GEOGRAPHY DOCTORAL SCHOOL - Faculty of Science
COLLEGE OF GEOMATICS SCIENCE
& SURVEYING ENGINEERING

Enhancing Semantic Segmentation of Large-Scale 3D Point Clouds with Deep Learning Techniques for Urban Digital Twin Creation

Supervisors:

Prof. R. Billen (Uliège)
Prof. R. Hajji (IAV HII)

BALLOUCH Zouhair

Dissertation presented in partial fulfillment of the requirements for the degree of Joint Doctor of Science (PhD):

Geography - Geomatics

November 2024

Enhancing Semantic Segmentation of Large-Scale 3D Point Clouds with Deep Learning Techniques for Urban Digital Twin Creation

Dissertation presented in partial fulfillment of
the requirements for the degree of Joint
Doctor of Science (PhD):

Geomatics – Geography

Morocco - Belgium

BALLOUCH Zouhair

This dissertation has been approved by:

Prof. Dr. Roland Billen	Co-director of research
Prof. Dr. Rafika Hajji	Co-director of research

The doctoral committee is composed of:

Prof. Dr. Roland Billen	University of Liège, Belgium
Prof. Dr. Rafika Hajji	IAV Hassan II, Morocco
Prof. Dr. Jonard François	University of Liège, Belgium
Dr. Poux Florent	University of Liège, Belgium

The jury is composed of:

Prof. Dr. Roland Billen, supervisor	University of Liège, Belgium
Prof. Dr. Rafika Hajji, supervisor	IAV Hassan II, Morocco
Prof. Dr. Khalid Zine-Dine, Chair	Mohammed V University, Morocco
Prof. Dr. François Jonnard, Secretary	University of Liège, Belgium
Dr. Poux Florent	University of Liège, Belgium
Prof. Dr. Loubna El Mansouri	IAV Hassan II, Morocco
Prof. Dr. Thierry Badard	Laval University, Canada
Prof. Dr. Rani EL Meouche	ESTP – Paris, France

This dissertation was conducted at the Geomatics Unit of the University of Liège and Hassan II IAV from January 2020 to September 2024.

Abstract

Classified point clouds often serve as the primary data source for decision-making scenarios. For example, these data can be used as the main layer for creating Digital Twins, as a basis for urban simulation studies (such as flood simulations, vegetation inventories, rooftop solar potential, etc.), as a reference for detecting object changes, or as a foundation for automatic 3D modeling of the urban environment… The applications are numerous and potentially growing, especially when considering classified point clouds as digital assets. However, the automatic and precise extraction of maximum semantic information from an urban environment (such as parking lots, street furniture, pedestrian pathways, etc.) remains a challenge. The growing development of LiDAR technology, in terms of precision and spatial resolution, provides a good opportunity to offer reliable semantic segmentation in large-scale urban environments. Additionally, the advancement of Deep Learning techniques has revolutionized the field of computer vision and demonstrated high performance in semantic segmentation. This thesis aims to address the challenges of precisely extracting urban details from airborne LiDAR point clouds using Deep Learning techniques, in order to meet the various needs of Urban Digital Twins. Several challenges related to object extraction from airborne point clouds are explored, particularly the adaptation of Deep Learning techniques, fusion of point clouds with corresponding images, efficient feature engineering and selection, semantic segmentation, automatic 3D modeling from semantic segmentation, as well as visualization and interaction with cognitive decision-making systems. Several fusion scenarios of point clouds and images were developed and evaluated, leading to a 3D semantic segmentation fusion approach that is less data-intensive, and one that effectively extracted the maximum semantic information from the urban environment, demonstrating good results in terms of both quality and quantity. Another fusion approach was recommended due to its performance in specific semantic classes. Furthermore, a new approach was developed to exploit enriched semantic 3D point clouds as an alternative to 3D models in urban simulations. This approach was designed to meet the needs of Digital Twins. Modeling procedures were implemented for each extracted object, enabling the automatic production of 3D urban models. Finally, a case study was conducted to create the foundational elements of a Digital Twin for the city of Liège, Belgium. Several concepts, algorithms, codes, and resources are provided to reproduce the results and expand current applications.

Résumé

Les nuages de points classifiés constituent souvent le support principal pour des scénarios d'aide à la décision. Par exemple, nous pouvons utiliser ces données comme couche principale pour la création des jumeaux numériques, comme base pour les études des simulations urbaines (simulation des inondations, inventaire de la végétation, potentiel solaire des toitures, etc), comme référence pour la détection de changements d'objets, comme fondement pour la modélisation automatique 3D de l'environnement urbain... Les applications sont nombreuses et potentiellement croissantes si l'on considère les nuages de points classifiés comme des actifs de réalité numérique. Cependant, l'extraction du maximum d'informations sémantiques à partir d'un environnement urbain (parking, mobilier urbain, chemin piétonnier, etc) de manière automatique et précise reste encore un défi. En effet, le développement croissant de la technologie LiDAR en termes de précision et de résolution spatiale offre une meilleure opportunité de fournir une segmentation sémantique fiable dans les environnements urbains à grande échelle. Ainsi, le développement des techniques de Deep Learning révolutionne le domaine de la vision par ordinateur et démontre des performances élevées en matière de segmentation sémantique. La thèse tente clairement de résoudre les problèmes d'extraction précise du maximum de détails urbains à partir des nuages de points LiDAR aéroportés en utilisant les techniques de Deep Learning, et ce, pour répondre aux différents besoins des jumeaux numériques urbains. Nous abordons plusieurs problématiques liées à l'extraction des objets à partir des nuages de points aéroportés, en particulier l'adaptation des techniques de Deep Learning, la fusion des nuages de points avec les images correspondantes, l'ingénierie et la sélection efficaces des caractéristiques, la segmentation sémantique, la modélisation 3D automatique à partir de la segmentation sémantique, la visualisation et l'interaction avec les systèmes cognitifs de décision. Nous avons développé et évalué plusieurs scénarios de fusion des nuages de points et des images, et nous avons abouti à une approche de fusion de segmentation sémantique 3D moins gourmande en données, et à une approche qui a permis d'extraire le maximum d'informations sémantiques présentes dans le milieu urbain, montrant des bons résultats en termes de qualité et de quantité. Une autre approche de fusion a été recommandée en raison de ses performances dans certaines classes sémantiques spécifiques. Par ailleurs, une nouvelle approche a été développée pour exploiter les nuages de points enrichis sémantiquement comme alternative aux modèles 3D dans les simulations urbaines. Elle a été conçue pour répondre aux besoins des Digital Twins. Des procédures de modélisation ont été mises en place pour chaque objet extrait, permettant ainsi de produire automatiquement des modèles urbains en 3D. Enfin, une étude de cas a été menée pour créer les bases fondamentales de Digital Twin pour la ville de Liège, en Belgique. Plusieurs concepts, algorithmes, codes et supports sont fournis pour reproduire les résultats et étendre les applications actuelles.

Contents

Acknowledgement

My first thoughts naturally go to Professors Roland Billen and Rafika Hajji. I would like to thank them not only for their supervision but also for giving me the opportunity to prove myself and for the countless opportunities they have provided. The cover of this thesis bears my name, but it is the result of the work of a trio.

I extend my gratitude to the members of my Thesis Committee (Prof. Dr. Jonard François, Prof. Dr. Ettarid Mohamed, and Dr. Poux Florent) for their support and the influence they have had on my developments and findings. I greatly benefited from their experience in writing this document.

I am deeply thankful to the members of my Thesis Jury (Prof. Dr. Thierry Badard, Prof. Dr. Rani El Meouche, and Prof. Dr. Loubna El Mansouri) for taking the time to show interest in this work and for their participation during the examination.

I would like to express my sincere appreciation to all the members, past and present, of the "Geomatics Unit." The positive and collaborative atmosphere within our team has played an instrumental role in fostering my daily work and facilitating my professional growth.

I am also grateful to all the members of the "Laboratory of Geodesy, GNSS, and Digital Topography." Their efforts have fostered an exceptional environment for collaboration and learning.

My deepest gratitude goes to my family, especially my parents, Mohamed and Mimouna, my brother, Lhaj, and my sisters, Khadija and Zineb, for their unwavering support, unconditional love, and constant encouragement throughout the entire duration of my thesis.

I would also like to extend a heartfelt thank you to my dear friends and colleagues—Abderrazzaq, Imane, Charline, Anass, Jean-Paul, Thomas, and Gilles-Antoine, who have been unwavering pillars of support.

Finally, my thanks extend more generally to all the people I have met during my years at the University of Liège and through various research projects. Each of them has indirectly contributed to the invaluable experience that a thesis project represents.

List of Figures

List of Tables

“They did not know it was impossible, so they did it.”

- Mark Twain

Introduction

1. Context & Motivations

Digital Twins (DTs) for cities represent a new trend for city planning and management, enhancing three-dimensional modeling and simulation of cities. Indeed, many cities around the world are building their Digital Twin Cities (DTCs) [1] to respond to many urban challenges such as environmental degradation, urban planning and management city resilience and forth. Semantic 3D city models built from LiDAR point clouds are relevant inputs for building DTCs both for academic and industry research [2,3]. Semantic segmentation allows the semantic enrichment of 3D city models, their updates, and the performance of multiple spatial and thematic analyses for city management, and decision-making.

Digital Twins (DTs) for urban environments are conceptualized as secure and dynamic virtual ecosystems that replicate all facets of a city, facilitating the generation of knowledge, supporting decision-making throughout the city's lifecycle, and yielding outcomes at the municipal level [4–6]. Moreover, from a technical standpoint, a tacit agreement has emerged from the majority of research endeavors concerning the essential components of a CDT within the geospatial domain, aligning with the principles outlined in previous Smart Cities initiatives [7]. Consequently, DTs for cities are grounded in (1) 3D models of urban environments enriched with both geometrical and semantic information, (2) frequently integrating heterogeneous data, often connected with historical and real-time sensor data (synchronized at an appropriate rate), thereby facilitating (3) a reciprocal link, such as data flow between the tangible urban counterpart and its virtual twin, (4) permitting updates and analyses through a suite of simulation, prediction, and visualization tools, and (5) furnishing an encompassing perspective of diverse datasets and models throughout their lifecycle. Such an integrated approach empowers the effective management and adaptation of current and future states of cities [4,5,7,8].

DTCs are considered as digital, realistic replica of urban environments encompassing all its distinctive features. This characteristic is readily validated by considering that a point cloud inherently constitutes a precise 3D geometric representation of urban landscapes, including cities. However, delving deeper into the definition, it becomes apparent that a DT must embed both semantic and geometrical information. While the geometrical dimension of a point cloud aligns with this requirement, there is a gap in integrating semantics into DTs. With regards to this issue, several methodologies have been proposed to enhance point cloud data with semantic capabilities. These approaches encompass techniques such as 3D semantic segmentation [9], the introduction of a conceptual data model termed "Smart Point Cloud Infrastructure"[10], and data integration involving Geographic Information System (GIS) data and 3D city models [11]. Despite the development of methods to address the semantic deficit in point cloud data, it still remains a challenge. Notably, the contemporary focus in advancing Digital Twins for cities lies predominantly in data integration methodologies. This involves associating and integrating both point cloud data and semantic 3D city models, as exemplified by the innovative "PointCloud" module introduced in CityGML 3.0 [11]. This module introduces a novel concept to bridge the gap between geometrically detailed point cloud data and enriched 3D semantic models. Through this integration, sets of points are intuitively assigned to corresponding objects. CityGML 3.0's approach provides an alternative for extending point cloud data to encompass additional semantic information beyond mere

classification, achieved through diverse methods. Consequently, integrating point cloud data with various datasets from GIS, Building Information Modeling (BIM), and 3D city models emerges as a strategic solution, effectively overcoming the limitations inherent in individual approaches and aligning with the requirements of building comprehensive Digital Twins.

A 3D urban model is a major input for DTCs and a building block for its development. It consists of a geometric and semantic representation of an object or a set of urban objects (buildings, infrastructure, vegetation, etc). 3D models find application in urban planning, enabling planners to conduct 3D simulations to assess the impact of their projects. They also serve as valuable tools in mobile telephony, where engineers can determine network coverage areas using propagation models. Additionally, 3D models play a significant role in archaeology, aiding in the conservation of sites and monuments. These models are also needed in the field of civil engineering for the production of realistic scenes during the design of large construction projects, as they are used in other areas such as military strategy, natural resource management, etc. 3D models of cities are emerging as a potential solution that goes beyond the current limitations of GIS models, placing them in front of new urban management needs. The production of 3D urban models has developed remarkably in recent years. Their richness and degree of accuracy depends on the mode of acquisition of geometric data and the adopted process for semantic segmentation and modeling.

The urban model allows to integrate, organize and exchange data between different stakeholders for an efficient management of cities. The interdisciplinarity and interoperability of the used data makes this urban model a tool for collaborative design, simulation, analysis, multitemporal management and decision-making. They are capable to meet several needs related to simulation and decision-making processes. However, most of 3D city models lack rich semantics about urban knowledge and are far to respond to several challenges about smart and sustainable cities. To respond to this need, semantic segmentation has a potential contribution for the elaboration of semantically rich 3D urban models. The 3D urban model not only enables the representation and 3D visualization of urban space but also facilitates robust semantic modeling. This capability supports various spatial and thematic analyses, providing planners, urbanists, and decision-makers with an effective tool for consultation and urban planning.

Based on limitations observed in existing 3D modeling methods, we argue that actively researching robust and efficient approaches to construct geometrically and semantically rich 3D urban models is still an active research trend. Our research aims to tackle this challenge by developing novel approaches to the semantic segmentation of 3D point clouds, with the perspective of constructing urban digital twins characterized by high levels of geometric and semantic detail.

Recent advancements have been directed towards optimizing the automatic reconstruction of semantic 3D city models. These advancements predominantly involve the integration of elevation data sourced from LiDAR (airborne, terrestrial, or mobile) or photogrammetry, coupled with 2D building footprints, to produce comprehensive city models [12–15]. Notably, “3dfier” represents an automated framework specifically designed for reconstructing LoD1.2 models based on predefined rules [14]. Similarly, another relevant initiative focuses on an automated workflow that delineates roof surfaces from point cloud data, generating buildings at LoD2.1 [13]. Despite the deployment of various methodologies to create precise semantic

3D city models for diverse spatial and thematic analyses, the city modeling process persists as a laborious and time-consuming task [16,17].

3D city models (3DCM) and Digital Twins (DTs) for urban environments have garnered considerable attention within the urban and geospatial domains [18–20]. These approaches are crafted through the integration of diverse datasets and techniques, primarily involving 3D reality capture and surveying technologies [14,21,22]. Notably, the utilization of 3D point cloud data derived from laser scanning serves as a potential input for generating both 3D semantic city models and geospatial Digital Twins [22–24]. Point clouds, characterized by a straightforward and manageable structure, faithfully replicate the physical features of cities based on point geometries. Recent advancements in aerial mapping technology enable the acquisition of 3D data at a high spatial resolution, facilitated by LiDAR (Light Detection And Ranging) technology, which captures geometric and radiometric information in the form of point clouds. This data acquisition method provides precise data with a high level of detail rapidly and reliably. However, the transition from point clouds to digital models remains a challenging task, marked by a tedious, manual, time-consuming process prone to errors due to the sheer volume of data and the complexity of automation. An ongoing challenge is the automation of processes involved in constructing 3D urban models from point clouds, aiming to reduce associated costs. Additionally, the integration of semantic data obtained during the semantic segmentation of point clouds holds potential benefits for urban space management. Challenges in the acquisition and processing phases, such as irregularities and rigid transformations, need to be addressed [25]. Pre-processing and registration are crucial intermediate steps before utilizing the acquired data, ensuring its consistency. The obtained data finds plenty of applications in various domains, including urban planning [26], outdoor navigation [27], and urban environmental studies [28]. These models are considered point-based, representing entities as sets of points. However, their discrete representation and lack of structure, topology, and connectivity make them easy to handle, but at the same time, they require costly processing, particularly for semantic enrichment through knowledge-based approaches, such as Machine Learning (ML) and Deep Learning (DL) approaches [29,30]. The current surge in Artificial Intelligence (AI) is revolutionizing 3D semantic segmentation (Figure 1), yielding highly satisfactory results [31,32]. Nevertheless, the success of newly developed DL approaches relies heavily on the consistency and semantic richness of training data.

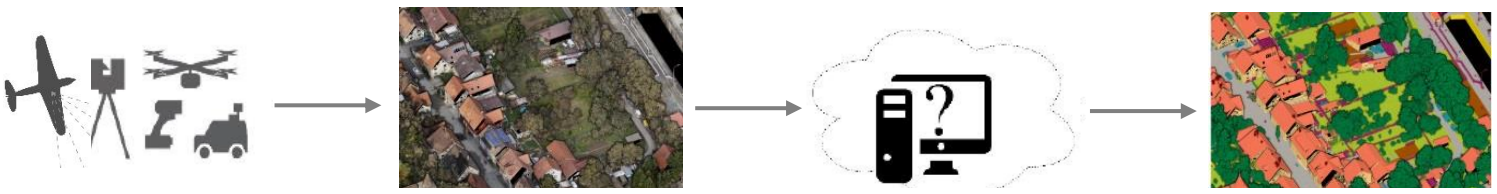


Figure 1. The sensor plays the role of our eyes, the spatial framework becomes a semantic representation, and the scene is tagged familiar using available knowledge.

The utilization of 3D LiDAR point clouds is increasingly pertinent across diverse urban applications, encompassing urban simulations, 3D visualisation through Virtual and Augmented Reality (VR and AR), Building Information Modeling (BIM), 3D urban mapping,

Smart Cities (SC), Urban Digital Twins (UDTs), and more. The swift acquisition of point clouds compared to other surveyed data enables regular updates for specific urban applications, delivering a detailed digital representation of urban settings with precise spatial information and extensive coverage, particularly when acquired through airborne sensors. Advances in LiDAR acquisition techniques have facilitated the creation of high-precision 3D point cloud representations of urban environments at a cost-effective rate. These point clouds adeptly capture objects of varying sizes, providing remarkably realistic depictions of cities and landscapes. Additionally, enhanced GPU capacity allows for the efficient rendering and instantaneous display of high-density 3D point clouds. A primary application of point clouds in urban settings is for autonomous driving, wherein recent advancements in Deep Learning (DL) techniques enable reliable navigation and decision-making through the use of dense, geo-referenced, and accurate 3D point cloud data from LiDAR. This real-time environment perception is crucial for creating high-definition maps and urban models, making it indispensable for autonomous vehicles [33]. Another significant application is 3D change detection in urban environments, facilitated by point clouds [34]. Indeed, recent developments in computer vision and machine learning enhance the automatic and intelligent detection of changes in urban settings. Moreover, point clouds find utility in virtual and augmented reality applications, offering a more immersive means of perceiving 3D digital objects [35]. Furthermore, 3D point cloud data serves as reference data for city modeling [36–39]. For instance, the 3D BAG (https://3dbag.nl/en/viewer) dataset provides multiple Levels of Detail (LoDs) of 3D buildings for the entire city of the Netherlands, generated based on building footprints from the BAG and height data from airborne laser scanning (ALS) [12,40,41]. Several cities worldwide have acquired 3D point cloud data to model their buildings, such as Helsinki, which used classified ALS point cloud data to determine elevation positions and roof shapes and created 3D building models using aerial images and airborne point cloud data to establish a CityGML model [42]. Additionally, in the creation of Urban Digital Twins for Singapore, known as "Virtual Singapore," an automatic tree modeling framework was proposed, combining airborne and mobile LiDAR scanning datasets with various remote sensing data to address the limitations of each acquisition technique. In recent years, the integration of Building Information Models (BIM) as input layers for Urban Digital Twins (UDTs) has been a subject of considerable attention in the Architecture Engineering Construction (AEC) field [43–45]. Various approaches within the AEC domain have been explored to automate and facilitate the creation of BIM models from point clouds, forming part of a scan-to-BIM workflow. The adoption of scan-to-BIM practices has yielded highly accurate data and expedited project delivery within the construction industry [46,47]. Efforts are ongoing in both industry and academia to enhance this process further by automating the segmentation of point clouds into individual building components and modeling them with continuous surfaces of solid geometries [46]. Despite notable progress, challenges persist, and existing approaches often necessitate manual modeling and reliance on proprietary software [43] . LiDAR point clouds also play a pivotal role in generating derived products, such as Digital Terrain Models (DTM), Digital Surface Models (DSM), or mesh models, which, in turn, find application in 3D city modeling and visualizations [48,49]. Significantly, advancements have been made to efficiently render massive point cloud data on the web, facilitating seamless data access [50]. Additionally, several tools are commonly employed that directly operate with point cloud data, bypassing the intricate and costly processes involved in deriving 3D city models from point clouds. The availability of an increasing volume of relevant point cloud data

is notable. However, working with point cloud data within the realms of 3D city modeling and UDTs remains challenging. Despite the enhancements introduced in CityGML 3.0, allowing the use of point cloud data to replicate city objects, the semantic information remains unaddressed even though various approaches are proposed to extend the semantic capabilities of 3D point cloud data [51].

The progression of computer vision technology has yielded more robust and reliable 3D semantic segmentation techniques. This advancement has significantly elevated the efficacy of point cloud semantic segmentation. In recent years, a multitude of approaches employing DL techniques have emerged for point clouds processing. Compared to traditional techniques, DL techniques demonstrate superior performance in terms of precision, processing speed, etc [25]. Moreover, numerous DL techniques have been developed for the semantic segmentation of LiDAR point clouds in recent times [52–54].

These DL techniques are designed to address complex tasks in various LiDAR applications, including classification, object detection, segmentation, etc. Notably, Deep Neural Networks (DNNs) have gained significant popularity and attention for their efficiency. On the other hand, it is widely recognized that learning models require an increased amount of labeled point clouds data for training. Driven by the heightened demand for training data, several datasets have been developed recently, with the majority of them being freely available online. Notable examples include Toronto-3D [55], SensatUrban [9], the Benchmark Dataset of Semantic Urban Meshes (SUM) [56], and Semantic3D [57]. The current emphasis lies in the formulation of new DL-based approaches aimed at enhancing the quality of semantic segmentation outcomes. Subsequently, it becomes imperative to conduct comparisons with existing approaches to identify the most suitable one for LiDAR point cloud processing.

While raw point clouds find widespread use, their utility is often constrained by their unstructured nature. In contrast, semantic point clouds assign a semantic label to each point (Figure 2), significantly enhancing the comprehension of scanned urban scenes and unlocking novel possibilities across various urban applications [58,59]. The pivotal role of semantic point clouds in the creation of 3D urban models, forming the foundational basis for Digital Twins (DTs), cannot be overstated. It provides a precise foundation for constructing semantic models in diverse formats such as CityGML and its encoded counterpart CityJSON, or Industry Foundation Classes (IFC) [60]. The adoption of semantic point clouds facilitates the accurate extraction of urban objects, a crucial step in the 3D modeling process of cities. Automation of object modeling, exemplified by the extraction and alignment of buildings with corresponding footprints to generate 3D models, is notably simplified with semantic point clouds [33]. Furthermore, an enriched semantic point cloud enhances 3D models by providing additional detailed information about the urban environment. The semantic richness of such point clouds proves valuable for swiftly identifying objects relevant to specific tasks or applications within the urban context. Recent advancements in 3D semantic segmentation techniques enable the extraction of comprehensive semantic information, covering elements such as vegetation, roads, railways, and more, so providing a rich 3D model that can serve for creating a Digital Twin of a city. Regular updates to the digital model are essential to faithfully mirror real-time changes in the urban environment and ensure the relevance of urban applications. Furthermore, semantic point clouds emerge as a compelling data source for

training DL techniques for semantic segmentation tasks. Leveraging semantically segmented point clouds allows for the creation of precise datasets, yielding high-performance pretrained models adaptable to various urban contexts and meeting the requirements of numerous urban applications. Additionally, semantic point clouds prove instrumental in extracting building footprints, a critical aspect of 3D building modeling. Airborne semantic point clouds, similarly, are employed to extract roofs, facilitating the creation of accurate models that align with the specific demands of urban applications. In addition, incorporating structured knowledge and semantics into 3D point clouds, beyond semantic segmentation, proves advantageous in addressing the diverse needs of urban applications [61]. On the other hand, 3D semantic segmentation plays a pivotal role in the continuous updating of DTs for cities and monitoring changes at the city scale. Specifically, 3D semantic point cloud data facilitates the real-time identification of changes occurring in the actual environment, allowing for the corresponding information to be promptly updated. Notably, point cloud data provides a comprehensive and realistic overview of the status of an urban object under construction. This is particularly beneficial in scenarios where the ongoing project lacks essential elements for generating a 3D model, such as a definitive footprint. The utility of semantic point clouds extends to urban planning and management, a common use case for DTs in cities. Enriched semantic point cloud data offers the advantage of extracting almost all urban classes, both static and dynamic objects. For specific applications, the retention of classes that are essential or require updates is prioritized, while non-crucial classes are disregarded. It is noteworthy that different use cases use various classes, aligning well with DTs requirements to capture all city objects in a single snapshot, customizing data for each specific need. Thus, semantic 3D point clouds enable accurate outlining of urban classes, improving semantic flexibility, enhancing modeling capabilities, providing new interpretability of data from diverse perspectives, and unlocking possibilities for numerous simulations and urban analyses.

Figure 2. 3D point cloud representation vs 3D semantic representation

In conclusion, the importance of airborne LiDAR point clouds in urban applications has grown considerably due to their rapid acquisition of precise spatial information in urban environments. These point clouds effectively capture the state of the city across large scales. However, improving the precision and richness of the semantics of the generated 3D models remains a crucial challenge. Certainly, semantic point clouds enhance the automatic creation of 3D city models with rich semantics and their updating. On the other hand, 3D semantic city models enriched with urban knowledge serve as a main input for City Digital Twins and enable meeting their requirements.

References

1. Shahat, E.; Hyun, C.T.; Yeom, C. City Digital Twin Potentials: A Review and Research Agenda. Sustainability 2021, 13, 3386, doi:10.3390/su13063386.

2. Ruohomäki, T.; Airaksinen, E.; Huuska, P.; Kesäniemi, O.; Martikka, M.; Suomisto, J. Smart City Platform Enabling Digital Twin. In Proceedings of the 2018 International Conference on Intelligent Systems (IS); September 2018; pp. 155–161.

3. White, G.; Zink, A.; Codecá, L.; Clarke, S. A Digital Twin Smart City for Citizen Feedback. Cities 2021, 110, 103064, doi:10.1016/j.cities.2020.103064.

4. Hristov, P.O.; Petrova-Antonova, D.; Ilieva, S.; Rizov, R. ENABLING CITY DIGITAL TWINS THROUGH URBAN LIVING LABS. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 2022, XLIII-B1-2022, 151–156, doi:10.5194/isprs-archives-XLIII-B1-2022-151-2022.

5. Nguyen, S.H.; Kolbe, T.H. PATH-TRACING SEMANTIC NETWORKS TO INTERPRET CHANGES IN SEMANTIC 3D CITY MODELS. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. 2022, X-4/W2-2022, 217–224, doi:10.5194/isprs-annals-X-4-W2-2022-217-2022.

6. Willenborg, B.; Pültz, M.; Kolbe, T. INTEGRATION OF SEMANTIC 3D CITY MODELS AND 3D MESH MODELS FOR ACCURACY IMPROVEMENTS OF SOLAR POTENTIAL ANALYSES. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2018, XLII-4/W10, 223–230, doi:10.5194/isprs-archives-XLII-4-W10-223-2018.

7. Stoter, J.E.; Arroyo Ohori, G.A.K.; Noardo, F. Digital Twins: A Comprehensive Solution or Hopeful Vision? GIM International: the worldwide magazine for geomatics 2021, 2021.

8. Würstle, P.; Padsala, R.; Santhanavanich, T.; Coors, V. VIABILITY TESTING OF GAME ENGINE USAGE FOR VISUALIZATION OF 3D GEOSPATIAL DATA WITH OGC STANDARDS. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. 2022, X-4/W2-2022, 281–288, doi:10.5194/isprs-annals-X-4-W2-2022-281-2022.

9. 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); IEEE: Nashville, TN, USA, June 2021; pp. 4975–4985.

10. Poux, F. The Smart Point Cloud: Structuring 3D Intelligent Point Data, 2019.

11. 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. Spatial Inf. Sci. 2021, VIII-4/W2-2021, 105–112, doi:10.5194/isprs-annals-VIII-4-W2-2021-105-2021.

12. Dukai, B.; Ledoux, H.; Stoter, J.E. A MULTI-HEIGHT LOD1 MODEL OF ALL BUILDINGS IN THE NETHERLANDS. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2019, IV-4-W8, 51–57, doi:10.5194/isprs-annals-IV-4-W8-51-2019.

13. Nys, G.-A.; Billen, R.; Poux, F. Automatic 3D Buildings Compact Reconstruction from LiDAR Point Clouds. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences; Copernicus, Goettingen, Germany, August 12 2020.

14. Ledoux, H.; Biljecki, F.; Dukai, B.; Kumar, K.; Peters, R.; Stoter, J.; Commandeur, T. 3dfier: Automatic Reconstruction of 3D City Models. Journal of Open Source Software 2021, 6, 2866, doi:10.21105/joss.02866.

15. Ortega, S.; Santana, J.M.; Wendel, J.; Trujillo, A.; Murshed, S.M. Generating 3D City Models from Open LiDAR Point Clouds: Advancing Towards Smart City Applications. In Open Source Geospatial Science for Urban Studies: The Value of Open Geospatial Data; Mobasheri, A., Ed.; Lecture Notes in Intelligent Transportation and Infrastructure; Springer International Publishing: Cham, 2021; pp. 97–116 ISBN 978-3-030-58232-6.

16. Girindran, R.; Boyd, D.S.; Rosser, J.; Vijayan, D.; Long, G.; Robinson, D. On the Reliable Generation of 3D City Models from Open Data. Urban Science 2020, 4, 47, doi:10.3390/urbansci4040047.

17. Naserentin, V.; Logg, A. Digital Twins for City Simulation: Automatic, Efficient, and Robust Mesh Generation for Large-Scale City Modeling and Simulation. 2022.

18. Mylonas, G.; Kalogeras, A.P.; Kalogeras, G.; Anagnostopoulos, C.; Alexakos, C.; Munoz, L. Digital Twins From Smart Manufacturing to Smart Cities: A Survey. IEEE Access 2021, PP, 1–1, doi:10.1109/ACCESS.2021.3120843.

19. Ferré-Bigorra, J.; Casals, M.; Gangolells, M. The Adoption of Urban Digital Twins. Cities 2022, 131, 103905, doi:10.1016/j.cities.2022.103905.

20. Ellul, C.; Stoter, J.; Bucher, B. LOCATION-ENABLED DIGITAL TWINS – UNDERSTANDING THE ROLE OF NMCAS IN A EUROPEAN CONTEXT. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. 2022, X-4/W2-2022, 53–60, doi:10.5194/isprs-annals-X-4-W2-2022-53-2022.

21. Deng, T.; Zhang, K.; Shen, Z.-J. (Max) A Systematic Review of a Digital Twin City: A New Pattern of Urban Governance toward Smart Cities. Journal of Management Science and Engineering 2021, 6, 125–134, doi:10.1016/j.jmse.2021.03.003.

22. Lehner, H.; Dorffner, L. Digital geoTwin Vienna: Towards a Digital Twin City as Geodata Hub. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science 2020, 88, 63.

23. Lu, Q.; Parlikad, A.K.; Woodall, P.; Xie, X.; Liang, Z.; Konstantinou, E.; Heaton, J.; Schooling, J. Developing a Dynamic Digital Twin at Building and City Levels: A Case Study of the West Cambridge Campus. Journal of Management in Engineering 2019, 36, doi:10.1061/(ASCE)ME.1943-5479.0000763.

24. Xue, F.; Lu, W.; Chen, Z.; Webster, C.J. From LiDAR Point Cloud towards Digital Twin City: Clustering City Objects Based on Gestalt Principles. ISPRS Journal of Photogrammetry and Remote Sensing 2020, 167, 418–431, doi:10.1016/j.isprsjprs.2020.07.020.

25. Zhang, J.; Zhao, X.; Chen, Z.; Lu, Z. A Review of Deep Learning-Based Semantic Segmentation for Point Cloud. IEEE Access 2019, 7, 179118–179133, doi:10.1109/ACCESS.2019.2958671.

26. Liu, C.; Zeng, D.; Akbar, A.; Wu, H.; Jia, S.; Xu, Z.; Yue, H. Context-Aware Network for Semantic Segmentation Toward Large-Scale Point Clouds in Urban Environments. IEEE Transactions on Geoscience and Remote Sensing 2022, 60, 1–15, doi:10.1109/TGRS.2022.3182776.

27. Jeong, J.; Song, H.; Park, J.; Resende, P.; Bradaï, B.; Jo, K. Fast and Lite Point Cloud Semantic Segmentation for Autonomous Driving Utilizing LiDAR Synthetic Training Data. IEEE Access 2022, 10, 78899–78909, doi:10.1109/ACCESS.2022.3184803.

28. Son, S.W.; Kim, D.W.; Sung, W.G.; Yu, J.J. Integrating UAV and TLS Approaches for Environmental Management: A Case Study of a Waste Stockpile Area. Remote Sensing 2020, 12, 1615, doi:10.3390/rs12101615.

29. Richter, R. Concepts and Techniques for Processing and Rendering of Massive 3D Point Clouds, 2018.

30. Döllner, J. Geospatial Artificial Intelligence: Potentials of Machine Learning for 3D Point Clouds and Geospatial Digital Twins. PFG 2020, 88, 15–24, doi:10.1007/s41064-020-00102-3.

31. Su, Z.; Zhou, G.; Luo, F.; Li, S.; Ma, K.-K. Semantic Segmentation of 3D Point Clouds Based on High Precision Range Search Network. Remote Sensing 2022, 14, 5649, doi:10.3390/rs14225649.

32. Wilk, Ł.; Mielczarek, D.; Ostrowski, W.; Dominik, W.; Krawczyk, J. SEMANTIC URBAN MESH SEGMENTATION BASED ON AERIAL OBLIQUE IMAGES AND POINT CLOUDS USING DEEP LEARNING. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2022, XLIII-B2-2022, 485–491, doi:10.5194/isprs-archives-XLIII-B2-2022-485-2022.

33. Li, Y.; Ma, L.; Zhong, Z.; Liu, F.; Chapman, M.A.; Cao, D.; Li, J. Deep Learning for LiDAR Point Clouds in Autonomous Driving: A Review. IEEE Transactions on Neural Networks and Learning Systems 2021, 32, 3412–3432, doi:10.1109/TNNLS.2020.3015992.

34. Kharroubi, A.; Poux, F.; Ballouch, Z.; Hajji, R.; Billen, R. Three Dimensional Change Detection Using Point Clouds: A Review. Geomatics 2022, 2, 457–485, doi:10.3390/geomatics2040025.

35. Alexiou, E.; Upenik, E.; Ebrahimi, T. Towards Subjective Quality Assessment of Point Cloud Imaging in Augmented Reality. In Proceedings of the 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP); October 2017; pp. 1–6.

36. Yan, J.; Zlatanova, S.; Aleksandrov, M.; Diakite, A.A.; Pettit, C. INTEGRATION OF 3D OBJECTS AND TERRAIN FOR 3D MODELLING SUPPORTING THE DIGITAL TWIN.

ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. 2019, IV-4/W8, 147–154, doi:10.5194/isprs-annals-IV-4-W8-147-2019.

37. Wang, Y.; Chen, Q.; Zhu, Q.; Liu, L.; Li, C.; Zheng, D. A Survey of Mobile Laser Scanning Applications and Key Techniques over Urban Areas. Remote Sensing 2019, 11, doi:10.3390/rs11131540.

38. Badenko, V.; Samsonova, V.; Volgin, D.; Lipatova, A.; Lytkin, S. Airborne LIDAR Data Processing for Smart City Modelling. In Proceedings of the Proceedings of EECE 2019; Anatolijs, B., Nikolai, V., Vitalii, S., Eds.; Springer International Publishing: Cham, 2020; pp. 245–252.

39. Huang, J.; Stoter, J.; Peters, R.; Nan, L. City3D: Large-Scale Building Reconstruction from Airborne LiDAR Point Clouds. Remote Sensing 2022, 14, doi:10.3390/rs14092254.

40. Dukai, B.; Peters, R.; Vitalis, S.; van Liempt, J.; Stoter, J. QUALITY ASSESSMENT OF A NATIONWIDE DATA SET CONTAINING AUTOMATICALLY RECONSTRUCTED 3D BUILDING MODELS. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2021, XLVI-4-W4-2021, 17–24, doi:10.5194/isprs-archives-XLVI-4-W4-2021-17-2021.

41. León-Sánchez, C.; Giannelli, D.; Agugiaro, G.; Stoter, J. TESTING THE NEW 3D BAG DATASET FOR ENERGY DEMAND ESTIMATION OF RESIDENTIAL BUILDINGS. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 2021, XLVI-4/W1-2021, 69–76, doi:10.5194/isprs-archives-XLVI-4-W1-2021-69-2021.

42. Soon, K.H.; Khoo, V.H.S. CITYGML MODELLING FOR SINGAPORE 3D NATIONAL MAPPING. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 2017, XLII-4/W7, 37–42, doi:10.5194/isprs-archives-XLII-4-W7-37-2017.

43. Deng, M.; Menassa, C.; Kamat, V. From BIM to Digital Twins: A Systematic Review of the Evolution of Intelligent Building Representations in the AEC-FM Industry. Journal of Information Technology in Construction 2021, 26, 58–83, doi:10.36680/j.itcon.2021.005.

44. Lehtola, V.V.; Koeva, M.; Elberink, S.O.; Raposo, P.; Virtanen, J.-P.; Vahdatikhaki, F.; Borsci, S. Digital Twin of a City: Review of Technology Serving City Needs. International Journal of Applied Earth Observation and Geoinformation 2022, 102915, doi:10.1016/j.jag.2022.102915.

45. Stojanovic, V.; Trapp, M.; Richter, R.; Hagedorn, B.; Döllner, J. TOWARDS THE GENERATION OF DIGITAL TWINS FOR FACILITY MANAGEMENT BASED ON 3D POINT CLOUDS.

46. Perez-Perez, Y.; Golparvar-Fard, M.; El-Rayes, K. Scan2BIM-NET: Deep Learning Method for Segmentation of Point Clouds for Scan-to-BIM. Journal of Construction Engineering and Management 2021, 147, 04021107, doi:10.1061/(ASCE)CO.1943-7862.0002132.

47. Soilán, M.; Justo, A.; Sánchez-Rodríguez, A.; Riveiro, B. 3D Point Cloud to BIM: Semi-Automated Framework to Define IFC Alignment Entities from MLS-Acquired LiDAR Data of Highway Roads. Remote Sensing 2020, 12, 2301, doi:10.3390/rs12142301.

48. Biljecki, F.; Stoter, J.; Ledoux, H.; Zlatanova, S.; Coltekin, A. Applications of 3D City Models: State of the Art Review. ISPRS International Journal of Geo-Information 2015, 4, 2842–2889, doi:10.3390/ijgi4042842.

49. Guth, P.L.; Van Niekerk, A.; Grohmann, C.H.; Muller, J.-P.; Hawker, L.; Florinsky, I.V.; Gesch, D.; Reuter, H.I.; Herrera-Cruz, V.; Riazanoff, S.; et al. Digital Elevation Models: Terminology and Definitions. Remote Sensing 2021, 13, 3581, doi:10.3390/rs13183581.

50. Oosterom, P.; Martinez Rubi, O.; Ivanova, M.; Horhammer, M.; Geringer, D.; Ravada, S.; Tijssen, T.; Kodde, M.; Goncalves, R. Massive Point Cloud Data Management: Design, Implementation and Execution of a Point Cloud Benchmark. Computers & Graphics 2015, 49, doi:10.1016/j.cag.2015.01.007.

51. Kutzner, T.; Chaturvedi, K.; Kolbe, T.H. CityGML 3.0: New Functions Open Up New Applications. PFG 2020, 88, 43–61, doi:10.1007/s41064-020-00095-z.

52. 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); IEEE: Seattle, WA, USA, June 2020; pp. 11105–11114.

53. 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; IEEE: Salt Lake City, UT, June 2018; pp. 4558–4567.

54. 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; Curran Associates, Inc., 2017; Vol. 30.

55. Tan, W.; Qin, N.; Ma, L.; Li, Y.; Du, J.; Cai, G.; Yang, K.; Li, J. Toronto-3D: A Large-Scale Mobile LiDAR Dataset for Semantic Segmentation of Urban Roadways.; 2020; pp. 202–203.

56. Gao, W.; Nan, L.; Boom, B.; Ledoux, H. SUM: A Benchmark Dataset of Semantic Urban Meshes. ISPRS Journal of Photogrammetry and Remote Sensing 2021, 179, 108–120, doi:10.1016/j.isprsjprs.2021.07.008.

57. Hackel, T.; Savinov, N.; Ladicky, L.; Wegner, J.D.; Schindler, K.; Pollefeys, M. Semantic3D.Net: A New Large-Scale Point Cloud Classification Benchmark. arXiv:1704.03847 [cs] 2017.

58. 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 Sensing 2022, 14, 3415, doi:10.3390/rs14143415.

59. 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, 2022; pp. 67–77 ISBN 978-3-030-80458-9.

60. 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. Spatial Inf. Sci. 2021, VIII-4/W2-2021, 105–112, doi:10.5194/isprs-annals-VIII-4-W2-2021-105-2021.

61. Poux, F.; Billen, R. Voxel-Based 3D Point Cloud Semantic Segmentation: Unsupervised Geometric and Relationship Featuring vs Deep Learning Methods. ISPRS International Journal of Geo-Information 2019, 8, 213, doi:10.3390/ijgi8050213.

2. Research questions

Currently, extracting 3D semantic objects from airborne LiDAR point clouds for urban applications presents several major limitations. These limitations include often insufficient precision and semantic richness, which do not always meet the demands of complex urban applications. Fusion approaches have demonstrated superior precision compared to non-fusion approaches. However, some fusion approaches are data-intensive, requiring considerable material and financial resources. These challenges highlight the need for low-cost automatic fusion approaches. Additionally, it is crucial to design approaches capable of extracting maximum urban details while improving precision and performance. Achieving this goal is essential to enhance the quality of semantic point clouds and generate 3D city models rich in semantic information. This is necessary for creating urban digital twins and meeting their needs, including updating and exploiting enriched semantic point clouds for urban simulations.

This research aims to address these challenges and answer the following key question:

"How to enhance the accuracy and richness of 3D semantic segmentation in urban environments through the fusion of airborne 3D point clouds and images using Deep Learning techniques?"

Additionally, this research aims to address the following complementary question:

"How to exploit enriched 3D semantic point clouds to build urban Digital Twins?"

3. Document outlines

The research outlines are structured around five reviewed publications (chapters 1, 2, 3, and 4), as well as complements to chapters 3 and 4. Each chapter begins with a specific preamble introducing a general introduction **outlining the general context of the thesis and research motivations (Figure 3), Chapter 1 delves into the state-of-the-art of semantic segmentation using deep learning techniques. This chapter examines existing approaches, developed algorithms, etc., and establishes initial guidelines for implementing a new fusion approach for semantic segmentation of airborne point clouds. Chapters 2 and 3 present new fusion approaches for point clouds and images for semantic segmentation that we have developed. Additionally, in Chapter 3, we propose a procedure for enriching semantic segmentation results by exploiting high-resolution spatial images and advancements in deep learning techniques for image processing. Subsequently, Chapter 4 aims to exploit the enriched 3D semantic point clouds to address the needs of city digital twins. This chapter is divided into two main sections, (A) and (B). Sub-section (A) explores the possibility of leveraging semantic segmentation results for urban simulation studies without resorting to modeling, while sub-section (B) presents a case study in which we adapted the segmentation approach developed in the previous chapters to segment an urban district in the city of Liège and integrates the obtained results into the modeling process to develop city model creation procedures. Following that, a complement is added to Chapter 4 presenting object modeling procedures that were not addressed in Chapter 4. Finally, the document concludes with Chapter 5, which presents the obtained results and perspectives for future work. The diagram of the thesis document structure is presented below:**

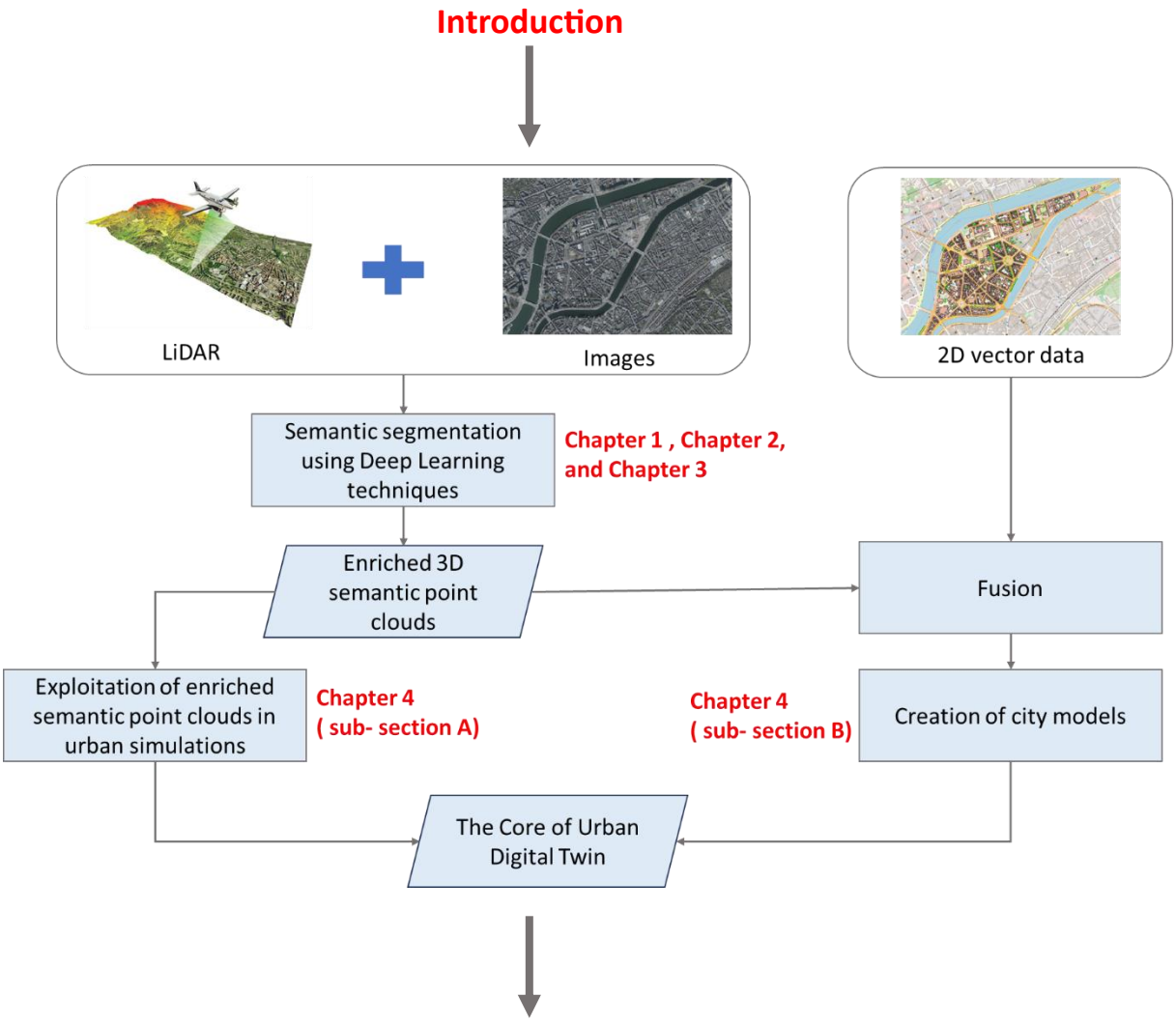


Figure 3. Thesis chapters diagram

A short description of each chapter is presented as follows:

Chapter 1: Toward a Deep Learning Approach for Automatic Semantic Segmentation of 3D Lidar Point Clouds in Urban Areas

Explores the contribution of deep learning to the semantic segmentation of large-scale 3D point clouds in urban areas. It thoroughly examines existing families of semantic segmentation approaches, assessing their performances and limitations. It also provides guidelines for an innovative fusion approach that integrates airborne LiDAR point clouds with corresponding images, employing deep learning techniques to improve the quality of semantic segmentation results. The preliminary obtained results of the proposed approach are presented.

Chapter 2: A Prior Level Fusion Approach for the Semantic Segmentation of 3D Point Clouds Using Deep Learning

Presents four significant contributions to the 3D semantic segmentation field. Firstly, it introduces a less data-intensive fusion approach for semantic segmentation, utilizing optical imagery and 3D point clouds. Secondly, the chapter presents a new airborne 3D LiDAR dataset specifically designed for semantic segmentation of airborne LiDAR point clouds purposes. Thirdly, the chapter adapts an advanced Deep Learning technique “RandLaNet” to enhance the performance of 3D semantic segmentation. Lastly, it addresses the issue of semantic class incoherence between LiDAR and image datasets during the fusion step, providing a solution to this challenge.

Chapter 3: Investigating Prior-Level Fusion approaches for Enriched Semantic Segmentation of Urban LiDAR Point Clouds

Proposes a new approach by developing and benchmarking three prior-level fusion scenarios to enhance the outcomes of point cloud enriched semantic segmentation. It compares the developed fusion approach with a baseline approach that used only the point cloud. In each scenario, specific prior knowledge (geometric features, classified images, or classified geometric information) and aerial images are fused into the neural network's learning pipeline with the point cloud data. The chapter adopts two Deep Learning techniques, "RandLaNet" and "KPConv," and optimizes their parameters for the different scenarios. Efficient feature engineering and selection during the fusion step facilitated the learning process, leading to improved enriched semantic segmentation results. The objective of this chapter was to identify the scenario that most significantly enhanced the neural network's knowledge, termed the "Efficient-PLF approach."

Furhtermore, Chapter 3 presents a practical methodology for extracting objects from high-resolution images and projecting them onto point clouds. This methodology has two objectives. The first is to extract a semantic class that does not exist in the LiDAR dataset used in a LiDAR-based approach. In other words, the developed image-based approach can

extract additional classes to LiDAR approaches. The second objective is to exploit this image-based approach for some objects where the extraction precision by the LiDAR approach is low. The developed methodology can compensate for the shortcomings of LiDAR approaches and improve their semantic enrichment.

Chapter 4: Exploiting enriched 3D semantic point clouds for urban Digital Twin requirements.

This chapter is divided into two sub sections (A and B). The first one explores the possibility of exploiting the enriched semantic point clouds in urban simulations. The aim is to meet the needs of urban Digital Twins without requiring 3D modeling, which is a costly step. The second subsection presents a processing pipeline for the automatic modeling of all urban objects extracted through semantic segmentation of 3D LiDAR point clouds. It first utilizes a deep learning-based semantic segmentation approach that integrates multiple training datasets to achieve precise extraction of target objects. Subsequently, open-source reconstruction tools are adapted for some objects, namely buildings, roads, etc., while Python codes and FME schemas have been developed for other objects, such as trees, ground, etc.

Chapter 1

Toward a Deep Learning Approach for Automatic Semantic Segmentation of 3D Lidar Point Clouds in Urban Areas

PREFACE

Advancements in deep learning techniques have revolutionized computer vision, with a growing number of tasks utilizing convolutional neural networks (CNNs), generative adversarial networks (GANs), recurrent neural networks (RNNs), and other related methods. In particular, these techniques have gained prominence in point cloud semantic segmentation due to their exceptional feature learning capabilities, which significantly enhance the accuracy of semantic segmentation results and the robustness of the trained models. On the other hand, the processing of large volumes of data requires significant computational resources and expertise. This is why the application of artificial intelligence becomes crucial, as it aids in structuring data, optimizing productivity, and automating repetitive tasks.

The primary focus of this chapter is to investigate how Deep Learning contributes to the semantic segmentation of 3D point-clouds in urban areas. Three distinct families of approaches for semantic segmentation were evaluated, namely Direct approaches, Derived Product Based Approaches, and Hybrid approaches. Furthermore, the chapter explores approaches that integrate 3D LiDAR data with other sources to enhance semantic segmentation accuracy. Nevertheless, these approaches do not accept large time differences between the acquisition of LiDAR and images data and require important storage capacity and processing time.

Additionally, the chapter thoroughly examines various deep learning techniques, such as PointNet, PointNet++, RandLA-Net, and others, applied to the semantic segmentation of airborne LiDAR data acquired in urban areas. Moreover, the chapter introduces guidelines for a novel approach that combines LiDAR data with other sources. This approach employs deep learning techniques to automatically extract maximum semantic information from point clouds. This proposed approache aims to improve the accuracy of semantic richness compared to existing approaches.

Based on Book Chapter [6]

Toward a Deep Learning Approach for Automatic Semantic Segmentation of 3D Lidar Point Clouds in Urban Areas

Abstract:

Semantic segmentation of Lidar data using Deep Learning (DL) is a fundamental step for a deep and rigorous understanding of large-scale urban areas. Indeed, the increasing development of Lidar technology in terms of accuracy and spatial resolution offers a best opportunity for delivering a reliable semantic segmentation in large-scale urban environments. Significant progress has been reported in this direction. However, the literature lacks a deep comparison of the existing methods and algorithms in terms of strengths and weakness. The aim of the present paper is therefore to propose an objective review about these methods by highlighting their strengths and limitations. We then propose a new approach based on the combination of Lidar data and other sources in conjunction with a Deep Learning technique whose objective is to automatically extract semantic information from airborne Lidar point clouds by enhancing both accuracy and semantic precision compared to the existing methods. We finally present the first results of our approach.

Keywords: Lidar, Deep learning, Semantic segmentation, Urban environment

Published in 2022

Type: Restricted Access Article

Publisher: Springer International Publishing

Book: Geospatial Intelligence: Applications and Future Trends

1. Introduction

Several challenges are facing contemporary cities such as urban sprawl, degradation, climate change, etc. Understanding these issues and predicting their impact can only be achieved through a deep and rigorous analysis of the urban environment. In this context, 3D city models are today positioned as powerful tools to address several needs about urban planning and sustainable development. Monitoring of the dynamics of cities, urban space management, construction design, and environmental studies are some appealing examples where a 3D city model is needed [1]. To respond to several city challenges, 3D city models are intended to be semantically rich to meet the requirements of urban planning and monitoring. Currently, Lidar techniques are recognized as powerful tools for producing 3D city models by offering very accurate and dense 3D point clouds at a large scale.
Semantic segmentation is an essential step to automatically design a rich 3D city model from Lidar data. It consists of assigning a semantic label for each group of point clouds (or a group of pixels in the case of images) based on homogeneous criteria [2] (Figure 4).

Figure 4. 3D semantic representation [3]

The segmentation of 3D Lidar point clouds has been widely investigated in the literature leading to several notable achievements. However, this is still an active research trend until the challenges about geometric and semantic accuracy as well as robustness and performance of the proposed methods are to be resolved.
Currently, there is a lot of interest in developing Deep Learning (DL) techniques for analyzing 3D spatial data. Thanks to their potential for processing huge amounts of data corresponding to large scale and complex urban areas with good performance in terms of accuracy and efficiency, DL methods revolutionizes the field of computer vision and are the state-of-the-art in object detection and semantic segmentation [4, 5].

According to the literature, several developments have been conducted in the field of segmentation of 3D Lidar point clouds. The developed methods can be classified into three families. The first one is based on the raw point cloud, the second one is based on a product derived from the cloud, mainly Digital Surface Model (DSM), while the third one combines original point clouds and other data sources (aerial image, land map, etc.) [1].

Several research teams have stated that the combination of Lidar data with other sources (aerial image, satellite image, etc.) is promising, thanks to the altimeter accuracy of the 3D point clouds and the planimetric continuity of the images [6]. This motivated us to conduct our research in this field where we propose to design a methodology based on the integration of Lidar data and other sources with the aim to enhance the quality of the semantic segmentation results for urban scenes.

In the next sections, we propose to give a global overview about the main developments in semantic segmentation by highlighting the strengths and the weakness of the developed approaches. Section 2 gives an overview about the main developed methods for automatic segmentation of Lidar point clouds. Then, Sect. 3 presents some DL approaches for semantic segmentation. The discussion of the main outcomes is the subject of Sect. 4. While Sect. 5 proposes the basic guidelines as well as the preliminary results of a new approach based on our investigations and the outcomes of the literature review. Finally, the paper ends with a conclusion.

2. Automatic Segmentation of 3D Point Clouds

Point cloud segmentation is an essential step for various applications. Besides clarifying the spatial relationships between point clouds and facilitating pattern recognition, the segmentation improves the quality of subsequent classifications. This process partitions a cloud of points into a set of segments characterized by spatial and/or geometric coherence. The definition of this coherence forms the critical part of the segmentation process [7]. Numerous segmentation approaches have been developed and applied to 3D Lidar data. In this section, we mainly focus on the general research methods that are widely used for the segmentation of 3D point clouds. Three families of approaches exist to perform a semantic segmentation of Lidar point clouds. The first one is based on the raw point cloud (Direct approaches). The second one is based on a product derived from the cloud (Derived Product Based Approaches). While the third one combines original point clouds and other data sources (aerial image, land map, etc.) (Hybrid approaches) [6].

2.1 Direct Approaches

Direct approaches are applied to 3D raw point clouds without any sampling method. Among the benefits of this family of approaches, we can cite the preservation of the original characteristics of data, including accuracy and topographical relationships. On the other hand, we raise some shortcomings and gaps that hinder the effectiveness and the relevance

of this family of approaches, mainly the need of a too high computing time and a rather large memory. In the literature, many studies have been based on direct approaches. Among the developed methods, Lee [8] has proposed a segmentation process based on 3D surface detection, specifically by using Lidar raw data directly without any prior interpolation. This method allows automatic division of the point cloud into two classes: ground and buildings, considered as the main objects of an urban scene. The method of [9] proposes a cluster analysis of 3D airborne Lidar data by using a slope adaptive neighborhood system based on accuracy, point density, and distance between 3D point clouds in order to define the neighborhood between the measured points. According to proximity and local continuity, points that are on the same surface are connected [9]. The method gives good results in extracting vertical walls and modeling objects with a precision of few centimeters. Lari [10] proposed a method for segmentation of planar patches using Lidar data. In this approach, the authors used an adaptive cylinder for establishing the neighborhood of each point by considering surface trend and density. This definition of neighborhood positively influences the calculation of segmentation attributes (vegetation, flat and gable roofs, walls …etc.). The approach demonstrates efficiency and reliability for both airborne Lidar and Mobile Mapping Systems data. Finally, a segmentation method applied to a mobile and airborne mapping system has been proposed by [11] where the main objective is to bypass the drawbacks of point-based classification techniques; its principle is based on grouping point clouds in regions with similar characteristics. The proposed methodology demonstrates a high potential in classification of both terrestrial and airborne Lidar data.

2.2 Derived Product Based Approaches

Since direct approaches require a very high processing time and large storage capacity, many researchers recommend the transformation of 3D data into 2D in order to have a regular form that is easy to manipulate. This is the principle of derived based approaches which are based on derived products from Lidar data such as DSM (Digital Surface Model) and 2D images. This family of approaches offers a wide range of advantages such as the ease of handling and the efficiency of data processing. However, these approaches require a 2D transformation of 3D data or voxels representations which result in a huge loss of geometric and radiometric information, and thus a loss of precision due to the resampling operation.

In the literature, there is a large number of approaches that have been developed for the segmentation of 3D point clouds from the regular data generated from the point clouds. Among these approaches, Yuan [12] proposed a new tech- nique called “Pointseg” that allows a real time semantic segmentation of road objects based on spherical images where the structure of the proposed network is based on SqueezeNet [13] and SqueezeSeg [14]. The proposed network has three main functional layers: (1) fire layer, (2) squeeze reweighting layer, and (3) enlargement layer. The results show compatibility with robot applications by achieving competitive accuracy with 90 frames per second on a single GPU (Graphics Processing Unit) and high efficiency when tested with KITTI 3D objet detection dataset. Milioto [15] proposed a semantic segmentation approach called RangeNet++. This approach has been applied to Lidar data recorded by a rotating Lidar sensor in order to enable the autonomous vehicles to make the best decisions in a timely manner. The authors

proposed a projection based 2D CNN (Convolutional Neural Network) processing of Lidar data and used a range image representation of each laser scan to perform the semantic inference. The results show that this method outperforms the state of the art both in runtime and accuracy. Moreover, a new approach for semantic labeling of unstructured 3D point clouds has been proposed in [16]. The authors proposed a framework that applies CNN on multiple 2D image views of the Lidar data based on two steps: (1) generation of two types of images: depth composite view and RGB view and (2) labeling each pair of bidimensional image views by means of CNNs. After that, they project back the semantized images. This approach showed good results when evaluated using a dataset called Semantic-8. Another method for Lidar data segmentation using voxel structure and graph-based clustering was pro- posed by [17]. The authors used a geometric method that not require any radiometric information. The process consists of three steps: (1) voxelisation of 3D point clouds, (2) calculation of geometric cues, and (3) the graph-based clustering. The method has demonstrated good results mainly for complex environment and non-planar areas, compared to several segmentation methods proposed in the literature. Riegler and Osman Ulusoy [18] proposed a method called “OctNet” as a novel tridimensional representation for point clouds labeling, which enables 3D CNN that are both high resolution and deep. The method was evaluated using Rue- Mong2014 dataset [19] and achieved good results. Finally, another work has been proposed by [20] where the authors evaluated various bidimensional image models using four datasets which are DUT1, NC, DUT2, and KAIST. The results, compared to those of direct approaches, show that the use of bidimensional image models give an interesting improvement in computational efficiency with a little loss of precision. Furthermore, the authors concluded that 2D image models are better suited to real-time segmentation of outdoor areas.

2.3 Hybrid Approaches

Despite the simplicity and the efficiency of Derived Product Based Approaches, several researchers argued that Lidar data need to be combined with other data sources (aerial photos, satellite images, etc.) to take benefits from the planimetric continuity of images and the altimetric precision of 3D point clouds [6]. Several investigations in this field have shown promising results in terms of accuracy and quality of the segmentation. However, despite their performance, these approaches have many disadvantages related to memory requirements, difficulty of handling and implementation, and the need to have a minimum difference in the time of acquisition of the two types of data.

The first method has been proposed by [21] for automatic building detection from 3D point clouds and multispectral imagery. This method is capable of detecting different urban objects (industrial buildings, urban residential, etc.) of different shapes with very high precision. The authors of [22] applied a multi-filter CNN for semantic segmentation based on the combination of 3D point clouds and high-resolution optical images, and then they used a MRS (Multi-Resolution segmentation) for delimiting the contours of objects. The results show that this approach improves the overall accuracy over other methods using Potsdam and Guangzhou datasets and is more suitable for the processing of objects with a regular shape such as cars and buildings. Furthermore, Xiu [23] proposed a new method to study the influ-

ence of integrating two types of data which are aerial images and 3D point clouds for semantic segmentation which shows an accuracy of 88%. Additionally, a new semantic segmentation study combining images and 3D point clouds has been proposed by [24] by adopting the DVLSHR (Deeplab-Vgg16 based Large-Scale and High-Resolution) model which is satisfactory for semantic segmentation of large-scale scenes when compared to other methods developed in the literature using CityScapes dataset. Another approach called SPLATNet was proposed in [25]. This approach has been tested with RueMonge2014 dataset [19] where an Intersection Over Union score was computed for all classes in order to evaluate the semantic segmentation results. The proposed approach scores well among the state-of-the-art algorithms for semantic segmentation. Recently, [26] proposed a new methodology for semantic segmentation which grasps bidimensional textural appear- ance and tridimensional structural characteristics in an integrated framework. The authors evaluated this approach using ScanNet Dataset [27]. The method has demonstrated good results compared to 3DMV (3D-Multi-View) and SplatNet (Sparse lattice Networks) approaches. Similarly, Li [28] designed a 3D real-time semantic map using 3D point clouds and images of road scenes. The method consists of using a CNN to segment 2D images acquired by a camera, and then the semantic segmentation results and the 3D point clouds are fused to generate a unified point cloud with an associated semantic information. The proposed technique is effective for several complex tasks including autonomous driving, robot navigation, etc.

2.4 Summary

3D Lidar data segmentation methods can be grouped into: Direct approaches, Derived Product Based Approaches, and Hybrid approaches. The direct approaches are the least used in the literature because they require a very large storage capacity and are very demanding in processing and computing time. Despite their limitations, their strengths lie in the preservation of the characteristics and the original topological relationships of the point cloud. Derived Product Based Approaches are the most dominant, simplest, and quickest approaches in the literature. However, the resampling operation applied to the point cloud causes a huge loss of information and so a loss of precision of the segmentation process. Finally, approaches combining 3D Lidar data and other sources allow improving the accuracy of the segmentation. However, these approaches do not accept large time differences between the acquisition of Lidar and images and require a very high storage capacity and a very important processing time (Table 1).

Actually, the development of DL methods offers a best opportunity to satisfy the need of computer vision field and demonstrates a high potential in semantic segmentation in terms of accuracy and efficiency. Their performance in segmentation process would enhance the quality of the results. The next section tries to give a brief overview of researches addressing DL in semantic segmentation.

3. Contribution of DL to Semantic Segmentation

Actually, DL methods revolutionize the field of computer vision and demonstrate good performance in semantic segmentation by solving a wide range of difficult problems in this field [29]. In this section, we examine some DL techniques used in semantic segmentation of Lidar data acquired in urban areas.

PointNet is a reference network which opened the way for the use of DL techniques for semantic segmentation of Lidar data [30]. Its performance, combined with its ease of implementation, makes it a perfect baseline for semantic segmentation of 3D point clouds. The core principle of PointNet is to implement the permutation invariance of the points in a cloud directly into the network. To evaluate its performance, the authors used the Stanford 3D dataset where data are annotated with 13 classes (floor, chair, table, etc.). PointNet has demonstrated satisfactory results compared to the literature. Similarly, Qi [31] proposed a hierarchical DL model called “PointNet++” in order to process a set of points that have been sampled in metric space in a hierarchical manner.

Table 1. Advantages and disadvantages of the different segmentation approaches.

Approach	Advantages	Disadvantages
Direct approaches	– Preserve the original topological relationships of point cloud	– Expensive – Few developed programs
Derived product based approaches	– Easy and fast drive Requires few parameters	– Loss of information and accuracy due to re-sampling – False data caused by resampling step Errors accumulation
Hybrid approaches	– Accurate – Efficient	– Expensive Require a minimum difference in time of acquisition of the two types of data

To test this approach, four datasets have been used, namely, ModelNet40, MNIST, SHREC15, and ScanNet. The results show that the proposed approach is more suitable to process point sets robustly and efficiently compared to other existing methods. Besides, this methodology introduced hierarchical feature learning and captures spatial features at different scales which is important in case of objects of different sizes. Another semantic segmentation approach named SegCloud was proposed in [32]. The proposed approach combines the advantages of trilinear interpolation, neural networks, and FC-CRF (Fully Connected Conditional Random Fields). The authors used the trilinear interpolation to transform voxels predictions to raw 3D points, then the FC-CRF allows overall consistency, and fine semantic segmentation. The authors evaluated the perfor- mance of the proposed algorithm using four multi-scale datasets about indoor or outdoor scenes (NYU V2, S3DIS, KITTI, and Semantic3D). The results show that CRF allows a significant improvement of the network and a high ability to extract the contours of objects in a very clear way. Moreover, a novel fully CNN approach for semantic seg- mentation of images named SegNet has been developed by [33]. It consists of an encoder-decoder structure based on the convolution layers of the VGG-16 algorithm. The architecture of SegNet is symmetrical and allows precise positioning of abstract features with good spatial locations. CamVid dataset has been used to evaluate the performance of the proposed method. This dataset is divided into two sets: the first contains 367 images used for training the model while the second contains 233 images used for performance evaluation. The results show that this algorithm gives good results and achieves very high scores in the case of semantic segmentation of road environments. Furthermore, Landrieu and Simonovsky [34] proposed a new Lidar approach applicable for large 3D Lidar data where the main objective is to divide the point clouds into simple forms. The process is based on three main steps: (1) a new concept called a superpoint graph to encode the relationships between object parts by edge attributes is proposed, (2) a neural network is used for the representation of each simple shape, and (3) two public datasets (S3DIS and Semantic3D) are used to improve the average of mIOU (mean Intersection Over Union). In addition, Qi [35] proposed a 3D object detection approach based on collaboration between Haugh Voting and point set network called

VoteNet. It is a geometric method that does not require any radiometric information but shows clear improvements over hybrid methods. Additionally, Yang [36] proposed a new large-scale urban semantic segmentation framework by integrating multiple aggregation levels (point-segment-object) of features and contextual features for road facilities recognition from 3D Lidar data. This study achieved very satisfactory results with an object recognition accuracy of more than 90%. Finally, Hu [37] developed a new neural network architecture called “RandLA-Net” that directly uses 3D Lidar data based on point sampling in a random manner. In order to reduce the point density, to avoid loss of information caused by the resampling step, the authors proposed a new local feature aggregation module. Compared to the literature, the proposed approach demonstrates a good performance in terms of precision, calculation time and is not demanding a fairly large memory.

4. Discussion

Today, 3D city models allow better understanding of urban spaces which is crucial for optimal management of cities. They are capable to meet several needs related to simulation and decision making processes. However, most of 3D city models lack rich semantics about urban knowledge and are far to respond to several challenges about smart and sustainable cities. In computer vision, semantic segmentation is defined as the assignment of a class to each coherent region of an image [2] or 3D point clouds. Many recent studies have shown the effectiveness of DL in this context [30, 34–37]. The first experiments of approaches dedicated to semantic segmentation of 3D point clouds began by the use of conventional image processing programs by transforming the 3D Lidar data into regular shapes (for example, series of images) as in the case of the approach proposed by [16] that requires a transformation of 3D point clouds to 2D images. Other DL techniques are based on the transformation of the Lidar data into a grid of voxels that have a regular form as the case of the SegCloud method that was proposed by [32]. These regular representations do not really allow a clear writing of the particular organization of Lidar data which limits the performance of this type of approach [34]. Besides, the voxel representation does not take into account the small details of 3D forms. Several research teams have proposed a range of dedicated approaches directly analyzing Lidar data. Among these approaches, the PointNet approach, proposed by [30], operates at the point level, which allows a very fine segmentation. This method is adapted to 3D point clouds acquired in indoor scenes, but it requires a necessary adaptation or additional training to be adapted to large datasets [32]. Similarly, the PointNet++ method is applied to the raw point clouds [25] without any sampling operation, which saves the initial information [35]. This method has demonstrated better performance in semantic segmentation and object classification [35]. However, it shows some limitations, namely, large computation and memory cost [38, 39]. Furthermore, this approach is not able to aggregate the scene context around the object centers due to more clutter and inclusion of neighboring elements [35], and also lacks a relevant specification of the spatial connectivity between points [25]. We note that “PointNet” and “PointNet++” have not been tested on data acquired by a large scale airborne mapping system that contains more complicated urban geographic features [23]. Recently, several approaches have been developed for processing of large-scale 3D point clouds. In this context, we find the SPG method that allows the preprocessing

of 3D Lidar data as super-graphs in order to subsequently apply a neural network to assign a semantic label for each group of points [15]. The main advantage of this approach is its ability to handle large point clouds simultaneously by cutting point clouds into simple shapes that are easier to classify than points, but despite the low number of network parameters, this approach is high demanding in terms of time of processing required by super-graph construction and geometric partitioning [15]. We can state that most of the existing semantic segmentation approaches require a variety of blocks partitioning steps, pre/post–processing as well as the construction of graphs. In contrary, the “RandLA-Net” approach is able to directly process large scale 3D Lidar data in a single pass with high efficiency (1 million points in a single pass) without any pre-processing or post-processing steps compared to the existing methods [39]. Finally, semantic segmentation is an active research trend which aims to reach robust methods to extract semantics from dense point clouds or images. The construction of these models from Lidar data requires designing new approaches capable of extracting the maximum amount of semantic information about a large-scale urban environment with high accuracy and efficiency. Our research tries to respond to this challenge by proposing an innovative hybrid approach which aims to enhance the quality of semantic segmentation of airborne Lidar point clouds.

5. Our Approach

The literature review about DL techniques that address semantic segmentation of Lidar point clouds shows that this is clearly a field that requires further research in order to improve the accuracy and the performance of the segmentation process. This has motivated us to conduct research in this field in order to propose an innovative approach for semantic segmentation of airborne Lidar data based on a hybrid solution. In this section, we expose the first guidelines and preliminary results of our proposed research in this context.

5.1 Methodology

We propose to design a DL approach based on the combination of 3D airborne Lidar data and aerial images for semantic segmentation of airborne Lidar point clouds corresponding to large-scale urban environments. Our methodology is expected to give better results in terms of precision and robustness to recognize 3D objects of urban scenes and associate them a rich semantic.

Figure 5 summarizes the general workflow of our approach. Our approach relies on the combination of the geometry of Lidar data and the spectral information of images. It is based on the use of raw data in order to retain the original characteristics and topological relationships of 3D point clouds. The first step consists of applying semantic segmentation to drone images which results will be integrated with Lidar data in order to refine the quality of the segmentation process (part 2). The test of the performance and the reliability of the proposed approach will be performed through several large-scale datasets. In the next section, we present and analyze the preliminary results related to the first step of the wokflow (Part1).

5.2 Preliminary Segmentation

Semantic segmentation from drone images is a first step of the general workflow. The results will be then integrated with Lidar point clouds to enhance the segmentation process. High spatial–resolution of data acquired by drones makes it possible to discriminate the different urban objects and associate them a semantic label. In this context, several DL Techniques applied to drone images have been proposed in the literature [40–43]. To our knowledge, there is no literature review about the evaluation of the existing techniques. This is why we had to conduct several tests to evaluate different models (Unet, Vgg_Unet, Resnet50_Unet, Segnet, Vgg_Segnet, and Resnet50_Segnet) in terms of precision and calculation time in order to choose the most suitable one for semantic segmentation of drone images.

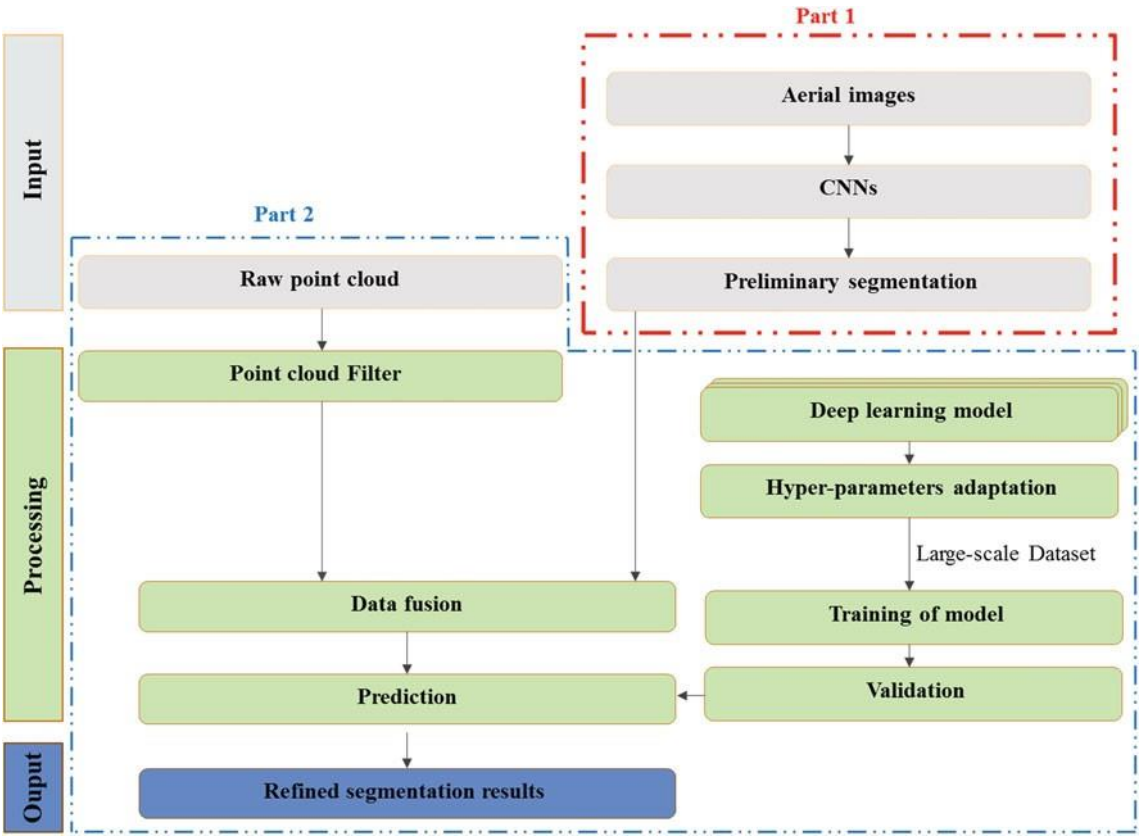


Figure 5. The general workflow of our approach

5.2.1 Data

The case study consists of 400 large-scale drone images with a high resolution of 6000 * 4000 px (24Mpx) and an altitude of 5–30 m above ground which are available for free download (https://dronedataset.icg.tugraz.at). The images are annotated with 20 classes: tree, gras, other vegetation, dirt, gravel, rocks, water, paved area, pool, person, dog, car, bicycle, roof, wall, fence, fence-pole, window, door, and obstacle. Some examples of the dataset are shown in Figure 6. Another data is used for the evaluation of the process. It is relative to an urban zone of the city of Nador (Morocco), where the images was acquired with a ground resolution at 100 m flight height of 3.5 cm and resolution of 12 MegaPixel.

5.2.2 Results

For the implementation of the DL models used in this study, we used the Keras library and Google Colaboratory as a cloud computing server. Google Colaboratory is a free Google tool that allows performing computational simulations with support of Python and some other libraries. For conducting the tests, 80% of the dataset is used for training the model while 20% serves as testing data. In this section, we present the results about the evaluation of both accuracy and time of calculation of the segmentation process applied to the selected models.

Accuracy assessment

The semantic segmentation realized according the tested models is evaluated through two parameters: (1) accuracy and (2) frequency weighted IU (f.w.IU). Accuracy metric is the ratio of the number of correct predictions to the total number of input samples. While the frequency weighted IU defines the variations on region intersection over union (IU) used in target detection [44]. These metrics are obtained using the equations below:

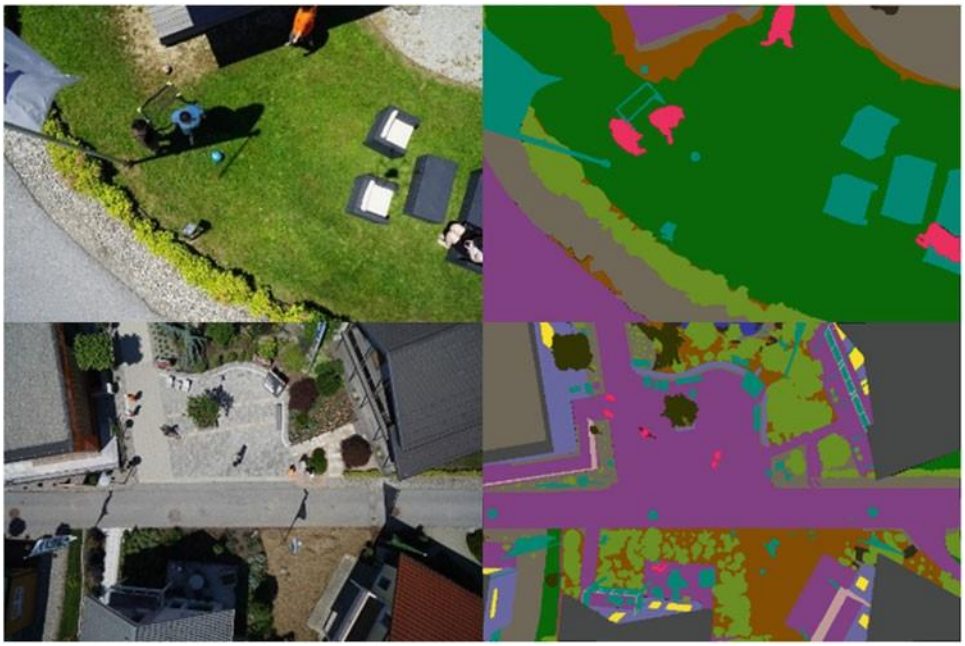


Figure 6. Examples of classified Drone images from the dataset

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions\ made}$$

$$\text{f. w. IU} = \Sigma i \left(\Sigma j U ij \cdot \Sigma k \Sigma j U ij\right) \frac{U ii}{\Sigma j U ij + \Sigma j U ji - U ii} [44]$$

where k represents the number of classes. The symbol u_{ij} corresponds to the number of

samples belonging to category i in ground truth and are classified in class j in segmentation results [44].
The evaluation results according accuracy and frequency weighted IU are reported in Table 2 and Table 3 respectively.
Even though we conducted the tests with a limited number of epochs, we reached good results for both accuracy and frequency weighted IU with the different models. According the results, we can say that the “Resnet50_Unet” model outperforms the others both in accuracy and fre quency weighted IU metrics.

Training duration

Besides the accuracy of the segmentation, we also evaluated the efficiency of the tested models in terms of processing time. The results are reported in Table 4.
According to the statistics in Table 4, we can state that the processing time is relatively negligible and all models require almost the same computation time with a bit difference of the Vgg-Unet model which requires slightly more time than the other ones.

The preliminary tests were necessary to test the performance of the selected models. According the results, the “Resnet50_Unet” has been elected as the most suitable model for semantic segmentation of drone images to be adopted in our approach. This model has been applied to the case study about the urban area in Morocco. The corresponding semantic segmentation results are shown in Figure 7 and validated by comparison to the field reality.

6. Summary

In the previous section, we presented the first results of the general workflow of our approach. It consists of semantic segmentation of drone images as a first step of the process. The general objective is to integrate the preliminary results of the image segmentation process with Lidar data in order to enhance the quality of the segmentation in terms of accuracy and performance. We performed a series of experiments to compare the capabilities of the different DL techniques for semantic segmentation of urban objects using Drone images.

Table 2. Comparison of accuracy between the DL models.

	Unet	Vgg_Unet	Resnet50_Unet	Segnet	Vgg_Segnet	Resnet50_Segnet
Accuracy	0.71	0.76	0.85	0.72	0.7215	0.82

Table 3. Comparison of frequency-weighted IU between the DL models.

	Unet	Vgg_unet	Resnet50 _Unet	Segnet	Vgg_seg net	Resnet50_Se gnet
Frequency_ Weighted_IU	0.56	0.63	0.76	0.58	0.56	0.72

Table 4. The required time for the segmentation process.

	Unet	Vgg-Unet	Resnet50-Unet	Segnet	Vgg-Segnet	Resnet50_Segnet
Epochs	1310 s	1403 s	1225 s	1217 s	1208 s	1281 s
Epoch 1	1269 s	1385 s	1202 s	1198 s	1160 s	1222 s
Epoch 2	1243 s	1366 s	1219 s	1178 s	1159 s	1175 s
Epoch 3	1248 s	1319 s	1229 s	1154 s	1161 s	1171 s
Epoch 4	1205 s	1287 s	1209 s	1152 s	1163 s	1172 s
Total time(s)	6275	6760	6084	5899	5851	6021
Total time(m)	105	113	101	98	97	100

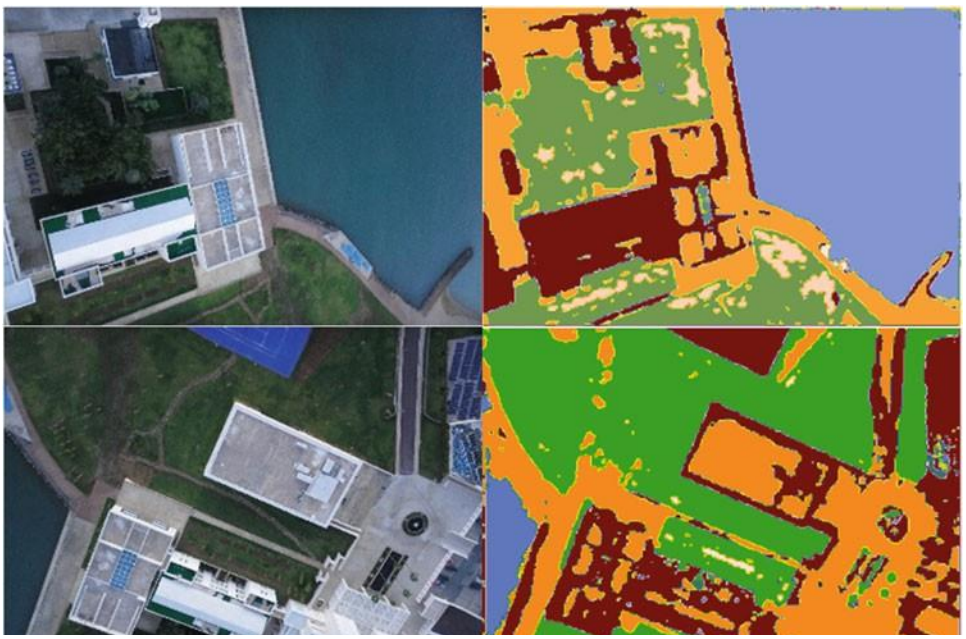


Figure 7. Examples of semantic segmentation results

The results show that that all tested models give good results in terms of accuracy and frequency weighted IU. However, the Resnet50_Unet model scores well in both parameters. Hence, it has been selected as the most suitable one for semantic segmentation of drone images among the others. We should note that the quality of the results can be further improved by using a powerful dataset with more training data and by augmenting the number of epochs. Finally, for a better evaluation of the performance of different DL models, we propose to use other types of datasets, as well as to apply the models to other images acquired in other different urban contexts.

7. Conclusion

In this paper, we have proposed a literature review about semantic segmentation methods of 3D Lidar point clouds based on DL. Several DL models have been presented and analyzed by highlighting their advantages and their limitations. We then presented the first guidelines about our proposed methodology which aims at developing a DL approach based on integrating 3D Lidar point clouds and aerial images for semantic segmentation in a large-scale urban environment. We aim to improve the object recognition accuracy and the efficiency of the existing methods.

As a first step of our approach, we investigated the performance of some DL models in terms of accuracy and performance for semantic segmentation of drone images by conducting several tests. In the next steps, our method will be tested on several datasets to confirm the reliability and the performance of the proposed approach.

References

1. A. Bellakaout, Extraction automatique des batiments, végétation et voirie à partir des données Lidar 3D. Thèse de docteur de l'institut agronomique et vétérinaire Hassan II, Maroc (2016)
2. L. Haifeng, Unsupervised scene adaptation for semantic segmen- tation of urban mobile laser scanning point clouds. ISPRS J. Photogramm. Remote. Sens. 169, 253–267 (2020)
3. B. Kim, Highway driving dataset for semantic video segmentation. School of Electrical Engineering Korea Advanced Institute of Science and Technology (KAIST), South Korea (2016)
4. J. Castillo-Navarro, Réseaux de neurones semi-supervisés pour la segmentation sémantique en télédétection. Colloque GRETSI sur le Traitement du Signal et des Images, Lille, France. hal-02343961 (2019)
5. A. Garcia-Garcia, A review on deep learning techniques applied to semantic segmentation. arXiv:1704.06857v1 [cs.CV] (2017)
6. M. Awrangjeb, Automatic detection of residential buildings using LIDAR data and multispectral imagery. ISPRS J. Photogram. Remote Sens. 65, 457–467 (2010)
7. J. Ravaglia, Segmentation de nuages de points par octrees et analyse en composantes principales. GTMG 2014, Mar 2014, Lyon, France. hal-01376473 (2014)
8. I. Lee, Perceptual organization of 3D surface points, photogram- metric computer vision. ISPRS Comm. III. Graz, Austria. XXXIV part 3A/B. ISSN 1682-1750 (2002)
9. S. Filin, Segmentation of airborne laser scanning data using a slope adaptive neighborhood. ISPRS J. Photogramm. Remote Sens. 60, 71–80 (2006). https://doi.org/10.1016/j.isprsjprs.2005.10.005 (2005)
10. Z. Lari, An adaptive approach for segmentation of 3D laser point cloud, in ISPRS Workshop Laser Scanning, Calgary, Canada (2011)
11. Z. Lari, A. Habib, Segmentation-based classification of laser scanning data, in ASPRS 2012 Annual Conference Sacramento, California, 19–23 Mar 2012
12. W. Yuan, PointSeg: real-time semantic segmentation based on 3D LiDAR point cloud. arXiv:1807.06288v8 [cs.CV] (2018)
13. F.N. Iandola, Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size. CoRR abs/1602.07360 (2016)
14. B. Wu, Squeezeseg: convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3d lidar point cloud. CoRR abs/1710.07368 (2017)
15. A. Milioto, RangeNet++: fast and accurate LiDAR semantic segmentation. German Research Foundation under Germany's Excellence Strategy, EXC-2070 - 390732324 (PhenoRob) as well as grant number BE 5996/1–1, and by NVIDIA Corporation (2019)
16. A. Boulch, SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks. Comput. Graph. (2017)
17. Y. Xu, Voxel- and graph-based point cloud segmentation of 3d scenes using perceptual grouping laws. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. IV-1/W1 (2017)

18. G. Riegler, A. Osman Ulusoy, Octnet: learning deep 3d represen- tations at high resolutions, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 3 (2017)
19. H. Riemenschneider, A. Bódis-Szomorú, Learning where to classify in multi-view semantic segmentation, in Proceedings of the European Conference on Computer Vision (ECCV) (2014)
20. Y. Liu, Comparison of 2D image models in segmentation performance for 3D laser point clouds. Neurocomputing (2017)
21. M. Awrangjeb, Automatic detection of residential buildings using LIDAR data and multispectral imagery. ISPRS J. Photogramm. Remote Sens. 65, 457–467 (2010)
22. Y. Sun, Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and LiDAR data. ISPRS J. Photogramm. Remote Sens. (2018)
23. H. Xiu, 3D semantic segmentation for high-resolution aerial survey derived point clouds using deep learning (Demonstration), in Information Systems (SIGSPATIAL'18), 6–9 Nov 2018, Seattle, WA, USA, ed. by F. Banaei-Kashani, E. Hoel (ACM, New York, NY, USA, 2018)
24. R. Zhanga, 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)
25. H. Su, V. Jampani, Splatnet: sparse lattice networks for point cloud processing, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018), pp. 2530–2539
26. H.-Y. Chiang, A unified point-based framework for 3D segmentation, in International Conference on 3D Vision (3DV) (2019)
27. A. Dai, Scannet: Richly-annotated 3d reconstructions of indoor scenes, in Proceedings of CVPR 2017 (2017)
28. J. Li, Building and optimization of 3D semantic map based on Lidar and camera fusion. Neurocomputing
29. Y. Li, Deep learning for remote sensing image classification: a survey. Wiley Interdisciplinary Reviews. Data Mining and Knowl- edge Discovery, vol. 8, p. 1264 (2018)
30. C.R. Qi, Pointnet: deep learning on point sets for 3d classification and segmentation. CoRR abs/1612.00593 (2016)
31. C.R. Qi, PointNet++: deep hierarchical feature learning on point sets in a metric space. arXiv:1706.02413v1 [cs.CV] (2017)
32. L.P. Tchapmi, Segcloud: semantic segmentation of 3d point clouds, in International Conference on 3D Vision (3DV) (2017), pp. 537–547
33. B. Vijay, SegNet: a deep convolutional encoder- decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 2481–2495 (2017)
34. L. Landrieu, M. Simonovsky, Large-scale point cloud semantic

segmentation with superpoint graphs, in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018), pp. 4558–4567

35. C.R. Qi, Deep Hough voting for 3D object detection in point clouds. arXiv:1904.09664v2 [cs.CV] (2019)
36. B. Yang, Computing multiple aggregation levels and contextual features for road facilities recognition using mobile laser scanning data. ISPRS J. Photogramm. Remote Sens. 126, 180–194 (2017)
37. Q. Hu, RandLA-Net: efficient semantic segmentation of large-scale point clouds. arXiv:1911.11236v3 [cs.CV] (2020)
38. Z. Yang, Std: sparse-to-dense 3d object detector for point cloud, in The IEEE International Conference on Computer Vision (ICCV) (2019)
39. Y. Cui, Deep learning for image and point cloud fusion in autonomous driving: a review. arXiv:2004.05224v2 [cs.CV] (2020)
40. A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv print. 14 p (2014)
41. O. Ronneberger, P. Fischer, U-Net: convolutional networks for biomedical biomedical image segmentation, in International Conference on Medical Image Computing and Computer-Assisted Intervention (2015), pp. 234–241
42. K. He, Deep residual learning for image recognition, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016), pp. 770–778
43. A. Chaurasia, Linknet: exploiting encoder representations for efficient semantic segmentation. arXiv preprint 1707.03718 (2017)
44. M.H. Wu, ECNet: efficient convolutional networks for side scan sonar image segmentation. Sensors 19(9), 2019 (2009). https://doi. org/10.3390/s19092009

Chapter 2

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

PREFACE

Following the previous chapter, which examined different families of semantic segmentation approaches and opted for fusion approaches due to their superior accuracy compared to other approaches. And also, the previous approach highlighted that fusion approaches can require diverse data sources, making them costly. Therefore, the objective of this chapter is to develop a less data-intensive fusion approach. To achieve this, a novel dataset was first created by manually labeling point clouds acquired in large-scale urban scenes. Then, this chapter introduces a new approach called Plf4SSeg (Prior-Level Fusion for Semantic Segmentation). This approach integrates geometric and intensity data from 3D point clouds with RGB information from aerial images for semantic segmentation. This approach consists of two primary steps: (1) image classification and (2) fusion of classified images and 3D point clouds. The classification of aerial images is based on training zones selected to align with the semantic classes of the LiDAR dataset, addressing the issue of semantic class inconsistency between LiDAR and image datasets. The approach begins with image classification, utilizing the Maximum Likelihood Classifier (MLC) as a supervised classification method. MLC is preferred due to its ability to consider variance-covariance within class distributions and its suitability for normally distributed data, leading to higher precision. This initial step provides a preliminary semantic segmentation based on the spectral information of objects. Combining this with 3D point clouds (X, Y, Z, and intensity) helps overcome their limitations. The training data is generated by assigning raster values from each classified image to the corresponding point cloud, based on (X, Y) coordinates. Finally, an advanced deep learning technique, "RandLaNet," is adopted for 3D semantic segmentation. RandLaNet is chosen for its direct and random processing of 3D point clouds, performing point sampling without the need for pre/post-processing operations.

Based on Article [1]

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

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

Published in 2022

Type: Open Access Article

Publisher: MDPI

Journal: Remote Sensing

Special Issue : Semantic Segmentation Algorithms for 3D Point Clouds

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], SensatUr- ban [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 dis- tinct 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 ap- proaches 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 segmen- tation 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 air- borne 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 geo-registration 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.

1. Materials and Methods

1.1 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 8). 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 (Figure 9 and Figure 10), and will be publicly available online.

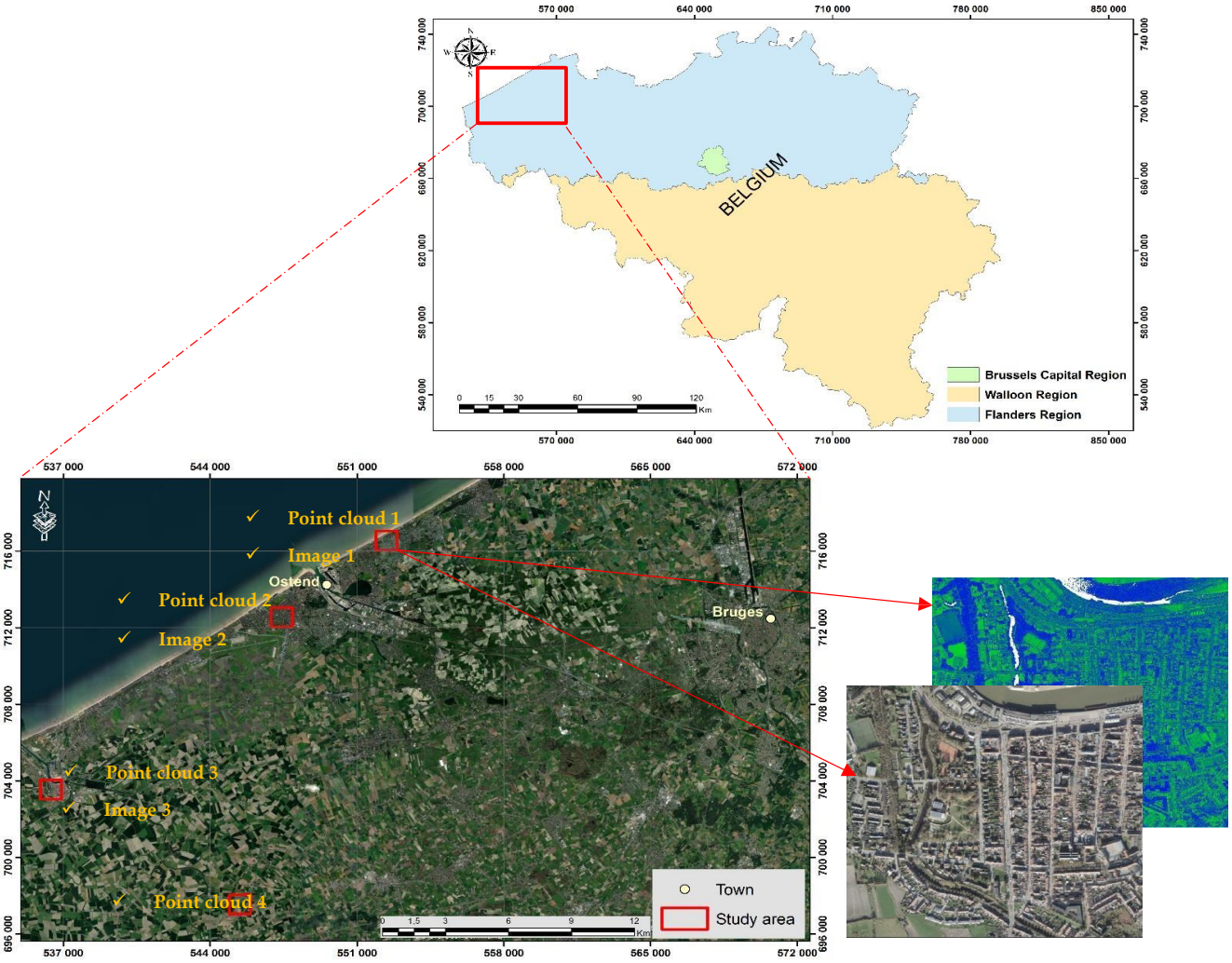


Figure 8. Location of datasets.

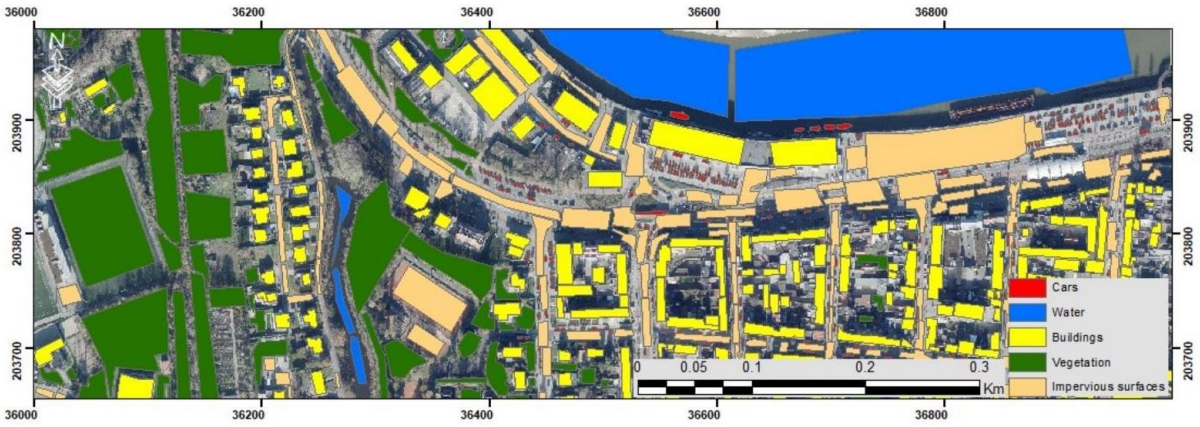


Figure 9. Example of classified point cloud from the created dataset.

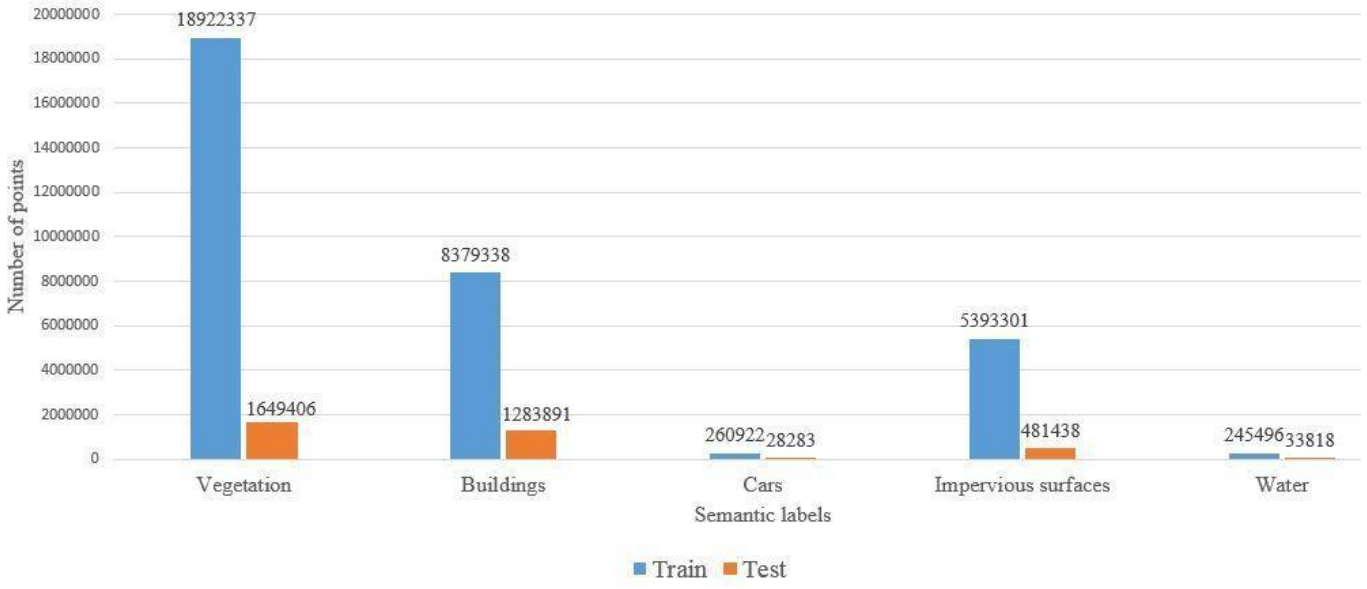


Figure 10. The distribution of different semantic classes in the created dataset.

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

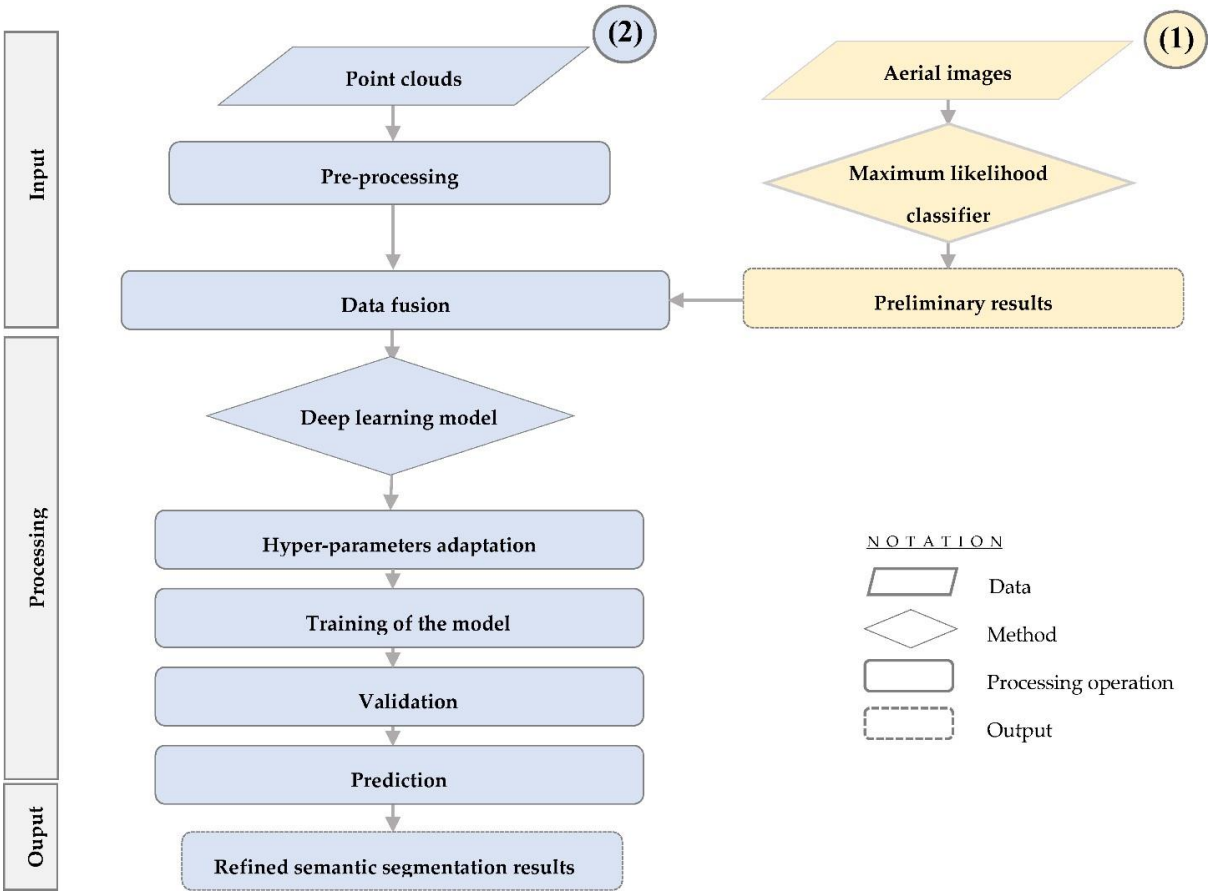


Figure 11. The general workflow of the proposed approach.

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 12 summarize the general process followed for image classification.

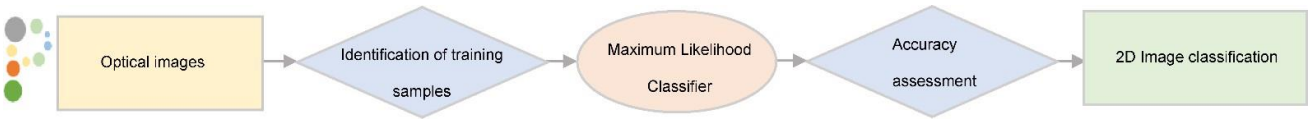


Figure 12. Methodological workflow for image classification.

The MLC is the most common statistical method used for image supervised clas- sification. 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 8) 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 13).

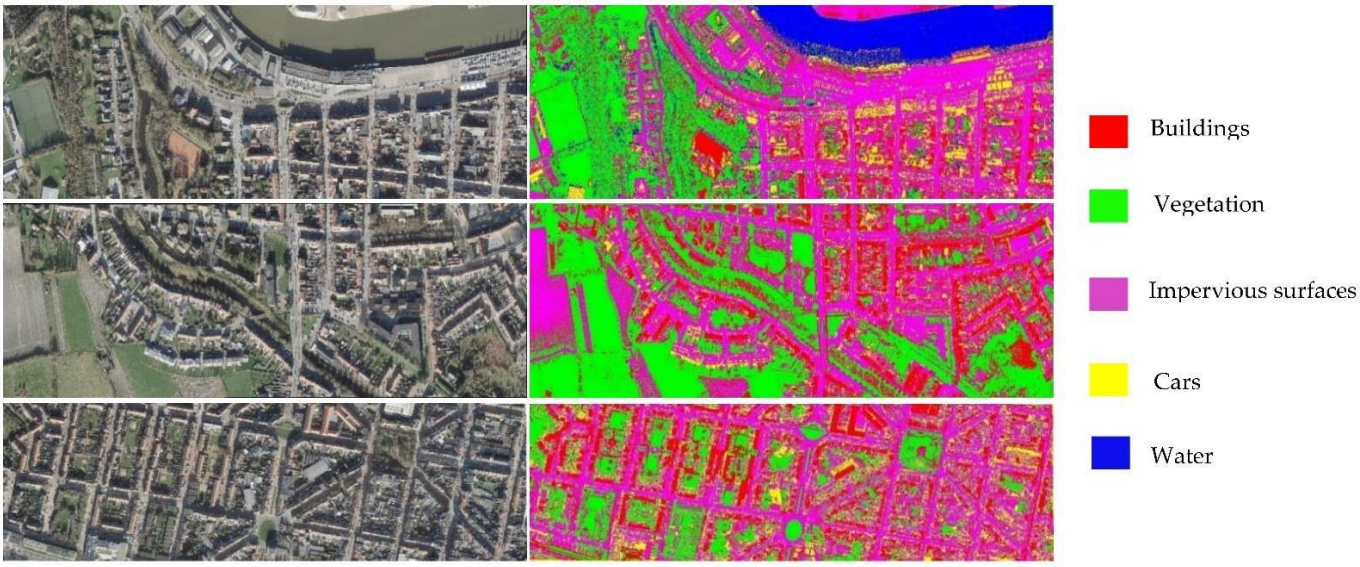


Figure 13. Examples of image classification results.

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:

Point cloud 1 (X1 + Y1 + Z1 + Intensity1 + Image classification 1) (1)

Point cloud 2 (X2 + Y2 + Z2 + Intensity2 + Image classification 2) (2)

Point cloud n (X_n + Y_n + Z_n + $Intensity_n$ + Image classification $_n$) (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 Rand- LaNet 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 14) precisely in terms of (XYZ) coordinates and intensity information.

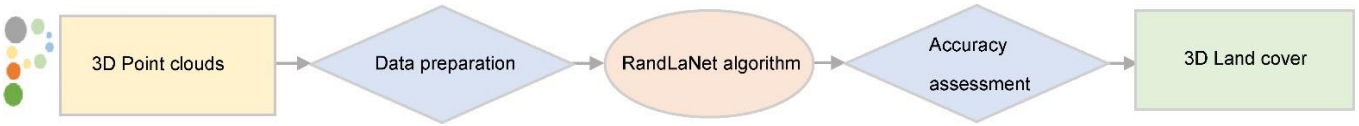


Figure 14. The general workflow of the non-fusion approach.

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.

$$\textbf{Accuracy} = (TN + TP) / (TP + FN + TN + FP)$$

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).

$$\textbf{Recall} = TP / (TP + FN)$$

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

$$\textbf{Precision} = TP / (TP + FP)$$

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

$$\textbf{IoU} = TP / (FP + TP + FN)$$

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.

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

- 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 5. Subsequently, the results obtained were compared with the non-fusion approach (Table 6). The objective was to study the contribution of data fusion to semantic segmentation quality.

A. Results of Plf4SSeg approach

Table 5. 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 6. 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 5 below report the resulting metrics.

From Table 5, 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 15) 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.

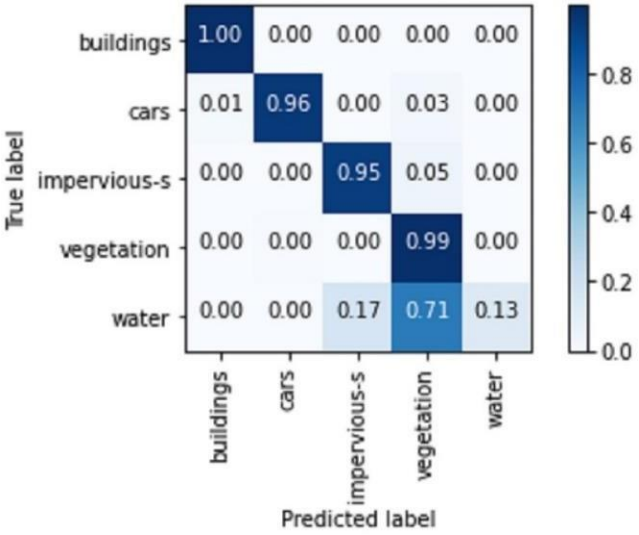


Figure 15. Normalized confusion matrix.

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 superim- posed on point clouds of the study area. The examples of point clouds (Figure 16A) and their corresponding semantic segmentation results (Figure 16B) are illustrated below (Figure 16).

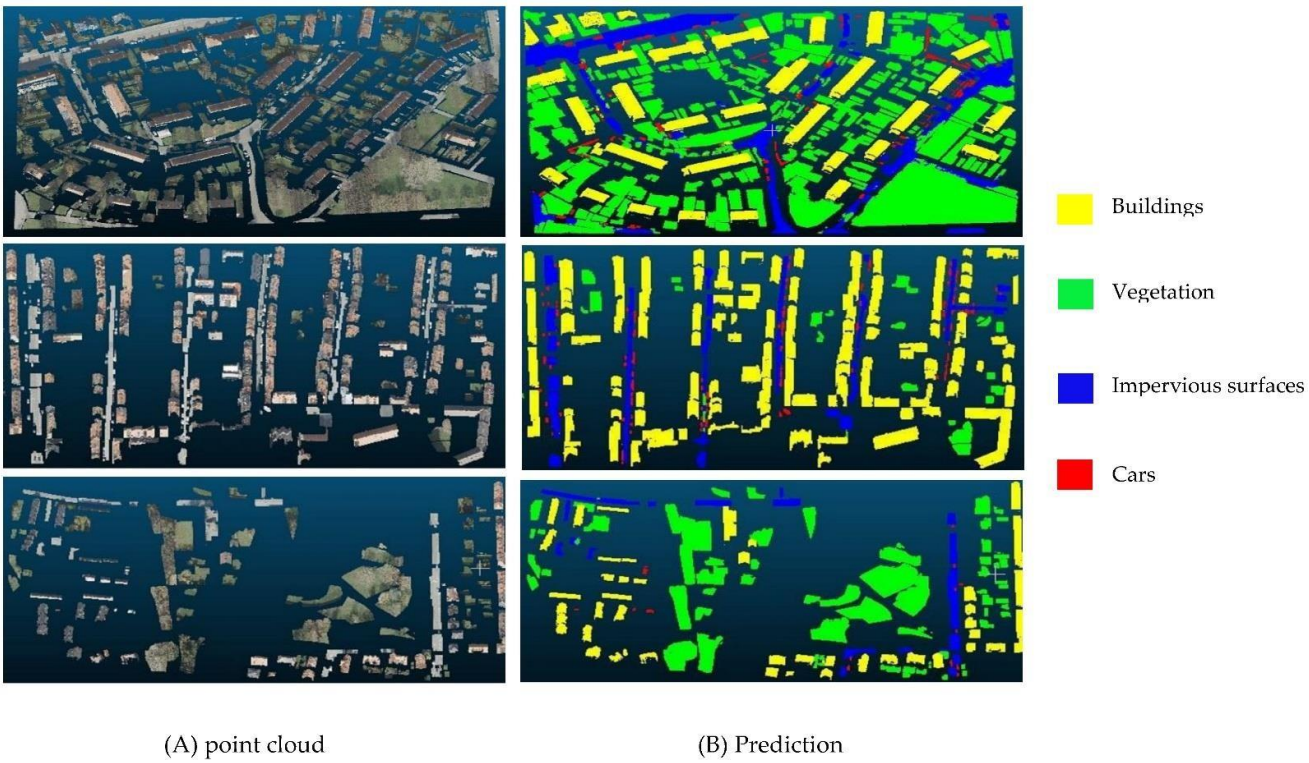


Figure 16. Examples of 3D semantic segmentation results obtained by the Plf4SSeg approach.

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 6 show the quantitative evaluation of the test results for different approaches.

Table 6 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 6). 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 17 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 5).

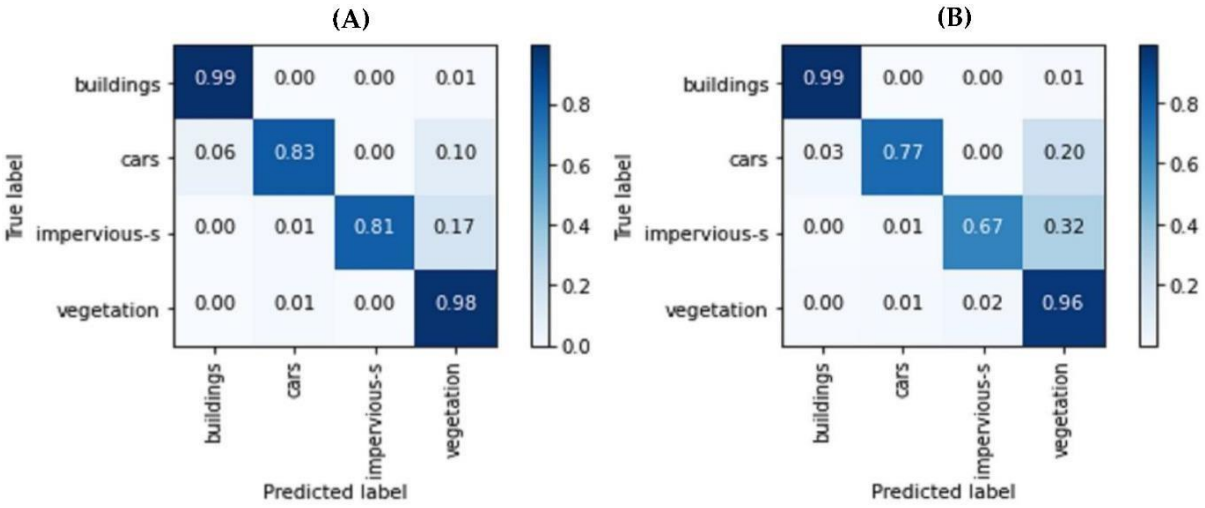


Figure 17. Normalized confusion matrix of the proposed approach (A) and the non-fusion approach (B).

Finally, it should be noted that this research work presents certain limitations, in- cluding 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]

Chapter 3

Investigating Prior-Level Fusion Approaches for Enriched Semantic Segmentation of Urban LiDAR Point Clouds

PREFACE

Following the previous chapter, which developed a less data-intensive fusion approach and demonstrated satisfactory semantic segmentation accuracies, the next objective is not only to improve precision but also to enhance the semantic richness of point clouds. This involves extracting maximum urban details (e.g., footpaths, parking, etc.) while maintaining superior accuracies for all classes. To do this, this chapter introduces a novel approach to extracting enriched semantic 3D objects from large-scale urban environments. This is achieved through the development and benchmarking of three prior-level fusion scenarios. This approach integrates fused knowledge into the deep learning technique at the prior level of the semantic segmentation pipeline. The developed fusion approach is motivated by the fact that semantic segmentation can significantly benefit from the fusion of point clouds with additional knowledge. This is particularly relevant in cases where distinguishing enriched semantic objects proves to be complex. Three distinct scenarios were conceived and investigated, each involving the fusion of point clouds, aerial images, and specific types of prior knowledge. The prior knowledge includes geometric features, classified images, or classified geometric information. To implement this, two deep learning techniques were adopted: "RandLaNet" and "KPConv." The scenario demonstrated most efficient in accurately extracting the maximum semantic information is derived as the "Efficient-PLF (Prior-Level Fusion) approach". The significance of this chapter lies in deploying an automated approach for enriching semantic segmentation of ALS point clouds. The presentation of all three scenarios in this chapter ensures a comprehensive comparative evaluation. This highlights the robustness and validity of the derived scenario. Each scenario, based on a specific workflow, delivers distinct performances. It is worth noting that even though the other two scenarios were not chosen as optimal, they prove useful in specific use cases. For instance, the second scenario, while not the primary choice, was recommended for its outstanding performance in certain specific classes.

Furthermore, this chapter adds a section to the publication, which outlines a methodology for extracting objects from high-resolution images and projecting them onto point clouds. This method has two main objectives. Firstly, it aims to detect semantic classes that are not present in the LiDAR dataset used in LiDAR-based approaches, thereby introducing additional classes. Secondly, it aims to improve the precision of object extraction for cases where LiDAR approaches may have limitations. This developed methodology addresses these challenges, contributing to the enhancement of semantic enrichment in the LiDAR-based process.

Based on Article [2]

Investigating Prior-Level Fusion approaches for Enriched Semantic Segmentation of Urban LiDAR Point Clouds

Abstract:

3D semantic segmentation is the foundation for automatically creating enriched Digital Twin Cities (DTCs) and their updates. For this task, prior-level fusion approaches show more promising results than other fusion levels. This article proposes a new approach by developing and benchmarking three prior-level fusion scenarios to enhance the outcomes of point cloud enriched semantic segmentation. The latter were compared with a baseline approach that used the point cloud only. In each scenario, specific prior knowledge (geometric features, classified images, or classified geometric information) and aerial images are fused into the neural network's learning pipeline with the point cloud data. The goal is to identify the one that most profoundly enhances the neural network's knowledge. Two Deep Learning techniques,"RandLaNet" and "KPConv," were adopted, and their parameters were modified for different scenarios. Efficient feature engineering and selection for the fusion step facilitated the learning process and improved the semantic segmentation results. Our contribution provides a good solution for addressing some challenges, particularly for more accurate extraction of semantically rich objects from the urban environment. The experimental results have demonstrated that Scenario 1 has higher Precision (88%) on the SensatUrban dataset compared to the baseline approach (71%), the Scenario 2 approach (85%), and the Scenario 3 approach (84%). Furthermore, the qualitative results obtained by the first scenario are close to the ground truth. Therefore, it was identified as the efficient fusion approach for point cloud enriched semantic segmentation which we have named the Efficient Prior-Level Fusion (Efficient-PLF) approach.

Keywords: prior-level fusion; enriched semantic segmentation; LiDAR point clouds; images; data fusion; prior knowledge; deep learning; urban environment.

Published in 2024

Type: Open Access Article

Publisher: MDPI

Journal: Remote Sensing

Special Issue: 3D and Semantic Reconstruction of the Urban Environment Using Multi-Modal and Multi-Resolution Remote Sensing Data

1. Introduction

Many cities worldwide are building their Digital Twin Cities (DTCs) [1]. Semantic 3D city models, essentially built from LiDAR point clouds through semantic segmentation, are the foundation for developing these DTCs both for academic and industry research [2,3]. Semantic segmentation allows for the semantic enrichment of 3D city models, their updates, and the performance of multiple spatial and thematic analyses for city management, urban planning, and decision making. Despite challenges in acquisition and processing, LiDAR technology has made significant advancements in capturing highly detailed three-dimensional data with substantial point density, finding versatile applications in urban planning, outdoor navigation, and urban environmental studies [4].

The advancement of computer vision technology and the widespread utilization of deep learning (DL) methods have resulted in the development of more robust and reliable 3D semantic segmentation techniques. Indeed, many DL techniques have been developed recently for 3D semantic segmentation [5–7]. DL techniques are proposed to handle complex tasks in various LiDAR applications. Among these techniques, we can cite the deep neural networks (DNNs), which have gained considerable popularity and attention due to their efficiency. The present focus is on developing new DL-based approaches to enhance the quality of semantic segmentation outcomes. Then, it is necessary to compare them with the existing approaches to derive the most suitable one for LiDAR point clouds processing.

In the literature, we observe that achieving the maximum amount of semantic information in the urban environment (i.e., extracting the maximum of urban objects such as Traffic Roads, Cars, etc.) with high precision remains a challenge. Most current approaches to semantic segmentation using LiDAR point clouds demonstrate good accuracy for easily extractable classes such as Buildings and Ground. However, extracting more detail and accurately identifying challenging classes, such as Parking and Street Furniture, remains an important research topic. To address this challenge, fusion approaches show higher accuracy compared to non-fusion approaches [8,9]. Within fusion approaches, prior-level fusion approaches exhibit better precision than point-level, feature-level, and decision-level fusion approaches, as explained later. This is why the objective was to delve into this family of prior-level approaches. To achieve this, we proposed an efficient prior-level fusion approach to enhance the knowledge of deep learning techniques for 3D semantic segmentation by integrating prior knowledge into the learning pipeline. This approach explicitly tackles the challenge of accurately extracting the maximum amount of urban objects (Footpaths, High Vegetation, etc.). It is motivated by the understanding that 3D semantic segmentation can gain advantages from the fusion of point clouds (PCs), aerial images, and prior knowledge, especially in cases where the differentiation between detailed urban objects is challenging. Some initiatives have been proposed in the literature [10], but to the best of our knowledge, no study has systematically developed and evaluated all possible scenarios of injecting prior knowledge and aerial images into point clouds, especially during the training phase of DL techniques. We have not only moved beyond traditional PC attributes but have also adopted advanced DL techniques, “RandLaNet [5]” and “KPConv [8]”, and optimized their parameters. For finding the efficient approach, three distinct scenarios were conceived and investigated. Each scenario involved the fusion of PCs, aerial images, and a specific type of prior knowledge. The efficient scenario that demonstrated the ability to extract the maximum amount of semantic information in an urban environment was identified from the evaluations. This scenario is derived as the “Efficient-PLF approach”. Our research’s potential lies in

deploying an automated enriched semantic segmentation pipeline with a high level of detail. While we have highlighted the optimal scenario, presenting all three scenarios not only ensures a comprehensive benchmark but also affirms the robustness and validity of our chosen approach. Each scenario is based on a specific workflow and provides different performances. It worth highlighting that even the two other scenarios that were not chosen as optimal are useful in specific use cases. For example, the second scenario, despite not being the primary choice, was recommended due to its outstanding performance on certain specific classes.

The following are the main contributions of this paper:

- ✓ Designing three prior-level fusion scenarios for 3D semantic segmentation that fuse PCs, aerial images, and prior knowledge into the DL pipeline;

- ✓ Evaluating the performance of each scenario in terms of enhancing DL techniques' knowledge;

- ✓ Enhancing semantic segmentation richness by detecting a maximum number of urban classes more efficiently and accurately;

The paper is organized as follows: Section 2 showcases the principal advancements made in fusion-based approaches for PC semantic segmentation. A detailed description of the fusion scenarios we developed is presented in Section 3. The experimental methodology and the obtained results are reported in Section 4. The discussion of our findings is in Section 5. Finally, the paper ends with a conclusion.

2. Related Works

The increasing need for automated urban assets extraction has resulted in 3D semantic segmentation of multi-sensor data becoming a rapidly growing and dynamic field of research. Although 3D urban semantic segmentation is based on 3D LiDAR data, other data sources (geometric features, classified images, etc.) can provide supplementary relevant information. The latter can compensate for the limits of 3D PCs; such as the confusion between artificial and natural objects and the fact that PCs are less suitable for delineating object contours. Promising results have been achieved in 3D semantic segmentation through the fusion of 3D PCs with other data sources, as demonstrated by several studies in the literature [9,10]. Furthermore, adding highly informative data is a major boost to semantic segmentation. The DL revolution has demonstrated that many three-dimensional semantic segmentation challenges (the automation of treatments, their speed, the precision of results, etc.) are addressed by DL techniques (PointNet++, SPGraph, etc.). On the other hand, it is well known that more training labelled PCs are required for learning models. Motivated by the high demand for training data, various datasets have been developed recently. The majority of

them are freely available online. We can list Toronto-3D, SensatUrban [11], Benchmark Dataset of Semantic Urban Meshes (SUM) [12], and Semantic3D [13]. Despite the efforts made, 3D semantic segmentation remains a delicate and complex task due to the spectral and geometric similarity between different urban classes. Due to the remarkable performance achieved lastly by fusion approaches in semantic segmentation tasks, it would be interesting to advance in this research niche. Fusion-based approaches are applied by fusing data from different sensors at different fusion levels. Fusion-based approaches can be categorized into four families that combine PCs with other sources: (1) Prior-level fusion approaches, (2) Point-level fusion approaches, (3) Feature-level fusion approaches, and (4) Decision-level fusion approaches.

2.1 Prior-Level Fusion Approaches

Fusing at the prior level assigns classified images to 3D PCs, enhancing LiDAR data semantic segmentation. This approach expedites convergence and reduces loss, thanks to direct image classification [14], but has challenges with non-overlapping areas and uncertainties [15]. There is a scarcity of prior-level fusion approach-based studies in the existing literature. Among them, ref. [16] proposed a fusion approach of images and LiDAR PCs for semantic segmentation. The proposed approach was compared with point-level, feature-level, and decision-level fusion approaches. The ISPRS dataset evaluation showed that the proposed approach outperformed all other fusion approaches with a good F1-score (82.79%). Ref. [17] proposes a fusion approach based on 2D images and 3D PCs to segment complex urban environments. The prior knowledge obtained from 2D images was mapped to PCs. Subsequently, the fine features of building objects were precisely and directly extracted from the PCs based on mapping results. Their results showed that the created model is adapted for high-resolution images and large-scale environments. Finally, a recent study [18] presented a new fusion approach for semantic segmentation in urban areas, which operates at the prior level. Their approach utilizes both aerial images and 3D PCs. Achieving an intersection over union of 96%, their results outperform the non-fusion approach, which only achieves 92%.

2.2 Point-Level Fusion Approaches

Point-level fusion assigns optical image spectral data to each point and uses a DL technique for 3D point cloud semantic segmentation. While these methods yield good results, they demand significant memory, computation time, and synchronized data acquisition times. Several point-level fusion processes are available for 3D semantic segmentation. [19] introduced PMNet, a DL architecture that merges optical images with PC, accounting for the permutation invariance properties of the latter. This approach has proven to be superior to observational- and global feature-level fusion approaches. Meanwhile, [20] developed a CNN-based approach for 3D semantic segmentation by integrating radiometric properties from image data. When tested on the SemanticKITTI dataset, their approach exhibited an 8.7% increased average accuracy in certain categories relative to a separate approach that combines image and PC, and it operated with a faster runtime. In another study, [21] investigated the benefits of blending CASI (Compact Airborne Spectrographic Imager) hyperspectral and airborne LiDAR data for land cover semantic segmentation, employing

PCA (Principal Components Analysis) and layer stacking. They used ML (Maximum Likelihood) and SVM (Support Vector Machine) classifiers for data categorization, observing that the fusion approach delivered an accuracy improvement of 9.1% and 19.6%, respectively, over approaches utilizing only LiDAR or CASI data.

2.3 Feature-Level Fusion Approaches

Feature-level fusion combines optical image and 3D point cloud features through neural networks for semantic segmentation. Such fusion delivers robust results, outperforming approaches using only radiometric or geometrical data [18]. However, drawbacks such as orthophoto wrapping and LiDAR's limitations in capturing occluded objects are notable. The importance of feature fusion in enhancing the quality of semantic segmentation is widely recognized in the literature. [22] employed spectral, texture, and shape features from hyperspectral images to minimize classification errors, emphasizing that it is challenging to find a singular optimal combination of features suitable for all datasets. They showed that even a basic combinatorial process using complementary features can be effective and highlighted the advantage of incorporating spatial information (shape features, texture, etc.) for improved semantic segmentation. In another study conducted by [23], a feature fusion approach was presented for classification tasks that utilized softmax regression. This approach took into account the likelihood of an object sample belonging to different classes and incorporated object-to-class similarity information. Experiments revealed that their method surpassed other baseline feature fusion methods like SVM and logistic regression, particularly in gauging feature similarity across multiple spaces, underscoring the potential of a softmax regression-based approach.

2.4 Decision-Level Fusion Approaches

Decision-level fusion merges the outcomes of semantic segmentation from individual neural networks, combining results from classifiers focused on either LiDAR space or pixel [24]. This fusion offers advantages like independent training and low complexity, given that each modality employs its own DL technique, capturing distinct feature representations. Yet, its reliance on both classifiers can inherit their limits, and it demands more memory and extra parameters due to its DL structure. The existing literature on decision-level fusion is sparse.[15] introduced a fusion approach for classification and object detection, fusing semantic segmentation results from unary classifiers via a CNN. Their approach, tested on the KITTI benchmark, achieved a 77.72% average precision. However, it had real-time application challenges and lower accuracies for "cyclists" and "pedestrians" classes because of sensor-derived incomplete data. Similarly, [25] suggested a fusion approach combining object-based image analysis on multiview very-high-resolution imagery and DSM. Their approach bolstered object recognition accuracy, showing improvements in kappa and overall accuracy metrics for DMC and WorldView-2 benchmarks. Yet, not all DMC benchmark class results were enhanced. Lastly, [26] presented a late fusion approach merging multi-modality information. The approach includes a pairwise CRF (Conditional Random Field) to enhance the spatial consistency of the structured prediction in a post-processing stage. Using the KITTI dataset for evaluation, their approach achieved a class accuracy of 65.4% and a per-pixel accuracy of 89.3%.

2.5 Summary

Previous research has highlighted the effectiveness of semantic segmentation approaches that leverage PCs combined with other data sources, such as satellite or aerial images. It demonstrates precise and high-quality visual outputs. In the literature, the commonly used fusion approaches of 3D LiDAR and image data can be categorized into four main types: prior-level, point-level, feature-level, and decision-level fusion approaches. The prior-level approaches are the new fusion approaches in the literature. They have enhanced the accuracy of semantic segmentation results. Additionally, they demonstrate good performances in semantic segmentation, especially in terms of precision. This precision was improved by the direct use of semantic knowledge from classified images. Moreover, they demonstrated the low-loss function in training and testing steps in comparison to other fusion approaches. Thus, because this approach type integrates semantic information from images, the loss reaches a stable state faster and becomes smaller. However, these processes are a bit long. The point-level fusion approaches are the most dominant, quickest, and simplest in the literature. However, these processes are not able to classify complex urban scenes containing a diversity of urban objects, especially, the geo-objects with geometric and radiometric similarity. The feature-level fusion approaches allow objective data compression. Consequently, they guarantee a certain degree of precision and retain enough important information. Nevertheless, the features extracted sometimes do not reflect the real objects. The decision-level fusion approaches are less complex and flexible. For that reason, the two semantic segmentation processes (one of the images and the other of the PCs) do not interfere. Nonetheless, these approach types can be affected by errors in both processes. In addition, decision-level fusion approaches require more memory since the DL structure fuses feature later. Additionally, layers need extra parameters for convolution and other operations. The performance and limitations of each approach can be accessed in Table 1 and are summarized on the following GitHub link: https://github.com/ZouhairBALLOUCH/Supplementary_Results_Article.git (accessed on 1 December 2023).

3. Materials and Methods

In this research, we adopted the prior fusion approaches that have demonstrated good results compared to the others. Therefore, we proposed and thoroughly evaluated three prior-level fusion scenarios to derive the Efficient-PLF approach, enhancing the DL technique knowledge.

3.1 Dataset

Our developed scenarios were evaluated using the SensatUrban dataset [11], which contains nearly three billion annotated points and was released at CVPR2021. The utilization of this dataset is justified by its high semantic richness compared to other existing airborne datasets. The 3D PCs were obtained by a UAV (Unmanned Aerial Vehicle) which follows a double-grid

flight path. Three sites of Birmingham, Cambridge, and York cities were covered. The dataset covers about six square kilometers of an urban area with a diversity of urban objects. The SensatUrban dataset contains 13 semantic classes: Street Furniture, Traffic Road, Water, Bike, Footpath, Car, Rail, Parking, Bridge, Wall, Building, Vegetation, and Ground. Each point contains six attributes: X, Y, Z, and RGB information. The allocation of semantic categories to objects within the dataset is extremely imbalanced, with the Bike and Rail classes collectively accounting for just 0.025% of the overall points present in the dataset. The SensatUrban dataset is freely available online at (https://github.com/QingyongHu/SensatUrban, accessed on 10 March 2023). However, it should be noted that the dataset's semantic labels for the testing data are not provided. Thus, to evaluate the proposed approach, the training data of SensatUrban were partitioned into new training and testing sets. In our experiments, a part of the training data (18 sets) were used to implement the first parts of the developed scenarios S1 and S3 (Section 3), while the remaining part of the data (16 sets) were used to implement the second steps of scenarios S1 and S3, S2, and the baseline approach (the main part of this work).

3.2 Methodology

Our study aims to create and evaluate three prior-level fusion scenarios to derive the Efficient-PLF one. Counting the baseline, the general work methodology includes four processes, as depicted in Figure 18. The first one consists of injecting classified images and spectral information as attributes into the PCs. The second is based on geometrical features (extracted from PCs), XYZ PCs, and aerial images (S2). The third classifies urban space using classified geometrical information, aerial images, and PCs (S3). The fourth process, known as the baseline approach, directly combines raw PCs and images. Afterwards, the advanced DL techniques "RandLaNet" and "KPConv" were adopted to implement the different four processes. An assessment of the results obtained by the different processes was performed based on metrics computation and visual investigations.

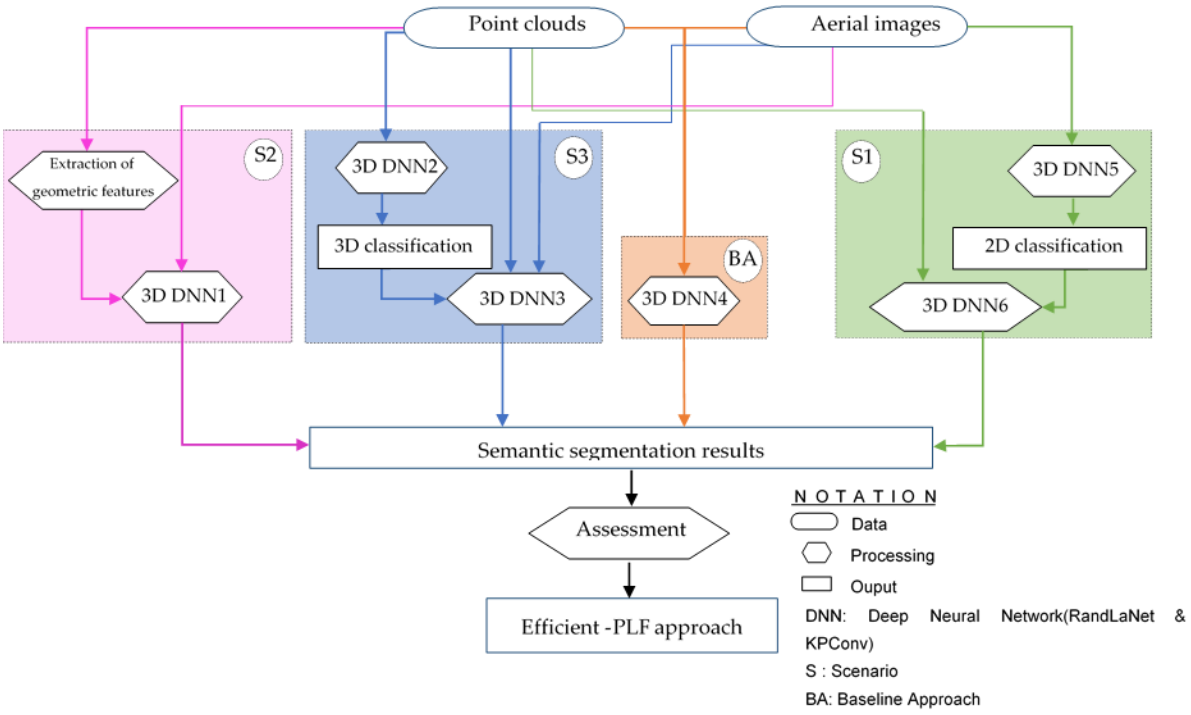


Figure 18. The general workflow.

In aerial image fusion, the Red, Green, and Blue bands were averaged into a single attribute for each PC. The aim is to propose a cost-effective scenario with fewer inputs, relying more on the investigation of the prior knowledge to identify the best-performing scenario. In this study, we used both the RandLaNet [5] and KPConv [8] techniques to evaluate the proposed scenarios. RandLaNet, introduced by [5], is a DL technique designed for large-scale PC, offering excellent computational and memory performance through random point sampling [27]. It requires no preprocessing or postprocessing, and incorporates a local feature aggregation module to retain geometric data details. KPConv, on the other hand, directly handles PCs and stands out for its ability to place convolution weights in Euclidean space using kernel points. This technique's adaptability ensures alignment with the point cloud's local geometry, offering precise results. Notably, KPConv outperforms traditional techniques, making it suitable for tasks requiring accuracy and resilience against density shifts. These techniques were not selected arbitrarily; their features directly align with our objective, and their efficacy has been empirically validated in numerous studies [28–31]. The mathematical formulas for the RandLA-Net technique are detailed in Sections 3.2 and 3.3 of [5], while those for the KPConv technique are outlined in Section 3 of [8]. We want to highlight that the objective of this work does not focus on the type of DL technique but rather on finding the right approach for selecting relevant features and the efficient fusion scenario.

3.2.1 Classified Images and PC-Based Scenario (S1)

The flowchart depicted in Figure 19 summarizes the first proposed scenario (S1), which uses 3D PC, aerial images, and classified images. In this scenario, the aerial images are extracted from the projection of the 3D point cloud into a 2D representation with colors. The

incorporation of aerial images into the point cloud has already been justified. However, the injection of classified images and spectral information as attributes of PCs into the DL technique during its training is justified by several reasons. Integrating classified images brings a semantic dimension to the scenario and provides detailed information about different object categories present in the urban environment. This prior knowledge enhances the neural network's knowledge during the learning pipeline. Furthermore, it can be valuable in guiding semantic segmentation by reducing false negatives and false positives. By leveraging this semantic information, this scenario can achieve more consistent results in object identification. This accelerates the convergence of this scenario, resulting in enhanced precision in urban object extraction.

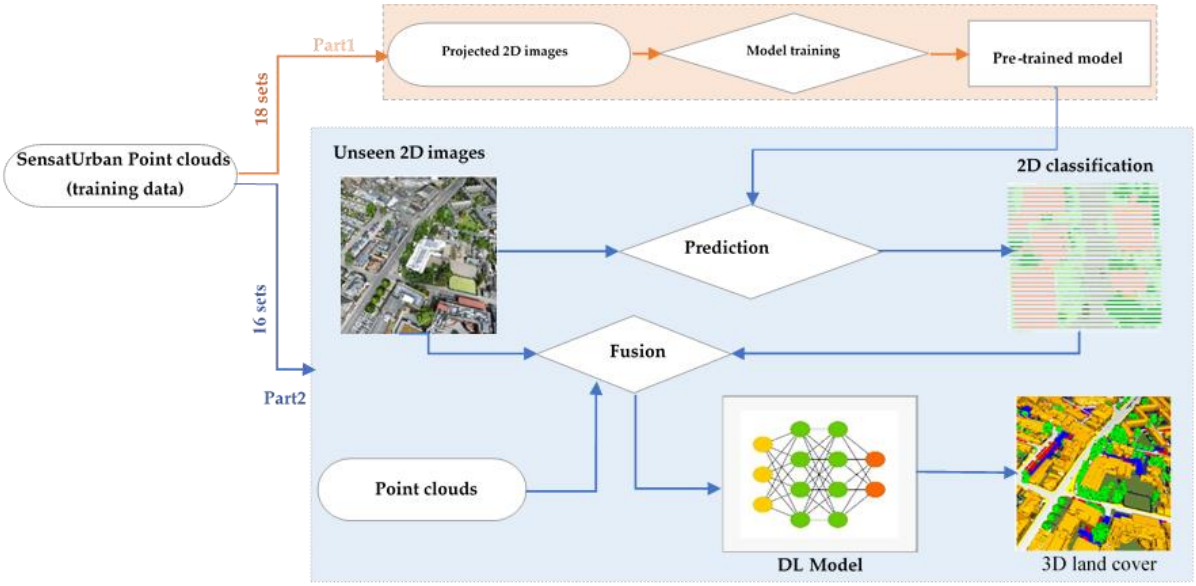


Figure 19. The first proposed scenario (S1).

To implement this scenario, we randomly divided the SensatUrban dataset into two parts: one containing 18 PCs and the other containing 16. First, the images were extracted from the colors of the 18 PCs of the dataset. The extraction of images was performed automatically using a batch processing script. Then, the technique was trained on these images and integrating them with their attributes (Red, Green, Blue). The use of RandLaNet instead of a 2D image classification technique is justified by the fact that we just multiplied the height by a scale factor of 0. So, our image is just a flattened point cloud, not pixels. After obtaining the trained model, we returned to the dataset containing 16 PCs (part 2) and extracted the images in the same manner. We then classified them using the trained model and merged them with the PCs (XYZ coordinates) and aerial images (RGB). Thus, each point cloud contained the following attributes: X, Y, Z, R, G, B, 2D classification. Finally, these prepared PCs were used to train the techniques (RandLaNet + KPConv). The fundamental hyperparameters of the original versions of the techniques have been adapted, and the techniques were evaluated using the test data.

3.2.2. Geometric Features, PC, and Aerial Images-Based Scenario (S2)

The idea of the second proposed scenario is to combine the geometric features, XYZ PCs, and aerial images. The aim of this scenario is to examine the contribution of geometric properties to improving the knowledge of the DL technique in the semantic segmentation pipeline. As shown in Figure 20, S2 mainly contains two parts: (A) Automatically selecting the appropriate geometric features for semantic segmentation; (B) Injecting selected geometric features with aerial images into PCs to improve knowledge of the techniques (RandLaNet + KPConv).

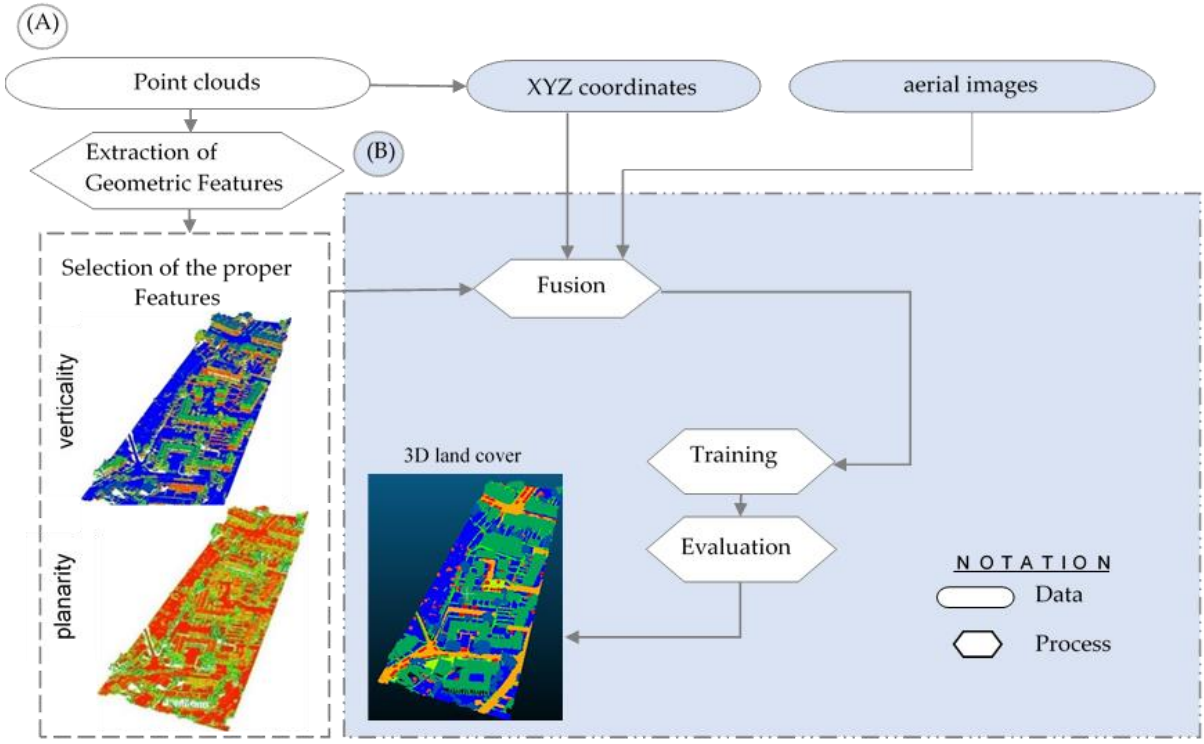


Figure 20. The second proposed scenario (S2). (A) Selection of the appropriate geometric features. (B) Data Training and Semantic Segmentation Using RandLaNet and KPConv Techniques.

(A) Selection of the appropriate geometric features

The use of geometric features can help elucidate the local geometry of PCs and is now commonly employed in 3D PC processing. Extracting these properties at multiple scales instead of a single scale aims to improve precision values. “Geometric features are calculated by the eigenvalues (λ1, λ2, λ3) of the eigenvectors (v1, v2, v3) derived from the covariance matrix of any point p of the point cloud” [32]:

$$cov(S) = \frac{1}{S}\sum_{p \epsilon S}(p - \bar{p})(p - \bar{p})^T$$

“Where p is the centroid of the support S” [32]. Several properties are calculated using eigenvalues: omnivariance, the sum of eigenvalues, eigenentropy, linearity, anisotropy, planarity, surface variation, verticality, and sphericity. A table summarizing the mathematical formulas for geometric features can be accessed via the following link: https://github.com/ZouhairBALLOUCH/Supplementary_Results_Article.git (accessed on 1 December 2023).

Geometric feature extraction is a crucial part of 3D semantic segmentation. Independent of the urban object to be semantically segmented and the data resolution, the geometric properties significantly impact the results. The geometric features have great importance by providing the DL structure with useful information about each urban class. Consequently, it helps the classifier to distinguish between different semantic classes. However, some of these geometric properties may mislead the semantic segmentation process. So, these errors should be considered during the analysis of results.
To select the geometric properties with the most positive impact on semantic segmentation results, all geometric features were initially calculated (anisotropy, planarity, linearity, etc.). Generally, to determine the importance of these features, automatic methods can be employed, such as the feature importance assessment offered by libraries like Scikit-learn. Consequently, planarity and verticality were selected for integration as attributes of PCs based on their importance to separate between classes. The geometric features with the least impact on the model training have been removed. The following are the geometric properties used in this study:
Planarity is a characteristic that is obtained by fitting a plane to neighboring points and computing the average distance between those points and the plane [33].
Verticality: The angle between the XY-plane and the normal vector of each point is calculated using its 3D surface normal values [33].

(B) Data Training and Semantic Segmentation Using RandLaNet and KPConv Techniques

In this scenario, selected geometric properties (planarity and verticality) and RGB from images were added as attributes to the PCs for the implementation of both the RandLaNet and KPConv techniques. To implement this scenario, we started with the preparation of the training data. As mentioned earlier, we divided our dataset into two parts. One of these parts contains 18 PCs, while the other contains 16. In the case of this specific scenario, we worked only with the set that contains 16 PCs. These are the same PCs that are used to implement

the second part of the other scenarios proposed in this work. The generation of training data was performed by calculating the geometric features (planarity and verticality) for each point cloud. The calculations were performed using the Cloud Compare software (version 2.12.4). The geometric features were computed with a 0.4 m radius sphere, representing support obtained with a radius of 4 m. Afterward, these geometric properties were merged with PCs and aerial images. This data preparation methodology was applied to all the PCs in the 16 datasets used.
Afterwards, during the training step, certain configurations and data representations were adjusted for both the original versions of RandLaNet and KPConv, including the format of the input tensor and data types. Some of the hyperparameters (such as the size of the first subsampling grid and the radius of the input sphere) were also modified. Finally, after training and validation of both the RandLaNet and KPConv techniques, the pre-trained models were used to predict the labels of the test data.

3.2.3. Classified XYZ PC, PC, and Optical Images-Based Scenario (S3)

We intend to explore a third scenario that also has not been previously examined in the literature. The proposed scenario (Figure 21) involves injecting classified point cloud (based only on XYZ coordinates) and radiometric information extracted from aerial images as attributes of PCs into the DL technique's learning pipeline. The use of PCs only in semantic segmentation may be insufficient due to the confusion between some semantic classes. To address this challenge, we decided to incorporate the described prior knowledge. This integration into DL technique's training would enable it to learn and enhance the delineation of 3D object contours more effectively. As a result, it becomes easier to differentiate between different objects. Furthermore, a rapid convergence was also expected in the training step.

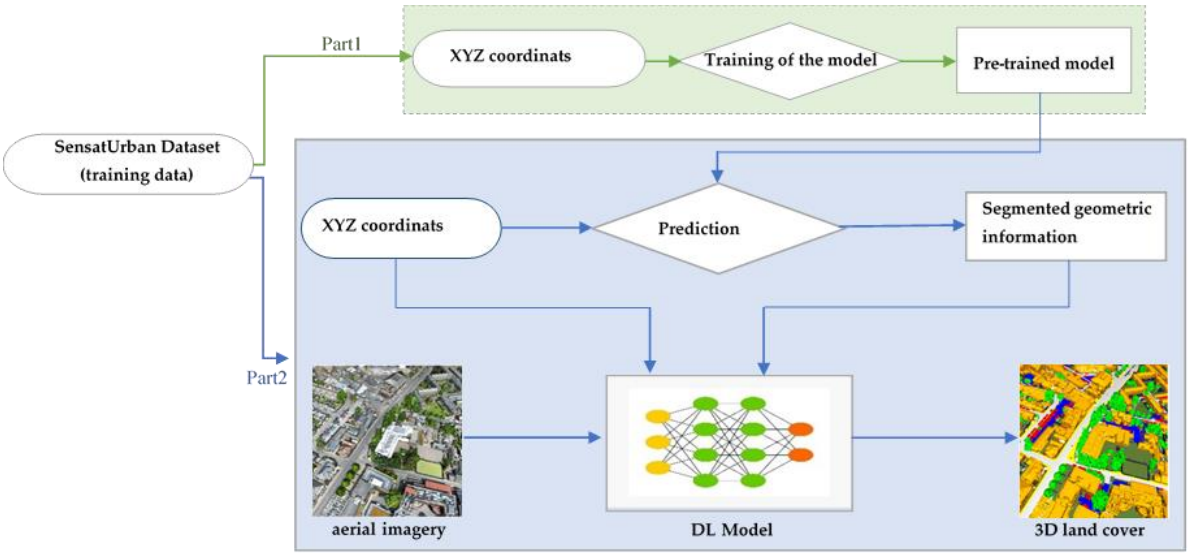


Figure 21. The third proposed scenario (S3).

For the implementation of S3, 18 sets of the SensatUrban dataset were used to perform the first part of the scenario and 16 sets to perform its second part (see Figure 21). The proposed process includes two main steps. Firstly, 18 sets of PCs that contain only the three attributes X, Y, and Z from the SensatUrban dataset were used in the training step. After that, the pre-trained model was used to predict all PCs (that contain also only XYZ coordinates) from the rest of the dataset (part 2 of the dataset that contains 16 clouds). The obtained results were considered to be prior knowledge to obtain refined semantic segmentation results. Secondly, this prior knowledge was assigned to PCs (XYZ + aerial image) based on its coordinates. The same process of data preparation was followed to prepare all PCs from 16 sets of the dataset. The merged data were then used to train the DL technique, where the fundamental hyperparameters of the original version were changed. Additionally, the basic input tensor was modified into several channels, including X, Y, Z, R, G, B, and classified geometric information. Finally, the trained model was utilized to predict the test data, which were prepared in the same manner as the training data, in order to evaluate the technique's performance.

3.2.4. Baseline Approach

The baseline approach [11] is a point-level fusion approach that directly combines aerial images and PC. It involves running both the RandLaNet and KPConv techniques using the following attributes: X, Y, Z, R, G, B. We compared the baseline approach with the developed scenarios to better understand how these scenarios improved the results of PC-enriched semantic segmentation. The benchmark was made with the baseline approach, which employs the most commonly used fusion method in the literature. The baseline approach includes two parts. The first one is the assignment of radiometric information from images to PC, while the second one is the adoption of both RandLaNet and KPConv to perform semantic segmentation. Figure 22 summarizes the general process followed for the implementation of the baseline approach.

Figure 22. The general workflow of the baseline approach.

To perform the RandLaNet technique, the same methodology of the existing approach [11] was followed with a slight difference. In our case, we used only 16 sets of the SensatUrban dataset to ensure a fair evaluation, similar to the developed scenarios. Additionally, we utilize the average of RGB instead of three separate columns containing the R, G, and B bands. That is, the basic input tensor was modified as follows: X, Y, Z, and average RGB. For the KPConv model, we followed a similar methodology as with RandLaNet, but tailored the input parameters and model configurations. It was crucial to ensure that both techniques were given equal footing for a fair comparison. Hence, we used the same 16 sets from the SensatUrban dataset for KPConv as well.

4. Experiments and Results Analysis

4.1. Implementation

The calculations for the study were carried out using Python programming language version 3.6, with Ubuntu version 20.04.3 as the operating system. Cloud Compare version 2.11.3 was used to calculate geometric properties and average RGB from images. Tensorflow-GPU v 1.14.0 was used as the code framework to implement the RandLaNet algorithm, with CUDA version 11.4 utilized to accelerate deep learning through parallel processing power of GPUs. All experiments were conducted on an NVIDIA GeForce RTX 3090, and a workstation with 256G RAM, a 3.70 GHz processor, and Windows 10 Pro for workstations OS (64-bit) was used for data processing. Furthermore, Scikit-learn, a free Python machine learning library, was employed to implement various processes, where optimized parameters remained unchanged throughout the study.

The RandLaNet technique is publicly available on GitHub at https://github.com/QingyongHu/RandLA-Net (accessed on 10 March 2023). The original version of the code was used to train and test the algorithm. For each scenario, the algorithm was adapted and trained six times using the provided data, and the hyperparameters were kept constant. The Adam optimization algorithm [34] was used for training with an initial learning rate of 0.01, an initial noise parameter of 3.5, and a batch size of 4. The technique was trained for 100 iterations, and all layers were included in the training. Every training process passes through two stages. The first is a forward pass, which deduces the prediction results and compares them with ground truth to generate a loss, while the second is a backward pass, in which the network weights are then updated by stochastic gradient descent. The obtained trained networks were used for the prediction of the blocks selected to test all processes. Consequently, a semantic label was assigned for each cloud point.

For our experimentation with the KPConv technique, which is publicly available on GitHub at https://github.com/HuguesTHOMAS/KPConv-PyTorch (accessed on 1 September 2023), we made specific adjustments to its parameters to optimize memory usage. We set the expected batch size order of magnitude to 10,000. The number of kernel points was designated as 15, and the radius of the input sphere was adjusted to 3.0 for memory efficiency. The size of the first subsampling grid was marked at 0.4, while the convolution radius was established at 2.5. We increased the deformable convolution radius to 5.0 to accommodate the kernel spread. Additionally, each kernel point's area of influence was defined at 1.2, with the behavior of convolutions was set to linear. Lastly, the aggregation function of KPConv was chosen to operate in sum mode.

In order to assess the efficacy of the developed scenarios, five metrics were adopted: precision, recall, F1 score, intersection over union, and confusion matrix. Precision gauges the percentage of points identified as positive in semantic segmentation. Recall evaluates the proportion of true positives in relation to all actual positive instances. F1 score represents the harmonic mean of precision and recall. Intersection over union (IoU) quantifies the extent of overlap between predicted and actual results. Evaluation of these metrics was conducted on Google Colaboratory.

4.2. Results

To highlight the semantic segmentation outcomes of the four processes, this section offers a dual analysis. In the first part, the results obtained using the RandLaNet technique are detailed. In the subsequent section, the results achieved using the KPConv technique are presented to validate and confirm the initial findings. For a comprehensive evaluation, described metrics are employed, along with a qualitative assessment that involves visually comparing predicted (synthetic) and observed (actual) data. Furthermore, we compare the Efficient-PLF approach with certain DL techniques from the literature in our results analysis.

4.2.1. Primary Semantic Segmentation Results Using RandLaNet

(A) Quantitative Assessments

In this subsection, we evaluate the scenarios S1, S2, and S3 with the baseline approach using test set data. The comparisons are reported in Table 7. Since several scenarios were evaluated in this work, the same data splits were used for the RandLaNet algorithm's training, validation, and testing to ensure a fair and consistent evaluation. Four urban scenes (four test sets) were used to evaluate the trained models and did not contribute to the training processes. We can see that all developed scenarios outperform the baseline approach in all evaluation metrics. The experimental results show that S1 delivers the best performance over other scenarios, which was manifested mainly in the higher IoU and highest precision in the obtained results. For example, in scene 1, the IoUs of S1, S2, S3, and the baseline approach were 80%, 77%, 75%, and 63%, respectively. Table 7 displays the achieved semantic segmentation accuracies.

Table 7. Quantitative results for developed scenarios and baseline approach using RandLaNet.

Urban	Processes	F1-Score	Recall	Precision	IoU
Scene 1	Baseline approach	0.71	0.77	0.71	0.63
	S1	0.87	0.87	0.88	0.80
	S2	0.85	0.86	0.85	0.77
	S3	0.83	0.84	0.84	0.75
Scene 2	Baseline approach	0.82	0.86	0.79	0.75
	S1	0.93	0.92	0.94	0.88
	S2	0.92	0.91	0.92	0.86
	S3	0.90	0.90	0.91	0.85
Scene 3	Baseline approach	0.75	0.78	0.74	0.67
	S1	0.86	0.85	0.88	0.79
	S2	0.84	0.83	0.87	0.77
	S3	0.83	0.82	0.86	0.76
Scene 4	Baseline approach	0.61	0.68	0.58	0.50
	S1	0.80	0.78	0.84	0.68
	S2	0.79	0.78	0.82	0.67
	S3	0.70	0.72	0.76	0.57

Based on the results from Table 7, S1 has obvious advantages, but the difference between it and S2 is relatively small. From the results of each metric, we can see that S1 achieved 88/80%, 94/88%, 88/79%, and 84/68% semantic segmentation precision/IoU in the four urban scenes. Compared to the baseline approach, S1 increases the semantic segmentation IoU of each scene by 17%, 13%, 12%, and 18%, respectively. Also, S2 increases the semantic segmentation IoU of each scene by 14%, 11%, 10%, and 17%, respectively. Additionally, S3 increases the semantic segmentation IoU of each scene by 12%, 10%, 9%, and 7%, respectively. The poor precision obtained by the baseline approach could be explained by the lack of prior knowledge from images or PC, which could provide useful information related to urban space. Therefore, it is difficult to obtain accurately diversified objects' semantic segmentation by the direct fusion of PCs and image data. On the other hand, S1 has advantages over both the scenarios with geometric features (S2) and with classified geometrical information (S3). The results obtained by S1 indicated that the integration of prior knowledge from images (image classification) improves the 3D semantic segmentation. It improved the semantic segmentation precision to around 94% in scene 2, for example. Additionally, with the help of prior knowledge from classified images in S1, we achieved about a 17% increase in overall precision compared to the baseline approach. Therefore, based on the evaluation metrics, we can conclude that the overall performance of S1 shows promising potential.

Having discussed the general evaluation metrics for semantic segmentation outcomes, Table 8 provides a comprehensive analysis of the performance for each semantic class obtained from the different scenarios and the baseline approach.

Table 8. Semantic segmentation performance of the baseline approach and developed scenarios (urban scene 2).

Semantic Segmentation Performance		Baseline Approach	S1	S2	S3
Ground	Precision	0.746	0.952	0.917	0.907
	Recall	0.990	0.921	0.927	0.917
	F1-score	0.851	0.936	0.922	0.912
High Vegetation	Precision	0.937	0.997	0.995	0.995
	Recall	0.998	0.992	0.995	0.993
	F1-score	0.967	0.994	0.995	0.994
Buildings	Precision	0.985	0.982	0.987	0.976
	Recall	0.909	0.955	0.938	0.951
	F1-score	0.946	0.968	0.962	0.963
Walls	Precision	0.790	0.769	0.766	0.725
	Recall	0.677	0.690	0.776	0.639
	F1-score	0.729	0.727	0.771	0.680
Parking	Precision	0.605	0.428	0.417	0.408
	Recall	0.123	0.757	0.727	0.722
	F1-score	0.205	0.547	0.530	0.522
Traffic Roads	Precision	0.000	0.840	0.828	0.803
	Recall	0.000	0.726	0.629	0.498
	F1-score	0.000	0.779	0.715	0.614
Street Furniture	Precision	0.325	0.250	0.259	0.230
	Recall	0.518	0.828	0.779	0.698
	F1-score	0.399	0.384	0.389	0.346
Cars	Precision	0.929	0.909	0.904	0.862
	Recall	0.721	0.937	0.956	0.935
	F1-score	0.812	0.922	0.929	0.897
Footpath	Precision	0.000	0.655	0.601	0.530
	Recall	0.000	0.664	0.557	0.530
	F1-score	0.000	0.660	0.578	0.530

After detailed analysis of the class-specific metrics, clear variations emerged across the scenarios. Using the F1-score as our main evaluation measure, S1 excelled in the “Ground” class with an F1-score of 0.94. For “High Vegetation”, S1, S2, and S3 all reached a similar high precision. In the “Buildings” category, S1 slightly led with an F1-score of 0.97, while for “Walls”, S2 was the best at 0.77. S1 was consistently ahead in “Parking” and “Traffic Roads” with scores of 0.55 and 0.78, respectively. The “Street Furniture” scores were modest but saw S1 and S2 closely matched and outperforming both the baseline and S3. In the “Cars”

class, S2 was the leader with 0.93, and for “Footpath”, S1 was the top performer with 0.66. Overall, while S1 showed strong results across multiple classes, S2 was more effective in specific categories like “Walls” and “Cars”.

The results obtained with different developed scenarios were studied in detail by computing a percentage-based confusion matrix using ground truth data. “This percentage-based analysis provides an idea about the percentage of consistent and non-consistent points” [18]. The percentage-based confusion matrix obtained by all scenarios for scene 1 is depicted in Figure 23. The corresponding confusion matrices for the other urban scenes (2, 3, and 4) can be found in Figures 1–3 on the following GitHub link: https://github.com/ZouhairBALLOUCH/Supplementary_Results_Article.git (accessed on 1 December 2023). The confusion matrices show that the developed scenarios significantly outperform the baseline approach and reveal the limitations of using only direct image and PC fusion for complex urban scene segmentation.

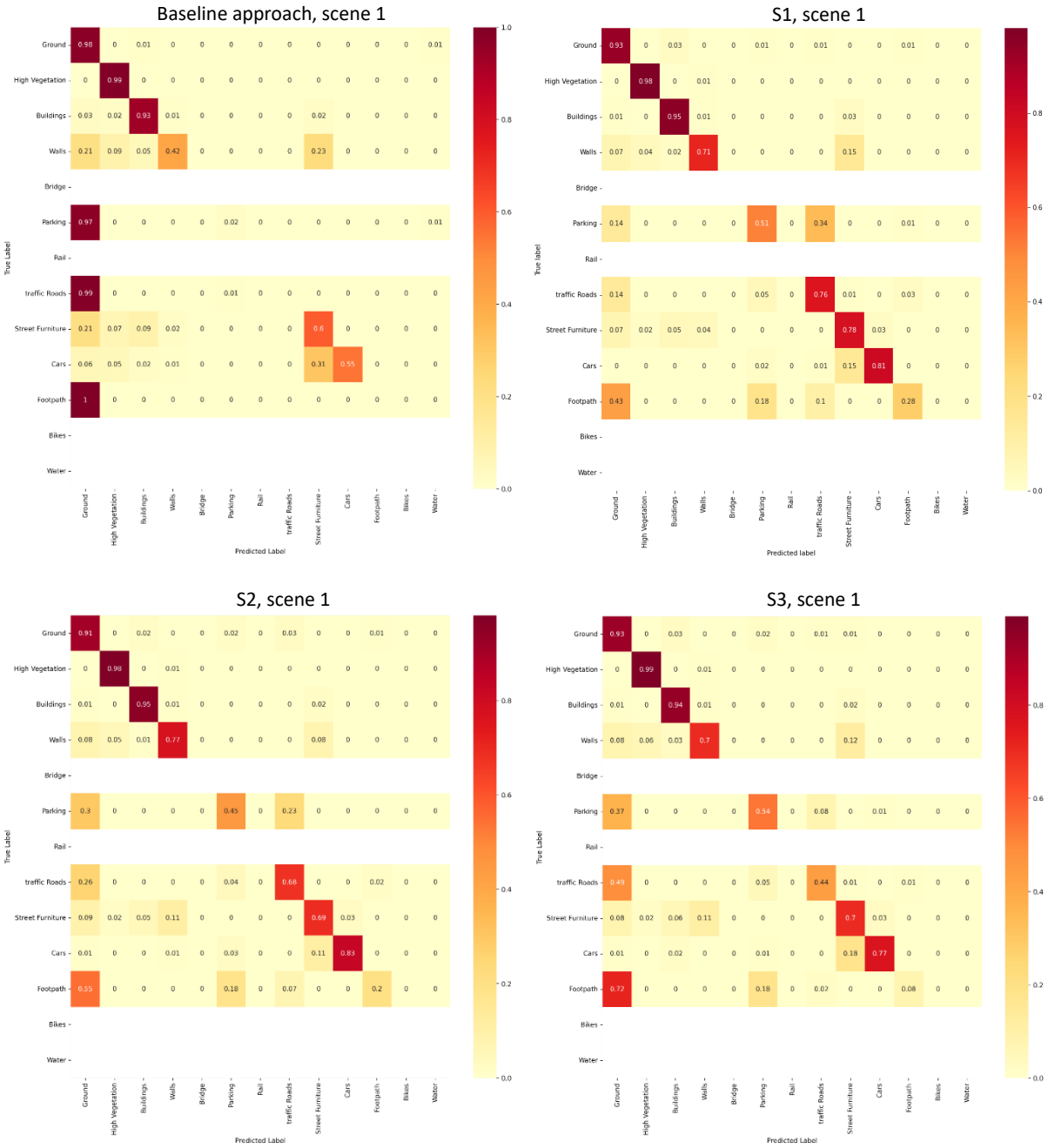


Figure 23. Normalized confusion matrix for proposed scenarios and the baseline approach in an urban scene using the RandLaNet technique.

The following are the detailed results of each semantic class independently:
Firstly, Ground and High Vegetation classes were successfully extracted in all scenes with all evaluated processes. This was due to their geometric and radiometric characteristics which are easy to recognize. That is, they are easily distinguished from other classes. This means that only the PCs and the aerial images fused in the baseline approach are sufficient to correctly segment the two classes. Secondly, the Building class was extracted accurately by S1, but the difference between it and other developed scenarios is relatively small. However,

despite its performance, a slight confusion was observed between this class and the Street Furniture object. Thirdly, by observing the four scenes, we can see that S1 has a good performance on the PC scenes containing Rail, Traffic Roads, Street Furniture, Footpath, and Parking objects. The five semantic classes were extracted precisely by this scenario, except for the Footpath class, and the precision of it was low. Additionally, the percentage of consistent points obtained by it surpassed all other developed scenarios and the baseline approach. The baseline approach failed to label these classes. For example, in scene 4, S1 increases the percentage of consistent of each class by 12% (Parking), 2% (Rail), 7% (Traffic Roads), 13% (Street Furniture), and 7% (Footpath), respectively, compared to S2. S1 increases the percentage of consistency of each class by 2% (Parking), 12% (Rail), 47% (Traffic Roads), 8% (Street Furniture), and 9% (Footpath), respectively, compared to S3. However, these semantic classes are often confused with others with similar characteristics. We can list the confusion between the Parking class with the Ground and Traffic Roads classes, as well as the confusion between the Rail with Street Furniture and Water objects. In addition to the confusion between the Traffic Roads class with Ground and Parking geo-objects, there is also confusion with Bridge class in scene 4. Fourthly, by observing the four scenes, we can see that S2 had good performance on the PC scenes containing Cars, Walls, and Bridge objects. The obtained results in these classes indicate that S2 generally performed better than the other scenarios. If we still take the example of scene 4, S2 increases the percentage of consistency of each class by 2% (Cars), 14% (Walls), 12% (Bridge), respectively, compared to S1. Additionally, it increases the percentage of consistency of each class by 5% (Cars), 4% (Walls), and 62% (Bridge), respectively, compared to S3. In addition, S2 increases the percentage of consistency of each class by 22% (Cars), and 42% (Walls), respectively, compared to the baseline approach. The Bridge class was not completely detected by the baseline approach. However, these semantic classes are often confused with other objects with similar characteristics. We can cite the confusion between the classes of Cars and Street Furniture in scenes 1 and 4 in addition to the confusion between the class Wall and Street Furniture. Thus, we noticed a slight confusion between the Wall object and the class Buildings (scene 4) and Ground (scene 1). Finally, we observed a confusion between the class Bridge and building in scene 4. Fifth, S1 was the only one to accurately detect the Water class, as reflected in the confusion matrix results. The Water class was mistaken for the Wall in S2 and for the Ground in S3. Finally, the Bike class was not detected by all scenarios due to the very-low percentage of Bike samples in the dataset.

(B) Qualitative Assessments

In addition to the quantitative evaluation, a qualitative analysis was performed by visualizing the semantic segmentation results in detail for the test data set. Figure 24 demonstrates the visual comparison of the predicted results obtained by the four processes with the corresponding ground truth. To show the semantic segmentation effect more intuitively, Figure 25 demonstrates some selected regions from 3D semantic segmentation maps of all evaluated processes. It can be observed from the figures that the results of S1 are closest to the ground truth. Additionally, its results are more accurate and coherent compared to the others, and classes were extracted precisely with clear boundaries.

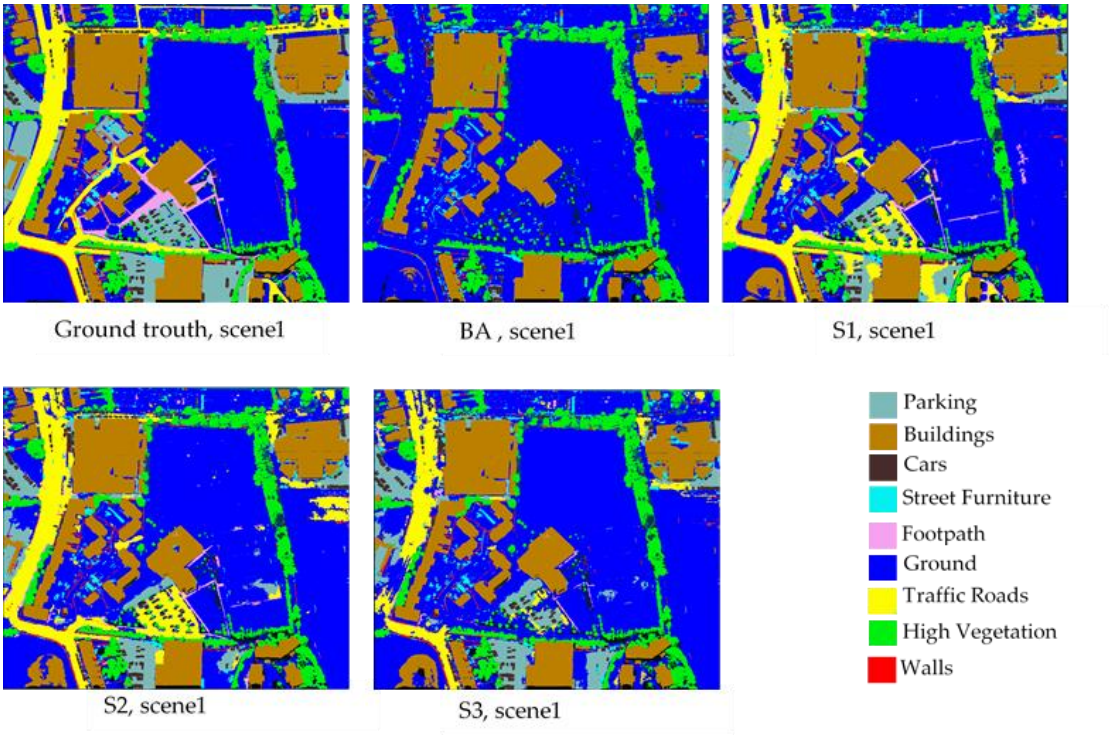


Figure 24. The 3D semantic segmentation results of the baseline and the three developed scenarios. Ground truth is also displayed.

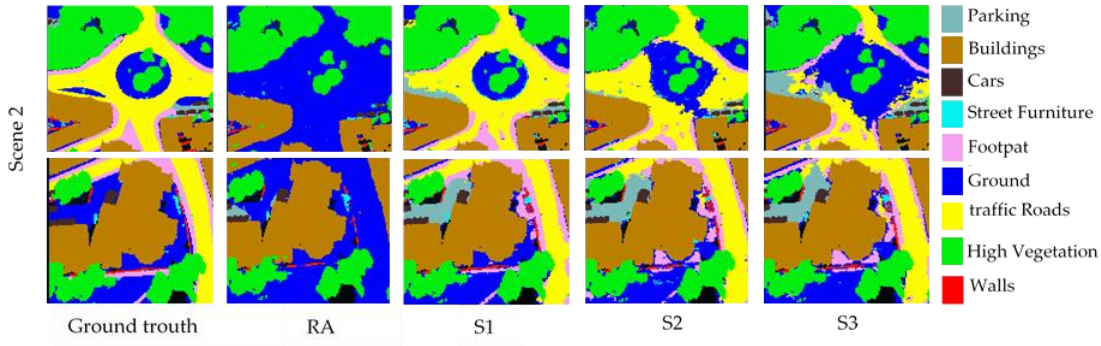


Figure 25. Selected regions from 3D semantic segmentation maps of the all evaluated processes.

The qualitative results of each class are further explained in the following paragraphs.
At first, the semantic segmentation results indicate that, in general, the Ground and High Vegetation classes were effectively segmented by all four processes. However, we observed that the baseline approach fails to label Rail, Traffic Roads, Street Furniture, and Parking classes effectively. These results were confirmed by the confusion matrix outcomes; for example, see the results for scene 4 at (https://github.com/ZouhairBALLOUCH/Supplementary_Results_Article.git, accessed on 1 December 2023).

Furthermore, as observed in the quantitative results, S1 shows better performance on these classes by producing very few miss-segmented points compared to others scenarios. Its errors in these classes were lower than those delivered by other scenarios for these semantic classes. On the other hand, in the cases of S2, S3, and the baseline approach, several Parking class points were miss-segmented as Ground. This was due to the similarity in their geometric and radiometric properties. Moreover, the three scenarios all confused certain points of Traffic Roads as a Ground class. The Street Furniture class shares a similar color to the Building and Wall classes; in fact, as shown in Figure 24, part of the Street Furniture was labeled as a Building in the semantic segmentation results of S2, S3, and the baseline approach. Finally, the Rail object was not detected by the baseline approach; additionally, S2 and S3 miss-classified it as Water and Street Furniture. Concerning the Building class, the visual evaluation shows that the different developed scenarios correctly extracted this object compared to the baseline approach. In the case of the baseline scenario, we observed a slight confusion between the Building class and those of Ground and High Vegetation. In addition, S1 errors were slightly lower than those delivered by S2 and S3 for the Building class.
Visually, we can observe in Figure 24 and Figure 25 that the Footpath object was difficult to recognize. S1, S2, and baseline scenario failed to label this class correctly, while S1 achieved an acceptable performance on it (scene 2). Concerning the Cars, Wall, and Bridge objects, thanks to the suitable geometric features calculated from PCs in S2, S2 errors were lower than those delivered by the other scenarios. The results indicated that the Bridge class was labeled as Buildings with the baseline approach. Additionally, a part of this class was labeled as Buildings in the segmentation results of S1 and S3. Moreover, as shown in Figure 24, various Car class points were miss-segmented as Street Furniture, especially in scene 4 (see confusion matrix results). In addition, the Wall was confused with several classes, mainly Street Furniture and Building geo-objects.
To conclude, based on visual comparison, the semantic segmentation of developed scenarios showed a very complementary effect compared to the baseline approach. The results also indicated that S1 generally outperformed S2 and S3. Particularly, S2 improved the semantic segmentation results of some classes (Wall, Cars, Bridge) more than the other scenarios.

4.2.2. Results Confirmation with KPConv

Following previous evaluations using the RandLaNet technique, further testing was conducted using the KPConv technique (Table 9) to validate and potentially reinforce the findings obtained by RandLaNet. The results presented in the Table 9 below were derived from the urban scene 2, which corresponds to the same urban scene studied in the initial tests conducted with RandLaNet (refer to Table 8). Upon reviewing the outcomes across four urban scenes by RandLaNet, Scenario 3's performance was consistently average when set against scenarios 1 and 2. Consequently, the discussion was primarily centered on the performances of scenarios 1 and 2.

Table 9. Results of semantic segmentation achieved using KPConv.

Semantic Segmentation Performance		BA	S1	S2
Ground	Precision	0.762	0.880	0.767
	Recall	0.946	0.931	0.949
	F1-score	0.844	0.905	0.849
High Vegetation	Precision	0.961	0.989	0.948
	Recall	0.889	0.986	0.987
	F1-score	0.924	0.987	0.967
Buildings	Precision	0.766	0.882	0.871
	Recall	0.936	0.975	0.926
	F1-score	0.843	0.926	0.903
Walls	Precision	0.456	0.540	0.760
	Recall	0.008	0.043	0.148
	F1-score	0.016	0.080	0.257
Parking	Precision	0.373	0.534	0.462
	Recall	0.280	0.357	0.352
	F1-score	0.320	0.428	0.400
Traffic Roads	Precision	0.475	0.727	0.558
	Recall	0.025	0.691	0.014
	F1-score	0.048	0.709	0.028
Street Furniture	Precision	0.334	0.344	0.606
	Recall	0.012	0.074	0.055
	F1-score	0.023	0.122	0.093
Cars	Precision	0.735	0.761	0.751
	Recall	0.399	0.719	0.634
	F1-score	0.517	0.739	0.681
Footpath	Precision	0.512	0.574	0.584
	Recall	0.028	0.208	0.023
	F1-score	0.053	0.305	0.043

After evaluating the semantic segmentation results obtained by the KPConv model, we found that the results matched the initial observations made by the RandLaNet algorithm. For the “Ground” class, while S1 has an F1-score of 0.90, it is closely followed by both the baseline approach and S2, each around 0.85. The “High Vegetation” category reaffirms previous conclusions with S1 standing out with an F1-score of 0.99, though S2’s 0.97 remains competitive. The “Buildings” semantic class witnesses S1 leading at 0.93, with S2 closely trailing. In “Walls”, despite modest scores overall, S2 shows a relative advantage with 0.26. The “Parking” results show improvements across scenarios compared to the baseline, with S1 achieving the highest score of 0.43. For “Traffic Roads”, S1 dominates with an F1-score of 0.71, a notable improvement over the other scenarios. “Street Furniture” and “Footpath” classes have modest F1-scores, yet some scenarios, especially S1, display improvements over the baseline approach. Finally, in the “Cars” category, S1 and S2 perform similarly well, with S1 slightly ahead at 0.74. In summation, the KPConv model’s results not only confirm

the previous findings but also highlight the potential of scenarios in semantic segmentation performance. For an overview of the general metrics achieved by the KPConv technique across two urban scenes, the results are available in Table 2 on this link: https://github.com/ZouhairBALLOUCH/Supplementary_Results_Article.git, accessed on December 1, 2023.

Following these insights, an in-depth analysis using percentage-based confusion matrices was carried out, showcasing advancements in the accuracy of semantic segmentation, especially for complex urban objects, as depicted in Figure 26.

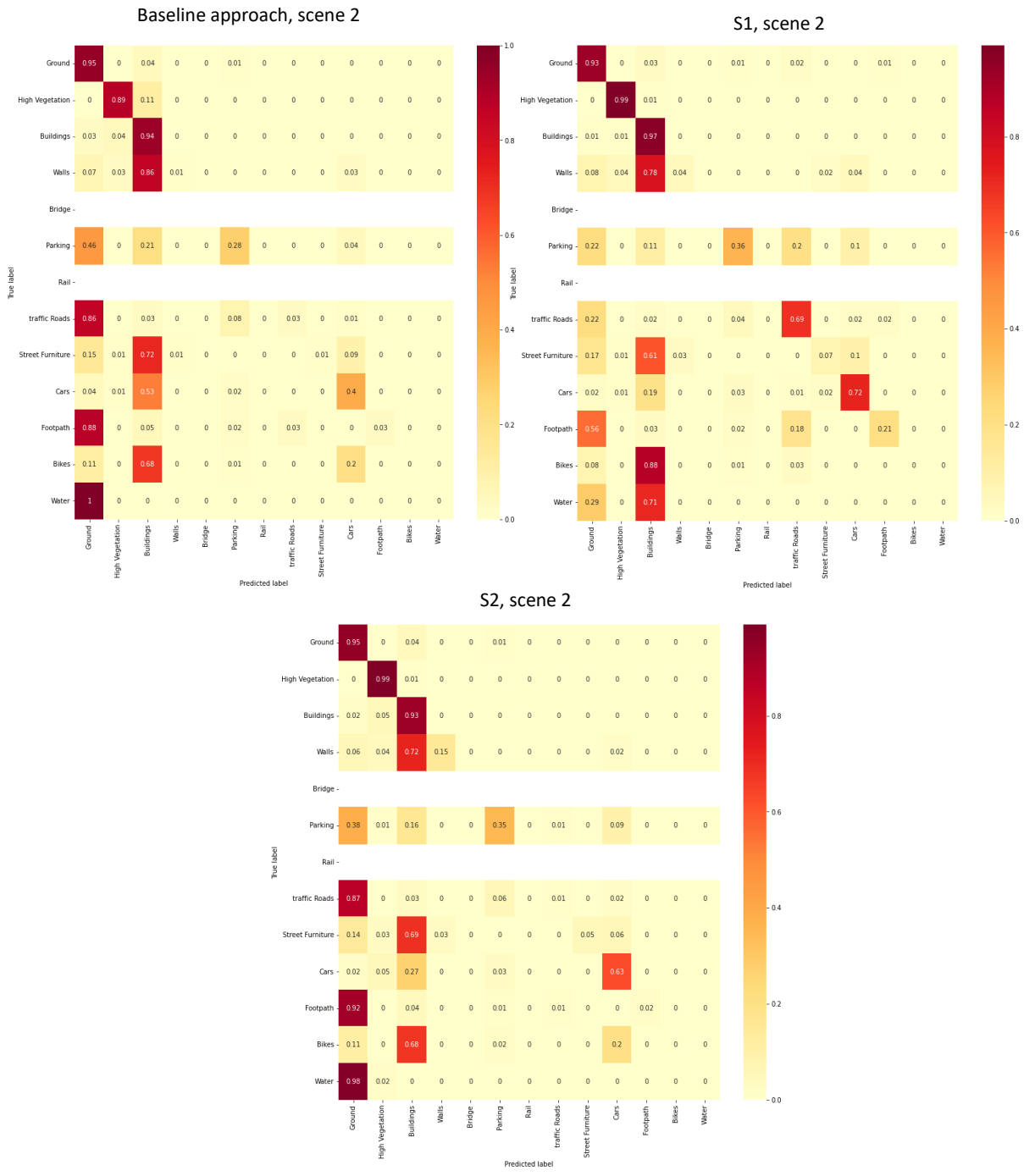


Figure 26. Normalized confusion matrix for the proposed scenarios and the baseline approach in an urban scene using the KPConv technique.

4.2.3. Comparison of Efficient-PLF Approach with DL Techniques from the Literature

The goal of this study does not concentrate on a particular type of DL technique but rather on finding an effective approach for selecting pertinent features and an efficient fusion scenario applicable to any DL technique. Despite using only a subset of the dataset (16 PC), RandLaNet adopted to our Efficient-PLF approach was compared with some DL techniques from the literature [11]. Note that the test data used to assess these DL techniques (PointNet [35], PointNet++ [36], TagentConv [37], and SPGraph [6]) differ from our test data (but data are from the same dataset; only the test samples differ). This difference is justified by the fact that the data they employed lack labels (ground truth) and are not openly accessible. The results can be found in Table 10.

Table 10. RandLaNet adopted to our Efficient-PLF approach vs. DL Techniques [11]: Per-class IoU (%) Comparison.

	Ground	High Vegetation	Buildings	Walls	Parking	Traffic Roads	Street Furniture	Cars	Footpath
PointNet [35]	67.96	89.52	80.05	0.00	3.95	31.55	0.00	35.14	0.00
PointNet++ [36]	72.46	94.24	84.77	2.72	25.79	31.54	11.42	38.84	7.12
TagentConv [37]	71.54	91.38	75.90	35.22	45.34	26.69	19.24	67.58	0.01
SPGraph [6]	69.93	94.55	88.87	32.83	15.77	30.63	22.96	56.42	0.54
RandLaNet adopted to our Efficient-PLF approach	85.42	97.33	90.81	49.22	42.06	56.00	35.00	77.97	19.86

4.3. Discussion

This work develops three prior-level fusion scenarios based on DL for 3D semantic segmentation. To summarize the performance of different developed scenarios, the results were compared to a baseline approach using both qualitative and quantitative assessments. Tables 7– 9 show that the semantic segmentation of the developed scenarios, especially S1, was significantly better than S2, S3, and the baseline approach across all urban scenes. To assess each semantic class individually, confusion matrices were computed using both the RandLaNet and KPConv techniques. By observing their results, it can be seen that the developed fusion scenarios achieved the best semantic segmentation compared to the baseline approach. Despite the good results of the baseline approach obtained in some classes such as Ground, it failed to label completely some others namely Bridge, Traffic Roads, and Footpath classes. Additionally, its results in Parking classes are not acceptable. Thus, it is quite difficult to detect these objects using only PCs and aerial images. As a first conclusion of this work, we point out that the direct fusion of PCs and aerial images is not sufficient for the semantic segmentation of complex scenes with a diversity of objects. Compared to the baseline approach, S2, and S3, we can see that S1 has the best

performance on the PC scenes containing Rail, Traffic Roads, Street Furniture, Footpath, and Parking objects. Despite the choice of the most appropriate geometric properties in S2 and the injection of classified geometric information in S3, these two scenarios did not succeed in obtaining the high accuracies that were obtained by S1. The prior knowledge selected in these scenarios was not enough to further distinguish these types of terrains. This could be due to the geometric similarity in these classes. The confusion matrices calculated have confirmed this situation. We can conclude here that the preliminary results of image classification guided the model to know these different classes and distinguish them precisely. On the other hand, the second scenario, S2, performed well on the Cars, Wall, and Bridge objects. It demonstrated the best precisions compared to S1, S3, and the baseline approach. The low accuracy obtained by S1 compared to those obtained by S2 may be due to the similarity in the radiometric information of these geo-objects. Nevertheless, the description of local geometric properties by selected geometric features has facilitated the distinction of these three classes in S2. The visual results confirm this situation. Figure 24 depicts the results of the four fusion scenarios. Overall, the developed scenarios outperformed the baseline in terms of visual quality and reduced semantic segmentation errors. Specifically, S1 closely mirrored the ground truth and outshined S2, S3, and the baseline for many classes. However, for geo-objects like Walls, Cars, and Bridges, S2 excelled, enhancing visual quality compared to the other scenarios. Additionally, S1 allows for the utilization of classified images from various sources, including drone and satellite images, and can be processed by different neural networks of image classification, making it a practical option. S1 is also not highly data-intensive, as satisfactory results were obtained by training the model with only a portion of the dataset, which reduces the financial resources and hardware required since it relies solely on aerial images and PCs. However, this scenario could be somewhat long, and classification errors in the images could negatively impact the 3D semantic segmentation results. Although S1 has several advantages, the difference between S1 and S2 is relatively small. Specifically, S2 excels at segmenting Walls, Cars, and Bridges, surpassing S1 and S3 based on both qualitative and quantitative findings. In addition, S2 is easier to handle than the other scenarios and does not require any prior knowledge. However, this scenario works best for classes with distinct geometries, but the issue with distinguishing geo-objects with similar geometrical features remains. Additionally, S2 necessitates the selection of features that have a positive impact on semantic segmentation. In regard to S3, it is better suited for geometrically distinct geo-objects. The uniqueness of this scenario lies in its direct use of semantic knowledge from geometric information, which enhances the distinction of such objects. However, a pre-classification step is required, which makes the process somewhat long. Moreover, the accuracy of its 3D semantic segmentation is relatively low, and classification errors in geometric information could have a negative impact on semantic segmentation outcomes. In conclusion, considering the good qualitative and quantitative results in all classes and its superior performance compared to other scenarios, S1 is the Efficient-PLF approach for semantic segmentation of PCs acquired on a large scale. In addition, we suggest considering S2 due to its high performance on certain semantic classes and its ease of handling. Finally, it should be noted that this research work presents certain limitations including the usage of only 16 sets of the SensatUrban dataset, which may not be sufficient to achieve the maximum accuracies of different scenarios. In addition, the developed fusion scenarios should be tested on other datasets that contain other semantic classes. As a perspective, we suggest investigating the derived Efficient-PLF approach in various urban contexts by choosing other urban objects and by also considering other dataset

types, especially, the terrestrial PCs. The goal is to evaluate the precisions and errors of the selected Efficient-PLF approach when confronted with other urban environments.

5. Conclusions

This article introduces a new prior-level fusion approach for semantic segmentation based on an in-depth evaluation of three scenarios, which involve fusing aerial images, prior knowledge, and PCs into the DL techniques' learning pipeline. Three proposed scenarios were evaluated based on their qualitative and quantitative results to identify the one that successfully extracted the maximum urban assets details. The derived scenario was named the "Efficient-PLF approach". Additionally, another contribution of this work was adopting advanced DL structures and tailoring their parameters to match the specific requirements of our research. Since S1 exhibits good scores in all classes and its performances surpass the other scenarios, we can conclude that S1 is the Efficient-PLF approach for the semantic segmentation of large-scale PC. Therefore, the preference for S1 is motivated by the accuracy of its results and the quality of its visual predictions. We also recommended S2 because of its high performance on some semantic classes and the simplicity of its processing. The experiments show that the derived Efficient-PLF approach can improve the knowledge of the DL techniques. It allows for good metrics, particularly for classes that are difficult to detect using the original DL architecture without prior knowledge. Additionally, it succeeds in reducing the confusion between different semantic classes. Furthermore, the Efficient-PLF approach can potentially be adapted for any 3D semantic segmentation DL techniques. So, we suggest investigating the semantic segmentation Efficient-PLF approach in other complex urban environments to evaluate its efficiency and limits in different urban contexts. Additionally, we recommend experimenting with adapting other DL techniques to the Efficient-PLF approach. Furthermore, regarding the image classification part, we propose testing the use of classified images from alternative sensors such as satellite imagery and drones.

Enrichment of semantic point clouds through classification of high-resolution spatial images

The current performance of artificial intelligence techniques in high-resolution image classification is notable [38–40] . These techniques excel in terms of precision, speed, and the visual quality of the results. Among these techniques, we can cite the "Segment Anything Model (SAM)" [41] ; a novel image segmentation model that has revolutionized the field of image processing [42,43]. These advancements present a significant opportunity to enhance semantic point clouds even further. Therefore, a new methodology based on the SAM technique was proposed. This methodology exploits high-resolution images corresponding to the point clouds. It is divided into two main steps: (1) segmentation of the images corresponding to the point clouds, and (2) projection of the segmentation results from the images onto the point clouds.

For the implementation, a series of high-resolution aerial images were extracted from the SensatUrban dataset [44] and utilized. We developed a Python code, which is available for open access at the following link: [link]. This code is based on Segment-Geospatial (samgeo) [45]. It's an open-source Python package designed to simplify the process of segmenting geospatial data with SAM. The developed code performs object extraction (cars, vegetation, etc.) from images initially. Then, it assigns the extracted objects to point clouds, resulting in the generation of classified 3D point clouds. In other words, it assigns the corresponding semantic label from the segmented image to each point in the cloud. The developed code has demonstrated good quality results, as shown in Figure 27. This image-based segmentation methodology can be useful in cases where some objects are not well classified using a LiDAR-based approach. Specifically, in certain objects like vegetation and cars, where this image-based methodology shows higher precision. It can also be used to extract an additional semantic class that is not present in the LiDAR dataset used. For example, in the "SensatUrban" LiDAR dataset used in this work, which contains 13 classes, this methodology can be employed to extract a class that is not included within these 13 classes. An instance of this is the grass class. This is illustrated in Figure 28. However, this methodology cannot replace the LiDAR approach but can complement it. This is due to some of its limitations. One of these is the confusion between certain classes with similar spectral information, such as roads and buildings.

Figure 27. Results of car (A) and building (B) detection.

Figure 28. Detection of grass areas in a neighborhood of the city of Liège

References

1. Shahat, E.; Hyun, C.T.; Yeom, C. City Digital Twin Potentials: A Review and Research Agenda. *Sustainability* **2021**, *13*, 3386. https://doi.org/10.3390/su13063386.
2. Ruohomäki, T.; Airaksinen, E.; Huuska, P.; Kesäniemi, O.; Martikka, M.; Suomisto, J. Smart City Platform Enabling Digital Twin. In Proceedings of the 2018 International Conference on Intelligent Systems (IS), Funchal, Portugal, 25–27 September 2018; pp. 155–161.
3. White, G.; Zink, A.; Codecá, L.; Clarke, S. A Digital Twin Smart City for Citizen Feedback. *Cities* **2021**, *110*, 103064. https://doi.org/10.1016/j.cities.2020.103064.
4. Zhang, J.; Zhao, X.; Chen, Z.; Lu, Z. A Review of Deep Learning-Based Semantic Segmentation for Point Cloud. *IEEE Access* **2019**, *7*, 179118–179133. https://doi.org/10.1109/ACCESS.2019.2958671.
5. 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.
6. 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.
7. Zhang, R.; Wu, Y.; Jin, W.; Meng, X. Deep-Learning-Based Point Cloud Semantic Segmentation: A Survey. *Electronics* **2023**, *12*, 3642. https://doi.org/10.3390/electronics12173642.
8. Thomas, H.; Qi, C.R.; Deschaud, J.-E.; Marcotegui, B.; Goulette, F.; Guibas, L.J. KPConv: Flexible and Deformable Convolution for Point Clouds. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6411–6420.
9. 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.
10. 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. https://doi.org/10.5194/isprs-archives-XLII-2-W13-1899-2019.
11. 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.
12. Gao, W.; Nan, L.; Boom, B.; Ledoux, H. SUM: A Benchmark Dataset of Semantic Urban Meshes. *ISPRS J. Photogramm. Remote Sens.* **2021**, *179*, 108–120. https://doi.org/10.1016/j.isprsjprs.2021.07.008.
13. 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:1704.03847.
14. 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; Toledo Santos, E., Scheer, S., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 1086–1098.
15. Oh, S.-I.; Kang, H.-B. Object Detection and Classification by Decision-Level Fusion for Intelligent Vehicle Systems. *Sensors* **2017**, *17*, 207. https://doi.org/10.3390/s17010207.
16. 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. https://doi.org/10.3390/rs13234928.
17. 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. https://doi.org/10.1016/j.isprsjprs.2018.04.022.
18. 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.
19. 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. https://doi.org/10.3390/rs11242961.
20. Ye, C.; Pan, H.; Yu, X.; Gao, H. A Spatially Enhanced Network with Camera-Lidar Fusion for 3D Semantic Segmentation. *Neurocomputing* **2022**, *484*, 59–66. https://doi.org/10.1016/j.neucom.2020.12.135.
21. 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.* **2016**, *8*, 3. https://doi.org/10.3390/rs8010003.
22. Mirzapour, F.; Ghassemian, H. Improving Hyperspectral Image Classification by Combining Spectral, Texture, and Shape Features. *Int. J. Remote Sens.* **2015**, *36*, 1070–1096. https://doi.org/10.1080/01431161.2015.1007251.
23. Bai, X.; Liu, C.; Ren, P.; Zhou, J.; Zhao, H.; Su, Y. Object Classification via Feature Fusion Based Marginalized Kernels. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 8–12. https://doi.org/10.1109/LGRS.2014.2322953.
24. Zhang, Y.; Chi, M. Mask-R-FCN: A Deep Fusion Network for Semantic Segmentation. *IEEE Access* **2020**, *8*, 155753–155765. https://doi.org/10.1109/ACCESS.2020.3012701.
25. Tabib Mahmoudi, F.; Samadzadegan, F.; Reinartz, P. Object Recognition Based on the Context Aware Decision-Level Fusion in Multiviews Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 12–22. https://doi.org/10.1109/JSTARS.2014.2362103.
26. Zhang, R.; Candra, S.A.; Vetter, K.; Zakhor, A. Sensor Fusion for Semantic Segmentation of Urban Scenes. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 1850–1857.
27. 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. https://doi.org/10.1109/TPAMI.2020.3005434.
28. Grzeczkowicz, G.; Vallet, B. Semantic Segmentation of Urban Textured Meshes through Point Sampling. *arXiv* **2023**, arXiv:2302.10635.
29. Li, W.; Zhan, L.; Min, W.; Zou, Y.; Huang, Z.; Wen, C. Semantic Segmentation of Point Cloud With Novel Neural Radiation Field Convolution. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. https://doi.org/10.1109/LGRS.2023.3307421.
30. Lin, Y.; Vosselman, G.; Cao, Y.; Yang, M.Y. Local and Global Encoder Network for Semantic Segmentation of Airborne Laser Scanning Point Clouds. *ISPRS J. Photogramm. Remote Sens.* **2021**, *176*, 151–168. https://doi.org/10.1016/j.isprsjprs.2021.04.016.

31. Song, H.; Jo, K.; Cho, J.; Son, Y.; Kim, C.; Han, K. A Training Dataset for Semantic Segmentation of Urban Point Cloud Map for Intelligent Vehicles. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 159–170. https://doi.org/10.1016/j.isprsjprs.2022.02.007.
32. Atik, M.E.; Duran, Z.; Seker, D.Z. Machine Learning-Based Supervised Classification of Point Clouds Using Multiscale Geometric Features. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 187. https://doi.org/10.3390/ijgi10030187.
33. Özdemir, E.; Remondino, F. Classification of aerial point clouds with deep learning. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W13*, 103–110. https://doi.org/10.5194/isprs-archives-XLII-2-W13-103-2019.
34. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
35. Qi, C.R.; Su, H.; Mo, K.; 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), Honolulu, HI, USA, 21–26 July 2017; pp. 652–660.
36. 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*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
37. 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, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3887–3896.

38. Hikuwai, M.V.; Patorniti, N.; Vieira, A.S.; Frangioudakis Khatib, G.; Stewart, R.A. Artificial Intelligence for the Detection of Asbestos Cement Roofing: An Investigation of Multi-Spectral Satellite Imagery and High-Resolution Aerial Imagery. Sustainability 2023, 15, 4276, doi:10.3390/su15054276.

39. Guo, W.; Yang, W.; Zhang, H.; Hua, G. Geospatial Object Detection in High Resolution Satellite Images Based on Multi-Scale Convolutional Neural Network. Remote Sensing 2018, 10, 131, doi:10.3390/rs10010131.

40. Wang, L.; Li, R.; Zhang, C.; Fang, S.; Duan, C.; Meng, X.; Atkinson, P.M. UNetFormer: A UNet-like Transformer for Efficient Semantic Segmentation of Remote Sensing Urban Scene Imagery. ISPRS Journal of Photogrammetry and Remote Sensing 2022, 190, 196–214, doi:10.1016/j.isprsjprs.2022.06.008.

41. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.-Y.; et al. Segment Anything.; 2023; pp. 4015–4026.

42. Ji, W.; Li, J.; Bi, Q.; Liu, T.; Li, W.; Cheng, L. Segment Anything Is Not Always Perfect: An Investigation of SAM on Different Real-World Applications. Mach. Intell. Res. 2024, doi:10.1007/s11633-023-1385-0.

43. Liu, S.; Zeng, Z.; Ren, T.; Li, F.; Zhang, H.; Yang, J.; Li, C.; Yang, J.; Su, H.; Zhu, J.; et al. Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection 2023.

44. 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); IEEE: Nashville, TN, USA, June 2021; pp. 4975–4985.

45. Wu, Q.; Osco, L.P. Samgeo: A Python Package for Segmenting Geospatial Datawith the Segment Anything Model (SAM). JOSS 2023, 8, 5663, doi:10.21105/joss.05663.

Chapter 4

Exploiting enriched 3D semantic point clouds and generated 3D models for creating urban Digital Twins-Case Study: Liege city

PREFACE

Cities worldwide are moving towards the implementation of Urban Digital Twins (UDTs). This innovative paradigm represents a new trend for city planning and management, enhancing three-dimensional city modeling and simulation. UDTs build a collaborative platform to help addressing city challenges, fostering everyday services, and improving the living conditions of the residents. In this context, this chapter explores how enriched semantic 3D point clouds can efficiently address city simulations and analysis without performing 3D modeling (section A) and, when 3D models are required, how to reconstruct them from semantic points clouds to develop UDTs (section B).

Section A: Enriched semantic 3D point clouds: An alternative to 3D City models for Digital Twin for Cities?

This section discusses a new reflection that argues on directly integrating the results of semantic segmentation to create the skeleton of the DTs and uses enriched semantically segmented point clouds to perform targeted simulations without generating 3D models. The paper discusses to what extent enriched semantic 3D point clouds can replace semantic 3D city models in the DTs scope. Ultimately, this research aims to reduce the cost and complexity of 3D modeling to fit some DTs requirements and address its specific needs. New perspectives are set to tackle the challenges of using semantic 3D point clouds to implement DTs for cities.

Section B : Towards a Digital Twin of Liege : The Core 3D Model based on Semantic Segmentation and Automated Modeling of LiDAR Point Clouds

This section presents fully automatic procedures to generate 3D city models, which are an inherent component of urban digital twins. The objective is to develop comprehensive processing workflows that automatically produce these 3D city models. This generation is done for each urban object since the way to create buildings differs from creating trees, roads, etc. The challenge is to extract maximum urban details (buildings, roads, vegetation, cars, etc.) accurately and model urban 3D objects accordingly. To achieve this, we explore the existing LiDAR data from the Walloon region along with corresponding orthophotos. We have created processing procedures enabling the extraction of semantic classes (roads, vegetation, cars, etc.). This extraction is achieved through the classification of point clouds, based on an automatic artificial intelligence approach. We fuse multiple data sources (LiDAR and spectral datasets) to obtain accurate classification results in order to create a detailed and complete city model. For the implementation, we chose the Outremeuse neighborhood as a study area. The developed pipeline allows the extraction of urban objects and the

creation of their 3D city models. Beyond visualization purposes, these models can be used in further simulations and urban analysis.

Section B adds a part to the publication that presents the different processes developed for modeling objects that were not previously addressed. These objects include ground, cars, walls, and bridges. These were extracted through the detailed semantic segmentation phase described earlier in Section B.

A.SEMANTIC 3D POINT CLOUD: AN ALTERNATIVE TO 3D CITY MODEL FOR DIGITAL TWIN APPLICATIONS.

Based on paper: Zouhair Ballouch And Imane Jeddoub **(equal contribution)** , Rafika Hajji, Roland Billen. 2023. "Enriched semantic 3D point clouds: An alternative to 3D city models for Digital Twin for Cities? ". 3D GeoInfo Conference.

Digital Twins (DTs) for cities are emerging as a novel tool for urban planning and management. They enhance the 3D modeling and simulation of urban environments. Despite advancements in this field, current literature predominantly focuses on employing semantic point clouds for developing 3D city models for DTs. However, this investigation introduces a novel reflection. It advocates for the direct integration of semantic segmentation outcomes to establish the main base of DTs. The objective is to use enriched semantic point clouds for targeted simulations, bypassing the need to generate 3D models. The reflection examines the potential of enriched semantic 3D point clouds to replace semantic 3D city models in the DTs field. Ultimately, this reflection aims to diminish the cost and complexity of city models. This is in order to meet certain DT requirements.

To address this key research question, we will split it into two sub-questions. Firstly, does the point cloud meet the DTs' requirements? (Section 1); Secondly, is the point cloud a good alternative to 3D city models? (Section 2). We finally give some research guidelines related to extending the use of semantic point clouds in DTs for cities (Section 3).

A.1 Semantic point cloud: An input layer to DTs for cities.

The presence discourse in the urban and geospatial context is predominantly about the relevance and the potentiality of considering semantic 3D city models as an input layer to create DTs for cities [1–4]. However, it is worth considering the potentialities and advantages of semantic point clouds to serve DTs needs as a fundamental in-put layer without going through the 3D city modeling process.

To tackle this research question, it is interesting to identify the requirements of DTs for cities. Indeed, DTs for cities are conceptualized as a risk-free, living virtual ecosystem that mimics all the city elements to generate knowledge, assist urban decision-making through the city lifecycle, and provide outcomes at the city level [4–6]. Furthermore, from technical perspective, most of the research led to a tacit agreement on what constitutes a DT for cities in the geospatial domain and the Smart Cities initiatives previously announced by [7]. Thus, DTs for cities are based on (1) 3D city models enriched with geometrical and semantic information, (2) often incorporate heterogenous data namely coupled with historical and sensor data in near or real time (at an appropriate rate of synchronization), thus enabling (3) a link (e.g., data flow between the real counterpart and the virtual twin and vice versa), (4) allowing updates and analysis through a set of simulations, predictions, and visualization

tools, and (5) providing an integrated view of the multiple datasets and models through their life cycle, enabling to manage and adapt cities' current and future states.

If we intend to unpack the DTs definition, we will first start from the assumption that the DT for cities is a digital realistic city replica that incorporates all its city features. Thus, we can clearly validate this characteristic since a point cloud by nature is a high geometrically 3D representation of urban environments such as cities and other landscapes. However, back to definition, a DT must have semantic and geometrical information. This is completely accurate from geometrical dimension of a point cloud but is not applicable for semantics. In this regard, various approaches are proposed to enrich the point cloud and extend its semantic capabilities, whether through 3D semantic segmentation [8], or a conceptual data model called "Smart Point Cloud Infrastructure" [9], or data integration (GIS data, 3D city models) [10].
Although possibilities exist to tackle the lack of semantics in point cloud data, the enrichment of such data remains critical and challenging. Indeed, the current advancement in the scope of building DTs for cities is more focused on data integration approaches, including the association and integration of both point cloud data and semantic 3D city models using for example the new “PointCloud” module of CityGML 3.0 [11]. This module provides a new concept to bridge the gap between the geometrically detailed point cloud data and the enriched 3D semantic model. The integration of both datasets intuitively assigns sets of points to the corresponding objects. The existing approach in CityGML 3.0 provides an alternative for extending point cloud data to cover more semantic information beyond classification using various methods. Thus, integration of point cloud data with different data sets from GIS, BIM, and 3D city models helps to overcome the limitations of each approach and meet the DT requirements.

At the same time, a widespread algorithm and approaches have emerged to extract 3D objects automatically and effectively by semantically segmenting LiDAR point clouds using supervised learning methods, including Machine Learning-based segmentation, as well as Deep Learning-based segmentation such as multi-view-based methods, voxel-based methods, and direct methods that consume point clouds directly. Recent advances in semantic segmentation allow the extraction of the main urban features, such as buildings, vegetation, roads, railways [12], and many more that are relevant for DT’s applications [13–15].

In another hand, 3D semantic segmentation is relevant to update DTs for cities and track the changes at city-scale. That said that 3D semantic point cloud data enable the identification of the changes as they appear in the real world and updating corresponding information. For example, point cloud data allows to have a realistic and big picture of the status of an urban object under construction, especially if the current project does not have the necessary elements to generate a 3D model (i.e., lack of definitive footprint that is mandatory to generate accurate model). This says that the semantic point cloud can help in urban planning and management which is one of the common use cases of DTs for cities. In addition, the advantage of enriched semantic point cloud data is that almost all urban classes are extracted (i.e., static, and dynamic objects) and for specific applications, classes that are required or need to be updated are simply retained. Nevertheless, the classes that are not crucial are

neglected. It is worth mentioning as well that for different use cases, different classes are deployed, which is completely in line with the DT's requirements that replicate all the city objects as one snapshot, and for each use case, the data will be derived. Hence, semantic 3D point cloud enables us to precisely define the urban classes, thus augmenting the performance of the semantic extendibility, improving modeling capabilities, giving new interpretability of the data from different perspectives, and opening new doors for various simulations and urban analysis.

Turning to one of the promising characteristics of a DT (i.e., the simulation feature), yet the available processes and simulation tools that involve the direct use of 3D point clouds are still limited. Few studies are conducted to explore the potential of this type of data. For example, the authors of [16] introduced a new approach based on the medial axis transform to performing visibility analysis. The approach could be used for any typical airborne LiDAR data, which gives more realistic results and effectively handles the missing parts of the point cloud (e.g., walls and roofs). Furthermore, performing visibility analysis is more insightful when working with point cloud data, as vegetation is considered. Another study case on an urban scale performs the visibility analysis for both surface model and point cloud data and puts them together for in-depth analysis according to their efficiency and accuracy. To ensure intervisibility between the reference points (i.e., the observer and target points), the authors of [17] generate cubes for each point to block the sight lines. The study concludes that consistent input data (i.e., dense and classified point clouds) will certainly improve the findings.
On the other hand, solar radiation is a relevant use case in 3D urban modeling. Historically, solar irradiance was measured using DSM. However, 3D city models gained a significant amount of interest to improve the sun exposition estimations. In addition, the authors have developed a tool that uses point cloud data to model illumination and solar radiation [18]. The algorithm is based on voxels and has shown its capabilities for green areas as well as urban environments.

Figure 29 depicts an illustrative example from our research works, demonstrating the simulation of solar radiation directly performed on semantic point clouds. The point cloud data utilized in the study was acquired in the Flanders region of Belgium. The pre-trained RandLA-Net model [19] on the Semantic3d dataset [20] was used for semantic segmentation of point clouds. The relevant semantic classes that have the potential to impact solar radiation were extracted, including high vegetation, low vegetation, buildings, and scanning artifacts. To perform the simulation, the "pcsrt" function proposed by [18] was used. The source code for this tool was adapted from its publicly available version on GitHub as an open-access resource (https://github.com/hblyp/pcsrt, accessed on August 2, 2022).

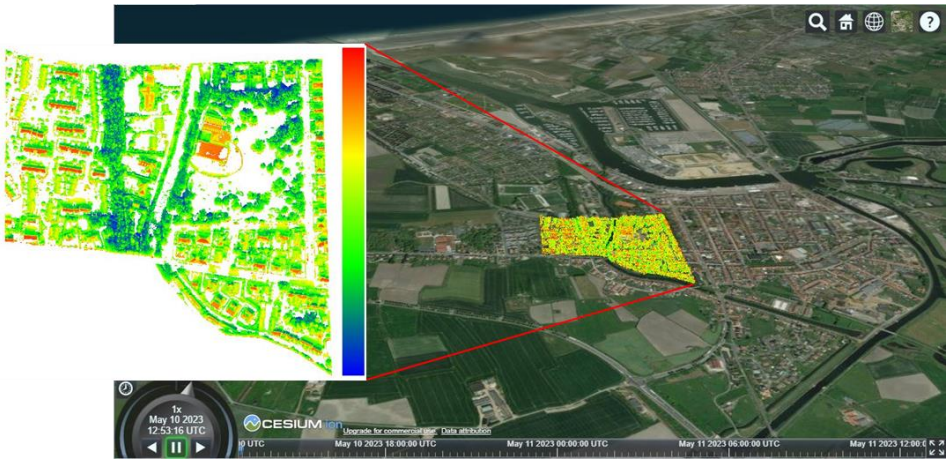


Figure 29. Example of solar radiation performed directly on semantic point cloud.

A further characteristic of DTs that is undoubtedly satisfied, is the visualization and the interactivity aspects. Point cloud data is supported through various visualization tools (i.e., web applications and game engines platforms). Additionally, point clouds are considered as a form of natural communication used as an input data to enhance immersivity and interactivity of Virtual Reality or Augmented Reality experiences. Moreover, for enhancing visualization, most of DT's initiatives tend to foster the ability to process, store, handle, and disseminate massive point clouds through the web, namely using the CesiumJS WebGL virtual globe. For instance, the Digital Twin of the City of Zurich sets a research agenda where further developments of the DTs for city are required namely, how to benefit from the derived mobile mapping point cloud data to improve the façades of the buildings as well as how to incorporate vegetation acquired from point cloud into the DTs. It is worth pointing that some visualization applications do not demand rich semantics, while others need specific attributes to perform simulations [21].

While the state of the art is well developed regarding the applications of 3D city models, some urban applications do not necessarily need a semantic 3D city model. Hence, enriched 3D semantic point cloud will certainly give new opportunities to perform some sophisticated analysis for DTs instead of creating surface models.

A.2 Semantic point cloud and semantic 3D city models: Advantages and Limitations.

While it is out of the scope of this article to compare 3D semantic model-based DTs and enriched semantic point cloud-based DTs, we will nevertheless highlight certain advantages and limitations of both semantic 3D city model and semantic 3D point clouds.

3D city models nowadays exhibit significant differences due to various factors including data acquisition, processing, storage, dissemination, use, and maintenance, as well as technical, socio-economic, political, and cultural variations. Consequently, it has become challenging to identify best practices, assess the quality of 3D city models, foster their appropriateness for specific use cases, and integrate effectively diverse datasets. Moreover, comparing multiple datasets present some difficulties, creating ambiguity in selecting the most suitable one. These concerns have implications for urban DTs, which rely on 3D city models as a key component [22]. Despite the availability of advanced 3D representation techniques and methods for creating 3D city models [23], significant challenges remain in achieving accurate and interactive 3D modeling of the urban environment. It is not just a matter of representing the environment in 3D, but also ensuring that the model is closer to the real world by attempting to represent as many urban objects of the physical world as possible without being restricted to a specific feature (i.e., buildings as they represent the identity of the city).

Research has identified several problems associated with 3D modeling [24], including limited data collection capabilities [25], reduced levels of automation [26], the lack of established modelling standards and rules [27], and limited applications for visualizing city models [28]. There are three types of modeling techniques: geometric modeling, mesh modeling, and hybrid modeling. Geometric modeling uses simple geometric primitives (planes, cylinders, lines, etc.) to represent objects, which reduces the volume of generated data and allows for semantic data to be embedded in the model. However, this method is dependent on the algorithms used and the resulting representation may lack fine details. Mesh modeling is useful for representing fine surface details, but the generated data remains voluminous, making interpretation and manipulation laborious for the user. Furthermore, 3D mesh models have limited analytic capabilities. However, few studies are conducted to improve the usability and applicability of mesh models by integrating semantic 3D city models with 3D mesh models [29]. Another related work enhances semantic segmentation of urban mesh using a hybrid model and a feature-based approach for semantic mesh segmentation in an urban scenario using real-world training data [30]. While meshes alone do not inherently allow semantic data to be embedded in the model since no shape or element recognition is performed. Semantic information could be introduced by modifying them or storing them using specialized data formats such as CityJSON that support semantics.

3D city modeling has different challenges that limit their full automation and usage. Firstly, there is an inconsistency between models generated using heterogenous dataset, reconstruction methods, and software, which affects geometry, appearance, and semantics. Standardization is the second challenge. Up to date, there are no common standards that

are established to handle DTs for cities from a technical point of view. However, we should take advantage of the existing standards by enabling convergence between them in a meaningful way with respect to the discrepancies (different geometries, semantics, structures and various spatial scales). Data quality is a major obstacle to create 3D city models, with many existing models containing errors that prevent their use in other software and applications. Data interoperability involves converting 3D models from one format to another. Language barriers may hinder understanding and interoperability. Indeed, public administrations often do not provide integrated and standardized 3D city models, making further analysis difficult. In addition, datasets may be managed in different standards and have different sets of information, making them unqualified for particular use cases. There is a lack of means to characterize data and their fit for purpose. In addition to the challenge posed by the heterogeneous nature of 3D city models in terms of making comparisons, another issue arises from the data integration approaches [22]. Data maintenance/governance is also a challenge, with governmental organizations lacking strategies for updating and maintaining different versions of the data. Lastly, implementing 3D data in the real world requires more precise definitions of specifications, validation mechanisms, clear semantics to address knowledge and skills gaps and integration of public and private sector models [7].

It is well known that in the scientific literature and in practice, the point cloud is considered a primary resource for reconstructing a semantic 3D city model. Indeed, 3D city models are by definition, a simplification of the real world (i.e., an abstraction at a certain LoD). With this in mind, 3D city models do not aim to represent all the features of the real world in the same detail as point cloud data. Thus, point cloud allows to avoid the abstraction needed for 3D city models, and objects such as trees are correctly rendered instead of being generalized according to city modeling standards. Furthermore, for a given point cloud, different 3D semantic model could be generated according to the use case, the standards and the quality of the acquired data. Moreover, recent advances in semantic segmentation and point cloud processing have made significant progress toward optimizing the algorithms and approaches. Another particularity of point cloud data is the lack of a specific standard to generate and process them. However, there may be variations in format and representation (e.g., voxels). In contrast, for 3D city models, there are many standards deployed to generate a semantically rich 3D model, namely CityGML and its JSON encoding, CityJSON. These standards are recognized as the foundation of DTs for cities. The existence of a range city modeling standards raises data interoperability issues. This does not mean that the standardization efforts are irrelevant, but this standard heterogeneity makes data integration challenging especially in the context of creating DTs in practice. This is also justified by having several 3D city models for the same urban scale from different stakeholders, but there is usually a single national LiDAR acquisition. Of course, for some cities, we may find more than one acquisition, however they are captured at different timescale having overlapped regions. It is also sometimes collected to fill some missing information for large scale aeras (i.e., urban land expansion). This extension of point cloud data to the temporal scale serves in the context of DTs given a 4D point cloud. However, this point cloud requires a high storage infrastructure, and detecting the changes is tricky since point-to-point corresponding is problematic.

Regarding the point cloud, another challenge that hinders its full potentials is the lack of topology, which can make simulating object behavior challenging. For instance, connections between different urban objects are difficult to represent without topology, which is why 3D models with a surface model are preferred for such representations, which are relevant for simulations namely for Computational Fluid Dynamics (CFD). Furthermore, 3D city models offer the possibilities to store attributes for objects (e.g., buildings) but also for surfaces, to build hierarchy (Building + Building Part) and to store the type of surfaces (namely used for energy modelling). It is also worth mentioning that 3D city models are significantly taking less space (compared to a raw point cloud, which is more than 10pts/m2 nowadays).

To conclude, semantic point clouds and semantic 3D city models are both great inputs to build DTs for cities. Both bring new opportunities but still have some weaknesses. However, all DT initiatives invest in hybrid models, enabling them to bridge the gap between different approaches and compensate for the limitations of the others.

A.3 Semantic point cloud: a new research field for DTs.

The potential benefits of implementing this new research path include reducing the cost of modeling, computation time, to take advantage of the semantic richness of the semantic point cloud since frequently we make large-scale acquisitions and heavy processing operations to end up exploiting only the buildings class in 3D modeling without other details of the urban environment (i.e., vegetation, roads). This approach also avoids the complexities of 3D modeling, particularly for other urban objects than buildings like transportation infrastructure and vegetation. It’s also advantageous for updating urban DTs and conducting specific simulations that require accurate and detailed information about the urban environment. The proposed reflection challenges the frequently used approach of relying solely on 3D modeling for DTs applications and suggests that semantic point clouds can be a viable alternative, particularly for addressing the limitations of 3D models and meeting the needs of DTs in an easy and effective way. However, it is important to note that while semantic 3D point clouds may be a useful input layer for some DT applications, they may not be a complete replacement for 3D city models in all cases. The choice between using semantic 3D point clouds or 3D city models as an input layer for DT applications will depend on the specific application purposes, the available resources, and the required level of accuracy and detail.

Further research is needed to explore the potential of semantic point clouds and develop new approaches for integrating them into DTs applications. As a first step of our reflection, we investigated the feasibility of some simulations that can be performed directly on point clouds. In the next steps, we will evaluate and validate our approach by comparing it with 3D city models that utilize the same data, in order to affirm its effectiveness and accuracy.

This work also suggests some perspectives to meet the requirements of DTs:

- Future research should focus on exploring the potential of semantic point clouds and developing integration methods for their use in DTs applications.
- It is important to consider the specific requirements of the application, available resources, and desired level of accuracy and detail when choosing between semantic 3D point clouds and 3D city models as an input layer for DT applications.
- Establishing standards for DTs could bring several benefits. Firstly, it would enable increased interoperability among different systems applying this concept, thereby facilitating collaboration and data exchange. Additionally, clearly defined standards could help ensure the security and protection of data, as well as the quality of the created models.
- Defining a preliminary LoD for semantic point clouds can help ensure the quality and usability of data for specific DTs applications.
- Developing new approaches and algorithms that enable the direct simulation of urban environments using semantically rich point clouds instead of generating 3D model, more precisely for sophisticated simulations such as computational fluid dynamics.
- Studying change detection and updating of DTs with semantically rich point clouds.

Conclusions

In this paper, we have proposed a research reflection on the use of semantic 3D point clouds as an alternative to 3D city models for DTs needs. We have introduced the limitations and performance of both 3D city models and semantic point clouds. Furthermore, we explain how semantic point clouds can overcome the limitations of 3D city models to create a DTs. We then presented the initial guidelines of the suggested reflection, which aims to answer the research question of whether a point cloud can meet the requirements of DTs by going beyond considering a semantic point cloud as input for modeling and performing simulations directly on it without resorting to 3D modeling. This research direction should be further explored to match point clouds to DTs' requirements and extend their urban applications. In short, semantic 3D point clouds appear as potential data that goes beyond the current deployment of creating 3D city models, which puts them at the forefront of new needs in urban simulations.

B. TOWARDS A DIGITAL TWIN OF LIEGE: THE CORE 3D MODEL BASED ON SEMANTIC SEGMENTATION AND AUTOMATED MODELING OF LIDAR POINT CLOUDS

Based on paper: Zouhair Ballouch, Imane Jeddoub, Rafika Hajji, Roland Billen. 2024. " Towards a Digital Twin of Liege: The Core 3D Model based on Semantic Segmentation and Automated Modeling of LiDAR Point Clouds". INTERNATIONAL CONFERENCE ON SMART DATA AND SMART CITIES.

Abstract:

The emergence of Digital Twins (DTs) in city planning and management marks a contemporary trend, elevating the realm of 3D modeling and simulation for cities. In this context, the use of semantic point clouds to generate 3D city models for Digital Twins proves instrumental in addressing this evolving need. This article introduces a processing pipeline for the automatic modeling of buildings, roads, and vegetation based on the semantic segmentation results of 3D LiDAR point clouds. It employs a semantic segmentation approach that integrates multiple training datasets to achieve precise extraction of target objects. Open-source reconstruction tools have been adapted for building and road modeling, while a Python code was optimized for tree modeling, leveraging a foundational code. The case study was conducted in the city of Liège, Belgium. The obtained results were satisfactory, and the schemas and geometry of the developed models were validated. An evaluation of the adopted reconstruction methods was conducted, along with their comparison to other methods from the literature.

1. Introduction

Digital twins for cities have become an efficient and collaborative decision-making tool that helps overcome cities' challenges [1,31]. The Urban Digital Twins (UDTs) serve city needs by integrating data, models, and processes into a one-stop platform, enabling two-way flows from the physical world to the digital replica and vice versa [14]. As we embark on the journey of implementing UDTs, semantic 3D city models gain perspective [7]. Over the past decades, many scholars have increasingly focused on the creation and use of 3D city models beyond simple visualization [24]. Indeed, semantic 3D city models offered many potential applications to urban and geospatial analysis and application at the city scale based on open standards such as CityGML [32]. Up to date, many city models are spread worldwide, implemented using different data and approaches, and serve various purposes. In contrast, there is a lack of a standard or common framework for 3D city modeling, which was one of the main motivations behind the work of [22]. The authors designed a holistic instrument to benchmark and evaluate 3D city models worldwide. Based on the findings, cities, such as Brussels and Namur, have invested in the creation of their 3D city models in the Belgian context.

Inspired by current digital technologies and recognizing the relevance of UDTs in the context of urban planning and management, the city of Liege has invested in the implementation of 3D city models as the first step toward the development of digital twins. This study presents the results of SEM3D, a project supported by Digital Wallonia and conducted in the Geomatics Unit with the collaboration of the city of Liège. The main contribution is to automatically extract 3D semantic objects for urban applications and explore the 3D modeling process to create 3D models of the derived urban objects. The paper proposes and tests an overall framework, from data preparation to 3D modeling. The particularity of this approach is that it is not restricted to a specific urban object (e.g., buildings) but also enables the modeling of other thematic objects (i.e., roads and vegetation) using open-source tools and the semantic segmentation results of 3D LiDAR data. The main contribution is to shed light on the relevance of the semantic segmentation of 3D airborne LiDAR data in the city modeling framework. The proposed framework uses the existing data and adapts available open-source tools to create a standardized CityJSON 3D city model of common urban objects. The paper is structured as follows: Section 2 gives an overview of the use of semantically segmented point cloud data in city modeling processes. Section 3 presents the proposed workflow, ranging from data preprocessing and transformation to 3D modeling. A description of the study area is provided in the same section. Section 4 discusses the findings. Section 5 concludes and gives an outlook for future perspectives.

2. Related Works

Digital twins for cities are data-hungry platforms [15,33]. They are based on heterogeneous data sets (geospatial data and sensor data, to name a few). 3D point cloud data from airborne acquisition is the most common input data for digital twins' implementation. They have shown their capabilities as an input layer for city modeling, namely for 3D building modeling [34]. For instance, 3D BAG is 3D reference building data covering the whole Netherlands in several output formats. The data are provided based on an automatic 3D reconstruction pipeline at different levels of detail (LoD). The workflow uses the 2D building footprints and the AHN point cloud data acquired by airborne laser scanning (ALS). Furthermore, 3dfier is an automatic workflow and open-source software that reconstructs LoD1 models using classified LiDAR point cloud data in LAS format and 2D semantic polygons (i.e., building footprints, water bodies, etc.) [35]. The workflow is based on a set of rules and uses a YAML configuration file to generate the 3D model. The software has support for various formats. In addition, the authors in [36] have proposed an automatic CityJSON workflow that extracts roof surfaces from LiDAR data and generates LoD2.1 building models. Another related work, City3D, was conducted by [37], presenting a large-scale 3D building reconstruction from the ALS point cloud. The authors propose an approach that infers the vertical walls of buildings from airborne LiDAR point clouds. In their work, the authors address in a comprehensive way

the challenges related to large-scale urban reconstruction from ALS data, namely: building instance segmentation, incomplete data, and complex structures. However, many of these processes do not involve advanced classification of the point cloud data. In this regard, enhancing semantic segmentation-based AI approaches improves the use of 3D point cloud data, thus creating 3D urban models [38,39]. It enables the automatic extraction of single and multiple city objects, which simplifies object modeling. Many DT initiatives have acquired point cloud data (airbone or UAV data) to create and enrich their semantic 3D city models [2,40,41]. In fact, optimizing the semantic segmentation process is of great interest to reconstruct 3D city models and implement UDTs efficiently and correctly.

3.Materials and Methods

This section explains the methodology used in the framework of this work. It consists of three main steps: data collection, data processing, and 3D modeling. The workflow is summarized in Figure 30. The pink boxes describe input data (namely point cloud data, topographical data, and images), while the blue ones refer to the intermediate transformations and processes. The resulting CityJSON 3D city model is presented in a green box. In the following, we first describe the input data. Then, we explain the semantic segmentation process and, finally, the reconstruction process for each city object.

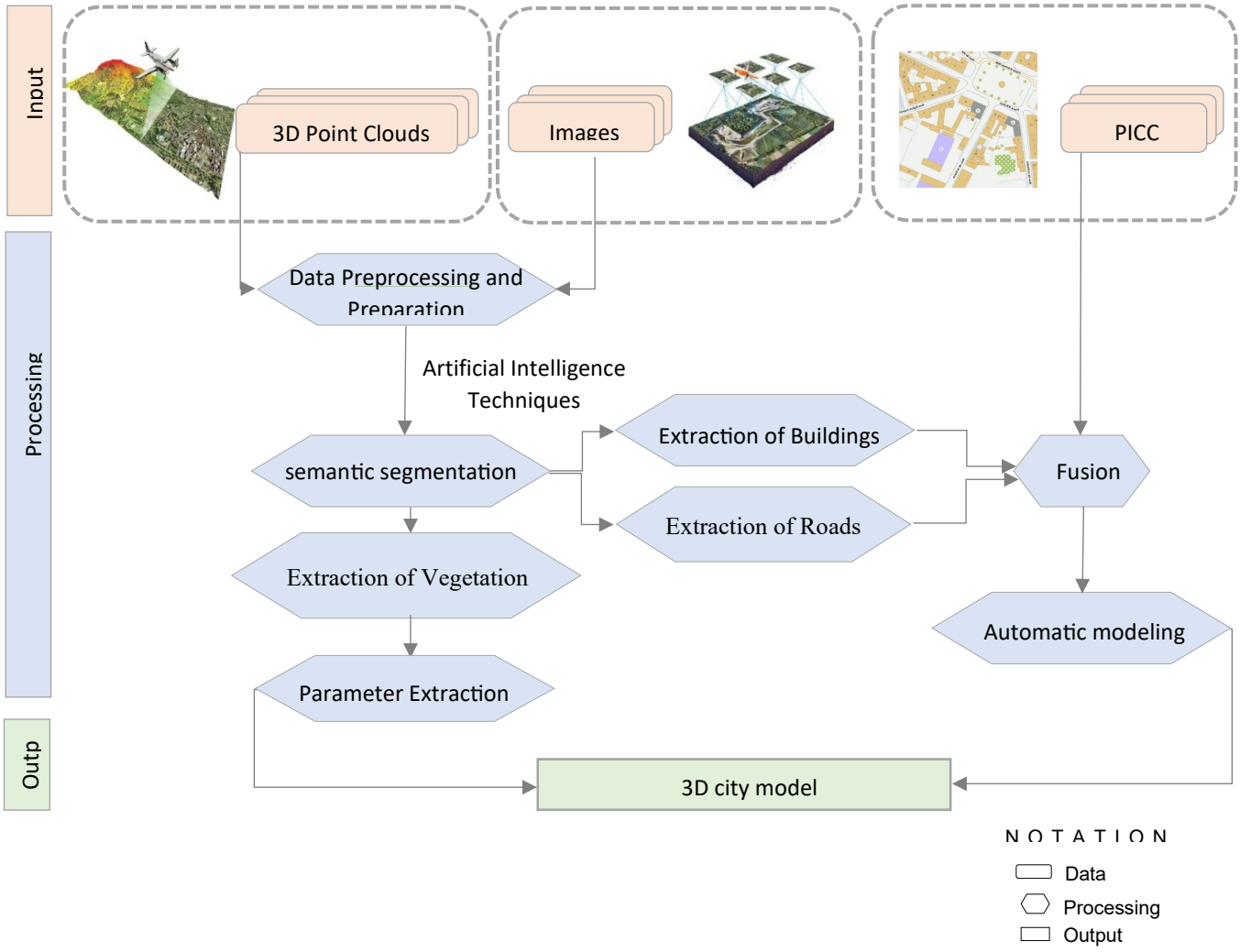


Figure 30. The general workflow.

The first step involves the collection of the input data sets. Then, the second step takes raw point cloud data and performs a semantic segmentation process. For this, an artificial intelligence-based approach is used. The approach fuses airborne LIDAR point clouds with corresponding aerial photos. It can accurately extract the main 3D objects within an urban scene with both geometric precision and semantic richness. Deploying a fusion approach with other sources (aerial photos, satellite images, etc.) allows for combining the spectral richness of images and the altimetric accuracy of 3D point clouds. Our aim is to automate the extraction of 3D objects, such as roads, vegetation, etc., in our study area, presented subsequently in section 3.2, with high accuracy and performance.

The third step is dedicated to the modeling process. For each urban object, an approach or an open-source tool is deployed. The extracted semantic classes from the semantic segmentation of 3D point cloud data are assigned to each modeling pipeline. For instance,

to model buildings using open standards (i.e., CityJSON), GeoFlow uses the building point cloud data as well as the building footprints to automatically generate the 3D building models at LoD2. For the road modeling, the class number of the corresponding road point cloud data is specified in the configuration file necessary to run the open-source tool 3Dfier[1]. The same logic is applied to the vegetation modeling. The derived vegetation point cloud data were integrated into the modeling process based on an adapted code. This code was based on the fundamental code previously available as open source at this link (https://github.com/RobbieG91/TreeConstruction).

3.1 Data collection

The data sources include LiDAR point clouds and PICC[2] (Plan d'Information sur le Cadre de Cartographie) data. The PICC serves as the three-dimensional digital cartographic reference for the entire Wallonia region in Belgium, with precision less than 25 cm, comprehensively capturing all identifiable elements of the Walloon landscape, such as buildings, structures, railway networks, hydrographic networks, roadways (including lanes, edges, sidewalks, etc.), and more. The datasets were provided by the Walloon region in Belgium. Additionally, other datasets, namely the SUM-Helsinki dataset and the SensatUrban dataset, were acquired through free downloads via links provided later (refer to https://github.com/QingyongHu/SensatUrban). Consequently, the Liège dataset was created by us based on the region's data. Table 11 provides a description of the data sources.

[1] https://github.com/tudelft3d/3dfier
[2] https://geoportail.wallonie.be/georeferentiel/PICC

Table 11. Data sources.

	Type of object	File type	Geometry/ Data type	Num. objects	Description
Semantic segmentation data	SUM-Helsinki dataset	.PLY	Mesh	19 M	This dataset covers approximately 4 km2 in Helsinki, Finland, featuring six classes: Terrain, Vegetation, Building, Water, Vehicle, and Boat. Derived from 2017's 3D textured meshes of Helsinki with a ground sampling distance of 7.5 cm, the dataset is obtained through oblique aerial images and processed using ContextCapture software. The study area is concentrated on the central region of Helsinki, encompassing 64 selected tiles.
	SensatUrban dataset	.PLY	Point	2847.1M	The SensatUrban dataset collected by UAV over Birmingham, Cambridge, and York cities covers six square kilometers of urban area and features 13 semantic classes with 6 attributes per point: X, Y, Z, and RGB information
	Liège dataset	.LAS	Point	25.635.237	The Liège dataset is derived from the data of the Walloon region described in section 3.1 and includes 12 semantic classes ('Ground', 'High Vegetation', 'Buildings', 'Walls', 'Bridge', 'Parking', 'Rail', 'Traffic Roads', 'Street Furniture', 'Cars', 'Footpath', 'Bikes', 'Water'). Each data point within the dataset is characterized by 6 attributes: X, Y, Z, and RGB information.
Modeling data	point clouds from the "Outremeuse" neighborhood in Liège	.LAS	Point	10.619.980	It is a point cloud of the Outremeuse neighborhood in Liège, extracted from four point cloud tiles covering this area, using vector data representing the administrative boundaries of the city of Liège
	PICC (building surfaces, road axes, road edges	.SHP	Building surfaces: Polygon	Building footprints :3897	These are vector data from the Walloon region available upon request. Each object (building,road axe,,etc) is represented by an identifier and attribute information relative to it
			Road axes: Ligne	Road axis:342	
			Road edges: Ligne	Road edges:1607	

3.2 Semantic segmentation of 3D LiDAR points clouds

The quality of semantic segmentation results plays a crucial role in the geometric accuracy of 3D urban models created based on these results. Therefore, the choice of a semantic segmentation approach that accurately extracts urban objects is essential. To achieve this, the LiDAR point clouds were fused with their corresponding images using the "RandLaNet" deep learning model [19]. This model was adopted for semantic segmentation due to its documented performance in the literature [19,42]. In this study, we trained this model on three different datasets: SensatUrban, accessible at https://github.com/QingyongHu/SensatUrban; a dataset from https://3d.bk.tudelft.nl/projects/meshannotation/; and a dataset created in the urban context of Liège city (access to this data is available upon request). Model parameters and hyperparameters were adjusted. The predictions made by the trained model were based on point cloud data from the "Outremeuse" neighborhood in Liège, Belgium. The location of this neighborhood is shown in Figure 31 below. The data used for creating the Liège dataset and working on the Outremeuse neighborhood are from recent LIDAR acquisitions in the Walloon region of Belgium (2021–2022).

The data characteristics include an average flight altitude (AGL) of 2400 m, density of 6.8 points/m², and the use of Double LMSQ780 and Double VQ780II-S equipment. The data were provided in 8 blocks in ". LAS" format, adhering to ETRS89 / Lambert Belgian 2008 planimetric coordinates, Second General Leveling altimetric coordinates, and planimetric accuracy with RMSE <= 1 m and altimetric accuracy with RMSE <= 0.4 m. A preprocessing step, including cleaning, was conducted to ensure data consistency. After adjusting projections and merging LiDAR point clouds with corresponding images (see Figure 31). The

Outremeuse neighborhood data were prepared for prediction, while data from other areas in Liège were utilized for creating the third training dataset. Data preprocessing was performed using CloudCompare, and data preparation and processing were carried out using the Ubuntu tool. The model "RandLaNet" has already been validated through our previous studies as well as by several studies in the literature using various evaluation criteria, including measures such as Accuracy, Intersection over Union (IoU), Recall, F1-score, and Confusion matrix [19,38,42]. Therefore, in this study, we have opted just for visual validation through a comparison of the model's results with the ground truth (see Figure 32), considering the comprehensive set of evaluation metrics used in prior research.

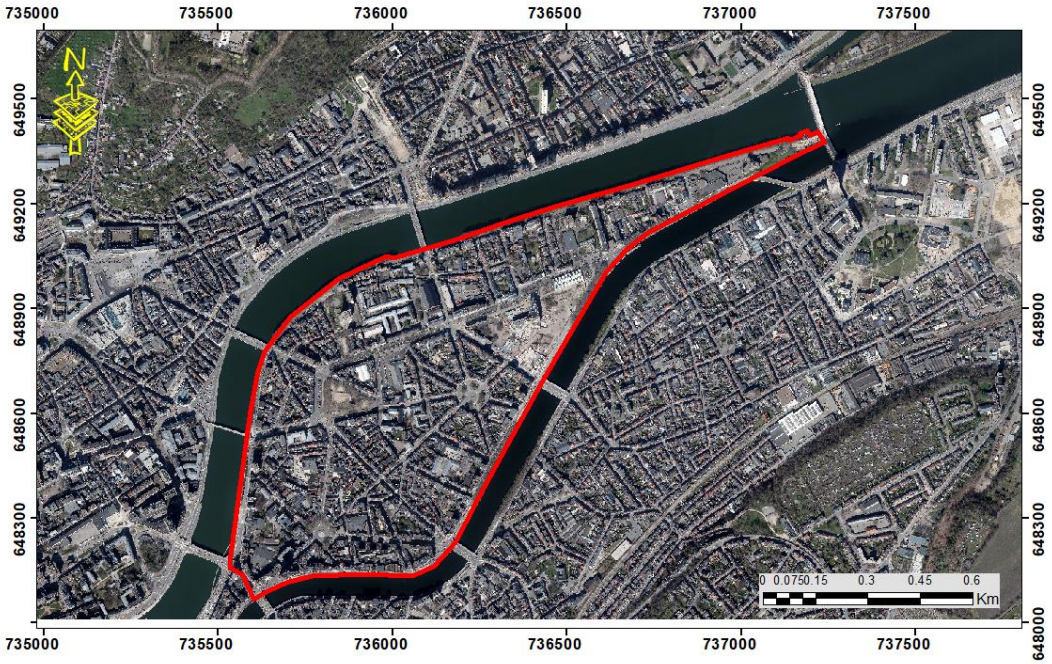


Figure 31. Geographical location of the Outremeuse district.

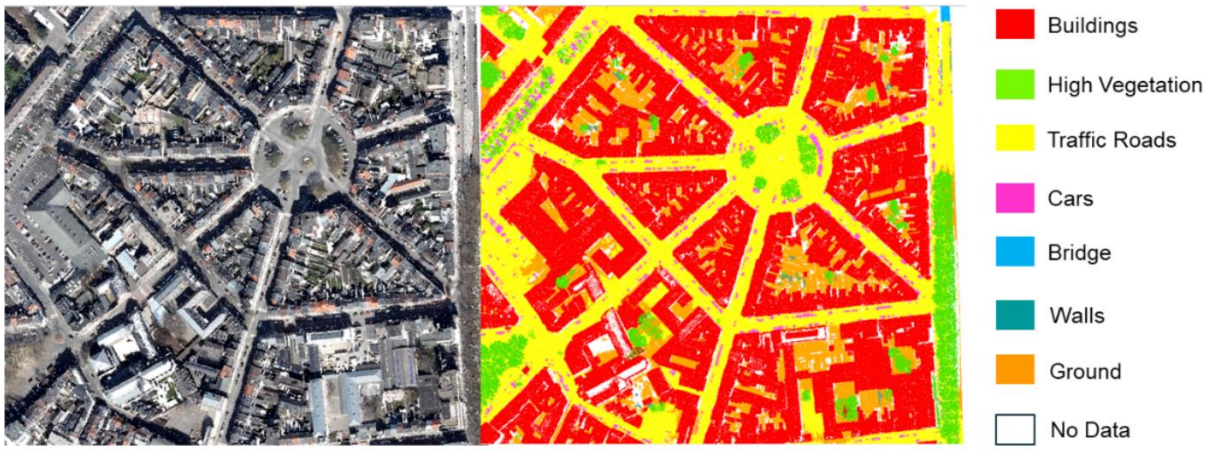


Figure 32. 3D point cloud representation and (B) example of 3D semantic segmentation outputs- Outremeuse district.

3.3 3D modeling workflow

This section outlines the processing pipeline we followed for generating 3D models from a classified point cloud. The required data include the classified point cloud obtained from Section 3.2 and the PICC.

3.3.1 Automatic 3D buildings modeling

Building modeling was conducted using the GeoFlow tool (the code is available for open access via this link: https://github.com/geoflow3d/geoflow-bundle), an open-source tool for 3D building model reconstruction from point clouds. The objective is to generate a realistic three-dimensional representation of buildings by harnessing point cloud data, vector data (PICC), and modeling functions provided by GeoFlow.

To execute the reconstruction from input data, both a JSON file containing a flowchart describing the logic of the reconstruction and the executable GeoFlow are necessary components. The flowchart outlines how different plugins and nodes connect, while GeoFlow executes the logic defined in the flowchart.

3.3.2 Automatic roads modeling

Studies and methods for 3D road modeling are still limited. The focus was historically on 3D building models. This is due to the lack of complete data and because most 3D roads have linear representation.

However, a recent study conducted by [43], has proposed an automatic process of 3D road modeling based on CityGML 3.0 specifications and CityJSON encoding. The authors produce a LoD2 3D road model using a semi-automatic extraction workflow based on mobile mapping LiDAR data.

Given the type of data provided in the scope of this work, we opted for the two approaches below: the first approach relies on developing an FME workbench. FME is an ETL (Extract, Transform, and Load) process that allows a series of data transformers. It also has the capability to read, convert, and write many data types and formats. Initially, we created an FME workspace (refer to Appendix) for the roads using only the 2D road axis and the georeferenced DEM generated from the LiDAR data using the "SurfaceModeller" transformer. We create a CityJSON v1.0.1-compliant model describing the road. The FME workbench is reusable. For now, the workbench allows the generation of the LoD1 road model. Further work will help extend it to produce higher LoDs. The second approach that seemed to be promising was the use of 3dfier. The tool allows you to generate smooth road surfaces in different output formats. To implement the practical modeling of roads with 3dfier using the classified point cloud and PICC data, we have followed a few steps. Firstly, the PICC data is initially linear, while 3dfier requires a set of topologically connected polygons as input data. To address this, we use QGIS tools to transform the linear representation into a polygonal result (see Figure 33). The axis and edges provided by the PICC data were transformed into surfaces using QGIS, thereby generating polygons representing the roads. These polygons were then used to perform the lifting based on the semantics of the road polygons. We then adjusted and adapted the 3dfier lifting options and setting parameters. Essentially, 3dfier relies on a binary classification of ground and non-ground (minimum requirements). However, in our case, a detailed classification was performed. We incorporated this detailed classification to accurately extract the "roads" class (use classes 2 and 11), which 3dfier will utilize during the lifting process.

Following the 3dfier requirements, we configured the (.yml) file used in the scope of this work.

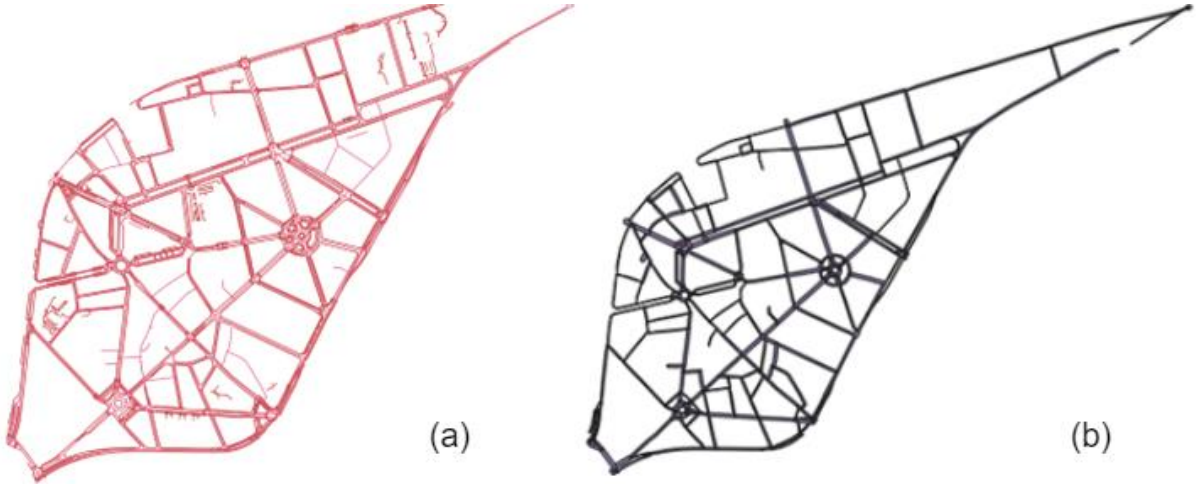


Figure 33. The data preparation for 3Dfier road modeling: (a) the shapefile raw data, linear representations (b) the polygonal representation based on QGIS tool.

3.3.3 Automatic vegetation modeling

To automatically generate 3D models of trees from airborne LiDAR point cloud data with a LoD2, a three-step process was followed: classification, segmentation, and modeling. Firstly, the point cloud must be exclusively classified as vegetation, after which individual trees need to be segmented. Finally, these segments of individual trees serve as the data source for constructing 3D tree models. The steps are illustrated in Figure 34.

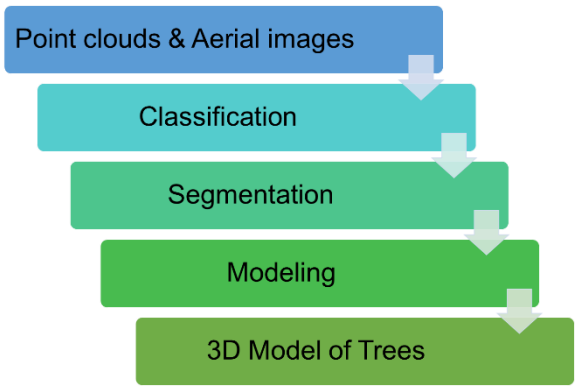


Figure 34. General Workflow for Tree Modeling.

A) Classification :

The classification phase has already been detailed in Section 3.2. To extract the vegetation point cloud, we utilized the CloudCompare tool. After importing the classified point cloud and displaying the scalar field corresponding to the classification, we proceeded to extract the "vegetation" class. This extraction can be performed in various ways, one of which involves accessing the main menu of CloudCompare and selecting the "Filter by Value" option.

B) Segmentation :

The aim of this step is to assign an identifier to each tree. One can opt to use available automatic codes, such as those presented on (https://github.com/r-lidar/lidR/tree/master), or choose tools like utilizing an algorithm integrated into the CloudCompare tool, as illustrated in this study. To employ CloudCompare, simply access the software's main menu, navigate to "Plugins," and select the "TreeIso" algorithm. Subsequently, we executed three types of segmentation: initial segmentation, intermediate segmentation, and refined segmentation. Adjustments to the parameters were made until achieving a satisfactory result. The selection of parameters depends on the type of data, data quality, etc. Finally, a data cleaning step is crucial, especially in situations where certain trees are not correctly segmented, particularly in dense forest areas. Various automatic or semi-automatic cleaning methods within the CloudCompare tool can be employed.

C) Modeling

The modeling process was based on the LoD specifications proposed by [44]. The required parameters are extracted from the segmented vegetation and modeled accordingly. These parameters include the tree top (the 99th percentile of the height), the tree base (ground height), the peripheral point (height range where most points are located), the base of the tree crown (the 5th percentile of the height), and two intermediate divisions for added detail. These divisions are determined using the midpoint between the peripheral point and the top, as well as the base of the tree crown, respectively [45]. These parameters are crucial for constructing individual plant objects, as illustrated in Figure 35.

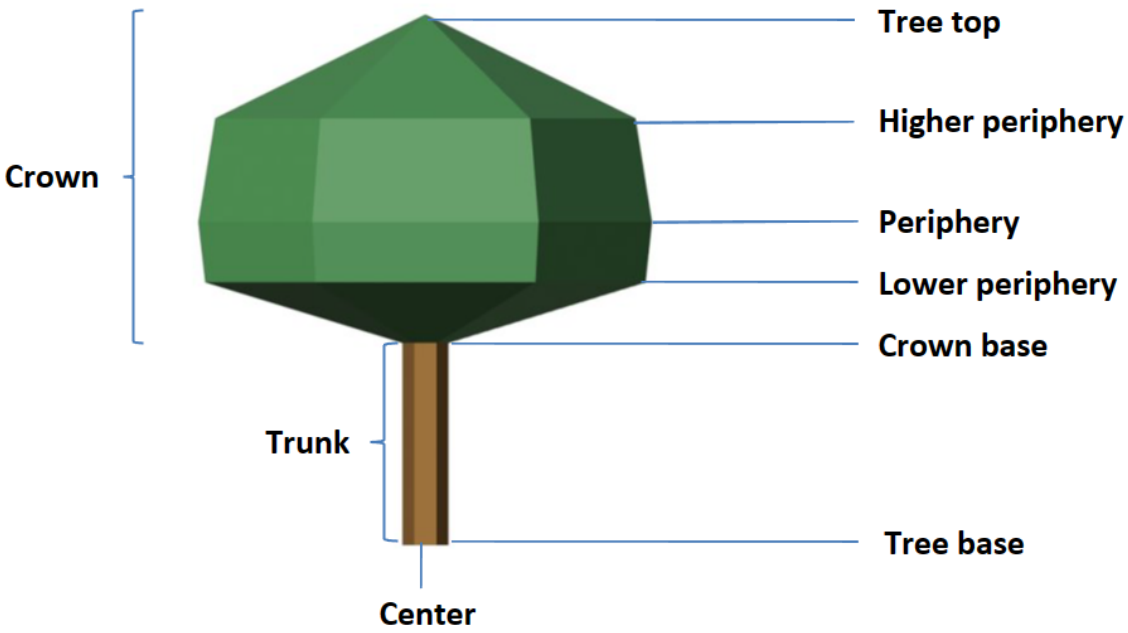


Figure 35. Tree construction parameters [45].

Each LoD employs a different combination of the extracted parameters to construct tree models. LoD0 uses only the peripheral radius and the tree base. LoD1 utilizes the peripheral radius, tree base, and tree top. LoD2 incorporates all the extracted parameters. The 3D tree models are constructed in accordance with CityJSON specifications. It is essential that vertices are arranged in a counter clockwise trigonometric order (CCW) when viewed from the outside, as it is a common rule in 3D modeling. This ensures that the faces have outward oriented normals. This guarantees that the constructed geometry is visible in any rendering software with 3D capabilities and adheres to ISO standards [International Organization for Standardization, 2019].

To initiate the extraction of parameters, the essential input data consists of a point cloud with attributes X, Y, Z, Tree Segment ID (a specific identifier for each tree obtained from the segmentation step), and the attribute "Height Above Ground," which was computed using the ground and tree classes. The calculations were performed using CloudCompare.

4. Results and discussion

The resulting 3D model for buildings, roads, and vegetation was compliant with CityJSON v1.1. All models were validated at the schematic and geometric levels. For that, CityJSON has a wide range of free and open-source tools and software that assist and facilitate the use

and manipulation of CityJSON data. For instance, Val3dity[3] is an open-source software dedicated to validating the 3D primitives (geometries) of the model. The software reports the geometric and topological errors by specifying the object in concern. For each 3D model, different errors are reported.

The Schema validation was fullfiled based on the official validator for CityJSON files, cjval[4]. Cjio[5] was also used to merge, upgrade, and validate the CityJSON files. The results were visualized using the web viewer ninja[6]. The Table 12 summarizes the validation process results of all city objects modelled in this work.

Table 12. Validation of the different 3D city objects using the open-source validator software.

File	Val3dity	Cjval	Ninja
Buildings.json	93,2%	100% valid	Semantic surfaces
Roads.json	87% valid	100% valid	No semantic surfaces
Vegetation.json	100% valid	100% valid	No semantic surfaces

4.1 3D building model

Geoflow has demonstrated its capabilities in providing good results both from geometric and semantic terms. The building model schema and geometry were both validated. Each building is represented by a specific and unique identifier derived from PICC data. Thus, semantic and attribute information has been accurately assigned to each building. The LoD2.2 is maintained for this work, showing various building elements such as building roofs, walls, etc. We also generated the LoD1.2 and LoD1.3 for future work. This automated modeling method offers significant advantages compared to other existing reconstruction methods in the literature. For instance, we use the same input data to generate the LoD1 building model using the 3Dfier tool, which may not be sufficient for certain urban applications. Furthermore, we conducted another approach using FME. We create a FME workbench that produce LoD1 building model. Despite the advantages it offers, this method is semi-automatic and requires human expertise. In addition, it poses challenges in automatically incorporating semantic information. Figure 36 provides an overview result of the three methods.

[3] https://github.com/tudelft3d/val3dity
[4] https://validator.cityjson.org/
[5] https://github.com/cityjson/cjio
[6] https://ninja.cityjson.org/

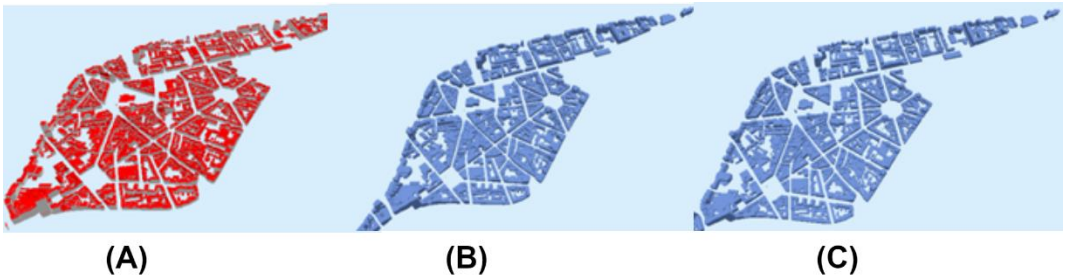


Figure 36. An Example of 3D building model: (A) LoD2 model based on Geoflow, (B) LoD1 based on 3Dfier and (C) LoD1 based on FME of the Outremeuse District.

The main errors reported from Val3dity are: CONSECUTIVE_POINTS_SAME, NON_PLANAR_POLYGON_NORMALS_DEVIATION.

4.2 Results of 3D Road Modeling

The produced 3D road model from 3Dfier (refer to Figure 37) was effectively validated both from geometric and schematic levels. Each road is represented by a unique identifier derived from PICC data; thematic attributes are handled by default by 3Dfier. The 3D road model is a LoD1 MultiSurface model. This automated modeling method is relevant while working with 2D polygonal data representation. The errors reported in val3dity are namely: CONSECUTIVE_POINTS_SAME and RING_SELF_INTERSECTION.

As we explained earlier, we created a FME road workbench as well. However, the result was invalid from a geometric perspective, and several errors were reported. To solve that, we use triangulate function of Cjio. The obtained model is valid and only CONSECUTIVE_POINTS_SAME error was reported for 2% of the 3D primitives.

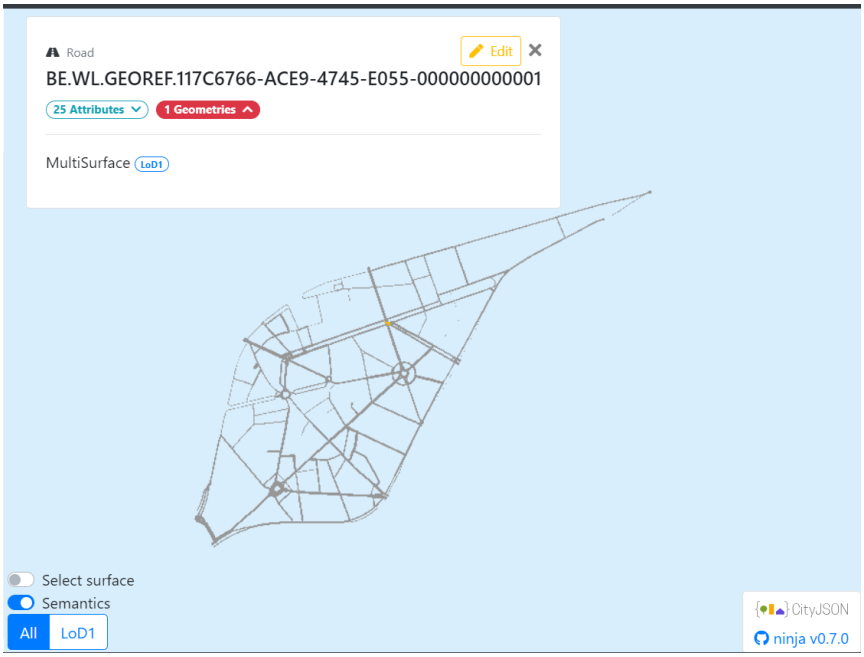


Figure 37. LoD1 road model of the Outremeuse District using 3Dfier.

4.3 Results of 3D Vegetation Modeling

The approach employed for tree modeling has yielded satisfactory geometric and semantic results (see Figure 38). The schema and geometry underwent through validation (refer to Table 12). Each tree is represented by a specific identifier, with parameters extracted from the tree and associated semantic information. The level of detail in tree modeling is LoD2, representing a realistic tree form (see Figure 38). This automated modeling approach offers significant advantages compared to other existing reconstruction methods/approaches in the literature. For instance, the use of 3dfier, while advantageous in automatically adding semantic and attributive information, is limited to presenting trees at LoD0, which proves insufficient for certain environmental and ecological applications. The 3D tree model was fully validated, and no geometric errors were reported.

Additionally, employing the tree reconstruction method with FME schemas presents some limitations, including scale issues and the manual addition of semantic information. However, the modeling approach utilized in this study also has its constraints, such as errors in the segmentation step, particularly in densely populated forest areas where precise tree differentiation poses a challenge.

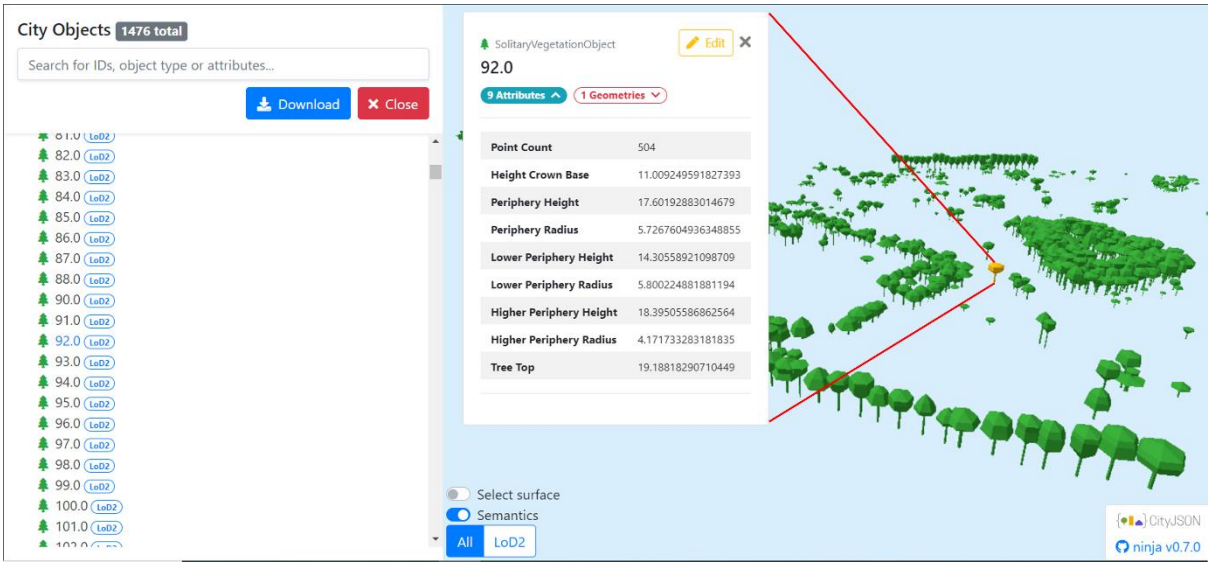


Figure 38. LoD2 tree models of Results of the Outremeuse District.

4.4 Discussion:

The aim of this work was to propose a general and reusable approach to generating 3D city models. The framework ranges from data preparation and pre-processing to 3D modeling. The methodology was implemented in a case study to illustrate the approach and to handle the challenges related to 3D city modeling. Especially since the literature mainly focuses on 3D building modeling, we presented a 3D modeling pipeline for buildings, roads, and vegetation.

Merging 3D buildings, roads, and vegetation could be achieved (refer to **Figure 39**)

Figure 39. 3D CITY MODEL OF OUTREMEUSE DISTRICT.

Table 13 below summarizes the findings according to various criteria. It will help guide the user through the reproducibility and applicability of the process.

Table 13. Basic information of 3D modeling of the city objects according to various criteria.

3D models	Buildings	Roads	Trees
Type of methods (automatic, semi-automatic, manual)	Automatic	Automatic	Semi-automatic
Input data	PICC data Point Clouds	PICC data Point Clouds	Point Clouds
Minimum required attributes in point clouds	X, Y, Z, Classification	X, Y, Z, Classification	X, Y, Z, Segment ID (for each tree), Height Above Grounds
Point cloud classification (basic or advanced)	Basic	Advanced	Advanced
LoD	LoD2	LoD1	LoD2
License, terms of use of the modeling tool	General Public License	General Public License	Not specified
Supported format (input/output)	Input: point cloud: LAS or. LAZ. 2D polygon: GeoPackage, ESRI Shapefile, or a connection to a PostGIS database Output: CityJSON	Input: point cloud: LAS or. LAZ Output is in the following formats: OBJ, CityGML, CityJSON, CSV (for buildings only, i.e. their ID and height (ground+roof) are output in a tabular format), PostGIS, and STL.	Input: point cloud: LAS or. LAZ. Output: CityJSON
Geometry type	Solid	MultiSurface	MultiSurface
Semantic handling	Yes	No	No
Minimum requirements	Classification into two categories:	Classification into two categories: ground and non-ground	Classification involving two categories: ground and trees

(configuration file)	ground and buildings		
Thematic attributes (Native/workaround)	Native support	Native support	-
Time (computational)	Very fast	Very fast	Very fast
Is there any report to guide the user.	Yes: https://github.com/geoflow3d/geoflow-bundle	Yes: https://github.com/tudelft3d/3dfier	Yes: https://github.com/RobbieG91/TreeConstruction

5. Conclusion :

This article presents a processing pipeline designed for the automated modeling of buildings, roads, and vegetation using semantic segmentation outcomes derived from 3D LiDAR point clouds. It has established the digital twin foundation of the city of Liège. The methodology employs a good semantic segmentation approach, ensuring precise extraction of target objects. The open-source reconstruction tools for buildings and roads modeling were adapted, and simultaneously, a Python code for tree modeling was optimized. The application of these methods in a case study conducted in Liège, Belgium, yielded satisfactory results, with validated schemas and geometry for the developed models. Furthermore, a evaluation of the adopted reconstruction methods, including a comparative analysis with other techniques from the existing literature, underscores the robustness and efficacy of the proposed approach. As a perspective, we recommend exploiting city models created in simulation tools by adding additional data to complete the digital twin of the city of Liège. Besides, we suggest investigating the proposed processing pipeline in other cities that do not yet have an urban model to evaluate its efficiency and limits in different urban contexts. Additionally, we recommend modeling other urban objects with the aim of producing highly detailed urban models rich in urban knowledge.

Enrichment of 3D urban modeling from semantic point clouds.

This section represents continuity by developing other approaches for creating 3D models of additional objects, namely the ground, bridges, walls, and cars. The developed approaches are standardized, following the conceptual data model “CityJSON”. These models are not created solely for visualization. The fusion of the different models enables the creation of an enriched full 3D model of a large-scale urban environment. This full 3D model represents a solid foundation of an urban Digital Twin onto which other data (for example dynamic data) will be grafted. Therefore, the semantic and geometric richness of this 3D information can improve simulation systems and tools for analyzing key urban challenges. Moreover, city models created can enhance simulations and urban analysis, thereby improving decision-making. Additionally, urban models are considered the main input for current building energy simulation, visibility analysis, flood studies, noise propagation simulation, etc. Furthermore, these city models can be directly used in virtual reality and augmented reality applications, highlighting participatory approaches (citizen implications).

A) Creation of the TIN

To generate the TIN, we first used the CityGML format. CityGML is both a data model and a standardized, open exchange format for storing 3D digital models of urban landscapes. Among the classes of objects in CityGML is the relief ("Relief Feature"). This class is subdivided into four subclasses that represent the different types of relief accepted in the CityGML model. The four types of relief are as follows:

- The "Raster Relief" type;
- The "TIN" type;
- The "BreaklineRelief" type;
- The "MasspointRelief" type.

In the context of this research, a TIN (Triangular Irregular Network) relief was created using the 3D point cloud (see Figure 40). To achieve this, firstly, we extracted the points from the cloud corresponding to the ground class. Once the study area was delimited, a TIN was generated. To obtain a TIN perfectly matching the shape of our study area, some additional processing was required. Afterward, specific attributes for the CityGML format were created. The level of detail and the role of the CityGML object were determined accordingly. For the relief, the level of detail is TIN, and the role is "Relief Component". After creating these attributes, a TIN was produced in the CityGML format.

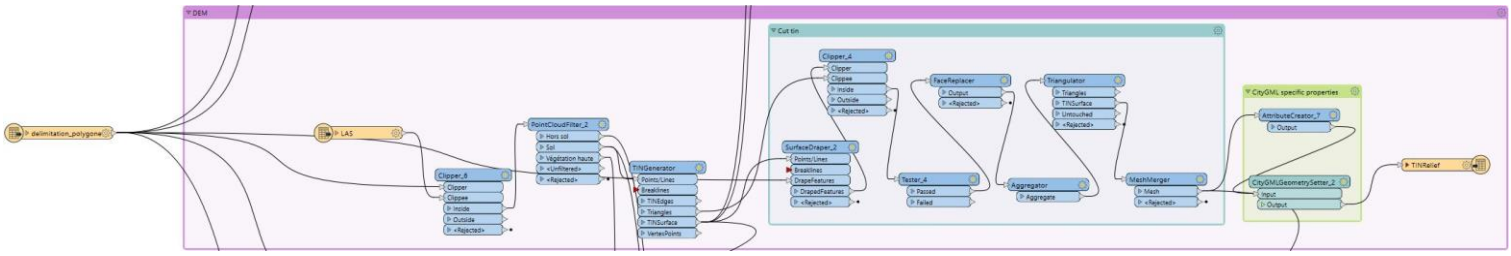


Figure 40. The FME schema followed for the creation of the TIN

Since all city models created in the context of this research are in CityJSON format, we converted the CityGML to CityJSON. The following figure represents the TIN created in CityJSON format (see Figure 41).

Figure 41. The TIN created in CityJSON format.

B) 3D modeling of Bridges

The produced 3D bridge model from 3Dfier tool (detailed in section I) was effectively validated both from geometric and schematic levels. The modeling of the bridges involved two main steps. Firstly, the geometry was created from point clouds. Secondly, the unique identifier of each bridge was retrieved from 2D vector data. The latter were created from the bounding boxes of the point cloud of the bridge object, then segmented to isolate each bridge and assign an identifier to each one. The IDs are handled by default by 3Dfier. The 3D bridge model is a LoD1 MultiSurface model (see Figure 42).

This automated modeling method is relevant while working with 2D polygonal data representation. The schema and geometry validation are 100% accurate.

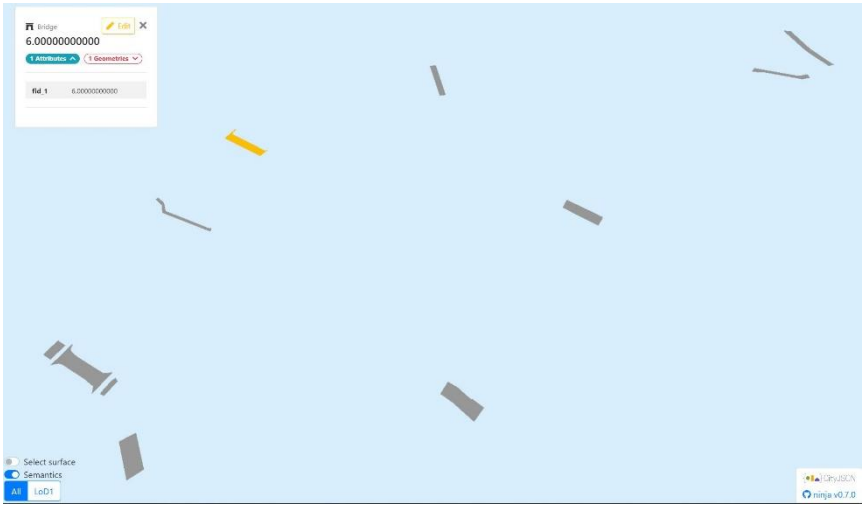


Figure 42. LoD1 bridges model of the Outremeuse District using 3Dfier.

C) 3D modeling of walls

The 3D walls model was generated by the 3Dfier tool, which was described previously. It was validated both in terms of geometry and structure. Each individual wall is uniquely identified based on vector data, which was derived from the bounding boxes of the wall object's point cloud. These vector data were segmented to isolate each wall and assign a unique ID to each one. Default handling of thematic attributes is performed by 3Dfier. The resulting 3D walls model conforms to LoD1 MultiSurface standards (see Figure 43). The schema and geometry validation are 100% accurate.

Figure 43. LoD1 walls model of the Outremeuse District using 3Dfier.

3D modeling of cars

For the 3D modeling of cars, to our knowledge, there is currently no specific car class in the thematic classes specified by the CityJson and CityGML formats. Therefore, we modeled the car class as a generic class in CityJson. To do this, first, we extracted the vector layer of detected cars in the study area. After that, we performed instance segmentation to assign a specific identifier to each car. Then, we calculated the centroids using QGIS. Next, we generated a layer containing the localization of the cars. Subsequently, we developed a Python code based on the car localization (in .shp format) and a car template in (.obj) format. The developed Python code is accessible at the following link (https://github.com/ZouhairBALLOUCH/3D-modeling-of-cars). The following figure shows an example of the results obtained (see Figure 44).

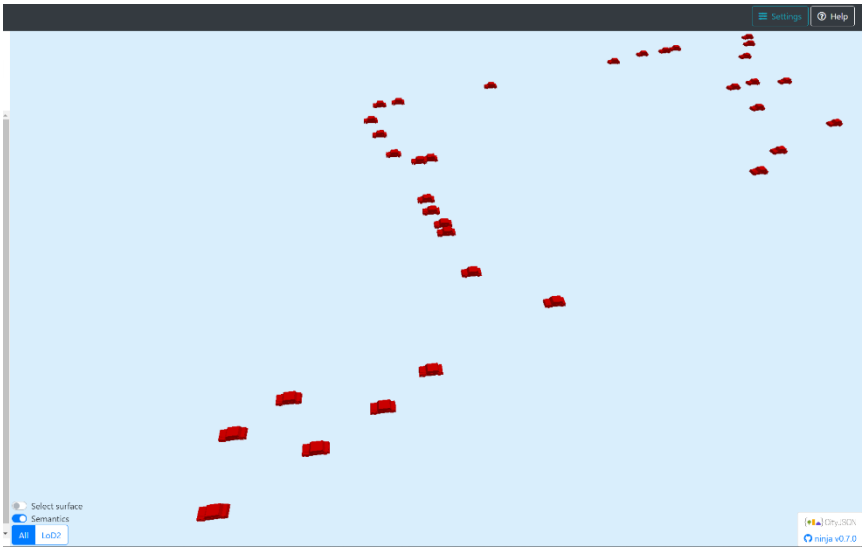


Figure 44. 3D Model of Cars in the Outremeuse District

References:

1. Ketzler, B.; Naserentin, V.; Latino, F.; Zangelidis, C.; Thuvander, L.; Logg, A. Digital Twins for Cities: A State of the Art Review. *Built Environment* **2020**, *46*, 547–573, doi:10.2148/benv.46.4.547.

2. DImitrov, H.; Petrova-Antonova, D. 3D City Model as a First Step towards Digital Twin of Sofia City.; 2021; Vol. 43, pp. 23–30.

3. Alva, P.; Biljecki, F.; Stouffs, R. USE CASES FOR DISTRICT-SCALE URBAN DIGITAL TWINS. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2022**, *XLVIII-4/W4-2022*, 5–12, doi:10.5194/isprs-archives-XLVIII-4-W4-2022-5-2022.

4. Würstle, P.; Padsala, R.; Santhanavanich, T.; Coors, V. VIABILITY TESTING OF GAME ENGINE USAGE FOR VISUALIZATION OF 3D GEOSPATIAL DATA WITH OGC STANDARDS. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2022**, *X-4/W2-2022*, 281–288, doi:10.5194/isprs-annals-X-4-W2-2022-281-2022.

5. Hristov, P.O.; Petrova-Antonova, D.; Ilieva, S.; Rizov, R. ENABLING CITY DIGITAL TWINS THROUGH URBAN LIVING LABS. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2022**, *XLIII-B1-2022*, 151–156, doi:10.5194/isprs-archives-XLIII-B1-2022-151-2022.

6. Nguyen, S.H.; Kolbe, T.H. PATH-TRACING SEMANTIC NETWORKS TO INTERPRET CHANGES IN SEMANTIC 3D CITY MODELS. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2022**, *X-4/W2-2022*, 217–224, doi:10.5194/isprs-annals-X-4-W2-2022-217-2022.

7. Stoter, J.E.; Arroyo Ohori, G.A.K.; Noardo, F. Digital Twins: A Comprehensive Solution or Hopeful Vision? *GIM International: the worldwide magazine for geomatics* **2021**, *2021*.

8. 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); IEEE: Nashville, TN, USA, June 2021; pp. 4975–4985.

9. Poux, F. The Smart Point Cloud: Structuring 3D Intelligent Point Data, 2019.

10. 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. Spatial Inf. Sci.* **2021**, *VIII-4/W2-2021*, 105–112, doi:10.5194/isprs-annals-VIII-4-W2-2021-105-2021.

11. 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 Annals of the*

Photogrammetry, Remote Sensing and Spatial Information Sciences **2021**, *VIII-4-W2-2021*, 105–112, doi:10.5194/isprs-annals-VIII-4-W2-2021-105-2021.

12. Zhou, Y.; Ji, A.; Zhang, L.; Xue, X. Sampling-Attention Deep Learning Network with Transfer Learning for Large-Scale Urban Point Cloud Semantic Segmentation. *Engineering Applications of Artificial Intelligence* **2023**, *117*, 105554, doi:10.1016/j.engappai.2022.105554.

13. Döllner, J. Geospatial Artificial Intelligence: Potentials of Machine Learning for 3D Point Clouds and Geospatial Digital Twins. *PFG* **2020**, *88*, 15–24, doi:10.1007/s41064-020-00102-3.

14. Lehtola, V.V.; Koeva, M.; Elberink, S.O.; Raposo, P.; Virtanen, J.-P.; Vahdatikhaki, F.; Borsci, S. Digital Twin of a City: Review of Technology Serving City Needs. *International Journal of Applied Earth Observation and Geoinformation* **2022**, 102915, doi:10.1016/j.jag.2022.102915.

15. Masoumi, H.; Shirowzhan, S.; Eskandarpour, P.; Pettit, C.J. City Digital Twins: Their Maturity Level and Differentiation from 3D City Models. *Big Earth Data* **2023**, *0*, 1–46, doi:10.1080/20964471.2022.2160156.

16. Peters, R.; Ledoux, H.; Biljecki, F. Visibility Analysis in a Point Cloud Based on the Medial Axis Transform. *Eurographics Workshop on Urban Data Modelling and Visualisation* **2015**, 6 pages, doi:10.2312/UDMV.20151342.

17. Zhang, G.; van Oosterom, P.J.M.; Verbree, E. Point Cloud Based Visibility Analysis: First Experimental Results. *Proceedings of the 20th AGILE Conference on Geographic Information Science* **2017**.

18. Pružinec, F.; Ďuračiová, R. A Point-Cloud Solar Radiation Tool. *Energies* **2022**, *15*, 7018, doi:10.3390/en15197018.

19. 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); IEEE: Seattle, WA, USA, June 2020; pp. 11105–11114.

20. Hackel, T.; Savinov, N.; Ladicky, L.; Wegner, J.D.; Schindler, K.; Pollefeys, M. Semantic3D.Net: A New Large-Scale Point Cloud Classification Benchmark. *arXiv:1704.03847 [cs]* **2017**.

21. Schrotter, G.; Hürzeler, C. The Digital Twin of the City of Zurich for Urban Planning. *PFG – Journal of Photogrammetry Remote Sensing and Geoinformation Science* **2020**, *88*, doi:10.1007/s41064-020-00092-2.

22. Lei, B.; Stouffs, R.; Biljecki, F. Assessing and Benchmarking 3D City Models. *International Journal of Geographical Information Science* **2022**, doi:10.1080/13658816.2022.2140808.

23. Toth, C.; Jóźków, G. Remote Sensing Platforms and Sensors: A Survey. *ISPRS Journal of Photogrammetry and Remote Sensing* **2016**, *115*, 22–36, doi:10.1016/j.isprsjprs.2015.10.004.

24. Stoter, J.E.; Ohori, G.A.; Dukai, B.; Labetski, A.; Kavisha, K.; Vitalis, S.; Ledoux, H. State of the Art in 3D City Modelling: Six Challenges Facing 3D Data as a Platform. *GIM International: the worldwide magazine for geomatics* **2020**, *34*.

25. Ledoux, H.; Biljecki, F.; Dukai, B.; Kumar, K.; Peters, R.; Stoter, J.; Commandeur, T. 3dfier: Automatic Reconstruction of 3D City Models. *Journal of Open Source Software* **2021**, *6*, 2866, doi:10.21105/joss.02866.

26. Park, Y.; Guldmann, J.-M. Creating 3D City Models with Building Footprints and LIDAR Point Cloud Classification: A Machine Learning Approach. *Computers, Environment and Urban Systems* **2019**, *75*, 76–89, doi:10.1016/j.compenvurbsys.2019.01.004.

27. Eriksson, H.; Johansson, T.; Olsson, P.-O.; Andersson, M.; Engvall, J.; Hast, I.; Harrie, L. Requirements, Development, and Evaluation of A National Building Standard—A Swedish Case Study. *ISPRS International Journal of Geo-Information* **2020**, *9*, 78, doi:10.3390/ijgi9020078.

28. Liamis, T.; Mimis, A. Establishing Semantic 3D City Models by GRextADE: The Case of the Greece. *J geovis spat anal* **2022**, *6*, 15, doi:10.1007/s41651-022-00114-0.

29. Willenborg, B.; Pültz, M.; Kolbe, T. INTEGRATION OF SEMANTIC 3D CITY MODELS AND 3D MESH MODELS FOR ACCURACY IMPROVEMENTS OF SOLAR POTENTIAL ANALYSES. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2018**, *XLII-4/W10*, 223–230, doi:10.5194/isprs-archives-XLII-4-W10-223-2018.

30. Tutzauer, P.; Laupheimer, D.; Haala, N. SEMANTIC URBAN MESH ENHANCEMENT UTILIZING A HYBRID MODEL. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2019**, *IV-2-W7*, 175–182, doi:10.5194/isprs-annals-IV-2-W7-175-2019.

31. Jeddoub, I.; Nys, G.-A.; Hajji, R.; Billen, R. Digital Twins for Cities: Analyzing the Gap between Concepts and Current Implementations with a Specific Focus on Data Integration. *International Journal of Applied Earth Observation and Geoinformation* **2023**, *122*, 103440, doi:10.1016/j.jag.2023.103440.

32. Biljecki, F.; Stoter, J.; Ledoux, H.; Zlatanova, S.; Coltekin, A. Applications of 3D City Models: State of the Art Review. *ISPRS International Journal of Geo-Information* **2015**, *4*, 2842–2889, doi:10.3390/ijgi4042842.

33. Batty, M. Digital Twins. *Environment and Planning B: Urban Analytics and City Science* **2018**, *45*, 817–820, doi:10.1177/2399808318796416.

34. Jeddoub, I.; Ballouch, Z.; Hajji, R.; Billen, R. Enriched Semantic 3D Point Clouds: An Alternative to 3D City Models for Digital Twin for Cities? In Proceedings of the Recent Advances in 3D Geoinformation Science; Kolbe, T.H., Donaubauer, A., Beil, C., Eds.; Springer Nature Switzerland: Cham, 2024; pp. 407–423.

35. Ledoux, H.; Biljecki, F.; Dukai, B.; Kumar, K.; Peters, R.; Stoter, J.; Commandeur, T. 3dfier: Automatic Reconstruction of 3D City Models. *JOSS* **2021**, *6*, 2866, doi:10.21105/joss.02866.

36. Nys, G.-A.; Billen, R.; Poux, F. AUTOMATIC 3D BUILDINGS COMPACT RECONSTRUCTION from LIDAR POINT CLOUDS.; 2020; Vol. 43, pp. 473–478.

37. Huang, J.; Stoter, J.; Peters, R.; Nan, L. City3D: Large-Scale Building Reconstruction from Airborne LiDAR Point Clouds. *Remote Sensing* **2022**, *14*, doi:10.3390/rs14092254.

38. 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 Sensing* **2022**, *14*, 3415, doi:10.3390/rs14143415.

39. Ballouch, Z.; Hajji, R.; Kharroubi, A.; Poux, F.; Billen, R. Investigating Prior-Level Fusion Approaches for Enriched Semantic Segmentation of Urban LiDAR Point. *Remote Sensing* **2024**, *16*, doi:10.3390/rs16020329.

40. Peters, R.; Dukai, B.; Vitalis, S.; van Liempt, J.; Stoter, J. Automated 3D Reconstruction of LoD2 and LoD1 Models for All 10 Million Buildings of the Netherlands. *Photogrammetric Engineering and Remote Sensing* **2022**, *88*, 165–170, doi:10.14358/PERS.21-00032R2.

41. Khawte, S.S.; Koeva, M.N.; Gevaert, C.M.; Oude Elberink, S.; Pedro, A.A. DIGITAL TWIN CREATION FOR SLUMS IN BRAZIL BASED ON UAV DATA. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2022**, *XLVIII-4/W4-2022*, 75–81, doi:10.5194/isprs-archives-XLVIII-4-W4-2022-75-2022.

42. Guo, Y.; Wang, H.; Hu, Q.; Liu, H.; Liu, L.; Bennamoun, M. Deep Learning for 3D Point Clouds: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2021**, *43*, 4338–4364, doi:10.1109/TPAMI.2020.3005434.

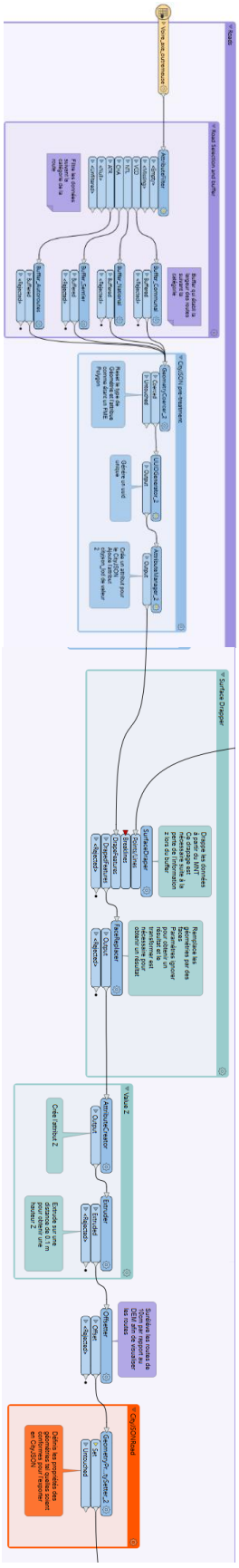
43. Yarroudh, A.; Nys, G.-A.; Hajji, R. 3D MODELING OF ROAD INFRASTRUCTURES ACCORDING TO CITYGML 3.0 AND ITS CITYJSON ENCODING. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2023**, *XLVIII-1-W2-2023*, 63–70, doi:10.5194/isprs-archives-XLVIII-1-W2-2023-63-2023.

44. Ortega-Córdova, L. Urban Vegetation Modeling 3D Levels of Detail. **2018**.

45. de Groot, R. Automatic Construction of 3D Tree Models in Multiple Levels of Detail from Airborne LiDAR Data. **2020**.

Appendix

FME workbench of road modeling



CHAPTER 5

Conclusion and research perspectives

5 .1 KEY FINDINGS AND CONTRIBUTIONS

The current limitations in extracting 3D semantic objects from point clouds highlight the need for low-cost fusion approaches, given the superior precision recently demonstrated by fusion approaches compared to non-fusion approaches. Additionally, it was imperative to develop approaches capable of extracting maximum urban details while enhancing accuracy and performance. This will improve the semantic richness of 3D point clouds and generate semantically rich 3D city models. These models form the basis for creating urban digital twins. This research aimed to address these challenges by answering the key question: "How to enhance the accuracy and richness of 3D semantic segmentation in urban environments through the fusion of airborne 3D point clouds and images using Deep Learning techniques?" Furthermore, it sought to address the complementary question: "How to exploit enriched 3D semantic point clouds to build urban Digital Twins?"

The thesis first explored the contribution of deep learning to the semantic segmentation of large-scale 3D point clouds in urban areas. It examined existing families of approaches and proposed an innovative fusion approach integrating airborne LiDAR point clouds and images. Subsequently, it presented a less data-intensive fusion approach, introduced a new airborne 3D LiDAR dataset, adapted the advanced technique 'RandLaNet', and resolved semantic class inconsistencies between LiDAR and image datasets. Following this, it developed and compared three prior-level fusion scenarios to enhance semantic segmentation richness, utilizing 'RandLaNet' and 'KPConv' to optimize these scenarios. Finally, it developed a practical methodology for extracting objects from high-resolution images and projecting them onto point clouds, to enrich the results of LiDAR approaches.

Moreover, the thesis thoroughly explored the utilization of enriched point clouds for urban simulations and 3D automatic modeling of urban objects. It proposed a new reflection on using enriched semantic point clouds for urban simulations, addressing the needs of urban digital twins without requiring 3D modeling, which can be costly. Additionally, the thesis presented an automatic processing pipeline for modeling urban objects extracted from point clouds. This methodology used deep learning techniques for precise object extraction and adapted open-source reconstruction tools for some objects like buildings and roads, while developing Python codes and FME schemas for other objects like trees and ground. This approach allowed for the creation of detailed and accurate urban models from enriched semantic point clouds, facilitating the creation of urban digital twins.

In summary, this thesis comprehensively addressed the research questions posed by developing innovative approaches and demonstrating their effectiveness in improving 3D semantic segmentation and creating urban digital twins. Below are the research questions and the corresponding answers:

How to enhance the accuracy and richness of 3D semantic segmentation in urban environments through the fusion of airborne 3D point clouds and images using Deep Learning techniques?

To address this main question, several sub-questions were proposed, and their respective answers are as follows:

What is the appropriate data fusion level of 3D LiDAR point clouds and airborne imagery to meet the spectral and geometrical information required for enriched semantic segmentation?

Before deriving the appropriate fusion levels, chapter 1 firstly presented the contribution of fusion approaches compared to direct approaches and derived product-based approaches. A thorough assessment of the performance and limitations of the different methodological families was investigated (Table 14). Subsequently, a general fusion approach of 3D point clouds and corresponding images was proposed. Chapter 1 addresses the first part of the approach, which is the classification of images. An in-depth evaluation of 6 deep learning architectures for image classification was conducted. This evaluation was carried out with the objective of deriving the most accurate architecture. For the implementation, a series of Drone images were used to evaluate the different architectures. The results indicate that all tested techniques exhibit acceptable results in terms of accuracy and frequency-weighted Intersection over Union (IU). However, the “Resnet50_Unet” technique outperformed the other in both metrics (refer to Table 15). Consequently, it has been identified as the most suitable technique for classification of drone images. This chapter emphasizes two key points. Firstly, the quality of the results could be further enhanced by increasing the quantity of training data. Secondly, extending the number of training epochs could also contribute to improving the results. In this chapter, only a limited number of epochs were used, as the primary objective was to evaluate different techniques rather than achieving maximum precision.

Table 14. Advantages and disadvantages of the different families of semantic segmentation approaches.

Approach	Advantages	Disadvantages
Direct approaches	Preserve the original topological relationships of point cloud	-Expensive -Few developed programs
Derived product based approaches	-Easy and fast drive -Requires few parameters	-Loss of information and accuracy due to re-sampling -False data caused by resampling step -Errors accumulation
Hybrid approaches	-Accurate -Efficient	-Expensive -Require a minimum difference in time of acquisition of the two types of data

Table 15. Comparing Deep Learning techniques for classification accuracy of Drone images.

	Unet	Vgg_Unet	Resnet50_Unet	Segnet	Vgg_Segnet	Resnet50_Segnet
Accuracy	0.71	0.76	0.85	0.72	0.7215	0.82

To determine the most appropriate fusion level for combining point clouds and their corresponding images, four possible fusion levels have been analyzed. The goal is to identify the most suitable one for meeting the spectral and geometrical information required for enhanced semantic segmentation richness. These levels include prior-level, point-level, feature-level, and decision-level fusion. Each fusion level presents strengths and limitations as outlined below:

A) Prior-level fusion approaches

Is a fusion approach that integrates classified images with 3D point clouds. Afterward, a Deep Learning technique is applied for 3D semantic segmentation, as depicted in Figure 45. This approach offers several benefits. Firstly, it allows for the direct utilization of semantic information from image classification. This allows for quicker convergence and reduced loss function during both training and testing phases. Secondly, integration of optical image

classification results accelerates loss stabilization and minimizes it more rapidly. Nonetheless, prior fusion approaches encounter challenges related to non-overlapping regions and uncertainties.

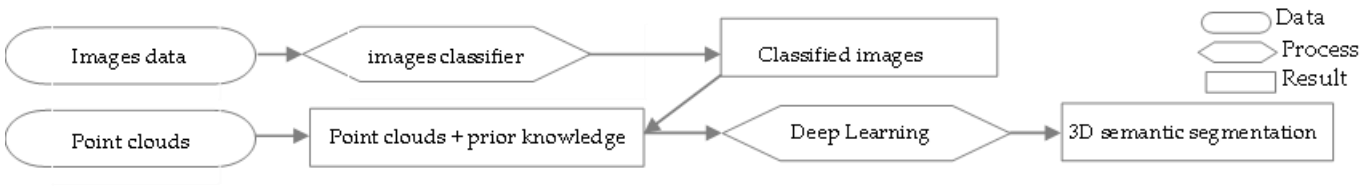


Figure 45. The general workflow of the prior-level fusion approaches.

B) Point-level fusion approaches

At the point level, fusion approaches involve assigning spectral data from images to each corresponding point in the clouds. Subsequently, deep learning techniques are employed for semantic segmentation of 3D point clouds with radiometric data, as illustrated in Figure 46. These approaches offer advantages such as good quality results and ease of use. However, they also entail drawbacks such as significant memory and computation requirements. Additionally, there is a necessity for simultaneous or minimally different acquisition times for both types of data.

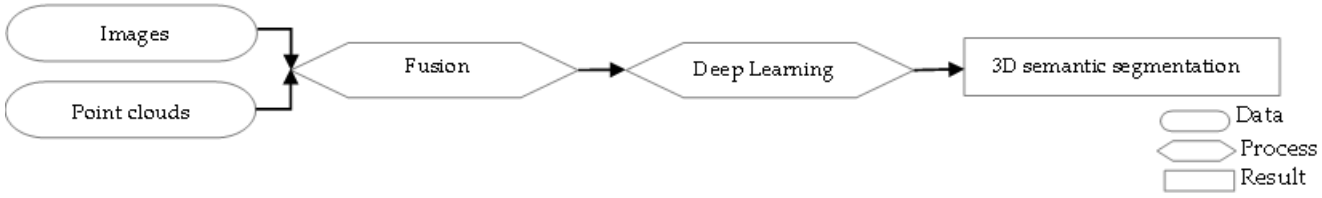


Figure 46. The general workflow of the point-level fusion approaches

C) Feature-level fusion approaches

In feature-level fusion approaches, features extracted from optical images and 3D point clouds are combined using neural networks (Figure 47). These combined features undergo processing with a Multi-Layer Perceptron (or other) to achieve semantic segmentation results. These approaches demonstrate improved precision compared to approaches using only radiometric or geometrical information. Additionally, feature-level fusion facilitates objective data compression while maintaining essential information. However, challenges such as the

wrapping phenomenon in orthophotos can arise. Additionally, the inability of LiDAR data to capture occluded objects like low-rise buildings can also be a concern.

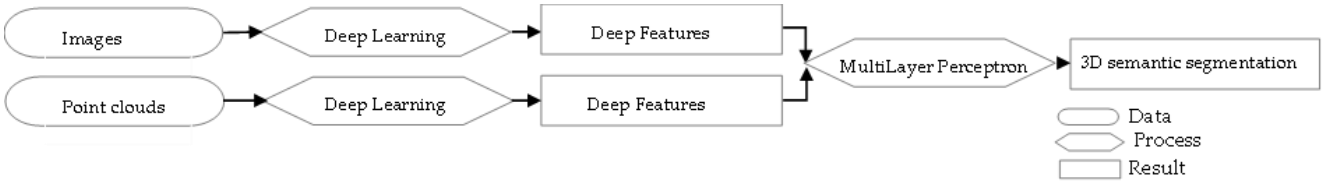


Figure 47. The general workflow of the feature-level fusion approaches

D) Decision-level fusion approaches

These approaches utilize a specific process to learn the final fusion layer using results from semantic segmentation (Figure 48). They combine outputs from two classifiers, each operating on LiDAR or pixel space. That is, one classifier processes spectral information for semantic segmentation of images, while the other segments LiDAR data. The two types of results are then fused using a heuristic fusion rule (or other). Decision-level fusion offers advantages such as independent training and validation of classification processes. As a result, this leads to flexibility and low complexity. Additionally, it can achieve good performance by employing each modality to train a single DL technique. This allows for the learning of independent features. However, relying on prior decisions from two classifiers can be affected by their shortcomings. While it may achieve modest improvements, this approach remains limited compared to other fusion approaches. Furthermore, decision-level fusion requires more memory as the DL structure combines features at a later stage. In addition, it demands additional parameters for convolutional layers and other operations.

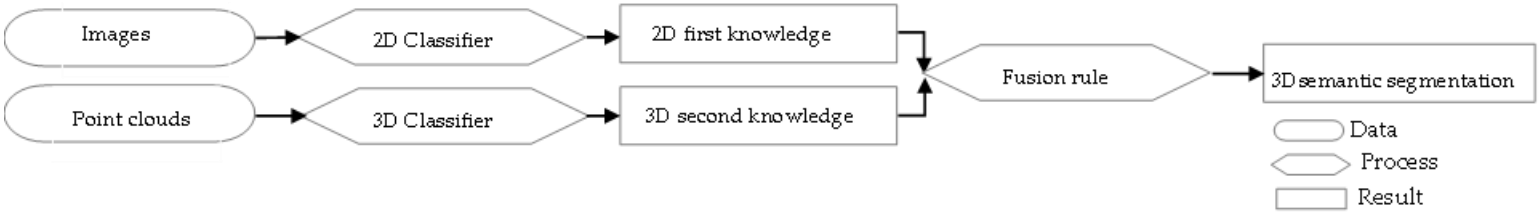


Figure 48. The general workflow of the decision-level fusion approaches.

The performance and limitations of each approach are summarized in Table 16:

Table 16. Performances and limitations of the different fusion approaches.

Fusion approach	Performances	Limitations
Prior-level	-Direct use of semantic information from images -Fast convergence -Low loss function -High classification accuracy.	-Problems of non-overlapping regions and uncertainties -Bit long process
Point-level	-Fast drive -Easy handling -No prior information is required.	High cost - Not able to classify diversified urban contexts Relatively low classification accuracy
Feature-level	-Objective data compression -Retaining enough important information	-Training loss higher -Features may not reflect the real objects.
Decision-level	-Non-interference of the two semantic segmentation processes -Good flexibility -Low-complexity -Learning the representation of independent features is allowed	-Impacted by the shortcomings of both classifiers. - Additional parameters for layers are required More memory requirement

In-depth evaluation of the different fusion levels shows that prior-level fusion is more accurate than point-level, feature-level, and decision-level fusion. This fusion level enables the integration of relevant features with a positive weight in semantic segmentation. It integrates sufficient geometric and spectral information required to enhance semantic segmentation results. Therefore, prior-level fusion was identified as the appropriate fusion level for integrating LiDAR point clouds and airborne imagery. For this reason, we have continued to develop fusion approaches at the prior level.

How to develop a less data-intensive fusion approach for 3D semantic segmentation using optical imagery and 3D point clouds? What solution addresses the issue of incoherence of the semantic classes present in the LiDAR and image datasets at the fusion step?

This question has been addressed in Chapter 2. It emphasizes that research in the field of data fusion for 3D semantic segmentation is moving towards the development of more data-intensive approaches. This includes multispectral images and hyperspectral images, beyond point clouds. However, these approaches require substantial financial and material resources. They also require significant computational memory and time. The necessity to collect the data within minimal time intervals to avoid changes adds to the complexity. Moreover, certain type of information may not significantly contribute to distinguishing urban objects. This requires developing a new fusion approach that uses less additional information but maintains high performance. In this regard, this chapter introduces a novel less data-intensive fusion approach utilizing optical imagery and 3D point clouds (named Plf4SSeg). It consists of two main steps: (1) image classification using the Maximum Likelihood Classifier (MLC). This step allows for the selection of training areas based on the classes present in the LiDAR dataset. (2) Raster values from images are assigned to LiDAR point clouds through prior-level fusion. In other words, classified images were considered as prior knowledge (see

Figure 49). This knowledge was integrated into the advanced Deep Learning technique «RandLaNet", which was optimized for 3D semantic segmentation.

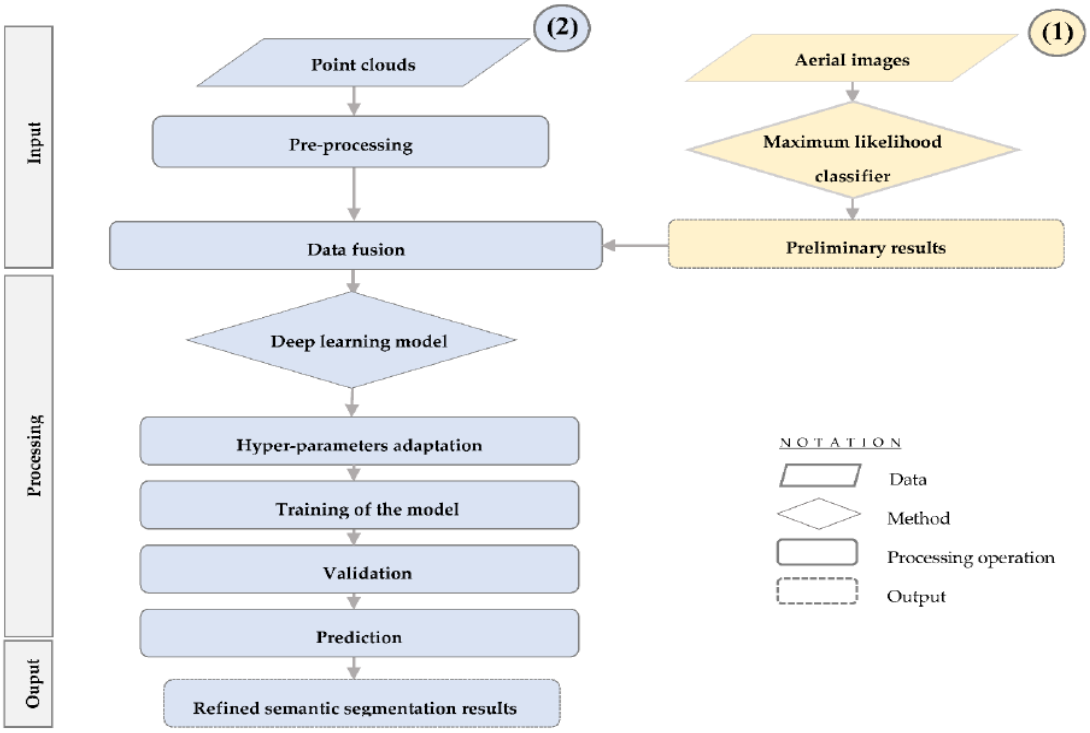


Figure 49. A less data-intensive fusion approach for semantic segmentation of 3D point clouds.

The developed approach offers several advantages. These include improved prediction results and flexibility in the types of images that can be utilized. For example, drone and satellite imagery can serve as alternatives. Moreover, it demonstrates good accuracy compared to non-fusion approaches. The Plf4SSeg approach shows promise for effectively delineating urban objects in ALS point clouds, particularly in large-scale urban environments.

Additionally, chapter 2 presents a solution for resolving semantic class (e.g., cars, trees, power lines) inconsistencies between LiDAR and image datasets during fusion. This solution involves utilizing a standard image classification method (e.g., MLC), where training areas are selected based on the classes present in the LiDAR dataset. That is, we choose the same classes present in the LiDAR dataset. Aligning these classes between the two types of datasets helps ensure coherence. However, this alignment may result in a reduction of semantic details in one of the datasets. For example, merging classes like "low vegetation," "shrub," and "tree" from the LiDAR dataset, in order for them to correspond to the "vegetation" class in the image dataset.

How to automatically and precisely extract the maximum semantic information from large-scale point clouds acquired in an urban environment? How can the performance of each scenario developed be assessed in terms of enhancing knowledge of deep learning techniques?

A) How to automatically and precisely extract the maximum semantic information from large-scale point clouds acquired in an urban environment?

To address this question, chapter 3 develops and investigates three prior-level fusion scenarios. These scenarios are specifically focused on accurately extracting maximum details of urban objects (vegetation, traffic Roads, etc). Two Deep Learning techniques, "KPConv" and "RandLaNet", were utilized. Their parameters were adapted to suit different scenarios. The goal is to identify the scenario that more profoundly enhances the Deep Learning technique's knowledge. In each scenario, specific prior knowledge is integrated into the Deep Learning technique. This includes geometric features, classified images, or categorized geometric data, along with aerial images, alongside the point cloud data. They integrate fused data into the deep learning technique during training for semantic segmentation pipeline. The efficient scenario was determined through evaluations based on several criteria, including qualitative and quantitative results. It demonstrates the capability to extract maximum semantic information with high precision compared to others. This scenario has been named the "Efficient Prior-Level Fusion (Efficient-PLF) approach". The derived efficient approach is presented in Figure 50. The most effective scenario for enhancing the semantic richness of 3D point clouds.

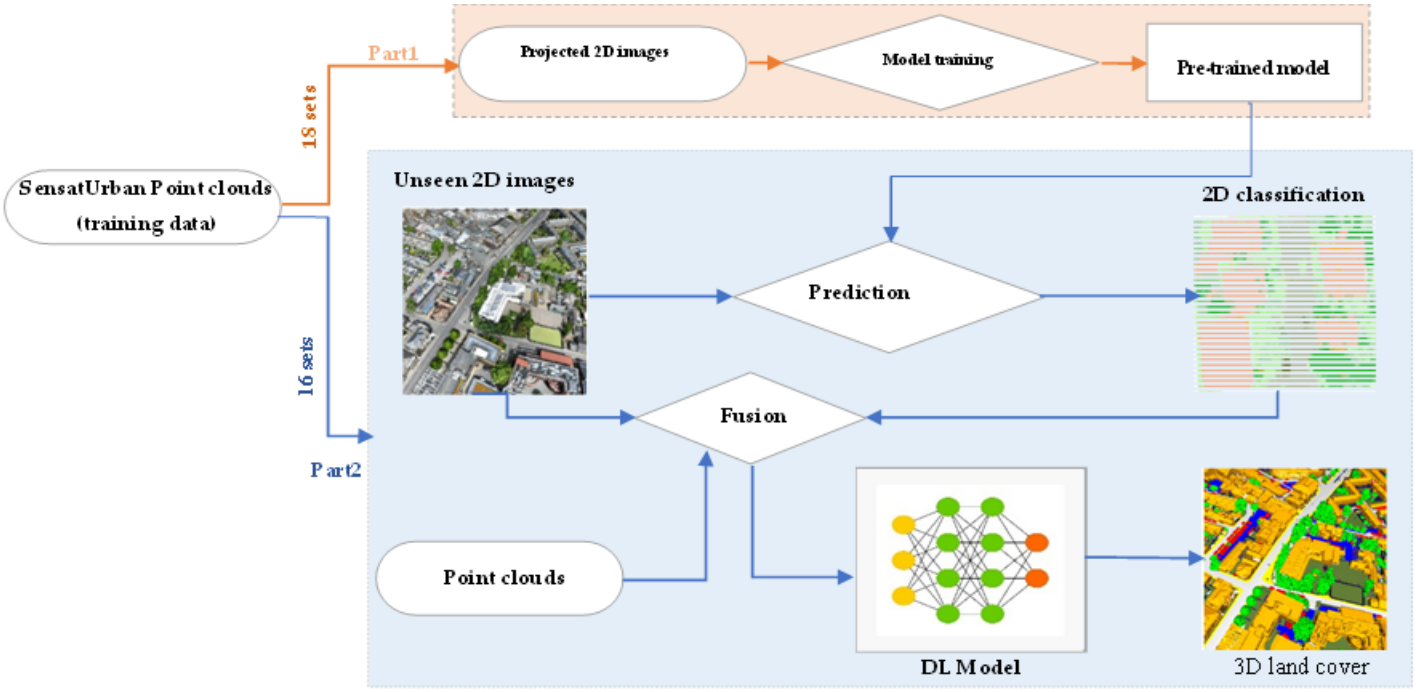


Figure 50. The most effective scenario for enhancing the semantic richness of 3D point clouds.

B) How can the performance of each scenario developed be assessed in terms of enhancing knowledge of deep learning techniques?

In Chapters 2 and 3, the assessment of each scenario's performance involved firstly the calculation of five key metrics: Precision, Recall, F1 score, Intersection over Union (IoU), and Confusion Matrix. Precision to measure the percentage of points correctly identified as positive in 3D semantic segmentation. Recall to assess the proportion of true positives among false negatives and true positives. The F1 score calculates the harmonic mean of Precision and Recall. IoU is used to quantify the percentage of overlap between predicted results and ground truth. Additionally, the confusion matrix was used to evaluate the performance of the Deep Learning technique by assessing the accuracy of its predictions. Each row represents a ground truth, while each column represents a predicted class. The explanation of the evaluation metrics results has led to a comprehensive quantitative analysis. This quantitative analysis was subsequently complemented by a qualitative evaluation. The latter comparing visually the data from the ground truth (actual) and the predicted one (synthetic). More precisely, a detailed study was conducted on the coherence of each semantic class between the results and the ground truth. Another visual comparison of the different scenarios was also conducted by overlaying the results on orthophotos. These complementary evaluation methods have provided a comprehensive understanding of the results. They elucidate how

each scenario has contributed to enhancing the knowledge of deep learning techniques. Therefore, they highlighted the most effective scenario for accurately extracting maximum urban details from point clouds.

How to exploit enriched 3D semantic point clouds to meet the requirements of urban Digital Twins?

To address this main question, two research questions were studied, and their responses are as follows:

Semantic 3D point cloud: An alternative to 3D city model for Digital Twin applications?

This research reflection was thoroughly explored in Chapter 4. Among the conclusions drawn, it was found that an enriched 3D semantic point cloud would enhance the manipulation and interpretation of 3D data, meeting the requirements of Digital Twins. Firstly, maintaining the initial geometric accuracy opens up new possibilities for conducting simulations directly on point clouds, rather than creating surface models. Secondly, point cloud data can be considered as an initial stage of Digital Twins, meeting their basic criteria by replicating all urban objects such as buildings, roads, vegetation, terrain, etc. Lastly, it's worth noting that point clouds can be easily updated over time to reflect changes in the urban environment, while updating a 3D city model may be more complex due to its hierarchical structure.

How to develop an enriched 3D urban model from the semantic segmentation of airborne LiDAR point clouds?

To address this question, chapter 4 (subsection B) presents a workflow for exploiting the results of semantic segmentation in the 3D modeling process. This workflow initially employed an artificial intelligence-based fusion approach for classifying 3D point clouds. This approach combines LiDAR data with corresponding aerial photos for 3D semantic segmentation. It merges three distinct datasets to train the Deep Learning technique to achieve the precise extraction of 3D urban objects. Subsequently, a modeling process was followed to create 3D city models for buildings, vegetation, and road objects.

Building modeling was conducted using the open-source tool "Geoflow". The objective was to generate a realistic 3D representation of buildings. This was achieved by leveraging point cloud, vector data, and modeling functions provided by Geoflow. A JSON file, along with the executable Geoflow, is used to execute the reconstruction from input data.

For road modeling, two processes were developed. The first process utilized a scheme developed in FME, while the second used the open-source tool "3dfier". For road modeling with 3dfier using classified point cloud and PICC data, several steps were followed. PICC data, initially linear, was transformed into polygonal surfaces in QGIS. Precisely, the axis and

edges provided by the PICC data were transformed into surfaces, thereby generating polygons representing the roads. Subsequently, these polygons were employed to execute the lifting process according to the semantics of the road polygons. Following this, modifications were made to the 3dfier lifting options and parameter settings. In essence, 3dfier relies on a basic binary classification of ground and non-ground (as per minimum requirements). However, in our case, an enriched classification was conducted. We integrated this enriched classification to precisely extract the roads class, which 3dfier utilized in the lifting process.

To automatically reconstruct 3D models of trees from airborne point clouds with a LOD2, a three-step process was implemented. This process included classification, segmentation, and modeling. The point cloud belonging to the vegetation class was extracted from the classification results. Individual trees were segmented. These segments served as the data source for constructing 3D tree models.

A part has been added to subsection (B) detailing automatic procedures for modeling other objects extracted from semantic segmentation that are not presented in subsection (B). These objects include TIN, bridges, walls, and cars.

5.2 RESEARCH PERSPECTIVES

This research opens several avenues for further exploration and enhancement. This section presents several research extensions to expand the effectiveness, applicability, and exploitability of enriched semantic segmentation of airborne 3D point clouds. This will pavethe way for advancements in urban Digital Twin research. These reflections include:

- Semantic 3D point cloud: An alternative to 3D city model for Digital Twin applications?
- How to automatically update 3D city models using enriched semantic 3D point clouds?
- Enriched semantic 3D point clouds: An alternative to manual and semi-automatic methods to produce 2D cadastral data?
- How can enriched semantic 3D point clouds be utilized to automatically detect changes in urban environments?
- How do geometric and temporal misalignments affect the quality of the results?
- What is the impact of image resolution on semantic segmentation performance?
- How can the analysis of label corrections enhance the interpretability of RandLaNet?

Semantic 3D point cloud: An alternative to 3D city model for Digital Twin applications.

In Chapter 5, this new research reflection was addressed. Experiments were conducted to estimate the energy potential and perform visibility analysis using only enriched semantic point clouds, without resorting to 3D modeling. The results obtained were satisfactory, but they require validation. Therefore, it' recommended to conduct further tests and experimentations by comparing the results obtained by the solely point cloud-based process and that from the 3D modeling-based process. This requires defining evaluation criteria to assess the results of both processes and comparing them with real data. We suggest testing both processes in multiple urban simulations. For example, estimating the energy potential of buildings, conducting flood studies, etc. This will help highlight the contribution of enriched

semantic point clouds in these types of simulations and confirm if they can replace 3D models in certain urban simulations.

Finally, we also recommend developing new approaches and algorithms that enable the direct simulation of urban environments using enriched semantic point clouds instead of generating 3D models, particularly for sophisticated simulations such as computational fluid dynamics.

How to automatically update 3D city models using enriched semantic 3D point clouds?

3D urban models form the foundation of digital twins, and their regular updates are a critical requirement. Semantic segmentation of LiDAR point clouds can effectively address this need. Its results can be utilized to automatically create enriched 3D urban models, and their updates. That is, the results of semantic segmentation can be used to automatically extract the target class (urban object). Subsequently, it can be integrated into the modeling process for updating. This need opens the way for research into how updates will be conducted. This includes determining whether updates will be made only where changes occur or applied on the entire layer. The Our proposed approaches with regards to semantic segmentation have demonstrated high capabilities in extracting the maximum semantic information from urban environments. Consequently, they can lead to the automatic updating of a large number of 3D city model objects. These new approaches have also resulted in the creation of high-performing trained deep learning models. These models are generative, meaning they can be applied to a wide range of urban contexts. They also adapt to different data qualities.

Figure 51 illustrates an example of results from applying a model trained on an urban context different from the one to which it is applied. Using ideally trained generative models on diverse urban datasets (various urban contexts) offers a valuable opportunity for automating 3D model updates. They enable the generation of updated models quickly and with fewer financial and material resources. Moreover, trained models offer a reproducible and readily deployable process that integrates new LiDAR acquisitions and corresponding aerial photos. This reusable characteristic is significant for continuously updating 3D models due to evolving on-the-ground situations.

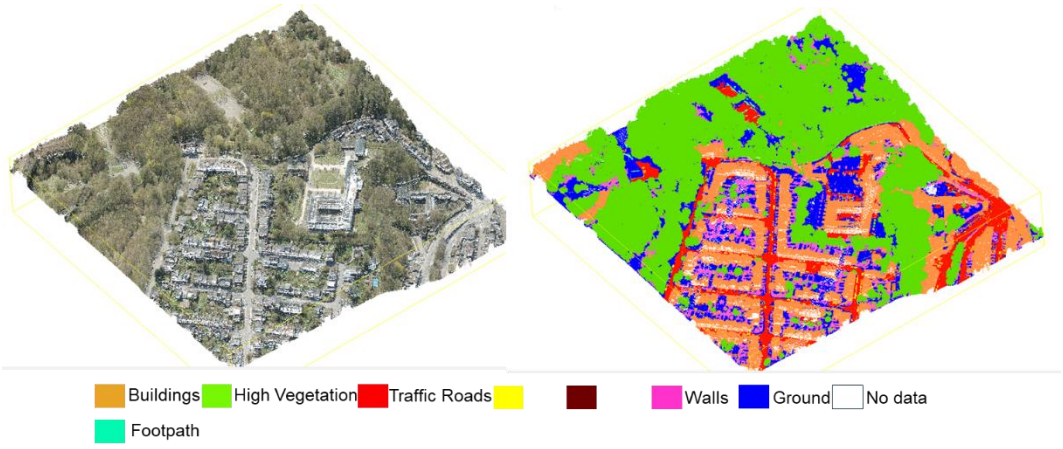


Figure 51. An example of results from applying a model trained.

Enriched semantic 3D point clouds: An alternative to manual and semi-automatic methods to produce 2D cadastral data?

The need for 2D vector data is crucial in several Geographic Information Systems (GIS) applications. Specifically, it is essential in spatial and multicriteria analysis, and other related tasks. In this context, enriched semantic point clouds can offer a good opportunity to meet this need. They provide the possibility of extracting the 2D vector layer related to each class in a precise and automatic manner (see Figure 52).

The extracted vector layer can be used to calculate spatial statistics for quantifying the characteristics of geographic entities (e.g., building area). Additionally, this data can serve as an alternative to cadastral data for modeling objects from point clouds. The extracted footprints also find application in updating 2D cadastral data and their verification. However, precisely segmenting vector data, such as buildings, to assign a specific identifier to each building still poses a challenge. This highlights the need for research to develop new segmentation algorithms to meet precision requirements. Furthermore, extracted vector data can be utilized to create thematic maps representing spatial data, such as population distribution or road networks.

In conclusion, vector data extracted from semantic point clouds can be a powerful tool for spatial analysis and decision-making across various applications. Their accuracy, flexibility, and interoperability can indeed make them an essential component of GIS.

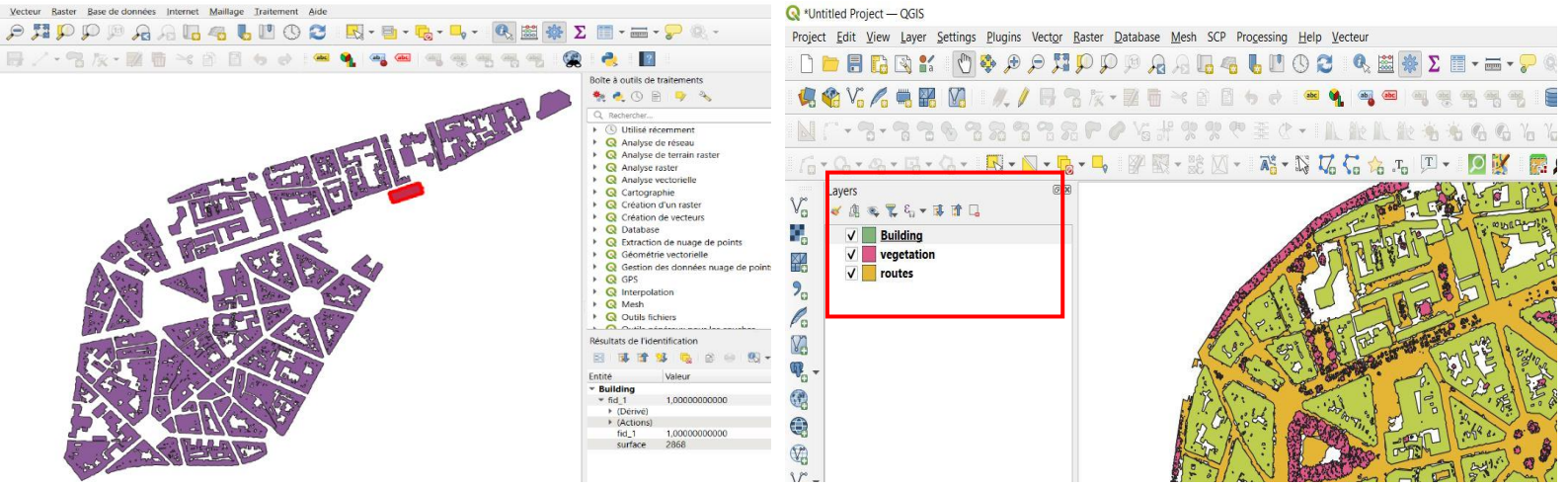


Figure 52. Examples of vector layers (buildings, vegetation, and roads) obtained from the results of semantic segmentation.

How can enriched semantic 3D point clouds be utilized to automatically detect changes in urban environments?

Change detection from 3D LiDAR point clouds is a versatile technique. It has applications in environmental monitoring, risk management, civil engineering, and archaeological research. It can monitor changes in the landscape and identify potential risks such as landslides or floods. Additionally, it assesses structural integrity in civil engineering projects and aids archaeological research by detecting alterations in sites. However, the effectiveness of change detection from point clouds relies on the precision of semantic point clouds. In other words, the better the quality of semantic segmentation results, the higher the quality of change detection outcomes. The focus on the precise extraction of urban objects in this thesis has led to the development of new semantic segmentation approaches. These approaches demonstrate satisfactory performance in accurately extracting detailed semantic information. Therefore, they can have the potential to elevate the quality of change detection. They can help to accurately detect changes in urban areas, aligning more closely with ground reality. These approaches can serve as operational methodologies, providing a reproducible process for identifying changes between the LiDAR point clouds acquired at different times. This reusability is particularly valuable for tracking and detecting changes on the ground, given the dynamic and rapidly evolving nature of the urban environment.

How do geometric and temporal misalignments affect the quality of the results?

The approach developed in this thesis relies on a key assumption: the datasets used, specifically LiDAR point clouds and images, are 1) perfectly aligned geometrically and 2) acquired under similar (or close) temporal conditions. However, in practice, it is common for these datasets to be acquired at different times, sometimes months or even years apart, which can introduce significant biases. Differences in sensor orientations, inaccuracies in ortho-rectification processes, and terrain changes due to natural or anthropogenic factors (e.g., appearance, disappearance, or modification of objects) further complicate this assumption. These limitations should be clearly highlighted, not only to define the conditions for applying the approach but also to encourage future studies addressing these challenges. For instance, developing algorithms capable of handling these geometric and temporal disparities could represent a significant advancement.

What is the impact of image resolution on semantic segmentation performance?

Another implicit assumption in this work concerns the use of high-resolution imagery. Although this concept is frequently mentioned, the resolution ranges suitable for the proposed approach are not clearly defined. This creates uncertainty about the performance of the approach when the resolution of the input data varies. A more detailed analysis, testing the approach across different resolutions, could help identify critical thresholds to avoid performance degradation or class confusion. While such an analysis is beyond the scope of this thesis, it paves the way for future research to better understand the optimal resolution ranges for applying the proposed approach in real-world contexts.

How can the analysis of label corrections enhance the interpretability of RandLaNet?

A promising direction for future research involves analyzing the label corrections made by the RandLaNet network. Examining the spatial distribution of label reversals (instances where the network corrects or modifies the initial pseudo-labels) could offer valuable insights into the model’s robustness and limitations. This analysis could be further enriched with explainability techniques, such as Shapley values, to better understand the influence of each input feature on the network's decisions. Such research could address the growing need for interpretability in the RandLaNet network and enhance its reliability for practical applications.

List of publications

Z. BALLOUCH, R. HAJJI, M. ETTARID, “The contribution of Deep Learning to the semantic segmentation of 3D point-clouds in urban areas”, IEEE International conference of Moroccan Geomatics (Morgeo), 2020.

Zouhair Ballouch, Rafika Hajji, “Semantic Segmentation of Airborne LiDAR Data for the Development of an Urban 3D Model”, Building Information Modeling for a Smart and Sustainable Urban Space, january 2021.https://doi.org/10.1002/9781119885474.ch7

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.

Ballouch, Zouhair, Rafika Hajji, Florent Poux, Abderrazzaq Kharroubi, and Roland Billen. 2022. "A Prior Level Fusion Approach for the Semantic Segmentation of 3D Point Clouds Using Deep Learning" Remote Sensing 14, no. 14: 3415. https://doi.org/10.3390/rs14143415

Kharroubi, Abderrazzaq, Florent Poux, Zouhair Ballouch, Rafika Hajji, and Roland Billen. 2022. "Three-Dimensional Change Detection Using Point Clouds: A Review" Geomatics 2, no. 4: 457-485. https://doi.org/10.3390/geomatics2040025.

Ballouch, Zouhair, Rafika Hajji, Abderrazzaq Kharroubi, Florent Poux, and Roland Billen. 2024. "Investigating Prior-Level Fusion Approaches for Enriched Semantic Segmentation of Urban LiDAR Point Clouds" Remote Sensing 16, no. 2: 329. https://doi.org/10.3390/rs16020329

Imane Jeddoub And Zouhair Ballouch (equal contribution), Rafika Hajji, Roland Billen. 2023. "Enriched semantic 3D point clouds: An alternative to 3D city models for Digital Twin for Cities? ". 3D GeoInfo Conference.

Kharroubi, Abderrazzaq, Zouhair Ballouch, Rafika Hajji, Anass Yarroudh, and Roland Billen. 2024. "Multi-Context Point Cloud Dataset and Machine Learning for Railway Semantic Segmentation" Infrastructures 9, no. 4: 71.https://doi.org/10.3390/infrastructures9040071

Ballouch, Z., Jeddoub, I., Hajji, R., Kasprzyk, J.-P., and Billen, R.: Towards a Digital Twin of Liege: The Core 3D Model based on Semantic Segmentation and Automated Modeling of LiDAR Point Clouds, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., X-4/W4-2024, 13–20, https://doi.org/10.5194/isprs-annals-X-4-W4-2024-13-2024, 2024.

Curriculum vitae