Sustainable Cities and Society 149 (2026) 107753

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Smart point cloud–guided 3D gaussian splatting for urban modeling and data-driven analysis

Yingwen Yu [a], Zhuoyue Wang [a], Guanting Zhang [b], Peter van Oosterom [a], Edward Verbree [a], Steffen Nijhuis [c], Yuyang Peng [c,*]

[a] Department of Architectural Engineering and Technology, Faculty of Architecture and the Built Environment, Delft University of Technology, Julianalaan 134, Delft, The Netherlands
[b] Nanjing Tech University, College of Architecture, Nanjing 211816, China
[c] Department of Urbanism, Faculty of Architecture and the Built Environment, Delft University of Technology, Julianalaan 134, Delft, The Netherlands

ARTICLE INFO

Keywords:
3D gaussian splatting (3DGS)
Point cloud
Street view imagery (SVI)
Visual exposure
Urban indicators
CityGML

ABSTRACT

Urban analysis often relies on separate geospatial mapping, street-level image analysis, point-cloud analysis and 3D urban modeling workflows, which can lead to semantic and spatial inconsistencies when structural, visual and volumetric results are interpreted together. This paper develops an SPC-guided Smart 3D Gaussian Splatting (S3DGS) analytical framework to address this fragmentation. Using the Aula–library block in Delft as a case study, we construct a CityGML-inspired Smart Point Cloud from laser-scanning data, generate street-level and aerial 3DGS components from multi-view imagery, align and fuse them into a common coordinate frame, and propagate SPC-derived semantics to Gaussian primitives. The resulting S3DGS base supports projection-based semantic composition, observer-centered visual exposure, volumetric spatial-presence aggregation and layered scene typing within the same reference. Cross-layer interpretation links volumetric and visual-exposure types in the same spatial units, revealing conditions such as physically present but visually recessive building mass. This layered reading allows surface composition, visual experience and three-dimensional spatial presence to be traced within the same urban unit. Evaluation results indicate a stable workflow, with an alignment RMSE of 0.0439 m and a kNN label consistency of 92.24%. By turning Gaussian scenes into semantically traceable and spatially co-referenced analytical bases, S3DGS offers a practical pathway from isolated urban measurements toward integrated, viewpoint-aware and three-dimensional interpretation of the built environment.

1. Introduction

The assessment and management of the urban built environment increasingly depend on digital representations that can support spatial computation, urban monitoring, and evidence-based decision-making (Engin et al., 2020; He & Chen, 2024; Mazzetto, 2024). Many planning, governance, and environmental assessment tasks require a representation of both what the city is, in terms of geometry, spatial structure, and semantic composition, and how the city is experienced from human-relevant viewpoints, including visibility, occlusion, and exposure to different urban elements (Geneletti et al., 2007; Peng et al., 2026). In practice, these varied dimensions are often addressed through separate analytical traditions. Structural-semantic representations are well suited to object-level, surface-level, and geometry-based analysis, whereas viewpoint-based approaches are more effective for measuring visual exposure and observer-centered experience (Kolbe, 2009; Liu & Nijhuis, 2020). As a result, urban indicators are frequently derived through fragmented workflows that rely on different data structures, semantic assumptions, and spatial references, reducing comparability, interpretability, and methodological consistency across analyses (Billen et al., 2015; Virtanen et al., 2021).

Recent developments in 3D Gaussian Splatting (3DGS) and Smart Point Clouds (SPCs) provide a route for connecting viewpoint-based observation with structural-semantic urban analysis (Chen et al., 2026; Poux et al., 2016; Yu et al., 2025; Yu et al., 2025). 3DGS (Fig. 1) represents a scene through optimizable Gaussian primitives that encode spatial position, spatial extent, opacity and appearance attributes (Fei et al., 2025; Mallick et al., 2024). These primitives are

* Corresponding author.
E-mail addresses: christinayu@tudelft.nl (Y. Yu), Z.Wang-111@student.tudelft.nl (Z. Wang), gtzhang@njtech.edu.cn (G. Zhang), p.j.m.vanoosterom@tudelft.nl (P. van Oosterom), e.verbree@tudelft.nl (E. Verbree), S.Nijhuis@tudelft.nl (S. Nijhuis), y.peng-1@tudelft.nl (Y. Peng).

https://doi.org/10.1016/j.scs.2026.107753
Received 17 January 2026; Received in revised form 27 April 2026; Accepted 26 July 2026
Available online 27 July 2026

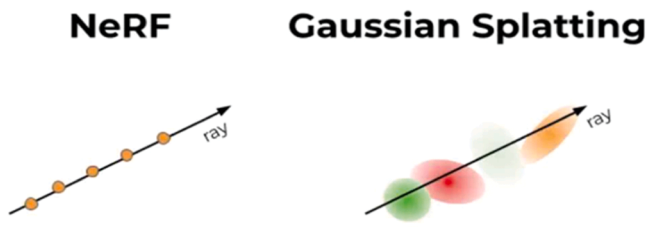

Fig. 1. Rendering principle of 3DGS in contrast to NeRF: NeRF synthesizes views by sampling an implicit neural radiance field along camera rays, whereas 3DGS represents the scene with explicit anisotropic Gaussian primitives carrying position, spatial extent, opacity and appearance attributes. During rendering, these Gaussians are projected and alpha-composited to generate view-consistent images. This explicit primitive-based representation makes 3DGS suitable for subsequent semantic enrichment, because SPC-derived labels can be attached to Gaussian primitives to construct the S3DGS analytical base (Kerbl et al., 2023).

optimized from calibrated multi-view images and can be rendered from arbitrary viewpoints, making 3DGS well suited to visual analyses involving visibility, occlusion, facade exposure and streetscape observation (Kerbl et al., 2023; M. Tancik et al., 2022). Recent large-scene Gaussian-splatting studies further show the potential of 3DGS-like representations for scalable outdoor and urban scene rendering, while also highlighting the need for partitioning, level-of-detail strategies and efficient scene management (Lin et al., 2024; Liu et al., 2024; Matthew Tancik et al., 2022). However, standard 3DGS does not normally provide explicit semantic structure: Gaussian primitives may reproduce the visual appearance of an urban scene, but they are not inherently linked to object classes, surface types or interpretable urban elements. This limitation is also reflected in recent semantic 3DGS work, which explicitly extends Gaussian Splatting with object-level, semantic or language-aware scene understanding (Qin et al., 2024; Ye et al., 2025).

SPCs provide a complementary structural-semantic layer. They organize point-cloud data into semantically meaningful computational units, such as patches, surfaces or object-related components, which can be linked to attributes such as semantic class, structural role and material information (Beil et al., 2021; Gröger & Plümer, 2012; Peng et al., 2022; Poux, 2019; Poux et al., 2016). This makes SPCs suitable for interpretable spatial analysis and aggregation, but they do not provide the view-consistent appearance and flexible virtual observation supported by 3DGS. The integration of SPCs and 3DGS therefore enables the construction of Smart 3DGS (S3DGS): a semantically enriched Gaussian scene in which Gaussian primitives inherit SPC-derived semantic information. In this study, S3DGS functions as the shared semantic-visual analytical base for deriving and combining urban indicators within a consistent spatial and semantic logic.

Therefore, this paper aims to develop an integrated SPC–3DGS analytical framework for urban analysis. The framework is designed to construct a shared semantic, spatial and visual base that preserves high-fidelity, view-consistent scene visualization while supporting the derivation, comparison and synthesis of different families of urban indicators. Rather than treating 3DGS as a visualization technique alone, the study positions the coupling of SPCs and 3DGS as a basis for urban analytical modeling, linking structural-semantic information with viewpoint-based observation. To this end, the study addresses three research questions:

RQ1: How can the complementary strengths of semantically structured point clouds and 3DGS be integrated into a coherent semantic-visual analytical framework for urban scenes?
RQ2: Which families of urban analytical tasks can be operationalized from such a shared S3DGS base, and how can they be organized as a coordinated framework rather than as separate workflows?
RQ3: What analytical value, reliability, and practical limitations are revealed when the framework is implemented in a real-world case study?

The framework is demonstrated through a real-world urban block case study in Delft, the Netherlands. By coupling the semantic and structural interpretability of SPCs with the view-consistent rendering capacity of 3DGS, and by propagating semantic information to Gaussian primitives, the study constructs a S3DGS analytical base for integrated urban assessment. This paper contributes an S3DGS-based analytical framework that addresses a persistent fragmentation in urban analysis: the separation between structural-semantic representations of urban form and viewpoint-dependent representations of urban experience. By using S3DGS as a shared semantic-visual analytical base, the framework enables different families of urban indicators to be derived, compared and synthesized within a consistent spatial and semantic reference.

2. Urban analytical traditions and the need for an S3DGS analytical base

Urban assessment has developed through several analytical traditions that rely on different representations of the built environment. Two directions are particularly relevant to this study: One uses spatial representations, and the other uses viewpoint-based and visual-perceptual representations. This section reviews these directions in terms of the urban analyses they support and the requirements they create for integrated interpretation.

2.1. Spatial-representation-based methods for geometric and semantic urban assessment

Spatial-representation-based methods provide an important basis for measuring the physical structure of the built environment. They include digital elevation and surface models, mesh models, point clouds, point-cloud-to-mesh reconstructions, 3D city models, City Information Models (CIM), urban digital twins and semantically enriched point-based datasets derived from surveying and mapping technologies (Cárdenas-León et al., 2024; Peng & Nijhuis, 2021; Xia et al., 2022). These representations differ in their balance between geometric fidelity and semantic richness. Point clouds, meshes and photogrammetric reconstructions may capture detailed geometry with limited semantic organization, while City Geography Markup Language (CityGML), CIM and digital twin objects may provide explicit semantic classes and object hierarchies with more abstracted geometry, texture or fine-scale appearance (Ledoux et al., 2019; Lei, Janssen, et al., 2023; Weil et al., 2023).

These spatial representations support a wide range of structural urban analyses. They are used to quantify urban form, morphology, building massing, height, surface configuration, vegetation visibility, and other geometry-based indicators (Labetski et al., 2023; Qi et al., 2024). When semantic attributes are available, they also support object-level or patch-level aggregation across parcels, blocks, infrastructure elements and functional zones (Peng, Li, et al., 2026). Recent work on urban digital twins and multi-scale digital modeling has further extended this tradition by linking GIS, BIM, semantic models, dashboards, sensors and decision-support workflows for sustainable urban design and management (Geneletti et al., 2007; Yu et al., 2026; C. Zhang et al., 2025).

Despite these advances, spatial-representation-based methods provide a less direct basis for interpreting how urban form is visually experienced from human-relevant viewpoints. Their analytical capacity depends on the balance between geometric fidelity, semantic richness and visual realism. Eye-level visibility, occlusion, facade detail, vegetation structure, street furniture, glazing, water surfaces and pedestrian perspective are often simplified, incomplete or indirectly represented in conventional spatial models (Ewing et al., 2006; Peng et al., 2026). This

becomes important when structural or semantic indicators need to be interpreted together with visual exposure, streetscape composition or human-scale environmental experience (Benedikt, 1979; Ewing et al., 2006; Lei et al., 2024).

2.2. Viewpoint-based visual methods for observer-centered urban assessment

Viewpoint-based visual methods examine the urban environment through images or rendered views acquired from human-relevant positions, routes or viewing conditions. They include street-view imagery (SVI), ground-level/ Unmanned Aerial Vehicle (UAV)-based photographs, panoramic images, oblique imagery and image-based semantic segmentation (Sánchez & Labib, 2024). These methods quantify what is visible from particular locations or paths, including streetscape composition, greenery exposure, building visibility, sky openness, facade presence and other visual indicators related to pedestrian-scale urban experience (Sánchez & Labib, 2024; Wu et al., 2023; Zhou et al., 2019).

SVI-based panoramic observation is important because they measure the city at the scale at which people encounter it (Biljecki & Ito, 2021; Ito et al., 2024). They have been widely used to estimate green view index, building exposure, sky openness, visual enclosure and related exposure-oriented indicators. These visual measurements also support perception-oriented studies on walkability, perceived safety, visual comfort, environmental quality and preference (Rundle et al., 2011). In addition, computer vision has increasingly been framed as a planning-relevant analytical tool, supporting data collection, issue identification, design evaluation, implementation monitoring and post-occupancy assessment (Marasinghe et al., 2024). Simulated views and virtual environments further extend this direction by allowing viewpoints, viewing paths, environmental conditions and design alternatives to be controlled, thereby supporting controlled assessment of urban form and visual perception (Shushan et al., 2016). More recently, neural rendering methods have added the possibility of high-fidelity visual reconstruction and flexible virtual observation from arbitrary viewpoints. Large-scene neural rendering studies have shown how urban-scale visual environments can be decomposed and rendered beyond discrete street-view samples, while 3D Gaussian Splatting further provides an explicit primitive-based representation for real-time, view-consistent rendering (Kerbl et al., 2023; M. Tancik et al., 2022).

These methods provide strong support for visual exposure and perception-oriented analysis, but their outputs often require geometric anchoring and semantic alignment when related to structural, volumetric or planning-oriented measures. Street-view and image-based workflows are constrained by camera availability, road-based sampling, seasonality, illumination, transient occlusions, coverage bias and segmentation uncertainty (Fan et al., 2025; Ito et al., 2024; Kim et al., 2021). Simulated environments provide greater control over viewpoints and scenarios, but their consistency with measured geometry and semantic datasets depends on how the scene is modeled. Neural rendering reduces some spatial sampling constraints through flexible virtual viewpoints, while standard 3DGS still provides limited semantic traceability and object-level interpretability because Gaussian primitives are optimized primarily for view-consistent rendering rather than explicit urban semantics (Kerbl et al., 2023). As a result, viewpoint-based indicators can be difficult to connect directly with the semantic units, spatial references and aggregation logic used in spatial-representation-based urban assessment (Biljecki & Ito, 2021; Lei et al., 2024).

2.3. From fragmented analytical needs to the S3DGS analytical base

The two traditions reviewed above have advanced urban assessment in complementary ways. Spatial-representation-based methods provide geometry, spatial reference and, in some cases, semantic organization. Viewpoint-based and visual-perceptual methods capture visibility, exposure and human-scale experience. The remaining challenge concerns analytical co-reference: structural, visual and volumetric indicators are often produced from different datasets, semantic classes, spatial units and sampling assumptions. This can make cross-indicator comparison and integrated interpretation less transparent (Billen et al., 2015; Jeddoub et al., 2023; Weil et al., 2023).

This situation creates a need for an analysis-ready base that supports three linked conditions. First, it should preserve semantic consistency across different indicator families. Second, it should connect flexible viewpoint-based observation with three-dimensional urban structure. Third, it should provide a shared spatial reference through which projection-based, observer-centered and volumetric indicators can be derived and synthesized within the same scene.

The proposed SPC–3DGS integration addresses these conditions by constructing a S3DGS analytical base. In this integration, the SPC provides the structural-semantic layer by organizing point-cloud data into meaningful patches and attaching CityGML-inspired semantic attributes (Beil et al., 2021; Poux, 2019; Poux et al., 2016). The 3DGS components provide the view-consistent appearance layer, enabling high-fidelity rendering and flexible virtual observation (Kerbl et al., 2023; Tancik et al., 2022). Through geometric alignment, model fusion and semantic propagation, SPC-derived labels are transferred to Gaussian primitives, so that the resulting S3DGS scene becomes both visually coherent and semantically structured.

The remainder of the paper develops this framework in two steps with a Dutch case. Section 3 describes how the S3DGS analytical base is constructed, including SPC construction, generation of street-level and aerial 3DGS components, geometric alignment, model fusion and semantic propagation. Section 4 explains how this base is operationalized for urban analysis through three indicator-generation modules and one synthesis module: projection-based semantic composition, observer-centered visual exposure, volumetric spatial-presence aggregation and layered scene typing.

3. Constructing the S3DGS analytical base

To demonstrate the proposed framework, we use the Aula–library block in Delft, the Netherlands, as a real-world urban case study (Fig. 3). The block contains the Aula, a post-war modernist building completed in 1966 and listed as a national monument, together with the main library, plazas, tree-lined streets, open spaces and adjacent campus buildings. This setting provides a suitable test case because it combines heritage architecture, pedestrian-scale public spaces, vegetation, roof structures and open courtyards within a compact urban block.

This section describes the construction of the S3DGS analytical base used in the subsequent analysis (Fig. 2). The workflow consists of four stages: (i) data collection and preprocessing; (ii) construction of a SPC with structured urban semantics; (iii) generation of street-level and aerial 3DGS components; and (iv) geometric alignment, model fusion and semantic propagation from the SPC to the fused 3DGS scene. The main software environments, algorithms and key parameter settings are summarized in Table 1.

3.1. Data collection and preprocessing

The case study combines geometry and imagery from multiple sources. UAV-based photography is used to capture roofs, upper building volumes and courtyard areas that are difficult to observe from street level. At pedestrian level, the XGRIDS/K1 system records street-level imagery, trajectory information and a spatially extensive point cloud along accessible routes. A GeoSLAM Horizon R1 scanner provides dense terrestrial laser scanning data for facades, ground surfaces and vegetation, while Actueel Hoogtebestand Nederland (AHN) airborne laser scanning data are used as a georeferenced spatial reference for cross-sensor registration.

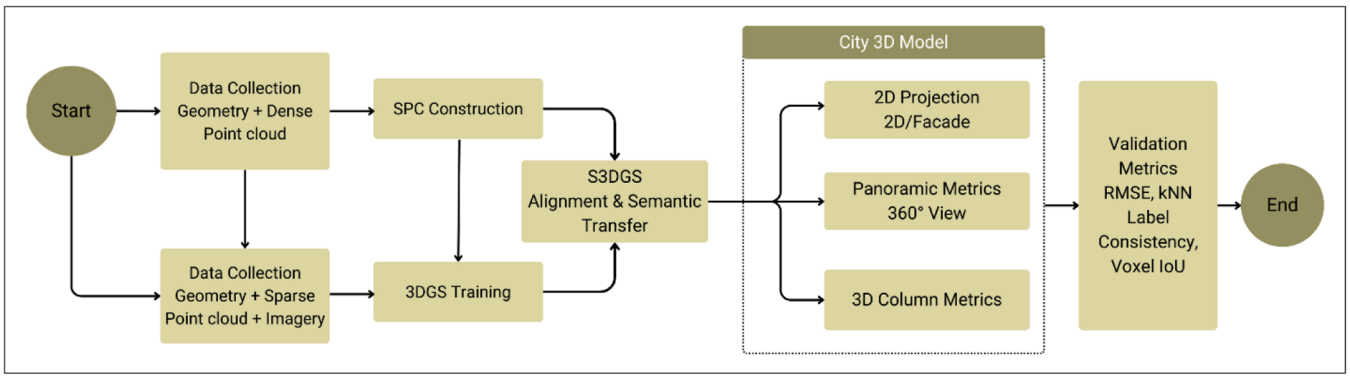

Fig. 2. End-to-End S3DGS Workflow.

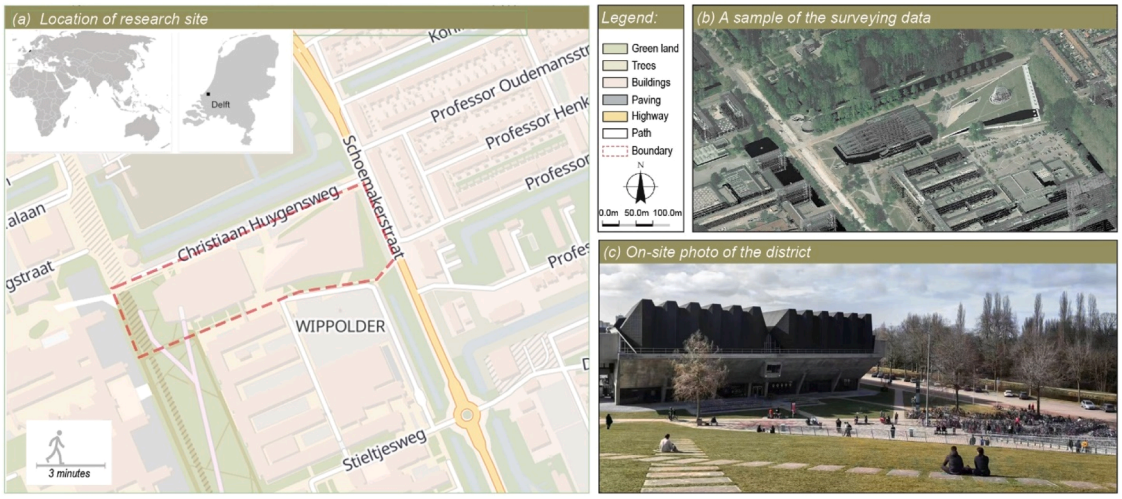

Fig. 3. Case site and origin AHN point cloud data.

The point-cloud datasets are cleaned, aligned and merged before SPC construction. The GeoSLAM data provide the primary dense geometric reference, while the XGRIDS/K1 data extend coverage into adjacent streets and open spaces. Corresponding points are manually selected to estimate an initial transformation from the mobile/terrestrial laser scanning datasets to the AHN reference, and the registration quality is checked using Root-mean-square error (RMSE). Ground points are separated in CloudCompare using the Cloth Simulation Filter (CSF), and major planar surfaces are extracted using CloudCompare's RANSAC-based primitive detection. The imagery from the UAV and XGRIDS/K1 system is retained for the separate aerial and street-level 3DGS reconstruction branches.

3.2. Construction of Smart Point Cloud (SPC)

The registered point cloud is converted into a SPC through segmentation and semantic enrichment. Each point stores 3D coordinates, RGB color, estimated surface normals and a semantic patch identifier. Initial segmentation is generated using geometric and radiometric cues, including spatial proximity, planar structure and color similarity. This first-pass segmentation is then refined through rule-based manual inspection to correct misclassified regions and adjust patch boundaries where necessary. The refinement step is limited to correcting the automatic grouping and defining meaningful urban patches, rather than manually redrawing the scene from scratch.

Each patch is assigned a unique patch_id, which links point-level geometry to an external semantic schema (Poux, 2019). Semantic information is stored in a structured JSON file indexed by patch_id. At the patch level, the schema records classes and roles such as building facade, roof, pavement, tree canopy, shrub, lawn, water surface and street furniture, together with structural attributes such as ground, vertical surface or canopy. The schema follows CityGML as a conceptual reference (Gröger & Plümer, 2012), allowing the SPC to serve as a semantically structured urban model and as the source of labels for the subsequent S3DGS construction. The refinement followed fixed semantic classes and patch-level rules, which improves procedural consistency, although inter-operator variability remains a limitation for larger-scale deployment.

3.3. Generation of street-level and aerial 3DGS components

Two 3DGS components are generated separately from the street-level and aerial data streams. The street-level branch is produced from the XGRIDS/K1 data inLCC Studio v1.9.0. This scanner-native reconstruction provides calibrated imagery, trajectory information and spatial reference for pedestrian-scale elements such as facades, pavements, plazas, vegetation and street furniture (Sevenich et al., 2025).

The aerial branch is produced from UAV imagery. The UAV images are first processed in COLMAP v3.12.6, where structure-from-motion (SfM) estimates camera intrinsics and extrinsics and multi-view stereo (MVS) generates an auxiliary dense reconstruction. This branch provides image-consistent camera geometry and spatial initialization for roofs, upper building volumes, courtyard areas and surfaces that are weakly observed from street level (Berra & Peppa, 2020; Meinen & Robinson, 2020). The resulting camera poses and reconstruction are then used in Postshot v1.0 to train the aerial 3DGS component.

Table 1
Summary of software, key operations and parameters used in the SPC-S3DGS workflow.

Workflow stage	Software / environment	Key operation and settings	Output
Point-cloud preprocessing and SPC construction	CloudCompare v2.13.2; Python	CSF ground filtering: cloth resolution = 2, max iterations = 500, classification threshold = 0.5; RANSAC plane detection; semi-automatic segmentation with rule-based manual refinement; JSON semantic schema linked by patch_id.	Cleaned and semantically structured SPC.
Street-level 3DGS component	LCC Studio v1.9.0	XGRIDS/K1 scanner-native reconstruction using calibrated imagery and trajectory information.	Street-level 3DGS component.
UAV 3DGS component	COLMAP v3.12.6; Postshot v1.0	COLMAP SfM-MVS for UAV camera poses and auxiliary dense reconstruction; Postshot training for aerial / roof-level 3DGS.	UAV-derived 3DGS component.
Alignment, fusion and semantic propagation	Custom Python; SciPy cKDTree	Control-point similarity transformation with RMSE validation; model-level fusion of transformed Gaussian primitives; k-nearest-neighbor (kNN) majority-vote propagation, k = 3; voxel IoU at 0.25 m.	Unified S3DGS with semantic labels.
Implementation profile	Dell Precision mobile workstation	Intel Core Ultra 7 155H CPU; 3DGS processing ≈ 2 h; semantic propagation and indicator derivation ≈ 1-2 h.	Indicative block-scale runtime.

This two-branch strategy reflects the different acquisition geometries of the XGRIDS/K1 scanner and the UAV image block. The SPC provides an accurate semantic and geometric reference, but it does not contain the image-consistent camera pose relationships required for optimizing Gaussian primitives from UAV imagery. SfM–MVS is therefore used for the UAV branch, while the street-level branch relies on the scanner-native reconstruction provided by LCC Studio. Both branches produce anisotropic Gaussian primitives with spatial, opacity and appearance attributes, which are exported for alignment, fusion and semantic propagation.

3.4. Geometric alignment, model fusion and semantic propagation

To construct S3DGS, the independently generated street-level and aerial 3DGS components are first transformed into a common world coordinate frame and fused into a unified non-semantic 3DGS scene. The SPC is anchored to the georeferenced point-cloud reference using AHN and ground-based laser scanning data. For each 3DGS component, corresponding control points are selected between the component model and the SPC/AHN reference, and a similarity transformation including scale, rotation and translation is estimated and refined through least-squares adjustment. For the final alignment and validation of the fused scene, six manually selected corresponding control points are used between the fused 3DGS scene and the SPC/AHN reference (Fig. 4). The key settings for alignment, fusion and semantic propagation are summarized in Table 1.

After transformation, the XGRIDS/K1-derived and UAV-derived Gaussian primitives are represented in the same coordinate system and merged into a single 3DGS scene. Overlapping regions are inspected to identify duplicated or locally misaligned areas while preserving the complementary coverage of the two components: the street-level branch contributes pedestrian-scale surfaces, whereas the UAV branch complements roofs, upper facades and courtyard spaces.

Semantic propagation is then performed from the SPC to the fused Gaussian scene. For each Gaussian center, neighboring SPC points are queried using a KD-tree, and the semantic label is assigned through k-nearest-neighbor majority voting with k = 3 (Cover & Hart, 1967). Through patch_id, each Gaussian inherits the semantic attributes stored in the SPC JSON schema, such as building facade, roof, pavement, tree canopy, shrub, lawn, water surface or street furniture. Gaussians without sufficient neighboring SPC support are marked as unassigned. The resulting S3DGS representation retains the photometric properties of the fused 3DGS scene while adding SPC-derived semantic structure.

The reliability of this construction is assessed using three metrics. Alignment RMSE is computed from the residuals of the control-point transformation. k-nearest-neighbor (kNN) label consistency compares each propagated Gaussian label with the majority label of its three nearest SPC neighbors. Voxel-based intersection-over-union (IoU) is computed by separately voxelizing SPC and 3DGS subsets at a voxel size of 0.25 m and comparing occupied voxel sets for corresponding semantic patches (Mescheder et al., 2019; Peng et al., 2022). Together, these metrics provide quality checks for geometric alignment, semantic transfer and spatial correspondence between the SPC and the fused S3DGS scene.

4. Using the S3DGS analytical base for multi-dimensional urban analysis

The previous section established the S3DGS analytical base. This section shows how it supports multi-dimensional urban analysis within a shared semantic, spatial and visual reference (Table 2). The framework includes three indicator-generation modules and one synthesis module. The indicator-generation modules correspond to projection-based semantic composition, observer-centered visual exposure and volumetric spatial-presence aggregation. These were selected because they represent three important and widely used analytical modes in urban assessment, namely surface composition, human-view exposure and three-dimensional spatial presence (Bonczak & Kontokosta, 2019; Labetski et al., 2023; G. Zhang et al., 2025). Together, they show how these usually separated perspectives can be analyzed and interpreted within the same S3DGS reference.

4.1. Projection-based semantic composition analysis

The first indicator-generation module demonstrates the projection-based analytical capacity of S3DGS. It derives semantic composition indicators through orthographic projections of the labeled S3DGS scene. This module represents a common family of urban analyses that quantify the semantic composition of surfaces, envelopes, and near-ground spatial conditions (Li et al., 2019; Murtiyoso et al., 2021).

In this study, facade-oriented semantic projections and height-specific horizontal semantic slices are selected as representative implementations. For selected facades, orthographic views are rendered from S3DGS, and each pixel is assigned the semantic label of the dominant underlying Gaussian primitive. In parallel, horizontal semantic slices are generated at selected heights above ground. In the present demonstration, 1 m and 3 m slices are used to capture different layers of the pedestrian environment. The 1 m slice is used to identify low-level vertical elements that may constrain, guide or partially obstruct

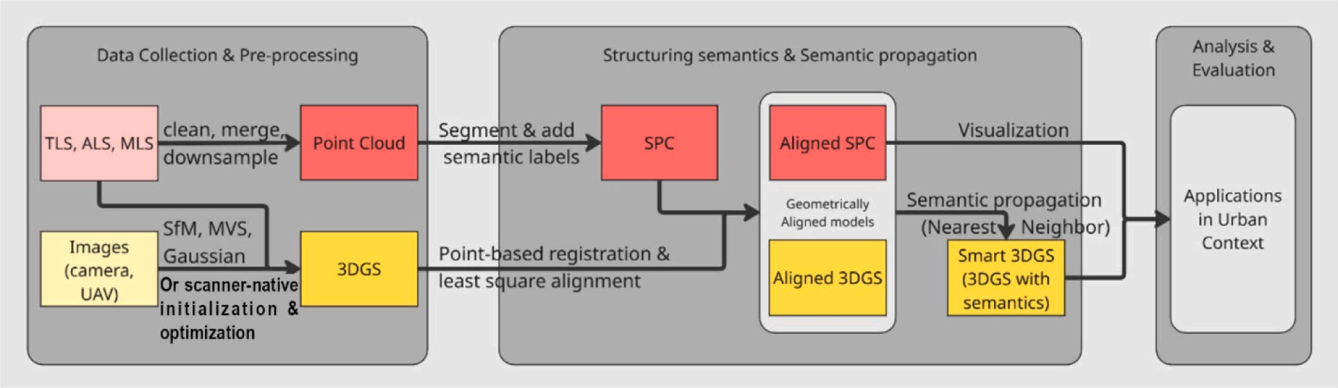

Fig. 4. Core workflow from SPC to S3DGS.

Table 2
Representative analytical modules implemented from the S3DGS analytical base.

Framework layer	Analytical module	Representative implementation in this study	Main output
Indicator generation	Projection-based semantic composition	Facade and ground-plane semantic projections	Class shares of facades, ground surfaces, vegetation, pavement and water; projection-based composition layer
Indicator generation	Observer-centered visual exposure	360° panoramic renderings at pedestrian eye-level and visual-exposure clustering	Green view index, building exposure, sky openness; visual-exposure types
Indicator generation	Volumetric spatial-presence aggregation	Column-based aggregation of labeled Gaussians and volumetric clustering	Green, building and hard-surface spatial-presence ratios; volumetric spatial-presence types
Synthesis and interpretation	Layered scene typing	Grid-cell-level stacking of projection-based composition, visual-exposure and volumetric spatial-presence layers within the S3DGS reference	Co-referenced layered scene profiles and spatial patterns

pedestrian movement, such as tree trunks, structural columns, walls and street furniture, thereby providing a human-scale indication of near-ground accessibility and open movement space (Shao et al., 2022). The 3 m slice is interpreted together with the 1 m layer to reveal overhead enclosure created by roof overhangs, building canopies, or tree crowns (Li et al., 2019; Pretzsch et al., 2015). This makes it possible to identify sheltered grey spaces and tree-shaded spaces within the same semantic scene.

From these projections and slices, planar class shares are calculated for green infrastructure, including tree canopy, shrubs, and lawns, as well as buildings, pavement, water, and other relevant urban elements. These outputs form a projection-based composition layer. Depending on the analytical purpose, each spatial unit can be interpreted through continuous class shares or summarized into a simple composition type, such as hardscape-dominated surface, green-surface-dominated area, facade-adjacent surface, or mixed surface composition.

4.2. Observer-centered visual exposure analysis

The second indicator-generation module demonstrates the observer-centered analytical capacity of S3DGS. It derives visual exposure indicators from pedestrian-level virtual viewpoints and is related to street-view-based measurements such as green view index, building exposure, facade visibility and sky openness (Peng et al., 2026; Peng et al., 2026; Sánchez & Labib, 2024).

In this study, virtual viewpoints are placed on a 5 × 5 m grid within ground-level accessible areas of the block (Labib et al., 2021), including pedestrian routes, plazas, courtyards and internal campus paths. At each viewpoint, a full 360° panorama is rendered from S3DGS at an eye height of 1.6 m, with each pixel carrying a semantic label inherited from the SPC. Class shares are then computed for vegetation, buildings, sky, pavement and other categories. These shares are used to derive indicators such as green view index, building visual exposure and sky openness.

These outputs form a visual-exposure layer. In the present demonstration, vegetation share, building share and sky share are used as a reduced visual-exposure signature because they represent three common dimensions of pedestrian-scale experience: greenery, built enclosure and openness. To summarize recurrent visual conditions, the feature vectors are standardized and k-means clustering is applied with $k = 5$, with the number of clusters selected based on elbow analysis. This layer provides the human-view component for the later layered scene typing.

4.3. Volumetric spatial-presence aggregation and layer typing

The third indicator-generation module demonstrates the capacity of S3DGS to support three-dimensional spatial aggregation. It aggregates semantic Gaussians within vertical spatial units to describe how urban elements are distributed above each grid cell. This type of analysis is related to voxel-based point-cloud analysis and 3D city-model metrics, but in the S3DGS framework it is derived from the same semantic Gaussian scene that also supports the projection and panoramic layers (Labetski et al., 2023; Ren et al., 2020).

In this study, the block is partitioned into vertical columns aligned with the same 5 × 5 m spatial grid used for the observer-centered visual-exposure analysis, so that the resulting volumetric layer can be directly overlaid with the visual layer in the subsequent layered scene-typing analysis. Within each column, the semantic point cloud is voxelized and aggregated according to its labels, including vegetation, buildings, ground, road surface, street furniture and unassigned elements. Specifically, the 3D space of each column is discretized into regular voxels with a voxel size of 0.25 m. Each occupied voxel is assigned the dominant semantic label of the SPC points falling within it, and class proportions are then calculated as the share of occupied voxels belonging to each semantic category. The resulting proportions describe how different semantic elements occupy the vertical space of each grid cell (Peng et al., 2024). Since the aggregation is based on voxelized

point-cloud occupancy, it provides a more spatially explicit volumetric estimate than simple point-count statistics. However, the values should still be interpreted as voxel-based volumetric composition indicators rather than exact solid-volume measurements, because they depend on voxel resolution, point-cloud completeness, segmentation quality and scene coverage.

For layer typing, each grid cell is represented by a reduced volumetric signature composed of four semantic proportions: ground share, vegetation share, building share and road-surface share. Before clustering, the features are standardized. The number of clusters is selected using elbow analysis, and k-means clustering is then applied with $k = 5$ to identify recurrent volumetric spatial-presence types across the study area. Based on the mean semantic composition of each cluster, the resulting types are interpreted as building-dominated, low-coverage / unassigned-mixed, ground-dominated, vegetation-dominated and road-surface-dominated columns. This volumetric layer can then be stacked with the visual-exposure layer in the scene-typing analysis.

4.4. Layered scene typing from co-referenced S3DGS layers

The synthesis module organizes the outputs of the preceding analytical modules within the same S3DGS spatial reference. The observer-centered module provides a visual-exposure layer, and the volumetric module provides a spatial-presence layer. Since these layers are linked to the same semantic reference and spatial grid, they can be interpreted as co-referenced descriptions of the same urban locations.

In the present implementation, layered scene typing is performed at the grid-cell level. For each spatial unit, the visual-exposure type derived from the panoramic layer and the volumetric spatial-presence type derived from the column layer are combined into a layered scene profile. The projection-based composition layer can also be overlaid within the same spatial framework, providing additional information on facade, ground and near-ground surface conditions (Peng et al., 2025). In this paper, the layered scene typing focuses on the visual-exposure and volumetric spatial-presence layers to keep the synthesis concise, while the projection-based layer is used as contextual support for interpretation. More extensive combinations of projection, visual, volumetric and other indicator layers could support richer scene profiles and additional analytical questions in future applications. The resulting layered scene profiles provide a compact way to examine how pedestrian-level visual exposure, three-dimensional spatial presence and surface composition relate to one another across the study area. The specific combined types and their spatial distribution are reported in the Results section.

5. Results

This section first reports the construction and quality of the S3DGS analytical base, including data products, visualization, semantic propagation and alignment evaluation. The subsequent subsections present three indicator layers: projection-based semantic composition, observer-centered visual exposure and volumetric spatial-presence aggregation. Finally, the layered scene-typing result illustrates how these outputs can support integrated cross-layer interpretation within the same spatial grid.

5.1. Data products and SPC construction results

The S3DGS analytical base was constructed from UAV imagery, XGRIDS/K1 street-level data, GeoSLAM terrestrial laser scanning data, and AHN airborne laser scanning reference data. Table 3 summarizes the main data products used in the case study, and Fig. 5 shows the acquisition coverage and preprocessing workflow for the Aula–library block.

After cleaning and registration, the unified point cloud used for SPC construction contains approximately 43 million points. The semi-automatic segmentation and refinement process produced 88 patches, with a median patch size of approximately 0.4 million points. Fig. 6 shows the raw point cloud, patch-level segmentation, and resulting semantic classes, while Table 4 summarizes the correspondence between SPC patch classes and CityGML-inspired concepts. Together, these results show that the SPC provides a semantically structured description of the block at patch level.

Table 3
Overview of data products used in the Aula-library block case study.

Component	Source	Volume / result
UAV imagery	DJI Mini 3	489 images; 1.91 GB
Street-level data	XGRIDS K1 scanner	219 images; approx. 11 million points; 2.70 GB
TLS point cloud	GeoSLAM Horizon R1	Approx. 153 million points; 1.37 GB
ALS reference	AHN5 GeoTiles 37EN2.11	Approx. 37 million points; 286 MB
Semantic data	Semi-automatic segmentation; CityGML-inspired schema	88 semantic patches

The fused 3DGS scene provides a dense and visually coherent representation of the study block. Following the two-branch workflow described in Section 3, the street-level component and the UAV-derived aerial component were generated separately and then aligned and fused into a common world coordinate frame. The resulting Gaussian scene reproduces the main architectural and environmental features of the block, including facades, roofs, ground surfaces, and major vegetation elements, with sufficient visual coherence for block-scale analysis (Fig. 7).

Semantic information was then propagated from the SPC to the fused 3DGS scene. The transferred labels correspond well to the original SPC segmentation, preserving the main boundaries between roofs, facades, pavements, and vegetation (Fig. 8). Taken together, the SPC, fused 3DGS scene, and propagated semantics form the S3DGS analytical base used in the subsequent indicator analysis.

5.2. Alignment, fusion and semantic-transfer quality

The alignment and semantic-transfer evaluation indicates that the fused S3DGS scene provides a sufficiently consistent basis for the subsequent projection-based, observer-centered and volumetric analyses. The quantitative results are summarized in Table 5, using the three metrics (RMSE, kNN label consistency, and voxel-based IoU).

The geometric alignment between the SPC reference and the fused S3DGS scene was validated using six manually selected corresponding control points. The resulting RMSE was 0.0439 m, indicating that the fused 3DGS scene and the SPC reference are aligned within a 5 cm tolerance. This level of accuracy is sufficient for detailed visualization, semantic transfer and block-scale indicator derivation. Importantly, this RMSE evaluates the final common coordinate frame after the street-level and UAV-derived 3DGS components had been aligned and fused, rather than a single-source reconstruction alone.

To assess local semantic-transfer consistency while accounting for point-density differences, each propagated Gaussian label was compared with the majority label of its three nearest SPC neighbors. The resulting label consistency reached 92.24%, showing that most Gaussian splats received semantic categories consistent with the local SPC neighborhood. The average distance of mismatched splats was 0.036 m, suggesting that most inconsistencies occurred near semantic boundaries, depth transitions or areas with limited spatial overlap between the SPC and the fused 3DGS scene.

A voxel-based IoU analysis was conducted to quantify spatial overlap between SPC and 3DGS patches. Using a voxel size of 0.25 m, the mean IoU across 12 patches was 0.258. This relatively low value reflects the different sampling principles of the two representations: the SPC is metrically dense and surface-oriented, whereas 3DGS samples are view-dependent and optimized for photorealistic rendering rather than complete geometric occupancy. The voxel IoU should therefore be

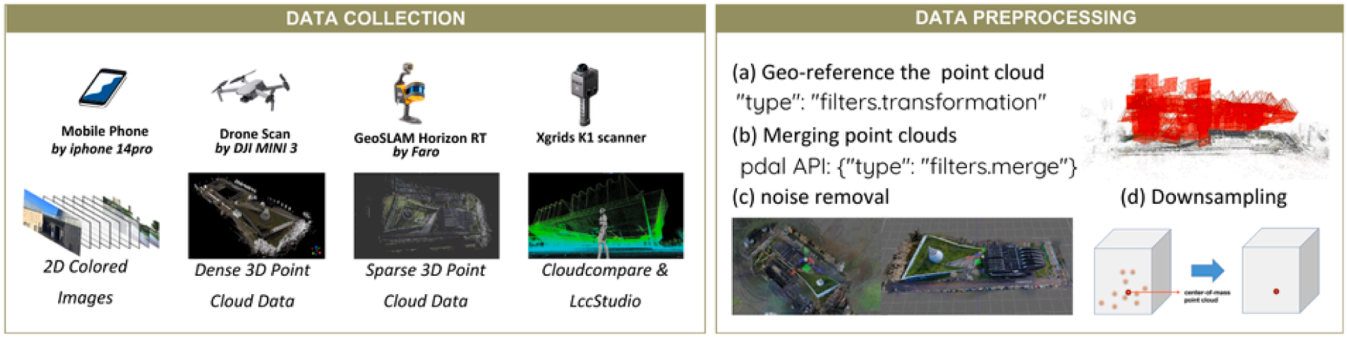

Fig. 5. Details of data collection and preprocessing.

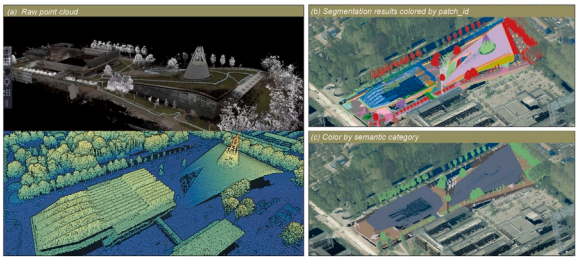

Fig. 6. Different types of raw point cloud and semantic data: (a) Raw point cloud, (b) patch-level segmentation, and (c) resulting semantic classes.

Table 4
SPC patch classes and correspondence to CityGML-inspired concepts.

SPC class		CityGML concept	Description	Patch count (88)
Building	Roof	bldg: RoofSurface	Roof planes of the aula and library	3
	Wall	bldg: WallSurface	Primary vertical facade surfaces	10
	Floor	bldg: FloorSurface	Exposed floor / slab surfaces	5
	Stair	bldg: BuildingInstallation (stair)	Exterior stair volumes represented as surfaces	4
	Railing	bldg: BuildingInstallation (railing)	Railings attached to stairs or balconies	6
	Structural column	bldg: StructuralElement	Exposed concrete columns around the aula	9
Road/Paved ground		dem: GroundSurface	Paved ground surfaces in the block	15
Street furniture		frn: CityFurniture	Benches, bollards or other street furniture	5
Green land/Unpaved ground		veg: PlantCover	Grass, planting beds and low vegetation	24
Tree canopy/ Vegetation		veg: SolitaryVegetationObject / veg: PlantCover	Tree crowns and taller vegetation	7

interpreted as a measure of spatial correspondence rather than as a direct indicator of physical volumetric accuracy. Despite the modest IoU value, the high label consistency confirms that semantics are transferred coherently across the two representations.

Together, these results indicate that the fused S3DGS scene provides a sufficiently reliable geometric and semantic basis for the multi-dimensional indicator analyses reported in the following sections.

5.3. Projection-based semantic composition results

The projection-based module produces facade-oriented semantic projections and height-specific horizontal slices from the S3DGS scene. Fig. 9**a-b** illustrate two horizontal semantic slices at 1 m and 3 m above ground. Rather than functioning as conventional top-down views, these slices reveal how semantic elements occupy different vertical levels of the pedestrian environment. The 1 m slice shows that most ground-level spaces within the block remain open and accessible for pedestrian movement, with low-level obstacles mainly concentrated around tree trunks, structural columns, walls and street furniture. When interpreted together with the 3 m slice, several semi-covered or shaded spatial conditions become visible. On the eastern side of the library, a noticeable amount of tree-shaded space is formed by the overlap between accessible ground and overhead canopy. The northern tree-lined edge shows larger and more continuous canopy coverage, while the southern edge contains smaller and more fragmented shaded areas. Around the Aula, the eastern and western sides form relatively large grey spaces associated with building overhangs and structural supports. The library area also appears to involve a sloped ground condition, which is reflected in the changing relationship between the 1 m and 3 m slices. These findings show that the multi-height slices can reveal accessibility, shade and semi-covered spatial conditions that are difficult to distinguish from a conventional plan view.

Fig. 9**c-d** present facade-oriented semantic projections. In these views, the building envelope is decomposed into structural walls, columns, roof edges, adjacent ground surfaces, and vegetated strips. Compared with point-cloud-only slice analysis, S3DGS improves the visual representation of visually salient but sparsely sampled elements, especially glass and water surfaces, which are often weakly captured in laser-scanning data because of transparency, reflection or poor returns. In this study, this improvement is mainly relevant for projection-based and exposure-oriented interpretation, while geometric boundaries and volumetric interpretation of these surfaces should still be treated with caution.

From these projections and slices, class shares can be computed for green infrastructure, buildings, pavement, water, and other urban elements. The results reveal continuous lawn and planting areas around the library, fragmented greenery along the internal street, and limited green cover in the central paved plaza between the Aula and the library.

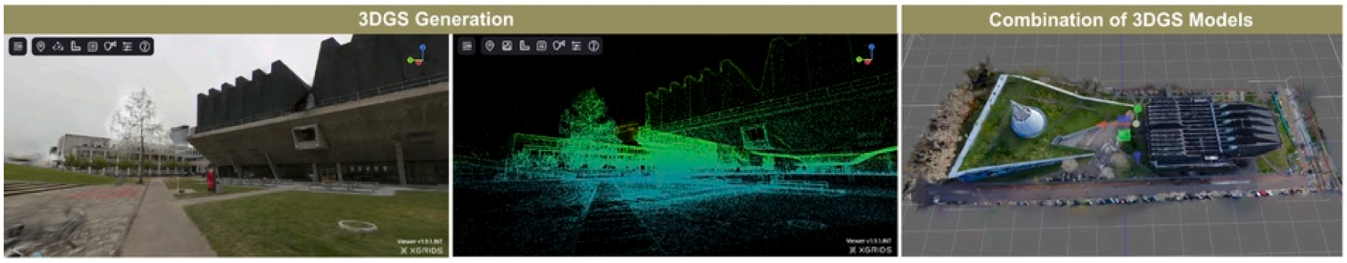

Fig. 7. 3DGS generation from Postshot and Lccstudio software.

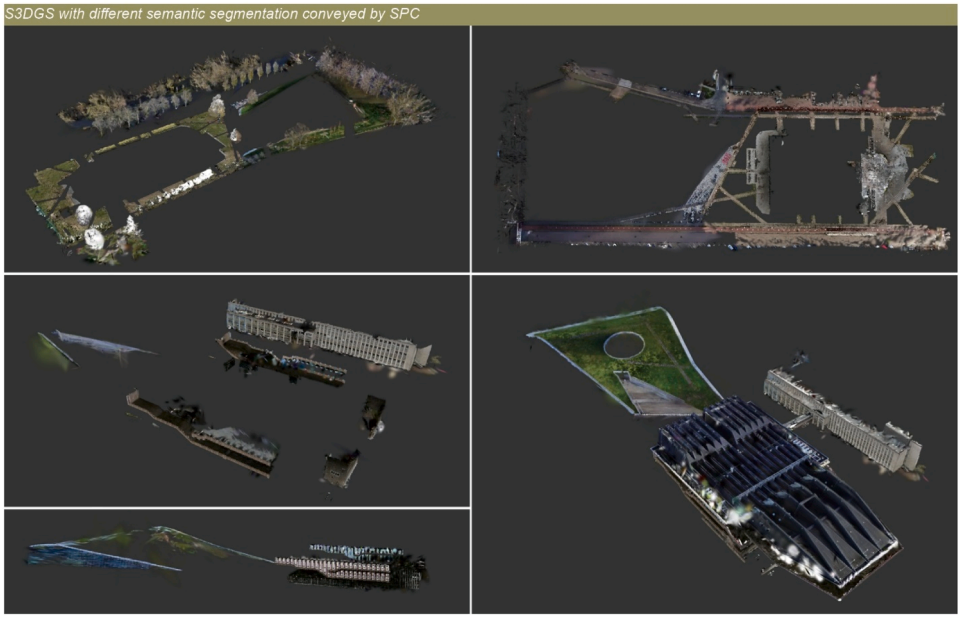

Fig. 8. SPC-guided semantic propagation to the fused S3DGS representation.

Table 5
Quantitative evaluation results for geometric alignment and semantic propagation between SPC and 3DGS.

Metric	Description	Result	Interpretation
RMSE (m)	Root-mean-square error between manually selected corresponding points in SPC and 3DGS	**0.0439**	Accurate geometric alignment (< 5 cm)
kNN label consistency (%)	Percentage of propagated Gaussian labels matching the local majority label of the three nearest SPC neighbors	**92.24 %**	High local semantic-transfer consistency
Mean mismatch distance (m)	Average distance between misclassified Gaussians and their nearest SPC reference	**0.036**	Errors confined to local boundaries
Mean voxel IoU	Intersection-over-union between SPC and 3DGS voxel grids (25 cm resolution)	**0.258**	Low volumetric overlap due to sampling and density differences

5.4. Observer-centered visual exposure results

The observer-centered module generates semantic panoramas from S3DGS and compares them with segmented SVI panoramas at corresponding accessible viewpoints. Fig. 10 shows representative examples of the two outputs. Compared with SVI segmentation, the S3DGS renderings show more coherent semantic boundaries. As highlighted by the red boxes, several SVI results contain broken or noisy semantic patches caused by the presence of the scanning vehicle, material shadows on the ground or local image-segmentation errors. Such artefacts often require image-by-image cleaning before SVI-derived indicators can be used reliably. In the S3DGS panoramas, these fragmented patches and local misclassifications are reduced because the semantic view is rendered from a labeled 3D scene rather than segmented independently from each image.

The S3DGS-based panoramas also provide a denser and more flexible observation field. While SVI is limited to existing camera locations, S3DGS allows virtual viewpoints to be placed across courtyards, internal campus paths and other locations where SVI coverage is incomplete or unavailable. As shown in Fig. 11**a-f**, this enables visual-exposure patterns to be mapped more continuously across the study area. As a result, this comparison is not intended as a full accuracy benchmark. Rather, it provides limited evidence that S3DGS can reduce image-specific segmentation artefacts and extend observer-centered analysis beyond available SVI locations.

The resulting S3DGS-based visual layer reveals clear spatial patterns. Tree-lined paths show stronger vegetation presence, while the central paved plaza between the Aula and the library is characterized by higher building and sky exposure. Around the library entrance, vegetation becomes more prominent again due to nearby planting beds and trees. Based on the reduced visual-exposure signature of vegetation, building and sky shares, the visual layer is further grouped into five recurrent visual-exposure types, with the number of clusters selected by elbow analysis. These types provide the observer-centered layer for the later

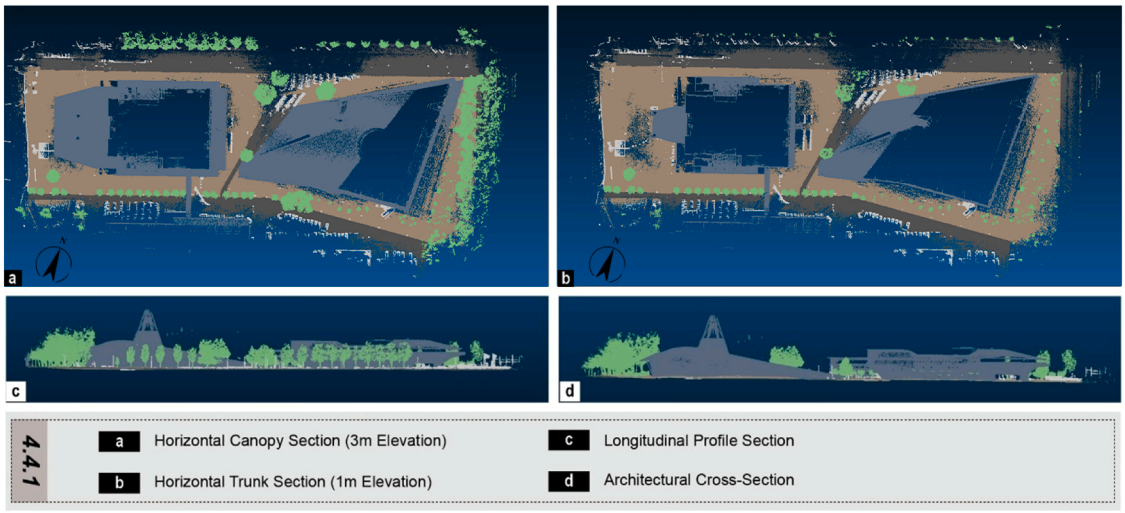

Fig. 9. Projection-based semantic composition results from the S3DGS scene, including horizontal semantic slices at 1 m and 3 m and facade-oriented semantic projections.

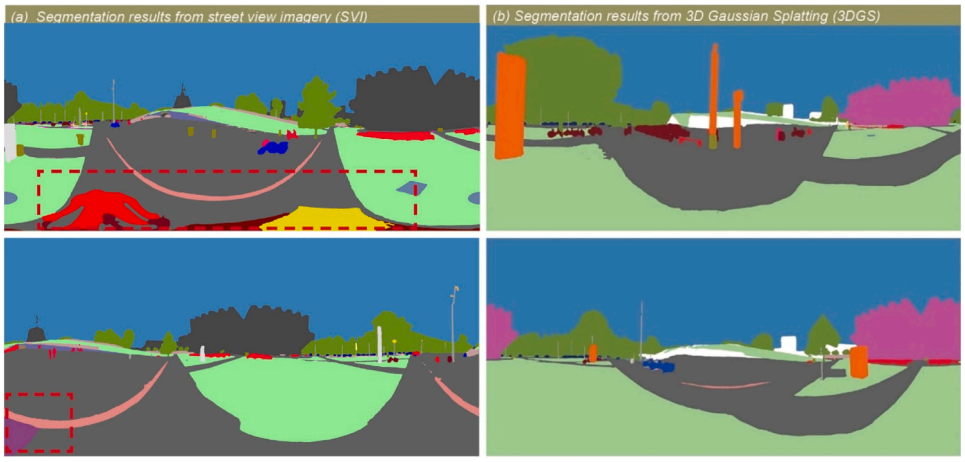

Fig. 10. Comparison between SVI semantic segmentation and S3DGS semantic panoramic rendering.

layered scene interpretation.

The clustering result summarizes the spatial variation of pedestrian-level visual exposure by grouping viewpoints with similar proportions **(E1-E5,** Fig. 11**g-h)**. Cluster E1 represents sky-open viewpoints, Cluster E2 green-open viewpoints, Cluster E3 ground-dominated open viewpoints, Cluster E4 building-enclosed viewpoints, and Cluster E5 street-furniture-affected viewpoints. The fifth type contains few samples and mainly results from a small number of grid-based viewpoints coinciding with localized street-furniture elements, so it should be interpreted as a minor sampling-related category.

5.5. Volumetric green and building exposure in 3D columns

The volumetric module produces column-based semantic composition results for the 5 × 5 m grid. Fig. 12**a** shows the spatial distribution of the five volumetric spatial-presence clusters, while Fig. 12**b** summarizes the stacked semantic composition of each cluster. Table 6 reports the overall semantic proportions across the study area. Building has the largest spatial-presence share, followed by road surface, vegetation and ground, while water and street furniture account for smaller proportions. These values should be interpreted as spatial-presence indicators rather than exact material volumes.

The five clusters (V1-V5) reveal distinct three-dimensional composition patterns (Fig. 13**a-b**). Cluster V1 is building-dominated, with building elements occupying almost the entire column composition. It corresponds mainly to the central built masses of the Aula and library. Cluster V2 is a mixed hard-surface type, combining road surface, ground, vegetation and a small building component. Spatially, it appears around transitional areas where circulation surfaces, open ground and nearby built or planted elements overlap. Cluster V3 is ground-dominated, with ground surfaces forming the largest component, and is associated with open accessible spaces and paved areas. Cluster V4 is vegetation-dominated and is concentrated along planted edges, tree-lined surroundings and green areas. Cluster V5 is road-surface-dominated, with road surface forming the main component, and is mainly distributed along circulation corridors and hard-surface movement spaces.

Compared with a map that shows only the dominant semantic class, the stacked composition reveals the internal structure of each volumetric type. It shows that some spatial units are compositionally mixed, especially at transitions between building edges, planted areas and circulation surfaces. These volumetric spatial-presence types provide the third analytical layer for the later layered scene interpretation.

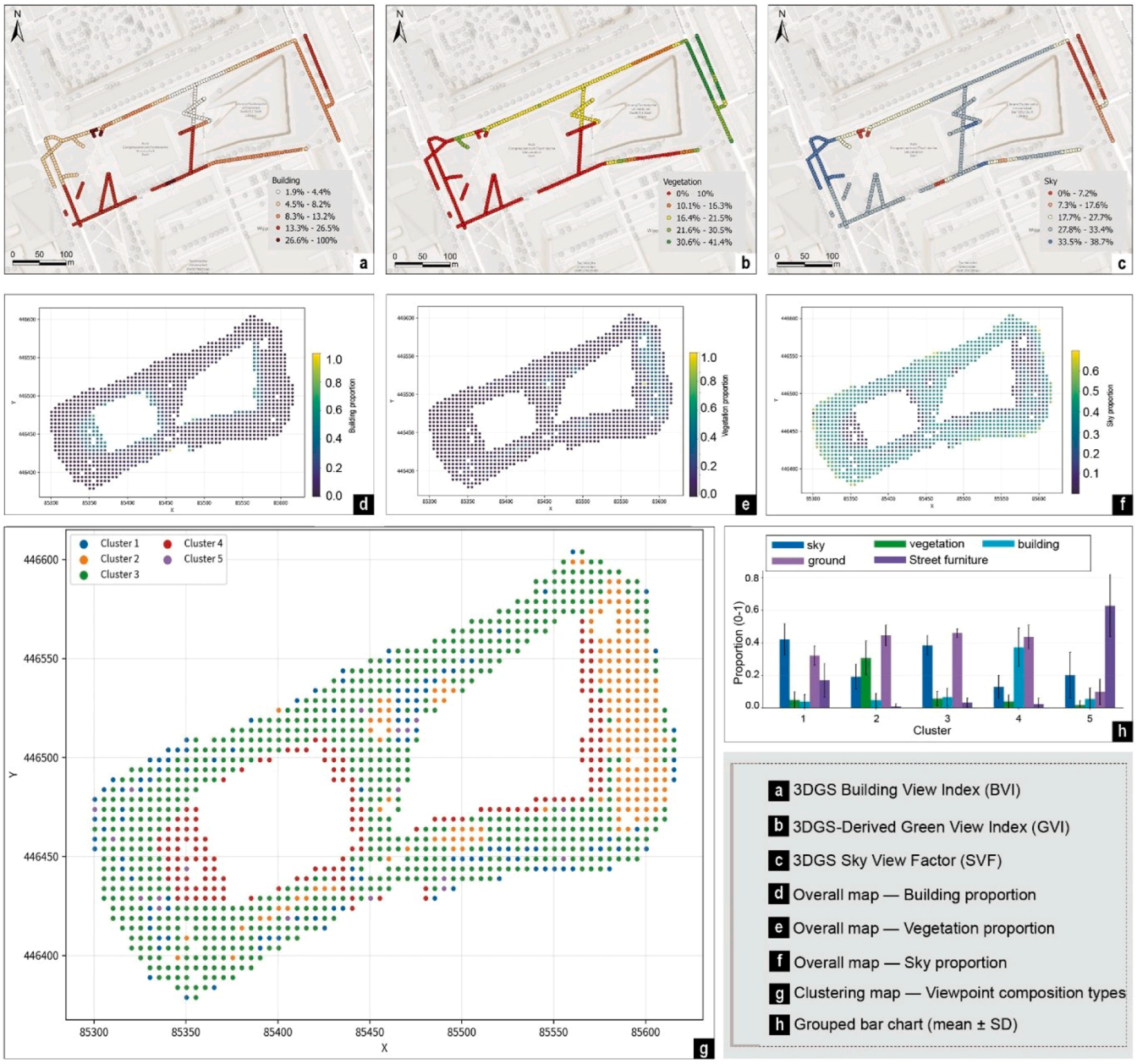

Fig. 11. Spatial patterns of observer-centered visual exposure and visual-exposure types derived from SVI and S3DGS panoramas.
Note: The visual-exposure type 5 contains only a small number of samples. This is mainly because the virtual viewpoints were placed according to the regular grid, and a limited number of viewpoints coincided with street-furniture elements. As a result, this type should be interpreted as a minor grid-sampling artefact rather than as a dominant visual condition of the site.

5.6. *Layered scene-typing results*

The final synthesis result is shown in Fig. 14. The map overlays the volumetric spatial-presence layer and the observer-centered visual-exposure layer within the same spatial units. In the figure, V1to V5 denote volumetric spatial-presence types, E1 to E5 denote visual-exposure types, and EU indicates units without a valid exposure assignment. Each combined code links the three-dimensional composition of a spatial unit with its pedestrian-level visual condition. This layered result shows the value of the shared S3DGS reference: volumetric and visual information can be read together for the same grid cell. It identifies areas where spatial presence and visual exposure are consistent, as well as areas where they diverge.

One example is V1-E3, which appears near the entrance side of the TU Delft Library. This type indicates a condition in which the volumetric layer records a strong building presence, while the visual-exposure layer is less dominated by the building and is more strongly associated with greenery or open ground in the field of view. In other words, the building mass remains present in three-dimensional space, but its visual exposure from the pedestrian perspective is reduced. This pattern is consistent with the design reading of the library as a substantial building volume that visually recedes into the surrounding campus landscape and maintains a modest relationship with the existing monumental context.

More broadly, the layered scene-typing result allows projection, visual, and volumetric outputs to be organized within a shared spatial reference. Each spatial unit can therefore be interpreted through multiple linked layers, providing a more nuanced description of urban scene types than a single indicator layer alone.

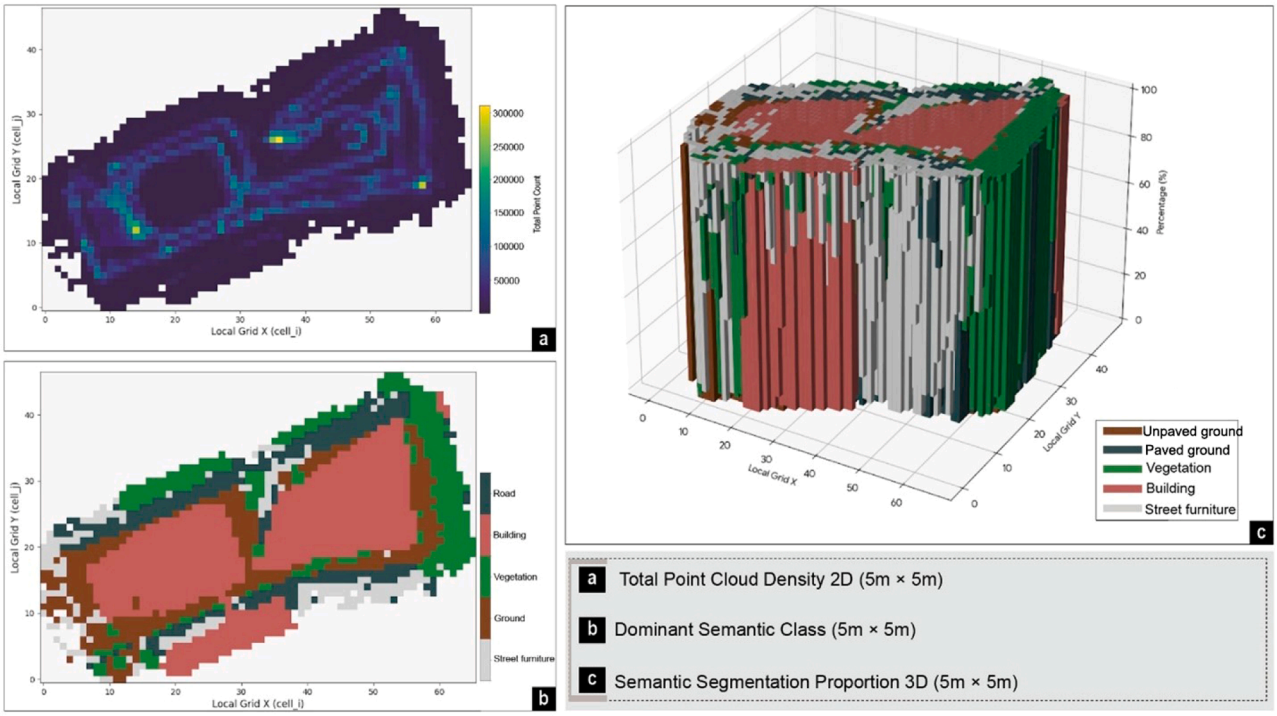

Fig. 12. Volumetric spatial-presence aggregation results in 3D columns.

Table 6
Semantic category proportions derived from S3DGS-based volumetric spatial-presence aggregation.

Semantic category	Voxel count	Voxel volume	Category volume percentage
street furniture	42256	660.25	0.027622
Green land/ Unpaved ground	240848	3763.25	0.157437
Vegetation	311365	4865.078	0.203532
Building	526484	8226.313	0.34415
Road/paved ground	326695	5104.609	0.213553
Water	5135	1283.75	0.053706
TOTAL	23903.25		

6. Discussions

The preceding sections demonstrated the proposed SPC–S3DGS workflow through the Aula–library block in Delft. The case shows how a shared S3DGS analytical base can support projection-based composition analysis, observer-centered visual exposure analysis, volumetric spatial-presence aggregation and layered scene typing within one semantic and spatial reference. The following discussion considers the analytical value of this shared reference, its possible application pathways, the uncertainty and transparency of the current workflow, and the limitations that remain for future work.

6.1. Shared analytical reference, layered interpretation and flexible observation

The added value of the proposed S3DGS framework lies in three aspects: analytical consistency, layered interpretation and flexible observer-centered analysis. Established GIS, SVI, point-cloud and 3D city-model workflows already support many individual indicator families. S3DGS adds a shared semantic, spatial and visual reference through which these indicator families can be derived and interpreted together. S3DGS organizes projection-based composition, observer-centered visual exposure and volumetric spatial presence within the same labeled Gaussian scene, allowing them to share semantic labels, spatial units and visual references.

First, S3DGS improves analytical consistency across indicator families. In separate workflows, projection-based, image-based and volumetric indicators are often produced from different data sources, acquisition times, coordinate assumptions and spatial sampling designs (Kolbe et al., 2015). Even within SVI-based analysis, repeated images reported at the same geographic location may still show positional shifts because of differences in camera trajectory, capture time, platform metadata or image selection (Fan et al., 2025; Sánchez & Labib, 2024). These inconsistencies become more pronounced when image-derived indicators are compared with 3D city models, point clouds or GIS layers constructed from different acquisition campaigns (Peng et al., 2026). Differences in projection methods, viewing geometry, temporal currency and semantic classification can introduce offsets between what is visible in images and what is represented in spatial models. In the present workflow, facade projections, horizontal slices, panoramic renderings and column-based volumetric measures are linked to the same semantic structure and spatial reference (Kutzner et al., 2020). This shared reference reduces the need for post-hoc reconciliation across separately produced analytical products and allows differences between layers to be interpreted as urban conditions rather than as artefacts of incompatible datasets.

Second, S3DGS supports layered interpretation. The layered scene-typing result shows how volumetric presence, visual exposure and surface composition can describe different aspects of the same location. Its significance lies in connecting physical presence, visible experience and surface condition within the same analytical unit (Peng et al., 2025). For example, a space may contain strong building mass in the volumetric layer while showing limited building dominance from the pedestrian viewpoint; such a condition indicates that built form is physically present but visually moderated by landscape, openness or viewing position. Similarly, shaded hard-surface spaces, vegetation-rich ground areas, visually green facade edges and mixed transitional zones can be distinguished by comparing visual, volumetric and projection-based layers. These relationships are difficult to identify from a single model or indicator layer, because planar maps, street-level images and volumetric

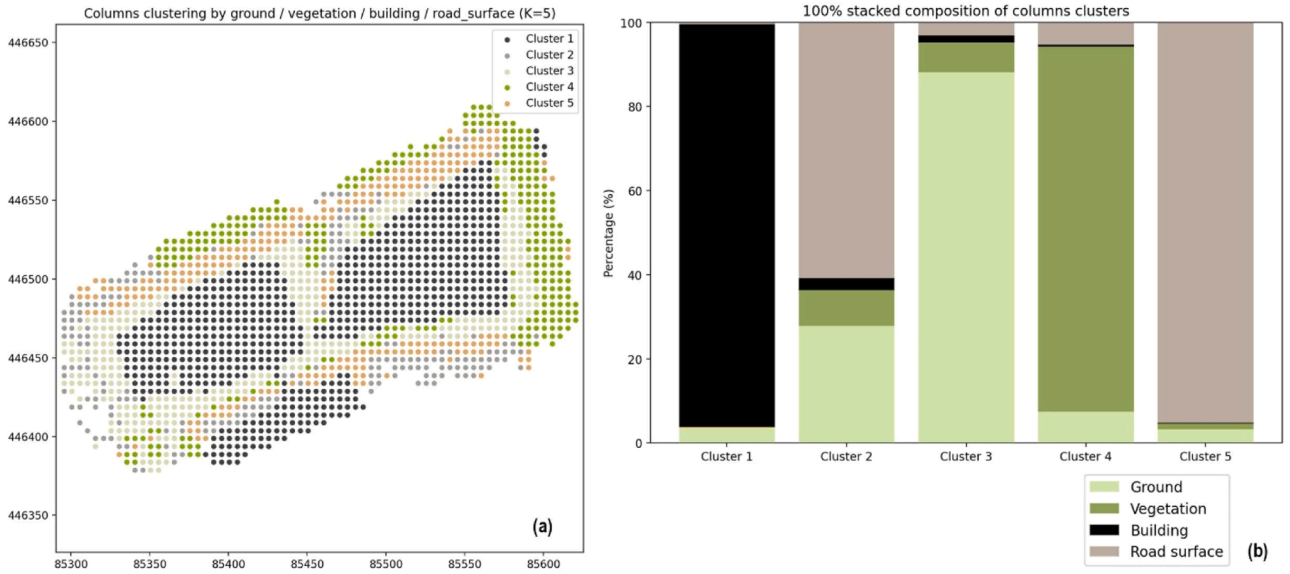

Fig. 13. Volumetric spatial-presence aggregation and clustering results: (a) Spatial distribution of the five volumetric column clusters across the 5 × 5 m grid. (b) Stacked semantic composition of each cluster based on ground, vegetation, building and road-surface proportions.
Note: Street furniture was excluded from the volumetric clustering features. It was retained in the overall semantic aggregation, but its localized small-scale volumes may dilute the relative shares of ground and road-surface elements. The clustering therefore uses ground, vegetation, building and road-surface shares as the main volumetric signature.

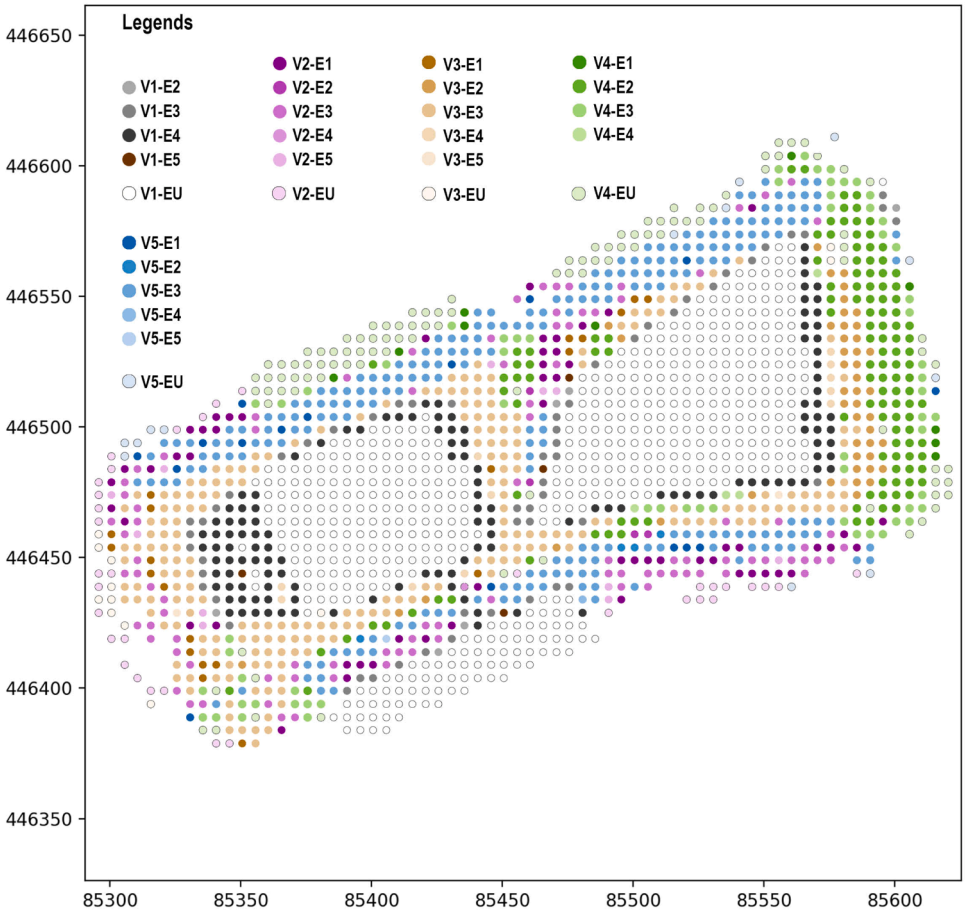

Fig. 14. Layered scene-typing result from co-referenced volumetric and visual-exposure layers: V1–V5 denote volumetric spatial-presence types, E1–E5 denote visual-exposure types, and EU indicates units without a valid exposure assignment.

representations each describe only part of the urban condition. The layered S3DGS reading therefore supports interpretation of how built

form, greenery, openness and surface composition interact at human scale.

Third, S3DGS extends observer-centered analysis beyond the limits of conventional viewpoint- and image-acquisition-based SVI workflows. SVI remains valuable for empirical street-level observation, but it is constrained by available camera positions, image quality, acquisition timing and segmentation uncertainty (Biljecki & Ito, 2021; Hou & Biljecki, 2022). In S3DGS, semantic panoramas can be rendered from controllable virtual viewpoints, including courtyards, internal paths and different viewing heights. Since the panoramas are rendered from a labeled 3D scene, vegetation edges, facade boundaries and small spatial transitions can appear more coherent than in image-segmentation results affected by fragmented patches or local misclassification. The SVI–S3DGS comparison therefore demonstrates an observation-design gain: S3DGS extends SVI-style visual analysis from discrete camera samples to a denser and more controllable observation field, while keeping the visual indicators aligned with the underlying 3D semantic structure. The 5 × 5 m grid used in this study is one operational scale for linking these layers. Finer spatial units could support detailed site diagnosis, while coarser units could support broader planning assessment. The analytical resolution can therefore be adjusted according to the scale and depth of the research question (Lei et al., 2023). Across these resolutions, the key role of S3DGS is to keep projection-based, observer-centered and volumetric outputs traceable to the same semantic scene.

Together, these results show that S3DGS adds value across both point-cloud-based and image-based analysis (Table 7). The image-driven Gaussian component improves the representation of visually salient elements that are weakly sampled by laser scanning, such as water surfaces and glazing (Whelan et al., 2018), while controllable virtual viewpoints extend SVI-style exposure analysis from fixed camera locations to a denser human-viewpoint sampling field. Both gains remain linked to the same semantic reference, allowing projection, visual and volumetric outputs to be interpreted within one analytical scene.

6.2. Implementation requirements and reliability

For practical implementation, the S3DGS workflow requires a minimum analytical package. This includes a sufficiently complete semantic point-cloud reference, a multi-view image set with adequate overlap and viewpoint diversity, a semantic schema aligned with the intended analysis, and a registration pathway that brings the point-cloud reference and Gaussian scene into a shared coordinate frame. A rendering and analysis pipeline is also needed to generate projections, semantic panoramas, voxel-based column aggregations and layered scene profiles.

The evaluation results reported in Section 5.2 indicate that the current implementation provides a stable basis for block-scale semantic transfer and indicator derivation, with an alignment RMSE of 0.0439 m, kNN label consistency of 92.24%, an average mismatch distance of 0.036 m and a mean voxel IoU of 0.258. These results are not intended as a general benchmark of 3DGS reconstruction accuracy; instead, they show that the fused SPC–S3DGS reference is sufficiently co-registered and semantically coherent for the analytical tasks demonstrated in this study.

For larger-scale deployment, S3DGS is more realistically implemented as a tile-based or block-based analytical layer rather than as a single city-wide model (Biljecki et al., 2015b). City-scale urban digital twins may retain coarser CityGML, CIM, GIS or point-cloud layers for general management and monitoring tasks, while S3DGS can be generated selectively for areas requiring high-resolution visual-semantic interpretation. Such areas may include heritage blocks, public-space redesign sites, campus environments, streetscape intervention zones and climate-adaptation hotspots. This selective strategy helps control segmentation effort and data-processing cost, because high semantic

Table 7
Main analytical contributions of S3DGS compared with separate workflows.

Contribution	Challenge in separate workflows	S3DGS-based implementation	Analytical gain
Analytical consistency	Projection-based, image-based and volumetric indicators are often derived from different datasets, semantic classes, coordinate systems and spatial units.	Facade projections, horizontal slices, semantic panoramas and column-based aggregations are generated from the same labeled S3DGS scene. The workflow achieved an alignment RMSE of 0.0439 m and a kNN label consistency of 92.24%.	Supports consistent comparison of structural, visual and volumetric indicators and reduces post-hoc reconciliation across separate analytical products.
Layered interpretation	Single indicator layers may describe surface composition, visual exposure or three-dimensional spatial presence separately, but may not show how these dimensions interact at the same location.	Projection-based composition, visual-exposure types and volumetric spatial-presence types are linked within the same 5 × 5 m grid cells.	Enables cross-layer readings such as physically present but visually recessive building mass, shaded hard-surface spaces and mixed transitional zones.
Flexible observation	SVI-based analysis is constrained by existing camera locations, accessible routes, image quality and segmentation artefacts.	Semantic panoramas are rendered from controllable virtual viewpoints in S3DGS, including courtyards, internal paths and different viewing heights.	Extends SVI-style exposure analysis from discrete image samples to a denser and more controllable observation field aligned with the 3D semantic scene.

and visual detail is applied where it directly supports planning or design questions (Biljecki et al., 2016; Biljecki et al., 2014). It also supports differentiated updating: relatively stable elements such as building massing and ground structure can be updated less frequently, while vegetation, construction zones, temporary interventions and rapidly changing public spaces can be refreshed seasonally or project by project. When applied across larger districts, this tile-based logic allows S3DGS outputs to be connected through a shared spatial grid, semantic schema and reference coordinate system (Lei, Stouffs, et al., 2023). Individual tiles can be generated, updated or replaced independently, while still contributing to a broader analytical framework. This makes S3DGS more suitable as a selective high-resolution layer within urban digital twins than as a fully continuous replacement for existing city-scale models.

6.3. Application pathways of the proposed S3DGS workflow

Building on these implementation requirements and reliability considerations, this section outlines the potential application pathways of S3DGS in urban analysis, planning assessment and digital twin workflows. The S3DGS framework can support urban analysis and design tasks that require structural, visual and spatial-presence information to be interpreted together. In public-space assessment, it can relate ground composition, visual openness and three-dimensional vegetation presence when evaluating comfort, enclosure or potential resting areas (Biljecki et al., 2015a). In streetscape evaluation, it can extend SVI-style visual indicators to internal paths, courtyards and pedestrian areas where street-view imagery is unavailable. In heritage and campus environments, it can support the analysis of facade visibility, visual dominance, shaded spaces and the relationship between new interventions and existing architectural settings (Salimi et al., 2023). In green-infrastructure planning, it can distinguish visually prominent greenery from spatially substantial vegetation and support more nuanced assessment of planting, canopy cover and hard-surface exposure (Casalegno et al., 2017; Peng, Li, et al., 2026).

The same logic can be embedded in broader technical workflows such as design review, scenario comparison, environmental assessment and urban digital twins. For example, alternative tree-planting, facade modification or plaza redesign scenarios could be evaluated through the same projection, visual-exposure and volumetric layers. In more advanced computational planning workflows, S3DGS can serve as a visual-semantic analytical layer linked to object-based models, dashboards, simulation tools or digital twin platforms (Dembski et al., 2020; Peng et al., 2026; Yu et al., 2020). Its main role is to provide an analysis-ready environment where structural, visual and spatial-presence information can be interpreted together across different planning and design tasks.

6.4. Limitations and future work

Several limitations remain. First, the empirical demonstration is based on a single, relatively well-structured campus block with high-quality input data, so the robustness of the framework in denser, more heterogeneous urban areas still needs further testing. Second, S3DGS quality depends strongly on image coverage, viewpoint diversity, resolution, illumination and occlusion. Dynamic objects such as pedestrians and vehicles may introduce transient artefacts, while reflective or transparent surfaces such as water and glass remain difficult to stabilize geometrically and semantically (Schönberger et al., 2016). Third, SPC construction currently involves semi-automatic segmentation with rule-based refinement (Li et al., 2023; Luiten et al., 2024). This supports interpretability in the present case, but may become costly at larger scales. In addition, scanner-native tools such as LCC Studio can support parts of the workflow, but format incompatibility and camera-model differences may still limit downstream fusion and semantic propagation. Finally, the volumetric indicators should be interpreted as voxel-based spatial-presence or volumetric composition indicators rather than exact material volumes, because they depend on voxel resolution, point-cloud completeness, segmentation quality and scene coverage (Gehrung et al., 2018; Zięba-Kulawik et al., 2021).

Future work should test the framework across larger and more diverse urban settings, including denser blocks, irregular street networks and mixed acquisition conditions. Further research should also compare different spatial and semantic resolutions to clarify when coarse semantic grouping is sufficient and when finer classification is needed for site-level diagnosis. At the implementation level, larger scenes will require tiling, hierarchical processing, Gaussian pruning, quantization or streaming strategies to reduce computational cost and memory footprint. Future work should also improve automation in SPC construction and semantic propagation, strengthen interoperability across open workflows such as RealityCapture, COLMAP and Postshot, and explore emerging UAV or mobile systems with stronger Gaussian-splatting-oriented acquisition capabilities. These developments would make the S3DGS workflow more accessible, scalable and suitable for operational urban digital twin and planning applications.

7. Conclusion

This paper developed an SPC-guided S3DGS analytical framework and demonstrated it through the Aula–library block in Delft, using laser-scanning point clouds and multi-view imagery to construct a shared semantic, spatial and visual base for urban analysis.

The study makes three main contributions. First, it shows how SPC semantics can be propagated to fused street-level and aerial 3DGS components, producing an S3DGS analytical base in which projection-based, observer-centered and volumetric analyses share the same semantic and spatial reference. This reduces the inconsistency that can arise when GIS, SVI, point-cloud and 3D-model outputs are combined after analysis. Second, it demonstrates a layered analytical workflow in which facade projections, height-specific horizontal slices, semantic panoramas, volumetric spatial-presence clusters and layered scene types can be derived and interpreted within the same scene. This enables linked interpretation of urban conditions that may be difficult to capture with a single analytical layer, such as visually recessive building mass, shaded hard-surface areas or mixed transitional spaces. Third, it demonstrates the broader observational flexibility of S3DGS for urban analysis. The framework supports controllable virtual viewpoints, different viewing heights, semantic panoramic rendering, facade-oriented projection and high-fidelity visualization of visually salient elements such as glass and water surfaces. These capacities provide a flexible basis for studying urban form, visual experience and spatial presence together.

Future work should test the framework across larger and more diverse urban settings, compare different spatial and semantic resolutions, improve automation and model efficiency, and explore emerging Gaussian-splatting-oriented acquisition workflows. These developments may further strengthen the use of S3DGS as an analysis-ready base for integrated urban assessment.

CRediT authorship contribution statement

Yingwen Yu: Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Zhuoyue Wang:** Software, Methodology. **Guanting Zhang:** Methodology. **Peter van Oosterom:** Supervision, Methodology, Conceptualization. **Edward Verbree:** Supervision, Formal analysis, Conceptualization. **Steffen Nijhuis:** Supervision, Methodology. **Yuyang Peng:** Writing – original draft, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Data availability

Data will be made available on request.

References

Beil, C., Kutzner, T., Schwab, B., Willenborg, B., Gawronski, A., & Kolbe, T. H. (2021). Integration of 3D point clouds with semantic 3D city models: Providing semantic information beyond classification. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, VIII-4/W2-2021*, 105–112. https://doi.org/10.5194/isprs-annals-VIII-4-W2-2021-105-2021

Benedikt, M. L. (1979). To take hold of space: Isovists and isovist fields. *Environment and Planning B: Planning and Design, 6*(1), 47–65.

Berra, E. F., & Peppa, M. V. (2020). Advances and challenges of UAV SFM MVS photogrammetry and remote sensing: Short review. In *2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)*.

Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning, 215*, Article 104217. https://doi.org/10.1016/j.landurbplan.2021.104217

Biljecki, F., Ledoux, H., Stoter, J., & Vosselman, G. (2016). The variants of an LOD of a 3D building model and their influence on spatial analyses. *ISPRS Journal of Photogrammetry and Remote Sensing, 116*, 42–54. https://doi.org/10.1016/j.isprsjprs.2016.03.003

Biljecki, F., Ledoux, H., Stoter, J., & Zhao, J. (2014). Formalisation of the level of detail in 3D city modelling. *Computers, Environment and Urban Systems, 48*, 1–15. https://doi.org/10.1016/j.compenvurbsys.2014.05.004

Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S., & Çöltekin, A. (2015a). Applications of 3D city models: State of the art review. *ISPRS International Journal of Geo-Information, 4*(4), 2842–2889. https://www.mdpi.com/2220-9964/4/4/2842.

Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S., & Çöltekin, A. (2015b). Applications of 3D city models: State of the art review. *ISPRS International Journal of Geo-Information, 4*(4), 2842–2889.

Billen, R., Cutting-Decelle, A.-F., Métral, C., Falquet, G., Zlatanova, S., & Marina, O. (2015). Challenges of semantic 3D city models: A contribution of the COST research action TU0801. *Int. J. 3D Inf. Model., 4*(2), 68–76. https://doi.org/10.4018/ij3dim.2015040106

Bonczak, B., & Kontokosta, C. E. (2019). Large-scale parameterization of 3D building morphology in complex urban landscapes using aerial LiDAR and city administrative data. *Computers, Environment and Urban Systems, 73*, 126–142. https://doi.org/10.1016/j.compenvurbsys.2018.09.004

Cárdenas-León, I., Koeva, M., Nourian, P., & Davey, C. (2024). Urban digital twin-based solution using geospatial information for solid waste management. *Sustainable Cities and Society, 115*, Article 105798. https://doi.org/10.1016/j.scs.2024.105798

Casalegno, S., Anderson, K., Cox, D. T. C., Hancock, S., & Gaston, K. J. (2017). Ecological connectivity in the three-dimensional urban green volume using waveform airborne lidar. *Scientific Reports, 7*(1), Article 45571. https://doi.org/10.1038/srep45571

Chen, W., Zhong, R., Wang, K., & Xie, D. (2026). Li-GS: a fast 3D Gaussian reconstruction method assisted by LiDAR point clouds. *Big Earth Data, 10*(1), 68–92. https://doi.org/10.1080/20964471.2025.2479428

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Dembski, F., Wössner, U., Letzgus, M., Ruddat, M., & Yamu, C. (2020). Urban digital twins for smart cities and citizens: The case study of Herrenberg, Germany. *Sustainability, 12*(6), 2307. https://www.mdpi.com/2071-1050/12/6/2307.

Engin, Z., van Dijk, J., Lan, T., Longley, P. A., Treleaven, P., Batty, M., & Penn, A. (2020). Data-driven urban management: Mapping the landscape. *Journal of Urban Management, 9*(2), 140–150. https://doi.org/10.1016/j.jum.2019.12.001

Ewing, R., Handy, S., Brownson, R. C., Clemente, O., & Winston, E. (2006). Identifying and measuring urban design qualities related to walkability. *Journal of Physical Activity and Health, 3*(s1), S223–s240. https://doi.org/10.1123/jpah.3.s1.s223

Fan, Z., Feng, C.-C., & Biljecki, F. (2025). Coverage and bias of street view imagery in mapping the urban environment. *Computers, Environment and Urban Systems, 117*, Article 102253. https://doi.org/10.1016/j.compenvurbsys.2025.102253

Fei, B., Xu, J., Zhang, R., Zhou, Q., Yang, W., & He, Y. (2025). 3D gaussian splatting as a new Era: A survey. *IEEE Transactions on Visualization and Computer Graphics, 31*(8), 4429–4449. https://doi.org/10.1109/TVCG.2024.3397828

Gehrung, J., Hebel, M., Arens, M., & Stilla, U. (2018). A voxel-based metadata structure for change detection in point clouds of large-scale urban areas. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 4*, 97–104.

Geneletti, D., Bagli, S., Napolitano, P., & Pistocchi, A. (2007). Spatial decision support for strategic environmental assessment of land use plans. A case study in southern Italy. *Environmental Impact Assessment Review, 27*(5), 408–423. https://doi.org/10.1016/j.eiar.2007.02.005

Gröger, G., & Plümer, L. (2012). CityGML – Interoperable semantic 3D city models. *ISPRS Journal of Photogrammetry and Remote Sensing, 71*, 12–33. https://doi.org/10.1016/j.isprsjprs.2012.04.004

He, W., & Chen, M. (2024). Advancing urban life: A systematic review of emerging technologies and artificial intelligence in urban design and planning. *Buildings, 14*(3), 835. https://doi.org/10.3390/buildings14030835

Hou, Y., & Biljecki, F. (2022). A comprehensive framework for evaluating the quality of street view imagery. *International Journal of Applied Earth Observation and Geoinformation, 115*, Article 103094.. https://doi.org/10.1016/j.jag.2022.103094

Ito, K., Kang, Y., Zhang, Y., Zhang, F., & Biljecki, F. (2024). Understanding urban perception with visual data: A systematic review. *Cities, 152*, Article 105169. https://doi.org/10.1016/j.cities.2024.105169

Jeddoub, I., Nys, G.-A., Hajji, R., & Billen, R. (2023). Digital Twins for cities: Analyzing the gap between concepts and current implementations with a specific focus on data integration. *International Journal of Applied Earth Observation and Geoinformation, 122*, Article 103440. https://doi.org/10.1016/j.jag.2023.103440

Kerbl, B., Kopanas, G., Leimkuehler, T., & Drettakis, G. (2023). 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph., 42*(4), 139. https://doi.org/10.1145/3592433

Kim, J. H., Lee, S., Hipp, J. R., & Ki, D. (2021). Decoding urban landscapes: Google street view and measurement sensitivity. *Computers, Environment and Urban Systems, 88*, Article 101626. https://doi.org/10.1016/j.compenvurbsys.2021.101626

Kolbe, T. H. (2009). Representing and exchanging 3D city models with CityGML. In J. Lee, & S. Zlatanova (Eds.), *3D Geo-Information Sciences* (pp. 15–31). Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-540-87395-2_2.

Kolbe, T.H., Burger, B., & Cantzler, B. (2015). CityGML goes to Broadway. Photogrammetric week.

Kutzner, T., Chaturvedi, K., & Kolbe, T. H. (2020). CityGML 3.0: New functions open up new applications. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 88*(1), 43–61. https://doi.org/10.1007/s41064-020-00095-z

Labetski, A., Vitalis, S., Biljecki, F., Arroyo Ohori, K., & Stoter, J. (2023). 3D building metrics for urban morphology. *International Journal OF Geographical Information Science, 37*(1), 36–67. https://doi.org/10.1080/13658816.2022.2103818

Labib, S. M., Huck, J. J., & Lindley, S. (2021). Modelling and mapping eye-level greenness visibility exposure using multi-source data at high spatial resolutions. *Science of The Total Environment, 755*, Article 143050. https://doi.org/10.1016/j.scitotenv.2020.143050

Ledoux, H., Arroyo Ohori, K., Kumar, K., Dukai, B., Labetski, A., & Vitalis, S. (2019). CityJSON: a compact and easy-to-use encoding of the CityGML data model. *Open Geospatial Data, Software and Standards, 4*(1), 4. https://doi.org/10.1186/s40965-019-0064-0

Lei, B., Janssen, P., Stoter, J., & Biljecki, F. (2023). Challenges of urban digital twins: A systematic review and a Delphi expert survey. *Automation in Construction, 147*, Article 104716. https://doi.org/10.1016/j.autcon.2022.104716

Lei, B., Liang, X., & Biljecki, F. (2024). Integrating human perception in 3D city models and urban digital twins. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 211–218. https://doi.org/10.5194/isprs-annals-X-4-W5-2024-211-2024

Lei, B., Stouffs, R., & Biljecki, F. (2023). Assessing and benchmarking 3D city models. *International Journal of Geographical Information Science, 37*(4), 788–809. https://doi.org/10.1080/13658816.2022.2140808

Li, Y., Jiang, L., Xu, L., Xiangli, Y., Wang, Z., Lin, D., & Dai, B. (2023). Matrixcity: A large-scale city dataset for city-scale neural rendering and beyond. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

Li, Y., Wang, W., Tang, S., Li, D., Wang, Y., Yuan, Z., Guo, R., Li, X., & Xiu, W. (2019). Localization and extraction of road poles in Urban Areas from mobile laser scanning data. *Remote Sensing, 11*(4), 401. https://www.mdpi.com/2072-4292/11/4/401.

Lin, J., Li, Z., Tang, X., Liu, J., Liu, S., Liu, J., Lu, Y., Wu, X., Xu, S., Yan, Y., & Yang, W. (2024). VastGaussian: Vast 3D gaussians for large scene reconstruction. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.ieeecomputersociety.org/10.1109/CVPR52733.2024.00494.

Liu, M., & Nijhuis, S. (2020). Mapping landscape spaces: Methods for understanding spatial-visual characteristics in landscape design. *Environmental Impact Assessment Review, 82*, Article 106376. https://doi.org/10.1016/j.eiar.2020.106376

Liu, Y., Luo, C., Fan, L., Wang, N., Peng, J., & Zhang, Z. (2024). CityGaussian: Real-Time High-Quality Large-Scale Scene Rendering with . In *Gaussians Computer Vision – ECCV 2024: 18th European Conference*. https://doi.org/10.1007/978-3-031-72640-8_15. September 29–October 4, 2024, Proceedings, Part XVI, Milan, Italy.

Luiten, J., Kopanas, G., Leibe, B., & Ramanan, D. (2024). Dynamic 3D gaussians: Tracking by persistent dynamic view synthesis. In *2024 International Conference on 3D Vision (3DV)*.

Mallick, S. S., Goel, R., Kerbl, B., Steinberger, M., Carrasco, F. V., & Torre, F. D. L. (2024). Taming 3DGS: High-Quality Radiance Fields with Limited Resources. In *SIGGRAPH Asia 2024 Conference Papers*. https://doi.org/10.1145/3680528.3687694

Marasinghe, R., Yigitcanlar, T., Mayere, S., Washington, T., & Limb, M. (2024). Computer vision applications for urban planning: A systematic review of opportunities and constraints. *Sustainable Cities and Society, 100*, Article 105047. https://doi.org/10.1016/j.scs.2023.105047

Mazzetto, S. (2024). A review of urban digital twins integration, challenges, and future directions in smart city development. *Sustainability, 16*(19), 8337. https://doi.org/10.3390/su16198337

Meinen, B. U., & Robinson, D. T. (2020). Mapping erosion and deposition in an agricultural landscape: Optimization of UAV image acquisition schemes for SfM-MVS. *Remote Sensing of Environment, 239*, Article 111666. https://doi.org/10.1016/j.rse.2020.111666

Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., & Geiger, A. (2019). Occupancy networks: Learning 3D Reconstruction in Function Space. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Murtiyoso, A., Lhenry, C., Landes, T., Grussenmeyer, P., & Alby, E. (2021). Semantic segmentation for building façade 3D point cloud from 2D orthophoto images using transfer learning. The international archives of the photogrammetry. *Remote Sensing and Spatial Information Sciences*.

Peng, Y., Li, W., Nijhuis, S., Yu, Y., & Wu, Z. (2026). Seeing heritage through green and blue: Assessing the visual influence of blue-Green infrastructure (BGI) in historic

urban areas (HUAs). *Environmental Impact Assessment Review, 118*, Article 108301. https://doi.org/10.1016/j.eiar.2025.108301

Peng, Y., & Nijhuis, S. (2021). A GIS-based algorithm for visual exposure computation: The west lake in Hangzhou (China) as example. *Journal of Digital Landscape Architecture, 6*, 424–435.

Peng, Y., Nijhuis, S., Geng, M., & Yu, Y. (2025). Enhancing visual attribute comprehension of urban heritage landscapes using combined GIS-based visual analysis methods: West Lake as a case study. *Environmental Impact Assessment Review, 115*, Article 108032. https://doi.org/10.1016/j.eiar.2025.108032

Peng, Y., Nijhuis, S., Zhang, G., Stoter, J. E., & Agugiaro, G. (2022). Towards a practical method for voxel-based visibility analysis with point cloud data for landscape architects: Jichang garden (Wuxi, China) as an example. *Journal of Digital Landscape Architecture, 7*(7), 682–691.

Peng, Y., Zhang, G., Nijhuis, S., Agugiaro, G., & Stoter, J. E. (2024). Towards a framework for point-cloud-based visual analysis of historic gardens: Jichang Garden as a case study. *Urban Forestry & Urban Greening, 91*, Article 128159. https://doi.org/10.1016/j.ufug.2023.128159

Poux, F. (2019). *The smart point cloud: Structuring 3D intelligent point data*. Belgium: Universite de Liege.

Poux, F., Neuville, R., Hallot, P., & Billen, R. (2016). Smart point cloud: Definition and remaining challenges. *IV-2*, 119. https://doi.org/10.5194/isprs-annals-IV-2-W1-1-2016

Pretzsch, H., Biber, P., Uhl, E., Dahlhausen, J., Rötzer, T., Caldentey, J., Koike, T., van Con, T., Chavanne, A., Seifert, T., Toit, B. D., Farnden, C., & Pauleit, S. (2015). Crown size and growing space requirement of common tree species in urban centres, parks, and forests. *Urban Forestry & Urban Greening, 14*(3), 466–479. https://doi.org/10.1016/j.ufug.2015.04.006

Qi, L., Hu, Y., Bu, R., Xiong, Z., Li, B., Zhang, C., Liu, H., & Li, C. (2024). Spatial-temporal patterns and influencing factors of the building green view index: A new approach for quantifying 3D urban greenery visibility. *Sustainable Cities and Society, 111*, Article 105518. https://doi.org/10.1016/j.scs.2024.105518

Qin, M., Li, W., Zhou, J., Wang, H., & Pfister, H. (2024). LangSplat: 3D language gaussian splatting. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.ieeecomputersociety.org/10.1109/CVPR52733.2024.01895.

Ren, C., Cai, M., Li, X., Shi, Y., & See, L. (2020). Developing a rapid method for 3-dimensional urban morphology extraction using open-source data. *Sustainable Cities and Society, 53*, Article 101962. https://doi.org/10.1016/j.scs.2019.101962

Rundle, A. G., Bader, M. D. M., Richards, C. A., Neckerman, K. M., & Teitler, J. O. (2011). Using google street view to audit neighborhood environments. *American Journal of Preventive Medicine, 40*(1), 94–100. https://doi.org/10.1016/j.amepre.2010.09.034

Salimi, S., Mirgholami, M., & Shakibamanesh, A. (2023). Visual impact assessment of urban developments around heritage landmarks using ULVIA method: (The case of Ark-e-Alishah monument in Tabriz). *Environment and Planning B: Urban Analytics and City Science, 50*(3), 678–693. https://doi.org/10.1177/23998083221123812

Sánchez, I. A. V., & Labib, S. M. (2024). Accessing eye-level greenness visibility from open-source street view images: A methodological development and implementation in multi-city and multi-country contexts. *Sustainable Cities and Society, 103*, Article 105262. https://doi.org/10.1016/j.scs.2024.105262

Schönberger, J. L., Zheng, E., Frahm, J.-M., & Pollefeys, M. (2016). *Pixelwise View Selection for Unstructured Multi-View Stereo*. Cham: Computer Vision – ECCV 2016.

Sevenich, J., Kemand, A., & Malkwitz, A. (2025). Comparison of mobile laser scanning systems for component-based scan-to-BIM evaluation: NavVis VLX2, XGRIDS L2 Pro, and XGRIDS K1. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-1/W6-2025*, 213–218. https://doi.org/10.5194/isprs-archives-XLVIII-1-W6-2025-213-2025

Shao, T., Qu, Y., & Du, J. (2022). A low-cost integrated sensor for measuring tree diameter at breast height (DBH). *Computers and Electronics in Agriculture, 199*, Article 107140. https://doi.org/10.1016/j.compag.2022.107140

Shushan, Y., Portugali, J., & Blumenfeld-Lieberthal, E. (2016). Using virtual reality environments to unveil the imageability of the city in homogenous and heterogeneous environments. *Computers, Environment and Urban Systems, 58*, 29–38. https://doi.org/10.1016/j.compenvurbsys.2016.02.008

Tancik, M., Casser, V., Yan, X., Pradhan, S., Mildenhall, B. P., Srinivasan, P., Barron, J. T., & Kretzschmar, H. (2022). Block-NeRF: Scalable large scene neural view synthesis. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.ieeecomputersociety.org/10.1109/CVPR52688.2022.00807.

Tancik, M., Casser, V., Yan, X., Pradhan, S., Mildenhall, B. P., Srinivasan, P., Barron, J. T., & Kretzschmar, H. (2022). Block-NeRF: Scalable large scene neural view synthesis. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Virtanen, J.-P., Jaalama, K., Puustinen, T., Julin, A., Hyyppä, J., & Hyyppä, H. (2021). Near real-time semantic view analysis of 3D city models in web browser. *ISPRS International Journal of Geo-Information, 10*(3), 138. https://doi.org/10.3390/ijgi10030138

Weil, C., Bibri, S. E., Longchamp, R., Golay, F., & Alahi, A. (2023). Urban digital twin challenges: A systematic review and perspectives for sustainable smart cities. *Sustainable Cities and Society, 99*, Article 104862. https://doi.org/10.1016/j.scs.2023.104862

Whelan, T., Goesele, M., Lovegrove, S. J., Straub, J., Green, S., Szeliski, R., Butterfield, S., Verma, S., Newcombe, R. A., & Goesele, M. (2018). Reconstructing scenes with mirror and glass surfaces. *ACM Trans. Graph., 37*(4), 102.

Wu, C., Ye, Y., Gao, F., & Ye, X. (2023). Using street view images to examine the association between human perceptions of locale and urban vitality in Shenzhen, China. *Sustainable Cities and Society, 88*, Article 104291. https://doi.org/10.1016/j.scs.2022.104291

Xia, H., Liu, Z., Efremochkina, M., Liu, X., & Lin, C. (2022). Study on city digital twin technologies for sustainable smart city design: A review and bibliometric analysis of geographic information system and building information modeling integration. *Sustainable Cities and Society, 84*, Article 104009. https://doi.org/10.1016/j.scs.2022.104009

Ye, M., Danelljan, M., Yu, F., & Ke, L. (2025). *Gaussian Grouping: Segment and Edit Anything in 3D Scenes*. Cham: Computer Vision – ECCV 2024.

Yu, L., Zhang, X., He, F., Liu, Y., & Wang, D. (2020). Participatory rural spatial planning based on a virtual globe-based 3D PGIS. *ISPRS International Journal of Geo-Information, 9*(12), 763. https://www.mdpi.com/2220-9964/9/12/763.

Yu, Y., Verbree, E., van Oosterom, P., & Pottgiesser, U. (2025). 3D gaussian splatting for modern architectural heritage: Integrating UAV-based data acquisition and advanced photorealistic 3D techniques. *AGILE: GIScience Series, 6*, 51. https://doi.org/10.5194/agile-giss-6-51-2025

Yu, Y., Wang, Z., van Oosterom, P., Pottgiesser, U., Verbree, E., & Peng, Y. (2026). Reframing the "H" in HBIM: From systematic review to UML-based conceptual modeling. *Frontiers of Architectural Research*. https://doi.org/10.1016/j.foar.2026.02.007

Zhang, C., Lin, G., Peng, Y., & Yu, Y. (2025). A workflow for urban heritage digitization: From UAV photogrammetry to immersive VR interaction with multi-layer evaluation. *Drones, 9*(10), 716. https://www.mdpi.com/2504-446X/9/10/716.

Zhang, G., Wang, Y., Wang, Y., & Peng, Y. (2025). Linking visual perception of urban greenery to resident preference in high-density residential areas through mobile point-cloud-based assessment. *Buildings, 15*(23), 4275. https://www.mdpi.com/2075-5309/15/23/4275.

Zhou, H., He, S., Cai, Y., Wang, M., & Su, S. (2019). Social inequalities in neighborhood visual walkability: Using street view imagery and deep learning technologies to facilitate healthy city planning. *Sustainable Cities and Society, 50*, Article 101605. https://doi.org/10.1016/j.scs.2019.101605

Zięba-Kulawik, K., Skoczylas, K., Wężyk, P., Teller, J., Mustafa, A., & Omrani, H. (2021). Monitoring of urban forests using 3D spatial indices based on LiDAR point clouds and voxel approach. *Urban Forestry & Urban Greening, 65*, Article 127324. https://doi.org/10.1016/j.ufug.2021.127324