

Synthesis Project - Geomatics

Integrating Gaussian Splatting with Semantic Labels for Heritage BIM

Marieke van Arnhem, Qiaorui Yang, Shawn Tew,
Xiaduo Zhao, Walter Kahn

November 2024



Integrating Gaussian Splatting with Semantic Labels for Heritage BIM

Marieke van Arnhem, Qiaorui Yang, Shawn Tew, Xiaduo Zhao, Walter Kahn

November 2024

Marieke van Arnhem, Qiaorui Yang, Shawn Tew, Xiaduo Zhao, Walter Kahn: *Integrating Gaussian Splatting with Semantic Labels for Heritage BIM* (2024)

© This work is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The work in this research was carried out in the:



Geo-Database Management Center
Delft University of Technology

In collaboration and with supervision from:



3D GeoData Academy
Florent Poux

Supervisors: Ir. E. Verbree
Clients: Dr. Florent Poux
MSc Ying Wen Yu

Abstract

In recent years, the need for heritage preservation and reconstruction has become evident as many mature buildings face the risk of deterioration, damage or loss due to factors such as urban development, environmental weathering as well as outdated infrastructure. This urgency has created surges of significant interest to find sustainable methods of heritage preservation. The rise of emerging digital technologies has introduced a multitude of innovative methods for storing, analysing, and showcasing building data. Technologies such as 3D LiDAR scanning, and Building Information Modelling enable detailed documentation and virtual exploration of heritage sites, while digital databases and archives facilitate the easy access and use of historical records. This project will attempt to address a new method of heritage preservation by using Gaussian Splatting in conjunction with segmentation methods to create a visually accurate model while also incorporating semantic labels. The project's outcomes and source code used can be found at <https://github.com/ShawnTew/Synthesis-Project-Group-4>.

Key Words: Heritage, Gaussian Splatting, Building Information Modelling, Preservation, Segmentation, 3D modelling

Acknowledgements

We would like to express our sincere gratitude to our three supervisors who played pivotal roles in guiding this research to completion.

To Dr. Florent Poux, our external supervisor, we extend special thanks for sharing his extensive technical expertise, particularly in implementing the SAM model for point cloud classification and developing ray casting solutions for back-side segmentation. His insights helped sharpen our research focus and provided clarity whenever we faced technical challenges or needed to choose between multiple potential solutions.

We are deeply grateful to Ir. Edward Verbree, our primary supervisor, for his dedicated mentorship throughout this project. Our frequent meetings were invaluable in refining our research direction. His enthusiasm for exploring new ideas and creative approaches, combined with his thoughtful guidance in academic writing, significantly enhanced the quality of our work. His ability to think alongside us and challenge our assumptions pushed us to achieve better results.

Special thanks to Phd candidate Yingwen Yu, whose expertise in BIM modeling and heritage building preservation broadened our perspective significantly. Her creation of the BIM model provided us with an essential ground truth for our validation process, and her insights into heritage preservation added valuable context to our technical work.

Their combined guidance and support were instrumental in shaping this research and helping us navigate the complexities of integrating modern technological solutions with heritage preservation needs.

Contents

1	Introduction	8
1.1	Research Sub-question	9
1.2	Client Requirements	11
2	Related Work	12
2.1	Heritage Building Information Modeling	12
2.2	3D Data Acquisition Methods for Heritage Buildings	13
2.3	Data pre-Processing and Management	13
2.4	Gaussian Splatting	14
2.4.1	Gaussian Splatting Overview	14
2.4.2	Neural Radiance Field	15
2.5	Gaussian splatting Post Processing	16
2.6	Gaussian Splatted Point Cloud Semantic Classification	17
2.7	Validation	18
2.8	Intermediate Conclusion	18
3	Methodology	19
3.1	Data Acquisition	20
3.2	Data Cleaning and Management	20
3.2.1	COLMAP Processing	20
3.2.2	RealityCapture Processing	21
3.2.3	Image Alignment and Sparse Point Cloud Generation	21
3.2.4	Exporting Camera Poses	21
3.2.5	Additional Data Preparation	21
3.2.6	File Format Management	21
3.2.7	Coordinated and Alignment	22
3.3	Gaussian Splatting	23
3.3.1	Inria Tool	23
3.3.2	Jawset Postshot	23
3.3.3	PolyCam	24
3.3.4	Scaniverse	25
3.4	Segmentation	25
3.4.1	Attributes of a Point Cloud	25
3.4.2	Attributes of Gaussian Splat	26
3.4.3	Segment Anything Model (SAM)	26
3.4.4	k-Means Clustering	28
3.4.5	Gaussian Mixture	28
3.4.6	Density Based Clustering	28
3.5	Semantic Labeling	29
3.6	Visualisation and Assessment	29
3.6.1	Label Quality	29
3.6.2	Visual Clarity	30
3.7	Intermediate Conclusion	31
4	Results	32
4.1	Data Collection	32
4.1.1	GeoSLAM	32
4.1.2	2D Photos	33

4.2	Gaussian Splatting	35
4.2.1	Polycam	35
4.2.2	Inria tool	35
4.2.3	Postshot	36
4.2.4	Level of Detail 3-Dimensional Gaussian Splatting (LOD 3DGS)	37
4.2.5	Edwardian Splatting	39
4.3	Segmentation	41
4.3.1	Point Cloud	41
4.3.2	Gaussian Splats	43
4.4	Intermediate Conclusion	46
4.5	Validation	46
4.5.1	Visual Validation	46
4.5.2	Accuracy Validation	49
4.6	Intermediate Conclusion	51
5	Discussion of Limitations	52
5.1	Challenges & Limitations	52
5.1.1	No BIM for Complex Buildings	52
5.1.2	Limited Range of Scaniverse	52
5.1.3	Limited Computational Power	53
5.1.4	Using Video as Input	53
5.1.5	GeoSLAM to Splat	53
5.1.6	Planar Surface	53
5.1.7	Gaussian Splatting not Deterministic	54
5.1.8	Alignment	54
5.1.9	Quality of Collection Methods	54
5.1.10	Using Splats for Segmentation	54
5.1.11	SfM from Gaussian Splatting Not Dense	55
6	Conclusion	56
7	Further Research	57
8	Task Distribution	59
8.1	Scoping/ Managing Expectanations: MoSCoW	59

List of Figures

1	Example of a Gaussian splatted Church (Best viewed in Adobe Acrobat).	9
2	Sparse Point Cloud	15
3	Gaussian Splat Example	15
4	Overall Synthesis Project Workflow	19
5	Spherical Image of Point Cloud	27
6	Segment Anything Model Mask Projected onto the Spherical Image	28
7	Panoramic Photo from the GeoSLAM	33
8	GeoSLAM Generated Point Cloud of Bouwpub's front-face	33
10	2D Images of Bouwpub's front-face	34
11	(a) Inria tool Gaussian splatting results	37
12	(b) Polycam Gaussian splatting results	37
13	(c) Postshot Gaussian splatting results	37
14	LOD3DGS Workflow	38
15	LiDAR Points & Images Results	39
16	Overview of the 3D reconstruction in Blender (Best viewed in Adobe Acrobat)	40
17	The resulting Edwardian splat (Best viewed in Adobe Acrobat)	40
18	Segmented GeoSLAM Point Cloud Results.	43
19	Segmented Gaussian splatted point cloud results.	45
20	Image of GeoSLAM Point Cloud in Greyscale	47
21	Laplacian Filter Applied to GeoSLAM Point Cloud	47
22	Image of Gaussian Splatted Point Cloud in Greyscale	48
23	Laplacian Filter Applied to Image of Gaussian Splatted Point Cloud	48
24	Heat Map for Precision Scores on all Methods.	49
25	Heat Map for Recall Scores on all Methods.	50
26	Barplot with Mean Precision and Recall with Error Bars on all Methods.	51

1 Introduction

In 2019 a huge fire broke out in the Notre Dame in Paris. For the reconstruction process of the Notre Dame a 3D Building Information Model (BIM) model was used based on a point cloud. It helped the process by showing the cathedral's structure, something that was missing from the old 2D drawings which were based on clay models from that era. But the model also helped by facilitating efficient site logistics, helping to plan the placement of cranes and worker access, while aiding in construction sequencing and risk management. Overall, these technologies accelerated the reconstruction process. (Fausto-Robledo [2021])

The example above shows the importance of capturing a 3D model of a heritage building such as the Notre Dame. Heritage buildings are significant parts of our history and culture. They include buildings, structures, objects, and places that have historical, artistic, societal, or architectural significance (Al-Sakkaf et al. [2020]). Documenting these buildings in 3D helps preserve them for future generations while keeping their artistic and architectural features intact (Khalil et al. [2021]). Additionally, 3D models allows the user to explore and understand the building's history more thoroughly (Khalil et al. [2021]).

The challenges in creating Heritage BIM include the need for new parametric modeling approaches, broader construction categories, specialised tools for heritage, enhanced information accuracy, and flexible representation of historical transformations (Cursi et al. [2022]). By using Gaussian ellipsoids, it enables rapid 3D reconstruction and real-time rendering, thus facilitating low-cost, high-quality content creation (Wu et al. [2024]), which could make it possible to capture the unique features of heritage buildings with high quality. Combining the Gaussian splats with labels, representing their building part or other features, could be used as a HBIM in a certain way. So creating a Smart Gaussian Splatted point cloud. This leads to the following research question: *To what extent can a segmented Gaussian splatted point cloud support the heritage Building Information Modeling workflow?*.

First, some key words are explained. BIM is a digital representation of the physical and functional characteristics of a facility, serving as a shared knowledge resource for decision-making throughout the lifecycle of a building, from design and construction to operation and maintenance (Internationl [2013]). Gaussian splatting is a 3D technique that builds a scene using small, overlapping shapes (Gaussian functions) to create a realistic view, refining the scene to look as close to real life as possible (Wu et al. [2024]). Figure 1 shows an example of a Gaussian splatted building. A segmented Gaussian splatted point cloud means the splats are segmented into clusters.

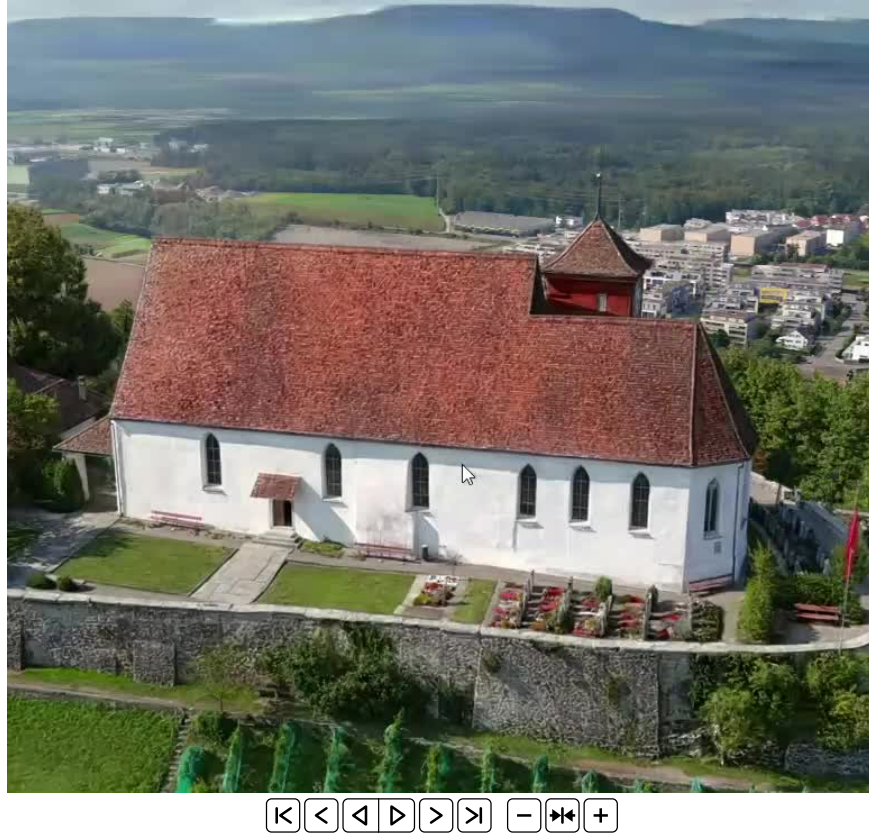


Figure 1: Example of a Gaussian splatted Church (Best viewed in Adobe Acrobat).

This report is structured as follows. The related work is covered in Chapter 2. Chapter 3 describes the methodology, it first explains the acquisition and preparation of the data, following the use of Gaussian splatting, segmentation methods and the labeling. The section ends with describing a visual- and label assessment. After the methods, the results of the data collection, Gaussian splatting methods and the segmentation methods are presented in Chapter 4. There will be an intermediate conclusion in that section followed by a chapter about the validation of the results. Chapter 5 explains the discussion: the limitations of this research and challenges are displayed here. And the final part, Chapter 6, is the conclusion of this research. Which will also describe further research.

1.1 Research Sub-question

Main research question: To what extent can a segmented Gaussian splatted point cloud support the heritage Building Information Modeling workflow?.

To address this research questions, the below five sub-questions give guidance. Each one targets a specific topic that helps clarify how Gaussian splatting and semantic labels might improve workflows in heritage BIM. The sub-questions are as follows:

What are the current digital preservation methods for heritage buildings, and how effective are they?

This section will start by explaining what heritage buildings are and why it is important to capture them using 3D models. It will explore what makes a 3D model of a heritage building suitable for different purposes. Next, it will review the current digital preservation methods, highlighting their strengths and weaknesses. By looking at these methods, a benchmark can be set to compare the method developed in this research with existing ones. Additionally, understanding the quality control of these methods will help in evaluating and testing the research approach.

How do different data collection devices (e.g., GeoSLAM, stereo cameras, phone cameras) impact the quality and usability of point clouds for heritage BIM restoration?

The data collection results may vary in terms of precision and capability as this research attempts to use different devices. High end devices such as the GeoSLAM are used in professional cases and can produce high-resolution and accurate point clouds. In contrast, stereo cameras as well as smart phones may offer greater accessibility but will end up sacrificing accuracy and detail, specifically when attempting to capture intricate details at larger distances.

By comparing the outputs of these devices, the project will assess how the difference in resolution, precision and depth capture can impact the overall quality of the output point clouds. This will help determine whether more accessible devices can be reliably used for capturing heritage buildings.

Does Gaussian splatting improve point cloud clarity and segmentation results compare to traditional point cloud rendering and original parameters?

The traditional method of representing point clouds is point cloud rendering, where each point is typically rendered with colour. However, these points may seem disconnected typically in regions of sparse point distribution. In contrast, Gaussian splatting treats each point as the center of a Gaussian distribution, allowing for smoother transition between points this may help improve overall visual clarity and reduce harsh edges or gaps seen in traditional methods (Cai et al. [2024]).

Furthermore, adding layers of information such as Gaussian splatting parameters to the dataset may enhance the categorical clarity of point clouds generated for classification such as the spherical harmonic features. Through the incorporation of blind classification, this research will aim to semantically label the point clouds according to specific heritage BIM standards. This additional data will not only make the smart point cloud a visual tool but also a source of semantic information allowing for a wider range of use and accessibility for all possible stakeholders.(Poux et al. [2017]).

How can the appearance and precision of the smart point clouds be assessed?

To determine if a segmented Gaussian splat can effectively function as part of Heritage Building Information Modeling (HBIM), the clusters and the visual quality of the Gaussian splats will be evaluated. The assessment tries to find metrics or methods to see if the points are clustered well and how well these splats represent the actual structure of the heritage building.

How can the potentially improved point cloud be applied for HBIM preservation?

Improving clarity and classification performance of Gaussian Splatted point clouds can be significant in the context of preservation and restoration for in the context of HBIM. By smoothing transitions between points as well as reducing noise, Gaussian splatting can ease the visualisation of heritage buildings clearly. This is crucial in the planning of restoration

or preservation as it allows for conservators or architects to have a detailed understanding of either the structural elements or design of a building (Cai et al. [2024]).

In addition, the interoperability of the produced smart point clouds is increased as it may be integrated with BIM platforms to streamline workflows. The semantics added through classification can aid in the conversion to parametric models which are important when representing complex geometries of heritage structures (Thomson and Boehm [2015]).

1.2 Client Requirements

This research is in cooperation with two clients. The first client is Ying Wen Yu, a PhD student at TU Delft. Ying Wen Yu, aims to use current technology and new digital innovations to address and improve the process of the challenges in the infrastructure around 3D models of heritage buildings. Below two of these challenges are mentioned, which will be useful for this project.

1. One of the challenges mentioned in the research proposal of Ying Wen is the need for high quality 3D models which is essential for the preservation, research, and educational purposes of the building (Megahed [2015]).
2. And another challenge is about effectively integrating with existing archives and datasets while ensuring that the 3D models remain compatible, easily accessible, and retain their integrity and detail (Shafique et al. [2020]).

This research is to help Ying Wen in the two mentioned challenges by using a Gaussian splatting pipeline combined with semantic label —a novel method that offers improved accuracy in point cloud reconstruction and classification. This approach aims to improve accuracy and enhance the workflow for Heritage Building Information Modelling.

The secondary client is Florent Poux from 3D Geodata Academy. Florent is a geo data specialist and has a large interest in point clouds as well as emerging digital technologies in point cloud generation and processing. The team will provide Florent Poux with insights and information on this pipeline, enabling him to share these advancements on his platform about 3D graphics, point clouds automation. This collaboration will help others implement similar techniques and address the second challenge of integrating new data with existing datasets while maintaining accuracy and detail.

2 Related Work

In this section, there are existing research related to Heritage Building Information Modeling (HBIM) in section 2.1 and 3D data acquisition for heritage conservation in section 2.2. This section explore various methods for capturing and managing 3D data, including photogrammetry, laser scanning, and Scan-to-BIM workflows, highlighting their significance in documenting architectural heritage in section 2.3. This section further examine Gaussian splatting and neural radiance field techniques, detailing their applications in enhancing HBIM reconstruction and analysis in section 2.4 2.5 2.6. Finally, we discuss validation methods in section 2.7 for HBIM and point cloud segmentation, assessing approaches to ensure accuracy and quality in heritage documentation.

2.1 Heritage Building Information Modeling

Building Information Models (BIM) are advanced modeling processes that generates, stores, and manages data throughout a building’s lifecycle, from design to demolition (Penjor et al. [2024]). It provides a digital representation of a building’s physical and functional characteristics, serving as a reliable source for decision-making (Martinelli et al. [2022]). A BIM is centered around a 3D model with a combination of the measurable, numerical information and the descriptive information of the building (Martinelli et al. [2022]) (Penjor et al. [2024]).

A significant challenge in conserving many heritage buildings is the absence of original construction designs, which are crucial for effective restoration and reconstruction efforts (Acierno et al. [2017]) (López et al. [2018]). To overcome this, technologies such as 3D scanning and photogrammetry are used to capture detailed models of the structures. These models are then processed and refined for integration into BIM platforms (López et al. [2018]). With the help of various software tools such as Autodesk Revit, ArchiCAD, Rhino, and Dynamo it is possible to convert these point cloud datasets and orthographic photos into a rich 3D model (López et al. [2018]). The main difference between HBIM and BIM is the intended purpose: BIM focuses on new construction, while HBIM targets existing buildings for conservation, restoration, and rehabilitation (López et al. [2018]). Another big difference is the huge popularity of BIM models and HBIM is still an emerging approach (López et al. [2018]).

The workflow of HBIM still contains challenges in every step. During data acquisition, various data sources must be integrated, often facing challenges such as inconsistencies in formatting and a lack of standardisation (López et al. [2018]). Besides that, the process of converting point clouds to parametric models has its limitations: library limitations, manual modeling and lack of detailed information (Radanovic et al. [2020]). Semantic interoperability is hindered by the absence of standardised ontologies (López et al. [2018]). The modeling process also struggles with the complexity of certain features (López et al. [2018]). And lastly, the quality assessment, now point clouds are a lot of times compared to models which are not the ground truth (Radanovic et al. [2020]).

2.2 3D Data Acquisition Methods for Heritage Buildings

In recent years, 3D data acquisition methods such as digital photogrammetry, laser scanning, and the Scan-to-BIM methodology have become fundamental for documenting and preserving architectural heritage. These technologies enable high-resolution, accurate representations of heritage structures, allowing for detailed digital records that support conservation efforts. By capturing complex geometries and textures, they facilitate monitoring, restoration, and impact assessment, thus offering invaluable tools for maintaining and studying historical buildings with precision and efficiency.

According to the study of [Moullou et al. \[2023\]](#), digital photogrammetry and laser scanning have become core tools for data acquisition and heritage management in archaeology. These methods are categorised into image-based, non-image-based, and hybrid techniques, each with distinct advantages. Image-based methods, such as UAV (Unmanned Aerial Vehicle, commonly known as a drone) photogrammetry, provide high-resolution 3D models with realistic textures; non-image-based methods, like terrestrial laser scanning, capture precise geometric data rapidly; and hybrid methods combine both strengths to achieve accurate geometry and detailed textures. With the advancement of technology, low-cost equipment and streamlined processes have made high-quality 3D modeling more accessible, offering new possibilities for the digital documentation and preservation of cultural heritage.

Similarly, the Scan-to-BIM methodology offers substantial benefits for architectural heritage conservation. Based on [Rocha et al. \[2020\]](#)'s research, by integrating photogrammetry and laser scanning into the BIM workflow, it enables accurate documentation of historical structures in their current state, which is essential for preservation efforts. This digital approach enhances the preservation of architectural heritage, making it possible to maintain, restore, and study historical buildings with greater precision and care.

As demonstrated by [Kerbl et al. \[2023\]](#), high-quality 3D scene representations can now be generated from sparse point clouds produced during camera calibration, such as Structure from Motion (SfM) data. This approach maintains continuous volumetric radiance fields for optimised scene capture.

2.3 Data pre-Processing and Management

The proliferation of LiDAR and photogrammetry in heritage documentation has led to challenges in managing and processing large point cloud datasets. Several authors have highlighted key issues in this area.

[Rocha et al. \[2020\]](#) emphasise the importance of data cleaning and preparation of heritage point clouds prior to analysis. In their case study, they found that pre-processing steps like noise reduction and removal of unnecessary elements resulted in an optimised point cloud with 12.6% fewer points, reducing computational requirements for subsequent modeling. They used Autodesk Recap for this cleanup process, removing artifacts, extraneous objects, and internal furniture. The management and organisation of large heritage point cloud datasets is another significant challenge. [Moullou et al. \[2023\]](#) note that while abundant point cloud data is now available for heritage sites, effectively managing this information remains difficult. They argue this has created a need for specialists focused on data management and IT infrastructure in heritage studies.

Proper data storage is critical for long-term preservation and reuse of heritage point cloud data. [Rocha et al. \[2020\]](#) highlight that the shift to digital acquisition has moved much of the data management burden from the field to the office, necessitating robust storage and organisation systems. [Moullou et al. \[2023\]](#) stress the importance of metadata and paradata in managing heritage point cloud data, arguing that this contextual information is crucial for ensuring research transparency and reproducibility.

2.4 Gaussian Splatting

This report will use Gaussian Splatting as a core component of HBIM reconstruction and preservation, this is crucial when capturing and documenting culturally significant structures. The application of Gaussian splatting in this context aims to improve the efficiency and quality of HBIM applications through smart point clouds.

2.4.1 Gaussian Splatting Overview

Gaussian splatting is a relatively new technique used to create a smooth continuous visual representation from discrete point clouds by converting each point with a Gaussian kernel. This method addresses some limitations of traditional point based rendering such producing hard edges and non-continuous visual representation.

This process begins with point conversion, where each point in the cloud is transformed into a Gaussian function, this allows each point to represent more than a fixed point position. Each Gaussian function provides information about the points position and the surrounding space. Following this, key parameters are defined for each Gaussian, including, the point's mean which determines its central location. The covariance matrix is used to define the shape, size, and orientation of the Gaussian. The colours of the Gaussian splat are determined by the spherical harmonic variables. These spherical harmonic attributes store light information and changes the colour based on the orientation of the splat relative to the light source. Opacity will also be defined as its own attribute, giving the points a transparency value. The overall parameters will influence the shape, appearance, and distribution of each Gaussian in space. This leads to the kernel definition stage where the points are converted into splats formed by the parameters to create smoother groups from a discrete point cloud that can be generated from images with SfM. ([Bao et al. \[2024\]](#)).

As the Gaussian splats are defined in the previous stage, they can now be projected onto a 2D image plane. This projection translates 3D point cloud data into a 2D space for rendering where their continuity allows for a smooth blending process, compared to harder edges seen in discrete point based rendering. Furthermore, point cloud rendering cannot create regions in the point cloud from unreconstructed areas in the SfM process. The continuity of the splats allows a blending that process takes overlapping Gaussians to create smooth transitions between points ([Kerbl et al. \[2023\]](#)).

Finally, the splats are rendered to generate a final, high quality 3D representation. By combining the Gaussian splats into an overall image, the method enables fine detail in high-density point regions. This makes Gaussian splatting well suited for complex scenes. This ability to handle point clouds efficiently can create a variety of real time applications in HBIM with high quality visualisations ([Wu et al. \[2024\]](#)). An example of this can be seen below in figure 2 and figure 3 from [Kerbl et al. \[2023\]](#).



Figure 2: Sparse Point Cloud



Figure 3: Gaussian Splat Example

2.4.2 Neural Radiance Field

NeRF (Neural Radiance Fields) is a method of creating high quality photorealistic views of complex 3D scenes. The overall NeRF process utilises a continuous volumetric scene function from a sparse set of input images with known camera poses. The scene is then represented using a fully connected deep neural network that takes 5D coordinates as input (3D spatial location as well as 2D viewing direction). The output of the network includes both volume density and the emitted radiance at each point in space. Queries are then

used on the 5D coordinates in combination with rendering techniques, a synthesised highly accurate view of the scene can be created. [Mildenhall et al. \[2020\]](#)

NeRFs are traditionally a computationally expensive process. However, when used in combination with Gaussian Splatting, the NeRF approach for visualisation becomes more efficient. Gaussian splats in this case, replace the neural networks reliance on volumetric ray marching with 3D Gaussian splats, which are representing the scene as discrete Gaussians that are directly projected onto the image plane. ([Kerbl et al. \[2023\]](#)). This is important as the readily available tools that will be used later in the methodology of this paper make use of this process.

2.5 Gaussian splatting Post Processing

Before point cloud segmentation, processing steps are crucial for improving segmentation accuracy and efficiency. In existing research, a series of processing steps are typically applied to point cloud data before segmentation. These steps include denoising, filtering, and downsampling, aimed at enhancing point cloud quality and reducing the computational burden of subsequent processing. For example, [Rusu and Cousins \[2011\]](#) proposed a statistical outlier removal algorithm that effectively eliminates noise points from point clouds. [Wang and Shan \[2009\]](#) developed a voxel grid-based downsampling method that significantly reduces data volume while preserving the main features of the point cloud. Furthermore, [Weinmann et al. \[2014\]](#) introduced a multi-scale feature extraction method that enhances the expressiveness of point clouds by computing local geometric features, laying a foundation for subsequent segmentation tasks.

In indoor point cloud scenarios, identifying transparent elements (such as glass doors and windows) is an important and challenging task. Currently, there are two main approaches to address this issue. The first approach is based on geometric and reflective properties. For instance, [Koch et al. \[2016\]](#) proposed a method combining laser reflection intensity and geometric features to detect glass surfaces. The second approach leverages deep learning techniques, which typically rely on specialised datasets containing annotated transparent elements. For example, [Funk et al. \[2018\]](#) constructed a large-scale indoor scene point cloud dataset with labeled transparent elements, including glass doors and windows, and developed a deep learning model to identify these elements based on this dataset. Another emerging method utilises multi-modal data fusion, such as combining RGB images with point cloud data. Research by [Yang et al. \[2016\]](#) demonstrates that this approach can significantly improve the accuracy of transparent object recognition.

However, directly applying Gaussian splatting techniques for point cloud segmentation is a relatively new research direction. This method first requires exporting Gaussian splats into standard point cloud formats, such as PLY files ([Jurski \[2024\]](#)). According to [Kerbl et al. \[2021\]](#), each Gaussian splat contains the following feature information:

- Position: represented as a 3D vector
- Rotation: represented as a quaternion
- Scale: represented as a 3D vector
- Opacity: represented as a real number between 0 and 1
- Direction-dependent color: represented as spherical harmonics

This representation method provides rich geometric and visual information for point cloud segmentation tasks, potentially improving the accuracy and robustness of segmentation. But the splat dataset for classification and segmentation is still not yet proposed.

2.6 Gaussian Splatted Point Cloud Semantic Classification

This section lays out the existing research that has been done into how pointclouds in general can be enhanced using deep learning techniques.

Semantic SLAM Systems Using Gaussian Splatting [Zhu et al. \[2024\]](#) introduced SemGauss-SLAM, which integrates semantic feature embedding with 3D Gaussian splatting for enhanced scene understanding in dense SLAM systems. Their work highlights the importance of semantic information in improving the accuracy of traditional SLAM systems. A key lesson learned is that incorporating multi-view semantic constraints reduces drift during tracking, resulting in more accurate map generation. Additionally, they demonstrate that feature-level loss optimisation can improve semantic segmentation without negatively affecting geometric reconstruction, emphasising the effectiveness of targeted optimisation techniques.

Similarly, [Ji et al. \[2024\]](#) presented NEDS-SLAM, a dense semantic SLAM framework that also leverages 3D Gaussian splatting. Their research focuses on the semantic consistency required for accurate mapping and pose estimation. The integration of semantic information greatly improves map quality, allowing for a richer environmental understanding, which is critical for robotics and augmented reality. Moreover, they emphasise efficiency in processing, using lightweight models and efficient algorithms to balance high performance with minimal computational cost—essential for real-time applications.

Feature Relevance and Optimisation in 3D Point Clouds [Kumar et al. \[2019\]](#) explored feature relevance in 3D point cloud classification using deep learning, emphasising that using features from multiple search radii and neighborhood statistics enhances the accuracy of classification compared to single-point features. This approach allows for better local feature extraction, which is critical for improving the generalisation of classification models.

[Maturana and Scherer \[2015\]](#) introduced VoxNet, a 3D convolutional neural network (CNN) for real-time object recognition, but their approach shows limitations in handling dense splatted point clouds. Although voxelisation can be useful for sparse point clouds, it is less effective for dense splats, as it may reduce the resolution of the classification output, requiring double voxelisation at the cost of higher computational complexity.

Semantic Segmentation Techniques in 3D Point Clouds [Tchapmi et al. \[2017\]](#) proposed SEG-Cloud, a pipeline for semantic segmentation of 3D point clouds. While their method offers a robust process for interior object classification, it does not outperform other similar techniques and does not leverage Gaussian splatting. However, SEGCloud provides useful insights into stepwise improvements in semantic segmentation pipelines.

[Qi et al. \[2017a\]](#) introduced PointNet, a seminal work in 3D point cloud segmentation that only utilises raw 3D coordinates (x, y, z) without additional benefits from Gaussian splatting. This method is notable for its ability to handle planar segmentation and its reliance on large training datasets, making it a promising approach for future improvements in SLAM applications. Building on this, PointNet++ [Qi et al. \[2017b\]](#) emphasised the importance of

hierarchical feature learning and adaptability to non-uniform data, showcasing how processing point sets at multiple scales improves performance by capturing both local and global features.

Challenges in Feature Selection and Data Representation The work by [Poux et al. \[2016\]](#) on smart point clouds highlights the importance of feature selection in 3D data classification. Their research found that multiple search radii help in generalisation, while balanced training data is essential for reducing misclassification, particularly in complex urban environments.

Lastly [Jurski \[2024\]](#), the Semantic 3D Segmentation of 3D Gaussian Splats research explores the challenges of obtaining ground truth data for new techniques like 3D Gaussian splatting. Since segmented point clouds may lack texture, scale, and images, it becomes difficult to compare results with traditional point cloud methods. Additionally, variations in covariance matrix representation across different methods suggest the need for standardising these representations for consistency in performance evaluation.

2.7 Validation

[Al-Bayari and Shatnawi \[2022\]](#) computes the accuracy of its scan-to-BIM process by comparing the measurements obtained from the laser scanner with those collected from a total station. [Sewasew and Tesfamariam \[2023\]](#) developed a HBIM of a heritage building in the port city of Massawa, Eritrea, by using an image-based method. The HBIM was validated by architectural drawings from 1927. They compared f.e. Window Sill Height, the overall width and the column capital height.

[Laparra et al. \[2016\]](#) show an image quality metric which might be interesting to assess the visual quality of the Gaussian splats compared to the point cloud. In their research they use a normalised Laplacian pyramid, which breaks an image into parts based on luminance and local contrast at multiple scales, aiming to capture image features more similarly to how humans perceive them.

[Zhang et al. \[2019\]](#) used the following indicators to evaluate the quality of deep learning-based semantic segmentation for point clouds: (1) execution time, (2) space complexity, (3) mean intersection over union, which looks at how well each category is labeled, and (4) overall accuracy, which counts the total number of correctly labeled points.

2.8 Intermediate Conclusion

This chapter highlights recent advancements in HBIM and 3D data acquisition for heritage conservation. HBIM, supported by methods like photogrammetry and laser scanning, enables precise digital documentation of historic structures but faces challenges in data integration and standardization. Gaussian splatting emerges as a promising technique, providing smooth, high-quality visualizations from point clouds, which is valuable for HBIM applications. Despite ongoing challenges in data management and validation, these advancements offer powerful tools for the detailed and sustainable preservation of cultural heritage.

3 Methodology

The primary method will be introduced in this section before being validated against 2 datasets produced later in section 4.5. Figure 4 below shows the overall workflow of how to create a Gaussian splatted point cloud from different sources. The different methods using different tools will produce multiple point clouds, these will be compared against each other before selecting the Gaussian splatted point cloud used for validation. In sections 3.1, 3.2, 3.3, and 3.4 further details are described of how each process will be carried out.

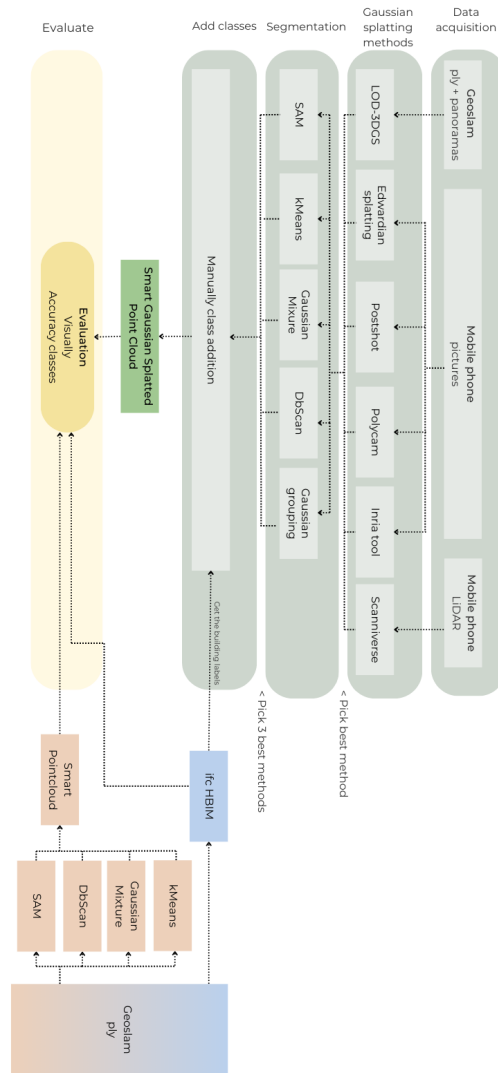


Figure 4: Overall Synthesis Project Workflow

3.1 Data Acquisition

Data will be collected manually using multiple devices, namely the GeoSLAM (Geographical Simultaneous Localisation and Mapping) and a mobile phone camera (iPhone 12 Pro). The GeoSLAM is a mobile LiDAR scanner that was created for both indoor and outdoor mapping and surveying, while the iPhone has built-in LiDAR scanners. SCANIVERSE (mobile phone LiDAR scanning application) was also considered. However, the iPhone LiDAR scanner has a limited range and did not allow for the scanning of a full building facade.

The iPhone camera utilises a 12 MP (Megapixel) camera to produce high quality images at a resolution of 3024x4032 pixels as well as a built in LiDAR scanner to produce point clouds with a range of 5 meters.

Meanwhile, the GeoSLAM, produces images at a resolution of 3840x1920 in a panoramic, fish eye perspective. The LiDAR scanner of the GeoSLAM produces a point cloud with a range of 100 meters.

The building chosen for this project is the Bouwpub, located behind the faculty of architecture building. Not only is it a heritage building ([Gemeente Delft \[nd\]](#)) but it also contains a variety of different building parts on one facade. This allows for the segmentation methods later in the project to be tested.

The GeoSLAM device requires a specific method of capturing the data where the start and end positions need to have the same correct orientation and position. This means that the user will have to circle back to the start position at the end of each scan. This ensures an accurate trajectory path when creating camera poses as well as the overall image and point cloud files.

3.2 Data Cleaning and Management

The data cleaning process for the heritage point cloud of the Bouwpub involved several steps using COLMAP, and RealityCapture. This multi-step approach was designed to address the unique challenges presented by the complex structure of the bouwpub and the various data acquisition methods employed.

3.2.1 COLMAP Processing

COLMAP ([Schönberger and Frahm \[2016\]](#)), an open-source SfM and Multi-View Stereo (MVS) algorithm, was used for the initial processing of images. The workflow included:

- **Image Input:** The input images include still photographs taken on-site (including mobile phone photos and GeoSLAM fisheye panoramic photos) and frame extractions from video footage.
- **SfM Processing:** COLMAP performed SfM to generate sparse point clouds and camera pose estimations.
- **Output:** The results were saved in binary(bin) file format.

3.2.2 RealityCapture Processing

RealityCapture is a photogrammetry software used to generate 3D models and obtain precise camera pose information from a series of overlapping images. In this project, RealityCapture was utilised to process input images, align them, and export a CSV file containing the camera poses. This section details the steps involved in the processing workflow with RealityCapture and explains the outputs relevant to further analysis.

3.2.3 Image Alignment and Sparse Point Cloud Generation

The first step in RealityCapture processing is the alignment of photos, where the software detects and matches common features between images to estimate the relative camera positions and orientations. This step produces a sparse point cloud, representing the structure of the scene based on key points detected in the images. The sparse point cloud provides a rough spatial structure and aids in verifying the quality of alignment between images.

3.2.4 Exporting Camera Poses

After image alignment, RealityCapture calculates the position and orientation (pose) of each camera in relation to the scene. These poses are crucial for applications requiring precise spatial registration, such as point cloud reconstruction and scene mapping. The camera poses, which include parameters such as camera location (X, Y, Z) and orientation (rotation angles or quaternion), are then exported as a CSV file. This CSV file serves as input for subsequent processing steps, allowing for accurate 3D transformations and further data integration.

3.2.5 Additional Data Preparation

In addition to the sparse point cloud and camera poses, RealityCapture provides options for exporting additional metadata and camera calibration parameters. These options facilitate integration with other tools or custom workflows that rely on consistent camera models. For projects involving dense reconstruction, RealityCapture's processing pipeline can also be extended to create dense point clouds and 3D meshes, though these were not covered in this project.

3.2.6 File Format Management

The diverse software tools and methods used in this research required careful management of different file formats. To ensure consistency throughout the workflow, a standardized output format was adopted: BIM models were stored in .ifc format, while point clouds and Gaussian splat data were standardized to the .ply format. These file formats serve not only as data storage but also as core representations for analysis and documentation.

For format conversion between different tools, several methods were implemented. Gaussian splat data was converted to .ply format using custom scripts to retain point attributes and spatial relationships. COLMAP outputs, which could be generated in either binary

(.bin) or text (.txt) format, were handled according to specific tool requirements—the INRIA method utilized binary files, while LOD3DGS required text format. COLMAP’s built-in conversion functionality facilitated seamless transitions between these formats.

Additionally, custom code was developed to integrate GeoSLAM camera data and trajectories into the Gaussian splatting pipeline by transforming trajectory data into camera poses using quaternion rotations and translations. This transformation aligned the data with the expected input format for Gaussian splatting, as detailed in Section 4.2.4. This standardized approach to file management ensured compatibility across different stages of the pipeline while preserving data integrity.

3.2.7 Coordinated and Alignment

The IFC model was constructed based on the LiDAR point cloud, ensuring they share the same coordinate system. However, the coordinate system from the SfM process in COLMAP differs significantly from this reference frame, primarily due to the arbitrary scale and orientation inherent in SfM reconstruction. This discrepancy needed to be addressed before any meaningful comparison or integration could be performed. To resolve these coordinate system differences, A point-to-point registration method was employed utilising Singular Value Decomposition (SVD). This approach relies on manually selected corresponding points between the two datasets, followed by a direct SVD-based solution to compute the optimal transformation matrix. The registration workflow consisted of three main steps:

1. **Point Correspondence:** Key points were manually identified in both datasets, ensuring accurate and reliable correspondence between the SfM-derived point cloud and LiDAR data. Let $\mathbf{P} = \{p_i\}_{i=1}^n$ and $\mathbf{Q} = \{q_i\}_{i=1}^n$ denote the corresponding point sets from SfM and LiDAR data respectively.

2. **SVD-based Scale and Transformation:** First, The point sets were centered by subtracting their centroids:

$$\bar{p} = \frac{1}{n} \sum_{i=1}^n p_i, \quad \bar{q} = \frac{1}{n} \sum_{i=1}^n q_i$$

$$p'_i = p_i - \bar{p}, \quad q'_i = q_i - \bar{q}$$

The covariance matrix was then constructed and decomposed using SVD:

$$\mathbf{H} = \sum_{i=1}^n p'_i q_i'^T = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$$

The diagonal values in $\mathbf{\Sigma}$ provided the scaling factors, and the rotation matrix was computed as:

$$\mathbf{R} = \mathbf{V} \mathbf{U}^T$$

The translation vector was then calculated as:

$$\mathbf{t} = \bar{q} - s \mathbf{R} \bar{p}$$

where s is the scale factor derived from $\mathbf{\Sigma}$.

3. Final Registration: The complete transformation for any point p can be expressed as:

$$p_{transformed} = s\mathbf{R}p + \mathbf{t}$$

This direct mathematical solution provided superior accuracy compared to iterative methods like ICP, as it avoids the risk of converging to local minima.

3.3 Gaussian Splatting

In this section, the cleaned point cloud will be Gaussian Splatted using readily available tools. Currently available tools are [Jawset Postshot](#), [Polycam](#), [Scaniverse](#), as well as a [Inria Tool](#) developed for real time radiance field rendering in combination with Gaussian Splatting (Kerbl et al. [2023]).

3.3.1 Inria Tool

The tool developed by Kerbl et al. [2023] applies memory efficient training using NeRF and Gaussian Splatting as one integrated program. The utilisation of NeRF is used to generate high quality 3D scene reconstructions from input data such as sparse point clouds and images. NeRF uses a volumetric approach to model complex geometries and lighting conditions using deep learning techniques. However, with Gaussian splatting, this replaces the ray marching process in NeRF reducing computational costs with 3D Gaussians as mentioned in 2.4.1.

Processing Software	Input Format	Files Required	Notes
COLMAP	*.bin, *.txt	cameras.bin, images.bin, points3D.bin, or their *.txt equivalents	Must include intrinsic/extrinsic camera parameters and sparse points.

Table 1: Input format for Command Line tool

As an output, it produces a **.ply** file and can be visualised in two ways, one with the system for image based rendering (SIBR) as well as a real time viewer, specifically designed for this Gaussian splatting process.

3.3.2 Jawset Postshot

Jawset Post Shot is a software created for fluid dynamics and visual effects (VFX) in 3D rendering workflows [Computing \[2024\]](#). Additionally, it utilises a similar method where Gaussian Splatting and NeRF is used in combination in order to produce a visually appealing output. Jawset Postshot can intake a variety of inputs as shown in table 2:

PostShot uses images and a combination of other related files in order to produce a Gaussian Splatted Point cloud as an output with the file extension a **.ply**. These input files can be created with preprocessing methods discussed in 3.2.

Processing Software	Input Format	Files Required	Notes
COLMAP	*.bin, *.txt	cameras.bin, images.bin, points3D.bin, or their *.txt equivalents	Must include intrinsic/extrinsic camera parameters and sparse points.
Reality Capture	*.csv, *.ply	*.csv (camera poses) *.ply (sparse point cloud)	Export camera poses and point cloud separately. Ensure correct settings for vertex colors and ASCII export.

Table 2: Input Formats for Jawset Post Shot

In the case of this project, the input images can be combined with the other input files such as **.csv**, **.bin** and **.ply** and imported into Postshot. In the case that only images are imported, the program first starts with a camera tracking step. This is where Postshot will calculate the image poses and create a sparse point cloud in order to begin the Gaussian splatting and NeRF process. After the camera tracking step, the training for NeRF begins. Here it is able to select the number of training steps to complete for training. In this project, the number of training steps were set to 30,000 as a starting point recommended by the Postshot user guide and the output can directly be viewed in Postshot. Results of this can be seen in section 4.2.

3.3.3 PolyCam

PolyCam offers a web-based implementation and viewer of Gaussian splatting for 3D scene reconstruction. The input consists of a set of overlapping images (.jpg or .png) or a video (mp4) from different angles. By taking closer shots of the objects, the resulting 3D model can have more fine-grained details. The process begins with Structure from Motion, which estimates the camera positions and generates a sparse set of 3D points. This input is used to generate a 3D splat model. Each image helps refining the scene by updating the parameters of each 3D Gaussian. This is done by minimising the error between the rendered image and the input image through back-propagation. This way, the scene is matched with the ground truth images. The implementation of Gaussian Splatting in PolyCam is based on the article of Kerbl et al. [2023]. (Polycam [2023])

It is not possible, at least for the free version, to (de)select any parameters. In addition, PolyCam allows users to visualise the result directly in their web browser and making it possible to measure distances in the scene. Besides that, it also provides the option to export the 3D reconstruction in a traditional 3D mesh format when you have a payed subscription, this costs \$26.99 for one month. (Polycam [2023])

3.3.4 Scaniverse

Scaniverse is a mobile scanning application that require devices equipped with LiDAR sensors or advanced depth sensors [Niantic \[2024\]](#). Users can capture their surroundings recording an object, and the app displays the scan in real-time, making it possible to do on-the-fly adjustments. After scanning, the app generates a cloud of Gaussians and optimises them to fit the scene. Users can then view the result and further improve the scan by retraining the model using the 'Enhance' button. The final scan can be shared directly through Scaniverse or exported as a video.

3.4 Segmentation

The fourth step of the pipeline is to take the best result from step 3 Gaussian splatting and segment the results. This segmentation was tested for both the dense point collected using the GeoSLAM and the sparse Gaussian splatted point cloud. A total of four methods of segmentation have been tried of which three methods have been passed on to the last step of the pipeline which is the labeling. The section starts with discussing the attributes present in a point cloud and the extra attributes calculated by Gaussian splatting. Then towards the end of the section each method gets explained.

3.4.1 Attributes of a Point Cloud

Segmentation often occurs based on a set of features or attributes that are associated with the segmentable elements of the object. In this case the object in question is a point cloud which when collected using a camera equipped GeoSLAM scanner or through the SfM process with the Phone camera has attributes assigned to each vertex. These attributes are as follows:

- X: Refers to the vertex position along the x-axis within a local coordinate system.
- Y: Refers to the vertex position along the y-axis within a local coordinate system.
- Z: Refers to the vertex position along the z-axis within a local coordinate system.
- R: Refers to the strength along the red color channel.
- G: Refers to the strength along the green color channel.
- B: Refers to the strength along the blue color channel.
- Intensity: Can be described as the brightness of the returned signal based on the specular, diffuse and ambient aspect of the captured point (only present with the GeoSLAM).
- Time: Refers to the time in which one of the reflected signals is returned with a valid distance (only present with the GeoSLAM).

These attributes tend to show patterns when visualised in an n-D space. These patterns can then be utilised by the segmentation methods to group them based on likeness.

3.4.2 Attributes of Gaussian Splat

As explained before Gaussian splatting uses sparse point clouds. However the training process does lead to an increase in the number of features present on each vertex. This increase in number of attributes leads to an increase in complexity due to the number of feature combinations. This set of attributes is not necessarily fixed as the .ply format allows for more columns to be created using scalar values. This is merely a list of attributes that are common to the Gaussian splatting methods that were applied during the research. For the visualisation in Blender further discussed in 3.6 a new attribute was added to signify the labeling. This labeling was created after the segmentation and does not impact the labeling itself. The attributes used for the segmentation are as follows:

- X: Refers to the vertex position along the x-axis within a local coordinate system.
- Y: Refers to the vertex position along the y-axis within a local coordinate system.
- Z: Refers to the vertex position along the z-axis within a local coordinate system.
- NX: Refers to the length of the normal vector along the x-axis. The combined normal vector is orthogonal the the geometriically visualised splat. The software outputs these as zero.
- NY: Refers to the length of the normal vector along the y-axis. The combined normal vector is orthogonal the the geometriically visualised splat. The software outputs these as zero.
- NZ: Refers to the length of the normal vector along the z-axis. The combined normal vector is orthogonal the the geometriically visualised splat. The software outputs these as zero.
- F_DC
- F_REST
- Opacity: Sometimes also referred to as the alpha value which defines the see-throughness of a splat in case of the captured point having opacity.
- Scale X: Refers to the scaling of the splat along the x-axis.
- Scale Y: Refers to the scaling of the splat along the y-axis.
- Scale Z: Refers to the scaling of the splat along the z-axis.
- Rotation X: Refers to the rotation of the splat around the x-axis.
- Rotation Y: Refers to the rotation of the splat around the y-axis.
- Rotation Z: Refers to the rotation of the splat around the z-axis.

As stated previously both the pointcloud and Gaussian splatting attributes are to be segmented using four different segmentation algorithms. The results of which are shown in section 4.

3.4.3 Segment Anything Model (SAM)

The Segment Anything Model (SAM) is a versatile and powerful image segmentation model designed to segment objects within an image efficiently. It leverages an extensive pre-trained foundation model that can handle a wide variety of segmentation tasks with minimal need for fine-tuning. SAM operates on the principle of promptable segmentation, where user-defined prompts, such as bounding boxes, key points, or free-form text, guide the segmentation process. The model uses a combination of convolutional and transformer-based architectures to extract detailed features and contextual relationships in the image, generating precise and high-quality segmentation masks. This allows SAM to be robust across

diverse image datasets and adaptable to varying object scales and complexities. The hyper-parameters for SAM include the size of the input image, the type of prompt provided for segmentation, and parameters controlling the fidelity and resolution of the segmentation masks.

This process first started with creating a spherical image of the point cloud as seen in figure 5. This was then followed by using the Segment Anything Model Automatic Mask generator with set parameters to create a mask of the segments over the image shown in figure 6. 3 model checkpoints were available with basic, large, and huge. The huge model checkpoint was selected in order to have the most accurate segments. The mask of the image was then reprojected back to the points using pixel mapping created during the creation of the mask and spherical image. However, this only reprojected points back onto the main face of the point cloud. Points that were behind the main face were not coloured. Initially, it was attempted to create 2 masks, one for the back face and one for the front face of the point cloud, before merging it into one. However, this caused many problems as the mask created for the front and the back did not align and different segments were created. To solve this, a simple nearest neighbour method was employed using raycasting to colour all corresponding unsegmented points. Results of this segmented point cloud can be seen in section 4.3.



Figure 5: Spherical Image of Point Cloud



Figure 6: Segment Anything Model Mask Projected onto the Spherical Image

3.4.4 k-Means Clustering

K-means is an iterative process that projects the points onto a nD -space and then using randomly assigned groups determines a mid point and tries to minimise the distance between the mid point centroid and the clusters. This when done on the same groups can be considered as a deterministic process meaning that it will always lead to the same results. The hyper parameters for a method like this are the number of clusters to be calculated and possible a stopping criteria based on the minimisation function.

3.4.5 Gaussian Mixture

Gaussian mixture refers to an iterative process that tries to estimate the Gaussian likelihood of a point based on point that have been projected onto an nD -space which then gets assigned to a set of Gaussian clusters. Each iteration the clusters and their assigned points get slightly adjusted. The hyper parameters for the method are the number of clusters, The covariance for the Gaussians width, the likelihood probability of a certain class and possible a stopping criteria based on the Gaussian density function.

3.4.6 Density Based Clustering

Density-Based Clustering (DB Clustering), often implemented through algorithms like DBSCAN (Density-Based Spatial Clustering of Applications with Noise), is a method used to identify clusters in a point cloud by examining the density of data points in the space. Instead of relying on a predetermined number of clusters, DB Clustering works by grouping together points that are closely packed while marking points in low-density areas as outliers or noise. The process begins by selecting an arbitrary point in the dataset and exploring its neighborhood to determine whether it has sufficient nearby points to be considered a core point. If so, a cluster is initiated, and all density-reachable points are recursively added to this cluster. This continues until no more points meet the density criteria. Key hyperparameters for DB Clustering include the radius of the neighborhood (ϵ) and the minimum number of points required to form a dense region (minPts).

3.5 Semantic Labeling

After the segments are created, each segment is assigned a label based on the building part it represents in the IFC model (e.g., wall, roof, stairs, door, window, or canopy). Labeling will be applied to a total of ten models:

- Results from k-Means clustering on the best Gaussian Splatting model, one with 7 and one with 14 clusters.
- Results from Gaussian Mixture on the best Gaussian Splatting model, one with 7 and one with 14 components.
- Results from k-Means clustering on the dense GeoSLAM point cloud, one with 7 and one with 14 clusters.
- Results from Gaussian Mixture on the dense GeoSLAM point cloud, one with 7 and one with 14 components.
- Results from the SAM method applied to both the best Gaussian Splatting model and the dense GeoSLAM point cloud.

This labeling is done manually by looking at the precision and recall scores for each cluster and each building part, as explained in Section 3.6.1.

3.6 Visualisation and Assessment

To address the third sub-question—whether Gaussian Splatting can improve point cloud clarity and segmentation results compared to traditional point cloud rendering and original parameters—two assessments will be conducted:

1. **Label Quality:** This assessment evaluates the accuracy of labels assigned to the points in both the Segmented Gaussian Splats and the Segmented GeoSLAM Point Cloud, where the GeoSLAM point cloud represents the traditional method. A ground truth, created by Yingwen Yu, is used to verify label accuracy. This ground truth is based on an IFC model of the Bouwpub's front facade, derived from the GeoSLAM point cloud. Details on calculating the label quality score are provided in Section 3.6.1.
2. **Visual Clarity:** This assessment scores the clarity of the point cloud to determine if Gaussian Splatting enhances visual clarity compared to traditional methods. A metric is described in 3.6.2.

3.6.1 Label Quality

To evaluate the accuracy of the point cloud model, a cluster-based analysis was used to examine how well the points in each cluster correspond to predefined building parts of an IFC model. From the IFC model, a set of building parts are defined. In this use case, the 6 parts are canopy, door, window, stairs, wall and roof, to match the IFC's Classes.

Following this, for each point in a cluster, the point's 3D location is observed and cross referenced to determine if it is inside one of the building parts. This is done by finding the closest boundary point and comparing the normal vector at the point with the normal vector

of the boundary. If the normal vector aligns with the normal vector of the boundary, it is inside the boundary. Conversely, if the normal vectors are misaligned, it is outside.

To validate label quality, precision and recall metrics are used. Precision indicates, for all points classified as a particular building part (e.g., wall), how many of these points are accurately classified as that building part (wall). Recall shows, for all points that should be classified as a particular building part (e.g., window), how many are correctly classified as that building part (window). In addition, the clusters will also be displayed visually to show the segmentation results. Higher percentages indicate more accurate clustering and segmentation.

3.6.2 Visual Clarity

Laplacian Filters

To assess and compare the output quality of the GeoSLAM point cloud and the Gaussian splatted point cloud, a laplacian filter will be employed to analyse the visual features of each output.

According to [Fisher et al. \[1996\]](#), the Laplacian filter, a second-order derivative operator, is particularly effective in highlighting areas of rapid intensity change, such as edges and fine details. It enhances high-frequency components in the data, making it a suitable tool for evaluating the sharpness and clarity of point cloud visualisations. The Laplacian filter will be applied onto the image where it calculates the second derivative at each pixel. This results in a new matrix where each value represents the change in intensity at that pixel. The variance is then calculated on the Laplacian output to calculate the overall Laplacian variance score.

The OpenCV package used approximates the Laplacian of an image using a discrete convolution kernel.

For an image I at pixel location (x, y) , the discrete Laplacian is generally computed as:

$$L(x, y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}. \quad (1)$$

Where:

- $L(x, y)$ is the Laplacian value at pixel (x, y)
- $\frac{\partial^2 I}{\partial x^2}$ and $\frac{\partial^2 I}{\partial y^2}$ represent the second partial derivatives of I with respect to x and y

These derivatives highlight areas of rapid intensity change, detecting edges in the image. However, rather than computing the Laplacian directly from this formula, OpenCV applies a convolution kernel to approximate it. The kernel used by OpenCV can be shown as ([Bradski \[2000\]](#)):

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}.$$

The Laplacian values for each pixel is stored as an array before applying the variance formula to find the overall variance of the image.

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (L_i - \mu)^2. \quad (2)$$

Where:

- σ^2 is the variance of the Laplacian values
- N is the total number of pixels in the Laplacian-filtered
- L_i is the Laplacian value at each pixel i
- μ is the mean of all Laplacian values, calculated as $\mu = \frac{1}{N} \sum_{i=1}^N L_i$.

Visualisation of Gaussian Splats and IFC

To further analyse the outputted point clouds visually, Blender was used to show Gaussian splats by installing the 3D Gaussian Splatting Blender Addon ([Carlier \[2023\]](#)). This addon processes attributes like position, opacity, scale and orientation, to construct a material using Blender's node system and shade system to define how each splat should be rendered. Since this research is working with segmented Gaussian splats, the [Gaussian Splat add-on](#) is modified so the label attribute remains and can be selected in Blender.

The other add on which was needed in Blender was [Bonsai](#) to be able to import an IFC object into Blender.

Besides Blender, also [SuperSplat](#) is used to visualise Gaussian Splats.

3.7 Intermediate Conclusion

This chapter explored advanced methods for capturing, processing, and segmenting 3D data for heritage building conservation, focusing on Gaussian splatting as an innovative approach. By combining traditional point cloud techniques with segmentation and validation methods, such as k-means, Gaussian mixture models, and SAM, we achieved enhanced clarity and segmentation accuracy. These improvements, supported by visualization tools like Blender and SuperSplat, demonstrate the potential of Gaussian splatting to create high-quality, semantically rich 3D models. This approach holds promise for more detailed and effective digital documentation, essential for preserving cultural heritage sites.

4 Results

The fourth section consists of results from the methodology introduced in section 3. The three main subsections that will be covered are Data Collection in subsection 4.1, Gaussian splatting in subsection 4.2 and Segmentation in subsection 4.3. Each subsection provides insights into the processing stages, from data acquisition to detailed analysis of segmentation techniques. The subsequent section will focus on validating these results, establishing a foundation for further discussion and interpretation.

4.1 Data Collection

In this subsection, the results of the discussed methodology in subsection 3.1 will be shown. Namely, the data collected from manual photography as well as LiDAR point cloud scans. This section presents the data collection methods used to obtain a comprehensive 3D model of the Bouwpub's front-facing structure. Two primary techniques were employed: GeoSLAM scanning and manual photography. GeoSLAM provided a 3D point cloud through a SLAM-based approach, along with geolocated panoramic images for spatial context. Additionally, standard 2D photos were taken with an iPhone 12 to capture finer visual details and mitigate fisheye distortion. These combined data sources form the foundation for the subsequent analysis and processing steps.

4.1.1 GeoSLAM

The GeoSLAM scanner was used to create a 3D point cloud model of the Bouwpub's front-facing structure through a Simultaneous Localization and Mapping (SLAM) approach. This method allows the device to generate the 3D model in real-time by combining pose estimation (localization) and mapping. Specifically, the SLAM process involves the continuous estimation of the scanner's position and orientation as it moves, while concurrently building the 3D map of the scanned environment, which here results in the point cloud model shown in figure 8. During the scanning process, GeoSLAM's panoramic camera captured two 180-degree panoramic images, which showed in figure 7. The two 180-degree panoramic images provided a full 360-degree visual context of the surroundings. These panoramic images are geolocated, meaning they are accurately positioned within the spatial model based on the scanner's location data. This geolocation is essential because it allows the visual and spatial data to be integrated precisely, enabling more accurate and context-aware analysis of the 3D point cloud model.



Figure 7: Panoramic Photo from the GeoSLAM

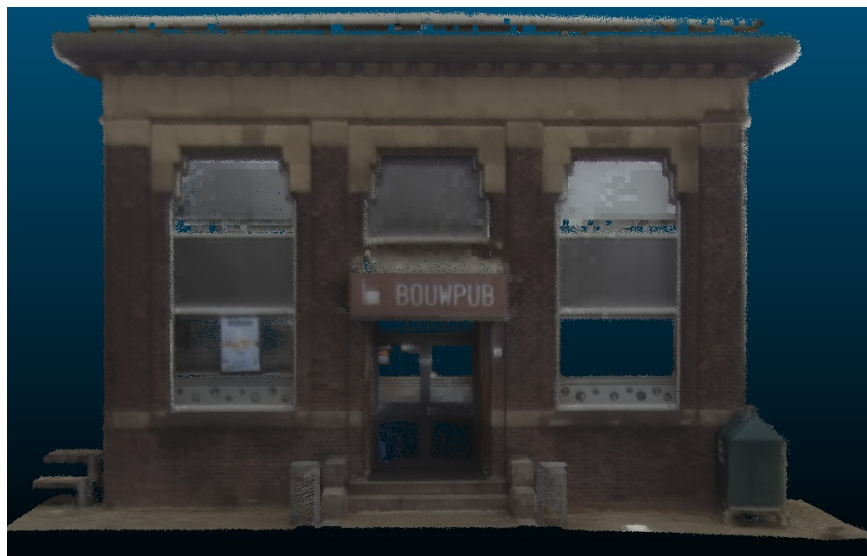


Figure 8: GeoSLAM Generated Point Cloud of Bouwpub's front-face

4.1.2 2D Photos

To compensate for the distortion of GeoSLAM's fisheye images, the Bouwpub front-face was also photographic captured by an iPhone 12, obtaining images from various angles around the structure, the examples are showed in figure 10.



(a) Photo 1 of Bouwpub's Front-Face



(b) Photo 2 of Bouwpub's Front-Face



(a) Top view photo 1 of Bouwpub



(b) Top view photo 2 of Bouwpub

Figure 10: 2D Images of Bouwpub's front-face

4.2 Gaussian Splatting

This subsection discusses four different approaches to Gaussian splatting for 3D reconstruction of a heritage building. Three of these approaches - Polycam, the original Inria implementation, and Jawset Postshot - utilise a traditional image-based pipeline, processing 124 high-resolution photographs captured with an iPhone. In contrast, the fourth approach, LOD 3DGS, explores a different methodology by incorporating LiDAR point cloud data alongside either panoramic images or high-resolution photographs. This fundamental difference in input data types represents two distinct paradigms in Gaussian splatting: pure image-based reconstruction versus LiDAR-assisted reconstruction.

4.2.1 Polycam

Polycam is a mobile application for iOS and Android devices as well as a web-browser based platform that leverages cloud-based computation for advanced Gaussian splatting processing. The project's implementation utilised an iPhone 12 Pro to capture 124 high-resolution images (3024x4032 pixels) of the heritage building. While the app facilitates convenient on-device capture, its key strength lies in its seamless integration with powerful cloud computing infrastructure.

Although the cloud-based processing system does not allow for manual parameter fine-tuning, Polycam's pre-configured parameters have been extensively optimised through rigorous testing on their high-performance servers. These default parameters significantly outperform what is typically achievable on local desktop systems.

As shown in [Figure 12](#), the reconstruction achieved exceptional visual quality. The geometric reconstruction exhibited remarkable accuracy, with superior preservation of fine architectural details. Despite the lack of manual parameter adjustment options, the cloud-based processing consistently produced better results that could achieve through manual parameter tuning on local desktop systems.

4.2.2 Inria tool

The original Gaussian splatting implementation from [Kerbl et al. \[2023\]](#) offers the most flexible parameter customisation among all tested tools. A total of 124 iPhone images were processed through COLMAP to generate the required input files (cameras.bin, images.bin, and points3D.bin). The implementation was run on a laptop equipped with an NVIDIA GTX 4060 GPU.

During initial testing, a significant amount of processing time was experienced due to the high resolution of the original images (3024x4032 pixels). The full-resolution images would have resulted in training times extending to several tens of hours, making it impractical for the research timeline. To address this limitation, A downsampling strategy was implemented in the parameter settings, reducing the image resolution to 25% of the original size.

The modified configuration ran for 30,000 training steps with the following key parameters:

- Image resolution: 1512x2016 pixels (downsampled from 3024x4032)

- Testing split percentage: 0.2
- Densification until iteration: 200,000
- Resolution down sampling: 2

A crucial parameter in the setup was setting `-densify_until_iter` to 200,000, which determines the iteration threshold for the densification process. This extended densification window allows the algorithm to continue adding new Gaussians to undersampled regions throughout the training process, resulting in better coverage of complex geometric features and improved detail preservation. While the actual training ran for 30,000 steps, the high densification threshold ensures the model maintains the ability to add new points where needed throughout the entire training process.

Figure 11 shows the results of this tool. This balanced approach allowed for the leveraging of the tool’s flexibility while managing computational resources effectively. The down sampling strategy, combined with the extended densification window, proved crucial in achieving practical processing times without severely compromising the reconstruction quality.

4.2.3 Postshot

Jawset Postshot represents an interesting implementation in the comparative analysis, building upon INRIA’s foundational work while offering distinct advantages in workflow and visualisation. The tool integrates SfM and Gaussian splatting into a streamlined pipeline, with its key differentiator being real-time visualisation of the training process. In the analysis, The following parameters were utilised:

- Image resolution: Downsampled to 1600 pixels
- Training steps: 30,000
- Maximum splats: 3,000,000

The real-time visualisation capability proves particularly value for parameter optimisation, allowing immediate adjustments and training restarts based on observed results. This represents a significant workflow improvement over tools like INRIA’s implementation, which requires complete training completion before results can be assessed. This interactive approach can substantially reduce the overall time needed for achieving optimal results. However, Postshot does present notable limitations. The tool offers a restricted set of adjustable parameters, primarily confined to:

- training time steps,
- image downsampling levels,
- maximum splat count.

This limited parameter space constrains fine-tuning capabilities, as users cannot adjust other potentially important variables. Furthermore, the analysis indicates that Postshot’s reconstruction quality, while competent, does not definitively surpass INRIA’s implementation, instead producing comparable results (see Figure 13).

The tool’s automated camera parameter optimisation during the SfM phase and view-dependent rendering optimisation continue to function effectively, maintaining reasonable color reproduction accuracy and consistency across viewpoints. However, these results

should be viewed within the context of the tool's constraints and its relationship to the original INRIA implementation.



Figure 11: (a) Inria tool Gaussian splatting results



Figure 12: (b) Polycam Gaussian splatting results



Figure 13: (c) Postshot Gaussian splatting results

4.2.4 Level of Detail 3-Dimensional Gaussian Splatting (LOD 3DGS)

The implementation of LOD 3DGS was based on the methodology presented in "LetsGo: Large-Scale Garage Modeling and Rendering via LiDAR-Assisted Gaussian Primitives"(Cui et al. [2024]). This approach differs from traditional Gaussian splatting by utilising LiDAR point clouds as input instead of SfM-generated sparse point clouds, theoretically offering higher geometric accuracy. The overall workflow can be seen in figure 14.

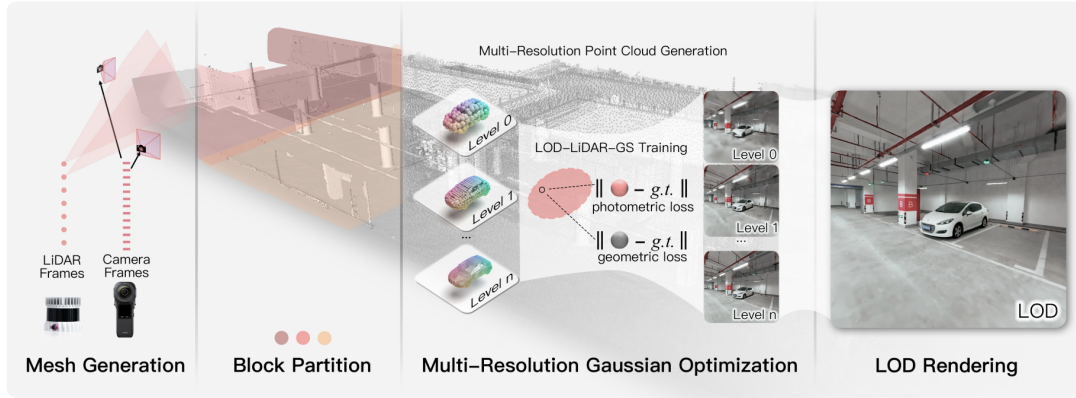


Figure 14: LOD3DGS Workflow

In the first approach, the GeoSLAM LiDAR point cloud was used as input along with camera poses obtained from GeoSLAM panoramic images. These images were transformed through cube mapping and re-projection.

To improve upon these results, a second approach was tested, combining the GeoSLAM LiDAR point cloud with high-resolution iPhone photographs. Since the coordinate system obtained from the camera’s SfM differs from the one generated by GeoSLAM’s built-in IMU sensor, this approach requires registering the GeoSLAM LiDAR point cloud to the camera motion coordinate system before applying Gaussian splatting. This method involved several preprocessing steps:

Using Structure-from-Motion (SfM) to process iPhone images and retrieve both camera poses and a sparse point cloud. This sparse point cloud served as the target alignment for the LiDAR point cloud, allowing it to be registered within the iPhone camera coordinate system. Cropping the GeoSLAM LiDAR point cloud to retain only the main structure of the building, thereby reducing computational overhead. Aligning the LiDAR point cloud with the SfM-derived sparse point cloud to ensure the LiDAR data was correctly positioned relative to the iPhone camera poses. Although this second approach produced slightly better results than the first (see [Figure 15a](#) and [Figure 15b](#)), the output still exhibited notable limitations. The reconstruction displayed significant variations in clarity across different viewing angles, and overall visual quality did not meet expectations.

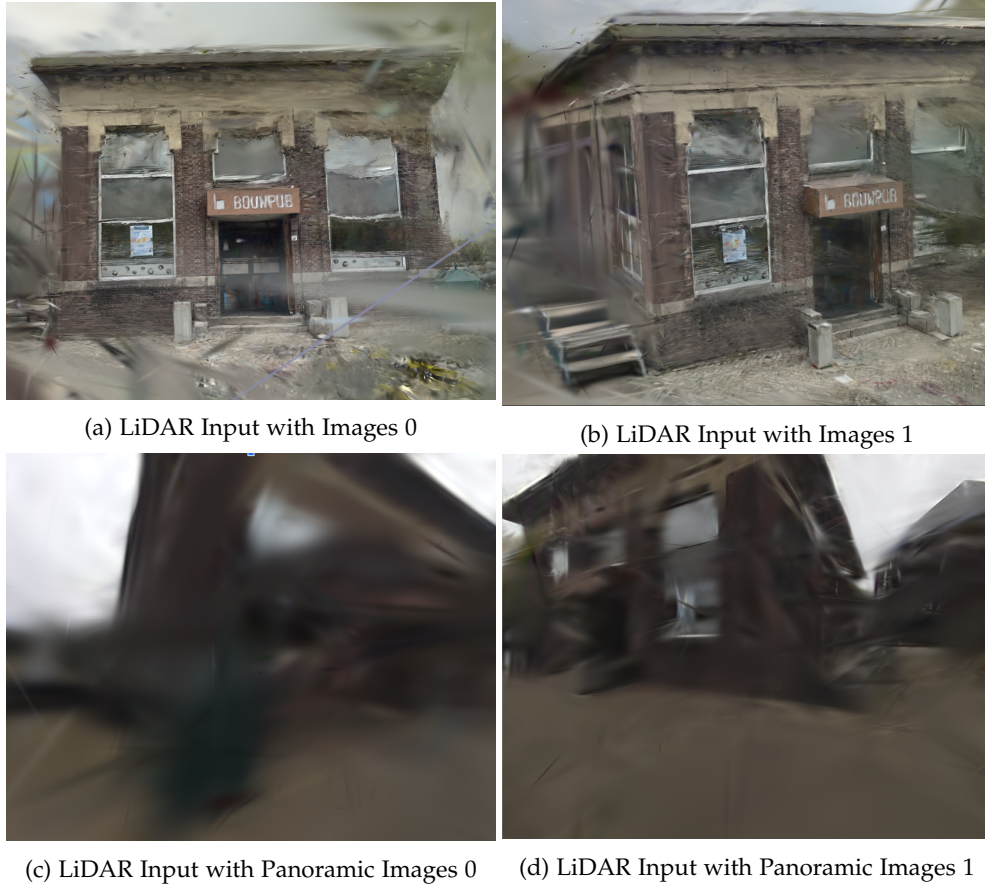


Figure 15: LiDAR Points & Images Results

4.2.5 Edwardian Splatting

So one of the original goals of the research was to devise a method that allows the use of the much denser GeoSLAM point cloud in combination with Gaussian Splatting. Due to the less than ideal results from LOD 3DGS visible in the previous subsection. This method places the point cloud into Blender accompanied by the original camera positions. This is visualised in Figure 16. To correctly transform the positions stored in a small bpy script used to create a small cube at the given coordinates. Then using a blender plugin called Camera Array Tool for Blender by Olli Huttunen a camera is added on the six faces of the cube thus resulting in a cube map from each of the six cardinal directions. The added advantage of this method verses the others is that it allows for the use of dense point clouds to be used without the need for matching that point cloud with the images by simply taking images of the point cloud. Figure 16 shows the setup in Blender as can be seen from the rendered viewport. In the gif the positions are visualised as black fustrums and the point cloud is added as a point cloud.

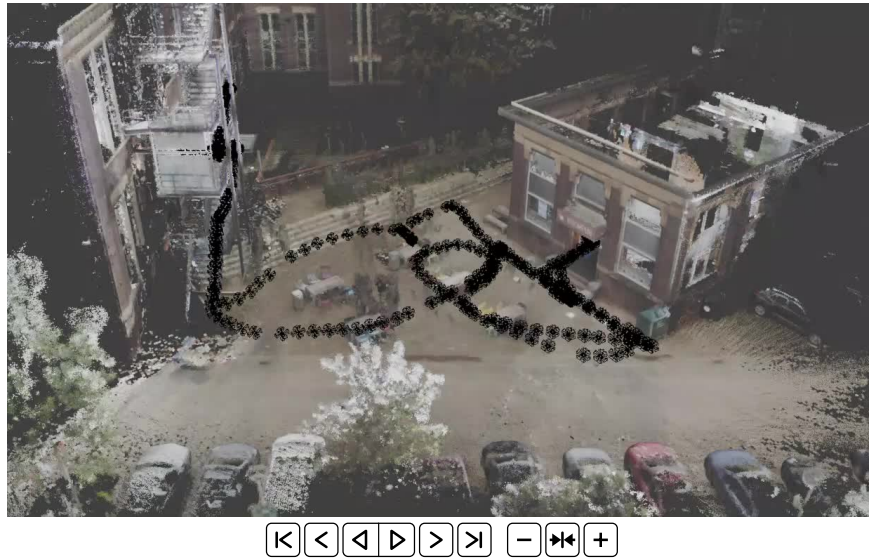


Figure 16: Overview of the 3D reconstruction in Blender (Best viewed in Adobe Acrobat)

This result was then rendered using blender resulting in a roughly 1800 images. These images were then brought into COLMAP and Postshot to create a Gaussian Splat. This is not ideal as a more optimal method would be to link the point cloud and original images directly as is done in Section 4.2.4. The result as shown in Figure 17 does show a better reconstruction when compared to LOD 3DGS.



Figure 17: The resulting Edwardian splat (Best viewed in Adobe Acrobat)

4.3 Segmentation

In this section, various segmentation methods are applied to both the point cloud models and Gaussian splats to analyze and compare their effectiveness in isolating architectural features. The segmentation aims to create distinct clusters of building components, which will enable validation and comparison between Gaussian splatting techniques and traditional point cloud models. By using models such as the Segment Anything Model, k-Means clustering, and Gaussian Mixture models, we evaluate each method's ability to delineate features like walls, windows, roofs, and other structural elements, providing insight into the strengths and limitations of each approach.

4.3.1 Point Cloud

In this subsection multiple segmentation methods are performed on the point cloud models. This is created in order to have a dataset that will allow for validation and comparison of Gaussian splats to traditional point clouds

Segment Anything Model

Figure 18a shows the final results of the segment anything model applied to the point cloud obtained by the GeoSLAM scanning device. The parameters used for the mask generator can be seen below as:

```
mask_generator = SamAutomaticMaskGenerator(  
    model=sam,  
    points_per_side=64,  
    pred_iou_thresh=0.9,  
    stability_score_thresh=0.95,  
    crop_n_layers=1,  
    crop_n_points_downscale_factor=2,  
)
```

From Figure 18a, it can be indicated that the Segment Anything Model was successful in creating clear segments in the point cloud relative to the project's goals. Out of all the buildings parts that needed to be segmented, only one window can be seen as not properly segmented. However, this was not segmented accurately in other methods that follow either. This can be due to the fact that there was a poster on the window, potentially creating a further segment, blocking out half of the window.

k-Means

Figure 18b and 18c show the results of performing k-Means clustering to the GeoSLAM generated point cloud.

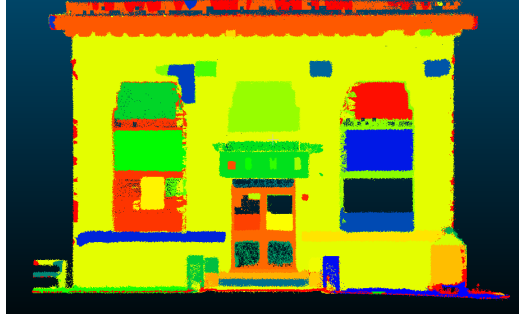
Figure 18b uses 7 clusters with randomly initiated centroids. This resulted in broad and generalised colour regions. However, the walls in this clustering are not segmented distinctly, and are blending in with other structural elements as well as an issue of creating multiple

segments for the wall split roughly in the center vertically. The lack of clear boundaries for walls and uniform colour representation creates challenges in isolating the building parts separately. Figure 18c uses 14 clusters with randomly initiated centroids to attempt to refine it at a higher resolution. Similarly, the issues still persist where the wall has still not been generated as one segment. On the other hand, it has segmented other building parts well, as the doors, roofs, and certain windows are now better segmented.

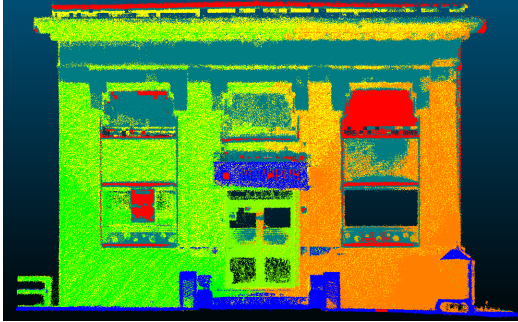
Gaussian Mixture

Figure 18d and 18e show the results of applying Gaussian Mixture clustering to the GeoSLAM generated point cloud.

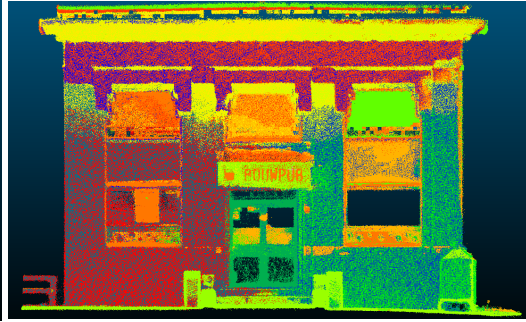
Figure 18d shows the application of a Gaussian Mixture Model with 7 clusters on the GeoSLAM point cloud. In this case, the segmentation results lack distinct boundaries, with considerable overlap between features such as walls, windows, and other structural elements. The limited cluster count struggles to differentiate between similar surfaces, resulting in a blurred segmentation where key building components are not well-separated. In Figure 18e, the number of clusters is increased to 14, providing a noticeable improvement in segmentation clarity. With more clusters, distinct features such as walls, roof, doors, and certain windows are better delineated. The additional clusters allow for finer segmentation, capturing more specific surface characteristics and enhancing the visual distinction between various architectural elements.



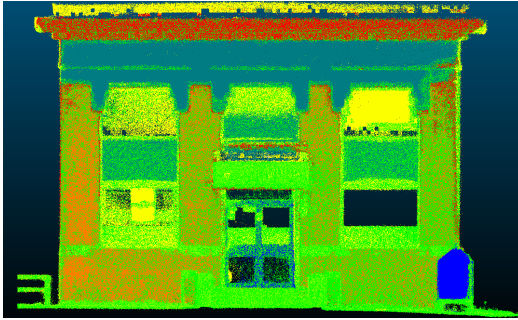
(a) Segment Anything Model applied to GeoSLAM point cloud.



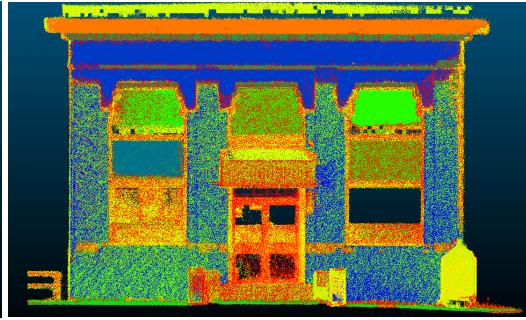
(b) k-Means clustering with 7 clusters applied to GeoSLAM point cloud.



(c) k-means clustering with 14 clusters applied to GeoSLAM point cloud.



(d) Gaussian Mixture with 7 clusters applied to GeoSLAM point cloud



(e) Gaussian Mixture with 14 clusters applied to GeoSLAM point cloud

Figure 18: Segmented GeoSLAM Point Cloud Results.

4.3.2 Gaussian Splats

In this section, the results of the segmented Gaussian splats are represented. Each cluster is shown in a different color. The Gaussian splat files are imported into Blender using the add-on, and more information is provided in subsection 3.6. By creating a new material in the *Shader Editor* and changing the material in *Set Material* in the *Geometry Node Editor*, it is possible to visualise the Gaussian splats using the colors of the clusters.

Segment Anything Model

Figure 19a shows the final results of the segment anything model applied to the Polycam Gaussian Splat.

The segmenting model divided the facade into a total of 10 segments. These segments do not clearly outline the building parts. There is a noticeable diagonal division, and one main cluster is visible. The uneven segment distribution may be due to two reasons: (1) the spherical projection of the Gaussian splat did not produce good results, and (2) the point cloud is sparse in some areas, such as inside the windows.

k-Means

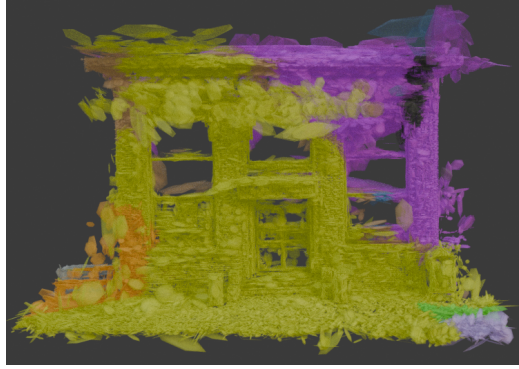
Figure 19b and 19c show the results of performing k-Means clustering to the Polycam Gaussian Splat.

Figure 19b shows the results of applying the k-Means model with 7 clusters to the Polycam Gaussian Splatted file. Visualising each cluster individually in Blender reveals that building parts are not segmented separately. Each cluster contains multiple building parts and even includes the floor, which is not part of the building structure. Figure 19c shows the results of applying the k-Means model with 14 clusters to the Polycam Gaussian Splatted file. Visualising each cluster individually in Blender reveals a somewhat clear separation between the ground and the building; however, the building parts are still not fully separated from each other. There is slightly more separation between the wall and roof, but the canopy, stairs, and windows remain grouped with either the wall or roof.

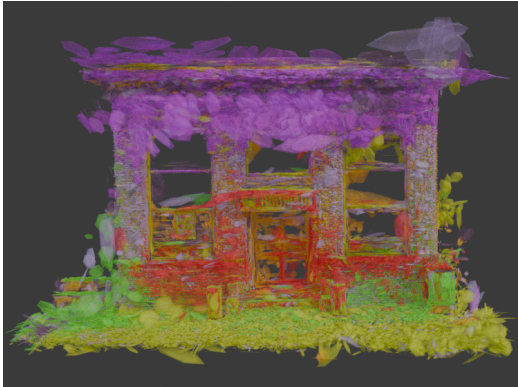
Gaussian Mixture

Figure 19d and 19e show the results of applying Gaussian Mixture clustering to the Polycam Gaussian Splat.

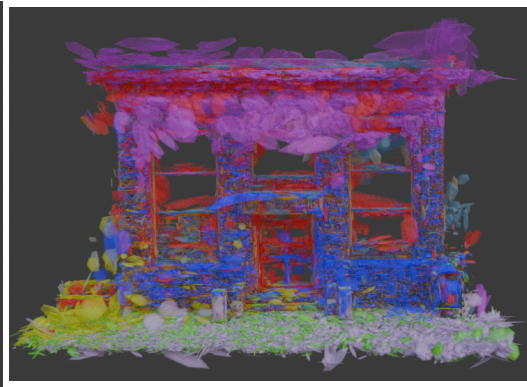
Figure 19d shows the results of applying the Gaussian Mixture model with 7 clusters to the Polycam Gaussian Splatted file. Visualising each cluster individually in Blender reveals a clear separation between the ground and the building. The roof is also well-separated from other building parts, but the window, canopy, and wall frequently appear together in several clusters, as do the door, stairs, and parts of the wall. Figure 19e shows the results of applying the Gaussian Mixture model with 14 clusters to the same file. Visualising each cluster in Blender reveals a good distinction between building parts and the ground. This time, the roof is even better clustered, as are the stairs. However, the model still struggles to fully separate the window, canopy, and wall; while some clusters contain only wall points, others combine all three of these elements.



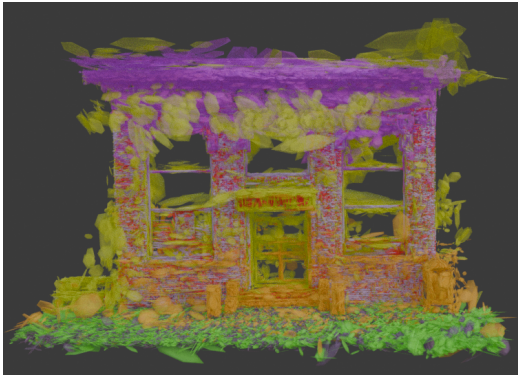
(a) Segment Anything Model applied to Gaussian splats.



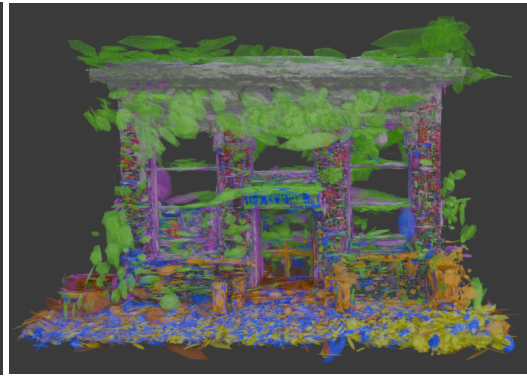
(b) k-Means clustering with 7 clusters applied to Gaussian splats.



(c) k-Means clustering with 14 clusters applied to Gaussian splats.



(d) Gaussian Mixture with 7 clusters applied to Gaussian Splats.



(e) Gaussian Mixture with 14 clusters applied to Gaussian Splats.

Figure 19: Segmented Gaussian splatted point cloud results.

4.4 Intermediate Conclusion

Based solely on the visual inspection of the segmentation results, it is implied that the SAM has provided the best results, followed by Gaussian mixture and lastly k-Means.

On the other hand, the segmented Gaussian splats visual inspection suggests that the Gaussian Mixture model with 14 clusters achieved the most effective segmentation. Although not perfect, this model provided the clearest division based on building parts.

Overall, there were some common challenges across both Gaussian splatted models as well as point cloud models. Windows were generally difficult to distinguish due to gaps in the Gaussian splats around window areas, while the roof was often successfully separated from other parts in several models. However, the door remained challenging to separate distinctly from other building parts. The door and window issues could be due to the fact that there were posters or fliers pasted onto the windows, creating further segments each time.

4.5 Validation

This section presents the validation of the segmentation and quality of the point cloud models generated from GeoSLAM and Gaussian splatting methods. The validation process involves both visual and accuracy assessments to evaluate the effectiveness of each approach. In section 4.5.1, the Laplacian filter is applied to compare the sharpness and detail of the models, using Laplacian Variance scores as a measure. In section 4.5.2, precision and recall metrics are calculated for specific building components to assess segmentation accuracy. These evaluations provide insights into the suitability of each model for applications requiring detailed and accurate representations, such as heritage building documentation.

4.5.1 Visual Validation

Figure 20 and 22 display a point cloud generated from the GeoSLAM and a Gaussian splatted point cloud, respectively. The application of the Laplacian filter resulted in the processed images shown in Figures 21 and Figure 23. The images were generated by taking screenshots of the front face of each point cloud and removing all colours unrelated to the point cloud, leaving only instances of the point cloud with no background. This allowed for a clear image without any background noise to focus on the comparison of the prominent features of the point clouds.

After applying the Laplacian filter, the GeoSLAM point cloud in figure 20 achieved a Laplacian Variance score of 1181.53. In contrast, the Gaussian splatted point cloud in figure 22 attained a significantly higher score of 2021.88. This substantial difference indicates that the Gaussian splatted point cloud contains more high-frequency information and sharper features than the GeoSLAM point cloud. For reference, the BIM model being used as ground truth receive a score of 3493.74.

Based on the study of [Chao et al. \[2021\]](#), a higher Laplacian Variance score for the Gaussian splatted point cloud suggests that it provides a visualisation with enhanced detail and clarity. This could be due to how Gaussian splatting effectively distributes point attributes over a Gaussian kernel, which smooths noise while preserving edges and fine details. This results in a point cloud that not only looks smoother but also retains critical structural information, making features more discernible.



Figure 20: Image of GeoSLAM Point Cloud in Greyscale

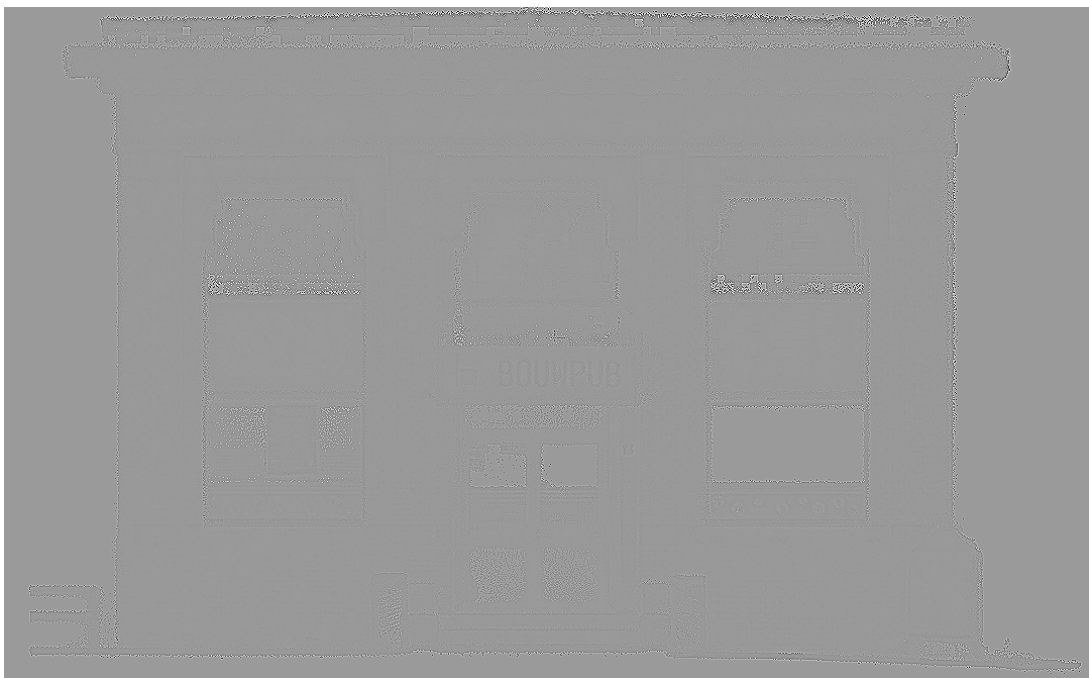


Figure 21: Laplacian Filter Applied to GeoSLAM Point Cloud



Figure 22: Image of Gaussian Splatted Point Cloud in Greyscale



Figure 23: Laplacian Filter Applied to Image of Gaussian Splatted Point Cloud

In summary, the Gaussian splatted point cloud demonstrates superior visualisation quality compared to the GeoSLAM point cloud. The use of the Laplacian filter and the analysis of the Laplacian Variance scores highlight the enhanced sharpness and feature definition achieved through Gaussian splatting. This makes the Gaussian splatted point cloud more suitable for applications that require high-resolution and detailed point cloud representations, such as precise modeling, inspection, and analysis tasks of heritage buildings. This gives us an overall visual quality representation of both 3D models as the images were taken directly from the specific 3D viewers.

4.5.2 Accuracy Validation

To evaluate the accuracy of the point cloud segmentation and labeling, precision and recall metrics were calculated for each building part, including canopy, door, roof, stairs, wall, and window. The precision heatmap in figure 24 shows the range of precision values across building parts, where higher values indicate a greater proportion of accurate classifications. For example, the GeoSLAM point cloud models tend to exhibit precision values ranging between approximately 10% and 77% for different building parts like the wall and roof. This range highlights effective identification of these elements in some configurations but also shows lower precision for other parts, like the canopy, possibly due to its more complex structure and fewer points.

The recall heatmap in figure 25 illustrates the percentage of points that should be classified as each building part and are correctly identified as such. Here, GeoSLAM point cloud configurations generally achieve higher recall values for the wall and stairs, indicating more successful segmentation coverage of these building elements. This higher recall for the wall may also be attributed to its higher point density, which provides more information for clustering and classification. Conversely, Gaussian Splatting models tend to show lower recall values, particularly for parts with fewer points, like the canopy, which can reduce detection accuracy.

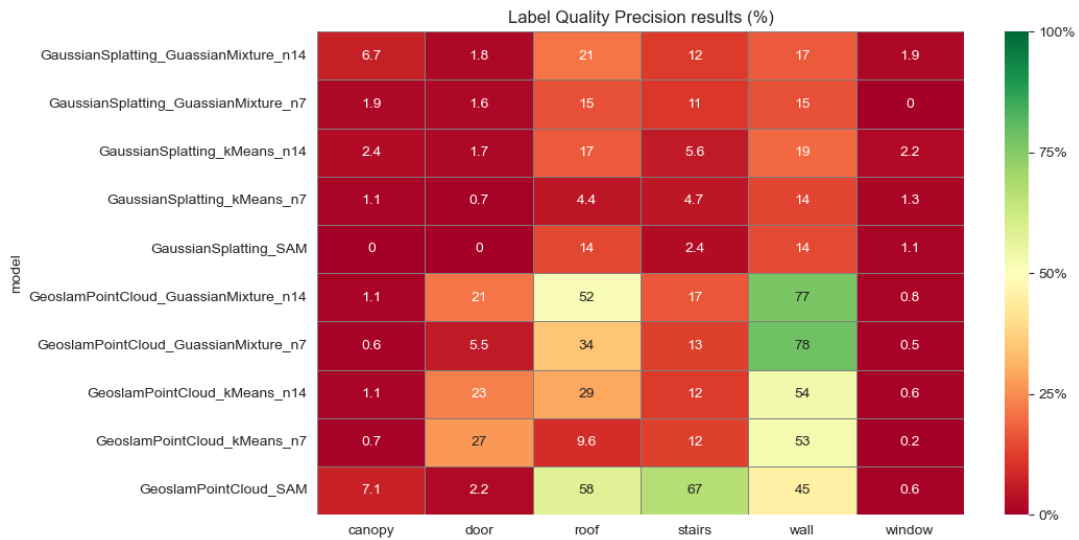


Figure 24: Heat Map for Precision Scores on all Methods.

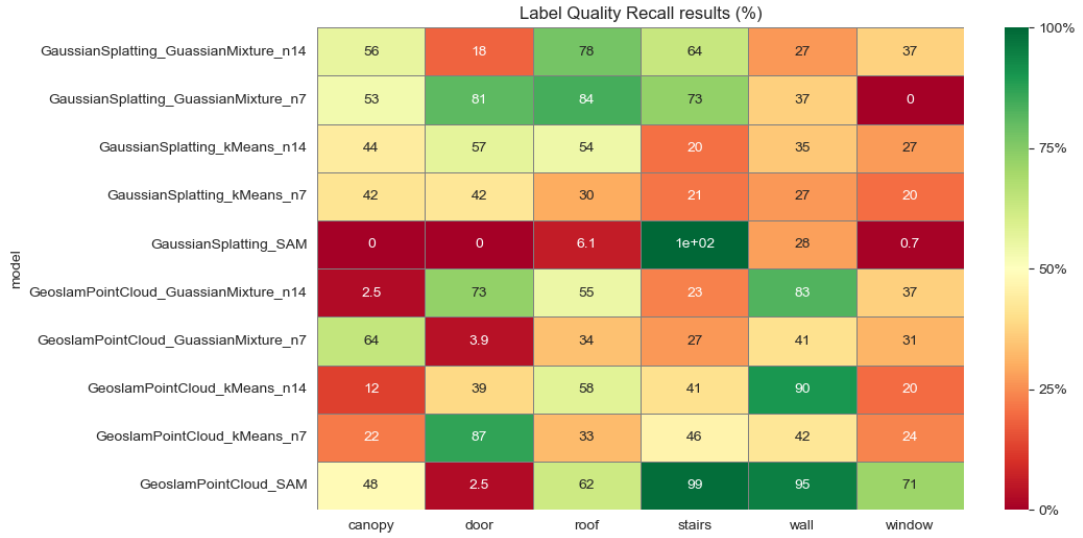


Figure 25: Heat Map for Recall Scores on all Methods.

The mean precision and recall bar chart in figure 26 provides an overview of these metrics across all building parts, with error bars representing standard deviations. It is observed that while Gaussian Splatting models demonstrate smaller standard deviations in precision across all configurations, 3 out of 5 GeoSLAM configurations achieve greater mean recall values. This suggests that GeoSLAM models perform more consistently in capturing a wider range of building elements, though Gaussian Splatting provides steadier precision with less variability.

Overall, these results suggest that while Gaussian Splatting methods offer a basic level of segmentation with stable precision across parts, GeoSLAM point cloud models, particularly with SAM configurations, achieve higher recall and are better suited for applications requiring comprehensive and accurate labeling. This trend is beneficial for heritage building documentation, where capturing extensive detail is crucial for effective segmentation and classification.

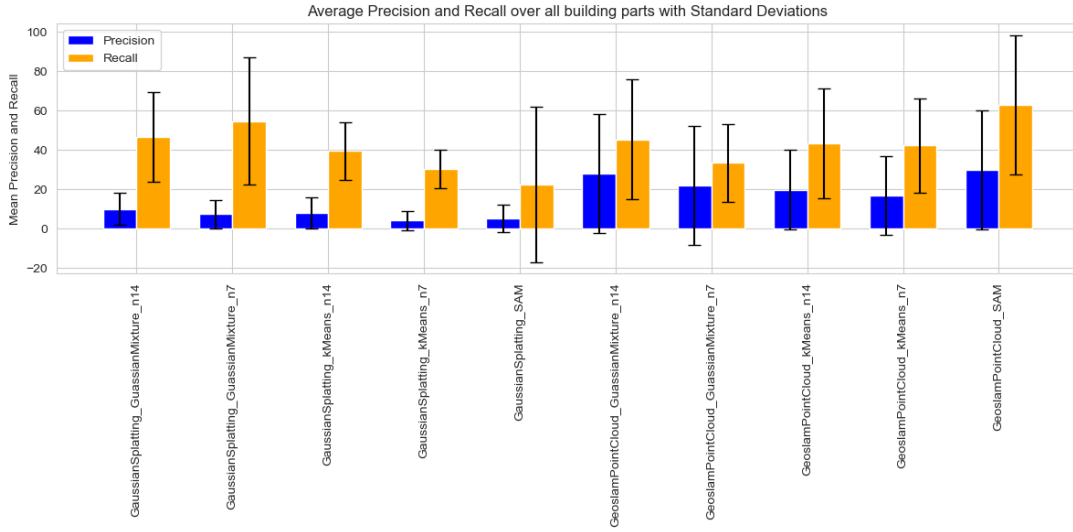


Figure 26: Barplot with Mean Precision and Recall with Error Bars on all Methods.

4.6 Intermediate Conclusion

This chapter demonstrates the effectiveness of Gaussian splatting and GeoSLAM techniques for point cloud segmentation and heritage building documentation. Visual and accuracy validations show that Gaussian splatting improves clarity, offering sharper detail through higher Laplacian variance scores, while GeoSLAM provides higher recall, capturing a broader range of building elements. Overall, Gaussian splatting enhances visual quality, making it ideal for applications demanding high-resolution models, whereas GeoSLAM excels in detailed labeling, supporting applications that require precise segmentation. These insights highlight each method's strengths, advancing digital documentation practices for heritage conservation.

5 Discussion of Limitations

The sixth section discusses the draw backs and limitations of the current research. The research was explorative in nature which led to trying a lot of different methods discussed in the 3. Some of these methods were successful and are shown and discussed in Section 4 and 4.5. The unsuccessful methods are discussed below with reasoning as to why these were unsuccessful. This section follows the MoSCoW as laid out in the Project Identification Document (PID).

5.1 Challenges & Limitations

The goal of the research was to use already existing methods and show drawbacks and gaps that are currently present within the pipeline. These challenges and limitations are summarised here.

5.1.1 No BIM for Complex Buildings

The overview of the pipeline as discussed first in section 3 requires a BIM or HBIM model to be present or created using software discussed in section 3 the research group was unfamiliar with these types of software. Luckily one of the supervisors was able to model a facade using BIM software.

The original intention was to scan three buildings. These buildings were laid out in the PID and were as follow:

1. Windmill
2. Aula
3. Bouwkunde

After scanning these buildings the workflow was further refined which eventually led to comparing against a HBIM as a baseline. As stated in section 3 this model would be used as a form of ground truth against which the different methods would be tested as shown in 3. This would mean that all initially scanned buildings needed to be modeled in BIM. Due to the complexity of the Windmill a simpler object was chosen. This leads to that one of the limitations that the current workflow has is that a BIM model needs to be present to go through it. This requires a substantial amount of work before hand.

5.1.2 Limited Range of Scaniverse

The second limitation that was originally set out in the PID was to use a variety of devices/methods. One of these methods was to create the Gaussian splat through the use of Scaniverse. As discussed in the Methodology Scaniverse is a mobile application for both iOS and android that takes pictures to use for SfM and create a Gaussian splat. Further more on Apple devices that are equipped with a LiDAR sensor also captures a LiDAR point cloud which is then used as well.

The limitation with Scaniverse was the maximum range which is only 5 meters. This is both a software and hardware problem as Scaniverse does not seem to use the LiDAR scanner for its Gaussian splatting training similar to other Inria based reconstructors. This is simply not

far enough to fully capture complex buildings at a larger scale. As such this limitation was seen as being too great leading to the method being cast aside in favor of the more successful methods shown in 4.5.

5.1.3 Limited Computational Power

One of the main hurdles faced during the project was the limited computational power available on modern laptops as it comes to Graphical Random Access Memory (RAM). Certain methods discussed were abandoned based on the requirements exceeding availability. This led to methods such as Edwardian splatting and Gaussian grouping not to be further explored.

5.1.4 Using Video as Input

One of the more common methods for Gaussian splatting using SfM is to take a continuous video of an object and split it up into separate frames of which a frame is picked at a given interval thus leading to a set of images. During the initial collection of data this was tried and it is introduced in section 3. However, this did not lead to a better results using the same camera versus individual images due to the introduction of movement blur the results could from a visual point of view even be deemed to be lesser than statically shot images.

5.1.5 GeoSLAM to Splat

One of the original goals set out during the PID was to somehow use the GeoSLAM point-cloud in addition to images to better reconstruct the Gaussian splat. Two approaches were tested. The first approach involved using LOD 3DGS which produced results but suffers from unusual artifacts in the Gaussian splat which results in less than ideal segmentation results. This could be due to a variety of reasons. One main suspected cause being due to the camera location on the GeoSLAM. However, trying to resolve this problem by using higher quality photos from the mobile phone did not solve this problem. The second method to use the GeoSLAM point cloud in Gaussian splatting was to load the .ply file in Blender and use the actual picture positions in blender to generate a new set of rendered images. This was first laid out in section 3. This did work but was not explored further during the research phase of the project.

5.1.6 Planar Surface

The current approach very much relies on the scanning, processing and labeling of a planar surface such as the Bouwpub. Initially the group had collected data for more intricate objects such as the Aula or a Mill. This leads to the pipeline as it currently stands to be optimised for planar surfaces such as a single facade but might not be necessarily be expanded to much more complex objects.

5.1.7 Gaussian Splatting not Deterministic

One of the challenges of Gaussian splatting is that running the process in the different software packages, laid out in section 3.3 does not necessarily lead to the same result each time the process is ran. This leads to the uncertainty that a Gaussian splat might not be the optimal representation of that object. Since some methods were taking hours, we were unable to run the same splat multiple times to determine which hyper parameter would be best used. This might lead to the challenge for others that might want to apply the method or recreate the results that their results might be entirely different. This stems from the fact that the learning process taking place during splatting is not deterministic.

5.1.8 Alignment

The current algorithm used for alignment does not produce adequate results this is due to the scaling factor that is present within SfM algorithms. This leads to less than ideal fitting of the point cloud and Gaussian splat back onto the HBIM model. Due to time constraints this was not further explored. Given the fact that all segmented results use the same alignment so the effect it has on results is similar in all cases. In the future this could be alleviated by either using targets that represent control points that are identifiable in each of the segmented representations. This would still not be ideal but would result in a better alignment as a whole.

5.1.9 Quality of Collection Methods

The current approach very much looks at the technical aspects of each collection method but not necessarily at the output where one used for the whole pipeline. Currently only the best from each step is used. It can be said that each method can be used for each step in the process depending on the choices the user makes. However the research looks less into the quality of the collection methods themselves. This is partially due to a lack in time and processing resources.

5.1.10 Using Splats for Segmentation

One of the challenges that was faced during the segmentation process is that the three segmentation methods explained in section 4 for the Gaussian splatting only works based on the point cloud's mean position and not the full splats. This means that only the added features for each point are used in the segmentation and not their actual geometric representation. The label accuracy is done by checking if the point clouds point position (x,y,z) are laying inside the BIM building part. Same for the splat, the x,y,z is looked at. So it was not considered when the Gaussian splat shape takes form. If it lies inside the building part but just the x,y,z . This might not be a fully complete method.

5.1.11 SfM from Gaussian Splatting Not Dense

The final challenge is closely related to the previous point. One of the aspects of Gaussian splatting is that it creates a sparse point cloud through SfM and then fills the geometric space with splats. Because the original point cloud is much denser even having the calculated features might not be enough to overcome the segmentation on the Gaussian splatted point cloud.

Overall, the research pipeline as it currently stands is robust and leads to results but based on the points discussed in this section it could be said that it does face limitations and that certain challenges still need to be overcome.

6 Conclusion

This section of the research report presents the conclusions based on the question stated in section 1.

To what extent can a segmented Gaussian splatted point cloud support the heritage Building Information Modeling workflow?

The first conclusion relates to the use of different data collection devices. The GeoSLAM captures the densest point cloud compared to the iPhone, while the smartphone captures the clearest images. This is likely due to the sensor size present on modern telephone devices. This allows for more detail to be captured through better light capture. The GeoSLAM device, utilising LiDAR-based SLAM, produces a denser point cloud than the iPhone's SfM-based approach. This is attributed to the high frequency of laser pulses and the rotating sensor in the GeoSLAM, which enables rapid and comprehensive spatial sampling. In contrast, the smartphone captures higher-resolution images with greater clarity, likely due to advancements in sensor size and light sensitivity in modern mobile cameras, facilitating more detailed visual information. These differences highlight the trade-offs between dense spatial coverage in LiDAR systems and image clarity in vision-based systems, each offering unique advantages for specific types of spatial data acquisition.

The second conclusion concerns the different Gaussian splatting methods that were applied during the second step of the pipeline. Most current implementations stem from the Inria base code, which uses COLMAP SfM positions for each photo along with the photos themselves. Among the methods tested, PolyCam produced the best reconstruction as visual representation. LOD 3DGS, which used LiDAR and photos as input, had lesser results. A drawback of PolyCam is the blackbox nature of the website and no available parameters to tune, unlike other methods. Each method always outputs a .ply file.

Regarding the best segmentation method, the results are inconclusive due to the large spread in the standard deviation for the different segmentation algorithms applied to all methods. This is partially due to three limitations described in Section 5. (1) Where the alignment plays a big role into the correctness of the results when compared to the HBIM ground truth. (2) The HBIM is based on the GeoSLAM generated pointcloud. (3) Gaussian splats are now simplified as points in the precision and recall metrics, but they are a distribution. Besides that, since the segmentation was only applied to a single facade, the hyperparameters were likely overfitted to the current use case. Based on this a conclusion about the best segmentation method cannot be made.

For the visual validation a Laplacian filter was used to assess the edge sharpness, which relates to a reconstruction of a building which itself is represented using sharp edges in the real world. The results suggest that the Gaussian splatted point cloud performs better as it better captures the real life facade.

For the final sub-question on improving HBIM preservation compared to a point cloud, it is essential first to determine if the initial BIM criteria outlined in Section 2 are met.

- The first aspect refers to whether the model is in 3D. Which in the case of a Smart Gaussian Splatted Point Cloud is in fact a three dimensional in space.

- The second aspect is the model’s ability to provide measurable, numerical and descriptive information to be derived about the building. The resulting splat can definitely be measured and would in the case of better segmentation be possible to also allow for descriptive information to be attained. However for the numerical information, such as number of a certain building part, that would also need to be improved.

Overall, a colorised Gaussian splat is in fact better for the visual experience of the user, but there are inconsistencies related splat size and IfcClass that a Gaussian splat can represent.

For the final conclusion related to the main question of the research, which concerns the extent to which a segmented Gaussian splatted could be used for the betterment of the heritage Building Information Modeling workflow. Compared to a simple point cloud, a Gaussian splatted model better captures the current state of the building. As it shows the results in a more photogrammetric representation when compared to the singular color valued points. One of the draw backs for Gaussian splatting, as it relates to the segmentation, which should then be used for better understanding of the object. This is however not entirely possible due to the singular label that can be assigned to a splat. A splat might be covering a larger area.

Overall, the method does impact certain aspects of the workflow but does not necessarily help in all aspects.

7 Further Research

In this research, only unsupervised learning methods are applied, but an alternative approach could involve supervised learning methods, such as VoxNet and PointNet++, as discussed in section 2.6. Supervised learning could potentially yield better results by training the model to make specific predictions, rather than relying on random guessing, and could allow for identifying patterns specific to each building component. PointNet++ architecture, for example, is particularly well-suited to directly process 3D point clouds through its hierarchical feature learning structure. By capturing both global and local geometric features with multi-scale grouping and sampling layers, PointNet++ is effective in identifying building parts, such as walls, roofs, and windows, based on their geometric characteristics.

Furthermore, there are pre-trained models for semantic segmentation on point clouds, such as RandLA-Net (Hu et al. [2020, 2021]) and TUM-FAÇADE (Wysocki et al. [2022, 2021]). Investigating whether these models could be adapted to work with Gaussian splatted files may provide additional insights into improving segmentation accuracy.

A second innovative approach integrates Large Language Models (LLMs) for semantic labeling. In this method, geometric features (such as planarity, verticality, and height) and images of each segmented point cloud cluster are converted into natural language descriptions, which are then processed by LLMs like GPT-4 via API calls. The LLM analyzes these descriptions to classify building components based on its architectural knowledge. A potential workflow involves extracting cluster features, formatting them as prompts (e.g., “This component has 0.95 verticality, 3.2m height, and is located on the building exterior...”), and using the LLM’s response to assign semantic labels. This method, combined with traditional supervised models, could enhance the semantic enrichment of segmented building point clouds.

The current validation approach treats Gaussian splats similarly to point clouds, focusing primarily on point-based metrics like precision and recall. However, this method does not fully capture the unique properties and advantages of splats as continuous representations rather than discrete points. A future validation method could directly analyze splat-specific attributes, such as scale, orientation, and opacity, which are intrinsic to Gaussian splats. By leveraging these attributes, validation metrics could be developed to assess not just the accuracy of individual points but the smoothness and continuity of splat-based segments, which would better reflect Gaussian splatting's strengths in creating seamless, high-fidelity models. Additionally, a splat-focused validation could incorporate metrics for evaluating splat overlap and blending quality, allowing for a more nuanced assessment of visual clarity and structural fidelity. This would be particularly beneficial for applications that require high-resolution and realistic representations, such as heritage conservation.

Another possible direction for future research would be to explore effective ways of visualising clusters in a Gaussian splatted file. Because Gaussian splats can overlap and some are large enough to cover significant areas, such as large sections of a building, a single Gaussian splat may encompass multiple building parts. In this research, Gaussian splats have been labeled and visualized individually, treating them as points. However, they should ideally be handled differently, as each splat is more than a point and represents a larger area.

To further improve the project, a method that could be used to store strings such as IfcClasses are saving these as numeric scalar field attributes in the .ply file and having a separate file such as a .json file or any other separate metadata file to store the strings attached to the numeric scalar field. Using this method will allow for the user to map each building part to the correct IfcClass while adhering to the file format standards of Gaussian splats.

An alternative approach in segmentation is to consider creating meshes for each splat to use in segmentation, rather than relying on the center points of each splat. While using center points may be simpler, it does not accurately represent the unique shapes and overlaps of individual splats. This simplification has impacted the accuracy of the segmentation method used in this project for Gaussian splatting. A possible improvement could involve integrating mesh-based segmentation, such as Poisson reconstruction, to enhance the precision and reliability of the segmentation process. Further exploration of this mesh-based approach could yield better results.

8 Task Distribution

Student name	Responsibilities in the report (e.g., responsible for chapter 3 or section 2.5, etc)	Role(s) performed (e.g., chair of the team, secretary, editor, programmer, GIS specialist, data acquisitionist, etc)
Marieke van Arnhem	Introduction, 2.1, 2.7, 3.5, 3.6, 4.3, Further Research	Chair of the Team, Programmer, Visualiser
Walter Kahn	Discussion, Conclusion, Task Distribution, 4.2, 3.4, 2.6	Data Acquisition, Processor, Editor
Shawn Tew	Abstract, 1.1, 2.4, 3.1, 3.3, 3.6, 4.3, 4.5, Further Research	Data Acquisition, Programmer, Editor, Processor
Qiaorui Yang	Introduction for sections and subsections, 2.2, 4.1, 4.5, 5.1	Data Acquisition, Secretary
Xiaduo Zhao	Section 2.3, 2.5, 3.2, 4.2, Introduction for subsection 3.2 and 4.2	Data Acquisition, Programmer, Processor

Table 3: Task distribution between team members and roles in the project

The research took place during september, oktober and november of 2024 at the TU Delft. The original scope was laid out in the Project Identification Document (PID) and was as follows.

8.1 Scoping/ Managing Expectanations: MoSCoW

While conducting this research, multiple limitations have been identified that may impact the overall study. Gaussian splatting is a relatively new technique and can present a multitude of new challenges due to the limited availability of usable information and tools for implementation. One major limitation is the inability to collect complete scans of a building, particularly exterior roof surfaces, without the use of drones or advanced aerial scanning methods. Furthermore, the classification process and the training speed of machine learning models applied to point clouds may be reduced by the computational power available, particularly when handling large point cloud datasets of heritage buildings.

While data collection methods using devices like smartphones and stereo cameras offer accessibility and portability, they often fall short in precision compared to high-end laser scanners, impacting model quality. Permission restrictions on scanning heritage buildings also present obstacles to full reconstructions. To bridge these limitations and the requirements for accurate, high-quality models, careful consideration of data collection methods, computational resources, and permissions is needed, along with leveraging Gaussian splatting's flexibility to enhance reconstruction quality.

1. **Must**

- Perform literature review on each step of the pipeline
- Use 2 devices to collect input data for gaussian splatting.
- Use gaussian splatting on the collected building data.
- Clean the outliers and noise in the point cloud dataset.
- Perform semantic labelling approaches on a simple level of BIM standards. This means the points are labelled in their building function. For instance, floor, wall and roof.
- Compare the different devices, algorithms and classification methods.
- Capture the whole pipeline of the process. Either in film or with a different method.

2. **Should**

- Use addition devices to collect a point cloud dataset.
- Clean the collected dataset by removing unnecessary components. These components are moving objects such as people.
- Output a smart point cloud with semantics up to a higher level of BIM standards. This means the points are labelled in for example their material type.

3. **Could**

- Show and share the Gaussian Splats on the web.
- Have the possibility to filter the semantics on the web display.
- Do semantically labelling approaches on a difficult level. This means the points are labelled when they are added to the building.

4. **Won't**

- Create a Gaussian splatting Algorithm and tool.
- Create an automated pipeline, it will have manual elements such as cleaning
- Creating a 3D viewer
- Scan outdoors, due to limitations of building heights

Most of the **Must** elements were achieved. The literature is shown in Chapter 2. Only two actual devices were used for the data capture. Both the GeoSLAM and the phone were capable and show promising results for both the capture as well as the segmentation of the Gaussian splat. This was part of the research as a variety of devices and types of software had to be used to compare and contrast and this is currently all captured on the project's outcomes and source code used which can be found at <https://github.com/ShawnTew/Synthesis-Project-Group-4>.

References

- Acierno, M., Cursi, S., Simeone, D., and Fiorani, D. (2017). Architectural heritage knowledge modelling: An ontology-based framework for conservation process. *Journal of Cultural Heritage*, 24:124–133.
- Al-Bayari, O. and Shatnawi, N. (2022). Geomatics techniques and building information model for historical buildings conservation and restoration. *The Egyptian Journal of Remote Sensing and Space Science*, 25(2):563–568.
- Al-Sakkaf, A., Zayed, T., and Bagchi, A. (2020). A review of definition and classification of heritage buildings and framework for their evaluation. *Scientific Committee*, page 194.
- Bao, Y., Ding, T., Huo, J., Liu, Y., Li, Y., Li, W., Gao, Y., and Luo, J. (2024). 3d gaussian splatting: Survey, technologies, challenges, and opportunities. *arXiv preprint arXiv:2407.17418*.
- Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Cai, Z., Yang, J., Wang, T., Huang, H., and Guo, Y. (2024). 3d reconstruction of buildings based on 3d gaussian splatting. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-4/W10-2024:37–43.
- Carlier, A. (2023). 3d gaussian splatting blender add-on. <https://github.com/ReshotAI/gaussian-splatting-blender-addon>. Blender add-on for 3D Gaussian Splatting.
- Chao, J. T., Roskelley, C. D., and Loewen, C. J. (2021). Maps: machine-assisted phenotype scoring enables rapid functional assessment of genetic variants by high-content microscopy. *BMC bioinformatics*, 22:1–19.
- Computing, J. V. (2024). Postshot. Version 0.4.156.
- Cui, J., Cao, J., Zhao, F., He, Z., Chen, Y., Zhong, Y., Xu, L., Shi, Y., Zhang, Y., and Yu, J. (2024). Letsgo: Large-scale garage modeling and rendering via lidar-assisted Gaussian primitives. *arXiv preprint arXiv:2404.09748*.
- Cursi, S., Martinelli, L., Paraciani, N., Calcerano, F., and Gigliarelli, E. (2022). Linking external knowledge to heritage bim. *Automation in Construction*, 141:104444.
- Fausto-Robledo, A. (2021). Two years in – autodesk’s assistance to notre-dame in paris.
- Fisher, R., Perkins, S., Walker, A., and Wolfart, E. (1996). Hypermedia image processing reference. *England: John Wiley & Sons Ltd*, pages 118–130.
- Funk, C., Dornhege, C., and Burgard, W. (2018). Semantic labeling of indoor environments from 3d rgb-d maps. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1871–1878. IEEE.
- Gemeente Delft (n.d.). Kaart Delft Basis Viewer. Accessed: 2024-11-06.
- Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., and Markham, A. (2020). Randla-net: Efficient semantic segmentation of large-scale point clouds. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., and Markham, A. (2021). Learning semantic segmentation of large-scale point clouds with random sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

- International, B. (2013). Industry foundation classes (ifc) for data sharing in the construction and facility management industries. *ISO*, 16739:2013.
- Ji, Y., Liu, Y., Xie, G., Ma, B., Xie, Z., and Liu, H. (2024). Neds-slam: A neural explicit dense semantic slam framework using 3d gaussian splatting. *IEEE Robotics and Automation Letters*.
- Jurski, K. (2024). Semantic 3D segmentation of 3D Gaussian Splats. Technical report.
- Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering.
- Kerbl, B., Tatzgern, M., Speciale, P., Masnadi, S., and Wimmer, M. (2021). Neural point cloud rendering via multi-plane projection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7829–7838.
- Khalil, A., Stravoravdis, S., and Backes, D. (2021). Categorisation of building data in the digital documentation of heritage buildings. *Applied Geomatics*, 13:29–54.
- Koch, T., Körner, M., and Fraundorfer, F. (2016). Automatic alignment of indoor and outdoor building models using 3d line segments. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 10–18.
- Kumar, A., Anders, K., Winiwarter, L., and Höfle, B. (2019). Feature relevance analysis for 3d point cloud classification using deep learning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4:373–380.
- Laparra, V., Ballé, J., Berardino, A., and Simoncelli, E. P. (2016). Perceptual image quality assessment using a normalized laplacian pyramid. *Electronic Imaging*, 28:1–6.
- López, F. J., Leronés, P. M., Llamas, J., Gómez-García-Bermejo, J., and Zalama, E. (2018). A review of heritage building information modeling (h-bim). *Multimodal Technologies and Interaction*, 2(2):21.
- Martinelli, L., Calcerano, F., and Gigliarelli, E. (2022). Methodology for an hbim workflow focused on the representation of construction systems of built heritage. *Journal of Cultural Heritage*, 55:277–289.
- Maturana, D. and Scherer, S. (2015). Voxnet: A 3d convolutional neural network for real-time object recognition. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 922–928. IEEE.
- Megahed, N. A. (2015). Towards a theoretical framework for hbim approach in historic preservation and management. *ArchNet-IJAR: International Journal of Architectural Research*, 9(3):130.
- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2020). Nerf: Representing scenes as neural radiance fields for view synthesis. *CoRR*, abs/2003.08934.
- Moullou, D., Vital, R., Sylaiou, S., and Ragia, L. (2023). Digital tools for data acquisition and heritage management in archaeology and their impact on archaeological practices. *Heritage*, 7(1):107–121.
- Niantic, I. (2024). Scaniverse. Version 4.0.

- Penjor, T., Banihashemi, S., Hajirasouli, A., and Golzad, H. (2024). Heritage building information modelling (hbim) for heritage conservation: Framework of challenges, gaps, and existing limitations of hbim. *Digital Applications in Archaeology and Cultural Heritage*, page e00366.
- Polycam (2023). Gaussian splatting tool. <https://poly.cam/tools/gaussian-splatting>. Accessed: 2024-09-26.
- Poux, F., Neuville, R., Hallot, P., and Billen, R. (2016). Smart point cloud: Definition and remaining challenges. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4(W1).
- Poux, F., Neuville, R., Hallot, P., Van Wersch, L., Luczfalvy Jancsó, A., and Billen, R. (2017). Digital investigations of an archaeological smart point cloud: A real time web-based platform to manage the visualisation of semantical queries. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-5/W1:581–588.
- Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017a). Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660.
- Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017b). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30.
- Radanovic, M., Khoshelham, K., and Fraser, C. (2020). Geometric accuracy and semantic richness in heritage bim: A review. *Digital Applications in Archaeology and Cultural Heritage*, 19:e00166.
- Rocha, G., Mateus, L., Fernández, J., and Ferreira, V. (2020). A scan-to-bim methodology applied to heritage buildings. *Heritage*, 3(1):47–67.
- Rusu, R. B. and Cousins, S. (2011). 3d is here: Point cloud library (pcl). In *2011 IEEE International Conference on Robotics and Automation*, pages 1–4. IEEE.
- Schönberger, J. L. and Frahm, J.-M. (2016). Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Sewasew, Y. and Tesfamariam, S. (2023). Historic building information modeling using image: Example of port city massawa, eritrea. *Journal of Building Engineering*, 78:107662.
- Shafique, K., Khawaja, B. A., Sabir, F., Qazi, S., and Mustaqim, M. (2020). Internet of things (iot) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5g-iot scenarios. *Ieee Access*, 8:23022–23040.
- Tchapmi, L., Choy, C., Armeni, I., Gwak, J., and Savarese, S. (2017). Segcloud: Semantic segmentation of 3d point clouds. In *2017 international conference on 3D vision (3DV)*, pages 537–547. IEEE.
- Thomson, C. and Boehm, J. (2015). Automatic geometry generation from point clouds for bim. *Remote Sensing*, 7(9):11753–11775.
- Wang, J. and Shan, J. (2009). Segmentation of lidar point clouds for building extraction. In *American Society for Photogrammetry and Remote Sensing Annual Conference*, pages 9–13, Baltimore, MD.

- Weinmann, M., Jutzi, B., and Mallet, C. (2014). Semantic 3d scene interpretation: A framework combining optimal neighborhood size selection with relevant features. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3:181–188.
- Wu, T., Yuan, Y.-J., Zhang, L.-X., Yang, J., Cao, Y.-P., Yan, L.-Q., and Gao, L. (2024). Recent advances in 3d gaussian splatting. *Computational Visual Media*, pages 1–30.
- Wysocki, O., Hoegner, L., and Stilla, U. (2022). Tum-faÇade: Reviewing and enriching point cloud benchmarks for faÇade segmentation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-2/W1-2022:529–536.
- Wysocki, O., Zhang, J., and Stilla, U. (2021). Tum-faÇade.
- Yang, B., Dong, Z., Liang, F., and Liu, Y. (2016). Automatic registration of large-scale urban scene point clouds based on semantic feature points. *ISPRS Journal of Photogrammetry and Remote Sensing*, 113:43–58.
- Zhang, J., Zhao, X., Chen, Z., and Lu, Z. (2019). A review of deep learning-based semantic segmentation for point cloud. *IEEE access*, 7:179118–179133.
- Zhu, S., Qin, R., Wang, G., Liu, J., and Wang, H. (2024). Semgauss-slam: Dense semantic gaussian splatting slam. *arXiv preprint arXiv:2403.07494*.

