remote sensing

MDPI

Article

A Prior Level Fusion Approach for the Semantic Segmentation of 3D Point Clouds Using Deep Learning

Zouhair Ballouch [1,2,*], Rafika Hajji [1], Florent Poux [2], Abderrazzaq Kharroubi [2] and Roland Billen [2]

[1] College of Geomatic Sciences and Surveying Engineering, IAV Hassan II, Rabat 6202, Morocco; r.hajji@iav.ac.ma
[2] Geomatics Unit, UR SPHERES, University of Liège, 4000 Liège, Belgium; fpoux@uliege.be (F.P.); akharroubi@uliege.be (A.K.); rbillen@uliege.be (R.B.)
* Correspondence: z.ballouch@iav.ac.ma; Tel.: +32-4-9939-1903

Abstract: Three-dimensional digital models play a pivotal role in city planning, monitoring, and sustainable management of smart and Digital Twin Cities (DTCs). In this context, semantic segmentation of airborne 3D point clouds is crucial for modeling, simulating, and understanding large-scale urban environments. Previous research studies have demonstrated that the performance of 3D semantic segmentation can be improved by fusing 3D point clouds and other data sources. In this paper, a new prior-level fusion approach is proposed for semantic segmentation of large-scale urban areas using optical images and point clouds. The proposed approach uses image classification obtained by the Maximum Likelihood Classifier as the prior knowledge for 3D semantic segmentation. Afterwards, the raster values from classified images are assigned to Lidar point clouds at the data preparation step. Finally, an advanced Deep Learning model (RandLaNet) is adopted to perform the 3D semantic segmentation. The results show that the proposed approach provides good results in terms of both evaluation metrics and visual examination with a higher Intersection over Union (96%) on the created dataset, compared with (92%) for the non-fusion approach.

Keywords: 3D point cloud; aerial images; semantic segmentation; data fusion; deep learning

Citation: Ballouch, Z.; Hajji, R.; Poux, F.; Kharroubi, A.; Billen, R. A Prior Level Fusion Approach for the Semantic Segmentation of 3D Point Clouds Using Deep Learning. *Remote Sens.* **2022**, *14*, 3415. https://doi.org/10.3390/rs14143415

Academic Editor: Joaquín Martínez-Sánchez

Received: 10 June 2022
Accepted: 14 July 2022
Published: 16 July 2022

1. Introduction

Three-dimensional city modeling has significantly advanced in recent decades as we move towards the concept of Digital Twin Cities (DTCs) [1], where 3D point clouds are widely used as a major input [2–4]. The development of a three-dimensional city model requires a detailed 3D survey of the urban fabric. Lidar technology is widely used for this purpose. It allows capturing geometric and spectral information of objects in the form of 3D point clouds. This acquisition system provides a large amount of precise data with a high level of detail, quickly and reliably. Nevertheless, the transition from 3D point clouds to the urban model is tedious, essentially manual, and time-consuming [2]. Today, the major challenge is to automate the process of 3D digital model reconstruction from 3D Lidar point clouds [3] while reducing the costs associated with it. Deep Learning (DL) methods are increasingly used to improve the semantic segmentation of 3D point clouds [4]. Semantically segmented point clouds are the foundation for creating 3D city models. The resulting semantic models are used to create DTCs that support a plethora of urban applications [5].

In the literature, different approaches to reconstructing 3D urban models from Lidar data have been proposed. Among the developed methods, Martinovic et al. [6] proposed a methodology for 3D city modeling using 3D facade splitting, 3D weak architectural principles, and 3D semantic classification. It is a technique that produces state-of-the-art results in terms of computation time and precision. Furthermore, Zhang et al. [7] used a pipeline with residual recurrent, Deep-Q, and Convolutional Neural Networks (CNN) to classify and reconstruct urban models from 3D Lidar data. Additionally, Murtiyoso et al. [8]

and Gobeawan et al. [9] presented two workflows for the generation of CityGML models for roof extraction and tree objects from point clouds, respectively. Moreover, several research teams have focused on merging the point clouds with other data sources to take advantage of the benefits of each. For instance, Loutfia et al. [10] developed a simple semi-automatic methodology to generate a 3D digital model for the urban environment based on the fusion of ortho-rectified imagery and Lidar data. In the proposed workflow, data semantic segmentation was carried out with an overall precision of almost 83.51%. The obtained results showed that the proposed methodology could successfully detect several types of buildings, and the Level of Detail (LoD2) was created by integrating the roof structures in the model [10]. Similarly, Kwak et al. [11] introduced an innovative framework for fully automated building model generation by exploiting the advantages of images and Lidar datasets. The main drawback of the proposed methodology was that it could only model the types of buildings that decompose into rectangles. Comparably, Chen et al. [12] obtained the buildings' present status and their reconstruction models by integrating Terrestrial Laser Scanning (TLS) and UAV (Unmanned Aerial Vehicle) photogrammetry.

Two main stages are essential to building a three-dimensional city model from 3D point clouds: semantic segmentation and 3D modeling of the resulting semantic classes. The first consists of assigning semantic information for each point based on homogeneous criteria [13]. In the literature, many developments were conducted in the field of 3D semantic segmentation of point clouds, which can be classified into three families. The first one is based on the raw point clouds; the second is based on a derived product from the point clouds; the third combines 3D point clouds and additional information (optical images, classified images, etc.). The richness and the accuracy of a 3D urban model created from point clouds depend on the acquisition, semantic segmentation, and modeling processes.

DL in geospatial sciences has been an active research field since the first CNN (Convolutional Neural Network) was developed for road network extraction [14]. Thanks to their capacity for processing large multi-source data with good performance, DL techniques revolutionize the domain of computer vision and are state-of-the-art in several tasks, including semantic segmentation [15,16]. Now, there is a lot of interest in developing DL algorithms for processing three-dimensional spatial data.

For the 3D semantic segmentation task, several papers have stated that the fusion of 3D point clouds with other sources (drone images, satellite images, etc.) is promising [17–20] thanks to the planimetric continuity of the images and the altimetric precision of point clouds. Currently, the scientific research in this niche of multi-source data fusion for semantic segmentation is oriented more towards the use of large amounts of additional information (point clouds, multispectral, hyperspectral, etc.). It requires significant financial and material resources, as well as a lot of computational memory and consequently a high computation time. Furthermore, these data-intensive approaches need to collect different types of data in a minimal time interval to avoid any change in the urban environment [21]. In addition, some information would not add much to the differentiation of urban objects. This motivates us to develop a new methodology of fusion that requires less additional information while ensuring high performance.

In this paper, a semantic segmentation approach was developed. It is based on multi-source data (raw point clouds and aerial images) and adopted an advanced deep neural network model. The proposed process can serve as an operational methodology to extract the urban fabric from point clouds and images with better accuracy. It uses a standard method for image classification, in which the training areas were chosen according to the classes present in the Lidar dataset. This technique solves the problem posed by the incoherence of the semantic classes present in the Lidar and image datasets.

To briefly summarize, this paper makes the following four major contributions:

- A less data-intensive fusion approach for 3D semantic segmentation using optical imagery and 3D point clouds;

- An adaptation of an advanced DL method (RandLaNet) to improve the performance of three-dimensional semantic segmentation;
- A solution to solve the problem of the incoherence of the semantic classes present in the Lidar and image datasets at the fusion step;
- A new airborne 3D Lidar dataset for semantic segmentation.

The present paper is structured as follows: In Section 2, the main developments in fusion-based approaches for semantic segmentation of Lidar point clouds are presented. Section 3 provides a comprehensive description of the proposed fusion approach. The experiments and results analysis are the subjects of Section 4. Finally, the paper ends with a conclusion.

2. Related Work

With the increasing demand for three-dimensional land use and urban classification, 3D semantic segmentation of multi-sensor data has become a current research topic. Data fusion methodologies have achieved good results in semantic segmentation [22], and several studies have demonstrated that fusing 3D point clouds and image data can improve segmentation results [23–25].

Various datasets available online, such as S3DIS [26], Semantic3D [27], SensatUrban [28], etc., have further boosted the scientific research of DL on 3D Lidar data, with an increasing number of techniques being proposed to address several problems related to 3D point cloud processing, mainly 3D semantic segmentation [4]. There has been an increasing number of research studies about adapting DL techniques or introducing new ones to semantically segment 3D point clouds. The developed methodologies can be classified into four methods: (1) projection of the point cloud into a 3D occupancy grid such as in [29]; (2) projection of the point cloud on images, and then the semantic segmentation of each image using DL techniques of image semantic segmentation [30]; (3) the use of CRFs to work more on graphs of the cloud as in the case of the SegCloud technique [31] or more by conducting convolutions on graphs as in the case of the SPGraph method [32]; (4) the use of networks that directly consume the point clouds and that can respect the ensemblist properties of a point cloud such as RandLaNet [33]. However, CNNs do not yet obtain similar performance on 3D point clouds as those achieved for image or voice analysis [32]. This opens the way to intensify the scientific research in this direction to enhance their performance.

Recently, research studies concluded that Lidar and multispectral images have distinct characteristics that render them better in several applications [23,34]. The fusion of multispectral images and 3D point clouds would achieve good performance in several applications compared to using a single type of data source. Indeed, the imagery, although relevant for the delineation of accurate object contours, is less suitable for the acquisition of detailed surface models. Lidar data, while considered a major input for the production of very detailed surface models, is less suitable for the delimitation of object limits [23] and can simply distinguish urban objects based on height values. Furthermore, due to the lack of spectral information, Lidar data can present semantic segmentation confusion between some urban objects (e.g., artificial objects and natural objects); consequently, the fusion of multispectral images and 3D point clouds can compensate for each other [23] towards more accurate and reliable semantic segmentation results [22].

Four fusion levels exist to merge Lidar and image data [35]. The first one is prior-level fusion. It assigns 2D land cover (prior knowledge) from a multispectral image to the 3D Lidar point clouds and then uses a DL technique to obtain 3D semantic segmentation results. The second is point-level fusion which assigns spectral information from image data to the points and then trains the classifier using a deep neural network to classify the 3D point clouds with multispectral information. The third is feature-level fusion which concatenates the features extracted from 3D points clouds and image data by a deep neural network and deep convolutional neural network, respectively. After concatenation, the features can be fed to an MLP (MultiLayer Perceptron) to derive the 3D semantic segmentation results. The

fourth is decision-level fusion, which consists of semantically segmenting the 3D Lidar data and multispectral image to obtain 3D and 2D semantic segmentation results, respectively. Subsequently, the two types of data are combined using a fusion technique as a heuristic fusion rule [36]. In this research, a new prior-level approach is proposed, in which the classified images and the raw point clouds are linked and then classified by an advanced deep neural network structure. The major objective is to improve the performance of 3D semantic segmentation.

The previous methods can be classified into two categories: (1) images based approaches and (2) point clouds-based approaches.

2.1. Image-Based Approaches

In these approaches, 3D point clouds represent auxiliary data for 2D urban semantic segmentation, while the multispectral image is the primary data. Point clouds are usually rasterized to Digital Surface Models (DSM) and other structural features, notably deviation angle and height difference.

Past research studies demonstrated the potential of the use of multi-source aerial data for semantic segmentation, where the 3D point cloud is transformed into a regular form that is easy to manipulate and segment [37]. The first study that showed the difficulty of differentiating regions with similar spectral features using only multispectral data was proposed by [38], where the authors used DSMs as a complementary feature to further improve the semantic segmentation results. They investigated four fusion processes based on the proposed DSMF (DSM Fusion) module to highlight the most suitable method and then designed four DSMFNets (DSM Fusion Networks) according to the corresponding process. The proposed methodologies were evaluated using the Vaihingen dataset, and all DSMFNets attained favorable results, especially DSMFNet-1, which reached an overall accuracy of 91.5% on the test dataset. In the same direction, Pan et al. [39] presented a novel CNN-based methodology named FSN (Fine Segmentation Network) for semantic segmentation of Lidar data and high-resolution images. It follows the encoder–decoder paradigm, and multi-sensor fusion is realized at the feature level using MLP (Multi-Layer Perceptron). The evaluation of this process using ISPRS (International Society for Photogrammetry and Remote Sensing) Vaihingen and Potsdam benchmarks shows that this methodology can bring considerable improvements to other related networks. Furthermore, Zhang et al. [40] proposed a fusion method for semantic segmentation of DSMs with infrared or color imagery. They deducted an optimized scheme for the fusion of layers with elevation and image into a single FCN (Fully Convolutional Networks) model. The methodology was evaluated using the ISPRS Potsdam dataset and the Vaihingen 2D Semantic Labeling dataset and demonstrated significant potential. Comparably, Lodha et al. [41] transformed Lidar data into a regular bidimensional grid, which they georegistered to grey-scale airborne imagery of the same grid size. After fusing the intensity and height data, they generated a 5D feature space of image intensity, height, normal variation, height variation, and Lidar intensity. The work achieved a precision of around 92% using the "AdaBoost.M2" extension for multi-class categorization. Furthermore, Weinmann et al. [42] proposed the fusion of multispectral, hyperspectral, color, and 3D point clouds collected by aerial sensor platforms for semantic segmentation in urban areas. The MUUFL Gulfport Hyperspectral and Lidar aerial datasets were used to assess the potential of the combination of different feature sets. The results showed good quality, even for a complex scene collected with a low spatial resolution. Similarly, Onojeghuo et al. [43] proposed a framework for combining Lidar data with hyperspectral and multispectral imagery for object-based habitat mapping. The integration of spectral information with all Lidar-derived measures produced a good overall semantic segmentation.

To sum up, previous studies state that although the networks have the strength to utilize the convolution operation for both elevation information and multispectral image, data may be distorted principally in case of sparse data interpolation. This distortion can affect the results of semantic segmentation depending upon transformation techniques

or the efficacy of the interpolation. In addition, the transformation of 3D point clouds into DSM or 2.5D data can provide obscure data, but, in terms of the prospects of fusion techniques by DL methods, these methods are relatively simpler and easier, as they consider the geometric information as a two-dimensional image representation [17].

2.2. Point Clouds Based Approaches

In these methods, 3D point clouds play a key role in 3D semantic segmentation; the multispectral image represents the auxiliary data, and its spectral information is often simply interpolated as an attribute of 3D point clouds [44].

Among the methodologies developed in this sense, Poliyapram et al. [17] proposed a neural network for aerial image and 3D points clouds point-wise fusion (PMNet) that respects the permutation invariance characteristics of 3D Lidar data. The major objective of this work is to improve the semantic segmentation of 3D point clouds by fusing additional aerial images acquired from the same geographical area. The comparative study conducted using two datasets collected from the complex urban area of the University of Osaka and Houston, Japan, shows that the proposed network fusion "PointNet (XYZIRGB)" surpasses the non-fusion network "PointNet (XYZI)" [17]. Another fusion method named LIF-Seg was proposed in [18]. It is simple and makes full use of the contextual information of image data. The obtained results show performance superior to state of the art methods by a large margin [18]. On the other hand, some research works are based on extracting features from the image data using a neural network and merging them with the Lidar data as in [19], which demonstrated that additional spectral information improves the semantic segmentation results of 3D points. Furthermore, Megahed et al. [34] developed a methodology by which Lidar data were first georegistered to airborne imagery of the same location so that each point inherits its corresponding spectral information. The georegistration added red, green, blue, and near-infrared bands to the Lidar's intensity and height feature space as well as the calculated normalized difference vegetation index. The addition of spectral characteristics to the Lidar's height values boomed the semantic segmentation results to surpass 97%. Semantic segmentation errors occurred among different semantic classes due to independent acquisition of airborne imagery and Lidar data as well as orthorectification and shadow problems from airborne imagery. Furthermore, Chen et al. [36] proposed a fusion method of semantic segmentation that combines multispectral information, including the near-infrared, red, etc., and point clouds. The proposed method achieved global accuracy of 82.47% on the ISPRS dataset. Finally, the authors of [20] proceed by mapping the preliminary segmentation results obtained by images to point clouds according to their coordinate relationships in order to use the point clouds to extract the plane of buildings directly.

To summarize, the aforementioned approaches, in which 3D point clouds are the primary data, show notable performance, especially in terms of accuracy. Among their benefits, they preserve the original characteristics of point clouds, including precision and topological relationships [37].

2.3. Summary

Scientific research is more oriented to the use of several spatial data attributes (X, Y, Z, red, green, blue, near-infrared, etc.) [34,36,42,43] by developing fusion-based approaches for semantic segmentation. These last ones have shown good performance in terms of precision, efficiency, and robustness. However, they are more data-intensive and require performant computing platforms [21]. This is due to the massive characteristics of the fused data, which can easily exceed the memory limit of desktop computers. To overcome these problems, it seems useful to envisage less costly fusion approaches based on less additional information while maintaining precision and performance. To achieve this objective, a prior-level fusion approach combining images and point clouds is proposed, which is able to improve the performance of semantic segmentation, including contextual image information and geometrical information.

3. Materials and Methods

3.1. Study Areas and Ground Data

To test the developed semantic segmentation process, the aerial images and Lidar point clouds data acquired by EUROSENSE Company are used. These are relative to four urban zones of the region of Flanders (Belgium), where the images were acquired with a resolution of 10 cm. The density of points in these four sites is greater than 128 points/m^2. The different data are acquired at the same time (December 2020) and in the same location (Figure 1). The Lidar data are used to develop a new dataset by manual labeling of point clouds. The created dataset contains labeled point clouds of urban scenes. All points in the clouds have RGB values, XYZ coordinates, and intensity values. The dataset consists of eight training scans with their labels and two test scans. The dataset contains five different classes, which are buildings, water, vegetation, cars, and impervious surfaces (Figures 2 and 3), and will be publicly available online.

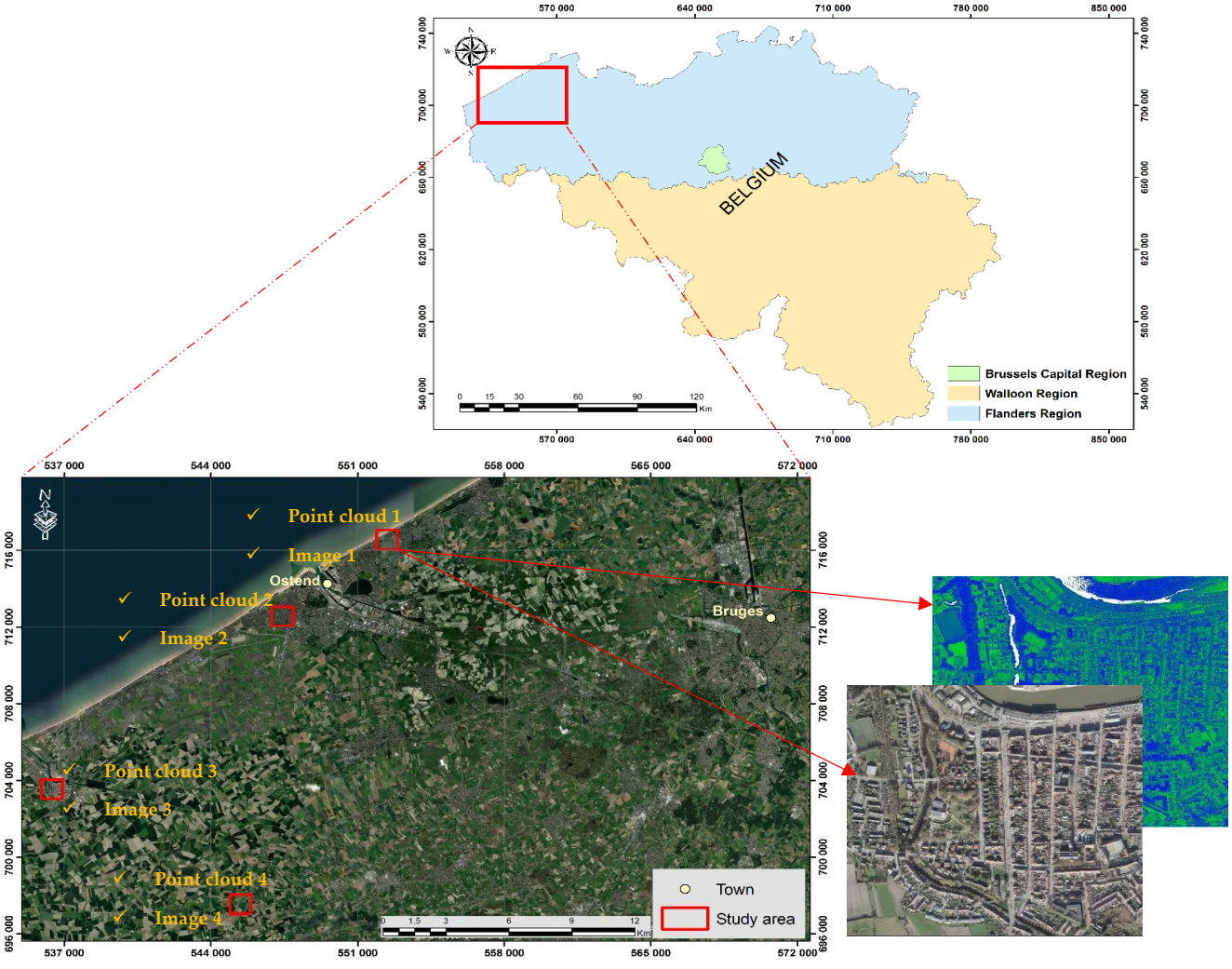

Figure 1. Location of datasets.

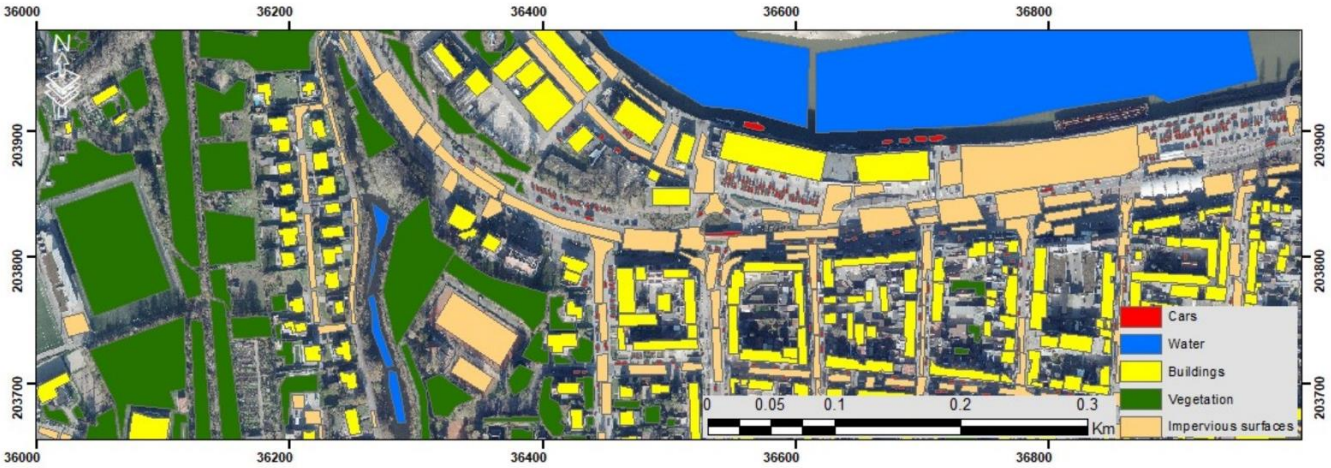

Figure 2. Example of classified point cloud from the created dataset.

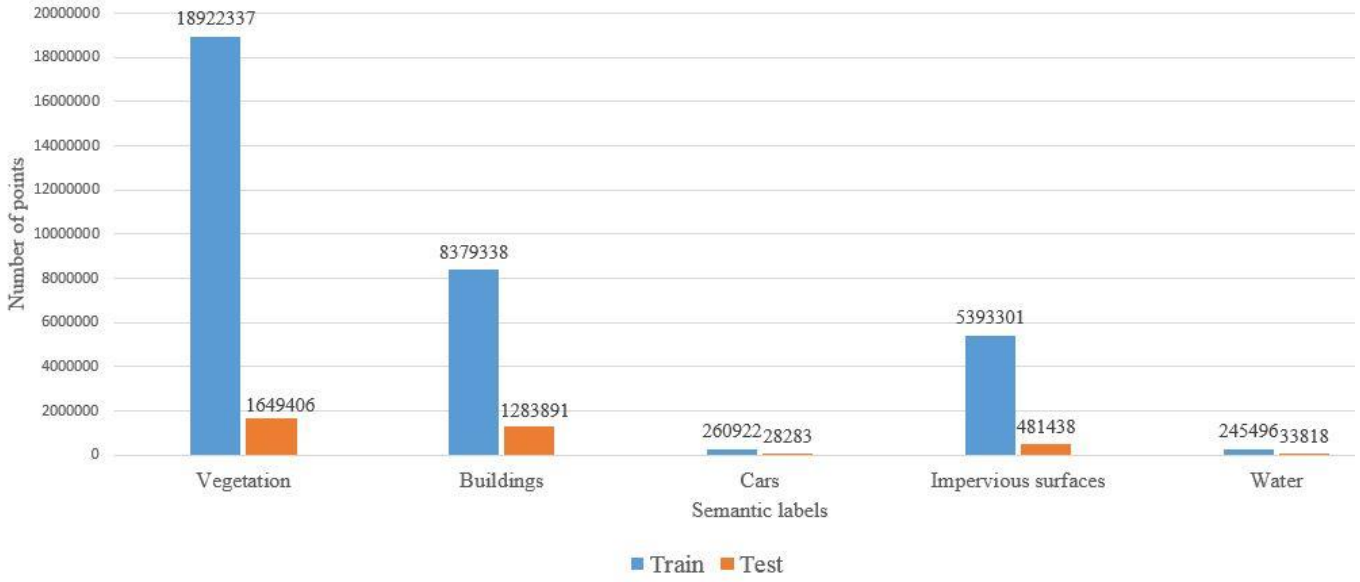

Figure 3. The distribution of different semantic classes in the created dataset.

3.2. Methodology

In the 3D semantic segmentation process, feature extraction from Lidar point clouds and image data plays a crucial role. It can significantly affect the final semantic segmentation results. The proposed approach, named Plf4SSeg (prior-level fusion approach for semantic segmentation), is based on combining geometric and intensity information from 3D point clouds and RGB information from aerial images for 3D urban semantic segmentation.

The methodology (Figure 4) includes two main steps: (1) image classification and (2) fusion of classified images and 3D point clouds.

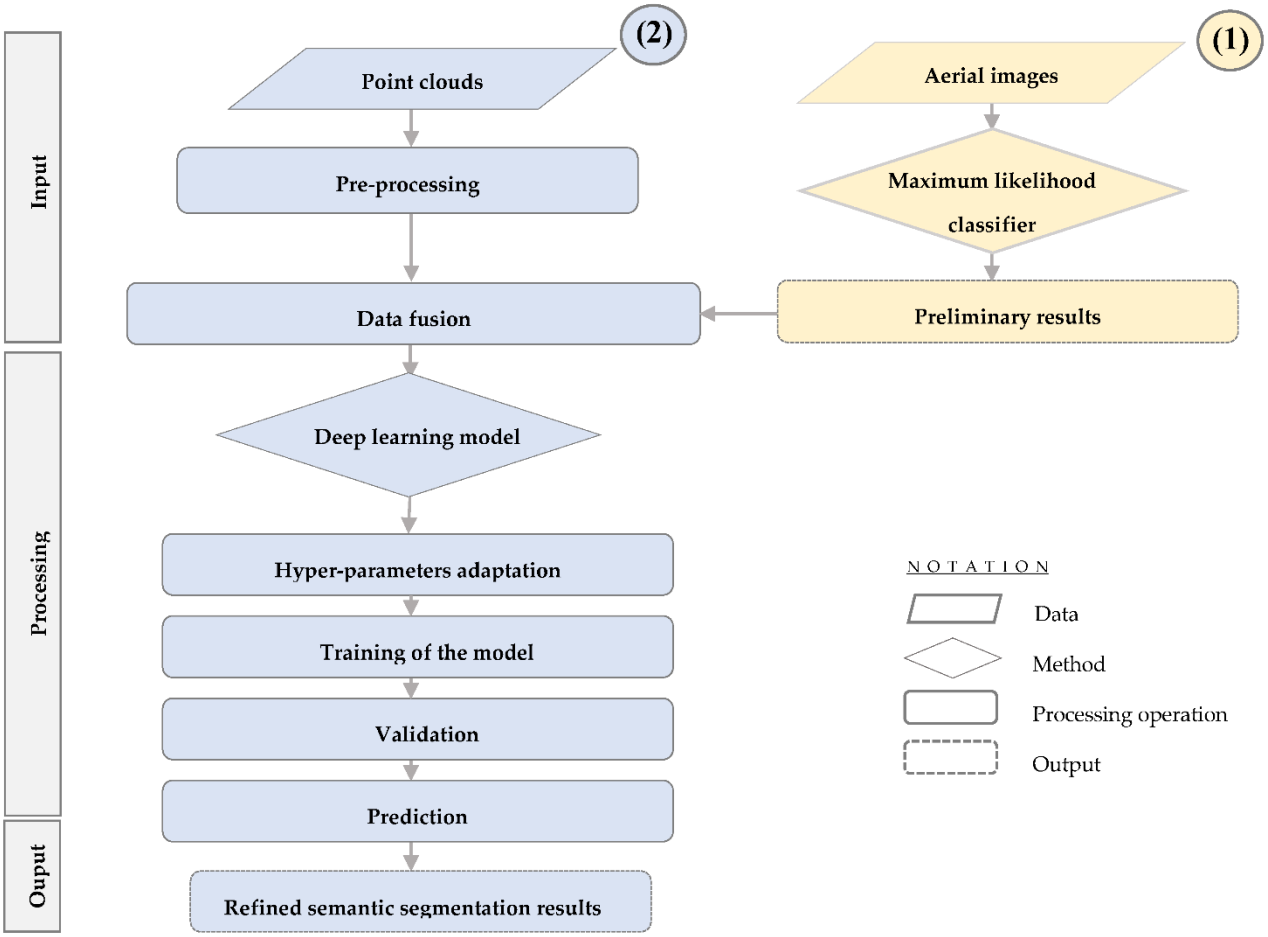

Figure 4. The general workflow of the proposed approach.

3.2.1. Image Classification (Called Prior-Knowledge from RGB-Images)

It is noteworthy to mention that the choice of inputs (X, Y, Z, red, green, blue, etc.) to integrate into the process of semantic segmentation has a significant impact on the quality of the results. In this regard, the image classification generated by a supervised classification algorithm was added as an attribute of the 3D point cloud.

For image classification from the study area, a supervised classification method was applied with the Maximum Likelihood Classifier (MLC). The latter was trained and classified using the ArcGIS 10.5 tool with default parameter settings. Figure 5 summarize the general process followed for image classification.

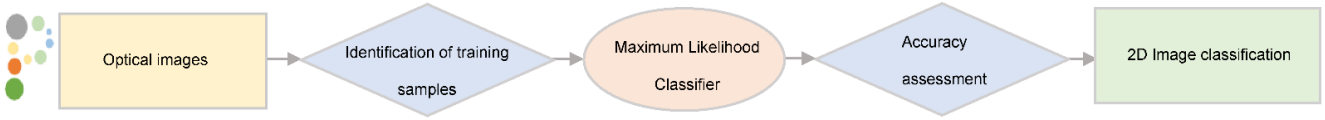

Figure 5. Methodological workflow for image classification.

The MLC is the most common statistical method used for image supervised classification. It is a parametric statistical technique where the analyst first supervises the classification by identifying land cover types, called training areas, as a source of reference data. The image classification process is a standard pixel-based method using a multivariate probability density function of semantic classes [45]. The selection of training samples must be conducted with separability as it has a significant impact on the classification results.

The image classification algorithm should take into consideration the risks of confusion between land use classes. Furthermore, it should be as automatic as possible to make the

image processing easily reproducible and dynamic over time. In this study, MLC was chosen as a parametric classifier that takes into account the variance–covariance within the class distributions as well as its adaptation for normally distributed data owing to its higher precision, as demonstrated by many recent papers [46–48].

The choice of using a non-DL method for image classification instead of a DL method is justified by the difference between the semantic classes (cars, trees, power lines, etc.) present in Lidar and image datasets. The creation of coherence between these classes by aligning them can reduce the semantic details of one of the datasets (for example, by matching the three classes "low vegetation", "shrub", and "tree" from the Lidar dataset to "vegetation" class from the image dataset). Furthermore, the use of the MLC as a supervised method offers the possibility to select the training zones (semantic classes) according to the type of classes present in the Lidar data; this allows obtaining the same semantic classes at the fusion level of classified images and labeled point clouds. Thus, unlike the standard method, DL methods require large amounts of training data.

The four images acquired at the beginning (Figure 1) were split into 10 images to simplify the manipulation of data (in the same way in the case of point clouds). The identification of the sampled site locations for each semantic class was performed by visual interpretation of RGB images. The training samples were populated for each class by creating new geometries using the several drawing tools provided by the ArcGIS tool. A total of five classes were defined: buildings, water, vegetation, cars, and impervious surfaces. The MLC is used depending on the created training sites.

At the end of all these operations of treatment and exploitation of data, the thematic images which highlight the different urban objects in the study area were obtained. The examples of RGB images and their corresponding classification results are illustrated below (Figure 6).

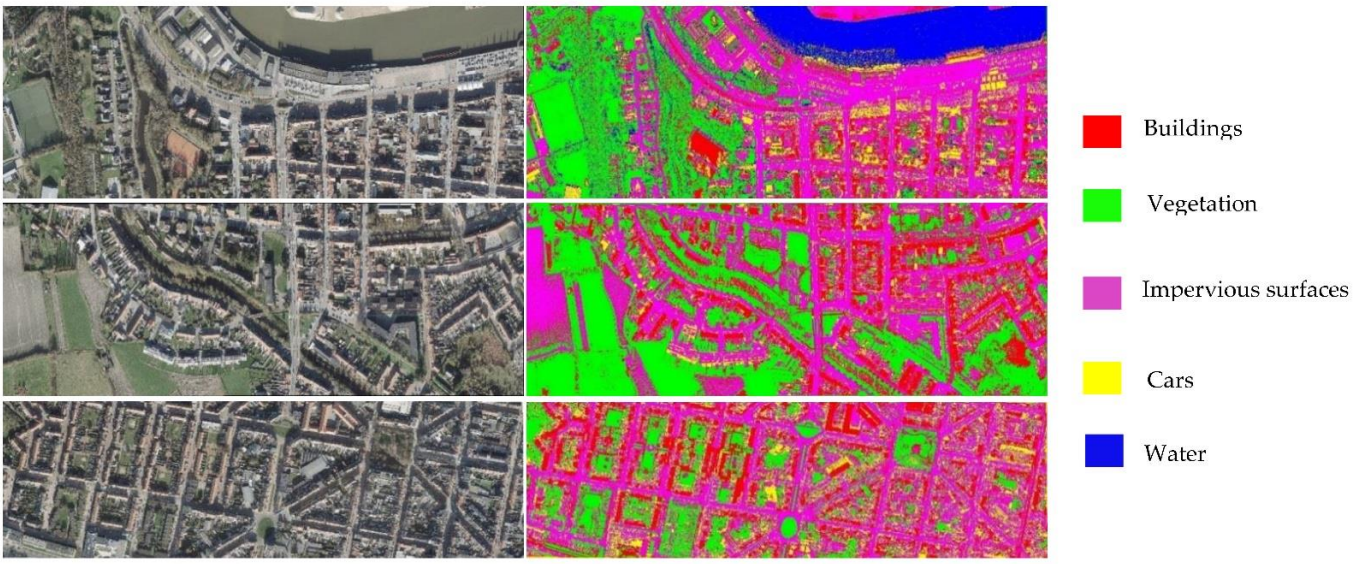

Figure 6. Examples of image classification results.

To summarize, image classification allows the distinction of spectrally homogeneous objects. The combination of this information already classified with point clouds (X, Y, Z, and intensity) can compensate for the limits of point clouds.

3.2.2. Fusion of Classified Images and 3D Point Clouds

A. Assignment of prior knowledge to 3D point clouds

The data acquired by the airborne Lidar contain geometric and radiometric information of objects in the form of point clouds, which vary in resolution and density, depending on the system's technical specifications. Before any exploitation of the raw data, it must be

preprocessed through several steps, including georeferencing, cleaning, etc. Subsequently, due to the manipulation of a set of images collected in different zones, the preliminary image classification results are obtained using the MLC described above.

Afterwards, the generation of training data is realized by assigning raster values from each classified image (.Tif) to the corresponding point cloud (.Las) in the Cloud Compare tool. It means that each classified image is added to the corresponding raw point cloud (XYZ, intensity) from the created dataset, based on its (X, Y) coordinates. That is to say, for each (x, y) position of the 3D point cloud, we search for its nearest pixel in the aerial image for data fusion. To do this, the images are first transformed into mesh format by Cloud Compare, and then the raster values from classified images are assigned to the corresponding clouds. The process is applied to all point clouds present in the dataset. The principle of data preparation according to the formalities of the developed process is illustrated below:

$$\text{Point cloud 1} \quad (X1 + Y1 + Z1 + \text{Intensity1} + \text{Image classification 1}) \tag{1}$$

$$\text{Point cloud 2} \quad (X2 + Y2 + Z2 + \text{Intensity2} + \text{Image classification 2}) \tag{2}$$

$$\text{Point cloud n} \quad (\mathbf{X_n} + \mathbf{Y_n} + \mathbf{Z_n} + \mathbf{Intensity_n} + \text{Image classification }\mathbf{_n}) \tag{3}$$

The linked classified images and point clouds are the inputs of the DL model adopted for 3D semantic segmentation. Finally, a high percentage of the data prepared is used for the model training step.

B. Three-Dimensional semantic segmentation

The 3D semantic segmentation algorithm used for this research is the RandLaNet algorithm [33], which is an advanced DL model for semantic segmentation. It treats directly and randomly 3D point clouds based on point sampling without requiring any pre/postprocessing operation. The performance of this DL technique has been evaluated on several public datasets, including Semantic 3D, S3DIS, and Semantic KITTI datasets. It has demonstrated very satisfactory qualitative and quantitative results [33].

Owing to its higher performance, the RandLaNet algorithm has proven itself to be one of the more effective semantic segmentation algorithms in several 3D laser-scanning system applications, including urban mapping, in which it achieves good results, as demonstrated by many recent papers [28,49,50].

The model was trained two times: the first to run the proposed approach; the second to run a process based only on point clouds. During these implementations, the same basic model hyper-parameters were kept after modifying the input tensor.

The choice of a prior-level approach (that is, the addition of the already classified images to the point clouds) is justified by its direct use of semantic information from image classification rather than the original spectral information of the aerial images. Therefore, it offers the fastest convergence. The difference between the predictions made by the Deep Neural Network and the ground truth of the observations used during the training process is minimal. That is, after embedding the semantic information from the image data, the loss reaches a stable state faster and becomes smaller. Thus, the Plf4SSeg approach can fill the gap between 2D and 3D dimensional land cover through a series form. Additionally, two-dimensional image semantic segmentation provides prior knowledge for 3D semantic segmentation, which could guide model-learning as it facilitates the distinction of the different semantic classes, with less confusion between them.

3.2.3. Non-Fusion Approach

To evaluate the proposed less data-intensive approach, it was compared with the approach based only on point clouds where all accomplished approaches used the RandLaNet algorithm and the same dataset (the created dataset) to ensure the fairness of the comparison as much as possible.

Unlike the Plf4SSeg approach, the process based only on point clouds, named the non-fusion approach, directly classifies the 3D point clouds (Figure 7) precisely in terms of (XYZ) coordinates and intensity information.

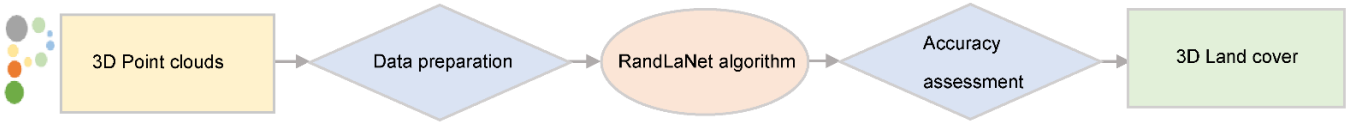

Figure 7. The general workflow of the non-fusion approach.

To properly evaluate both approaches, the same process was followed for data preparation. In addition to the same hyperparameters (batch size, learning rate, epochs, etc.), the same techniques (metrics and visual quality) were employed for the evaluation of model predictions. After training and model validation in both cases, a set of test data from the created dataset was used to evaluate the quality of predictions by comparing the field reality and the model output in both approaches.

4. Experiments and Results Analysis

4.1. Implementation

The RandLA-Net model described above was used for the implementation of the Plf4SSeg approach. This choice is justified by the fact that this model uses random point sampling instead of more complex point selection methods. Therefore, it is computationally and memory efficient. Moreover, it introduces a local feature aggregation module in order to progressively increase the receptive field for each tridimensional point, thus, preserving the geometric details.

Additionally, "Ubuntu with python" was used to perform both approaches: it is a GNU/Linux distribution and a grouping of free software that can be adapted by the user. For Python libraries, the choice is not obvious. Indeed, many DL frameworks are available; each has its limitations and its advantages. The Scikit-Learn library was chosen due to its efficiency: this is a free Python library for machine learning, which provides a selection of efficient tools for machine learning and statistical modeling, including semantic segmentation, regression, and clustering via a consistent interface in Python. The TensorFlow deep learning API was used for the implementation of DL architecture. It was developed to simplify the programming and the use of deep networks.

All computations were processed by Python programming language v 3.6, on Ubuntu v 20.04.3. Cloud Compare v 2.11.3 was used to visualize the 3D Lidar point clouds. The code framework of the RandLaNet model adopted was Tensorflow-gpu v 1.14.0. The code was tested with CUDA 11.4. All experiments were conducted on an NVIDIA GeForce RTX 3090. Data analysis was carried out on a workstation with the following specifications: Windows 10 Pro for workstations OS 64-bit, 3.70 GHz processor, and memory of 256G RAM.

The RandLaNet model used for the implementation of the Plf4SSeg approach was implemented by stacking random sampling layers and multiple local feature aggregation. A source code of its original version was used to train and test this DL model. It was published in open access on GitHub (https://github.com/QingyongHu/RandLA-Net (accessed on 15 June 2022)); this code was tested using the prepared data (Each cloud contains: XYZ coordinates, intensity information, and corresponding classified image as an attribute of the cloud). Furthermore, the basic hyper-parameters were kept as they are crucial for the performance, speed, and quality of the algorithm. The Adam optimization algorithm was adopted with an initial learning rate equal to 0.01, an initial noise parameter equal to 3.5, and batch size during training equal to 4. During the test phase, two sets of point clouds (from the created dataset) were prepared according to the formalities of the Plf4SSeg approach (i.e., each point cloud must contain the attributes X, Y, Z, intensity, and image classification). Subsequently, these data were introduced into the pre-trained

network to deduce the semantic labels for each group of homogeneous points without any pre/postprocessing such as block partitioning.

4.2. Results

The performance of the Plf4SSeg approach was evaluated using the created dataset. Several evaluation criteria were adopted. In addition to the metrics (accuracy, recall, F1 score, and overall accuracy), the visual quality of the results was also considered. This section demonstrates the obtained results and provides a comparative analysis with the non-fusion approach, which uses the raw point clouds only.

4.2.1. Metrics

The accuracy of the semantic segmentation results is influenced by several factors, such as the urban context, the DL technique, and the quality of the training and evaluation data. Precision, recall, accuracy, intersection over union, and F1 score are often used to evaluate the effect of a point cloud semantic segmentation [51]. The following are the evaluation metrics that were used to assess the semantic segmentation results:

- Accuracy score is defined as the ratio of true negatives and true positives to all negative and positive observations.

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}}.$$

TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively.

- Recall of a class is the fraction of true positives (TP) among true positives and false negatives (FN).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

- Precision is calculated as the fraction of true positives (TP) among true and false positives (FP).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$

- The intersection over union (IoU) metric is used to quantify the percentage of overlap between ground truth and model output.

$$\text{IoU} = \frac{\text{TP}}{\text{FP} + \text{TP} + \text{FN}}.$$

TP, FP, and FN are true positive, false positive, and false negative, respectively.

- The F1 score of a class is the harmonic mean of the precision rate (P) and recall (R). It combines these two indicators as follows.

$$\text{F1} - \text{score} = \frac{2\,(\text{R} * \text{P})}{\text{R} + \text{P}}.$$

- A confusion matrix is a good indicator of the performance of a semantic segmentation model by measuring the quality of its results. Each row corresponds to a real class; each column corresponds to an estimated class.

4.2.2. Quantitative and Qualitative Assessments

As already mentioned, the results of the evaluation of both metrics and visual examination of the proposed process are presented in Table 1. Subsequently, the results obtained were compared with the non-fusion approach (Table 2). The objective was to study the contribution of data fusion to semantic segmentation quality.

A. Results of Plf4SSeg approach

Table 1. Quantitative results of Plf4SSeg approach.

The Dataset Class	F1-Score	Intersection over Union
Buildings	0.997	0.996
Vegetation	0.994	0.990
Impervious surfaces	0.945	0.901
Cars	0.952	0.913
Water	0.224	0.126

Table 2. Comparison of the Plf4SSeg approach and the non-fusion approach.

	Non-Fusion Approach	Plf4SSeg Approach
Accuracy	0.959	0.980
F1-score	0.956	0.977
Recall	0.959	0.980
Precision	0.960	0.981
IoU	0.924	0.962

The quality assessment of the semantic segmentation was evaluated through the aforementioned metrics by comparing the output of the model and the reference test data that were labeled. Table 1 below report the resulting metrics.

From Table 1, it appears that the quality of predictions of the different classes is significantly better on the reference samples except for the water class. Additionally, the metrics obtained for the building and vegetation classes are slightly higher than the cars and impervious surfaces classes. The obtained results indicate that the model is reliable for the prediction of unseen data. It should be noted that the low metrics obtained in the water class are justified by its confusion with vegetation classes since they present almost the same altitude. In addition, the Plf4SSeg approach tends to fail in the water class due to the lack of water surfaces in the study area.

The confusion matrix presented below (Figure 8) shows that the model very accurately classified buildings (100% correct), cars (96% correct), impervious surfaces (95% correct), and vegetation (99%). The analysis of this matrix also shows that the confusion between the different semantic classes is low, except for the water class, which is strongly confused with vegetation.

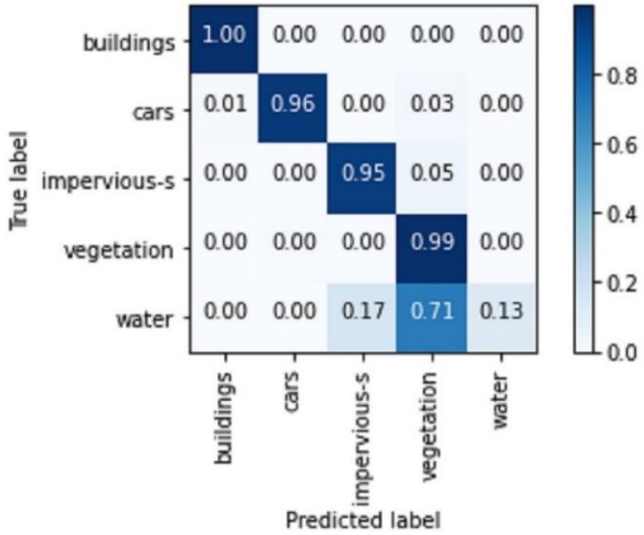

Figure 8. The Normalized confusion matrix.

Finally, the semantic segmentation approach based on data fusion of raw point clouds and classified images highlights the different urban objects present in the study area. To better visually evaluate these semantic segmentation results, these last ones were superimposed on point clouds of the study area. The examples of point clouds (Figure 9A) and their corresponding semantic segmentation results (Figure 9B) are illustrated below (Figure 9).

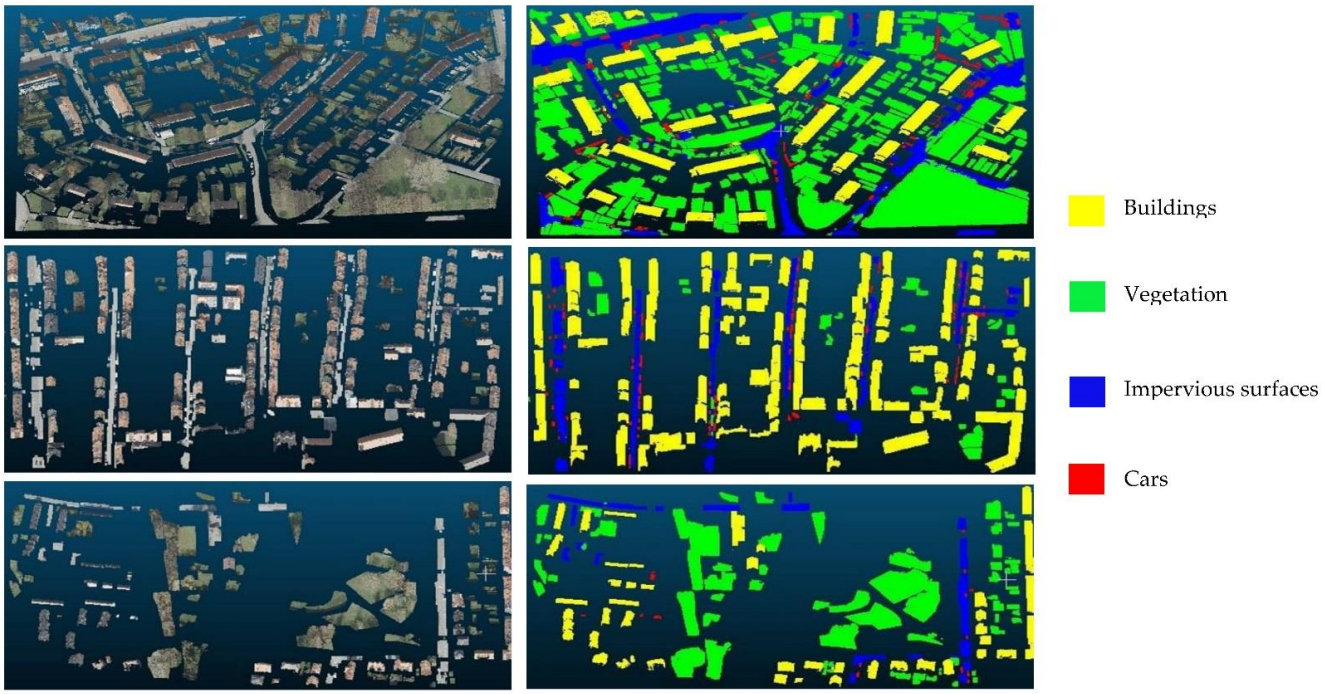

Figure 9. Examples of 3D semantic segmentation results obtained by the Plf4SSeg approach.

At first sight, the obtained predictions are very close to the reference image. This leads us to conclude that the Plf4SSeg approach is successful in associating semantic labels for the different urban objects with better quality, where buildings, vegetation, cars, and impervious surfaces were extracted accurately with clear boundaries.

B. Comparison with the non-fusion approach

In this research, the contribution of classified images in the 3D semantic segmentation using as attributes the raw point clouds and the classification of the corresponding images was studied. The obtained results were then compared with the non-fusion approach, which uses XYZ coordinates and intensity only. Table 2 show the quantitative evaluation of the test results for different approaches.

Table 2 uses metrics such as precision, F1 score, accuracy, recall, and intersection over union to evaluate the performance in detail. RandLaNet (X, Y, Z, intensity information, image classification) shows a significant improvement compared to RandLaNet (X, Y, Z, I) in terms of both precision (0.98) and F1 score (0.97), and hence, it demonstrates that the fusion method is more performant than the one using only (X, Y, Z, I) (Table 2). It significantly outperforms the other process in terms of accuracy (0.98) and IoU (0.96).

The calculation of the different metrics allows us to quantitatively evaluate the quality of the semantic segmentation results produced in the two study cases. The results show a clear improvement in the case of the Plf4SSeg approach compared to the non-fusion methodology with an intersection over union of 0.96 and an F1 score of 0.97. The overall accuracy of the semantic segmentation improves (98%) as well as the other calculated metrics. Consequently, the potential attributes proposed are important to include in the

segmentation process, given their interest in the differentiation of the urban objects present in the captured scene.

To summarize, an adequate parameterization of the DL model with an appropriate choice of the different attributes to be included is relevant for a very good performance of semantic segmentation.

4.3. Discussion

Three-dimensional Lidar semantic segmentation is a fundamental task for producing 3D city models and DTCs for city management and planning. However, semantic segmentation is still a challenging process which requires high investment in terms of material and financial resources. In this paper, a new less-data-intensive fusion DL approach based on merging point clouds and aerial images was proposed to meet this challenge.

The particularity of the Plf4SSeg fusion approach compared to most existing fusion methods is that it requires less additional information by combining Lidar point clouds and classified images. The latter was obtained by a classification of RGB images using the MLC. The majority of users avoid using fusion approaches due to their high cost in terms of additional information, as well as required hardware resources for processing and computing. The Plf4SSeg method offers the possibility of using classified images from different data sources, namely satellite images, UAV images, etc., which increases its feasibility and usability. In addition, the developed methodology is adapted to different Lidar datasets. Indeed, the use of a standard method for image classification offers the possibility to choose the semantic categories according to those present in the 3D Lidar datasets. This technique conserves the semantic richness of the Lidar datasets instead of opting for an adaptation of the semantic classes present in the Lidar and image datasets. Furthermore, compared to the methods from the literature that transform the point cloud into a regular shape, the Plf4SSeg approach treats the 3D Lidar data without any interpolation operation and preserves its original quality.

The Plf4SSeg approach takes into consideration geometric and radiometric information. Additionally, the merging of different data sources was conducted during the data preparation step. This way of combination improves the learning of the DL method, which can positively influence the model prediction results. Finally, the developed semantic segmentation process applies to airborne data acquired in large-scale urban environments, so it is very useful to highlight the different urban objects present in the city scale (buildings, vegetation, etc.). On the other hand, for the training, validation, and testing of the DL technique, an airborne Lidar dataset was created, and that will be published online later. The created dataset presents the main semantic classes that are very useful for different urban applications, which are buildings, vegetation, impervious surfaces, cars, and water. The results are satisfactory for all semantic classes except for the water class, representing a very small percentage in the dataset. The comparative study shows that the Plf4SSeg approach improves all metrics over the non-fusion approach using the test data.

Three-dimensional semantic segmentation results were studied in detail by computing a percentage-based confusion matrix with a ground truth label. In Figure 10 below, A (the Plf4SSeg approach) and B (non-fusion approach) show the percentage-based confusion matrix for a point cloud from the test data, respectively. This percentage-based analysis provides an idea about the percentage of consistent and non-consistent points. The Plf4SSeg approach shows a higher percentage of consistency than the non-fusion approach. Additionally, in the case of the non-fusion approach, confusion in some semantic classes was observed, for example, cars and impervious surfaces with vegetation. However, in the case of the proposed approach, low confusion between these classes was obtained. The height consistency obtained can be justified by the addition of already classified spectral information, which facilitated the distinction of the different classes.

The evaluation of the Plf4SSeg approach that requires less additional information compared to data-intensive approaches combining large amounts of additional information (point clouds, multispectral, hyperspectral, etc.) shows that the developed methodology

can achieve compared or superior results against these expensive methodologies. Some examples of common semantic classes are taken; for example, in the case of the class buildings, higher accuracy was obtained compared to those obtained by [43] at the level of the built-up area class, with all tested techniques using the merged Eagle MNF Lidar datasets. Similarly, in the case of the class of cars, higher accuracy was achieved compared to the one obtained by [36] (71.4), which used the ISPRS dataset. Another example is the revealed confusion between the two semantic classes, buildings and vegetation, in [34], contrary to this work, in which the two semantic classes are well classified (Table 1).

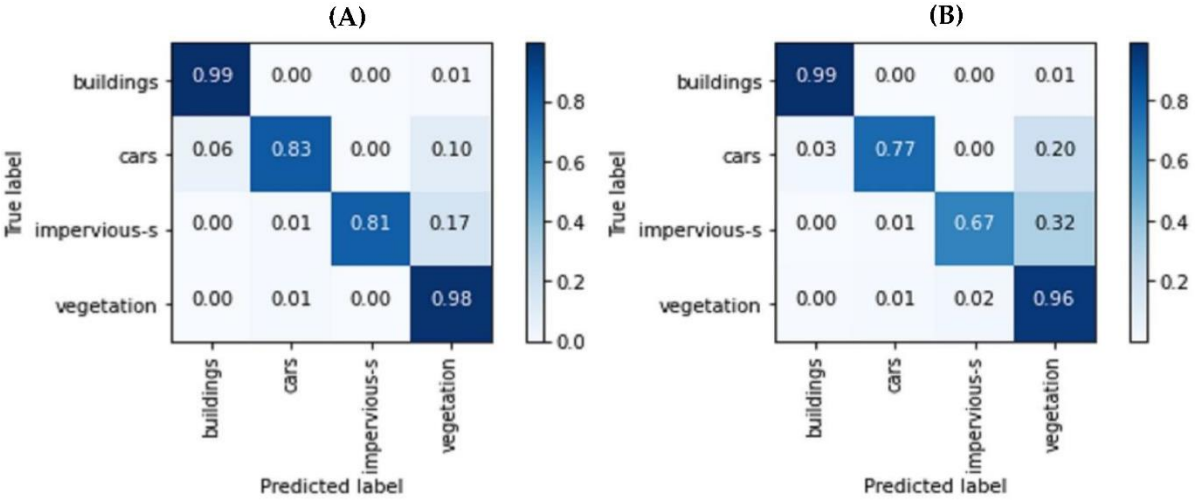

Figure 10. Normalized confusion matrix of the proposed approach (**A**) and the non-fusion approach (**B**).

Finally, it should be noted that this research work presents certain limitations, including the choice of the training zones that is conducted manually in the case of image classification. Additionally, the Plf4SSeg approach should be tested in other urban contexts that contain numerous objects. As a perspective, we suggest investigating the proposed semantic segmentation process in several urban contexts by choosing numerous semantic classes and by also considering the case of other terrestrial and airborne datasets. The objective is to evaluate the performance and the limitations of the proposed approach when confronted with other contexts.

5. Conclusions

In this study, a prior-level and less data-intensive approach for 3D semantic segmentation based on images and airborne point clouds was proposed and compared with a process based only on point clouds. The proposed approach assigns the raster values from each classified image to the corresponding point cloud. Moreover, it adopted an advanced deep neural network (RandLaNet) to improve the performance of 3D semantic segmentation. Another main contribution of the proposed methodology is that the semantic segmentation of aerial images is based on training zones selected accordingly to the semantic classes of the Lidar dataset, which allows solving the problem of the incoherence of the semantic classes present in the Lidar and image datasets. Consequently, the proposed approach was adapted for all Lidar dataset types. Another advantage of the proposed process was its flexibility in the choice of image type to use; that is, all types of images, including satellites, drones, etc., can be used. The Plf4SSeg approach, although it is based on less additional information, demonstrated good performance compared to both the non-fusion process based only on point clouds and the state-of-the-art methods. The experimental results using the created dataset show that the proposed data-intensive approach delivers a good performance, which is manifested mainly in intersection over union (96%) and F1 score (97%) metrics that are high in the 3D semantic segmentation results. Therefore, an adequate parameterization of the DL model with an appropriate choice of the different attributes

to be included allowed us to achieve a very good performance. However, the proposed process was a bit long, and the image classification part required a little human intervention when manual identification of training zones. Low precision was obtained in the water class due to the lack of water surfaces in the study area. We suggest investigating the proposed approach in other urban contexts to evaluate its performance and limitations when confronted with other contexts.

Author Contributions: Conceptualization, Z.B., F.P., R.H. and R.B.; methodology, Z.B., F.P., R.H. and R.B.; validation, Z.B., R.H. and R.B.; writing—original draft preparation, Z.B., R.H. and R.B.; writing—review and editing, Z.B., F.P., R.H., A.K. and R.B.; visualization, Z.B.; supervision, R.H. and R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the EUROSENSE Company for providing the raw data.

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

References

1. Yan, J.; Zlatanova, S.; Aleksandrov, M.; Diakite, A.; Pettit, C.J. Integration of 3D Objects and Terrain for 3D Modelling Supporting the Digital Twin. In Proceedings of the 14th 3D GeoInfo Conference, Singapore, 24–27 September 2019.
2. Wang, R.; Peethambaran, J.; Chen, D. LiDAR Point Clouds to 3-D Urban Models: A Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 606–627. [CrossRef]
3. Macher, H.; Landes, T.; Grussenmeyer, P. From Point Clouds to Building Information Models: 3D Semi-Automatic Reconstruction of Indoors of Existing Buildings. *Appl. Sci.* **2017**, *7*, 1030. [CrossRef]
4. 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.* **2021**, *43*, 4338–4364. [CrossRef] [PubMed]
5. Beil, C.; Kutzner, T.; Schwab, B.; Willenborg, B.; Gawronski, A.; Kolbe, T.H. Integration of 3D Point Clouds with Semantic 3D City Models—Providing Semantic Information Beyond Classification. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *VIII-4/W2-2021*, 105–112. [CrossRef]
6. Martinovic, A.; Knopp, J.; Riemenschneider, H.; Van Gool, L. 3D All The Way: Semantic Segmentation of Urban Scenes From Start to End in 3D. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 4456–4465.
7. Zhang, L.; Zhang, L. Deep Learning-Based Classification and Reconstruction of Residential Scenes From Large-Scale Point Clouds. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 1887–1897. [CrossRef]
8. Murtiyoso, A.; Veriandi, M.; Suwardhi, D.; Soeksmantono, B.; Harto, A.B. Automatic Workflow for Roof Extraction and Generation of 3D CityGML Models from Low-Cost UAV Image-Derived Point Clouds. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 743. [CrossRef]
9. Gobeawan, L.; Lin, E.S.; Tandon, A.; Yee, A.T.K.; Khoo, V.H.S.; Teo, S.N.; Yi, S.; Lim, C.W.; Wong, S.T.; Wise, D.J.; et al. Modeling Trees for Virtual Singapore: From Data Acquisition to CityGML Models. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII-4/W10*, 55–62. [CrossRef]
10. Loutfia, E.; Mahmoud, H.; Amr, A.; Mahmoud, S. 3D Model Reconstruction from Aerial Ortho-Imagery and LiDAR Data. *J. Geomat.* **2017**, *11*, 9.
11. Kwak, E. Automatic 3D Building Model Generation by Integrating LiDAR and Aerial Images Using a Hybrid Approach. Ph.D. Thesis, University of Calgary, Calgary, AB, Canada, 2013. [CrossRef]
12. Chen, X.; Jia, D.; Zhang, W. Integrating UAV Photogrammetry and Terrestrial Laser Scanning for Three-Dimensional Geometrical Modeling of Post-Earthquake County of Beichuan. In Proceedings of the 18th International Conference on Computing in Civil and Building Engineering, São Paulo, Brazil, 18–20 August 2020; Toledo Santos, E., Scheer, S., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 1086–1098.
13. Luo, H.; Khoshelham, K.; Fang, L.; Chen, C. Unsupervised Scene Adaptation for Semantic Segmentation of Urban Mobile Laser Scanning Point Clouds. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 253–267. [CrossRef]
14. Marmanis, D.; Wegner, J.D.; Galliani, S.; Schindler, K.; Datcu, M.; Stilla, U. Semantic Segmentation of Aerial Images with an Ensemble of CNSS. In Proceedings of the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Prague, Czech Republic, 12–19 July 2016; Halounova, L., Schindler, K., Limpouch, A., Šafář, V., Pajdla, T., Mayer, H., Oude Elberink, S., Mallet, C., Rottensteiner, F., Skaloud, J., et al., Eds.; Copernicus Publications: Göttingen, Germany, 2016; Volume III–3, pp. 473–480.
15. Castillo-Navarro, J.; Le Saux, B.; Boulch, A.; Lefèvre, S. Réseaux de Neurones Semi-Supervisés Pour La Segmentation Sémantique En Télédétection. In Proceedings of the Colloque GRETSI sur le Traitement du Signal et des Images, Lille, France, 26–29 August 2019.

16. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A Review on Deep Learning Techniques Applied to Semantic Segmentation. *arXiv* **2017**, arXiv:170406857.
17. Poliyapram, V.; Wang, W.; Nakamura, R. A Point-Wise LiDAR and Image Multimodal Fusion Network (PMNet) for Aerial Point Cloud 3D Semantic Segmentation. *Remote Sens.* **2019**, *11*, 2961. [CrossRef]
18. Zhao, L.; Zhou, H.; Zhu, X.; Song, X.; Li, H.; Tao, W. LIF-Seg: LiDAR and Camera Image Fusion for 3D LiDAR Semantic Segmentation. *arXiv* **2021**, arXiv:210807511.
19. Meyer, G.P.; Charland, J.; Hegde, D.; Laddha, A.; Vallespi-Gonzalez, C. Sensor Fusion for Joint 3D Object Detection and Semantic Segmentation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019; pp. 1230–1237.
20. Zhang, R.; Li, G.; Li, M.; Wang, L. Fusion of Images and Point Clouds for the Semantic Segmentation of Large-Scale 3D Scenes Based on Deep Learning. *ISPRS J. Photogramm. Remote Sens.* **2018**, *143*, 85–96. [CrossRef]
21. Ballouch, Z.; Hajji, R.; Ettarid, M. The Contribution of Deep Learning to the Semantic Segmentation of 3D Point-Clouds in Urban Areas. In Proceedings of the 2020 IEEE International Conference of Moroccan Geomatics (Morgeo), Casablanca, Morocco, 11–13 May 2020; pp. 1–6.
22. Khodadadzadeh, M.; Li, J.; Prasad, S.; Plaza, A. Fusion of Hyperspectral and LiDAR Remote Sensing Data Using Multiple Feature Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2971–2983. [CrossRef]
23. Zhang, J.; Lin, X. Advances in Fusion of Optical Imagery and LiDAR Point Cloud Applied to Photogrammetry and Remote Sensing. *Int. J. Image Data Fusion* **2017**, *8*, 1–31. [CrossRef]
24. Ghamisi, P.; Rasti, B.; Yokoya, N.; Wang, Q.; Hofle, B.; Bruzzone, L.; Bovolo, F.; Chi, M.; Anders, K.; Gloaguen, R.; et al. Multisource and Multitemporal Data Fusion in Remote Sensing: A Comprehensive Review of the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 6–39. [CrossRef]
25. Luo, S.; Wang, C.; Xi, X.; Zeng, H.; Li, D.; Xia, S.; Wang, P. Fusion of Airborne Discrete-Return LiDAR and Hyperspectral Data for Land Cover Classification. *Remote Sens.* **2015**, *8*, 3. [CrossRef]
26. Armeni, I.; Sener, O.; Zamir, A.R.; Jiang, H.; Brilakis, I.; Fischer, M.; Savarese, S. 3D Semantic Parsing of Large-Scale Indoor Spaces. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1534–1543.
27. Hackel, T.; Savinov, N.; Ladicky, L.; Wegner, J.D.; Schindler, K.; Pollefeys, M. Semantic3D.Net: A New Large-Scale Point Cloud Classification Benchmark. *arXiv* **2017**, arXiv:170403847. [CrossRef]
28. Hu, Q.; Yang, B.; Khalid, S.; Xiao, W.; Trigoni, N.; Markham, A. Towards Semantic Segmentation of Urban-Scale 3D Point Clouds: A Dataset, Benchmarks and Challenges. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 4977–4987.
29. Xu, Y.; Hoegner, L.; Tuttas, S.; Stilla, U. Voxel- and Graph-Based Point Cloud Segmentation of 3D Scenes Using Perceptual Grouping Laws. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *IV-1/W1*, 43–50. [CrossRef]
30. Boulch, A.; Saux, B.L.; Audebert, N. Unstructured Point Cloud Semantic Labeling Using Deep Segmentation Networks. In Proceedings of the Eurographics Workshop 3D Object Retrieval, Lyon, France, 23–24 April 2017. [CrossRef]
31. Tchapmi, L.; Choy, C.; Armeni, I.; Gwak, J.; Savarese, S. SEGCloud: Semantic Segmentation of 3D Point Clouds. In Proceedings of the 2017 International Conference on 3D Vision (3DV), Qingdao, China, 10–12 October 2017; pp. 537–547.
32. Landrieu, L.; Simonovsky, M. Large-Scale Point Cloud Semantic Segmentation with Superpoint Graphs. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4558–4567.
33. Hu, Q.; Yang, B.; Xie, L.; Rosa, S.; Guo, Y.; Wang, Z.; Trigoni, N.; Markham, A. RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11105–11114.
34. Megahed, Y.; Shaker, A.; Yan, W.Y. Fusion of Airborne LiDAR Point Clouds and Aerial Images for Heterogeneous Land-Use Urban Mapping. *Remote Sens.* **2021**, *13*, 814. [CrossRef]
35. Ghassemian, H. A Review of Remote Sensing Image Fusion Methods. *Inf. Fusion* **2016**, *32*, 75–89. [CrossRef]
36. Chen, Y.; Liu, X.; Xiao, Y.; Zhao, Q.; Wan, S. Three-Dimensional Urban Land Cover Classification by Prior-Level Fusion of LiDAR Point Cloud and Optical Imagery. *Remote Sens.* **2021**, *13*, 4928. [CrossRef]
37. Ballouch, Z.; Hajji, R.; Ettarid, M. Toward a Deep Learning Approach for Automatic Semantic Segmentation of 3D Lidar Point Clouds in Urban Areas. In *Geospatial Intelligence: Applications and Future Trends*; Barramou, F., El Brirchi, E.H., Mansouri, K., Dehbi, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 67–77. ISBN 978-3-030-80458-9.
38. Cao, Z.; Fu, K.; Lu, X.; Diao, W.; Sun, H.; Yan, M.; Yu, H.; Sun, X. End-to-End DSM Fusion Networks for Semantic Segmentation in High-Resolution Aerial Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1766–1770. [CrossRef]
39. Pan, X.; Gao, L.; Marinoni, A.; Zhang, B.; Yang, F.; Gamba, P. Semantic Labeling of High Resolution Aerial Imagery and LiDAR Data with Fine Segmentation Network. *Remote Sens.* **2018**, *10*, 743. [CrossRef]
40. Zhang, W.; Huang, H.; Schmitz, M.; Sun, X.; Wang, H.; Mayer, H. Effective Fusion of Multi-Modal Remote Sensing Data in a Fully Convolutional Network for Semantic Labeling. *Remote Sens.* **2017**, *10*, 52. [CrossRef]

41. Lodha, S.K.; Fitzpatrick, D.M.; Helmbold, D.P. Aerial Lidar Data Classification Using AdaBoost. In Proceedings of the Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007), Montreal, QC, Canada, 21–23 August 2007; pp. 435–442.
42. Weinmann, M.; Weinmann, M. Fusion of Hyperspectral, Multispectral, Color and 3D Point Cloud Information for the Semantic Interpretation of Urban Environments. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W13*, 1899–1906. [CrossRef]
43. Onojeghuo, A.O.; Onojeghuo, A.R. Object-Based Habitat Mapping Using Very High Spatial Resolution Multispectral and Hyperspectral Imagery with LiDAR Data. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *59*, 79–91. [CrossRef]
44. Yousefhussien, M.; Kelbe, D.J.; Ientilucci, E.J.; Salvaggio, C. A Multi-Scale Fully Convolutional Network for Semantic Labeling of 3D Point Clouds. *ISPRS J. Photogramm. Remote Sens.* **2018**, *143*, 191–204. [CrossRef]
45. Siljander, M.; Adero, N.J.; Gitau, F.; Nyambu, E. Land Use/Land Cover Classification for the Iron Mining Site of Kishushe, Kenya: A Feasibility Study of Traditional and Machine Learning Algorithms. *Afr. J. Min. Entrep. Nat. Resour. Manag.* **2020**, *2*, 115–124.
46. Asad, M.H.; Bais, A. Weed Detection in Canola Fields Using Maximum Likelihood Classification and Deep Convolutional Neural Network. *Inf. Process. Agric.* **2020**, *7*, 535–545. [CrossRef]
47. Gevana, D.; Camacho, L.; Carandang, A.; Camacho, S.; Im, S. Land Use Characterization and Change Detection of a Small Mangrove Area in Banacon Island, Bohol, Philippines Using a Maximum Likelihood Classification Method. *For. Sci. Technol.* **2015**, *11*, 197–205. [CrossRef]
48. Berila, A.; Isufi, F. Two Decades (2000–2020) Measuring Urban Sprawl Using GIS, RS and Landscape Metrics: A Case Study of Municipality of Prishtina (Kosovo). *J. Ecol. Eng.* **2021**, *22*, 114–125. [CrossRef]
49. Cortinhal, T.; Tzelepis, G.; Erdal Aksoy, E. SalsaNext: Fast, Uncertainty-Aware Semantic Segmentation of LiDAR Point Clouds. In *Advances in Visual Computing, Proceedings of the 15th International Symposium on Visual Computing, San Diego, CA, USA, 5–7 October 2020*; Bebis, G., Yin, Z., Kim, E., Bender, J., Subr, K., Kwon, B.C., Zhao, J., Kalkofen, D., Baciu, G., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 207–222.
50. Xu, C.; Wu, B.; Wang, Z.; Zhan, W.; Vajda, P.; Keutzer, K.; Tomizuka, M. SqueezeSegV3: Spatially-Adaptive Convolution for Efficient Point-Cloud Segmentation. In *Computer Vision—ECCV 2020, Proceedings of the 16th European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 1–19.
51. Li, Y.; Tong, G.; Du, X.; Yang, X.; Zhang, J.; Yang, L. A Single Point-Based Multilevel Features Fusion and Pyramid Neighborhood Optimization Method for ALS Point Cloud Classification. *Appl. Sci.* **2019**, *9*, 951. [CrossRef]