Landscape Architecture and Sustainability 3 (2026) 100034

Contents lists available at ScienceDirect

Landscape Architecture and Sustainability

journal homepage: www.keaipublishing.com/LAS

Full Length Article

From comparison to combination: Street view imagery (SVI) and 3D model-based analyses for urban visual environment assessment

Yuyang Peng[a,*], Steffen Nijhuis[a], Zaichen Wu[a], Yingwen Yu[b]

[a] Department of Urbanism, Delft University of Technology, Delft 2628 CD, the Netherlands
[b] Department of Architectural Technology and Engineering, Delft University of Technology, Delft 2628 CD, the Netherlands

ARTICLE INFO

Keywords:
Urban landscape
Built environment
Street view imagery (SVI)
Semantic segmentation
Field-of-view (FOV)
Visual perception
Blue-green infrastructure

ABSTRACT

Street view imagery (SVI) is widely used in urban visual analysis and often treated as equivalent to eye-level perception. Yet its limitations and contextual applicability remain underexplored. This paper conducts a diagnostic viewpoint-level comparison of an image-based SVI pipeline and a 3D model-based field-of-view (FOV) method to clarify their respective weaknesses, strengths, and how they can be combined in practice (rather than treated as interchangeable or numerically fused). Using the West Lake ring road in Hangzhou as a case study, we analyze 2140 panoramas at 1075 viewpoints. The comparison shows systematic differences: SVI produces higher green shares (+0.16 on average), while FOV yields higher paved ground (+0.13) and building shares (+0.08). Sky differs little overall, water remains minor, and cross-method consistency varies by segment; SVI displays greater local variability linked to canopy occlusion and near-field heterogeneity. A small perception survey validates these findings. Terrain relief and building height were recognized more consistently in FOV, while vegetation and water abundance aligned more closely with SVI. Participants also judged overall ambience more easily from FOV's structural stability, even though SVI conveyed greater visual realism. These results reveal clear complementarities: FOV provides structure-aware metrics, SVI emphasizes appearance cues, and neither alone captures lived perception. On this basis, we propose a combination-oriented three-layer workflow, with perception as a required validation layer to support reliable applications in skyline and openness control, interface and character management, greenery maintenance, and equity assessment.

1. Introduction

Visual analysis of the built environment has long been one of the core topics in urban landscape studies (Liu and Nijhuis, 2020; Gobster et al., 2019). Such research contributes to evaluating spatial quality and providing empirical foundations for urban design, environmental psychology, and policy-making (Kyttä et al., 2011; Khan et al., 2014; van Kamp et al., 2003). It deepens our understanding of how urban spatial characteristics relate to human perception and daily experience, ultimately guiding more human-centered and aesthetically oriented urban interventions (Wang, 2023). Generally, research in this area follows two main pathways: one based on high-precision digital modeling (such as terrain surface models, 3D urban models) (Li et al., 2023; Cinnamon and Jahiu, 2021), and the other utilizing emerging data sources enabled by the "Internet Plus" era (Wu et al., 2024; Xu et al., 2021). Among these, street view images (SVI) have become a significant data source for visual studies due to their accessibility, massive quantity, wide coverage, and high visual fidelity (Gaw et al., 2022). With computer vision advances, SVI has emerged as a new form of data enabling urban visual analysis (Ito et al., 2025; Li et al., 2022).

While the strengths of this SVI-based approach have been widely recognized, its limitations and context-specific applicability have not been systematically examined (Biljecki et al., 2023). Furthermore, the relationship between SVI-based methods and the established digital modeling approaches remains underexplored. This paper addresses these gaps by quantitatively comparing SVI-based and digital model-based methods to clarify their respective limitations and complementarities and explore their integrated potential for urban visual analysis and management.

1.1. SVI-based visual analysis of the built environment

The large-scale application of SVI in urban visual analysis began in the early 2010s (Ito et al., 2025; Li et al., 2022). Its methodological core

Peer review under the responsibility of Editorial Office of *Landscape Architecture and Sustainability*.
* Corresponding author.
E-mail address: Y.Peng-1@tudelft.nl (Y. Peng).

https://doi.org/10.1016/j.las.2026.100034
Received 30 September 2025; Received in revised form 5 February 2026; Accepted 6 February 2026

lies in computer vision, particularly deep learning-based semantic segmentation, object detection, labeling and counting. These analyses often extend to statistical measurements and clustering for urban classification. Specific research applications include, but are not limited to, the measurement of urban green space exposure, streetscape greenery, urban color systems (Zhou et al., 2023) and spatial-visual characteristics (Liang et al., 2023). Numerous review papers have praised SVI for its precision, stability and wide applicability in urban studies (He and Li, 2021; Kang et al., 2020). Despite these advantages, several studies have also noted SVI's limitations. These include the limited frequency and spatial coverage of data collection (Rzotkiewicz et al., 2018), image quality inconsistencies in certain datasets (Hou et al., 2024; Fan et al., 2025), and challenges such as vegetation occlusion and terrain interference (Zhao et al., 2025), which may lead to discrepancies between semantic segmentation results and actual human perception.

While these issues have been individually addressed in the literature, there is still a lack of systematic inquiry into how such limitations affect the reliability and scope of SVI applications. Furthermore, strategies for mitigating or correcting these technical biases remain underdeveloped. Therefore, a deeper investigation into SVI's limitations and ways to optimize its performance is warranted.

1.2. Digital model-based visual analysis of the built environment

Compared to SVI-based approaches, digital model-based visual analysis has an earlier origin, gaining popularity with the widespread adoption of geographic information systems (GIS) (Inglis et al., 2022; Llobera, 2003). Common modeling techniques include digital elevation models (DEM), digital surface models (DSM), and digital landscape models (DLM) that incorporate vertical landscape elements such as vegetation and buildings (Guth et al., 2021; Orlof et al., 2025). In recent years, modeling tools like CityEngine, as well as photogrammetry and LiDAR-based point cloud techniques, have further advanced the realism and precision of urban models (Wang et al., 2018; Župan et al., 2023). Analytical methods in this category include viewshed analysis, field-of-view (FOV) calculations from human eye-level perspective, and line-of-sight (LoS) assessments (Peng et al., 2024; Peng and Nijhuis, 2021). These tools are widely used to evaluate skyline composition, high-rise building impacts, urban canyon openness, and solar exposure, among other applications (Puspitasari and Kwon, 2021; Oke, 1981; Gong et al., 2018).

Nevertheless, digital model-based methods also have clear limitations. Their abstraction cannot fully replicate the dynamic, real-life variability of urban environments, such as seasonal changes or ephemeral human activities (Eriksson and Harrie, 2021; Lei et al., 2024). Moreover, a trade-off often exists between model precision and computational or acquisition efficiency: highly accurate models demand substantial resources, while lower-resolution models are easier to manage but less reliable (Morgan et al., 2024; Schoorl et al., 2000). Despite these challenges, digital models maintain a strong advantage in geometric accuracy and spatial quantification. Thus, the integration of digital models and image-based data, particularly the increasingly dominant SVI data, presents a promising direction. Exploring how these two data forms can complement each other is a critical task for advancing urban visual analysis.

1.3. Research gaps and objectives

In summary, two primary research gaps can be identified: (1) Although SVI-based methods are becoming mainstream in urban visual analysis, their technical limitations, contextual applicability, and potential for improvement remain insufficiently clarified (Li et al., 2022); (2) The potential complementarity and combined use of digital modeling and SVI-based approaches have yet to be fully explored, especially regarding how digital models' spatial accuracy can support diagnostic interpretation and validation of image-based measures.

To address these gaps, this paper aims to: First, conduct a diagnostic, viewpoint-level comparison between SVI and 3D model-based FOV approaches (Ghasemian Sorboni et al., 2024), to evaluate their respective strengths and weaknesses across varied urban environments (Luo et al., 2025); Second, develop a combination-oriented framework for using the two methods together (rather than treating them as interchangeable or forcing numerical fusion), and propose a practical workflow for their combined use in urban visual analysis, with perception-based evidence serving as a required validation layer. To achieve these goals, the West Lake ring road in Hangzhou has been selected as a case study. This site includes a rich mix of urban interfaces, cultural landmarks, varied topography, and natural landscapes, making it an ideal testing ground for assessing the applicability and combined use of both SVI and digital modeling methods.

This study makes two main contributions. First, it clarifies the limitations and contextual applicability of widely used SVI-based urban visual metrics by conducting a diagnostic, viewpoint-level comparison with a 3D-model-based FOV approach, thereby showing when SVI can be treated as a reasonable proxy and when it systematically deviates. Second, based on the observed complementarities, it proposes an easy-to-adopt, combination-oriented workflow for practice, in which model-based FOV is used as a structural supplement to SVI, and perception evidence is used as a required validation layer, enabling more reliable interpretation and application without forcing numerical fusion.

2. Methods

To address the above research objectives, this paper follows a three-step methodological framework (Fig. 1). First, we collect urban geographic data and SVIs for the study area. Second, we conduct parallel analyses using a digital model-based approach and an SVI-based approach. Finally, we compare the results from both approaches to reveal their differences and applicability, thereby informing their potential combination in future urban visual analysis.

2.1. Study area and data

Hangzhou, a megacity with a permanent population exceeding ten million, is endowed with abundant urban green space resources. The ring road around West Lake traverses diverse urban landscape settings (Fig. 2), including mountainous sections, lakeside segments, waterfront-urban transition zones, and building-dominated urban interfaces, embodying a complex and representative interplay of natural and human-made elements (UNESCO). This continuous loop corridor not only facilitates directional segmentation and spatial comparison but also features notable elevation variation, extensive street-view image coverage, and access to open-source base geographic data, making it an ideal testbed for examining the applicability and complementarity of SVI- and digital model–based methods in heterogeneous environments.

To ensure comprehensive coverage of the study area, all distinct geographic coordinates located along the West Lake ring road were collected. Whenever possible, viewpoints captured between April and October were prioritized to minimize the influence of seasonal variation on tree canopy conditions. SVIs were obtained from Baidu Maps (https://map.baidu.com/) at approximately 5 m intervals along the corridor, each with embedded geolocation information (latitude and longitude recorded for every image). A total of 2140 images were collected: May 2014 (103), April 2015 (20), August 2017 (1020), September 2017 (822), and June 2020 (175). After removing duplicate viewpoints, 1075 unique locations were retained, giving priority to images captured in August and September 2017.

Geospatial data were obtained from two main sources: (1) official topographic survey data from the local government, including elevation points and 1 m contour lines (with elevation point density of approximately 100 points/ha) (Fig. 2b); and (2) open-source vector datasets from Baidu Maps, which include building footprints, an associated

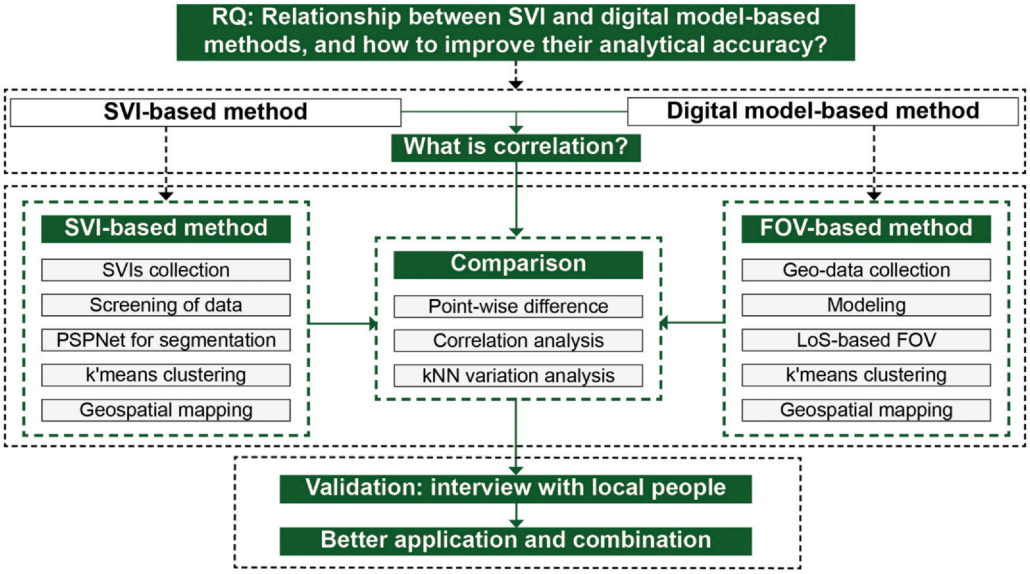

Fig. 1. Research framework for comparing SVI-based and digital model-based methods of urban visual analysis.

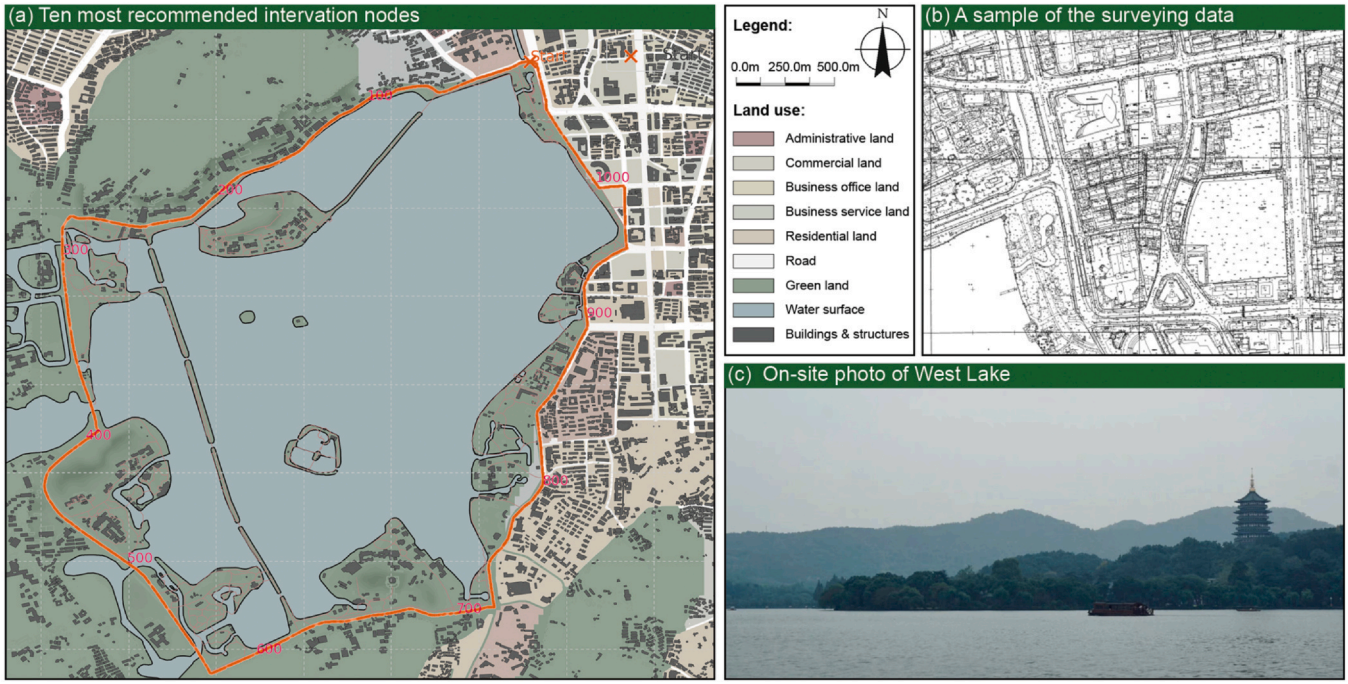

Fig. 2. Study area of the West Lake Ring Road in Hangzhou: (a) Locations of 10 recommended intervention nodes along the West Lake ring road, overlaid on land-use categories; (b) A sample of the detailed surveying data used for modeling; (c) On-site photo of West Lake, showing the landscape context of the study area.

building height attribute, road centerlines, land-use polygons, and water bodies (Fig. 2a). The Baidu-derived building height attribute was used for LOD-1 building extrusion in the 3D scene; because the platform does not provide survey-grade metadata on how this height attribute is derived (e.g., acquisition method and vertical accuracy), it was treated as an attributed height proxy for corridor-scale structural representation. For analysis purposes, these vector data were simplified into four unified categories: paved ground, unpaved ground, buildings, and water surfaces.

2.2. 3D model-based FOV analysis

We implement a field-of-view (FOV) analysis in a simplified 3D scene to quantify structural visual composition driven by terrain and built forms along the corridor. The 3D scene is intentionally a bare model. The elevation input represents ground, bare-earth topography and does not contain vegetation canopy geometry, and no independent vegetation 3D dataset is available to reconstruct trees or shrubs as occluders. Therefore, vegetation is not modeled as 3D geometry in the FOV simulation. Instead, we include an unpaved-ground layer to represent the plan-view green base, meaning ground cover in plan-view rather than seasonal canopy occlusion. This design matches the study goal of comparing structure-oriented 3D visibility with the canopy-sensitive SVI pipeline.

2.2.1. Modeling the built environment around the West Lake

All vector and elevation datasets were transformed to a common projected coordinate system and vertical datum. Basic topology issues, including duplicates, gaps, and slivers, were corrected. Terrain was generated by interpolating official elevation points and 1-m contour

lines into a digital terrain model. Local spikes and pits were smoothed, and the shoreline was regularized.

Buildings were modeled as LOD-1 extrusions using Baidu Maps building footprints and an associated building-height attribute. Because survey-grade metadata for this height attribute, including acquisition method and vertical accuracy, are not publicly available, it was treated as a height proxy suitable for corridor-scale structural representation rather than a field-validated absolute measure. Roads were polygonized from centerlines using available or default widths and assigned the local median terrain elevation. The lake was represented as a horizontal surface at the official water level. The unpaved-ground layer was derived by intersecting designated green land-use classes, including parks, green spaces, and forested areas, with the unpaved-surface category, and was used as the plan-view green baseline.

A quality-assurance pass ensured that building bases aligned with the terrain, that roads and water surfaces did not intersect building volumes, and that unpaved-ground polygons did not overlap paved classes. To quantify the vertical accuracy of the final terrain surface and avoid circular validation, we reserved 100 official elevation points as independent check points and excluded them from terrain interpolation. Check points were selected via stratified random sampling across low-relief and high-relief areas to ensure representative spatial coverage. Elevations were sampled from the final terrain surface at the check-point locations and compared with the surveyed elevations. We report mean bias, MAE, RMSE, and the 95th percentile absolute error as summary indicators of vertical error (Appendix A1).

2.2.2. Mapping the FOV

Viewpoints are placed at the SVI coordinates and elevated to 1.6 m. At each viewpoint, rays are cast across azimuth from 0° to 355° and elevation −80° to + 85° with 5° increments (72 azimuth angles by 34 elevation angles, 2448 rays per viewpoint), with a maximum range of 5000 m. The chosen elevation bounds balance coverage of near-field ground and upward openness while avoiding redundant sampling near the zenith. Each ray is labeled by the first intersected class, unpaved ground, paved ground, buildings or structures, or lake surface, or as sky when there is no hit within range. Ray counts are aggregated into per-viewpoint element shares, and summary statistics including mean, minimum, maximum, and standard deviation are reported.

Because the FOV module is computed directly in the angular domain using a fixed azimuth–elevation sampling scheme, it is not affected by the equirectangular projection bias that can influence pixel-based SVI estimates. We therefore perform a local sensitivity analysis on a representative subset of viewpoints by varying the angular step size (10°, 5°, 2° and 1°) relative to the default setting (5°), while holding all other parameters constant. The resulting element shares and the derived openness indicator remain qualitatively stable across step sizes, indicating that a 5° sampling resolution provides an acceptable accuracy-efficiency trade-off. In addition, to assess the potential bias introduced by viewpoint height, we recomputed the FOV at 20 representative locations using a 2.5 m viewpoint height to simulate the Baidu street-view camera and compared the resulting class shares with the baseline setting (Appendix A2). We further vary the maximum ray range and observe only marginal changes once the range reaches approximately 5000 m, confirming that $R = 5000$ m is sufficient to cover the modeled domain and that the results are not driven by range truncation.

2.2.3. Clustering and visualization

Each viewpoint is represented by a five-dimensional share vector, unpaved ground, paved ground, buildings or structures, water, and sky. We apply k-means clustering with $k = 3$ (chosen via the elbow method together with interpretability considerations). Euclidean distance is used, with a fixed random seed for reproducibility. Cluster labels are mapped back to viewpoint coordinates and visualized along the West Lake ring road. For each cluster, we report its mean share profile and its frequency, providing a compact summary of structural patterns for the cross-method comparison in Section 2.4. Because the FOV modality aims to summarize coarse structural typologies, k is selected within this modality and is not forced to match the SVI clustering. Cross-reading between FOV and SVI clusters is therefore based on their co-occurrence across viewpoints, rather than on one-to-one matching of cluster labels.

2.3. SVI-based analysis

This pipeline estimates the apparent visual composition from SVI, capturing vegetation canopy and other transient elements that are absent from the 3D model. Data acquisition, coordinate selection, and temporal prioritization follow Section 2.1; here we describe preprocessing, segmentation, and the computation of projection-corrected class shares (Figs. 3 and 4).

2.3.1. Segmentation of SVI

To reduce transient occlusions, each panorama is cleaned using semi-automatic foreground removal. People and vehicles above a small area threshold are masked and filled via mask-guided inpainting, followed by a light manual quality check to avoid over-editing. All panoramas are then normalized to a common equirectangular resolution to ensure consistent inference. Semantic segmentation is performed using PSPNet (ResNet-101 backbone, pretrained on ADE20K) under a standard inference setup. Raw labels are re-mapped into seven semantic categories: Vegetation, Unpaved ground, Paved ground, Buildings, Infrastructure, Water, and Sky. Because equirectangular panoramas do not represent equal solid angles per pixel, we compute per panorama class shares using a solid-angle-weighted pixel aggregation. For class c, the projection-corrected share is computed as:

$$s_c = \frac{\sum_{u,v} \cos(\phi(v))\, \mathbb{1}[y(u,v) = c]}{\sum_{u,v} \cos(\phi(v))}, \tag{1}$$

$$\phi(v) = \frac{\pi}{2} - \pi \frac{v + 0.5}{H}, \tag{2}$$

These projection-corrected shares are used as the primary metrics in subsequent analyses and visualizations. For transparency and sensitivity checks, unweighted pixel shares are also retained and compared against the corrected metrics in agreement diagnostics reported in the cross-method analysis (Appendix A3).

2.3.2. Analysis and visualization

To examine whether within-leaf-on season variability could affect the SVI-derived composition metrics, we stratified panoramas into two intra-seasonal groups based on acquisition month: Apr–Jun (early leaf-on) and Aug–Oct (late leaf-on). Because landscape composition along the West Lake corridor is spatially heterogeneous and sampling coverage may differ across months, we further conducted a location-aware sensitivity check. We then compared Apr–May (2014–2015), June (2020) and Aug–Sep (2017) within each sector to reduce confounding from spatial sampling imbalance. This stratification is used as a robustness check rather than as a causal attribution of temporal change (Appendix A4).

For pattern summarization, we adopt the same clustering and visualization procedure as in the FOV analysis (Section 2.2.3). Specifically, each viewpoint is represented by its seven-dimensional share vector, clustered by k-means with $k = 5$ (chosen via the elbow method together with interpretability considerations), and the resulting labels are joined back to viewpoint coordinates for color-coded maps along the West Lake ring road. For each cluster we report its mean share profile and its distribution across the ring road, enabling a cross-reading comparison with the FOV-based clusters in the cross-method analysis.

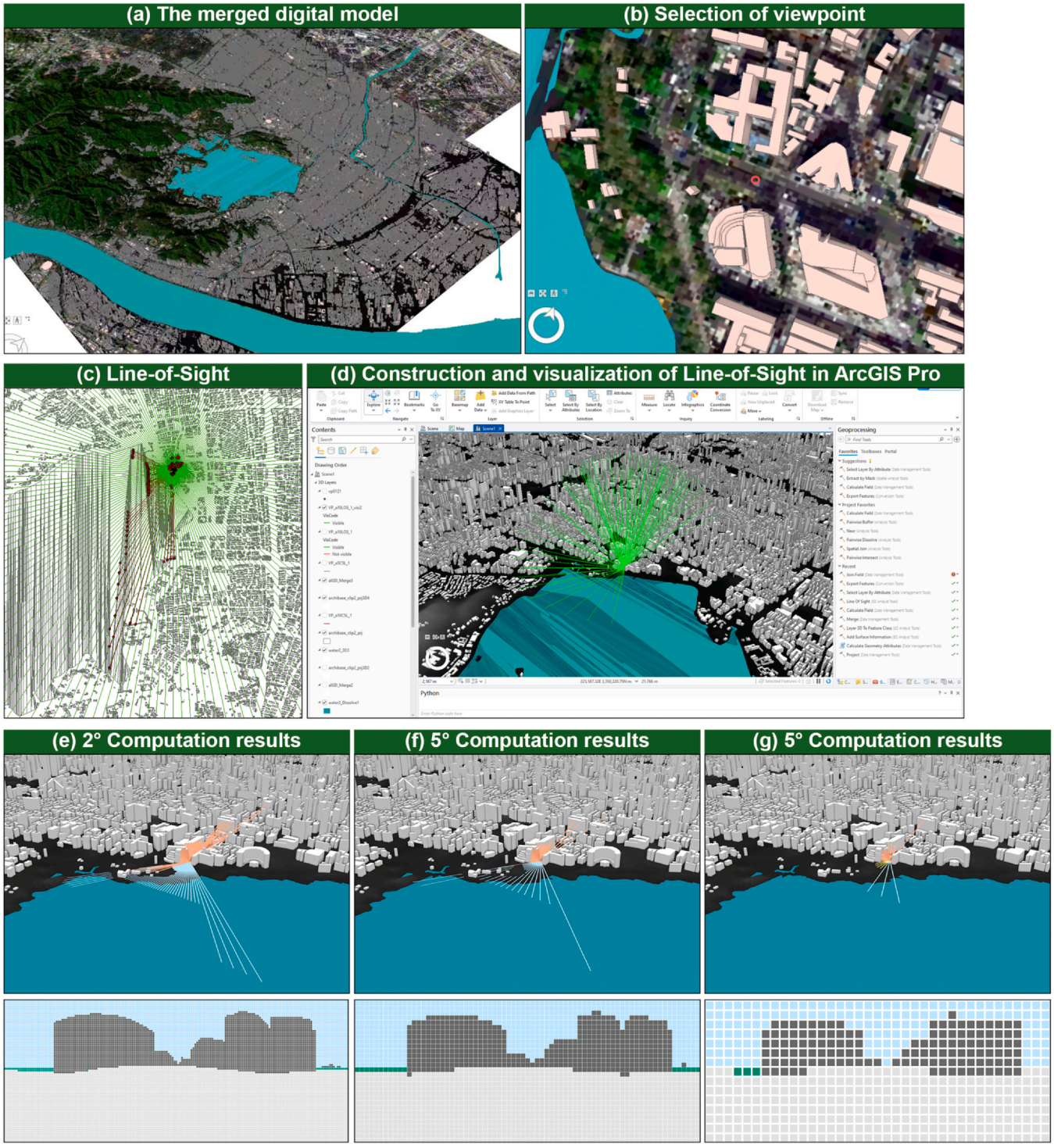

Fig. 3. Workflow of 3D model-based line-of-sight (LoS)/ field-of-view (FOV) analysis and sensitivity to angular sampling: (a) The merged digital model (terrain, lake surface, and buildings/structures); (b) Selection of analysis viewpoints aligned with street-level locations; (c) Radial LoS casting from each viewpoint to determine visibility (5000 m); (d) LoS construction and visualization in ArcGIS Pro; (e–g) Example FOV outputs under different angular step sizes, illustrating that coarser sampling yields a smoother, less detailed visible boundary, while finer sampling captures more local occlusion/visibility variations.

Here $k = 5$ is used to capture finer, appearance-oriented variation in SVI. Given the different goals and dimensionality of the two modalities, we do not enforce a unified k. Instead, the cross-reading is implemented by summarizing how SVI cluster labels co-occur with FOV cluster labels across the same viewpoints in the cross-method results.

We interpret SVI and FOV clusters as modality-specific typologies in light of the paired cross-method comparison, and provide the comparative interpretation in Section 3.5.

2.4. Cross-method comparison and validation

This section assesses the consistency between two independent pipelines applied to the same corridor viewpoints: a 3D model-based FOV analysis and an SVI-based image analysis. We use a paired, viewpoint-level design so that outputs from both pipelines can be compared element-by-element at the same location. Evidence is organized in two layers: (1) paired statistical consistency as the primary evidence, and

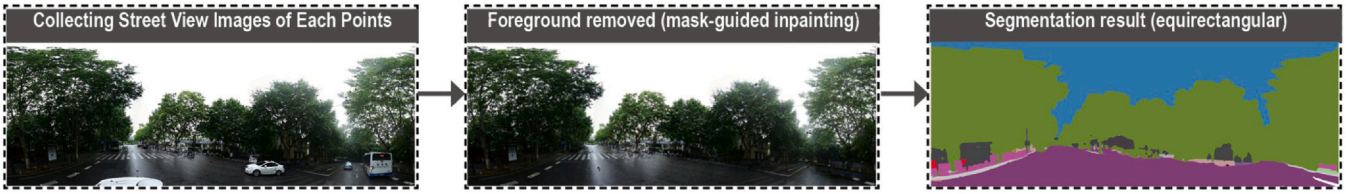

Fig. 4. The example of SVI data processing: from raw collected image to segmented data. (Many identified elements have been merged to seven types [e.g., roads and pathways, and other paved grounds have been merged into "paved ground"]).

(2) a small human perception check as supporting evidence to contextualize and explain the convergence or divergence between methods.

2.4.1. Paired cross-method consistency

For each matched viewpoint, we compare element shares across the two pipelines using a consistent set of pairings. Water and Paved ground are assessed one-to-one between FOV and SVI. We operationalize openness as the Sky proportion and compare Sky directly across methods as the paired openness metric. For green-related composition, we compare FOV Unpaved ground with three SVI specifications, namely SVI Unpaved ground, SVI Vegetation, and their combined share (SVI Vegetation + SVI Unpaved ground), so that both the planar green base (FOV) and the canopy-sensitive green signal (SVI) are represented in a comparable way. For the built environment, we compare FOV Buildings with SVI Buildings and SVI Infrastructure separately, and also with their combined share (SVI Buildings + SVI Infrastructure), to reflect differences in semantic partitioning between the two pipelines.

Cross-method association is quantified using Pearson's r and Spearman's ρ. Pearson's r summarizes linear agreement, while Spearman's ρ captures monotonic agreement that is less sensitive to non-normality in bounded share variables. Beyond corridor-wide summary statistics, we compute pointwise differences using a fixed sign convention (Δ = FOV − SVI) for each paired metric and map these differences in geographic space. The purpose of the difference maps is to reveal spatial patterns of both convergence and divergence, including segments where the two pipelines produce consistently similar estimates and segments where disagreement is systematically elevated. We then interpret these spatial patterns in relation to local scene conditions and candidate mechanisms identified from the corridor context, and we evaluate these mechanism hypotheses by examining whether different discrepancy metrics co-vary across space in the expected directions. This combination of corridor-level association, pointwise difference mapping, and cross-metric co-variation provides a coherent framework for diagnosing where the two pipelines agree, where they diverge, and what scene factors are most plausibly associated with the observed differences.

2.4.2. Human perception check

To provide a perception-oriented reference for interpreting the paired results, we conduct a small human assessment as supporting evidence rather than geometric ground truth. We recruit 10 local participants familiar with the corridor and present a set of geotagged scenes on an interactive map. Participants first confirm that they recognize each marked viewpoint. For each scene, they rate terrain undulation (flat, gentle, marked), vegetation amount (5-point scale), presence of water (yes, no, uncertain), building height (low, mid-rise, high-rise), and spatial openness (5-point scale from enclosed to open). Responses are summarized by item, and inter-rater consistency is reported to indicate reliability. These judgments are used to contextualize method-sensitive elements such as sky and openness, and boundary-sensitive classes such as unpaved ground versus vegetation, rather than to designate a single pipeline as "correct." Ethical approval for this study was obtained from the authors' affiliated research institution.

In addition, to provide a closed-loop, governance-oriented demonstration of marginal perceptual gain, we include three paired comparisons that functions as a modest quasi-intervention using seasonal canopy change. We select a representative segment where the FOV pipeline indicates high openness while the SVI pipeline indicates low openness in a leaf-on condition. At the same geotagged viewpoint and viewing direction, we obtain two street-view panoramas captured in September (leaf-on) and February (leaf-off), treating the February scene as a proxy for a light canopy-clearing or pruning intervention that primarily reduces near-field occlusion while leaving the structural setting unchanged. We then recompute SVI element shares for both dates under the same preprocessing and segmentation settings, and we recompute the paired openness metric (Sky share) together with the relevant class comparisons used in the corridor-wide analysis. Because the FOV module does not include explicit canopy geometry, its openness estimate at this site serves as a structural baseline expected to remain stable across the two dates, enabling a direct check of whether SVI and perceived openness move toward the FOV baseline when canopy occlusion is reduced.

We repeat the perception check for this site using a paired design. Participants evaluate both the September and February scenes for openness (and the same auxiliary items), with presentation order randomized, and we additionally record a direct paired judgment of which scene appears more open. We report the before–after change in computed shares and the corresponding change in perceived openness to demonstrate the direction and magnitude of marginal gains along the image, model, and perception chain. Finally, we integrate this site-level paired test with the corridor-level association results and the segment-level difference maps to form a coherent validation suite that documents where the two pipelines agree, where they diverge, and how occlusion-related mechanisms help explain those patterns.

3. Results

This section presents the outcomes of the two analytical pipelines. We first report the descriptive statistics and clustering of the SVI-based analysis (Fig. 5), followed by the FOV-based results (Fig. 6). We then compare the two methods element by element, including openness and local variability, and conclude with human validation. The structure moves from separate descriptions to cross-method comparisons and finally to alignment with on-site perception, thereby providing a complete picture of the consistency and complementarity between the two approaches.

3.1. SVI-based visual analysis

We applied k-means clustering to the per-viewpoint SVI class-share vectors, yielding five interpretable streetscape types (Fig. 5h). The separation is driven mainly by the trade-off among Vegetation, Sky (openness), Paved ground, and Buildings; Unpaved ground and Infrastructure remain minor in most clusters, while Water is generally small but helps distinguish lake-visible viewpoints (Fig. 5a–g).

Cluster-S1 high-vegetation low-sky, enclosed canopy (n = 259): Vegetation is highest (0.616) and Sky is near-zero (0.013), with very low buildings (0.009), indicating strongly occluded, semi-enclosed corridors

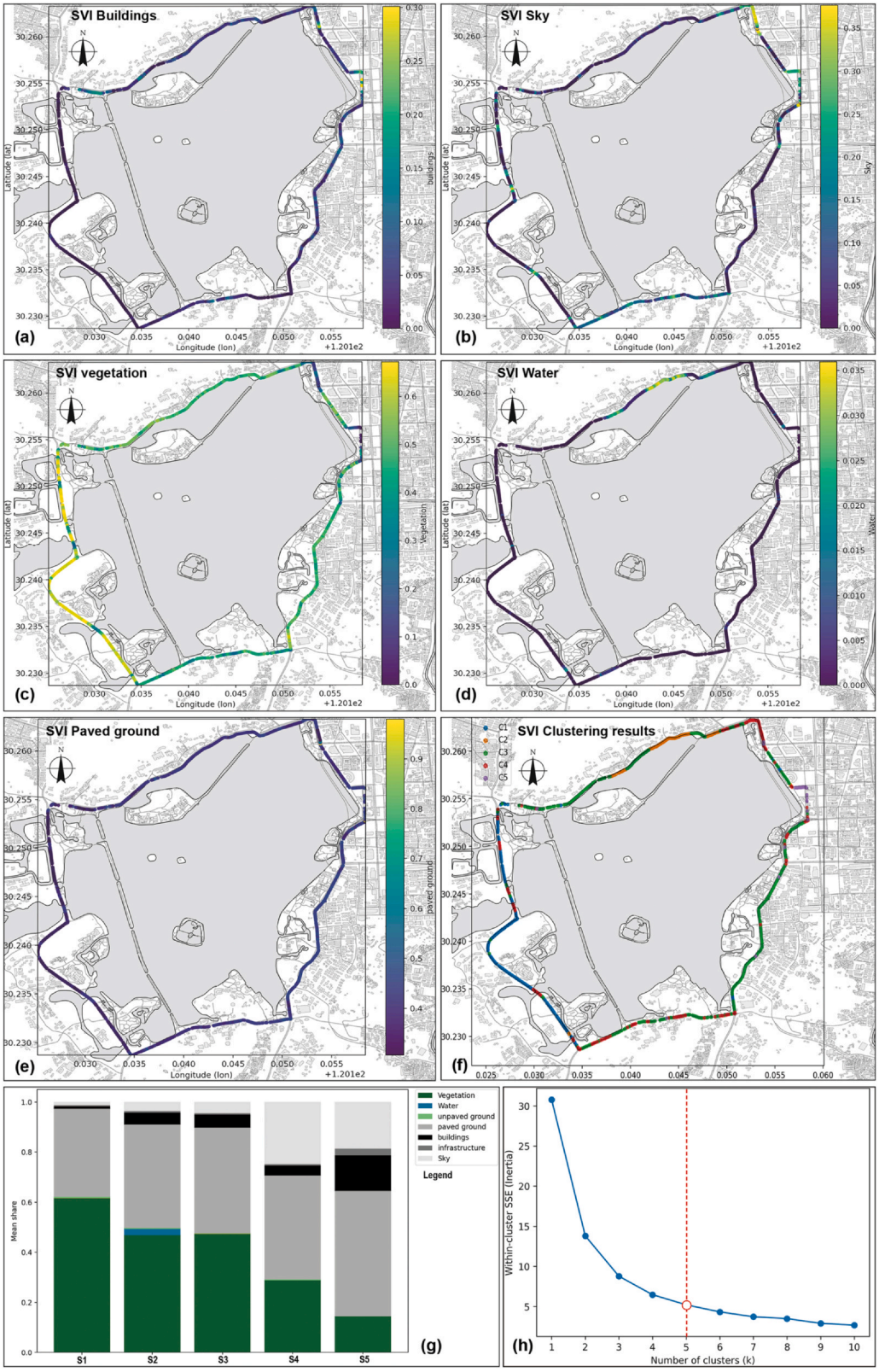

Fig. 5. Results of SVI-based visual analysis along the West Lake ring road: (a–f) Spatial distribution of class shares for Buildings, Infrastructure, Sky, Vegetation, Unpaved ground, Water, and Paved ground (k = 5); (g) Mean class composition of the clusters; (h) Elbow plot for k-means clustering.

dominated by canopy cover. Spatially, Cluster-S1 forms long, continuous runs along the west and southwest lakeside stretches of the loop.

Cluster-S2 vegetation plus paved ground with visible water (n = 94): Vegetation (0.468) and Paved ground (0.413) dominate, but Water is clearly higher (0.025), marking viewpoints where the lake surface becomes visually present. On the map, Cluster-S2 appears intermittently along shoreline exposures, most evidently on parts of the northern edge.

Cluster-S3 vegetation plus paved ground, water-trace-mixed (n = 504): The most common type, dominated by Vegetation (0.474)

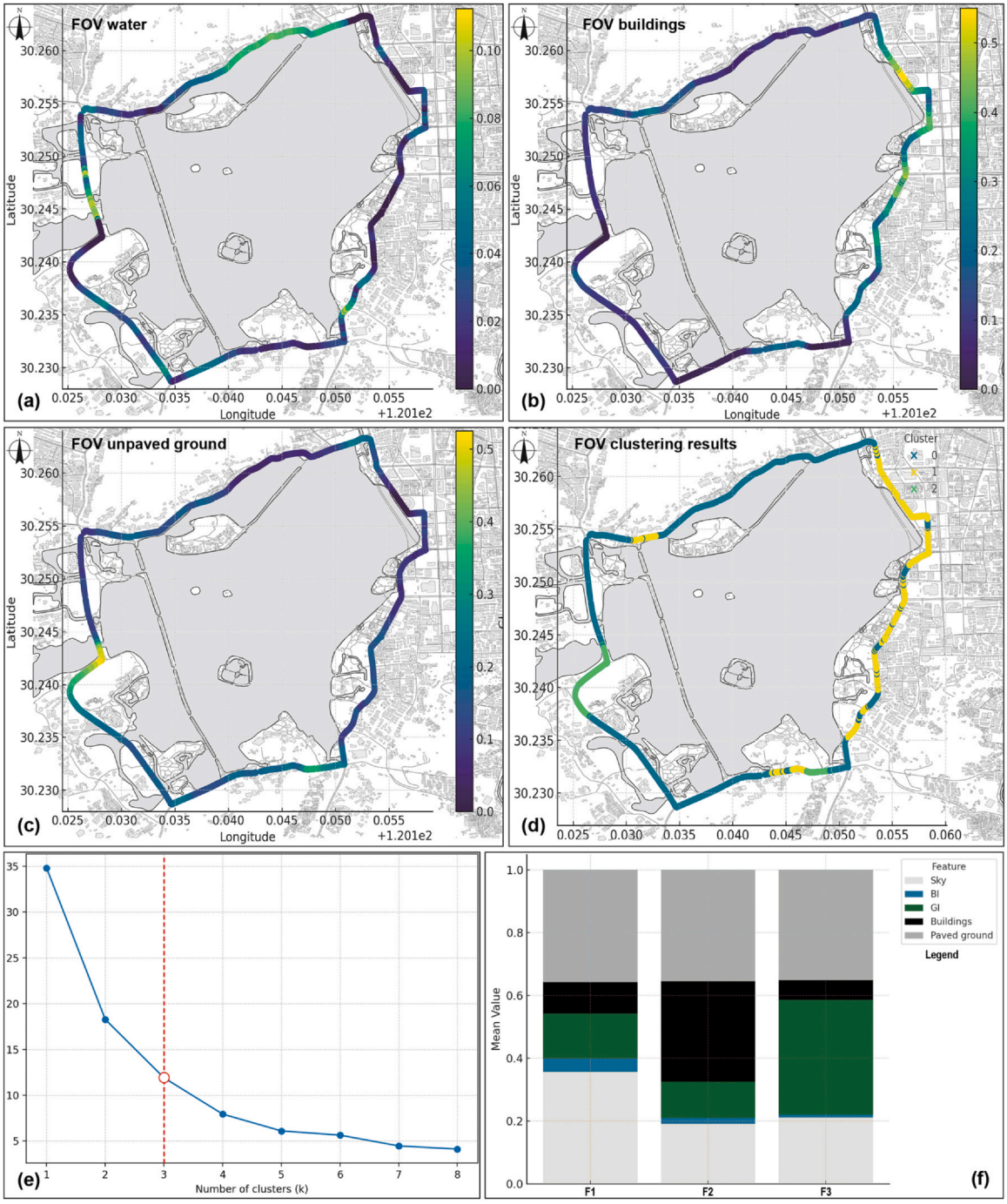

Fig. 6. Results of FOV-based visual analysis along the West Lake ring road: (a–d) Spatial distribution of class shares for Water, Unpaved ground, and Buildings; (e) Elbow plot for *k*-means clustering ($k = 3$); (f) Mean class composition of the clusters derived from FOV-based *k*-means classification.

and Paved ground (0.422), with Water effectively absent (0.001), representing the prevailing everyday composition around the loop. It is widely distributed around the ring road and constitutes the prevailing condition across extended east and south sections.

Cluster-S4 high-sky, open views with moderate vegetation ($n = 177$): Sky is highest (0.248) and Vegetation is moderate (0.289), with low Buildings (0.039), capturing more open scenes with reduced occlusion. It characterizes open lakeside and causeway-like stretches, with a notable concentration near the northeast arc where openness increases.

Cluster-S5 high-building high-paved, low-vegetation urban interface ($n = 41$): Paved ground peaks (0.501) and Buildings increase markedly (0.141), while Vegetation is lowest (0.145); it also shows a higher Infrastructure share (0.028), consistent with hard-edged, city-facing scenes. It is sparse overall and clusters primarily along the east to northeast urban frontage, with only isolated occurrences elsewhere.

3.2. FOV-based results

The model-based FOV captures the structural visual composition governed by terrain and building massing. Unlike SVI, it does not represent canopy occlusion, so the outputs emphasize geometry-driven openness (Sky) and the relative prominence of ground surfaces and built form. At the corridor scale ($n = 1075$), the mean class shares are: Paved ground 0.358, Sky 0.297, Unpaved ground 0.156, Buildings 0.158, and Water 0.033 (Fig. 6a–c). *k*-means clustering ($k = 3$) on the

Table 1
Paired cross-method comparisons between FOV and SVI class shares (correlation and error metrics).

Pair	Pearson r	Spearman	Mean (FOV − SVI)	MAE
FOV Sky vs SVI Sky	−0.017040	0.087754	0.225149	0.238759
FOV Paved ground vs SVI Paved ground	0.067008	0.105299	−0.049420	0.055054
FOV Water vs SVI Water	0.492673	0.452914	0.029992	0.030331
FOV Unpaved ground vs SVI Unpaved ground	−0.059940	−0.284930	0.154621	0.154800
FOV Unpaved ground vs SVI Vegetation	0.152081	0.101593	−0.311510	0.317686
FOV Unpaved ground vs SVI (Vegetation + Unpaved ground)	0.149126	0.095339	−0.314190	0.320056
FOV Buildings vs SVI Buildings	0.385096	0.455206	0.115423	0.119101
FOV Buildings vs SVI Infrastructure	0.266647	0.250793	0.148931	0.149165
FOV Buildings vs SVI (Buildings + Infrastructure)	0.399944	0.474326	0.108466	0.113633

five FOV share vectors yields three interpretable structural types, and the map projection of cluster labels (Fig. 5) shows a clear spatial organization around the ring road (Fig. 6d).

Cluster-F1 (open views, $n = 666$) is characterized by the highest Sky share ($M = 0.358$, $SD = 0.055$) with modest Buildings ($M = 0.099$, $SD = 0.048$) and moderate Unpaved ground ($M = 0.142$, $SD = 0.051$); spatially, it forms long, continuous runs along the more open arcs of the loop, consistent with lake-edge and causeway-like sightlines.

Cluster-F2 (urban interface, $n = 304$) is dominated by Buildings ($M = 0.319$, $SD = 0.093$) and has the lowest Sky ($M = 0.192$, $SD = 0.055$), with limited Unpaved ground ($M = 0.116$, $SD = 0.053$); on the map it concentrates along the more urbanized frontage of the ring road, producing a structurally city-facing profile.

Cluster-F3 (green-structured corridors, $n = 105$) shows the highest Unpaved ground share ($M = 0.362$, $SD = 0.086$) with small Buildings ($M = 0.064$, $SD = 0.065$) and mid-range Sky ($M = 0.213$, $SD = 0.047$); mapped locations appear in more park- and foothill-adjacent segments, where terrain and green land cover dominate the view.

Across the three types, Paved ground remains comparatively stable ($M = 0.352$–0.358), so the separation is driven mainly by the trade-off among Sky, Buildings, and Unpaved ground, while Water acts as a secondary marker of lake visibility. In addition, the influence of viewpoint height is also evident. Under a 2.5 m viewpoint height (simulating the street-view vehicle), the estimated Paved ground share decreases, while the Sky share increases markedly. The Unpaved ground and Water shares show slight increases, whereas changes in the Buildings share are mixed and location-dependent.

3.3. Cross-method comparison and spatial diagnosis

3.3.1. Paired comparisons under the cross-walk

We compared paired class shares under the intended cross-walk between the two pipelines, with openness operationalized as the Sky proportion. Cross-method association is summarized using Pearson's r and Spearman's ρ, and systematic offsets are reported using the signed mean difference and MAE (Table 1, Fig. 7).

Water exhibits the strongest agreement. FOV Water and SVI Water show the highest correlation (Pearson $r \approx 0.49$), and the average difference is small (mean $\Delta \approx 0.03$), indicating that both pipelines capture lake visibility in a broadly comparable way at matched viewpoints. Built form shows moderate correspondence. FOV Buildings versus SVI Buildings reaches a mid-range correlation (Pearson $r \approx 0.39$), and FOV tends to be higher on average. Importantly, agreement improves slightly when SVI Buildings and Infrastructure are combined into an aggregated built indicator, suggesting that the FOV Buildings metric aligns more closely with total built presence than with a single SVI built subclass.

In contrast, green-related measures and openness show weak pointwise correspondence and clear systematic offsets. For green, FOV Unpaved ground is weakly related to SVI green whether SVI is represented as Unpaved, Vegetation, or their sum, and the direction of bias depends on which SVI definition is used. In particular, relative to SVI (Vegetation + Unpaved), FOV Unpaved is consistently lower (mean $\Delta \approx -0.31$), reflecting that SVI Vegetation is strongly influenced by near-field canopy occlusion and image masking, while FOV Unpaved represents a planar ground-based component and intentionally excludes canopy geometry. For openness, FOV Sky and SVI Sky show near-zero association at the viewpoint level, while FOV reports substantially higher sky shares on average (mean $\Delta \approx +0.23$), consistent with canopy occlusion suppressing visible sky in SVI but not in the structural FOV baseline. Paved ground shows comparatively small systematic offsets (mean $\Delta \approx -0.05$), but remains weakly correlated at the point level, consistent with perspective and occlusion differences between pixel-based and ray-based measurements.

Overall, the corridor-wide comparison indicates that the two pipelines are most comparable for water visibility and for built form when SVI built subclasses are aggregated, while they diverge for green and sky/openness due to the central role of canopy occlusion in SVI and its intentional absence from the FOV representation.

3.3.2. Spatial patterns of agreement and disagreement

We compute per-viewpoint differences using a fixed sign convention (Δ = FOV − SVI) and map these values onto geographic coordinates. The maps reveal spatially coherent patterns of both convergence and divergence along the ring road (Fig. 8). Higher agreement tends to occur in two settings: building-dominated interfaces with relatively small street trees, and terrain-dominated natural segments. By contrast, where street trees are large, disagreement increases even when the surrounding setting is otherwise open, building-dense, or lake-open, indicating that near-field canopy and interface occlusion can dominate the SVI composition and weaken alignment with the geometry-driven FOV baseline.

Element-wise difference maps clarify how these patterns arise. Water discrepancies are concentrated where water is potentially visible and can be intermittently blocked by vegetation, whereas segments with no water visibility show consistently small differences. Built-related discrepancies are larger in urban segments and are amplified where canopy occlusion suppresses visible building surfaces in SVI. Green-related and sky-related discrepancies closely track the spatial distribution of vegetation, with larger mismatches in segments with denser street trees.

3.3.3. Mechanism check: vegetation occlusion

We then test the vegetation-occlusion mechanism quantitatively by relating discrepancy metrics to vegetation. Sky mismatch increases with SVI vegetation share, showing a clear positive rank association (Spearman $\rho \approx 0.5$). This relationship is not only visible in the maps but also persists under multivariate controls that account for structural and contextual composition, indicating that vegetation remains a meaningful explanatory factor beyond co-occurring scene features. A similar vegetation-linked structure is reflected in the overall mismatch index that aggregates absolute differences across unpaved, buildings,

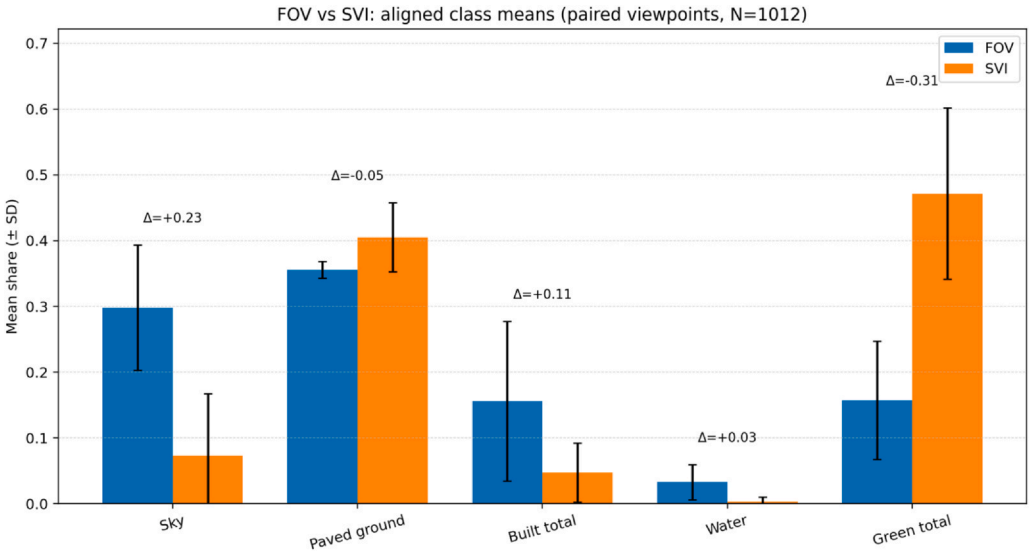

Fig. 7. Mean class-share comparison between FOV and SVI ($M \pm SD$). Note: To reduce redundancy and potential bias from near-duplicate viewpoints, locations (n = 63) that were too close (within 5-ms) to each other were removed (spatial thinning), so that highly similar scenes would not disproportionately influence the comparison.

sky, and water. Together, these results support canopy occlusion as a primary mechanism shaping where and why the geometry-based FOV baseline diverges from image-based SVI composition.

3.4. Human validation

3.4.1. General validation across methods

This section provides perception-based validation for the two approaches using 10 participants and 10 geotagged viewpoints. Across interviews and scene-based judgments, participants' responses show that site character can diverge from what image segmentation alone suggests under strong canopy occlusion. In several viewpoints, street trees partially blocked building facades, and most participants still described the place as city dominated or mixed, even when SVI segmentation indicated a large green share. Conversely, some viewpoints that appeared relatively urban in computed outputs were recalled as clearly green dominated. These contrasts indicate that near-field occlusion can depress the apparent building share in SVI and, in turn, that segmentation does not always represent overall urban character at the site scale.

Participants nevertheless responded consistently to different cue types. They reliably distinguished hilly versus flat terrain and low-rise versus high-rise contexts, signals that are preserved more explicitly in the 3D model-based FOV outputs and support cross-scene comparability. At the same time, perceived vegetation tended to follow SVI results more closely, and water was reported reliably only at lakeside viewpoints, with higher uncertainty elsewhere. Overall, these responses suggest complementary sensitivities: FOV better retains structural context for comparing scenes, while SVI better captures immediate, near-field components, especially canopy and water, that shape what people notice in a single view (e.g., Fig. 9).

3.4.2. Seasonal validation of perceived openness

To specifically validate seasonal effects, we selected five locations that share a consistent conflict condition: all exhibit high structural openness by FOV, yet SVI indicates lower openness because vegetation occupies a large share of the view (that is strong near-field canopy occlusion). In September (leaf-on), this pattern is visually apparent in the segmentation masks: canopy pixels dominate and suppress visible sky and water components, yielding an SVI openness estimate that is lower than the FOV-based structural openness at the same locations (Fig. 9).

In February (leaf-off), the SVI signal shifts toward the structural baseline captured by FOV. Segmentation results show a redistribution away from canopy and toward increased visibility of sky and water, and SVI-derived openness becomes closer to the corresponding FOV values. Crucially, the perception check mirrors this change: participants consistently reported that sky and water visibility increased noticeably in the February scenes. Taken together, the five sites provide converging evidence that, under broadly stable corridor geometry, seasonal canopy conditions materially modulate experienced openness, and that SVI captures this perceptually salient shift through transparent semantic components (sky-water versus vegetation).

3.5. Interpreting modality-specific clusters and their cross-reading value

Building on the paired cross-method comparison and the perception evidence, we interpret FOV and SVI clusters as modality-specific typologies rather than labels that should match one-to-one. FOV clusters summarize structure-oriented regimes driven by terrain and built form, and they therefore provide a relatively stable baseline segmentation. SVI clusters summarize appearance-oriented regimes that are more sensitive to canopy occlusion, near-field objects, and acquisition conditions, and they capture fine-grained local variability. The cross-reading value of clustering lies in using one typology to explain variation inside the other, and in identifying one-to-many or many-to-one relationships that point to different sources of disagreement.

Concretely, the clusters can be used for cross-checking at the viewpoint level. For example, when a viewpoint is assigned to FOV Cluster-1 (open views) but to SVI Cluster-S1 (high-vegetation low-sky, enclosed canopy), the cross-reading indicates that the underlying structure is relatively open, yet the panorama is locally occluded by canopy and near-field vegetation. In this case, the SVI-derived sky and openness signal at that viewpoint is likely to be underestimated, while the vegetation share may be inflated, so the viewpoint should be flagged as a caution case for interpretation. In practice, this viewpoint-level cross-reading provides an interpretable basis for flagging caution segments and prioritizing targeted perception sampling when cross-method disagreement implies elevated interpretation risk.

4. Discussions

4.1. Core findings and mechanisms

This paper showcases several essential findings, including: (1) Relative to the model-based FOV, the image-based SVI reports systematically higher green shares and lower building and sky shares. Formally, $\Delta_{\text{total green}} = \text{SVI} - \text{FOV} > 0$, $\Delta_{\text{building}} < 0$, $\Delta_{\text{sky}} < 0$. Off

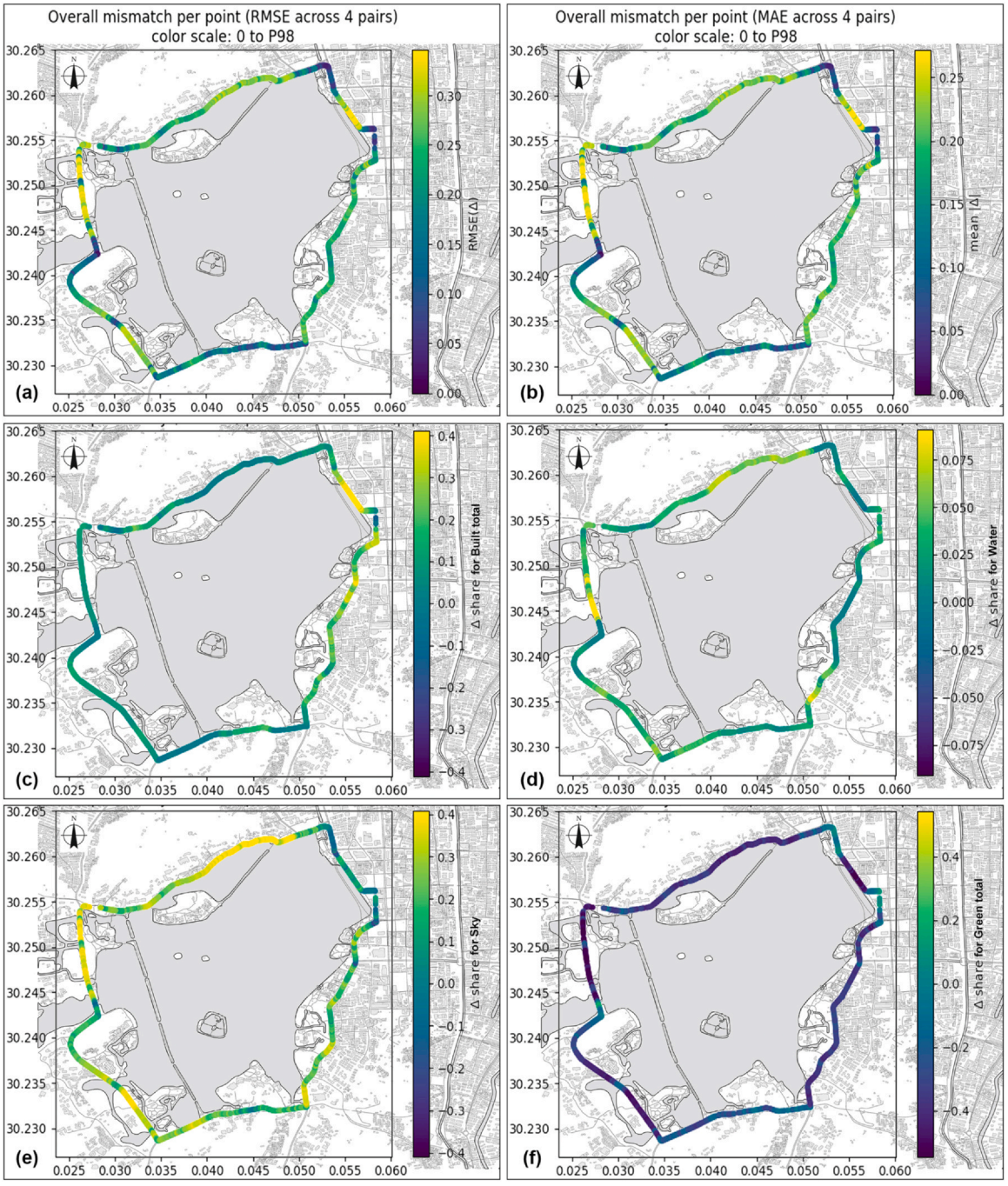

Fig. 8. Spatial distribution of FOV–SVI discrepancies along the West Lake ring road: overall mismatch and class-wise share differences: (a) Overall mismatch per viewpoint measured by RMSE across the four matched pairs (color scale 0–P98); (b) Overall mismatch per viewpoint measured by MAE across the four matched pairs (color scale 0–P98); (c) Difference in built share; (d) Difference in water share; (e) Difference in sky share; (f) Difference in green share.

lakeside segments, water is often higher in FOV than in SVI. The sign of these gaps is stable across scenes, magnitudes increase with canopy occlusion, and the eastern corridor shows the largest differences. SVI also exhibits higher local variability at the segment scale. (2) Agreement with lived perception is complementary rather than uniform. Structural cues that underpin scene structure, including terrain relief, building height, and geometric openness, align more closely with FOV. Near-field cues, especially vegetation and open water, align more closely with SVI.

The mechanisms behind these discrepancies fall into three groups. First, imaging and data geometry: street-view cameras sit higher than a 1.6 m eye height and near the carriageway centerline, which narrows the visible ground band and makes sky and facades sensitive to small pitch and roll; panorama stitching and equirectangular projection distort the top and bottom of the frame and alter pixel proportions near the edges; dense canopy and seasonal foliage hide facades and distant terrain and inflate green while deflating building and sky in SVI; the FOV model omits canopy geometry, which suppresses seasonal noise but lowers estimated greenness. These effects are strongest in the eastern corridor and coincide with low agreement and hotspot locations. Second, algorithms and semantics: a pretrained segmentation model without local fine-tuning can drift on local tree colors, facade materials, and sharp sky edges, which raises green and lowers building and sky; the two pipelines encode green infrastructure differently, with

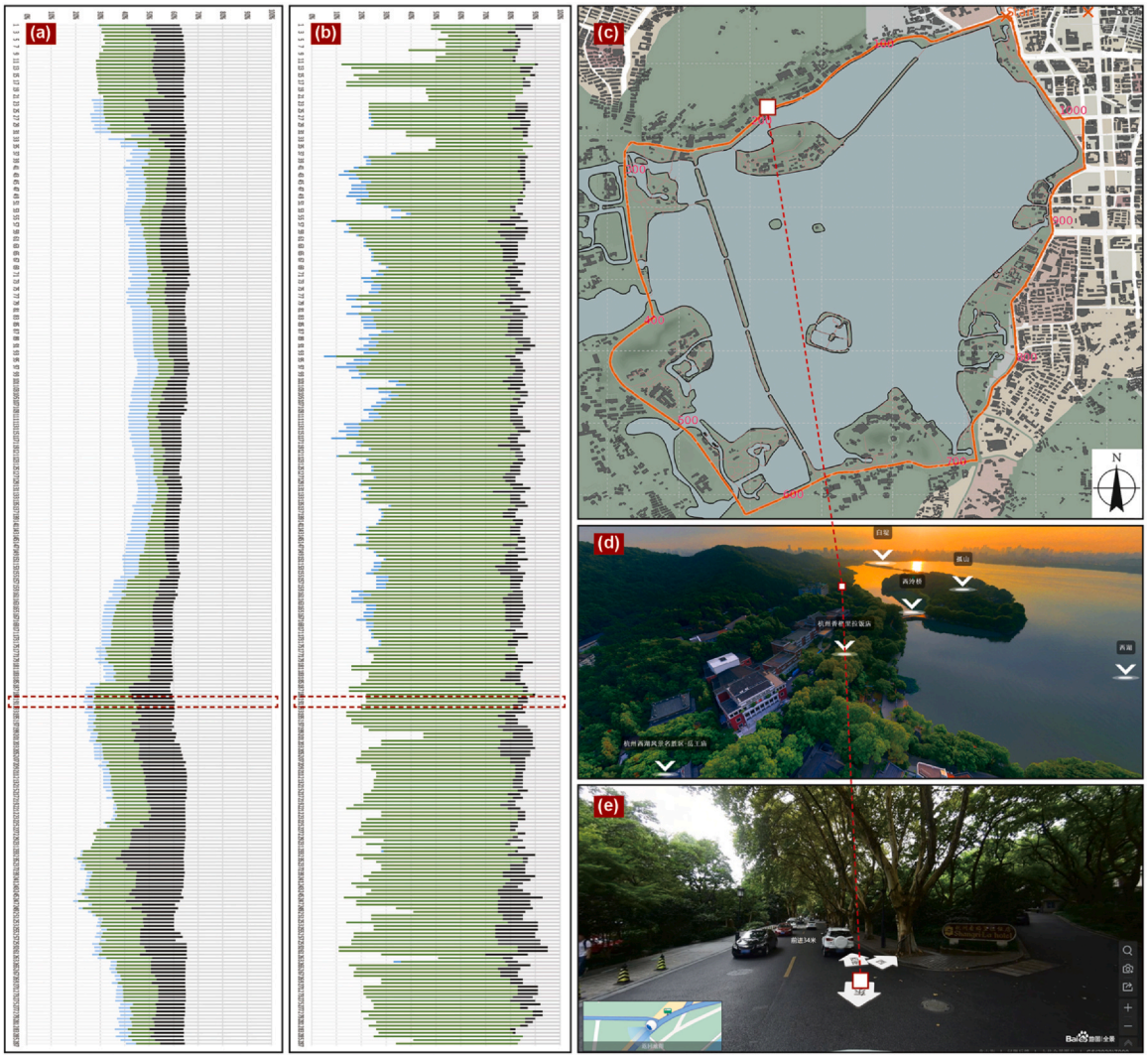

Fig. 9. One example of perception sampling and comparison between FOV- and SVI-based results in the North segment of the West Lake ring road: (a) FOV-based composition profile for the North segment; (b) SVI-based composition profile for the North segment; (c) Location of the sampled perception experiment point; (d) Aerial view of the sampled site, showing surrounding terrain and water bodies; (e) SVI of the same site, illustrating the tree canopy and occlusion conditions. (This bar chart is not intended to suggest that FOV and SVI are interchangeable or that one can replace the other. Rather, it highlights their systematic inconsistencies and complementary nature, reflecting that they capture different aspects of the visual environment).

FOV using a planar green base within green land use and SVI combining vegetation with unpaved ground, which amplifies systematic green gaps; foreground removal and small-component filtering can trim roads and facades and slightly underestimate paved or building shares. Third, human perception: observers use multi-viewpoint sampling and motion parallax to see past near-field occluders and infer terrain and height from the horizon, skyline breaks, and landmarks; cross-modal cues such as sound, airflow, humidity, smell, and prior knowledge help identify lakeside scenes as watery and open, while non-lakeside scenes elicit more uncertainty; recall bias and anchoring around strong landmarks increase dispersion where distinctiveness is weak. These perceptual factors explain why SVI shows higher local variability in general and why its advantage fades in building-dense or strongly undulating corridors where canopy hides height and relief in the images while FOV preserves the structural signal.

In sum, FOV provides stable structure-aware priors for terrain, height, and geometric openness, while SVI supplies high-fidelity near-field appearance cues for canopy and water. Each method has condition-dependent bias when used alone. Used together they balance structure and appearance and align more closely with interview-based perception.

4.2. SVI-based pipeline

SVI is an analytic approach to the built environment that leverages large, readily available street-level imagery and offers a pedestrian-eye viewpoint, but it is not a direct proxy for lived perception (Lu et al., 2025). Panoramic projection and non-human camera heights distort the angular proportions of sky and ground, and dense canopy can occlude façades and undulating terrain. In our corridor, this combination reduces the sensitivity of the SVI pipeline to structural cues such as terrain relief and building height at some urban interfaces, even where interviewees consistently recognized those differences on site. At the same time, it tends to amplify near-field canopy signals relative to paved ground and façades. These mechanisms help explain the higher local variability we observe in SVI and the element-wise gaps that cluster at canopy-sky disagreement hotspots (Alpherts et al., 2025).

A further methodological caveat concerns how class shares are computed on equirectangular panoramas. Simple pixel counting over-weights high-zenith zones and under-weights the equatorial band, which can inflate sky and canopy unless angular distortion is corrected (Liang et al., 2017). Work on hemispherical estimation shows why equal-solid-angle sampling or cosine weighting is needed to obtain

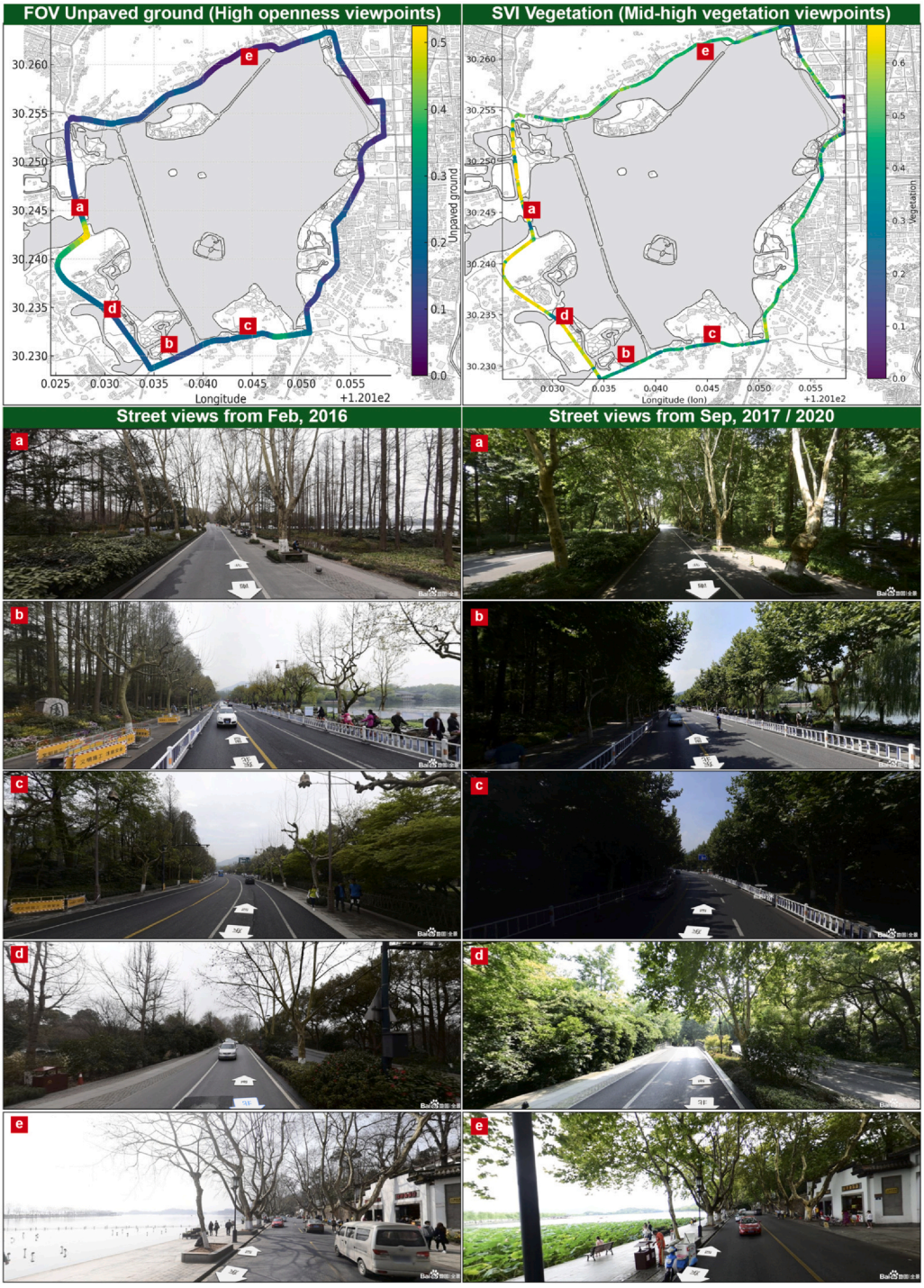

Fig. 10. Seasonal change in streetscape vegetation and openness, illustrated with paired street-view examples and viewpoint selection context: (Left) FOV Unpaved ground share along the ring road, shown to contextualize and select high-openness viewpoints (locations a–e); (Right) SVI vegetation share along the ring road, shown to contextualize and select mid-to-high vegetation viewpoints (locations a–e); (a–e) Paired street-view images at the selected locations a–e, comparing a leaf-off/low-canopy season (Feb 2016) with leaf-on conditions (Sep 2017/2020). The examples highlight seasonal canopy growth that increases image-based vegetation visibility and reduces perceived openness at the same viewpoints.

projection-aware shares (Liang et al., 2017; ANON, 2001). Together with platform heterogeneity and mixed vintages, these factors mean that SVI segmentation is strongest for near-field vegetation and streetscape details, but it is less reliable for structural cues unless calibrated against a geometric baseline (Kim and Jang, 2023). Even so, SVI's advantages are substantial: public platforms expose massive archives at negligible marginal cost, modern deep models enable high-throughput extraction of greenery, blue-green exposure and façade features, and street-level greenery proxies such as the green view index (GVI) have proven both scalable and perception-adjacent in planning and environmental applications (Li et al., 2015).

These properties situate SVI for specific kinds of work. It is well suited to rapid, city-scale audits of blue-green exposure; screening and prioritization (e.g., locating canopy gaps or identifying streetscape types); and before–after checks where acquisition dates are known and stable. In practice, projection-aware sampling (equal-solid-angle/cosine weighting), explicit documentation of provider and vintage, and harmonized label taxonomies materially improve robustness, while cross-method calibration, such as the element-wise and openness comparisons we report, helps localize uncertainty and supports targeted field checks.

4.3. 3D model-based FOV pipeline

Model-based FOV analysis trades some photorealism for geometric control and reproducibility. Once the scene and parameters are fixed

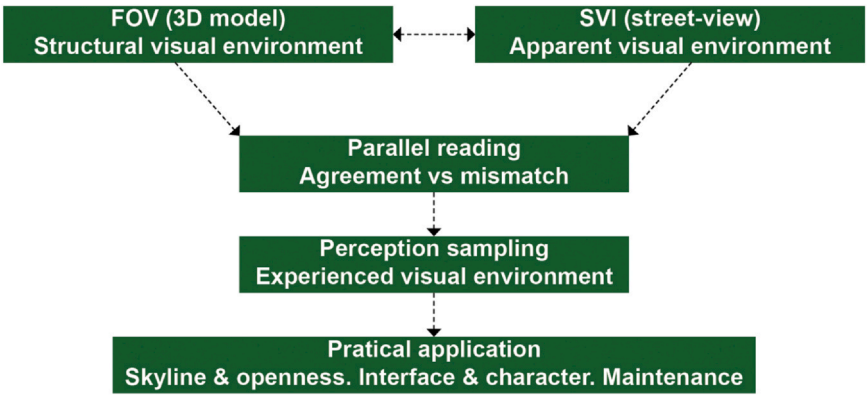

Fig. 11. The unified workflow: The integration of SVI-based analysis, 3D model-based analysis and perception-based tools for better understanding of urban visual environment.

(eye height, ray density, range), the method provides parameter-traceable outputs that are highly sensitive to structural cues, elevation information, that can be masked in images by canopy or perspective effects. In our corridor the FOV pipeline preserved terrain and building-height signals more faithfully than SVI in several segments and yielded stable local-variability profiles where the corridor's structure is regular. These patterns align with the human-validation results and with long-standing visibility scholarship on geometry-driven visual metrics. At the same time, classic studies show that DEM/DSM-based viewsheds are sensitive to elevation errors, resolution, observer height and algorithmic choices (Fisher, 1993), and that different implementations can produce measurable differences for the same site.

Scene realism is a design choice with cost implications. A simplified scene that omits vegetation, used here for comparison, cannot reproduce seasonal canopy or other short-lived elements, which helps explain systematic offsets in the "green infrastructure" component between methods. By contrast, enriching the scene with LiDAR or photogrammetry reduces occlusion artefacts and improves façade and canopy realism, yet it increases data-collection, modelling, and compute costs that are often prohibitive at corridor-to-city scale (Parent and Lei-Parent, 2023; Orlof et al., 2024). Furthermore, FOV carries little information about transients: lighting, traffic, human presence, unless they are explicitly modelled or coupled to other models.

Within those bounds, the FOV pipeline is the natural choice when structural fidelity and scenario control are paramount. Typical applications include view-corridor and skyline protection, and "what-if" analyses that remove or add canopy, alter height caps or vary setbacks. Digital scenes also integrate readily with other 3D analytics, overshadowing and solar/glare studies, acoustic and radio line-of-sight, where photographic appearance is secondary to geometry (Llobera, 2003; Chamberlain and Meitner, 2013). Practically, a resolution ladder is effective: screen large areas with DSM/DTM for speed, then escalate to LiDAR/photogrammetry only in disagreement hotspots identified by cross-method maps, thereby aligning modeling accuracy with decision needs.

4.4. Framework for combined use and practical applications

This study moves integration beyond application-level method comparison by treating model-based FOV and street-view based SVI as two coupled but non-equivalent representations of the urban visual environment. We distinguish three constructs that are often conflated: structural openness, apparent openness, and experienced openness. Structural openness is governed by terrain, building height, and corridor geometry and is captured by model-based FOV. Apparent openness is what is visually exposed at eye level under near-field occlusion and seasonal canopy conditions and is captured by SVI semantic segmentation. Experienced openness is how people actually perceive permeability and visibility on site and is used here as a validation anchor. Under this framing, disagreement between FOV and SVI is interpreted as diagnostic information rather than noise. In addition, clustering within each modality can be considered as an optional way to summarize viewpoint-level heterogeneity into a few interpretable types for corridor-scale reading.

4.4.1. Three-layer framework and unified workflow

The structural layer provides stable, cross-scene comparable cues and supports corridor-level diagnosis and tasks such as skyline control and view-corridor identification. Its limitation is representational: when canopy and other near-field elements are not encoded in the 3D model, FOV can overstate what is seen in leaf-on conditions. The visual layer captures near-field components such as canopy, occluding edges, and surface conditions and is therefore sensitive to seasonal and maintenance-related changes that shape immediate perception. Its limitation is that canopy-occlusion can distort site character inference, for example by suppressing apparent building share when facades are blocked, and camera and phenology effects can reduce comparability if segmentation outputs are read without context.

Perception sampling forms the validation layer and is triggered when the two computational layers disagree in decision-relevant ways (Figs. 10 and 11). It is implemented through small-sample checks at selected viewpoints, recording cues such as openness, vegetation abundance, sky-water visibility, terrain undulation, and building-height impression. The workflow proceeds from task definition to scope alignment on shared elements, parallel reading of both outputs, and a reliability label for each segment or viewpoint. Where useful, cluster summaries may be reported alongside the maps to provide a compact description of typical structural or appearance patterns, without implying one-to-one correspondence across methods. Agreement cases are treated as low-risk, while mismatch cases are flagged for verification. In particular, high-FOV and low-SVI cases are interpreted as occlusion-dominated and are suitable targets for actions such as selective canopy pruning or interface clearing, with perception checks used to confirm expected perceptual gains.

4.4.2. Application scenarios

The framework supports planning and management without forcing numerical fusion. For openness evaluation and skyline-related decisions, FOV provides a structural baseline, while SVI reveals where canopy closure suppresses sky-water visibility in lived experience, guiding targeted opening of view windows where structural potential exists. For interface governance and greenery maintenance, SVI flags occlusion-dominated locations for verification and prioritization, while FOV ensures interventions align with corridor geometry rather than image artifacts. Overall, the combined structural-visual-validated reading provides an interpretable basis for diagnosing mismatches, prioritizing field checks, and linking micro-interventions to perceptual outcomes.

4.5. Limitations

This study provides a comparative framework for quantifying streetscape visual composition using complementary model-based and image-based pipelines. Several limitations remain in terms of data quality, algorithmic uncertainty, and the current stage of empirical integration. First, the simplified 3D environment is constrained by model fidelity. Building geometries are represented at a level of detail that may not fully capture façade articulation, overhangs, fine-scale structures, or small objects that can affect visibility. As a result, some element shares, particularly those related to buildings and structures, may be sensitive to geometric generalization. Vegetation geometry is intentionally excluded to isolate a planar baseline rather than to reproduce canopy occlusion. Second, the SVI pipeline inherits uncertainty from semantic segmentation. Despite consistent preprocessing and class re-mapping, misclassification can occur under challenging illumination, shadowing, motion blur, and visually ambiguous boundaries, and these errors propagate to class-share estimates. Third, the FOV ray-casting procedure introduces discretization error because visibility is sampled at finite angular intervals. Although our sensitivity tests indicate qualitative stability at the chosen resolution, denser angular sampling may recover additional local occlusion and visibility details in complex scenes at a higher computational cost. Finally, the current study focuses on quantifying and comparing visual compositions rather than fully fitting the metrics to downstream behavioral or planning outcomes. Calibration against independent ground truth and end-to-end validation in applied planning settings remain important next steps.

5. Conclusions

This paper conducted a systematic comparison between SVI and digital model-based FOV approaches along the West Lake ring road. The results show that the two pipelines capture different but complementary constructs of the urban visual environment. The FOV analysis provides a stable structure-based baseline governed by terrain and built form, whereas the SVI analysis captures appearance-based composition that is strongly affected by canopy occlusion, acquisition season, and near-field clutter. Given the widespread use of SVI as a default proxy for visual exposure and perceived environment, a key contribution of this study is to systematically reveal these limitations and the conditions under which SVI-derived indicators can be misinterpreted as stable structural signals.

Building on these findings, we propose a three-layer framework that integrates the image layer, the model layer, and a perception layer into a unified workflow. Rather than forcing numerical fusion, the framework supports parallel reading of structural and appearance signals, classification of segments into reliable and caution categories, and targeted perception sampling when cross-method disagreement indicates interpretability risks. This workflow is applicable to tasks such as skyline and openness control, interface and character management, roadside greenery maintenance, equity assessments of green distribution, and route design to enhance public experience.

Overall, the study shows that neither image-based nor model-based analysis alone is sufficient, but their combination offers a balanced and reproducible way to capture both structural and appearance dimensions of urban landscapes. Future work will improve 3D model fidelity and segmentation robustness and will replicate the framework across additional settings to further establish when SVI is reliable, when it should be used with caution, and how it can be responsibly interpreted in planning and management.

CRediT authorship contribution statement

Yuyang Peng: Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Steffen Nijhuis:** Supervision, Conceptualization. **Zichen Wu:** Software, Methodology. **Yingwen Yu:** Visualization, Software.

Declaration of competing interest

Given his role as Associate Editor, Prof. Steffen Nijhuis had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.las.2026.100034.

References

Alpherts, T., Ghebreab, S., Noord, Nv, 2025. Artifacts of idiosyncracy in global street view data. Proc. 2025 ACM Conf. Fairness Account. Transpar. Assoc. Comput. Mach. 416–437.

ANON Comparison of methodologies for computing sky view factor in urban environments. Los Alamos National Laboratory (LANL), Los Alamos, NM (United States); 2001.

Biljecki, F., Zhao, T., Liang, X., Hou, Y., 2023. Sensitivity of measuring the urban form and greenery using street-level imagery: a comparative study of approaches and visual perspectives. Int. J. Appl. Earth Obs. Geoinf. 122, 103385. https://doi.org/10.1016/j.jag.2023.103385

Chamberlain, B.C., Meitner, M.J., 2013. A route-based visibility analysis for landscape management. Landsc. Urban Plan. 111, 13–24. https://doi.org/10.1016/j.landurbplan.2012.12.004

Cinnamon, J., Jahiu, L., 2021. Panoramic street-level imagery in data-driven urban research: a comprehensive global review of applications, techniques, and practical considerations. ISPRS Int. J. GeoInf. 10 (7), 471. ⟨https://www.mdpi.com/2220-9964/10/7/471⟩.

Eriksson, H., Harrie, L., 2021. Versioning of 3D city models for municipality applications: needs, obstacles and recommendations. ISPRS Int. J. GeoInf. 10 (2), 55. ⟨https://www.mdpi.com/2220-9964/10/2/55⟩.

Fan, Z., Feng, C.-C., Biljecki, F., 2025. Coverage and bias of street view imagery in mapping the urban environment. Comput. Environ. Urban Syst. 117, 102253. https://doi.org/10.1016/j.compenvurbsys.2025.102253

Fisher, P.F., 1993. Algorithm and implementation uncertainty in viewshed analysis. Int. J. Geogr. Inf. Syst. 7 (4), 331–347. https://doi.org/10.1080/02693799308901965

Gaw, L.Y., Chen, S., Chow, Y.S., Lee, K., Biljecki, F., 2022. Comparing street view imagery and aerial perspectives in the built environment. ISPRS Ann. Photo Remote Sens. Spat. Inf. Sci. X-4/W3-2022, 49–56. https://doi.org/10.5194/isprs-annals-X-4-W3-2022-49-2022

Ghasemian Sorboni, N., Wang, J., Najafi, M.R., 2024. Fusion of google street view, LiDAR, and orthophoto classifications using ranking classes based on F1 score for building land-use type detection. Remote Sens. 16 (11), 2011. ⟨https://www.mdpi.com/2072-4292/16/11/2011⟩.

Gobster, P.H., Ribe, R.G., Palmer, J.F., 2019. Themes and trends in visual assessment research: introduction to the Landscape and Urban planning special collection on the visual assessment of landscapes. Landsc. Urban Plan. 191, 103635. https://doi.org/10.1016/j.landurbplan.2019.103635

Gong, F.-Y., Zeng, Z.-C., Zhang, F., Li, X., Ng, E., Norford, L.K., 2018. Mapping sky, tree, and building view factors of street canyons in a high-density urban environment. Build. Environ. 134, 155–167. https://doi.org/10.1016/j.buildenv.2018.02.042

Guth, P.L., Van Niekerk, A., Grohmann, C.H., Muller, J.-P., Hawker, L., Florinsky, I.V., et al., 2021. Digital elevation models: terminology and definitions. Remote Sens. 13 (18), 3581. ⟨https://www.mdpi.com/2072-4292/13/18/3581⟩.

He, N., Li, G., 2021. Urban neighbourhood environment assessment based on street view image processing: a review of research trends. Environ. Chall. 4, 100090. https://doi.org/10.1016/j.envc.2021.100090

Hou, Y., Quintana, M., Khomiakov, M., Yap, W., Ouyang, J., Ito, K., et al., 2024. Global Streetscapes — A comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics. ISPRS J. Photogramm. Remote Sens. 215, 216–238. https://doi.org/10.1016/j.isprsjprs.2024.06.023

Inglis, N.C., Vukomanovic, J., Costanza, J., Singh, K.K., 2022. From viewsheds to viewscapes: trends in landscape visibility and visual quality research. Landsc. Urban Plan. 224, 104424. https://doi.org/10.1016/j.landurbplan.2022.104424

Ito, K., Zhu, Y., Abdelrahman, M., Liang, X., Fan, Z., Hou, Y., et al., 2025. ZenSVI: An open-source software for the integrated acquisition, processing and analysis of street view imagery towards scalable urban science. Comput. Environ. Urban Syst. 119, 102283. https://doi.org/10.1016/j.compenvurbsys.2025.102283

van Kamp, I., Leidelmeijer, K., Marsman, G., de Hollander, A., 2003. Urban environmental quality and human well-being: towards a conceptual framework and demarcation of concepts; a literature study. Landsc. Urban Plan. 65 (1), 5–18. https://doi.org/10.1016/S0169-2046(02)00232-3

Kang, Y., Zhang, F., Gao, S., Lin, H., Liu, Y., 2020. A review of urban physical environment sensing using street view imagery in public health studies. Ann. GIS 26 (3), 261–275. https://doi.org/10.1080/19475683.2020.1791954

Khan, A.Z., Moulaert, F., Schreurs, J., Miciukiewicz, K., 2014. Integrative spatial quality: A Relational Epistemology of Space and Transdisciplinarity in Urban Design and Planning. Taylor & Francis, pp. 393–411.

Kim, J., Jang, K.M., 2023. An examination of the spatial coverage and temporal variability of Google Street View (GSV) images in small- and medium-sized cities: a people-based approach. Comput. Environ. Urban Syst. 102, 101956. https://doi.org/10.1016/j.compenvurbsys.2023.101956

Kyttä, M., Kahila, M., Broberg, A., 2011. Perceived environmental quality as an input to urban infill policy-making. URBAN Des. Int. 16 (1), 19–35. https://doi.org/10.1057/udi.2010.19

Lei B., Su Y., Biljecki F. Humans As Sensors in Urban Digital Twins. 2024. p. 693-706.

Li, M., Dai, W., Song, S., Wang, C., Tao, Y., 2023. Construction of high-precision DEMs for urban plots. Ann. GIS 29 (2), 193–203. https://doi.org/10.1080/19475683.2023.2182360

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., Zhang, W., 2015. Assessing street-level urban greenery using google street View and a modified green view index. Urban For. Urban Green. 14 (3), 675–685. https://doi.org/10.1016/j.ufug.2015.06.006

Li, Y., Peng, L., Wu, C., Zhang, J., 2022. Street View Imagery (SVI) in the Built Environment: A theoretical and systematic review. Buildings 12 (8), 1167. 〈https://www.mdpi.com/2075-5309/12/8/1167〉.

Liang, J., Gong, J., Sun, J., Zhou, J., Li, W., Li, Y., et al., 2017. Automatic sky view factor estimation from street view photographs—a big data approach. Remote Sens. 9 (5), 411. 〈https://www.mdpi.com/2072-4292/9/5/411〉.

Liang, X., Zhao, T., Biljecki, F., 2023. Revealing spatio-temporal evolution of urban visual environments with street view imagery. Landsc. Urban Plan. 237, 104802. https://doi.org/10.1016/j.landurbplan.2023.104802

Liu, M., Nijhuis, S., 2020. Mapping landscape spaces: methods for understanding spatial-visual characteristics in landscape design. Environ. Impact Assess. Rev. 82, 106376. https://doi.org/10.1016/j.eiar.2020.106376

Llobera, M., 2003. Extending GIS-based visual analysis: the concept of visualscapes. Int. J. Geogr. Inf. Sci. 17 (1), 25–48.

Lu, X., Li, Q., Ji, X., Sun, D., Meng, Y., Yu, Y., et al., 2025. Impact of streetscape built environment characteristics on human perceptions using street view imagery and deep learning: a case study of changbai island, shenyang. Buildings 15 (9), 1524. 〈https://www.mdpi.com/2075-5309/15/9/1524〉.

Luo, J., Liu, P., Kong, X., Shen, J., Wu, Q., Xu, D., 2025. Urban digital twins for citizen-centric planning: a systematic review of built environment perception and public participation. Int. J. Appl. Earth Obs. Geoinf. 143, 104746. https://doi.org/10.1016/j.jag.2025.104746

Morgan, G.R., Zlotnick, D., North, L., Smith, C., Stevenson, L., 2024. Deep learning for Urban tree canopy coverage. Anal. A Comp. Case Study Geomat. 4 (4), 412–432. 〈https://www.mdpi.com/2673-7418/4/4/22〉.

Oke, T.R., 1981. Canyon geometry and the nocturnal urban heat island: Comparison of scale model and field observations. J. Climatol. 1 (3), 237–254. https://doi.org/10.1002/joc.3370010304

Orlof, J., Ozimek, P., Łabędź, P., Widłak, A., Ozimek, A., 2024. Generating viewsheds based on the digital surface model (DSM) and point cloud. PloS One 19 (12), e0312146.

Orlof, J., Ozimek, P., Łabędź, P., Widłak, A., Ozimek, A., 2025. Generating viewsheds based on the digital surface model (DSM) and point cloud. PLOS ONE 19 (12), e0312146. https://doi.org/10.1371/journal.pone.0312146

Parent, J.R., Lei-Parent, Q., 2023. Rapid viewshed analyses: a case study with visibilities limited by trees and buildings. Appl. Geogr. 154, 102942. https://doi.org/10.1016/j.apgeog.2023.102942

Peng Y., Nijhuis S.A. GIS-based algorithm for visual exposure computation: the west lake in Hangzhou (China) as example. 2021,

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 For. Urban Green. 91, 128159. https://doi.org/10.1016/j.ufug.2023.128159

Puspitasari, A.W., Kwon, J., 2021. A reliable method for visibility analysis of tall buildings and skyline: a case study of tall buildings cluster in Jakarta. J. Asian Archit. Build. Eng. 20 (3), 356–367. https://doi.org/10.1080/13467581.2020.1787839

Rzotkiewicz, A., Pearson, A.L., Dougherty, B.V., Shortridge, A., Wilson, N., 2018. Systematic review of the use of Google Street View in health research: Major themes, strengths, weaknesses and possibilities for future research. Health Place 52, 240–246. https://doi.org/10.1016/j.healthplace.2018.07.001

Schoorl, J.M., Sonneveld, M.P.W., Veldkamp, A., 2000. Three-dimensional landscape process modelling: the effect of DEM resolution. Earth Surf. Process. Landf. 25 (9), 1025–1034. https://doi.org/10.1002/1096-9837(200008)25:9<1025::AID-ESP116>3.0.CO;2-Z

Wang, J., 2023. The digital presentation of human-oriented urban design. Sustain. Cities Soc. 97, 104746. https://doi.org/10.1016/j.scs.2023.104746

Wang, R., Peethambaran, J., Chen, D., 2018. LiDAR Point Clouds to 3-D Urban Models$:$ A Review. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (2), 606–627. https://doi.org/10.1109/JSTARS.2017.2781132

Wu, C., Yu, X., Ma, C., Zhong, R., Zhou, X., 2024. Integrating geospatial data and street-view imagery to reconstruct large-scale 3D urban building models. Trans. GIS 28 (5), 1326–1352 https://doi.org/10.1111/tgis.13192.

Xu, F., Jin, A., Chen, X., Li, G., 2021. New data, integrated methods and multiple applications: a review of urban studies based on street view images. 2021 IEEE Int. Geosci. Remote Sens. Symp IGARSS.

Zhao, T., Xiucheng, L., Biljecki, F., Tu, W., Cao, J., Li, X., et al., 2025. Quantifying seasonal bias in street view imagery for urban form assessment: a global analysis of 40 cities. Comput. Environ. Urban Syst. 120, 102302. https://doi.org/10.1016/j.compenvurbsys.2025.102302

Zhou, Z., Zhong, T., Liu, M., Ye, Y., 2023. Evaluating building color harmoniousness in a historic district intelligently: an algorithm-driven approach using street-view images. Enviro. Planning B Urban Analy. City Sci. 50 (7), 1838–1857. https://doi.org/10.1177/23998083221146539

Župan, R., Vinković, A., Nikçi, R., Pinjatela, B., 2023. Automatic 3D building model generation from airborne lidar data and openstreetmap using procedural modeling. Information 14 (7), 394. 〈https://www.mdpi.com/2078-2489/14/7/394〉.