial some Reflectif
and hasRoughness some LowRoughness
  
```

3.3.2. Knowledge-Based Workflow

The workflow of the KB approach consisted of dividing the complexity of the point cloud by regrouping the points in homogeneous segments (segmentation), then extracting some relevant features from these segments (feature extraction) to then finally identify segments as objects or parts of objects (classification) [39]. Figure 3 shows the workflows of the KB approach.

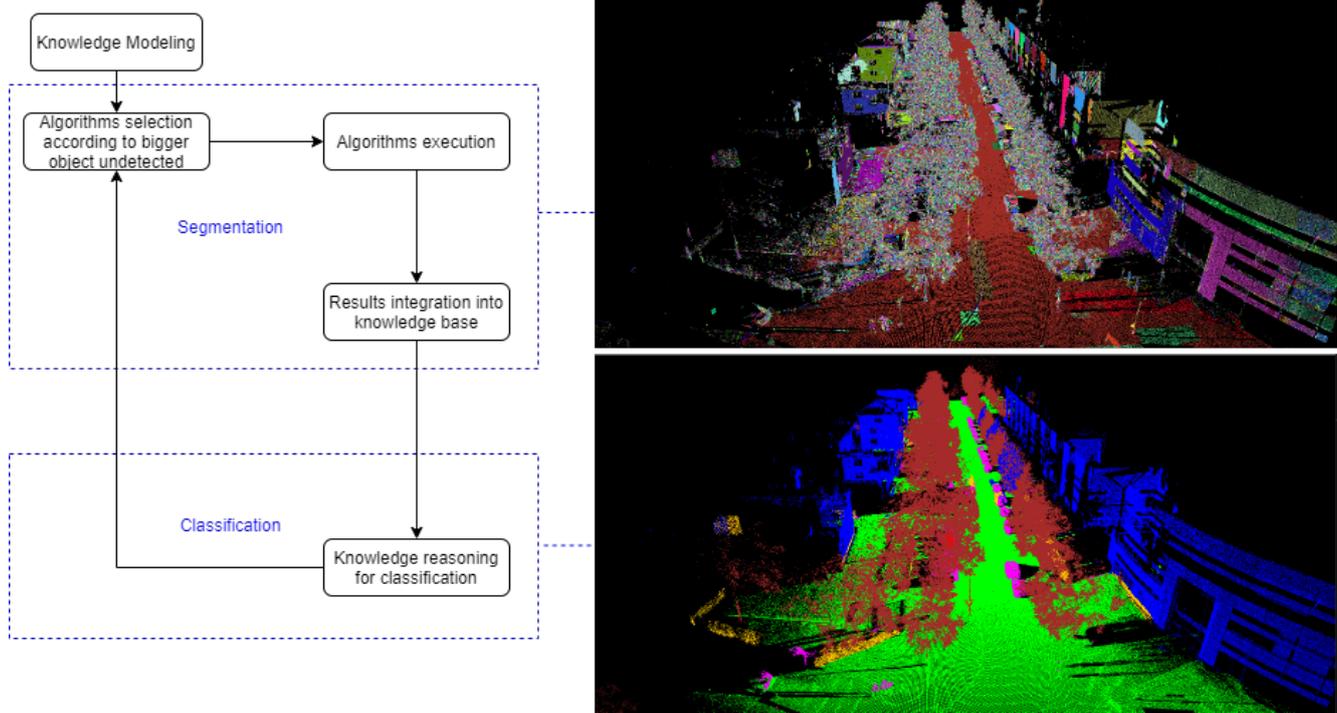


Figure 3. Overview of knowledge modeling.

3.3.3. Segmentation Process

The combination of algorithms(such as “Color-based Region Growing Segmentation”, “Estimating Surface Normals”, “VoxelGrid filter”, “Supervoxel Clustering”, “Region Growing Segmentation”) from libraries such as PCL 1.11.1 (<https://github.com/PointCloudLibrary/pcl> accessed on 9 April 2020) and OpenCV 4.5.2 (<https://opencv.org/> accessed on 9 April 2020) performed the segmentation and feature extraction steps. Such an approach allowed the iterating of segmentation and classification according to different objects targeted, such as explained in [40]. Such a strategy can be considered a refined segmentation process. The selections and configurations of algorithms were automatically designed according to objects and data descriptions, as explained in [41]. More precisely, the description of object characteristics aimed at guiding the strategy of semantic segmentation. The simpler the object shape was, the simpler its semantic segmentation strategy could be. In contrast, objects with complex shapes required more elaborate semantic segmentation strategies.

Three main parts composed the semantic description of objects: the object’s characteristics, its geometry, and the scene that it belonged to. These parts aim at facilitating the adaptation of semantic segmentation strategies. The feature extraction step described essential geometric characteristics such as their centroid, orientation and dimensions (height, length, width). These features were then integrated into the ontology.

4. Test Data Set

The two approaches were applied to the data set illustrated in Figure 4 and compared the results obtained for each approach based on the most relevant literature metrics. The data set used for the comparison was an urban outdoor point cloud acquired by a mobile mapping system and was composed of 7.5 M points. Figure 5 illustrates the point cloud.

We chose to study the semantic segmentation of the most widespread object categories in urban environments, corresponding to the six following categories: car, bush, tree, ground, streetlight and building.

The car category here represented any visible automobile, without any other larger vehicles, such as trucks. A bush was the visible conjugate structure, the same for the tree class, whereas crown and trunk are combined during classification. In some areas, trees grow within large bushes. Here the two approaches only classified the visible and distinguishable tree parts; the remaining section was classified as a bush. The class of ground included everything on a ground plane regardless of the surface character. Streetlights were classified as such when they were distinguishable for the human eye. For the building class, every part of a house and its additions, like gutters, were included.

First, the roughness (through the approach [42]), the density (through the approach [43]) and the curvature (through the approach [44]) of the data set to highlight the challenges were estimated. Figure 5 shows the estimation of these three characteristics of the data set.

Another essential characteristic of the data set was noise. Figure 6 shows the estimated noise of the data set.

Trees and bushes had a lot of noise, low density, and high surface roughness. These characteristics made their shapes very difficult to predict. Therefore, their semantic segmentation was a challenge. The semantic segmentation of cars was also challenging due to the varying shapes. Likewise, building and ground semantic segmentation was challenging due to density variation and incompleteness caused by occlusion. Finally, streetlight semantic segmentation's main challenge remained in the low density and the thinness of their portrayal.



Figure 4. Point cloud: acquired by mobile mapping system, used for comparison.

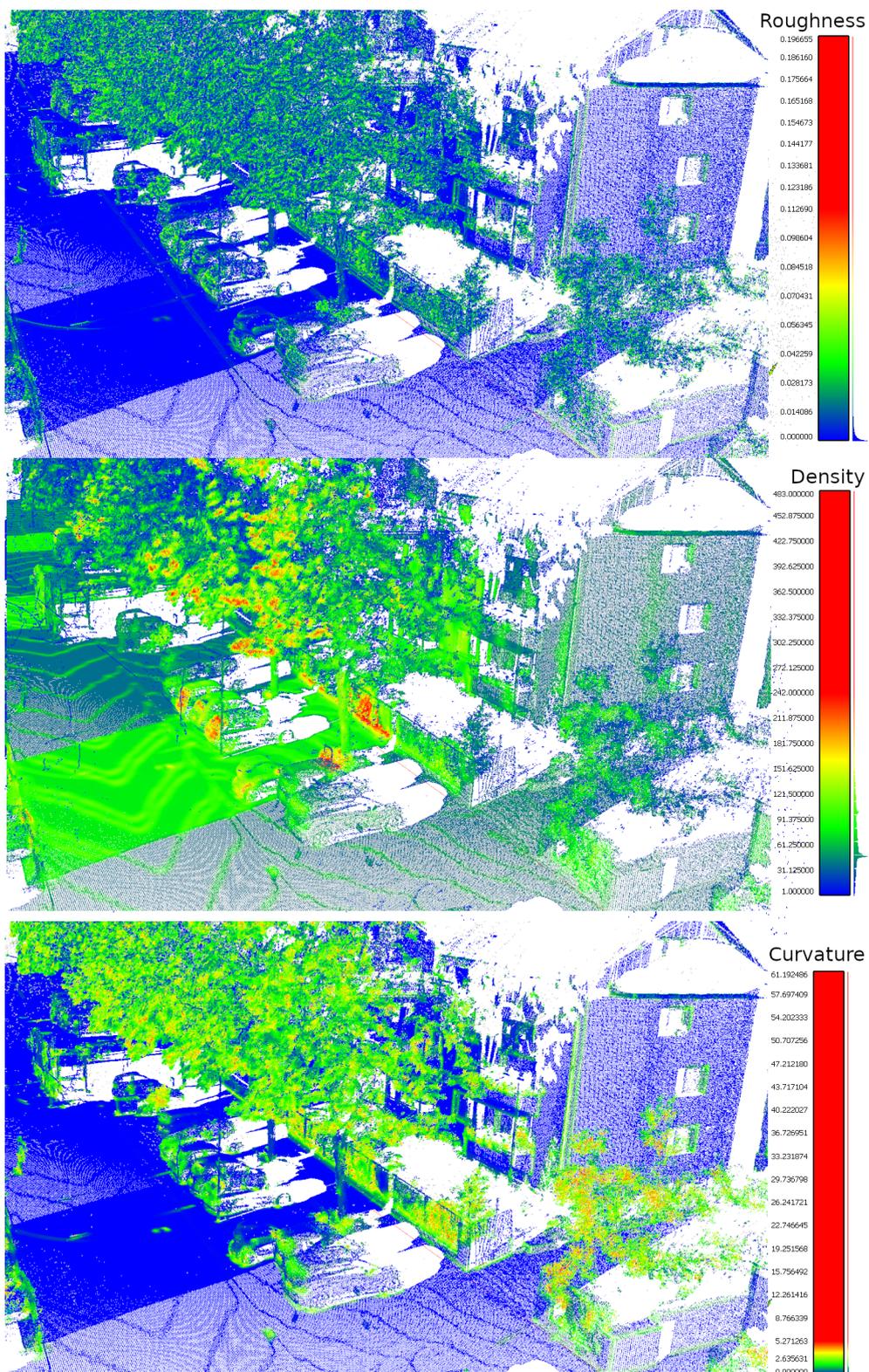


Figure 5. Point cloud used for metrics comparison: (top) roughness estimation, (center) density estimation, (bottom) curvature estimation.

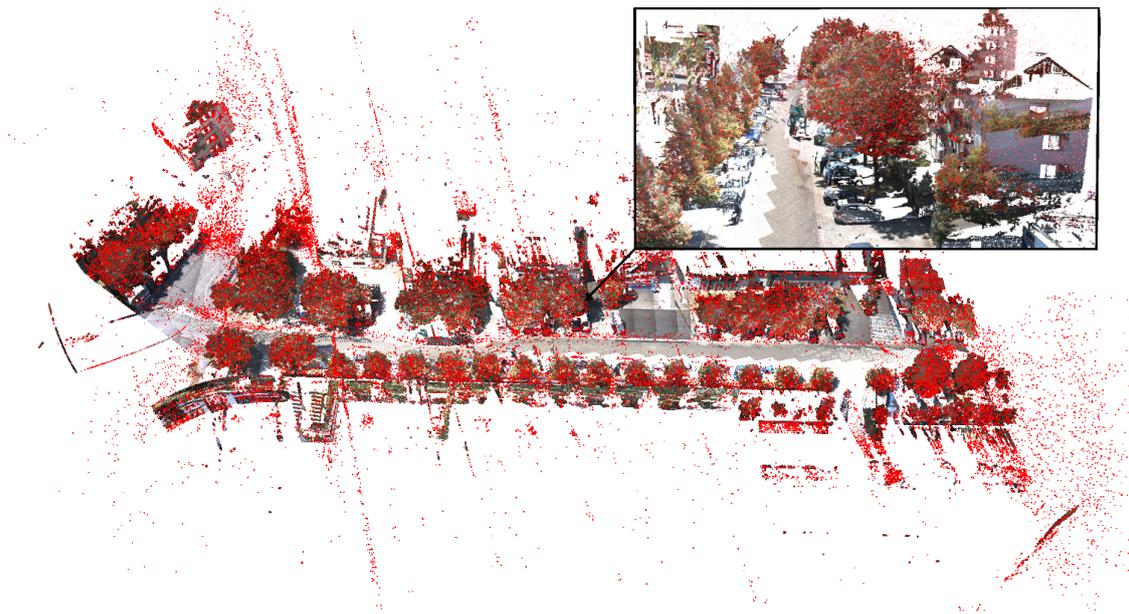


Figure 6. Noise illustration in red on the point cloud acquired by the mobile mapping system.

5. Results

The object identification consisted of assigning a label (represented by a color) to each category and applying this label to each point in the point cloud identified as belonging to one of these categories (car: labeled in magenta; bush: labeled in yellow; tree: labeled in brown; ground: labeled in green; streetlight: labeled in red; building: labeled in blue).

Among the different quantitative metrics existing in the literature for assessing object semantic segmentation in discrete sets the most often used metrics chosen were: precision, recall, F1 score, and IoU. Precision measured the ability to correctly classify a segment, while recall measured the ability not to miss segments for a class. The F1 score, which is a harmonic average of recall and precision, allowed the measurement of the classifier's average performance. In contrast, the IoU score measured the worst-case performance of the classifier.

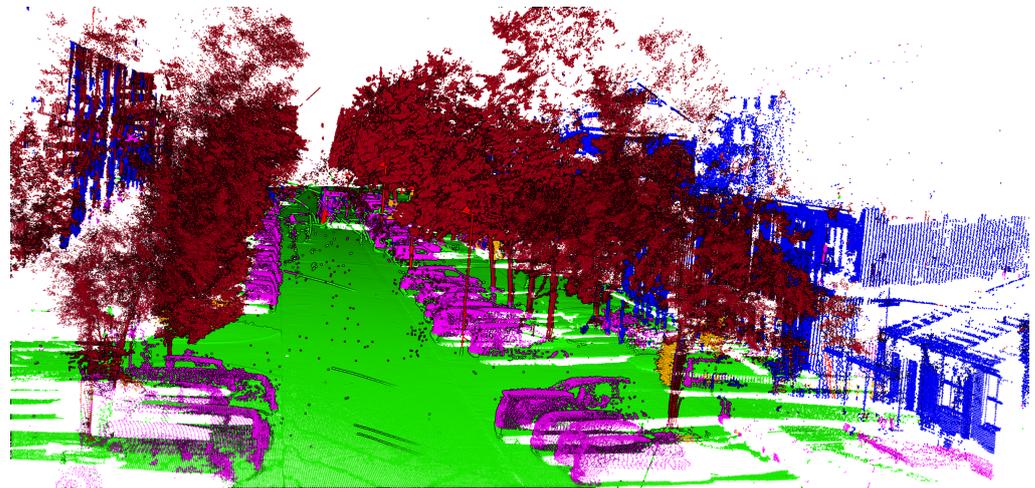
5.1. Global Results

The results were obtained by applying the two approaches—KB and DL—to the test data set. Neither of these approaches was optimized for this data set. Score metrics were calculated by comparing the results of each of the two approaches with a ground truth that was entirely generated by manual annotation.

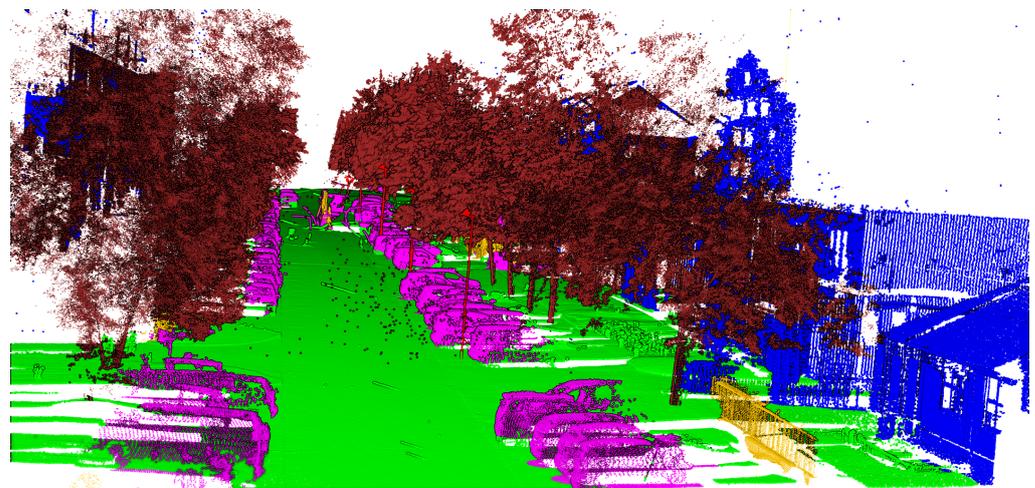
Figure 7 illustrates the global results view of the semantic segmentation performed by the DL approach and the KB approach on this data set. The two approaches yielded an average F1 score higher than 78% (KB) and 66% (DL), and an IoU score (which indicated the “worst-case” encountered) higher than 65% (KB) and 51% (DL). Both approaches reach a high score for the semantic segmentation of ground and tree. They both had low precision but good recall for the semantic segmentation of streetlights. This meant that most streetlights were correctly identified but incorporated other elements. Bushes, cars and buildings were segmented well, on the whole, reaching an F1 score between 56% and 76.7%. The statistical results showed that the KB approach was more efficient in detecting all the classes except the bush class, for which the approach presented by IPM was more efficient. Among the different results, we observed that the DL approach had better precision for the semantic segmentation of buildings and bushes and better recall for bushes than the KB approach. In the next section, the results obtained are further detailed through illustrative examples.



(a) Original point cloud



(b) DL semantic segmentation results



(c) KB semantic segmentation results

Figure 7. Point clouds used for the comparison.

The comparison between the KB approach and the DL approach is presented in Tables 1 and 2 for the classification of the six main classes in the data.

Table 1. Results of the statistics analysis of the deep learning (DL) approach.

Deep Learning Approach (DL)				
Class	Precision	Recall	F1 Score	IoU
Car	63.71%	70.44%	66.90%	50.27%
Bush	80.42%	70.23%	74.98%	59.98%
Tree	86.47%	67.98%	76.11%	61.44%
Ground	87.17%	83.75%	85.43%	74.56%
Streetlight	24.90%	78.52%	37.81%	23.31%
Building	85.52%	41.82%	56.17%	39.05%

Table 2. Results of the statistics analysis of the knowledge-based (KB) approach.

Knowledge-Based Approach (KB)				
Class	Precision	Recall	F1 Score	IoU
Car	71.97%	79.51%	75.55%	60.71%
Bush	75.02%	63.04%	68.51%	52.10%
Tree	92.73%	88.08%	90.35%	82.39%
Ground	90.83%	91.93%	91.38%	84.12%
Streetlight	56.29%	81.00%	66.42%	49.73%
Building	67.39%	89.02%	76.71%	62.21%

5.2. Detailed Results

Bushes were described semantically in the studied KB approach according to their geometry and relationship with other elements. However, bushes did not have a constant geometry or a recursive relational link with other elements. Hence, their identification was more complicated for the KB approach.

Figure 8 shows an example of an incorrect bush semantic segmentation by the KB approach compared to the DL approach.

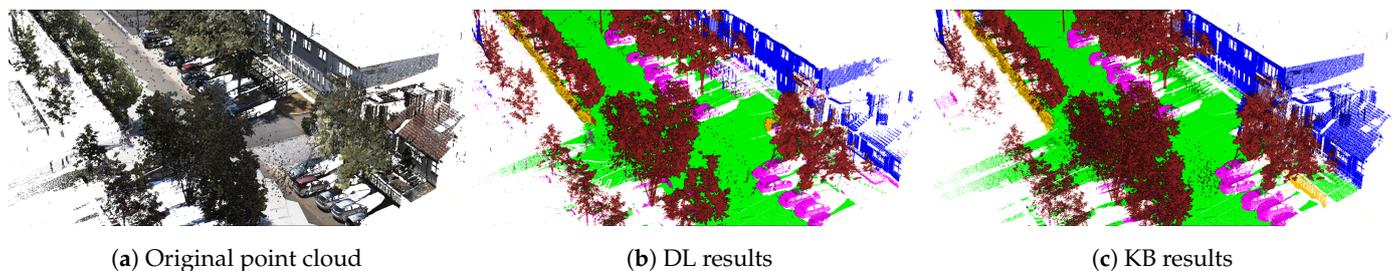
**Figure 8.** Example of bush and barriers semantic segmentation results.

Figure 8 illustrates that the DL approach detected the bush (in yellow) correctly. In contrast, the knowledge-based approach did not dissociate barriers from bushes and identified them as bushes. This lack of segmentation was mainly caused by the logical description of bushes, which was not specific enough to exclude other elements.

Categories that could be formally described geometrically or with relational links such as cars, trees, ground, buildings, and streetlights were detected better by the KB approach.

Figure 9 shows that the DL approach missed building and shaft portions. Besides, some portions of the ground were classified as cars. Additionally, some portions of the buildings were classified as trees. This was—in large part—caused by projection errors and misclassification of images.

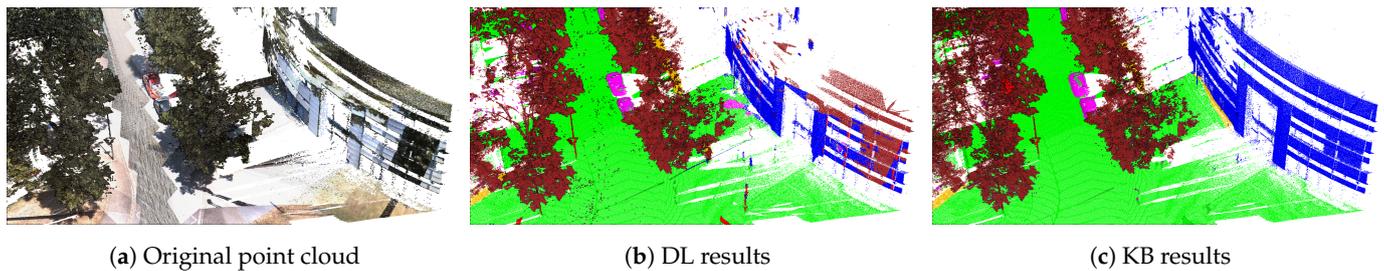


Figure 9. Sample semantic segmentation results

Each sensor of the acquisition vehicle had to be calibrated and positioned in relation to the vehicle: each camera and lens had to be corrected and calibrated. Every sensor needed to be placed in a dependent manner to the vehicle, and eventually, the vehicle itself needed to be placed within a world coordinate system. This was a very fragile process, where one small error propagated in a large quantity and produced incorrectly classified points.

As seen in Figure 10, the pure classification result of the acquired RGB images were—by contrast—correct and much more precise than the resulting point cloud. For the complete processing pipeline, an offset up to 100 pixels was measured, depending on the image region, as the distortion increased at the image's borders.



Figure 10. Original RGB image (left). Semantic segmentation produced by DL approach (right).

In contrast, the semantic-based approach detected more trees and buildings by using the topological relations that elements had between them, allowing classification of portions according to their proximity to each other. This made it possible to relate unclassified portions to already classified portions and thus avoid misclassification. Moreover, the descriptive geometries of trees, buildings and cars were different from each other. The semantic approach correctly dissociated these classes.

The difference in accuracy between the KB approach and the DL approach was maximal for streetlight semantic segmentation. On the one hand, this could be explained by the projection errors in the DL approach, where the thinnest elements were the most difficult to project well. On the other hand, the constant distance between two streetlights ensured a strong relational link for semantic-based approaches, allowing the KB approach to deduce a streetlight position relative to another detected streetlight. Figure 11 illustrates an isolated view of the streetlight semantic segmentation for both studied approaches.

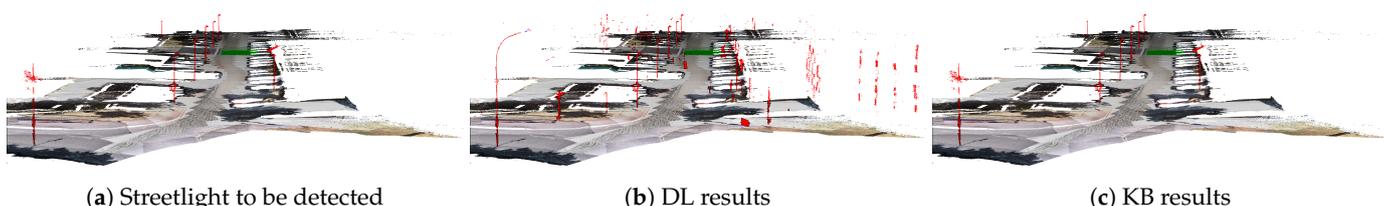


Figure 11. Streetlight semantic segmentation results (red).

The DL approach was more precise regarding the semantic segmentation of the class of building. Buildings were mainly identified using criteria of their geometry in the KB approach. The geometry of buildings was not semantically defined precisely enough in the KB approach. This is the main reason why the semantic-based approach did not detect the buildings. Thus, groups of points composed of bush and tree presented the same geometric criteria and were misidentified by the KB approach, as shown in Figure 12. If enough semantic rules could be expressed, then the KB approach could detect complex building structures and differentiate them from other elements.

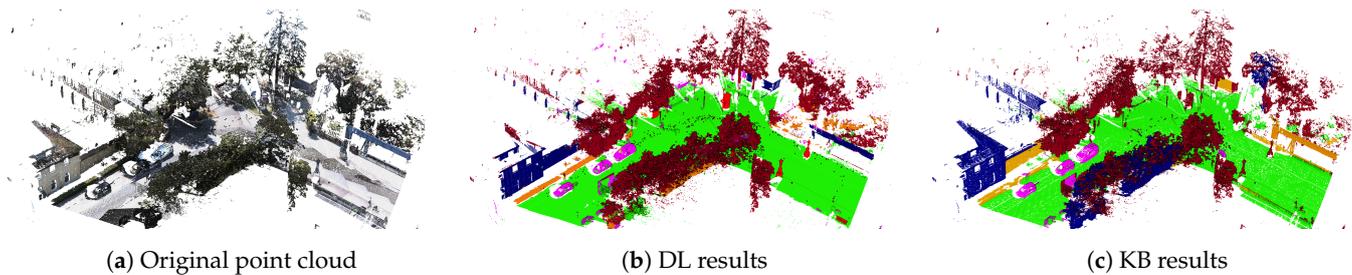


Figure 12. Semantic segmentation sample of a challenging Point cloud.

However, the use of the building's geometric characteristics allowed the studied KB approach to detect a larger set of buildings (better recall) than the DL approach, as shown in Figure 13.

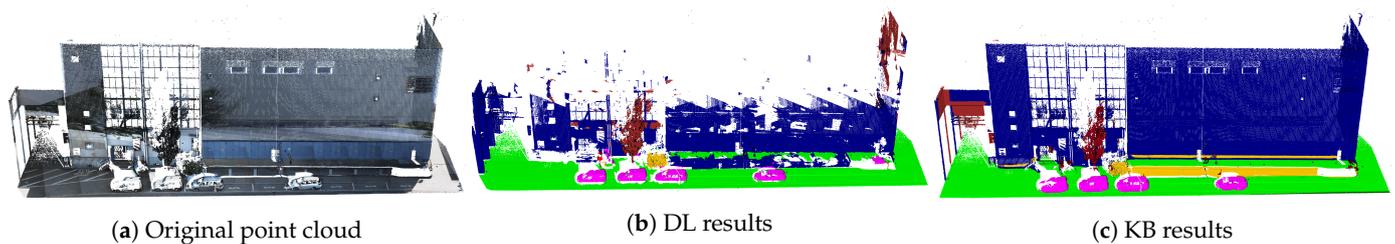


Figure 13. Semantic segmentation sample of Point cloud representing the façade of a building.

Besides the projection errors, the DL approach classification was limited by the cameras' fields of view. Especially high and/or narrow buildings, which were not entirely captured by the RGB cameras, were unable to be processed and therefore unable to be detected correctly. As seen in Figure 13 (unclassified points in blue), less than half of the façade was classified correctly. On the original image in Figure 14, bottom right, the field of view of the current camera can be seen. This translates to the classification result on the bottom left side.

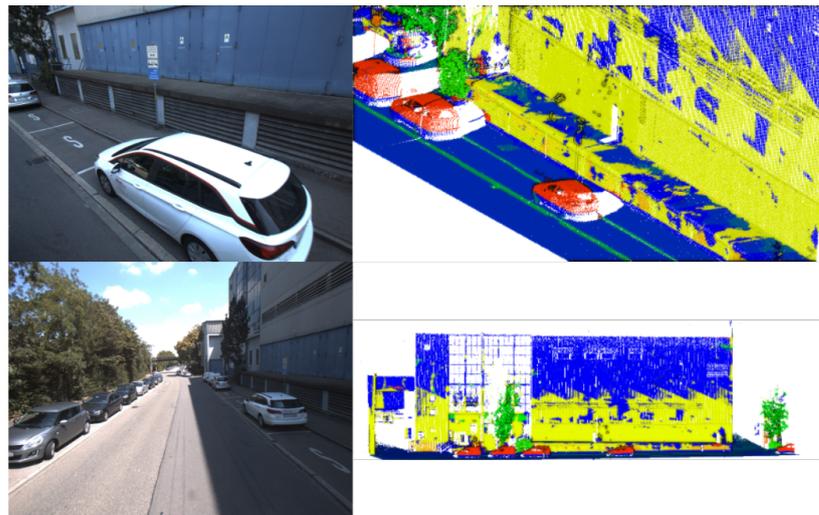


Figure 14. Original image (left), semantic segmentation produced by DL approach (right).

6. Discussion

DL and KB approaches present opposite strategies for semantic segmentation of point clouds. DL approaches use a bottom-up strategy (from data to object representation), whereas the KB approaches use a top-down strategy (from the ontological object representation to the data). Due to the acquisition process, the elements contained in the data are partially acquired (one side), discontinuous (occluded) and noisy. These characteristics represent common challenges in the semantic segmentation of elements and lead to high processing complexity. Such complexity allows the studying of the robustness of both DL and KB approaches. Therefore, it is pertinent chosen to compare the efficiency of these two approaches by studying the results obtained by implementing each of them on a point cloud acquired by mobile mapping technology.

The two studied implementations of the DL and KB approaches obtained a mean IoU score of 51.43% and 65.21% , respectively, on the studied data set. The other approaches in the literature [45] used on similar data sets have a mean IoU between 39% and 63% . Compared to these approaches, both DL and KB implementations have a good quality with slightly better results for the KB approach on the data set studied.

The KB implementation was particularly effective in detecting elements with clearly defined geometrical, topological, or relational characteristics, such as buildings, trees, cars, and streetlights. The implementation of DL, on the other hand, is more effective for detecting elements with characteristics that are more difficult to describe explicitly, such as bushes. For this type of object, the color information helps in particular to detect the objects. It might also explain the better score for DL, as the KB approach only uses geometries and does not consider color information. This fact is not a general characteristic; it is rather a practical one due to the actual state of development of the KB approach.

The difference in quality is mainly due to the fact that the DL approach has many projection errors, “invisible” chunks, where no 2D images were acquired, and so on. The projection of the 2D classification errors into the 3D point cloud decreases the DL approach’s efficiency, mainly by diminishing the recall score. Better quality for the registration between images and point cloud should reduce this effect.

A general difference in both approaches is the way they formulate the base for the segmentation. DL uses implicit content represented in the images and accompanied with their annotations. This base is fixed through the existing images and the training process and only contains information represented there. The KB approach, in turn, formulates the base in an explicit way. It uses characteristics such as geometrical and topological aspects and many other elements belonging to objects, data, environment, acquisition process, and more with their logical connection. This provides the possibility to add and/or modify the knowledge at any time, providing great flexibility and room for optimizations as needed.

For that reason, the quality of the logical representations also has a strong impact on KB approaches' results. However, expert users can always improve the logical descriptions to increase the quality. In contrast, DL approaches use patterns learned within a training phase to represent the detected elements. Therefore, these approaches depend directly on the quality of their learning in relation to the data to be processed. Changes or additions are more complex to perform for DL approaches.

However, image-based pattern recognition's strength allows the DL approach to detect even weakly structured objects, such as bushes, very well. The geometrical characteristics are not decisive here. Certainly, this is critical for the KB approach. The more similar characteristics of objects to be separated, the more difficult it is to describe them uniquely. KB approaches cannot formulate object descriptions in such a way as to represent all cases in the data, and they face difficulties detecting objects that have no or few explicit characteristics. Thus, when elements are not distinctively described, the KB approach generates results below the DL approach level.

Considering the extension aspect of these two approaches, both require the intervention of humans. Updating the DL pipeline with a new class usually requires complete retraining and new annotations, whereas updating KB approaches requires adding explicit knowledge of a new class. Therefore, both approaches are extendable but with different efforts and consequences. New annotations for the DL approach require human input and computation effort for the training, whereas the KB approach requires experts to define explicit knowledge. However, this enrichment process has the advantage that it can be automated by approaches such as knowledge-based self-learning [41,46].

To conclude, the DL approach using exclusively RGB images for semantic segmentation provides good segmentation in images but is limited in 3D regarding projection errors. It also has the advantage of detecting "unstructured" classes like bushes because it does not only rely upon geometry but on other patterns, too. Objects with unconventional geometric characteristics such as a curved garden fence or a very stylized building should be detected. Nevertheless, when there are elements that cause occlusion, such as trees in front of a building, the DL approach has difficulty projecting the points due to the camera's field of view, which causes projection errors. Additionally, DL needs a sufficient amount of data in order to derive reliable patterns. This results in particular problems for smaller objects, such as poles or streetlights.

In contrast, KB approaches directly process every point and can segment every type of class and deal with occlusion by integrating sensing process knowledge [41]. This allows the integration of knowledge about position and viewing angle of the moving acquisition system in order to identify areas with occlusions, as well as knowledge of the reflectivity of the elements and of the type of acquisition system to determine the density of some areas. For example, high reflectivity objects are acquired with very low density by systems based on laser scanners.

The KB approach shows strength in the semantic segmentation accuracy of structured objects that have distinct characteristics, whereas the DL approach using RGB images for the segmentation shows strength in detecting less structured objects with low geometric characteristics.

In general, the results are compared to given data in a given application case. Modifications in the database will change the outcome, as the impact of the registration error between point cloud and images has shown. Additionally, the knowledge base is not adapted to this particular comparison, so its modification will most probably significantly improve the outcome. Nevertheless, the results presented should be typical for the different natures of the two approaches.

7. Conclusions

This research compares a machine learning approach developed by IPM and a knowledge-based approach called KnowDIP, developed by i3mainz. The DL approach is a supervised deep learning solution that uses a VGG6 Encoder with an FCN8 Architecture

on pure RGB images with a projection of the results onto the point cloud. The KB approach uses standard semantic web technologies such as SPARQL and OWL2 to guide the selection of algorithms from the standard library (PCL) [47] and the classification process according to objects or data characteristics in an iterative process.

The comparison of the two approaches on the same data set highlights strengths and weaknesses for both approaches. This comparison is based on the semantic segmentation of the classes of car, tree, bush, streetlight, ground and building in an urban outdoor point cloud. A mobile mapping system has acquired this point cloud. This point cloud's characteristics are estimated by its color, roughness, density, curvature, and noise. These characteristics have allowed us to highlight challenges in the semantic segmentation.

This application case presents several challenges due to irregular shapes, partially acquired objects, geometrical problems, erroneous colors, and high noise. Still, both approaches provide results close to state-of-the-art approaches [45]. The results have been assessed using quantitative metrics: recall, precision, F1, and IoU scores.

Globally, in the different categories of objects to semantically segment, the KB approach has shown strength in the semantic segmentation of objects with distinct characteristics, such as streetlights. In contrast, the DL approach better detects less structured objects like bushes. The DL approach's efficiency is decreased by projection errors, which impacts the recall score; the KB approach shows a higher precision for four categories out of six. In case of a desired extension or adaptation to another scenario, effort has to be invested in both approaches. On the DL side this is mainly new or extended training, whereas for KB it is extension of the knowledge model. Still, the KB approach has more leeway to improve its results by improving and refining the object description compared to DL and its huge training efforts.

However, it would be interesting to compare the results with additional approaches from these two categories to further study the advantages and disadvantages of each category. Therefore, future work includes (1) extending the study to more data and objects to confirm the results, (2) improving the DL approach by reducing the projection error and improving the KB approach by adding more data features such as color, (3) exploring an approach that combines knowledge and deep learning. Combining the two approaches would enhance the weak streetlight semantic segmentation of the DL approach and the weak bush semantic segmentation of the KB approach. For example, the KB approach could enhance DL approaches using RGB images as a segmentation pipeline. For areas where the FOV of the camera is unable to capture everything, if one already knows some of the points of one façade, the KB approach could easily extract features and recognize the rest of the wall.

In addition, DL could provide implicit characteristics for the segmentation of unstructured elements like bushes, thus increasing the separation from other classes, such as fences. In addition, an integration of DL inside the pipeline of a KB approach could allow the knowledge to drive deep learning segmentation instead of algorithms to fulfill the limit of "unstructured" object segmentation and thus benefit from both approaches.

Author Contributions: Conceptualization, Jean-Jacques Ponciano; methodology, Jean-Jacques Ponciano, Alexander Reiterer, and Frank Boochs; validation, Jean-Jacques Ponciano and Moritz Roetner; formal analysis, Jean-Jacques Ponciano and Moritz Roetner; data curation, Jean-Jacques Ponciano and Moritz Roetner; writing—original draft preparation, Jean-Jacques Ponciano, Moritz Roetner, Alexander Reiterer and Frank Boochs; writing—review and editing, Jean-Jacques Ponciano, Moritz Roetner, Alexander Reiterer and Frank Boochs; visualization, Jean-Jacques Ponciano and Moritz Roetner. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external public funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Ashish Karmacharya and Claire Prudhomme for proofreading the article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DL Deep Learning
KB Knowledge-Based

References

1. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef] [PubMed]
2. Hommeaux, E.P.; Seaborne, A. SPARQL Query Language for RDF. W3C Recommendation. 2008. Available online: <http://www.w3.org/TR/rdf-sparql-query> (accessed on 9 April 2020).
3. Grau, B.C.; Horrocks, I.; Motik, B.; Parsia, B.; Patel-Schneider, P.; Sattler, U. OWL 2: The next step for OWL. *Web Semant.* **2008**. [CrossRef]
4. Díaz-Vilariño, L.; Conde, B.; Lagüela, S.; Lorenzo, H. Automatic detection and segmentation of columns in as-built buildings from point clouds. *Remote Sens.* **2015**, *7*, 15651–15667. [CrossRef]
5. Anagnostopoulos, I.; Pãtrãucean, V.; Brilakis, I.; Vela, P. Detection of Walls, Floors, and Ceilings in Point Cloud Data. Construction Research Congress 2016: Old and New Construction Technologies Converge in Historic San Juan. In *Proceedings of the 2016 Construction Research Congress, CRC 2016*; American Society of Civil Engineers: Reston, VA, USA, 2016; pp. 2302–2311. [CrossRef]
6. Hu, P.; Dong, Z.; Yuan, P.; Liang, F.; Yang, B. Reconstruction of 3D models from point clouds with hybrid representation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. ISPRS Arch.* **2018**, *42*, 449–454. [CrossRef]
7. Drost, B.; Ulrich, M.; Navab, N.; Ilic, S. Model globally, match locally: Efficient and robust 3D object recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2010; pp. 998–1005. [CrossRef]
8. Tombari, F.; Salti, S.; Di Stefano, L. A combined texture-shape descriptor for enhanced 3D feature matching. In *Proceedings of the 2011 18th IEEE International Conference on Image Processing*; IEEE: New York, NY, USA, 2011; pp. 809–812. [CrossRef]
9. Lawin, F.J.; Danelljan, M.; Tosteberg, P.; Bhat, G.; Khan, F.S.; Felsberg, M. Deep projective 3D semantic segmentation. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer International Publishing: Ystad, Sweden, 2017; Volume 10424 LNCS, pp. 95–107. [CrossRef]
10. Nguyen, A.; Le, B. 3D point cloud segmentation: A survey. In *Proceedings of the 2013 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM)*; IEEE: New York, NY, USA, 2013; pp. 225–230.
11. Guo, Y.; Wang, H.; Hu, Q.; Liu, H.; Liu, L.; Bennamoun, M. Deep learning for 3D point clouds: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *1*. [CrossRef]
12. Charles, R.Q.; Su, H.; Kaichun, M.; Guibas, L.J. PointNet: Deep learning on point sets for 3D classification and segmentation. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2017; Volume 2017-Janua, pp. 77–85. [CrossRef]
13. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In *Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017*; Volume 2017-Decem, pp. 5100–5109. Available online: <https://arxiv.org/pdf/1706.02413.pdf> (accessed on 9 April 2020).
14. Engelmann, F.; Kontogianni, T.; Leibe, B. Dilated Point Convolutions: On the Receptive Field Size of Point Convolutions on 3D Point Clouds. *Proc. IEEE Int. Conf. Robot. Autom.* **2020**, 9463–9469. [CrossRef]
15. Streiffer, C.; Raghavendra, R.; Benson, T.; Srivatsa, M. DarNet: A deep learning solution for distracted driving detection. In *Proceedings of the Middleware 2017—Proceedings of the 2017 International Middleware Conference (Industrial Track)*; Zhu, X., Roy, I., Eds.; ACM Press: New York, NY, USA, 2017; pp. 22–28. [CrossRef]
16. Liu, J.; Ni, B.; Li, C.; Yang, J.; Tian, Q. Dynamic points agglomeration for hierarchical point sets learning. *Proc. IEEE Int. Conf. Comput. Vis.* **2019**, 2019, 7545–7554. [CrossRef]
17. Rosu, R.A.; Schütt, P.; Quenzel, J.; Behnke, S. LatticeNet: Fast Point Cloud Segmentation Using Permutohedral Lattices. 2019. Available online: <https://arxiv.org/pdf/1912.05905> (accessed on 9 April 2020).
18. Tatarchenko, M.; Park, J.; Koltun, V.; Zhou, Q.Y. Tangent Convolutions for Dense Prediction in 3D. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; Computer Vision Foundation: Salt Lake City, UT, USA, 2018; pp. 3887–3896. [CrossRef]
19. Milioto, A.; Vizzo, I.; Behley, J.; Stachniss, C. RangeNet ++: Fast and Accurate LiDAR Semantic Segmentation. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*; IEEE: New York, NY, USA, 2019; pp. 4213–4220. [CrossRef]
20. Rethage, D.; Wald, J.; Sturm, J.; Navab, N.; Tombari, F. Fully-Convolutional Point Networks for Large-Scale Point Clouds. 2018. Available online : http://openaccess.thecvf.com/content_ECCV_2018/papers/Dario_Rethage_Fully-Convolutional_Point_Networks_ECCV_2018_paper.pdf (accessed on 9 April 2020).

21. Wang, J.; Sun, B.; Lu, Y. MVPNet: Multi-View Point Regression Networks for 3D Object Reconstruction from A Single Image. 2019. Available online: <https://arxiv.org/pdf/1811.09410> (accessed on 9 April 2020).
22. Belgiu, M.; Tomljenovic, I.; Lampoltshammer, T.J.; Blaschke, T.; Höfle, B. Ontology-based classification of building types detected from airborne laser scanning data. *Remote Sens.* **2014**, *6*, 1347–1366. [[CrossRef](#)]
23. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [[CrossRef](#)]
24. Tsarkov, D.; Horrocks, I. FaCT++ description logic reasoner: System description. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 4130 LNAI, pp. 292–297. [[CrossRef](#)]
25. Durand, N.; Derivaux, S.; Forestier, G.; Wemmert, C.; Gançarski, P.; Boussaid, O.; Puissant, A. Ontology-based object recognition for remote sensing image interpretation. In *Proceedings of the Proceedings—International Conference on Tools with Artificial Intelligence, ICTAI*; IEEE: New York, NY, USA, 2007; Volume 1, pp. 472–479. [[CrossRef](#)]
26. Maillot, N.; Thonnat, M.; Hudelot, C. Ontology based object learning and recognition: Application to image retrieval. In *Proceedings of the International Conference on Tools with Artificial Intelligence, ICTAI*, Boca Raton, FL, USA, 15–17 November 2004; pp. 620–625. [[CrossRef](#)]
27. Ponciano, J.J.; Boochs, F.; Trémeau, A. Knowledge-based object recognition in point clouds and image data sets. *Gis Science-Die Zeitschrift für Geoinformatik* **2017**, *3*, 97–104.
28. Florkova, Z.; Duris, L.; Veselovsky, M.; Sedivý, S.; Kovalova, D. Three-dimensional mobile mapping system and its use in road engineering. *MATEC Web Conf.* **2018**, *196*, 4082. [[CrossRef](#)]
29. Reiterer, A.; Leidinger, M. Mobile Urban Mapping System: MUM. Available online: <https://www.ipm.fraunhofer.de/content/dam/ipm/en/PDFs/product-information/OF/MTS/mobile-urban-mapping-system-MUM.pdf> (accessed on 9 April 2020).
30. Fraunhofer-Institut für Physikalische Messtechnik. Clearance Profile Scanner CPS. Available online: <https://www.ipm.fraunhofer.de/content/dam/ipm/en/PDFs/product-information/OF/MTS/Clearance-Profile-Scanner-CPS.pdf> (accessed on 9 April 2020).
31. Reiterer, A.; Wäschle, K.; Störk, D.; Leydecker, A.; Gitzen, N. Fully automated segmentation of 2D and 3D mobile mapping data for reliable modeling of surface structures using deep learning. *Remote Sens.* **2020**, *12*, 2530. [[CrossRef](#)]
32. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional Architecture for Fast Feature Embedding. 2014. Available online: <https://arxiv.org/pdf/1408.5093v1> (accessed on 9 April 2020).
33. Xu, B.; Jiang, W.; Shan, J.; Zhang, J.; Li, L. Investigation on the weighted RANSAC approaches for building roof plane segmentation from LiDAR point clouds. *Remote Sens.* **2016**, *8*, 5. [[CrossRef](#)]
34. Ben Hmida, H.; Cruz, C.; Boochs, F.; Nicolle, C. From 3D point clouds to semantic objects: An ontology-based detection approach. In *Proceedings of the KEOD 2011—International Conference on Knowledge Engineering and Ontology Development*, Paris, France, 26–29 October 2011; pp. 255–260.
35. Karmacharya, A.; Boochs, F.; Tietz, B. Knowledge guided object detection and identification in 3D point clouds. In *Proceedings of the Videometrics, Range Imaging, and Applications XIII*, Munich, Germany, 22–23 June 2015. [[CrossRef](#)]
36. Dietenbeck, T.; Torkhani, F.; Othmani, A.; Attene, M.; Favreau, J.M. Multi-layer ontologies for integrated 3D shape segmentation and annotation. In *Studies in Computational Intelligence*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; Volume 665, pp. 181–206. [[CrossRef](#)]
37. Ponciano, J.J.; Karmacharya, A.; Wefers, S.; Atorf, P.; Boochs, F. Connected Semantic Concepts as a Base for Optimal Recording and Computer-Based Modelling of Cultural Heritage Objects. In *Structural Analysis of Historical Constructions*; Aguilar, R., Torrealva, D., Moreira, S., Pando, M.A., Ramos, L.F., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 297–304.
38. Ponciano, J.J.; Boochs, F.; Tremeau, A. Identification and classification of objects in 3D point clouds based on a semantic concept. In *Oldenburger 3D-Tage 2019*; Wichmann: Oldenburger, Germany, 2019; pp. 98–106.
39. Ponciano, J.J.; Boochs, F.; Trémeau, A. 3D object recognition through a process based on semantics and consideration of the context. In *Photogrammetrie, Laserscanning, Optische 3D-Messtechnik, Beiträge der Oldenburger 3D-Tage 2020*; Wichmann: Oldenburg, Germany, 2020; pp. 1–12.
40. Ponciano, J.J. Object Detection in Unstructured 3D Data Sets Using Explicit Semantics. Ph.D. Thesis, University of Lyon, Lyon, France, 2019.
41. Ponciano, J.J.; Trémeau, A.; Boochs, F. Automatic detection of objects in 3D point clouds based on exclusively semantic guided processes. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 442. [[CrossRef](#)]
42. Tonietto, L.; Gonzaga, L.; Veronez, M.R.; Kazmierczak, C.d.S.; Arnold, D.C.M.; da Costa, C.A. New Method for Evaluating Surface Roughness Parameters Acquired by Laser Scanning. *Sci. Rep.* **2019**, *9*, 1–16. [[CrossRef](#)]
43. Lari, Z.; Habib, A. Alternative methodologies for the estimation of local point density index: Moving towards adaptive LiDAR data processing. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *39*, B3. [[CrossRef](#)]
44. Zhang, X.; Li, H.; Cheng, Z. Curvature estimation of 3D point cloud surfaces through the fitting of normal section curvatures. *Proc. ASIAGRAPH* **2008**, *2008*, 23–26.
45. Hackel, T.; Savinov, N.; Ladicky, L.; Wegner, J.D.; Schindler, K.; Pollefeys, M. Semantic3D.Net: A New Large-Scale Point Cloud Classification Benchmark. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. *Gottingen* **2017**, *4*, 91–98. [[CrossRef](#)]

-
46. Poux, F.; Ponciano, J.J. Self-learning ontology for instance segmentation of 3d indoor point cloud. In *ISPRS—International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*; Copernicus: Bethel Park, PA, USA, 2020. [[CrossRef](#)]
 47. Rusu, R.B.; Cousins, S. 3D is here: Point Cloud Library (PCL). In Proceedings of the IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011. [[CrossRef](#)]