buildings

MDPI

Article

NERF2BIM: AI-Driven Detailing-on-Demand Through Sustainable Point Cloud Surveys and Semantic 3D Understanding for Advanced Modeling of Existing Buildings

Ivan Bratoev [1,*], Omar Faig Orujlu [1], Ziyang Xu [1], Ziya Erkoc [2], Matthias Nießner [2], Christoph Holst [1] and Frank Petzold [1]

[1] TUM School of Engineering and Design, Technical University of Munich, 80333 Munich, Germany
[2] TUM School of Computation, Information and Technology, Technical University of Munich, 80333 Munich, Germany
[*] Correspondence: bratoev@tum.de; Tel.: +49-89-289-22174

Abstract

The refurbishment and energy-efficient renovation of existing buildings, specifically those before 1945, with complex architectural building elements, require a level of building information that is often unavailable, incomplete or imprecise. As they represent a large percentage of current building stock (up to 25% of all buildings), it is crucial to address such an issue, as such buildings are ideal targets for renovation and energy retrofitting projects. This paper presents a conceptual pipeline, developed through the NERF2BIM research project, focusing on an Artificial Intelligence (AI)-supported holistic pipeline for the creation of as-is Building Information Modeling (BIM) models of existing buildings, expanding upon existing methodologies with the embedding of knowledge-driven semi-automatic detailing-on-demand task. The proposed pipeline integrates uncertainty-aware spatial capture, semantic interpretation and reconstruction, and knowledge-based reasoning within a BIM-oriented workflow. The paper provides an overview of current advancements in the respective aspects of the pipeline, highlighting current gaps. The proposed conceptual pipeline aims at addressing these issues through novel applications of AI and knowledge-driven solutions. Key contributions include: (1) a conceptual approach in addressing the imprecision of more sustainable data gathering approaches; (2) a context-aware BIM reconstruction process, providing multiple data output types; and (3) a formalization of architectural and construction knowledge and its utilization in a detailing-on-demand approach of the reconstructed BIM models. Through the integration of uncertainty-aware data gathering, context-aware reconstructions and domain expertise into existing reconstruction pipelines, the proposed pipeline bridges the data gap for existing buildings, enabling more efficient and knowledge-driven renovation processes.

Keywords: point clouds; Building Information Modeling; uncertainty evaluation; multi-modal Large Language Models; knowledge graphs; detailing-on-demand

Academic Editor: Yasser Mohamed

Received: 31 May 2026
Revised: 9 July 2026
Accepted: 11 July 2026
Published: 14 July 2026

1. Introduction

In the last few decades, BIM has seen a continuous increase in adoption in the Architecture, Engineering, and Construction (AEC) sector. It has supported the shift from fragmented, paper-based and manual labor-intensive workflows toward more automated and data-driven processes. With the increase in optimization, retrofitting, and refurbishment projects of existing building stock, especially in countries such as Germany, where they constitute 80% of all construction projects, the need for quality BIM-based solutions for the documentation of existing buildings (as-is) is imperative [1,2]. This trend is especially evident for older residential stock, such as buildings constructed before 1918 (>15%) and between 1920 and 1940 (∼13%), where the renovation and modernization potential is even higher. Such building stock does not usually have reliable architectural plans, and construction documentation is incomplete, outdated or in most cases, missing. This leads to complex and time-consuming manual on-site surveying procedures.

While new projects have BIM processes accompanying them during their full life cycle, the continuous alterations, repairs and adaptations of older building stock remain poorly documented. This shifts their analysis to a reverse design and planning process, requiring the approximation of the built environment from physical sampling of the building, such as with laser scanners and the analysis of their results. The outcome of such a process is not a precise representation of the captured environment, but rather an approximation, encoding in its representation an inherent level of uncertainty and vagueness. This representation consists only of the visible surfaces, both indoor and outdoor, of the building. Information regarding the materials of the building elements, the underlying structures and their composition remains unknown. This leads to the additional time-consuming and often manual tasks of reviewing existing documentation, if present, or contracting experts to provide additional assessments.

Current Scan-to-BIM and reverse-design approaches focus on the geometric reconstruction and segmentation of 3D point clouds. They often focus on the use of high-precision surveying setups that are costly, time-consuming and difficult to scale across multiple existing buildings. While emerging methods and solutions for uncertainty estimation in data acquisition, semantic scene understanding and reconstruction, and knowledge-driven detailing are being continuously developed, they are rarely integrated into coherent, holistic pipelines, and work in isolated, idealized scenarios.

A further gap in current Scan-to-BIM methodologies lies in how data acquisition and its utilization are linked. While sustainable, low-cost surveying methods are gaining popularity, the produced data is less precise, with unquantified uncertainty. This is in stark contrast to the semantic understanding and reconstruction pipelines, where methods are developed with the assumption of precise and idealized data. Such approaches lack the capabilities of integrating the vagueness of the obtained data and rarely provide additional mechanisms for guiding the reconstruction process. This leads to the necessity that existing methodologies evolve past purely geometry-centric tasks and explicitly incorporate domain knowledge into the process.

This paper proposes the conceptual pipeline NERF2BIM as an integrated, holistic, knowledge-driven, and uncertainty-aware process in order to address these discrepancies. The remainder of this paper is structured as follows. Following this introduction, Section 2 provides a comprehensive review of the current state of the art, covering existing approaches in Scan-to-BIM, semantic scene understanding, uncertainty estimation, and knowledge-driven reconstruction methods. Section 3 presents the proposed architecture in detail and introduces its main components, including uncertainty-aware point cloud acquisition, panoptic segmentation, knowledge graph construction, and semantic reasoning. Section 4 is devoted to the discussion of potential challenges, limitations, and implementation pitfalls that may arise during the development and deployment of the pipeline, as well as possible directions for future improvements. Finally, Section 5 concludes the paper by summarizing the key contributions of this proposed pipeline

https://doi.org/10.3390/buildings16142791

and outlining future research perspectives for advancing knowledge-enhanced Scan-to-BIM methodologies.

The key contribution of this proposed work can be summarized as follows:

- Uncertainty-aware point cloud acquisition for improved handling of occluded objects and missing points
- Panoptic segmentation for enhanced spatial awareness and accurate dimensional extraction
- Knowledge-driven inference of internal structure and material composition from external domain knowledge

2. State of the Art

The process of capturing a physical environment—either from aerial or ground-based scanning—and converting the acquired data into an intelligent 3D model using BIM software (e.g., Autodesk Revit, Graphisoft ArchiCAD) is referred to as Scan-to-BIM [3].

This approach is particularly valuable for generating as-is models of existing assets, which often lack information, have outdated drawings, or have incomplete documentation. Scan-to-BIM addresses these limitations by transforming point cloud data into a structured 3D reconstruction, where building elements are identified and labeled. The overall workflow typically consists of data acquisition, object recognition or semantic segmentation, and BIM modeling [4–7]. In the data acquisition stage, Light Detection and Ranging (LiDAR) technologies [8] are widely used for their high-accuracy point clouds. Photogrammetry can also be used; however, the accuracy of scans can degrade over large areas, and data acquisition can become time-consuming. Although the obtained laser point clouds are usually highly accurate, they can contain noise and redundant information. Here, data cleaning and downsampling approaches are applied to remove noise and errors, thereby improving the accuracy and efficiency of subsequent steps [9]. Following these steps, semantic segmentation of the point clouds is applied [10]. In semantic/instance segmentation, the elements in the scene are detected and classified. In the final stage, the segmented point clouds are converted into a parametric BIM model. This is typically achieved by fitting geometric primitives—such as planes for walls, floors, and ceilings, cylinders for columns and pipes, and bounding boxes for openings—to each segmented instance to recover its dimensions and orientation [6]. The fitted primitives are then mapped to standardized BIM element classes (e.g., IfcWall, IfcColumn, IfcSlab, IfcDoor) and organized into the project–site–building–story hierarchy required by the Industry Foundation Classes (IFC) schema, typically using open-source libraries such as IfcOpenShell [11]. Complementary techniques include rule-based topological inference to reconstruct relationships between elements (e.g., wall-hosts-window, slab-supports-wall) [12], model-fitting approaches that align pre-defined parametric building templates to the point cloud [13], and recently learning-based methods that directly predict structured BIM representations from segmented input [14]. Once assembled, the resulting BIM model is exported through open data formats such as IFC or gbXML, which ensure interoperability and facilitate information exchange between BIM platforms [11,15].

The methods surveyed in Table 1 reveal two consistent gaps in the current state of the art. First, the majority of automated Scan-to-BIM pipelines—including Campagnolo et al. [4], Roman et al. [11], and Chamseddine et al. [16]—produce geometrically accurate IFC models but carry no uncertainty information from the acquisition stage into the downstream representation, meaning that errors introduced by occlusions, noise, or scanner limitations are silently propagated without any formal record. Second, even approaches that incorporate semantic enrichment, such as Werbrouck et al. [17], rely on manual classification decisions and Linked Data provenance metadata but do not apply domain knowledge to infer building properties that are unobservable from the scan alone, such as structural composition or period-appropriate material configurations. The only method that addresses uncertainty at the acquisition level—Jarząbek-Rychard & Maas [18]—quantifies geometric uncertainty through geodetic stochastic propagation but does not carry this information forward into a semantically enriched or ontology-governed BIM output.

For high-precision engineering, accurately quantifying point cloud uncertainty is paramount. Managing data quality at the source ensures this critical uncertainty information is preserved for downstream processing tasks, such as fine registration, classification, and segmentation. Without incorporating structural semantics, the resulting BIM models are of limited utility for applications such as renovation, structural assessment, and maintenance planning.

To address these challenges, the proposed conceptual NERF2BIM pipeline is designed to jointly tackle all limitations. Per-point uncertainty is estimated directly at the acquisition stage and carried forward as an explicit attribute through segmentation, reconstruction, and reasoning, so that downstream decisions can be informed by the quality of the underlying measurements. A Large Language Model (LLM)-based joint 3D reconstruction and panoptic segmentation, leveraging large-scale generative priors, produces the model's geometric backbone. A Graph Convolutional Network (GCN)-based neuro-symbolic reasoning engine then combines this observed geometry with domain ontologies encoding the historical construction context—in the present case study, the 1895 Munich Bauordnung—to infer plausible structural configurations and material properties that cannot be observed directly from the scan. The resulting IFC model explicitly distinguishes measured from inferred attributes, making the epistemic status of every element traceable and reviewable by domain experts, and supporting advanced downstream applications such as renovation planning, structural analysis, and historical documentation.

2.1. 3D Laser Scanning and Uncertainty Evaluation

The rapid evolution of 3D laser scanning platforms has heightened the necessity for rigorous uncertainty evaluation to ensure data fidelity in high-precision surveying and mapping. Methodologies to quantify and mitigate these spatial uncertainties are broadly classified into three categories: forward modeling, backward modeling, and learning-based modeling.

Analytical frameworks utilizing forward modeling rely fundamentally on the law of variance–covariance propagation to quantify spatial uncertainty. This paradigm entails a rigorous theoretical decomposition of the entire laser scanning measurement mechanism, wherein discrete error sources are mathematically isolated to characterize the stochastic properties of the resulting point cloud. To account for the heterogeneous error budgets introduced by multi-sensor integration, researchers have executed several forward modeling evaluations that map sensor-specific variances to the final spatial coordinates [19–21]. Furthermore, established frameworks leverage these analytical error budgets as a primary mechanism for point cloud quality control and system calibration [22]. Beyond rigid geometric configurations, investigations have also focused on radiometric attributes, utilizing the intensity values to model and predict range measurement precision [23,24]. Although forward modeling remains highly valuable for its physical determinism, its implementation encounters substantial practical limitations when modeling complex, dynamic environments, prompting a distinct shift in contemporary research toward backward modeling alternatives.

Table 1. Comparison of Scan -to-BIM Methods Across Key Pipeline Dimensions.

Method	Data Acquisition	Uncertainty Estimation	3D Reconstruction	Segmentation-to-IFC Mapping	Knowledge Graphs/ Ontologies
Sacks et al. [12] (2017)	N/A	None	N/A	Rule-based topological inference	None
Werbrouck et al. [17] (2020)	TLS + Photogrammetry	Metadata-level (provenance, assumptions)	Point cloud geometry (Rhinoceros)	Manual + Linked Data classification	BOT, ifcOWL, RDF
Campagnolo et al. [4] (2023)	LiDAR	None	Primitive fitting (RANSAC, DBSCAN)	DNN instance seg. (BIM-Net++)	None
Jarząbek-Rychard & Maas [18] (2023)	TLS	Geodetic stochastic propagation	Plane fitting (indoor)	Rule-based	None
Roman et al. [11] (2024)	LiDAR/ Photogrammetry	None	DL-driven parametric fitting	DL-based IFC class prediction	None
Mehranfar et al. [13] (2024)	LiDAR	None	End-to-end AI; parametric fitting (Manhattan)	Point Transformer + rule-based IFC	None
Chamseddine et al. [16] (2026)	LiDAR	None	Hybrid learning + topology refinement	Semantic seg. $\rightarrow$ IFC (topology-aware)	Construction ontology (labeling only)
Chen et al. [14] (2026)	LiDAR/ Photogrammetry	None	DL-based seg. pipeline	Semantic seg. $\rightarrow$ parametric BIM	None
NERF2BIM (Proposed)	LiDAR (per-point σ)	Per-point, propagated throughout pipeline	NeRF + panoptic seg. + NURBS fitting	GCN knowledge enrichment + SHACL validation	Bauordnung rules, Architectural Elements typology, BOT

In contrast to analytical propagation, backward modeling relies on high-fidelity reference datasets as ground truth (GT) to empirically quantify spatial discrepancies. The foundational workflow centers on the precise and efficient calculation of predefined uncertainty metrics relative to this baseline. Standard methodologies frequently compute cloud-to-cloud (C2C), cloud-to-mesh (C2M), and multiscale model-to-model cloud comparison (M3C2) distances to characterize the geometric deviations between the acquired scans and the ground truth [25–29]. While these distance-based assessments are computationally straightforward and highly cost-effective, they fail to resolve six Degrees of Freedom (6 DoF) transformation parameters, which prevents a truly comprehensive three-dimensional representation of structural uncertainty.

To address this limitation, researchers have introduced point-based and line-based extraction techniques to capture more complex localized spatial variations [30,31]. Plane-based strategies are also extensively utilized, leveraging planar primitives within the scene to establish robust geometric correspondences and isolate systematic distortions [32,33]. The core mechanism of this advanced backward modeling relies on extracting distinct geometric features from the point clouds to guarantee correct data association, thereby enabling the rigorous quantification of uncertainty across all six DoF components. Within Mobile Laser Scanning (MLS) workflows, this capability is especially critical for isolating and evaluating trajectory estimation errors [34].

Due to its high adaptability and operational simplicity, backward modeling remains the dominant paradigm for quality assurance in the field. Although its direct dependence on ground truth verification ensures empirical reliability and accuracy, this requirement simultaneously constrains its scalability across expansive, unstructured real-world environments. Consequently, learning-based modeling has emerged as a promising alternative to overcome these operational bottlenecks.

Learning-based uncertainty evaluation represents an emerging paradigm within geodetic science, developing from concerted efforts to mitigate or entirely eliminate the strict dependence on physical GT data through data-driven estimation techniques. The foundational methodologies in this domain utilized ensemble learning methods to estimate the ranging precision of Terrestrial Laser Scanning (TLS) instruments by analyzing localized features such as backscattered intensity and beam incidence angles [35–37]. Subsequent investigations extended these concepts into the realm of deep learning, employing neural network architectures to predict and calibrate raw TLS ranging accuracy [38]. However, these pioneering studies were primarily restricted to isolating TLS distance measurement variations under controlled laboratory conditions, leaving the complex spatial dynamics of MLS point clouds unaddressed. Most recently, ensemble learning frameworks have been adapted to evaluate MLS uncertainty to determine point quality within complex, real-world operational scenarios [39,40].

To the best of our knowledge, most existing research on point cloud uncertainty mainly focuses on forward and backward modeling. Few publications apply machine learning to TLS for this purpose, suggesting that the field is still in its early stages of development. A further analytical summary can be seen in Table 2.

2.2. 3D Panoptic Segmentation and Reconstruction

The pursuit of holistic 3D scene understanding represents a foundational challenge in computer vision, robotics, and embodied artificial intelligence. Historically, the overarching objective of scene parsing was compartmentalized into discrete, loosely coupled tasks: geometric reconstruction mapped the spatial layout, semantic segmentation categorized amorphous background materials known as “stuff” (such as walls, roads, and ceilings), and instance segmentation identified and tracked distinct, countable foreground objects known as “things” (such as chairs, vehicles, and humans). Panoptic segmentation unifies these disparate objectives into a cohesive framework, requiring algorithms to assign both a continuous semantic class label and a unique, discrete instance identifier to every spatial element within a given scene.

Table 2. Comparison of 3D Point Cloud Uncertainty Evaluation Strategies.

	Forward Modeling	Backward Modeling	Learning-Based Modeling
Core Mechanism	Theoretical decomposition of scanning mechanisms based on error propagation laws.	Empirical quantification of spatial discrepancies against a high-precision GT data.	Data-driven estimation utilizing ensemble or deep learning to map local features to spatial uncertainty.
GT Dependence	No GT required	Strong dependence	Very low dependence
Core Limitations	Fails in complex, dynamic environments due to unmodeled physical variables.	Limited scalability because GT is hard to obtain in large-scale, unstructured environments.	Still in its infancy; early works were restricted to controlled, static laboratory settings.
Innovation & Trends	Physical determinism and solid theoretical foundation.	Empirical engineering and remains the current dominant paradigm for field quality assurance.	Data-Driven revolution with GT elimination and bypasses complex physical modeling. Real-World Scaling and evolve from TLS to MLS.

A seminal and foundational advancement in this domain is Panoptic Lifting [41]. Traditional approaches to 3D scene segmentation either required direct, labor-intensive 3D supervision (such as hand-annotated bounding boxes) or relied heavily on sparse ground-truth semantic annotations linked to active depth sensors. In stark contrast, the framework operates exclusively on standard, in-the-wild Red Green Blue (RGB) image inputs accompanied by off-the-shelf, machine-generated 2D panoptic segmentation masks inferred from a pre-trained 2D network, such as Mask2Former [42]. The underlying panoptic radiance field is designed to model radiance, continuous semantic class probability distributions, and discrete instance IDs for every coordinate in 3D space. To resolve the severe cross-view instance ID fluctuations inherent in 2D predictions, the methodology maps a machine-generated 2D instance to a global 3D surrogate identifier by solving a complex linear assignment problem using the Hungarian algorithm.

While explicit Gaussian representations are exceptionally fast and highly editable, they present a unique structural vulnerability: conventional 3D Gaussian Splatting (3DGS) lacks the inherent spatial smoothness regularization naturally provided by the continuous mathematical parameterization of Multi-Layer Perceptrons (MLPs) in Neural Radiance Fields (NeRF) frameworks. An MLP, by its nature, produces smooth interpolations between points in space. Gaussians, however, are independent entities. Consequently, when naively supervised with noisy 2D panoptic masks, standard 3DGS models are highly susceptible to generating spurious “floater” Gaussians, producing noisy 3D assets that do not resemble coherent objects, or suffering from severe over-segmentation where a single object is shattered into dozens of disjointed clusters. To mathematically mitigate this lack of implicit smoothness, advanced Gaussian architectures such as PanoGS [43] and Panoptic-Splatting [44] introduce highly sophisticated spatial regularization, attention mechanisms, and graph-based grouping protocols.

Frameworks such as Segment Any 3D Gaussians (SAGA) [45] immediately capitalized on this explicit structure for panoptic segmentation. Rather than querying a continuous field, SAGA directly integrates the features of 2D foundation models (like Segment Anything Model (SAM)) with the 3DGS representation by appending scale-related affinity features directly to the data structure of each individual Gaussian primitive. Through sophisticated Gaussian grouping algorithms and cross-view label voting mechanisms, object instances are explicitly segmented and tracked. Because the representation is explicit, this allows for immediate, real-time spatial interaction. If a user queries an object, the system instantly identifies the exact subset of Gaussians representing that object, allowing the user to delete, scale, or physically translate those specific primitives instantly without requiring network re-training.

To completely eliminate the need for per-scene optimization, researchers developed purely feed-forward architectures capable of predicting panoptic segmentation instantly. PanSt3R exemplifies this state-of-the-art approach [46]. It serves as a unified, feed-forward architecture that predicts both dense 3D geometry and multi-view consistent panoptic segmentation in a single forward pass, entirely bypassing the NeRF optimization loop. PanSt3R leverages the robust 3D representations generated by MUSt3R [47] (a scalable, multi-view geometric foundation model based on the DUSt3R [48] paradigm) to extract globally aligned 3D geometric features from completely unposed input images. Simultaneously, it utilizes DINOv2 [49] to extract deep 2D semantic features. The architecture concatenates these features to construct enriched frame tokens and mask features.

The most significant contemporary evolution in 3D scene parsing is the decisive shift toward "open-vocabulary" segmentation. Traditional models, bound by fully supervised learning paradigms, were intrinsically constrained to closed-set taxonomies. For example, networks trained on the SemanticKITTI [50] dataset could only recognize the exact 19 predefined classes (8 "things" and 11 "stuff" categories) present in the training corpus. However, real-world autonomy, robotic navigation, and embodied AI demand the ability to segment, identify, and reason about entirely novel object categories and terrain types on the fly, without access to direct 3D training examples for those novel classes. Achieving this requires embedding rich, generalized linguistic knowledge into the geometric 3D structure. Panoptic Vision-Language Feature Fields (PVLFF) tackles open-vocabulary 3D panoptic segmentation by learning a continuous radiance field alongside semantic and instance feature heads [51]. It effectively bridges modalities by distilling 2D vision-language embeddings directly into the 3D feature field and subsequently clustering the learned features to isolate specific instances.

To clearly delineate the contributions of the proposed pipeline against the state-of-the-art, five tangible performance and architectural dimensions critical for engineering-grade 3D scene understanding are established:

1. **Open-Vocabulary Semantic Flexibility:** The ability to segment and label arbitrary, out-of-distribution, or long-tail architectural classes without retraining or closed-set taxonomy constraints.
2. **Optimization-Free Inference:** The ability to process scenes in a feed-forward, single-pass pipeline without requiring costly per-scene optimization or neural radiance/splatting field training loops.
3. **Generative Geometric Completion:** The capability to actively predict, reconstruct, or retrieve high-fidelity geometry for heavily occluded, missing, or sparse architectural elements rather than merely parsing visible structures.
4. **Uncertainty-Aware Processing:** The capacity to natively ingest and propagate spatial unreliability or sensor noise metrics to adaptively guide the segmentation and reconstruction heads.
5. **Real-Time Spatial Interactivity:** The structural capacity to immediately edit, scale, or manipulate explicit 3D primitives dynamically post-inference.

As detailed in Table 3, existing frameworks demonstrate remarkable strengths in specific areas such as the real-time editing capabilities of Gaussian-based models or the rapid deployment of purely feed-forward architectures. However, a distinct structural gap remains: no current approach unifies open-vocabulary reasoning with generative, uncertainty-guided geometric reconstruction. The proposed pipeline is specifically designed to address these combined deficits.

Table 3. Deficit Analysis of Related 3D Panoptic Segmentation and Reconstruction Frameworks.

Method	Open-Vocab.	Opt.-Free	Generative Completion	Uncertainty Aware	Real-Time Interactive
Panoptic Lifting [41]	Closed	No	None	Partial	No
PanoGS [43]/Panoptic Splatting [44]	Yes	No	None	No	Partial
SAGA [45]	Partial	No	None	No	Yes
PanSt3R [46]	Closed	Yes	Interpolation	No	No
PVLFF [51]	Yes	No	None	No	No
Ours (Proposed)	**Yes**	**Yes**	**Full**	**Yes**	**No**

2.3. Knowledge Graph and Ontology-Based Detailing-on-Demand

The concept of a knowledge graph as a structured, machine-readable representation of entities and their interrelations originates in the semantic web research of the early 2000s [52]. A knowledge base is a structured dataset that captures real-world facts and semantic relationships as subject-predicate-object triplets. When these triplets are organized as a graph, the nodes represent entities and edges represent the relations between them. The result is referred to as a knowledge graph. The two terms are broadly considered equivalent and are frequently used interchangeably in the literature. The schema governing a knowledge graph is typically defined through an ontology, which formally specifies the concepts of a given domain and the relationships that hold among them. Consequently, ontology construction constitutes a foundational step in the development of any knowledge graph [53]. As formally characterized in the literature, a knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge that is not explicitly asserted but is logically implied by the combination of instance data and schema-level axioms [54]. This capacity for inference distinguishes knowledge graphs from conventional relational databases and makes them particularly powerful for domains where domain expertise must be encoded formally and reused automatically.

In the AEC domain, the earliest systematic applications of ontology-based knowledge representation appeared in the context of the IFC standard, which serves as the primary open schema for BIM. Researchers recognized early that IFC, expressed in the EXPRESS language, lacks the semantic expressiveness needed for inference, and began converting building models to Web Ontology Language (OWL)-based representations—a direction formalized through the ifcOWL ontology [55].

The motivations for applying semantic web technologies in AEC are threefold, as noted by [56]. First, interoperability: graph-based representations allow data from heterogeneous sources, such as point clouds, BIM files, regulatory texts, sensor streams, geographic information systems, to be integrated without imposing a single rigid schema, using Linked Data principles so that any dataset can reference and extend another. Second, inference: formal ontological rules allow implicit knowledge to be made explicit, for example, deriving energy-relevant properties from spatial topology, or inferring structural roles from element geometry, without requiring re-authoring of models for every downstream application. Third, traceability: the provenance of every inferred triple can be recorded, which is critical in as-is modeling contexts where reconstruction involves assumptions and uncertainties that must remain auditable [56].

The development of standardized, reusable ontologies for the building domain has been a core activity of the W3C Linked Building Data (LBD) Community Group, an international initiative established to develop interoperable Linked Data vocabularies for the AEC sector [57]. Among the outputs of this community, the Building Topology Ontology

(BOT) has emerged as the de facto lightweight backbone for representing the topological structure of buildings [58]. BOT defines a minimal set of classes and relations capturing the fundamental spatial composition of a building together with adjacency and containment relationships. Its deliberately compact design enables modular extension: BOT is intended to be combined with complementary ontologies addressing product information (e.g., the Building Product Ontology), properties (e.g., the Ontology for Property Management), geometry, sensors, or geographic information, allowing each application to assemble exactly the vocabulary it needs without committing to a monolithic schema. However, BOT intentionally omits any representation of the geometric detail of building elements, leaving these aspects to be supplied by external ontologies or, in the present pipeline, inferred from segmented as-is geometry using ontological rules.

Werbrouck et al. [17] proposed a framework to integrate Semantic Web tools into the Scan-to-BIM approaches, with the goal of including the information about uncertainties, sources, and modeling decisions. The core idea of this work is to represent existing buildings not just as geometric BIM models derived from point clouds, but as rich, interconnected knowledge graphs that include geometry, metadata, sources, and uncertainties. The authors develop a workflow and data model to link multiple data sources, document modeling assumptions and uncertainties, and enable interoperability across disciplines. They validate the approach through a case study of the Gravensteen Castle, where a point cloud-based reconstruction is semantically enriched and stored as a graph. In *SeeBIM*, the authors also proposed to semantically enhance the BIM model with a rule-based inference method [12]. The paper is significant in establishing that domain expert knowledge encoded as rules can drive automatic semantic completion of geometrically observed models. This approach was applied to a bridge and obtained very strong classification results. However, this work has shortcomings regarding scalability, adaptability, and manual rule selection. The rule library complexity grows rapidly with object type diversity, and validating large rule sets is difficult, particularly for partially occluded elements in point clouds. Lilis et al. [59] proposed a method that uses Geometric Relation Checking (GRC) on BIM models to extract topological information, which later uses it to enhance or generate a knowledge graph. Their approach comprises four steps: geometry extraction from IFC files, GRC processing, relation detection, and knowledge graph generation or enhancement. This work showed that geometric information can be used to infer missing semantic information. The work has promising results; however, its accuracy depends on the quality of the geometric model. It cannot recover the primary data, and errors in the BIM export can lead to false positives/negatives.

To clearly delineate the contributions of this subproject against the state-of-the-art in knowledge-enhanced Scan-to-BIM, five tangible dimensions critical for the targeted heritage building reconstruction are established:

1. **Scan-Based Input:** The ability to operate directly on as-is point cloud data, rather than requiring a pre-existing IFC or BIM model as input.
2. **Uncertainty-Aware Reasoning:** The capacity to ingest per-point acquisition uncertainty and use it to modulate the balance between observed geometry and ontology encoded priors during inference.
3. **Material & Substructure Inference:** The ability to infer attributes that are not directly observable from the scan, such as material composition, internal substructure, and load-bearing role.
4. **Period-Specific Ontology:** The use of a knowledge base grounded in the historical construction rules and architectural typology corresponding to the building's period of construction.

5. **Learning-Based Reasoning:** The use of a learned reasoning engine that combines symbolic priors with observed geometric evidence, rather than relying solely on manually authored rule sets.

As detailed in Table 4, existing frameworks demonstrate strengths in specific dimensions—such as the graph-based traceability and provenance handling of Werbrouck et al. [17], the rule-based inference of SeeBIM, or the topology-driven enrichment of Lilis et al. [59]. However, no current approach unifies scan-based input, uncertainty-aware reasoning, and inference of unobservable structural and material properties grounded in period-specific knowledge. The proposed subproject is specifically designed to address these combined deficits.

Table 4. Deficit Analysis of Related Knowledge-Enhanced Scan-to-BIM Frameworks.

Method	**Scan-Based Input**	**Uncertainty-Aware Reasoning**	**Material & Substructure Inference**	**Period-Specific Ontology**	**Learning-Based Reasoning**
Werbrouck et al. [17]	Yes	Partial	No	No	No
SeeBIM [12]	No	No	Partial	No	No
Lilis et al. [59]	No	No	No	No	No
Jarząbek-Rychard & Maas [18]	Yes	Partial	No	No	No
Ours (Proposed)	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

3. Overall Pipeline Architecture

To address the challenges of current Scan-to-BIM solutions, the NERF2BIM project proposes to combine learning-based uncertainty-aware data acquisition, AI-driven 3D panoptic segmentation and reconstruction of point clouds, and knowledge-based detailing-on-demand into a holistic, unified pipeline. To achieve this, the pipeline is subdivided into three key subprojects, each focusing on a specific stage of the pipeline. The following subsections outline the goals, proposed methodologies and the progress of each subproject. An overview of the pipeline is in Figure 1. As a preliminary case study, parts of the Main Campus of the Technical University of Munich (TUM) have been selected due to their distinct, well-documented architectural and construction principles.

3.1. Learning-Based Uncertainty Evaluation of MLS Point Clouds

3.1.1. Overall Goal

As an efficient and accurate point cloud acquisition technology, MLS has been widely adopted in numerous fields, including AEC, engineering geodesy, Scan-to-BIM, and Geographic Information System (GIS). In particular, over the past couple of decades, the advancement of MLS systems and corresponding point cloud processing algorithms has been a notably active research area, yielding substantial scientific achievements. In high-precision engineering tasks, it is crucial to accurately acquire and understand point clouds' uncertainty. Controlling point cloud quality at its source by evaluating uncertainty ensures this information is well-preserved through later data processing, such as registration, classification, and segmentation. This leads to consistent results across tasks and supports effective decision-making and process optimization. However, compared to the booming development in point cloud acquisition and post-processing research, studies on point cloud quality and uncertainty have fallen considerably behind, leaving a critical research gap. In other words, most current point cloud users either simply accept the accuracy specifications claimed by the scanning systems by default or overlook the importance and necessity of Point Cloud Quality Control (PCQC).

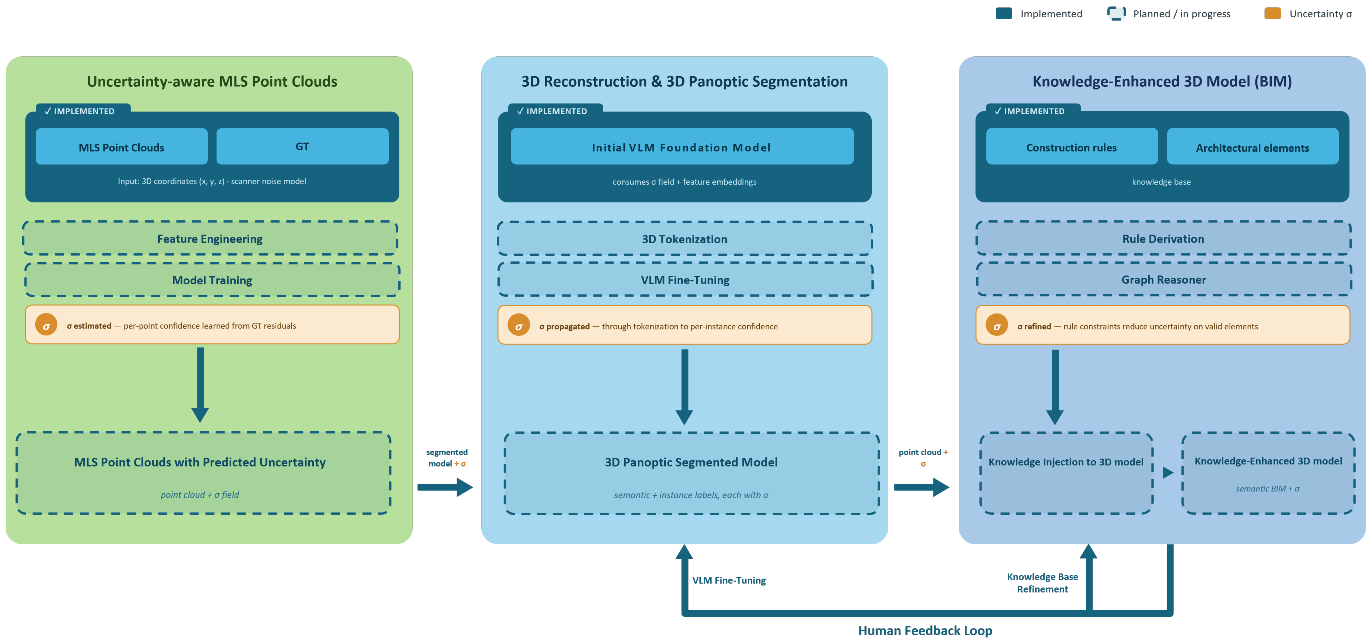

Figure 1. NERF2BIM conceptual pipeline.

To fill this gap, this subproject is dedicated to establishing a learning-based theoretical framework for evaluating MLS point clouds' uncertainty, and to exploring practical algorithmic foundations and engineering applications that bridge machine learning techniques with classical error quantification within the context of engineering geodesy. The specific objectives are as follows:

1. Analysis and comparison of different machine learning and optimization methods for point cloud error prediction.
2. Exploration of novel learning methodologies that integrate physical and geometric prior knowledge.
3. Development of a learning-based MLS point cloud uncertainty prediction framework to quantify point-wise errors and test its generalization performance.
4. Evaluation of the scientific insights and practical engineering gains of learning-based point cloud uncertainty evaluation, to further strengthen the research ecosystem for PCQC in urban digital twins.

3.1.2. Proposed Concept

Figure 2 illustrates the conceptual diagram of the proposed framework. The framework is designed as a single-pass pipeline primarily consisting of data input, model training, and error prediction components. The core of the framework lies in the model training component, specifically the learning-based error modeling. This component is designed to address the central question: "Are point cloud errors learnable and predictable?" Once training is complete, the framework enables the prediction of point-level errors for new point cloud inputs.

Data Input and Feature Representation. This module determines the way in which point cloud data is ingested by the trained model for error prediction, a matter that remains an open research question. Generally, conventional point cloud deep learning tasks, such as classification and semantic segmentation, utilize raw 3D coordinates as input for feature extraction and representation. Similarly, certain architectures employ radiometric or RGB attributes as auxiliary inputs to enhance performance. In the context of learning and predicting errors for MLS point clouds, however, the problem becomes significantly more complex due to the highly sophisticated nature of MLS error sources. The final point clouds represent a compound manifestation of various error sources that propagate and merge, inherently exhibiting a high degree of non-linear complexity.

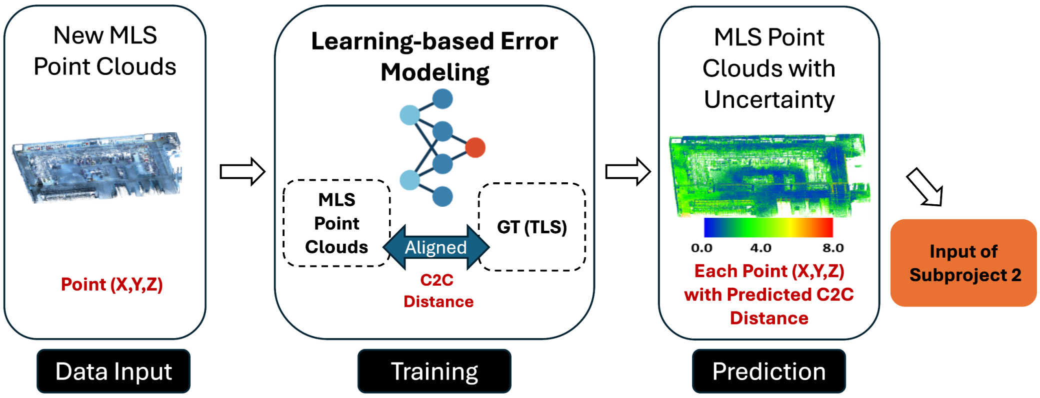

Figure 2. Proposed Learning-based Uncertainty Evaluation Framework for MLS Point Clouds. The core of Subproject 1 is to evaluate the uncertainty of the raw MLS point cloud through a learning-based framework. Afterward, these point clouds, which have been assigned uncertainty scores, will serve as input for Subproject 2.

Furthermore, the entire data acquisition and processing pipeline requires consideration of extensive physical properties and geometric constraints. For this reason, utilizing only the spatial coordinate information of each point may not be sufficient to support network convergence. Consequently, it is analyzed whether additional metadata from the data acquisition process can be incorporated as supplementary inputs, establishing this as a core research direction of the subproject.

Architectural Training and Prediction. This core technical module focuses on identifying and designing the optimal learning architecture tailored for MLS point cloud error modeling. Unlike traditional point cloud tasks, such as classification, segmentation, and registration, learning directly from point cloud errors represents a novel paradigm and consequently introduces three major challenges:

Complexity of MLS Error Propagation. For MLS platforms, uncertainty primarily originates from four main categories of error: instrumental, atmospheric, object- and geometry-related, and trajectory estimation. These error sources are highly correlated, and the underlying physical mechanisms and mathematical models of specific components remain partially uncharacterized.

Data Scarcity and Calibration Limits. To meet the rigorous training requirements of deep neural networks, extremely high-precision TLS datasets to serve as the ground truth baseline are required. This necessitates that alignment errors are minimized to the absolute limit. These training data must be harvested from complex, real-world scenarios and span a diverse array of error distributions. Currently, no public benchmarks are available for this specific task.

Diversity of Quantitative Metrics. Commonly utilized metrics for the quantification of uncertainty in MLS point clouds include C2C, M3C2, and C2M distances. Each metric exhibits specific characteristics and is often optimized for distinct application requirements. Consequently, adopting different distance metrics as prediction targets for identical training data yields varying uncertainty definitions. For instance, C2C distances comprehensively reflect positional deviations caused by combined systematic and random errors, whereas M3C2 distances suppress local noise to isolate the systematic bias of the MLS platform. The proposed learning framework must balance predictive accuracy for a single metric with generalizability across cross-metric transfers.

Evaluation Protocol. To comprehensively assess the proposed learning-based MLS point cloud uncertainty evaluation framework, the GT data will be derived from high-precision TLS point clouds that have undergone rigorous quality assurance. In addition, cross-sensor and cross-scenario tests will be conducted to further investigate the generalization capability of the model.

To address these challenges and achieve the overarching objectives, the technical route of this subproject is structured into three sequential parts:

Advanced Quantification Algorithms: Geometry- and data-driven point cloud uncertainty evaluation and mitigation algorithms are developed to establish reliable, high-precision solutions for quantifying point cloud errors. This provides the methodological foundation for subsequent learning-based solutions.

Benchmark Dataset Construction: An MLS point cloud uncertainty evaluation dataset specifically tailored for the built environment is constructed. This establishes a targeted data benchmark to facilitate subsequent network training.

Deep Learning Paradigm Exploration: Potential paradigms for deep learning-based point cloud uncertainty quantification and prediction are explored. This involves a comparative analysis of existing machine learning approaches, refining or proposing specialized network architectures optimized for MLS point cloud uncertainties, and validating these models using real-world empirical data.

3.1.3. Current Progress

The completed milestones establish a progressive foundation that bridges traditional error quantification with advanced data-driven prediction. To do so, a fully automated, plane-based framework dedicated to evaluating and mitigating trajectory estimation errors was first introduced. The mathematical reliability and structural robustness of this method were validated through two independent empirical experiments, proving its efficacy in real-world scenarios [34]. Parallel to this algorithmic development, the critical shortage of standardized benchmarks by engineering a specialized MLS point cloud dataset for complex indoor environments is also addressed. This repository captures multi-platform kinematic data paired with high-precision TLS baselines as ground truth, and it has been integrated into the public TUM2TWIN initiative to facilitate open-source validation [60].

Building upon these foundational data quality layers, the uncertainty frameworks are extended to critical downstream engineering applications. L2M-Reg, a novel plane-based LiDAR-to-Model fine registration framework optimized for individual building structures was developed. By explicitly modeling the generalization uncertainties inherent to Level of Detail (LoD) 2 semantic city models, L2M-Reg overcomes conventional alignment biases and delivers superior registration performance across five distinct real-world datasets [61]. Finally, to transition away from a strict dependence on physical ground truth, a machine learning pipeline that predicts point-level spatial uncertainty directly from localized geometric features was developed. A controlled benchmark between XGBoost and Random Forest implementations confirmed comparable predictive accuracy, with XGBoost achieving a significantly lower computational runtime that is highly favorable for scaling to massive datasets [39,40].

In summary, the completed efforts form a cohesive trajectory: moving from diagnosing and fixing core platform trajectory errors to building robust multi-sensor datasets, subsequently injecting uncertainty metrics into advanced object-level registration, and ultimately validating a machine learning proof of concept for direct error prediction. With these structural pieces firmly in place, the next objective centers on deep neural network architectures to further mature this learning-based paradigm.

3.2. 3D Panoptic Segmentation and Reconstruction

3.2.1. Overall Goal

In the AEC field, the creation of accurate BIM models from real-world environments is heavily reliant on 3D point clouds captured via LiDAR or photogrammetry. However, raw point clouds are fundamentally unstructured and frequently suffer from occlusions and sensor noise. Translating this raw spatial data into actionable, semantically rich architectural models currently requires intensive manual labor or fragmented computational pipelines that handle geometry reconstruction and object segmentation as isolated tasks.

Furthermore, traditional segmentation models are restricted to closed-set, predefined categories (e.g., "wall," "door"), failing to comprehend complex architectural nuances or user-specific queries (e.g., "load-bearing columns showing damage" or "Heating, Ventilation, and Air Conditioning (HVAC) ducts interacting with the dropped ceiling"). This open-vocabulary capability will support the human-in-the-loop process of the knowledge graph generation.

This subproject introduces a unified, multi-modal framework that leverages LLMs to perform simultaneous 3D panoptic segmentation and geometric reconstruction directly from point clouds. The primary objective is to transition from rigid, purely geometric processing to an interactive, semantic-driven understanding of architectural spaces. By utilizing a multi-modal LLM, the aim is to

1. Reconstruct missing geometry caused by occlusions (e.g., the floor space beneath built-in cabinetry).
2. Provide dense panoptic segmentation, simultaneously distinguishing continuous architectural elements ("stuff" like walls and floors) and distinct instances ("things" like individual doors, windows, and fixtures).
3. Enable open-vocabulary, language-driven interaction, allowing architects to query and segment spaces based on implicit intentions rather than hard-coded class labels. This will aid the human-in-the-loop process during knowledge graph generation.

3.2.2. Proposed Concept: The Multi-Modal 3D Pipeline

The proposed architecture integrates 3D geometry processing with the reasoning capabilities of a Vision-Language Model (VLM) (Figure 3). The pipeline is designed as a single-pass framework where spatial reconstruction and semantic reasoning mutually inform each other.

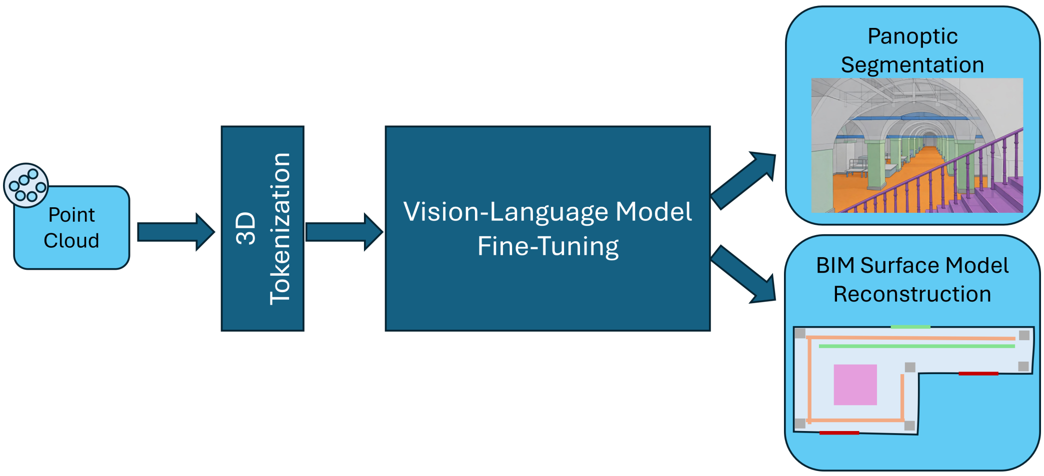

Figure 3. Proposed Vision-Language Model Framework for the panoptic segmentation and reconstruction of point clouds.

3D Feature Extraction and Tokenization. The pipeline ingests raw, unstructured, and potentially incomplete 3D point clouds augmented with pre-computed point-level uncertainty scores. Mathematically, the input is defined as $P = \{(p_i, u_i)\}_{i=1}^{N}$, where $p_i \in \mathbb{R}^d$ represents the spatial coordinates and local features (e.g., color, normals) of the i-th point, and $u_i \in [0,1]$ is its associated geometric uncertainty. Because LLMs natively process sequential tokens, a hierarchical point-transformer network [62] (functioning as a 3D encoder) to extract local and global geometric features is applied. This network clusters the raw point cloud into M local patches via Farthest Point Sampling (FPS). For each patch m with centroid $p_m^{(c)}$, the strategy generates a discrete, information-dense 3D visual token $v_m \in \mathbb{R}^D$ and an aggregate token-level uncertainty score U_m via

$$v_m = \text{MaxPool}_{i \in \text{patch}(m)} \left(\text{MLP}(p_i \oplus (p_i - p_m^{(c)})) \right)$$

$$U_m = \frac{1}{|\text{patch}(m)|} \sum_{i \in \text{patch}(m)} u_i$$

where $\oplus$ denotes concatenation. This tokenization process preserves crucial spatial relationships and multi-scale structural geometries while translating the data into a format the LLM can interpret.

Multi-Modal LLM Reasoning and Alignment. The core of the framework is a frozen, pre-trained multi-modal LLM [63,64], specifically adopting the LLaMA-3 architecture (e.g., the 8B parameter variant) for its robust zero-shot spatial reasoning and contextual understanding capabilities. The 3D tokens are projected into the LLM's embedding space alongside optional user-provided text prompts (e.g., "Segment all structural supports and reconstruct missing floor areas"). To ensure robustness against noisy scans, the aggregate uncertainty U_m is natively integrated into this reasoning phase. During cross-attention, U_m is designed to act as a dynamic confidence weight; tokens with high uncertainty would be down-weighted in the spatial attention matrix. Through cross-attention mechanisms, the LLM fuses the spatial 3D data with textual semantic embeddings. This allows the network to perform high-level spatial reasoning, resolving ambiguities in the point cloud by leveraging the LLM's vast prior knowledge of how physical objects and architectural spaces are typically organized.

Joint Decoder: Panoptic Segmentation and Reconstruction. Instead of branching into completely separate tasks, the outputs from the LLM feed into a joint decoder module that simultaneously tackles reconstruction and segmentation. First of all, Query-Based Panoptic Segmentation utilizes learnable object and region queries guided by Mixed-parameterized Positional Embeddings (MPE). These queries interact iteratively with the fused LLM features to group the point cloud into distinct masks. The LLM assigns an open-vocabulary semantic label and an instance ID to each mask, achieving full panoptic coverage (delineating every individual desk, door, wall, and ceiling). Secondly, for geometry reconstruction, a generative module (such as a coarse-to-fine masked autoencoder) predicts missing geometry. Here, the propagated uncertainty explicitly regulates the generative branch. Regions where the token uncertainty U_m exceeds a predefined threshold τ are structurally treated as "masked" patches. Instead of relying on degraded local geometry, the decoder relies on semantic priors. Crucially, this reconstruction is semantically conditioned. Because the LLM knows a specific cluster of points represents a "column", the generative branch can intelligently reconstruct the occluded back of that specific column, rather than blindly interpolating shapes.

Training Strategy. Low-Rank Adaptation (LoRA) is applied to the LLM's attention layers. The model will be trained on domain-specific architectural instructions paired with segmented point clouds (combining datasets like ScanNet, S3DIS, and custom AEC scans).

The loss function is a combination of focal loss for semantic classification and bipartite matching loss (Hungarian loss) for precise instance mask prediction.

For the generative branch, the model will be trained to output retrieval tokens mapped to a curated library of high-fidelity 3D architectural assets, with Geometric reconstruction losses (e.g., Chamfer Distance) to be coupled with semantic matching constraints to ensure retrieved objects perfectly fit the spatial and programmatic context of the room. Alternatively, reconstruction methods such as SAM3D [65] can be applied. To enforce these losses, reinforcement learning-based post-training is also explored.

Integration of Point Cloud Uncertainty. To address the inherent noise, sparsity, and artifacts in raw 3D scans, the proposed pipeline is designed to natively ingest uncertainty metrics calculated in the previous stage. The input point cloud is not merely treated as spatial coordinates and colors, but as augmented data where each point is paired with an external uncertainty score. The proposed framework actively propagates this signal through the tokenization, reasoning, and decoding stages to ensure robust performance even in degraded regions.

During the *3D Feature Extraction* phase, the point-transformer clusters the raw points into 3D visual tokens. Concurrently, the point-level uncertainty scores are aggregated (e.g., via feature-weighted averaging) to produce a unified confidence score for each resulting token.

In the *Multi-Modal LLM Reasoning* stage, these token-level confidence scores directly modulate the cross-attention mechanisms. The pipeline uses the uncertainty values as a dynamic attention mask; tokens with high uncertainty (e.g., indicating highly reflective surfaces or severe occlusions) are down-weighted. This intentionally forces the LLM to rely more heavily on the user-provided textual prompts and its own semantic priors to infer the spatial context of that region, effectively bridging unreliable geometric gaps with high-level world knowledge.

Finally, within the *Joint Decoder*, the propagated uncertainty is designed to dictate the division of labor between the segmentation and reconstruction heads. Regions characterized by uncertainty exceeding a defined threshold would be treated analogously to "masked" patches in a masked autoencoder framework. Instead of forcing the panoptic segmentation head to guess labels based on unreliable geometry, the generative module would be tasked with reconstructing those specific areas from scratch, strictly conditioned on the LLM's semantic understanding of the surrounding space. The final panoptic segmentation will also output the uncertainty value per face.

Computational Requirements. The primary computational bottleneck of the architecture stems from the dual burden of the hierarchical 3D point-transformer encoder and the autoregressive inference of the LLaMA-3 backbone. Processing a typical indoor scan requires significant Video Random-Access Memory (VRAM) allocations during the forward pass. The 3D point-transformer scales quadratically with respect to the number of localized patches before token aggregation. While LoRA significantly reduces the trainable parameter footprint (<1%), the model cannot achieve real-time inference on consumer-grade hardware. Local execution is constrained to high-end workstation Graphics Processing Units (GPUs) (e.g., NVIDIA RTX 4090 or enterprise A100/H100 clusters (NVIDIA Corporation, Santa Clara, CA, USA)), where a single inference pass spanning tokenization, LLM alignment, and joint decoding is expected to complete in less than a minute depending on scene complexity.

Scalability Challenges. Scalability is bounded heavily by the context window limits of the underlying LLM and the spatial density of the point clouds. While architectures like LLaMA-3 natively support large context windows, representing massive, multi-story architectural environments as sequential 3D tokens causes rapid sequence length explosion.

To scale the pipeline from individual rooms to full building scales, a spatial chunking strategy is employed. The global point cloud is partitioned into localized, overlapping volumetric blocks (e.g., 5 m × 5 m regions). However, this chunking introduces a significant scalability challenge: the loss of long-range global context. For example, a structural beam spanning across multiple chunks may be misidentified or inconsistently reconstructed at block boundaries. To mitigate this, a rolling coordinate cache or downsampled global topological map must be appended to the text prompt to maintain macro-spatial awareness across chunks, increasing prompt management complexity.

3.2.3. Reliability, Verification, and Deployment Limitations

While the integration of VLMs and LLMs provides unprecedented zero-shot spatial reasoning for 3D processing, deploying these architectures in strict engineering contexts necessitates addressing their inherent reliability challenges and computational overhead.

Justification and Reliability. Traditional deterministic 3D segmentation models often fail when confronted with highly ambiguous, heavily occluded, or context-dependent geometries. The use of multi-modal LLMs specifically is justified due to their ability to leverage broad semantic world-knowledge to resolve these physical ambiguities (e.g., deducing that a partially scanned flat plane above four legs is a desk). However, because LLMs are generative by nature, they are susceptible to physical hallucinations. To ensure reliability, the proposed pipeline is designed to restrict the LLM's role to semantic reasoning and region proposition, while anchoring the actual spatial output strictly to the deterministic coordinates of the input 3D visual tokens.

Verification Mechanisms. To meet the stringent safety and accuracy requirements of engineering applications, the multi-modal outputs must be physically verified rather than blindly trusted. The proposed framework introduces geometric verification mechanisms acting as a post-processing guardrail: reconstructed geometries generated by the joint decoder would be subjected to free-space conflict checks against the original point cloud's known ray-casting bounds. Furthermore, semantic labels assigned by the LLM (e.g., "load-bearing column") can be cross-verified against expected programmatic topologies (e.g., structural elements must intersect the floor and ceiling planes). Any generated geometry that violates these hard deterministic physical constraints would be flagged for user review, ensuring human-in-the-loop verification.

Practical Deployment Limitations. Despite its reasoning capabilities, the proposed pipeline faces distinct limitations regarding practical deployment. First, the tokenization of large-scale, high-density point clouds results in massive sequence lengths that approach or exceed the context windows of current LLMs. Consequently, large architectural spaces must be chunked and processed sequentially, which can break global contextual understanding. Secondly, the computational footprint and memory requirements of querying a frozen, multi-modal LLM prohibit real-time, edge-device processing (e.g., running inference locally on mobile LiDAR scanners). Current practical deployment is therefore limited to asynchronous, cloud-based workflows where scan data is uploaded, processed offline, and returned to the user. Future work will explore distilling the LLM's spatial reasoning capabilities into smaller, edge-deployable traditional networks.

3.2.4. Evaluation Protocol

To rigorously assess the performance of the proposed Multi-Modal 3D Pipeline, a comprehensive evaluation framework focusing on both panoptic segmentation and semantically-conditioned reconstruction is established. The evaluation protocol is designed to measure the quantitative accuracy of the joint decoder outputs, compare against state-of-the-art baselines, and qualitatively validate the reasoning capabilities of the multi-modal architecture.

Datasets and Benchmarks. Evaluation will be conducted on the standard validation and hidden test splits of prominent 3D indoor scene datasets, specifically ScanNet [66] (including ScanNet200 [67] for long-tail open-vocabulary assessment) and ScanNet++ [68]. This will assess the model's generalization capabilities across standard indoor environments and specialized architectural spaces.

Quantitative Metrics. For panoptic segmentation, the standard metrics will be employed: Panoptic Quality (PQ), Segmentation Quality (SQ), and Recognition Quality (RQ). Additionally, the mean Intersection over Union ($mIoU$) for semantic segmentation and Average Precision (AP) across varying thresholds (AP_{25}, AP_{50}) for instance tracking will be reported. For the geometry reconstruction and retrieval branch, structural fidelity will be measured using Chamfer Distance (CD) between the predicted geometries and the ground-truth models.

3.2.5. Current Progress

Preliminary experiments developed an initial 3D tokenization module and achieved stable integration with the chosen multi-modal LLM backbone. The scene synthesis capabilities have also been advanced by fine-tuning LLMs to generate complete room-scale environments utilizing 3D object retrieval.

By leveraging the LLM's understanding of spatial co-occurrence patterns (e.g., placing task seating specifically near desks and power access), the system can populate and reconstruct predicted layouts with high-fidelity, semantically appropriate 3D assets drawn from extensive libraries. This generative retrieval step effectively bridges the gap between abstract semantic understanding and actionable architectural layout creation, demonstrating strong proof-of-concept for handling complex interior programming.

The next step is to tightly couple the loss functions of the point-cloud reconstruction and segmentation branches. This ensures that the panoptic queries actively constrain the geometric generation, preventing structural hallucinations and improving boundary precision between retrieved assets and existing architecture.

A transition from standard indoor benchmarks to complex, noisy AEC datasets is planned as well. The model will be fine-tuned on point clouds derived from active construction sites, heritage building scans, and complex Mechanical, Electrical, and Plumbing (MEP) installations to improve the LLM's domain-specific reasoning.

The pipeline will also be optimized to handle massive, building-scale point clouds via localized fragment processing. Concurrently, the interactive prompt-interface allowing users to dynamically query, swap retrieved objects, and edit the segmented BIM model will be developed.

3.3. Knowledge Injection and Detailing-on-Demand

3.3.1. Overall Goals

As discussed above, predicting the internal structure of elements in the given scene is a fundamental challenge in Scan-to-BIM frameworks. Remote sensing methods can capture only the surfaces of surrounding elements, not the materials and structures they are built upon [11,69]. Consequently, enhancing the resulting 3D model with information about material composition, internal substructure, and load-bearing properties is currently a manual, time-consuming process that requires substantial input from domain experts [6].

The overall goal of this subproject is to automate this enrichment step through a knowledge-driven detailing process. Specifically, the uncertainty-aware, panoptically segmented 3D model produced by the upstream components of the pipeline is intended to be augmented with semantic, structural, and material information inferred from a domain-specific knowledge base. This transforms the segmented geometric representation into a semantically complete and application-relevant BIM model, suitable for downstream tasks such as renovation planning, structural analysis, and historical building documentation. The overall framework is illustrated in Figure 4 and is detailed in the following subsections.

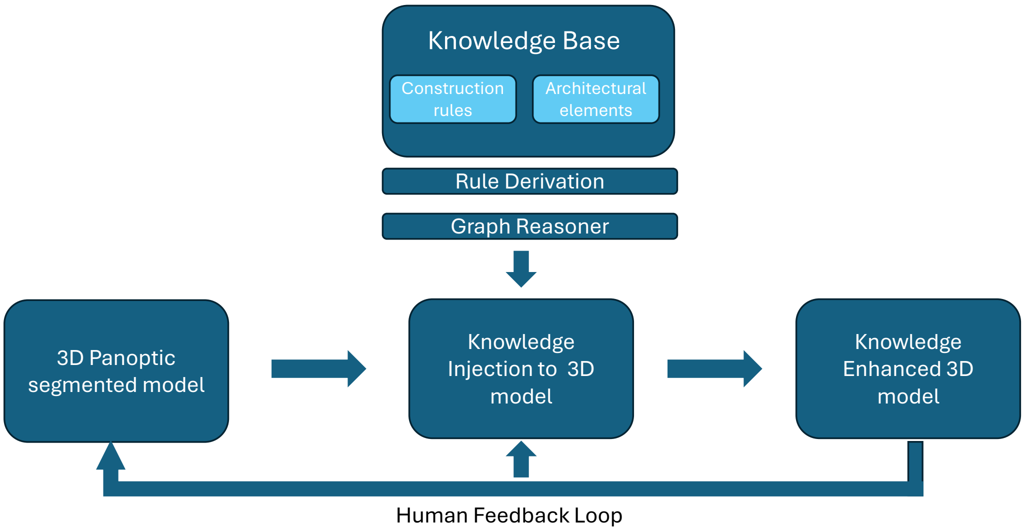

Figure 4. Proposed Framework for Knowledge injection into the 3D model.

3.3.2. Proposed Concept

The proposed concept is centered on a structured knowledge base composed of two interconnected ontologies: *Construction Rules* and *Architectural Elements*.

Ontologies. The *Construction Rules* ontology formalizes historical construction regulations and engineering practices that governed building design at the time of construction. In the present work, the construction-rules ontology is derived from the building regulations and construction standards specific to the case study under investigation. By drawing on the regulatory framework that governed building design at the time and location of the investigated structure, the ontology provides a contextually accurate basis for reasoning about structural design principles, material selection, and dimensional constraints relevant to the targeted building. The same conceptual approach is transferable to other buildings and construction periods: in each case, the ontology is populated with the historical regulations applicable to the specific context, ensuring that the encoded knowledge reflects the construction practice of the period rather than relying on general assumptions. The *Architectural Elements* ontology describes the geometric, proportional, and relational properties of architectural components. It encodes knowledge about structural and non-structural building elements such as walls, columns, beams, floors, ceilings, windows, and doors, including their dimensional constraints, spatial dependencies, and characteristic proportions [70,71]. This ontology captures the architectural design patterns commonly used during the building's construction. Based on these ontological representations, a unified knowledge base is constructed, from which logical rules and inference constraints are derived. These rules define valid relationships between building components, permissible dimensional ranges, likely material compositions, and structural dependencies between elements [17,56].

Graph-based Reasoning Engine. A reasoning block is planned to automatically identify building elements, their interrelations, and additional attributes such as material and internal structure. The reasoner will be trained on the axioms and rules encoded in the Construction Rules and Architectural Elements ontologies, together with the panoptically segmented scene graph produced by the upstream pipeline. The ontologies provide symbolic priors on element types, dimensional ranges, spatial dependencies, and material associations, while the scene graph provides the geometric evidence observed in the actual building. By combining both sources, the reasoner is designed to infer attributes that are not directly observable from the scan, such as material composition, internal substructure, load-bearing role, and permissible dimensional constraints. Per-point uncertainty from the acquisition stage enters the reasoning process as a node-level signal: in regions of low uncertainty, the reasoner relies more on the observed geometry; in regions of high uncertainty caused by occlusions or noisy returns, it would shift weight toward the ontology-encoded priors. This step can also be seen as 3D Scene Graph Generation, in which the model learns not only relationships between elements but also attributes of their material and structure [72,73]. The intended result is a knowledge-enhanced scene graph in which each node represents a building element enriched with geometric attributes, material properties, structural roles, dimensional constraints, and uncertainty estimates originating from the acquisition stage [72]. The final pipeline output would therefore be not merely a geometric reconstruction but a semantically enriched BIM model, suitable for downstream applications such as renovation planning, structural analysis, maintenance assessment, and historical building documentation [11].

Uncertainty Propagation. The per-point uncertainty σ estimated at the acquisition stage is carried forward as a first-class attribute through every subsequent stage of the pipeline rather than being consumed locally. During panoptic segmentation, per-point σ values are aggregated into per-instance confidence scores that down-weight points with high acquisition uncertainty in the segmentation loss. In the Non-Uniform Rational B-Splines (NURBS)-based surface reconstruction stage, point-level σ propagates into surface-fitting confidence bounds, yielding per-surface uncertainty envelopes that are stored alongside the fitted geometry. These envelopes are then attached to the corresponding nodes of the 3D scene graph as explicit uncertainty attributes, making the epistemic status of every geometric measurement visible to the reasoning engine. The GCN-based reasoner uses these attributes directly as input features and as edge-weighting signals during message passing: in regions of high σ—typically caused by occlusions, grazing-angle returns, or material-induced noise—the reasoner shifts weight from the geometric evidence toward the ontology-encoded priors, so that inferred values are produced from domain knowledge rather than unreliable measurements. The final IFC output preserves this distinction through the $*$ convention, with measured (σ) and inferred (σ^*) values exported as separate property sets, ensuring that uncertainty information remains auditable through the full BIM generation workflow rather than terminating at acquisition.

Human-in-the-Loop. To further improve reliability and adaptability, the framework is designed to incorporate a human-in-the-loop feedback mechanism. Domain experts would inspect the generated model, identify inconsistencies, and provide corrections where necessary. These corrections would be applied not only to the final reconstruction but are also fed back into the system: expert feedback is used to refine the reasoning rules, update the knowledge base, and adjust the parameters of the underlying neural components [74]. This continuous feedback loop enables iterative improvement of both the semantic reconstruction quality and the system's overall generalization capability.

3.3.3. Current Progress of the Subproject

At the current stage of the subproject, the focus has been placed on establishing the foundational components of the knowledge-driven detailing pipeline: the two underlying ontologies and the initial design of the graph reasoning engine. As a case study, a building located on the main campus of the Technical University of Munich was selected. Parts of this campus were constructed in the 1870s under well-defined construction principles that were strictly followed at the time, providing a representative testbed for evaluating the proposed approach on a historically significant building.

The first ontology, formalizing the historical construction rules and regulations applied in Munich at the time of the building's construction, has been implemented based on the 1895 Munich *Bauordnung (Building Regulations)* [75]. The ontology currently encodes a structured representation of regulatory constraints, including the minimum thicknesses of various wall and ceiling types, fire prevention requirements, and material-related restrictions. The second ontology, representing the architectural elements characteristic of the investigated building, has likewise been implemented in its initial form. It encodes a taxonomy of structural and non-structural components—including walls, columns, vaults, staircases, beams, and ceilings—together with their typical proportions, dimensional ranges, and material associations as documented in the relevant architectural literature of the period [70,71].

Both ontologies are being further refined as the historical documentation of the testbed building is investigated in greater depth. Since several parts of the TUM main campus building have been repaired or reconstructed using design and construction philosophies that differ from the original 1870s and 1890s practices, the ontology development is necessarily an iterative process. In the next stage, the two ontologies will be merged into a unified knowledge base from which logical inference rules and dimensional constraints can be derived systematically.

In parallel with the ontological work, the graph reasoning engine that operates on the knowledge base is currently under development. The reasoner is being implemented as a GCN [76] that takes the panoptically segmented scene as input and produces, in combination with the rule constraints derived from the knowledge base, an enriched 3D Scene Graph. By formulating the reasoning task as message passing over a heterogeneous graph, the GCN-based architecture is able to combine the observed geometric evidence with the symbolic priors encoded in the ontologies, thereby providing a principled mechanism for injecting domain knowledge into the scene understanding pipeline [72,77].

The immediate next steps of the subproject are threefold: (i) extending and refining the two ontologies with additional historical sources and period-specific construction principles; (ii) consolidating both ontologies into a unified knowledge base and deriving a complete set of inference rules; and (iii) finalizing the GCN-based reasoning engine and validating its behavior on the segmented representations provided by the upstream components of the NERF2BIM pipeline.

3.3.4. Validation Strategy

The validation strategy treats measured and inferred quantities under separate evaluation protocols, since they are observable to fundamentally different degrees. Geometric accuracy of the reconstructed surfaces will be assessed against the georeferenced point cloud using Chamfer distance, surface-to-surface deviation, and per-class Intersection over Union (IoU) on the panoptically segmented input. Scene graph quality will be evaluated at two levels: node-level element classification using precision, recall, and F1 against expert-annotated ground truth, and edge-level relation prediction (e.g., *hosts*, *supports*, *adjacentTo*) using accuracy and per-relation F1, with held-out rooms of the case-study building serving as test data. Inferred attributes such as material composition, internal substructure, and load-bearing function cannot be measured from the scan and will instead be validated against the as-is building documentation, historical archival records, and review by domain experts in historic construction, with inter-rater agreement quantified using Cohen's κ. Structural validity of the final IFC export is verified through closed-world constraint checks, reporting the fraction of element nodes that pass all checks prior to export. Because the case study consists of a single building ($n = 1$), generalization claims are explicitly bounded; cross-validation is therefore performed across architecturally distinct sub-regions of the building, such as different wings, vault types, and corridor configurations, rather than across separate buildings. Finally, an ablation study comparing the full neuro-symbolic system against a geometry-only baseline will isolate the contribution of the knowledge layer to the final reconstruction quality.

4. Discussion

While the proposed pipeline introduces a holistic approach for an uncertainty-aware, semantically enriched BIM reconstruction, there are distinct challenges associated with its applications to real-world scenarios. The following sections discuss limitations, potential improvements, and future directions for each subproject.

4.1. Challenges and Future Work in Uncertainty Evaluation of MLS Point Clouds

Model generalization ability. In the context of 3D point cloud analysis, model generalization ability remains a paramount challenge, particularly when deploying automated algorithms across diverse operational environments. Neural networks often exhibit a phenomenon known as domain shift, where models trained on data from one specific sensor or scene geometry suffer significant performance degradation when applied to unseen data. This vulnerability is especially critical in applied measurement and engineering domains, where structural diversity, varying point densities, and distinct sensor noise profiles are inevitable. An effective learning framework must therefore transcend the mere memorization of source-domain distributions, establishing robust representations that maintain predictive fidelity across highly heterogeneous datasets. The capacity to generalize is not simply a measure of average empirical accuracy; rather, it reflects the model's intrinsic understanding of underlying spatial structures independent of the acquisition hardware. Without robust generalization, the transition from controlled laboratory benchmarks to unstructured, real-world environments introduces unpredictable error margins, thereby compromising the integrity of downstream tasks like construction quality control. Thus, bridging the generalization gap by enabling models to dynamically adapt to target-domain characteristics is the fundamental prerequisite for scaling point cloud intelligence into universal, sensor-agnostic solutions.

Uncertainty in Model Predictions. Addressing uncertainty in model predictions is an indispensable component of reliable point cloud processing, serving as a critical boundary between confident algorithmic execution and potential catastrophic failure. Standard deep learning architectures typically output deterministic point estimates, projecting a false sense of absolute certainty even when confronted with highly ambiguous, sparse, or occluded spatial data. In realistic scenarios, point clouds are inherently corrupted by sensor artifacts, reflective surfaces, and registration errors, making deterministic predictions not only misleading but fundamentally dangerous in safety-critical applications. By rigorously quantifying predictive uncertainty, a framework provides actionable statistical guarantees, allowing downstream decision-making systems to differentiate between trustworthy measurements and regions requiring human intervention or secondary sensor validation. This paradigm shift from blind prediction to calibrated risk assessment ensures that the model can gracefully acknowledge its own limitations, flagging out-of-distribution anomalies or noisy patches rather than hallucinating reliable geometries. Ultimately, embedding robust uncertainty quantification within point cloud pipelines transforms raw spatial coordinates into a probabilistic safety envelope, establishing a verifiable trust metric that is essential for deploying automated quality assessment systems in strict engineering domains.

Physical-informed or Geometry-guided Error Learning. The integration of physical-informed or geometry-guided error learning represents a vital evolution from purely data-driven statistical modeling toward domain-aware artificial intelligence. Traditional machine learning methodologies often treat point cloud error prediction as an abstract numerical regression task, entirely divorced from the physical realities of the spatial environments they represent. This mathematical blindness frequently results in statistically optimal but physically absurd outputs, such as unbounded confidence intervals that suggest negative distances or structural deformations that violate spatial logic. By embedding explicit geometric constraints and physical priors into the learning process, models are inherently forced to respect the spatial integrity of the scanned objects. Geometry-guided frameworks leverage the intrinsic topological features of the point cloud, such as surface normals, local curvatures, and spatial continuity, to anchor the statistical learning process in physical reality. This approach not only explicitly truncates impossible negative bounds but also naturally sharpens the predictive intervals around complex architectural features. Consequently, fusing physical awareness with statistical learning bridges the critical gap between abstract algorithmic optimization and practical engineering compliance, ensuring that the generated safety bounds are physically plausible, rigorously verifiable, and directly applicable to structural tolerance assessments.

Practical Deployment Limitations. Although the proposed learning-based uncertainty evaluation system shows great promise, it still faces numerous practical limitations. First, large-scale, high-quality training data is scarce, and acquiring such data is extremely costly. Second, different application scenarios often require different metrics for quantifying point cloud uncertainty; therefore, developing a flexible framework that can accommodate various quantification metrics is of great practical significance.

4.2. Challenges and Future Work in 3D Panoptic Segmentation and 3D Reconstruction

While the integration of multi-modal Large Language Models with 3D panoptic segmentation and generative reconstruction offers a paradigm shift for the generation of BIM models from 3D Point Clouds, several significant technical challenges remain. Addressing these bottlenecks will form the basis of the future research directions.

Computational Bottlenecks and Scalability. Architectural point clouds are inherently massive. The standard paradigm of tokenizing 3D point clusters and feeding them into dense LLM transformer blocks creates extreme memory and computational overhead. Scaling this architecture from single-room environments (e.g., ScanNet datasets) to multi-story building complexes within reasonable inference times remains a critical bottleneck.

Geometric Fidelity versus Semantic Hallucination. While LLMs excel at understanding spatial co-occurrences (e.g., knowing a window belongs inside of a wall), they lack an inherent understanding of rigorous physics and hard geometric boundaries. A major challenge in retrieval-augmented 3D reconstruction is preventing "hallucinations" where retrieved assets physically intersect with existing architecture (e.g., a generated door not properly aligned with the opening in the wall) or violate dimensional constraints.

Domain-Specific Data Scarcity. Current multi-modal LLMs are predominantly trained on generic, open-world image-text pairs or simplified indoor 3D datasets. There is a severe lack of high-quality, densely annotated datasets specific to the AEC domain.

To overcome these limitations and push the pipeline toward real-world deployment, future work will focus on the following core areas:

Integration of Mixture-of-Experts (MoE) Architectures. To resolve the computational bottlenecks, a transition from the LLM backbone to a sparse 3D MoE architecture is planned. The aim is to drastically reduce the number of activated parameters during inference by routing specific types of architectural queries (e.g., "reconstruct missing geometry" vs.

"identify opening elements") to specialized, lightweight expert layers, thus enabling near real-time interaction with large-scale point clouds.

Transition to Continuous 3D Representations. Shifting the underlying data structure from discrete point clouds to 3DGS will be explored. By embedding semantic LLM features directly into the covariance and opacity attributes of 3D Gaussians, continuous, high-fidelity scene representations can be achieved. This will allow for smoother panoptic boundaries compared to discrete voxel or point-based methods.

Physics-Informed Collision Constraints. To solve geometric hallucinations during object retrieval, a differentiable physics engine will be integrated into the generative loss function. This will enforce hard boundary constraints, ensuring that any LLM-driven reconstructed geometry strictly adheres to collision volumes, structural logic, and the precise dimensional bounds of the scanned architectural environment.

Online Panoptic Mapping for Dynamic Sites. Future iterations of the pipeline will move from offline, static processing to online, sliding-window mapping. This will allow the LLM to continuously update its semantic understanding of a space in real-time as a robot or surveyor walks through an active construction site, dynamically resolving occlusions and tracking state-changes over time.

4.3. Challenges and Future Work in Knowledge-Enhanced 3D Model

While the proposed knowledge-driven detailing framework offers a promising route toward semantically rich BIM models of existing buildings, several open challenges remain.

Scalability of Ontology Construction. The most immediate challenge concerns the scalability of building the underlying knowledge base. In the present work, the construction-rules ontology is derived from a single, well-documented regulatory source—the 1895 Munich *Bauordnung* [75]—and the architectural-elements ontology is grounded in established reference works on historical construction [70,71]. While this provides a coherent foundation for the case study area, generalizing the approach to other buildings, regions, or construction periods requires the systematic acquisition and formalization of substantial volumes of heterogeneous historical material. Building regulations, technical handbooks, and pattern books vary considerably across jurisdictions and time periods, are often only available as scanned analog documents, and are typically written in period-specific terminology that is not directly machine-readable. Manually authoring an ontology for each new context is therefore prohibitively expensive and represents a fundamental bottleneck for any knowledge-driven Scan-to-BIM approach. Recent advances in document understanding and large language model–assisted ontology engineering [78,79] may help mitigate this bottleneck, but their reliability for safety-critical structural reasoning remains to be established.

Outdated and Inconsistent Documentation. A second, closely related challenge is the quality of the historical sources themselves. Construction guidelines and pattern books reflect the building practice of their time, but real-world buildings rarely conform to a single regulatory regime throughout their lifecycle. The case study area, for instance, has undergone substantial repair, partial reconstruction, and renovation following both wartime damage and later modernization phases, each guided by different design philosophies and construction standards. The 1895 *Bauordnung* therefore describes only one of several construction regimes coexisting in the same physical building. Reasoning over such a building requires a knowledge base to represent multiple, sometimes contradictory, historical priors, and the reasoning engine must be capable of selecting or weighting the most plausible rule set per region of the scene. This temporal heterogeneity is not adequately addressed by current scene-graph generation methods, which typically assume a single, internally consistent domain prior.

Long-Tail Distribution of Architectural Elements. A third challenge stems from the long-tail distribution of architectural elements in historically significant buildings. Common components such as walls, floors, and standard windows are abundantly represented in both training data and reference documentation, while rare but architecturally important elements, such as ornate vaults, period-specific staircases, atypical column orders, or unique decorative features, appear only sporadically [80]. This imbalance affects both the upstream data-driven panoptic segmentation and the knowledge-driven reasoning stage: the ontology may underspecify rare elements simply because they are infrequently documented in the consulted sources, and the reasoner may fall back to majority-class priors for ambiguous cases. Addressing this issue will likely require explicit modeling of class imbalance in the inference step, as well as targeted enrichment of the knowledge base by documenting rare architectural features.

Computational Requirements and Deployment. The practical applicability of the proposed framework depends on three deployment factors. First, the upstream stages—LiDAR acquisition, panoptic segmentation, and NURBS surface fitting—dominate the runtime and memory footprint, since they operate on dense point clouds at the scale of hundreds of millions of points per building; GPU acceleration is required for the deep components, while a single workstation suffices for the surface-fitting stage. Second, the knowledge-driven reasoning stage is comparatively lightweight: the GCN operates on the panoptic scene graph (on the order of 10^2–10^3 nodes per building), and the symbolic rule engine executes Shapes Constraint Language (SHACL) and SPARQL Protocol and RDF Query Language (SPARQL) rules over the same graph, so the entire reasoning pipeline runs in seconds to minutes on a standard Central Processing Unit (CPU) and adds negligible overhead relative to the geometric stages. Third, the framework is designed for offline as-is documentation rather than real-time deployment, which relaxes latency constraints but introduces scalability challenges along a different axis: scaling to building portfolios requires not raw compute but the scalable ontology construction discussed above, since each new construction context demands a separately curated knowledge base. The pipeline is modular by design—acquisition, segmentation, reconstruction, and reasoning communicate through well-defined intermediate representations (point cloud with σ, panoptic surfaces, scene graph, IFC)—so individual stages can be replaced, parallelized across buildings, or deployed on heterogeneous hardware without reworking the overall workflow.

Validation and Ground Truth. A further limitation concerns validation. The targeted properties for inference, such as internal substructure or material composition, are by definition not visible in the scan data. Validating them, therefore, requires either a careful comparison against the as-is documentation of the real building, where such records exist, or direct consultation with domain experts familiar with the structure and its construction history. Neither option is destructive, but both are extremely time-consuming and rely heavily on manual effort. The proposed human-in-the-loop feedback mechanism can mitigate this issue to some extent, since domain experts can flag implausible inferences based on their experience. However, expert judgment alone is not a substitute for systematic validation. How to validate knowledge-driven BIM reconstructions in a principled way, perhaps through sparse non-destructive probes or confidence-calibrated outputs, is still an open question.

5. Conclusions

The proposed learning-based framework for MLS uncertainty evaluation addresses a critical research gap in PCQC. The completed milestones establish a progressive, cohesive trajectory bridging traditional error quantification with data-driven prediction. By delivering an automated trajectory error mitigation framework, engineering a specialized indoor

MLS benchmark dataset for the TUM2TWIN initiative, and introducing the uncertainty-aware L2M-Reg registration pipeline, the subproject demonstrates the value of embedding quality control directly into downstream engineering applications. Furthermore, the initial machine learning implementations validate the core premise that point cloud spatial errors are learnable and predictable directly from geometric features, eliminating the strict reliance on physical ground truth data.

Moving forward, the next research phases will focus on deep neural network architectures to advance this learning-based paradigm. Optimal methods to integrate physical and geometric prior knowledge as multi-modal input to resolve network convergence challenges caused by complex error propagation will be explored. Ultimately, by balancing predictive accuracy with cross-metric generalizability, this research strengthens the quality assurance ecosystem for urban digital twins and provides a robust foundation for high-precision engineering applications.

The proposed reconstruction and panoptic segmentation method represents a fundamental shift in how the AEC industry processes and interacts with 3D spatial data. By unifying open-vocabulary panoptic segmentation and geometric reconstruction within a single, multi-modal LLM pipeline, the limitations of fragmented, closed-set perception models can be addressed. The ability to seamlessly translate unstructured, noisy point clouds into semantically rich 3D scenes, while intelligently resolving occlusions through retrieval-augmented scene generation, bridges the gap between raw spatial capture and semantically enriched BIM models. This pipeline will be further refined to address scalability and physical constraints. This proposed solution will ultimately transform 3D scenes from static geometric records into dynamic, interactive design environments, empowering architects to analyze, query, and modify complex built spaces through intuitive natural language.

This subproject contributes the knowledge-driven detailing component of the NERF2BIM pipeline, transforming the uncertainty-aware, panoptically segmented representation of the scanned building into a semantically enriched 3D scene graph. The proposed approach formalizes historical construction regulations and architectural design principles—derived in the present work from the 1895 Munich *Bauordnung* and period-specific reference literature—into two interconnected ontologies, which together form a domain-specific knowledge base. A GCN-based graph reasoning engine, currently under development, will operate on this knowledge base to inject inferred structural, dimensional, and material information into the scene graph, thereby bridging the gap between purely geometric reconstruction and a semantically complete BIM model. A human-in-the-loop feedback mechanism further enables iterative refinement of both the knowledge base and the reasoning engine over time. While important challenges remain—particularly regarding the scalability of ontology construction, the heterogeneity of historical documentation, and the long-tail distribution of architectural elements in heritage buildings—the proposed pipeline offers a principled route toward automated, knowledge-rich BIM models of existing buildings, with direct applications to renovation planning, structural assessment, and historical building documentation.

Author Contributions: Conceptualization, I.B., O.F.O., Z.E. and Z.X.; methodology, O.F.O., Z.E. and Z.X.; software, O.F.O., Z.E. and Z.X.; validation, O.F.O., Z.E. and Z.X.; formal analysis, O.F.O., Z.E. and Z.X.; investigation, O.F.O., Z.E. and Z.X.; resources, O.F.O., Z.E. and Z.X.; data curation, O.F.O., Z.E. and Z.X.; writing—original draft preparation, I.B., O.F.O., Z.E. and Z.X.; writing—review and editing, I.B., O.F.O., Z.E., Z.X., M.N., C.H. and F.P.; visualization, O.F.O., Z.E. and Z.X.; supervision, M.N., C.H. and F.P.; project administration, I.B., M.N., C.H. and F.P.; funding acquisition, M.N., C.H. and F.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The research presented is supported by the TUM Georg Nemetschek Institute Artificial Intelligence for the Built World.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
BIM	Building Information Modeling
AEC	Architecture, Engineering, and Construction
LiDAR	Light Detection and Ranging
IFC	Industry Foundation Classes
RDF	Resource Description Framework
DNN	Deep Neural Network
DL	Deep Learning
NURBS	Non-Uniform Rational B-Splines
LLM	Large Language Model
C2C	Cloud-to-Cloud
C2M	Cloud-to-Mesh
M3C2	Multiscale Model-to-Model Cloud Comparison
DoF	Degrees of Freedom
MLS	Mobile Laser Scanning
TLS	Terrestrial Laser Scanning
3DGS	3D Gaussian Splatting
MLP	Multi-Layer Perceptron
NeRF	Neural Radiance Fields
SAM	Segment Anything Model
SAGA	Segment Any 3D Gaussians
PVLFF	Panoptic Vision-Language Feature Fields
OWL	Web Ontology Language
LBD	Linked Building Data
BOT	Building Topology Ontology
GRC	Geometry Relation Checking
PCQC	Point Cloud Quality Control
GIS	Geographic Information System
GT	Ground Truth
LoD	Level of Detail
HVAC	Heating, Ventilation, and Air Conditioning
VLM	Vision-Language Model
PQ	Panoptic Quality
SQ	Segmentation Quality
RQ	Recognition Quality
IoU	Intersection over Union
mIoU	mean Intersection over Union
AP	Average Precision
CD	Chamfer Distance
FPS	Farthest Point Sampling
MPE	Mixed-parameterized Positional Embeddings
GPU	Graphics Processing Unit
VRAM	Video Random-Access Memory
LoRA	Low-Rank Adaptation

MEP	Mechanical, Electrical, and Plumbing
GCN	Graph Convolutional Network
MoE	Mixture-of-Experts
SHACL	Shapes Constraint Language
SPARQL	SPARQL Protocol and RDF Query Language
CPU	Central Processing Unit
RGB	Red Green Blue
RANSAC	Random Sample Consensus
DBSCAN	Density-Based Spatial Clustering of Applications with Noise

References

1. Liu, Z.; Lu, Y.; Peh, L.C. A review and scientometric analysis of global building information modeling (BIM) research in the architecture, engineering and construction (AEC) industry. *Buildings* **2019**, *9*, 210. [CrossRef]
2. Azhar, S. Building information modeling (BIM): Trends, benefits, risks, and challenges for the AEC industry. *Leadersh. Manag. Eng.* **2011**, *11*, 241–252. [CrossRef]
3. Rashdi, R.; Martínez-Sánchez, J.; Arias, P.; Qiu, Z. Scanning technologies to building information modelling: A review. *Infrastructures* **2022**, *7*, 49. [CrossRef]
4. Campagnolo, D.; Camuffo, E.; Michieli, U.; Borin, P.; Milani, S.; Giordano, A. Fully automated scan-to-bim via point cloud instance segmentation. In *Proceedings of the 2023 IEEE International Conference on Image Processing (ICIP)*; IEEE: New York, NY, USA , 2023; pp. 291–295.
5. Abreu, N.; Pinto, A.; Matos, A.; Pires, M. Procedural point cloud modelling in scan-to-BIM and scan-vs-BIM applications: A review. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 260. [CrossRef]
6. Mehranfar, M.; Vega-Torres, M.A.; Braun, A.; Borrmann, A. Automated data-driven method for creating digital building models from dense point clouds and images through semantic segmentation and parametric model fitting. *Adv. Eng. Inform.* **2024**, *62*, 102643. [CrossRef]
7. You, K.; Chen, K.; Xue, F. Automated scan-to-BIM for construction digital transformation: Conceptual framework, processing methods and best-practice guidelines. *Autom. Constr.* **2026**, *182*, 106755. [CrossRef]
8. Rodriguez, E.V.; Bosché, F.; Esposito, M.B. Laser scanning for BIM. *J. Inf. Technol. Constr.* **2022**, *27*, 486–495. [CrossRef]
9. Qiu, Q.; Wang, M.; Guo, J.; Liu, Z.; Wang, Q. An adaptive down-sampling method of laser scan data for scan-to-BIM. *Autom. Constr.* **2022**, *135*, 104135. [CrossRef]
10. Perez-Perez, Y.; Golparvar-Fard, M.; El-Rayes, K. Segmentation of point clouds via joint semantic and geometric features for 3D modeling of the built environment. *Autom. Constr.* **2021**, *125*, 103584. [CrossRef]
11. Roman, O.; Bassier, M.; De Geyter, S.; De Winter, H.; Farella, E.M.; Remondino, F. BIM Module for Deep Learning-driven parametric IFC reconstruction. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2024**, *48*, 403–410. [CrossRef]
12. Sacks, R.; Ma, L.; Yosef, R.; Borrmann, A.; Daum, S.; Kattel, U. Semantic enrichment for building information modeling: Procedure for compiling inference rules and operators for complex geometry. *J. Comput. Civ. Eng.* **2017**, *31*, 04017062. [CrossRef]
13. Mehranfar, M.; Braun, A.; Borrmann, A. From dense point clouds to semantic digital models: End-to-end AI-based automation procedure for Manhattan-world structures. *Autom. Constr.* **2024**, *162*, 105392. [CrossRef]
14. Chen, T.; Wang, H.; Li, D.; Liu, J.; Liu, P.; Wu, Z.; Xu, C.; Zhang, M. An automated framework for converting point cloud data to building information modeling with segmentation and refinement. *Eng. Appl. Artif. Intell.* **2026**, *163*, 112915. [CrossRef]
15. Pătrăucean, V.; Armeni, I.; Nahangi, M.; Yeung, J.; Brilakis, I.; Haas, C. State of research in automatic as-built modelling. *Adv. Eng. Inform.* **2015**, *29*, 162–171. [CrossRef]
16. Chamseddine, M.; Kaufmann, F.; Schellen, M.; Glock, C.; Stricker, D.; Rambach, J. BIMStruct3D: A Fully Automated Hybrid Learning Scan-to-BIM Pipeline with Integrated Topology Refinement. *arXiv* **2026**, arXiv:2604.24311.
17. Werbrouck, J.; Pauwels, P.; Bonduel, M.; Beetz, J.; Bekers, W. Scan-to-graph: Semantic enrichment of existing building geometry. *Autom. Constr.* **2020**, *119*, 103286. [CrossRef]
18. Jarząbek-Rychard, M.; Maas, H.G. Modeling of 3D geometry uncertainty in Scan-to-BIM automatic indoor reconstruction. *Autom. Constr.* **2023**, *154*, 105002. [CrossRef]
19. Barber, D.; Mills, J.; Smith-Voysey, S. Geometric validation of a ground-based mobile laser scanning system. *ISPRS J. Photogramm. Remote Sens.* **2007**, *63*, 128–141.
20. Goel, S.; Lohani, B. Relative contribution and effect of various error sources on the performance of Mobile Mapping System (MMS). *J. Indian Soc. Remote Sens.* **2015**, *43*, 639–645. [CrossRef]
21. Xu, S. Research on Calibration of Mobile Mapping System for Land Vehicle and Its Accuracy Assessment. Ph.D. Thesis, Wuhan University, Wuhan, China, 2016.

22. Habib, A.; Bang, K.I.; Kersting, A.P.; Lee, D.C. Error budget of LiDAR systems and quality control of the derived data. *Photogramm. Eng. Remote Sens.* **2009**, *75*, 1093–1108. [CrossRef]
23. Stenz, U.; Hartmann, J.; Paffenholz, J.A.; Neumann, I. High-precision 3D object capturing with static and kinematic terrestrial laser scanning in industrial applications—Approaches of quality assessment. *Remote Sens.* **2020**, *12*, 290. [CrossRef]
24. Wujanz, D.; Burger, M.; Mettenleiter, M.; Neitzel, F. An intensity-based stochastic model for terrestrial laser scanners. *ISPRS J. Photogramm. Remote Sens.* **2017**, *125*, 146–155. [CrossRef]
25. Heinz, E.; Eling, C.; Wieland, M.; Klingbeil, L.; Kuhlmann, H. Development, calibration and evaluation of a portable and direct georeferenced laser scanning system for kinematic 3D mapping. *J. Appl. Geod.* **2015**, *9*, 227–243. [CrossRef]
26. Lague, D.; Brodu, N.; Leroux, J. Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). *ISPRS J. Photogramm. Remote Sens.* **2013**, *82*, 10–26. [CrossRef]
27. Stenz, U.; Hartmann, J.; Paffenholz, J.A.; Neumann, I. A framework based on reference data with superordinate accuracy for the quality analysis of terrestrial laser scanning-based multi-sensor-systems. *Sensors* **2017**, *17*, 1886. [CrossRef] [PubMed]
28. Zahs, V.; Winiwarter, L.; Anders, K.; Williams, J.G.; Rutzinger, M.; Höfle, B. Correspondence-driven plane-based M3C2 for lower uncertainty in 3D topographic change quantification. *ISPRS J. Photogramm. Remote Sens.* **2022**, *183*, 541–559. [CrossRef]
29. Xu, Z.; Fischer, S.; Holst, C. Uncertainty-aware Evaluation and Fusion of Point Clouds for Simultaneous Scanning of Two State-of-the-art Indoor MLS Systems. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2025**, *48*, 1603–1610. [CrossRef]
30. Hauser, D.; Glennie, C.; Brooks, B. Calibration and accuracy analysis of a low-cost mapping-grade mobile laser scanning system. *J. Surv. Eng.* **2016**, *142*, 04016011. [CrossRef]
31. Poreba, M.; Goulette, F. Line segment-based approach for accuracy assessment of MLS point clouds in urban areas. In Proceedings of the 8th International Symposium on Mobile Mapping Technology, Tainan, Taiwan, 1–3 May 2013.
32. Shahraji, M.H.; Larouche, C.; Cocard, M. Analysis of systematic errors of mobile LiDAR systems: A simulation approach. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *1*, 253–260. [CrossRef]
33. Shi, B.; Bai, Y.; Zhang, S.; Zhong, R.; Yang, F.; Song, S.; Li, G. Reference-plane-based approach for accuracy assessment of mobile mapping point clouds. *Measurement* **2021**, *171*, 108759. [CrossRef]
34. Xu, Z.; Hackl, M.; Holst, C. PL4U: Automated plane-based uncertainty evaluation and reduction method for indoor Mobile Laser Scanning systems. *ISPRS J. Photogramm. Remote Sens.* **2025**, *228*, 467–488. [CrossRef]
35. Hartmann, J.; Heiken, M.; Alkhatib, H.; Neumann, I. Automatic quality assessment of terrestrial laser scans. *J. Appl. Geod.* **2023**, *17*, 333–353. [CrossRef]
36. Hartmann, J.; Alkhatib, H. Uncertainty modelling of laser scanning point clouds using machine-learning methods. *Remote Sens.* **2023**, *15*, 2349. [CrossRef]
37. Hartmann, J.; Neumann, I.; Alkhatib, H. Modeling the distribution of TLS distance related uncertainties. *Measurement* **2026**, *270*, 120908. [CrossRef]
38. Hartmann, J.; Ernst, D.; Neumann, I.; Alkhatib, H. PointNet-based modeling of systematic distance deviations for improved TLS accuracy. *J. Appl. Geod.* **2024**, *18*, 613–628. [CrossRef]
39. Xu, Z.; Wysocki, O.; Holst, C. Point-level uncertainty evaluation of mobile laser scanning point clouds. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2025**, *XLVIII-1/W6-2025*, 243–249. [CrossRef]
40. Xu, Z.; Wysocki, O.; Holst, C. From propagation to prediction: Point-level uncertainty evaluation of MLS point clouds under limited ground truth. *arXiv* **2025**, arXiv:2511.03053.
41. Siddiqui, Y.; Porzi, L.; Bulò, S.R.; Müller, N.; Nießner, M.; Dai, A.; Kontschieder, P. Panoptic Lifting for 3D Scene Understanding With Neural Fields. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 9043–9052.
42. Cheng, B.; Misra, I.; Schwing, A.G.; Kirillov, A.; Girdhar, R. Masked-attention mask transformer for universal image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 1290–1299.
43. Zhai, H.; Li, H.; Li, Z.; Pan, X.; He, Y.; Zhang, G. Panogs: Gaussian-based panoptic segmentation for 3D open vocabulary scene understanding. In Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TN, USA, 11–15 June 2025; pp. 14114–14124.
44. Xie, Y.; Yu, X.; Jiang, C.; Mao, S.; Zhou, S.; Fan, R.; Xiong, R.; Wang, Y. Panopticsplatting: End-to-end panoptic gaussian splatting. In *Proceedings of the 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE: New York, NY, USA, 2025; pp. 4067–4074.
45. Cen, J.; Fang, J.; Yang, C.; Xie, L.; Zhang, X.; Shen, W.; Tian, Q. Segment any 3D gaussians. *Proc. AAAI Conf. Artif. Intell.* **2025**, *39*, 1971–1979. [CrossRef]
46. Zust, L.; Cabon, Y.; Marrie, J.; Antsfeld, L.; Chidlovskii, B.; Revaud, J.; Csurka, G. Panst3r: Multi-view consistent panoptic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Honolulu, HI, USA, 19–20 October 2025; pp. 5856–5866.

47. Cabon, Y.; Stoffl, L.; Antsfeld, L.; Csurka, G.; Chidlovskii, B.; Revaud, J.; Leroy, V. Must3r: Multi-view network for stereo 3D reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 11–15 June 2025; pp. 1050–1060.
48. Wang, S.; Leroy, V.; Cabon, Y.; Chidlovskii, B.; Revaud, J. Dust3r: Geometric 3D vision made easy. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 20697–20709.
49. Oquab, M.; Darcet, T.; Moutakanni, T.; Vo, H.; Szafraniec, M.; Khalidov, V.; Fernandez, P.; Haziza, D.; Massa, F.; El-Nouby, A.; et al. DINOv2: Learning Robust Visual Features without Supervision. *arXiv* **2024**, arXiv:2304.0719.
50. Behley, J.; Garbade, M.; Milioto, A.; Quenzel, J.; Behnke, S.; Stachniss, C.; Gall, J. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9297–9307.
51. Chen, H.; Blomqvist, K.; Milano, F.; Siegwart, R. Panoptic vision-language feature fields. *IEEE Robot. Autom. Lett.* **2024**, *9*, 2144–2151. [CrossRef]
52. Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. In *Linking the World's Information: Essays on Tim Berners-Lee's Invention of the World Wide Web*; ACM: New York, NY, USA, 2023; pp. 91–103.
53. Peng, C.; Xia, F.; Naseriparsa, M.; Osborne, F. Knowledge graphs: Opportunities and challenges. *Artif. Intell. Rev.* **2023**, *56*, 13071–13102. [CrossRef]
54. Ehrlinger, L.; Wöß, W. Towards a definition of knowledge graphs. *SEMANTiCS (Posters Demos SuCCESS)* **2016**, *48*, 2.
55. Beetz, J.; Van Leeuwen, J.; De Vries, B. IfcOWL: A case of transforming EXPRESS schemas into ontologies. *AI EDAM* **2009**, *23*, 89–101.
56. Pauwels, P.; Zhang, S.; Lee, Y.C. Semantic web technologies in AEC industry: A literature overview. *Autom. Constr.* **2017**, *73*, 145–165. [CrossRef]
57. Pauwels, P.; Costin, A.; Rasmussen, M.H. Knowledge graphs and linked data for the built environment. In *Industry 4.0 for the Built Environment: Methodologies, Technologies and Skills*; Springer: Cham, Switzerland, 2021; pp. 157–183.
58. Rasmussen, M.H.; Lefrançois, M.; Schneider, G.F.; Pauwels, P. BOT: The building topology ontology of the W3C linked building data group. *Semant. Web* **2020**, *12*, 143–161. [CrossRef]
59. Lilis, G.N.; Wang, M.; Katsigarakis, K.; Mavrokapnidis, D.; Korolija, I.; Dimitrios, R. BIM-based semantic enrichment and knowledge graph generation via geometric relation checking. *Autom. Constr.* **2025**, *173*, 106081. [CrossRef]
60. Wysocki, O.; Schwab, B.; Biswanath, M.K.; Greza, M.; Zhang, Q.; Zhu, J.; Froech, T.; Heeramaglore, M.; Hijazi, I.; Kanna, K.; et al. TUM2TWIN: Introducing the large-scale multimodal urban digital twin benchmark dataset. *ISPRS J. Photogramm. Remote Sens.* **2026**, *232*, 810–830. [CrossRef]
61. Xu, Z.; Schwab, B.; Yang, Y.; Kolbe, T.H.; Holst, C. L2M-Reg: Building-level uncertainty-aware registration of outdoor LiDAR point clouds and semantic 3D city models. *ISPRS J. Photogramm. Remote Sens.* **2026**, *233*, 467–486. [CrossRef]
62. Zhao, H.; Jiang, L.; Jia, J.; Torr, P.; Koltun, V. Point Transformer. *arXiv* **2021**, arXiv:2012.09164 .
63. Grattafiori, A.; Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; Schelten, A.; Vaughan, A.; et al. The llama 3 herd of models. *arXiv* **2024**, arXiv:2407.21783.
64. Bai, J.; Bai, S.; Chu, Y.; Cui, Z.; Dang, K.; Deng, X.; Fan, Y.; Ge, W.; Han, Y.; Huang, F.; et al. Qwen technical report. *arXiv* **2023**, arXiv:2309.16609.
65. Chen, X.; Chu, F.J.; Gleize, P.; Liang, K.J.; Sax, A.; Tang, H.; Wang, W.; Guo, M.; Hardin, T.; Li, X.; et al. Sam 3D: 3Dfy anything in images. *arXiv* **2025**, arXiv:2511.16624.
66. Dai, A.; Chang, A.X.; Savva, M.; Halber, M.; Funkhouser, T.; Nießner, M. Scannet: Richly-annotated 3D reconstructions of indoor scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5828–5839.
67. Rozenberszki, D.; Litany, O.; Dai, A. Language-grounded indoor 3D semantic segmentation in the wild. In *Computer Vision—ECCV 2022, Proceedings of the European Conference on Computer Vision* ; Springer: Cham, Switzerland, 2022; pp. 125–141.
68. Yeshwanth, C.; Liu, Y.C.; Nießner, M.; Dai, A. Scannet++: A high-fidelity dataset of 3D indoor scenes. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 12–22.
69. Chen, Y.; Luo, X. Indoor scan-to-BIM workflows: Progress, challenges, and future directions (2014–2024). *Autom. Constr.* **2026**, *181*, 106578. [CrossRef]
70. Ahnert, R.; Krause, K.H. *Typische Baukonstruktionen von 1860 bis 1960*; Verlag für Bauwesen: Berlin, Germany, 1991.
71. Cole, E. (Ed.) *The Grammar of Architecture*, 1st U.S. ed.; Bulfinch Press: Boston, MA, USA, 2002.
72. Wald, J.; Dhamo, H.; Navab, N.; Tombari, F. Learning 3D semantic scene graphs from 3D indoor reconstructions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3961–3970.
73. Strader, J.; Hughes, N.; Chen, W.; Speranzon, A.; Carlone, L. Indoor and outdoor 3D scene graph generation via language-enabled spatial ontologies. *IEEE Robot. Autom. Lett.* **2024**, *9*, 4886–4893. [CrossRef]

74. Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. Training language models to follow instructions with human feedback. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 27730–27744. [CrossRef]
75. Englert, F. *Die Münchener Bauordnung vom 29. Juli 1895: Mit Erläuterungen und einem Anhange*; Lindauer: Munich, Germany, 1895.
76. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.
77. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4–24. [CrossRef]
78. Babaei Giglou, H.; D'Souza, J.; Auer, S. LLMs4OL: Large language models for ontology learning. In *The Semantic Web—ISWC 2023, Proceedings of the International Semantic Web Conference*; Springer: Cham, Switzerland, 2023; pp. 408–427.
79. Mateiu, P.; Groza, A. Ontology engineering with large language models. In *Proceedings of the 2023 25th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*; IEEE: New York, NY, USA, 2023; pp. 226–229.
80. Matrone, F.; Grilli, E.; Martini, M.; Paolanti, M.; Pierdicca, R.; Remondino, F. Comparing machine and deep learning methods for large 3D heritage semantic segmentation. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 535. [CrossRef]

https://doi.org/10.3390/buildings16142791