Author-accepted manuscript (postprint).
Published in the International Journal of Geographical Information Science*, 2026.*
DOI: 10.1080/13658816.2026.2656269

Vision2Slope: Estimating urban road slopes with street view imagery

Yang Chen[a,b,c], Zicheng Fan[b], Hao Li[d], Wufan Zhao[e], Xin Yang[a,c,f], Guoan Tang[a,c,f], and Filip Biljecki[b,g]

[a]State Key Laboratory of Climate System Prediction and Risk Management, Nanjing Normal University, Nanjing, China; [b]Department of Architecture, National University of Singapore, Singapore; [c]Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing, China; [d]Department of Geography, National University of Singapore, Singapore; [e]Urban Governance and Design Thrust, Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China; [f]Jiangsu Centre for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, China; [g]Department of Real Estate, National University of Singapore, Singapore

ARTICLE HISTORY
Compiled April 19, 2026

Citation (Version of Record): Chen Y, Fan Z, Li H, Zhao W, Yang X, Tang G, Biljecki F (2026). Vision2Slope: estimating urban road slopes with street view imagery. *International Journal of Geographical Information Science*, 1–45. https://doi.org/10.1080/13658816.2026.2656269

ABSTRACT
Road slopes shape mobility patterns and drive the reliability of urban simulations. Yet in most cities, road-level slope information remains scarce. We introduce Vision2Slope, a framework that leverages panoramic street view imagery to estimate road slopes using computer vision techniques. The workflow consists of three steps: (i) projecting panoramic images into road-aligned views; (ii) semantic-prompted image deskewing to correct geometric distortion induced by camera orientation; and (iii) a two-level slope estimation strategy that extracts point- and segment-level slope and relief characteristics from road-edge geometry and iterative regression to reduce outliers. Using Google Street View images from San Francisco and New York City, the framework estimates slopes for over 60,000 locations and 17,000 street segments. Point- and segment-level MAEs are 0.81°/0.57° and 0.72°/0.78°, respectively, with segment relief errors of 1.70 and 1.66 m. Conditional bias analysis reveals the influence of street-level environmental features on estimation accuracy. The proposed framework significantly outperforms the omnipresent 30 m digital elevation models and maintains robustness under simulated changes in camera orientation and imaging conditions. As an open and scalable workflow, Vision2Slope emphasizes the potential of street view imagery for cost-effective, detailed urban road slope mapping, enriching foundational data for vertical-aware urban analytics.

KEYWORDS
Road network; 3D geographic information; Urban terrain; Hilly city; Ground-level imagery

CONTACT Filip Biljecki. Email: filip@nus.edu.sg

1. Introduction

Roads weave through urban environments in complex spatial patterns (Xue *et al.* 2022). Besides a plethora of research on their two-dimensional layout, their third dimension — slope or relief — which reveals how urban form adapts to underlying terrain, is progressively being brought into focus (Higgins 2019). As cities increasingly expand toward sloped landscapes driven by population growth and land scarcity (Yang *et al.* 2022, Shi *et al.* 2025), inclined road segments have become common features in hilly and mountainous urban areas, such as San Francisco and Hong Kong (Figure 1). Despite the role of road slope in shaping urban form, current urban digital twins and 3D city modeling efforts focus primarily on buildings and vegetation rather than road networks, which often remain represented as flat surfaces (Ying *et al.* 2020, Zhao *et al.* 2023). This simplification hinders a wide range of urban analytics and applications. For example, road slope shapes human and vehicular mobility (Faria *et al.* 2019, Jiang *et al.* 2025), influences flood routing (van Ginkel *et al.* 2021), affects energy consumption (Rosero *et al.* 2021), and constrains transportation planning toward sustainable development (Larrañaga *et al.* 2016), but most of such analyses have not evolved into 3D like those focusing on other urban features. A key hindrance to such developments is the lack of widely available data suitable to support them. These gaps call for accurate, spatially explicit estimation of road slope in urban areas for both enhancing existing urban analysis and management and developing new use cases.

Beyond traditional survey measurements, various geospatial data sources have been used to estimate road slope. Broadly, existing methods fall into two categories: overhead and ground-level observations. From the overhead perspective, digital elevation models (DEMs) derived from satellite or aerial imagery provide large-scale elevation coverage. However, their moderate resolution, typically 10–30 m, tends to smooth out the fine undulations of urban roads with mixed effects, particularly in complex terrain or densely vegetated areas (Liu *et al.* 2018, Xiong *et al.* 2025). Ground-level sensing offers higher fidelity. On-vehicle Light Detection and Ranging (LiDAR) systems can capture highly detailed 3D point clouds of road surfaces (Yadav *et al.* 2017), but their operational cost and limited spatial coverage hinder broad application. Crowdsourced GPS trajectories from smartphones and vehicles offer a cost-effective alternative (Gupta *et al.* 2020), though they are often compromised by signal noise and multi-path interference in urban canyons. While these approaches have advanced our capacity to quantify road slope, they remain constrained by cost, coverage, and observation perspective. Further, when accurate data is collected by authorities and companies, it is rarely available openly. Consequently, there is a growing need for cost-efficient and robust data sources that can perceive slope to bridge the gap between fine geometric accuracy and large-scale applicability.

Street View Imagery (SVI), primarily captured and maintained by proprietary services such as Google Street View (GSV)[1], Baidu Street View[2], and crowdsourced platforms such as Mapillary[3], provides a promising alternative for extracting urban road information. Unlike overhead imagery, SVI captures scenes from the ground-level perspective at a high spatial granularity, and it provides a direct and high-resolution insight into the geometry and visual context of road surfaces. This viewpoint, thanks to the usually high spatial frequency of acquisition and predominantly high quality of imagery, preserves subtle vertical variations that are often smoothed out or obscured

[1]https://www.google.com/maps
[2]https://map.baidu.com/
[3]https://www.mapillary.com/

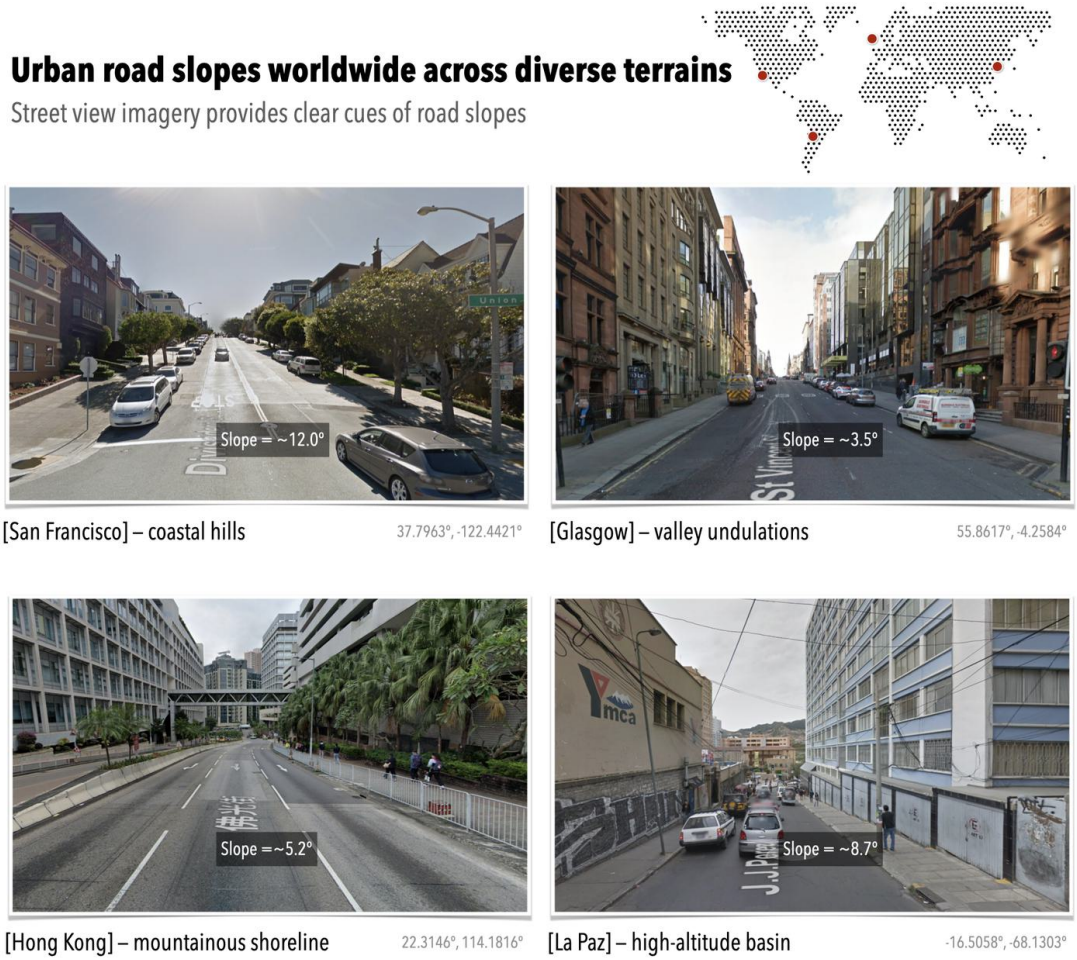

Figure 1. Four examples of slope roads from around the world. Source of imagery: Google Street View.

in aerial data, particularly in dense urban environments. Therefore, SVI creates a unique opportunity to recover fine-grained road geometry from readily accessible imagery. SVI provides rich geometric and semantic information, and advances in computer vision have greatly expanded what can be extracted from it (Fan *et al.* 2023). Deep learning models now support a wide range of street-level applications, including street function classification (Fang *et al.* 2022, Huang *et al.* 2024), infrastructure assessment (Szcześniak *et al.* 2022, Ma *et al.* 2025), and perception modeling (Gebru *et al.* 2017, Danish *et al.* 2025). Despite this progress, most efforts center on 2D scene understanding, and the use of SVI to derive 3D road geometry remains limited. While stereoscopic and multi-view SVI have enabled urban 3D reconstruction (Torii *et al.* 2009, Micusik and Kosecka 2009, Qin *et al.* 2024), these approaches depend on controlled camera settings, dense viewpoints, or computationally heavy pipelines, constraining their scalability for citywide use. This gap highlights the need for a scalable approach that harnesses widely available single-view SVI to directly estimate road slope in urban environments.

Among existing SVI-based studies, very few provide a unified approach for converting single-view street images into reliable, citywide estimates of road slope. Earlier attempts were typically limited to small-scale demonstrations, which relied on explicit

camera metadata or employed simplified heuristics that do not generalize to complex urban environments. Moreover, very few studies have analyzed the bias effects of urban forms. Inspired by them, we are determined to explore the application potential of SVI in large-scale road slope estimation and raise three research questions:

(1) To what extent do urban street-level visual environments encode information about road slope?
(2) How do image-related uncertainties, such as perspective distortion and visual ambiguity, affect the inference of road slope from single-view imagery?
(3) What characteristics of urban morphology systematically contribute to biases in slope estimation?

To answer these questions and address the mentioned limitations, our framework offers a fully image-driven, metadata-free pipeline that scales to entire cities while systematically correcting the geometric distortions inherent in panoramic SVI.

The main contributions of this study are as follows:

- Developing a new perspective and scalable framework, Vision2Slope, for urban road slope estimation using a single panoramic SVI. The framework integrates data acquisition and processing, semantic-prompted image deskewing, and two-level slope estimation within a unified and explicit workflow (Section 3).
- Conducting intensive experiments in two major U.S. cities with distinct urban contexts: San Francisco and New York City. A total of over 60,000 points and 17,000 road segments are estimated and mapped. The results demonstrate that our method achieves accepted accuracy in road slope estimation, compared to ground-truth slopes derived from high-resolution DEMs (Section 4.3 to Section 4.6).
- Systematically analyzing the uncertainties associated with urban built-up features and camera parameters, and uncovering the effects of image quality degradation (pitch, roll, brightness, field of view) to assess the robustness and generalizability of the framework and future application (Section 4.7).

In summary, this study introduces a data- and knowledge-driven framework for urban road slope estimation from street-view imagery using computer vision techniques. The proposed solution offers a cost-effective and scalable approach to obtaining large-scale 3D road network information, overcoming the limitations of conventional elevation and surveying data. By validating the reliability of SVI for extracting vertical road geometry, this work advances the contribution of street-level imagery to urban 3D modeling and supports the development of accurate, fine-grained, and continually updatable urban digital twins.

2. Related work

2.1. *Traditional data sources for road slope estimation*

Road slope estimation has been conducted using various data sources, each characterized by distinct differences in spatial resolution, road coverage (i.e., the effective coverage of road surfaces usable for slope estimation), cost, and observation perspective, as summarized in Table 1.

Traditional field surveys capture detailed elevation points along the road but are labor-intensive and costly, making them unsuitable for large-scale applications. Remote

sensing techniques address this limitation by enabling large-scale mapping of road slopes from an overhead perspective. Unmanned Aerial Vehicles (UAVs) equipped with high-resolution cameras or LiDAR sensors can capture detailed 3D information on road surfaces and provide flexible data acquisition and rapid coverage of target areas (Gao *et al.* 2021). However, their operational costs, regulatory constraints, and limited flight endurance hinder widespread use. Satellite-based observations offer a cost-effective alternative for large-scale road slope estimation. Medium-resolution DEMs derived from satellite observation, such as Shuttle Radar Topography Mission and Advanced Land Observing Satellite World 3D - 30 m, are widely used (Liu *et al.* 2018, Wang *et al.* 2021, Chen *et al.* 2022). Yet their spatial resolution (typically 30 m or coarser) is insufficient to capture the fine-scale variations of roads, especially in dense urban environments (El Masri and Bigazzi 2019). Although some regional DEMs provide higher resolution (e.g., 5 m or 10 m) and finer road elevation change (Sun *et al.* 2019, Hosseini *et al.* 2024), their coverage is limited and often unavailable. Notably, while overhead observations provide efficient ways to capture large-scale road slope, they all face difficulties in areas with tall buildings and dense vegetation canopies. These limitations not only degrade the original data quality but also impose a high cost on processing large datasets, as automatic editing of vegetation and buildings is still not ideal.

Ground-based data provides another perspective for road slope estimation. This data type refers to measurements collected on the ground, typically within road network, rather than from an overhead view. Compared with overhead observations, it captures road information more directly without obstruction from the top canopies. Mobile laser LiDAR is a representative ground-based data that includes detailed 3D point clouds of roads. It offers good coverage along road networks and achieves relatively high vertical accuracy (Yang *et al.* 2013, Salazar Miranda *et al.* 2022). Although it enables the reconstruction of complex 3D road structures, its widespread use is constrained by the high cost of equipment and the complexity of data processing. In contrast, GPS trajectories from smartphones or vehicles provide a cost-effective option for estimating road slopes (Yazdani Boroujeni and Frey 2014, John *et al.* 2017, McKenzie and Janowicz 2017). They record the 3D coordinates (longitude, latitude, elevation) of points over time, from which the road slope can be inferred by calculating horizontal and vertical differences. However, the positional accuracy of consumer-grade GPS receivers is often degraded by signal noise and multipath effects, especially in dense urban canyons (Yang *et al.* 2023). Despite various algorithms developed to address these issues, substantial efforts are still required for data cleaning and map matching. Furthermore, the availability of large-scale GPS datasets is restricted by privacy concerns (Seidl *et al.* 2016). Similar limitations apply to in-vehicle sensors, such as accelerometers, gyroscopes, and fuel consumption data, which measure the variations of value changes along trajectories (Du *et al.* 2016, Jauch *et al.* 2018, Fan *et al.* 2022). Road slopes can be estimated by fitting the trend of these measurements with distance. Yet, their accuracy depends heavily on sampling frequency, as short spatial intervals are needed to capture slope variations. Consequently, the derived slope profiles are often discontinuous, and the data quality varies significantly across different vehicles and sensor modes.

These limitations highlight the need to find a trade-off among different data sources to balance resolution, road coverage, and cost for effective road slope estimation. It is also essential to develop methods that address the obstruction issues associated with overhead perspectives. In this regard, SVI presents a promising alternative and has unique advantages. 1) SVI is organized as dense sampling points, each with a unique

Table 1. General comparison of data sources for road slope estimation.

Data Source	Resolution	Road Coverage	Cost	Perspective
Field survey	●●●●	●●○○	●○○○	Ground-based
Mobile LiDAR	●●●●	●●●●	●●○○	Ground-based
UAV photogrammetry	●●●○	●●●○	●●○○	Overhead
GPS & in-vehicle sensors	●●○○	●●●○	●●●○	Ground-based
Regional DEMs	●●○○	●●●○	●●●○	Overhead
Global DEMs	●○○○	●○○○	●●●●	Overhead

Note: More filled circles indicate *better*. For **Cost**, more means *more affordable*.

identifier. Although the distance between points is not fixed, the participatory uploads by residents and time-series updates help fill spatial gaps, providing adequate resolution for road slope estimation; 2) In terms of road coverage, platforms such as Google, Baidu, and Mapillary have extensively covered drivable roads worldwide. SVIs inherently contain rich visual information about road surfaces, including color and geometry, similar to mobile LiDAR; 3) The cost of SVIs is relatively low. Commercial sources such as GSV or Baidu are more affordable than LiDAR systems or field surveys. Several open projects also offer freely accessible annotated SVI datasets, e.g. NUS Global Streetscapes (Hou *et al.* 2024); 4) as trajectory-based data, SVIs largely mitigate the adverse effects of vegetation canopies and high-rise buildings when capturing road information, like point clouds from mobile LiDAR. Considering these advantages, SVIs are likely to represent a valuable data source for large-scale road slope estimation. However, their potential in this field remains underexplored, particularly for large-scale urban analyses.

2.2. *SVI-based approaches for 3D urban information extraction*

With the increasing availability of SVI, it has become a key data source for analyzing urban environments from a ground perspective (Biljecki and Ito 2021). SVI provides rich visual cues about urban geometry and structures, enabling 3D characterization of streets, buildings, and public spaces (Zhang *et al.* 2021). Depending on the acquisition mode, SVI can be categorized into single-frame and panoramic imagery. Among them, panoramic imagery, which captures the full 360-degree street scene, offers comprehensive spatial context and is particularly valuable for 3D urban reconstruction and spatial understanding tasks (Fan *et al.* 2025).

Existing methods for extracting 3D information from SVI generally fall into two categories: multi-view stereo reconstruction and single-view inference. Multi-view stereo methods utilize pairs or sequences of SVIs captured from different viewpoints to recover 3D scene geometry through feature matching, camera pose estimation, and dense point cloud generation (Schönberger and Frahm 2016). Frameworks such as OpenSfM[4] and COLMAP[5] represent typical implementations. However, their performance heavily depends on dense spatial sampling and accurate camera metadata, which are often difficult to guarantee in large-scale research due to limited overlap between adjacent

[4]https://github.com/mapillary/OpenSfM
[5]https://github.com/colmap/colmap

viewpoints, camera tilt, and calibration errors. Moreover, the computational demands of multi-view reconstruction limit its scalability to city-scale applications.

Single-view inference methods aim to derive 3D geospatial information from a single SVI (Pang and Biljecki 2022). Compared with multi-view approaches, single-view methods are more efficient because they do not require image-pair matching or precise viewpoint alignment. Recent advances in deep learning have enabled detailed monocular reconstruction of building models directly from individual SVIs (Bacharidis *et al.* 2020). For instance, several studies have applied deep neural networks to estimate block-level building heights, leveraging façade features and geometric transformations captured in SVIs (Yan and Huang 2022, Fan *et al.* 2024, Yan *et al.* 2025). Similarly, the lowest elevations of residential buildings have been inferred by integrating depth estimation with visual features from SVI (Ning *et al.* 2022b). Despite these advances, most existing studies have focused on building-related features, whereas the road surface itself, particularly road slope, has received little attention.

Very few studies have explored the use of SVI for road slope estimation. Lu and Karimi (2015) applied edge detection and calculated the angular differences between edges to infer slopes. This attempt illustrated that SVI could, in principle, be used for slope inference, but the approach remained heuristic, was demonstrated only on a very limited area, and did not account for the visual complexity of real urban environments. The pitch angle recorded in commercial SVI metadata also provides information about road slope under certain conditions (Ning *et al.* 2022a). The pitch angle refers to the angle between the camera's optical axis and the horizontal plane; when the camera points toward the road surface, it approximately corresponds to the road slope. However, pitch angle data are often unavailable in open-access SVI datasets. Even in commercial datasets where such metadata exist, data-sharing restrictions prevent public access. Furthermore, differences in camera models and mounting positions can introduce inconsistencies, making pitch angles unreliable as direct indicators of road slope. Therefore, a more robust and accessible method for estimating road slopes from single SVI images is needed. Building on these gaps, our approach departs from prior work by correcting the inherent geometric distortions in panoramic SVI and estimating the road's vertical geometry through a combination of deep learning and regression-based inference. Rather than relying on camera metadata, we directly recover road slope from a single SVI using a deskewing–regression pipeline, enabling large-scale application without any external high-resolution elevation data.

3. Methodology

Our core idea is to leverage the semantic and geometric features found in SVI to capture variations in road edges for estimating road slopes at the point-level and segment-level in lateral views from panoramas. We focus on lateral views because they explicitly capture road curbs and pavement boundaries on both sides of the street, which provide stable geometric references under most urban settings, as illustrated in the top panels of Figure 2. In an idealized 3D scene where the camera is mounted rigidly above the vehicle and perfectly aligned with the road surface (Figure 2, bottom panel), the apparent inclination of road edges in lateral views would directly correspond to the true road slope. However, in real-world SVI, lateral images are often skewed. This skewing arises from two coupled factors: (1) the actual longitudinal slope of the road and (2) uncertainties in camera geometry, including pitch and roll deviations caused by vehicle motion, uneven pavement, or imperfect sensor calibration. Importantly, both factors

manifest as apparent angular deviations of vertical or near-vertical structures in the lateral views. While these effects are inseparable at the raw image level, road curbs remain a critical and consistently observable cue for slope estimation. Our method is therefore designed to disentangle road-induced inclination from camera-induced distortion through a geometry-aware correction procedure, followed by robust aggregation of curb-derived slope estimates at both point and segment levels.

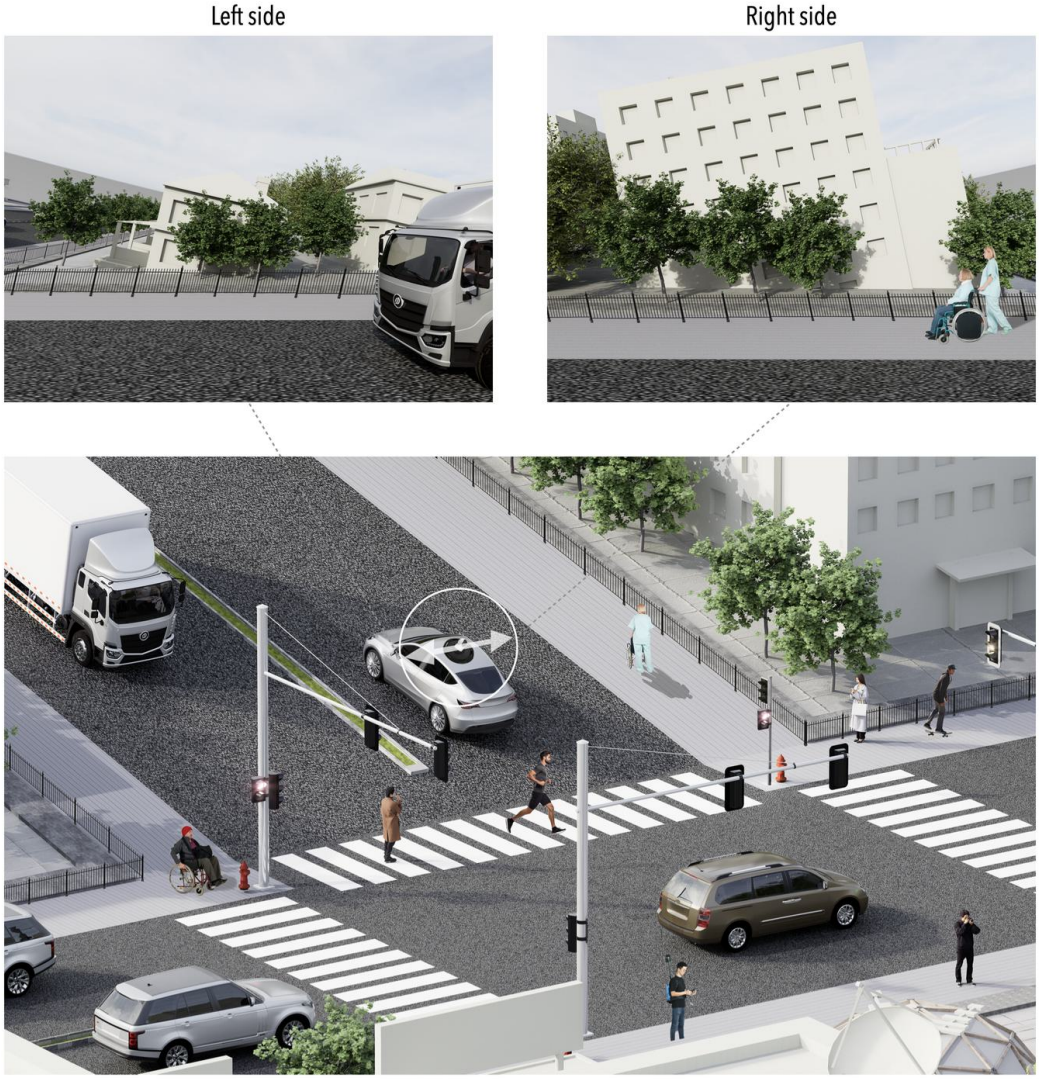

Figure 2. Conceptual illustration of an urban intersection viewed from street-level perspectives. The top panels show left-side and right-side street-view scenes with distinct visibility conditions. The bottom panel provides a plan-view context of the same intersection.

The proposed framework, as shown in Figure 3, consists of several key components: data collection and processing, semantic-prompted SVI deskewing, and multi-level road slope estimation.

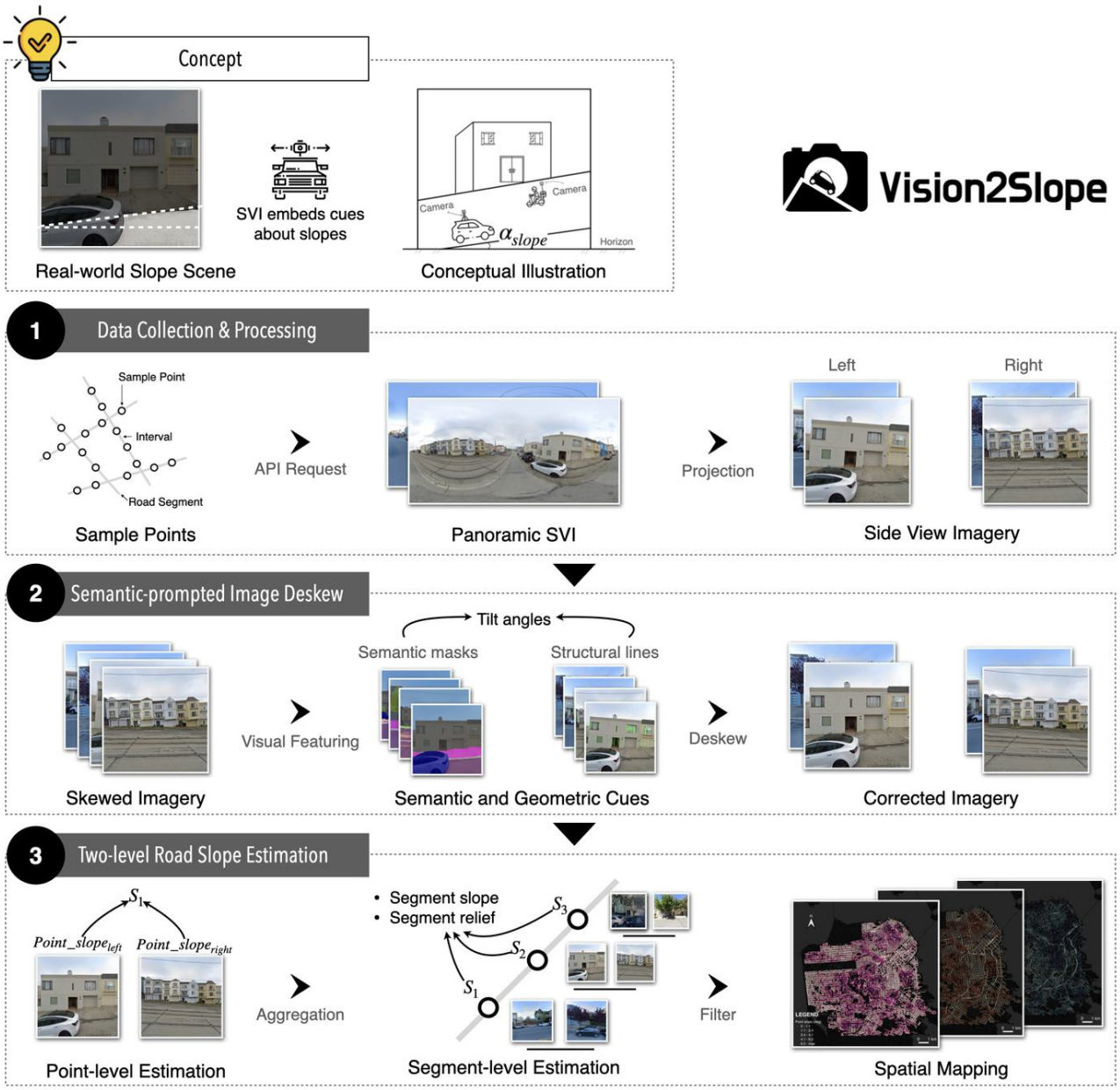

Figure 3. Conceptual modeling and general workflow of the Vision2Slope framework, including data collection, semantic-prompted deskewing, and multi-level road slope estimation. SVI data source: Google Street View.

3.1. *Data collection and processing*

3.1.1. Data collection and sampling points

The first step of our framework was the collection of SVI data. Panoramic images were obtained through the GSV API along the road network, sampled at regular intervals (i.e., every 30 m) to ensure sufficient coverage. The data source of the road network is OpenStreetMap (OSM)[6], which is accessed by OSMnx (Boeing 2025). Considering

[6]https://www.openstreetmap.org/

the sparse distribution of roads in natural areas such as Central Park in NYC, the sampling was focused on built-up regions with dense road networks. To further reduce potential uncertainties, sampling points located within a 15 m buffer of road intersections were excluded, as local slope variations tend to be lower near intersections because intersection areas are usually relatively flat. After spatial filtering, a final quality control was performed to ensure the reliability of the visual data. Panoramas with poor resolution or evident artifacts were manually discarded and replaced with the latest available images. Following these procedures, a total of 32,680 panoramic images were collected for San Francisco and 28,506 for NYC (accessed in June 2025).

3.1.2. Transforming panoramic images to perspective views

The panoramic images were converted into perspective projections using the pinhole camera model. Specifically, a fixed field of view (90°) was defined, and each panorama was projected onto a 2D plane and selected with the side view images through the ZenSVI package (Ito *et al.* 2025). Only the left- and right-side views were retained for further analysis, while the front- and rear-view images were excluded, as they contain limited geometric information for reliable slope estimation.

3.2. *Semantic-prompted SVI deskewing*

SVI often exhibits vertical tilt distortions along road surfaces due to uneven road conditions and varying camera mounting angles. These distortions can substantially reduce the accuracy of road slope estimation. To mitigate this issue, we develop a semantic-prompted deskewing method that estimates the tilt angle from structural lines and corrects the image through geometric transformation. As illustrated in Figure 4, the workflow comprises three main steps: semantic segmentation of urban elements, extraction of structural lines, and tilt angle estimation followed by correction.

3.2.1. Semantic segmentation of urban elements

We adopted the Mask2Former model to segment essential urban features to prompt the imagery deskewing. Mask2Former is developed by Meta with a transformer-based architecture, which captures long-range dependencies within the image by modeling global context through multi-scale attention mechanisms (Cheng *et al.* 2022). This capability allows the model to distinguish between complex and visually similar urban elements. The generated segmentation masks serve as the foundation for subsequent structural feature extraction, enabling the isolation of human-made elements (e.g., buildings, walls, guard rails) that provide semantic prompts for tilt angle estimation, while excluding trees that may introduce noise into the analysis.

3.2.2. Structural line extraction and filtering

To detect candidate structural lines that could serve as reliable references for tilt estimation, we first converted the imagery to grayscale and applied a Gaussian blurring operation to suppress high-frequency noise and enhance edge continuity. Subsequently, the Canny edge detector was employed to extract edge maps that delineate prominent intensity transitions in SVIs. This preprocessing helps stabilize line detection, particularly in complex urban scenes where textures, shadows, and reflections might otherwise lead to fragmented or spurious edges.

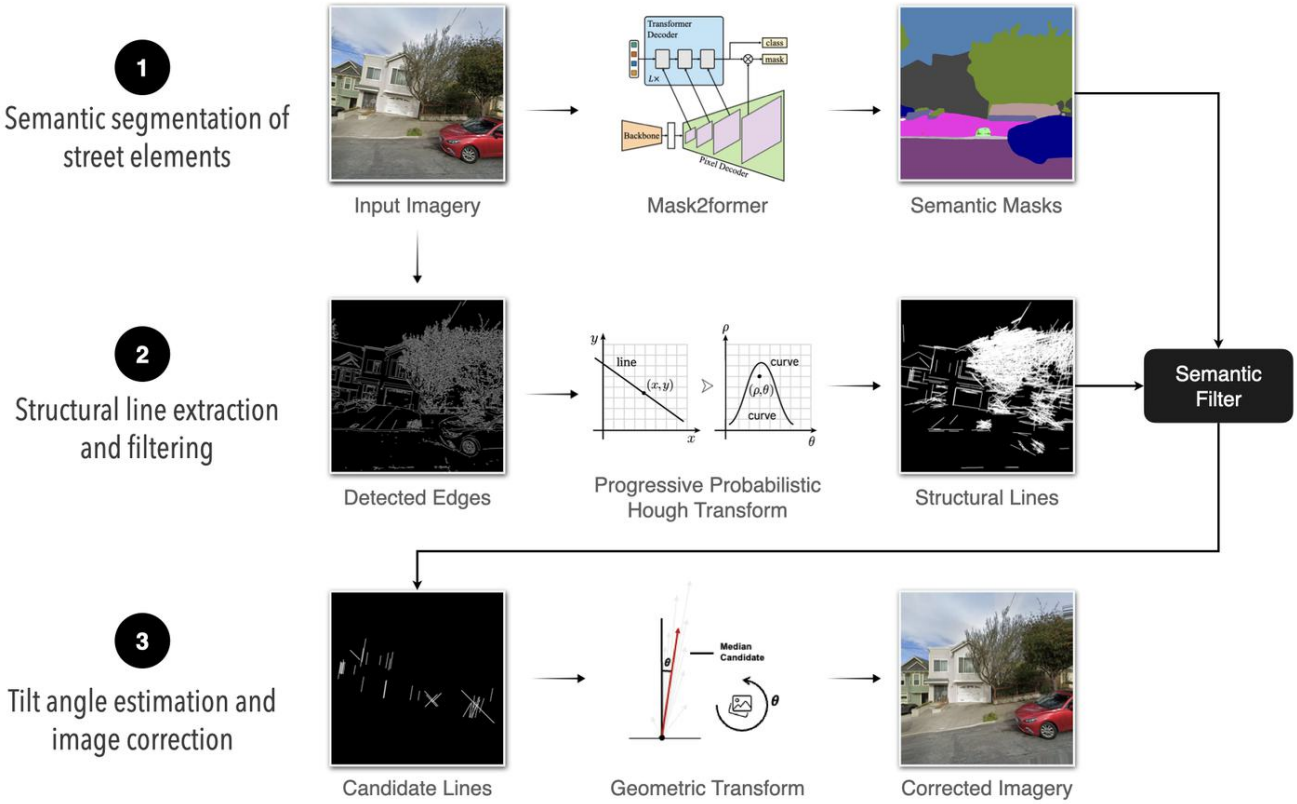

Figure 4. Workflow of the semantic-prompted deskewing module. Source of the imagery: Google Street View.

After edge extraction, we adopted the Progressive Probabilistic Hough Transform (PPHT) to identify line segments. Unlike the common Hough Transform, which accumulates votes in the entire parameter space, PPHT performs iterative probabilistic sampling of edge points and dynamically updates the line parameters as more evidence accumulates (Matas *et al.* 2000). The result of this step is a set of candidate linear features that represent the dominant structural elements within each image.

To refine these candidate lines, we applied a two-step filtering strategy based on geometric and semantic constraints. Geometrically, only lines whose orientations deviated by less than 10 degrees from the vertical direction and whose lengths exceeded 50 pixels were retained, ensuring that the remaining features corresponded to prominent, near-vertical structures. Semantically, we cross-referenced these lines with object masks derived from semantic segmentation. Lines located within building, wall, fence, and guard rail masks were preserved as structurally reliable, while those intersecting tree masks were discarded. Tree branches, despite their linear appearance, rarely maintain consistent vertical alignment and could introduce significant noise into the tilt estimation.

3.2.3. Tilt angle estimation and image correction

The tilt angle of SVI was estimated based on the orientation of the detected structural lines after filtering. For each line segment, we calculated its deviation angle from the true vertical direction. To enhance robustness against noise and irregular features, we

aggregated all deviation angles within the image and used the median value as the representative tilt angle, which effectively suppresses outliers caused by non-structural edges or occlusions. Once the tilt angle was determined, we corrected the vertical distortion by rotating the entire SVI around its center using the warpAffine function in OpenCV. This affine transformation realigns the vertical structures, ensuring that the corrected image approximates a true horizontal viewing geometry. During the rotation process, blank boundary areas inevitably appear due to the transformation. To maintain visual continuity and avoid introducing artificial patterns, we filled these areas using the nearest pixel values from surrounding regions. This interpolation method preserves the local texture and color consistency of the original image.

3.3. *Two-level road slope estimation*

After obtaining the deskewed SVI, we estimated the road slope in two-level, i.e., point-level and segment-level. In this stage, three attributes of road slope were derived: point-level slope, segment-level slope, and segment-level relief. The main idea is to first infer the local slope at each sampled point by modeling the variations in the coordinates of points along the road edge within the SVI, and then aggregate these point-level estimates to characterize the overall slope and relief of entire road segments, as shown in Figure 5.

3.3.1. Point-level slope estimation

For each sampling point, we estimated the road slope using the corresponding two side view deskewed perspective images at that point. Similar to the deskewing step, we used the Mask2Former model to generate a road mask through semantic segmentation. Because the segmentation model produces separate classes for the main road surface, road markings, and crosswalks, we merged these three classes to obtain a single binary road mask. To fill small holes in the mask, we applied a morphological closing operation, which is a dilation followed by an erosion. We then extracted the uppermost and largest contour of the closed mask to approximate the road edge. The upper boundary of the contour is converted to candidate points. A straight line was fitted to these points using the RANSAC algorithm. RANSAC iteratively selects random subsets of the data, fits a model to these hypothetical inliers, and tests all points against the model. The consensus set of inliers is used to update the model parameters. This procedure is robust to outliers because the fitted line is determined primarily by points that agree with the model. We configured the maximum number of iterations and the distance threshold in the scikit-learn implementation and computed the slope of the final fitted line as the arctangent of its angle relative to the horizontal axis. Each sampling point has two side-view images and thus yields two independent slope estimation values. To mitigate the influence of occlusion and noise, we combine the two estimates using a weighted average, where the weight is proportional to the road area in the corresponding mask, giving higher confidence to masks with larger road area. Notably, the absolute value of the slope was used for the final estimation, as the slope direction depends on the camera's viewing orientation.

3.3.2. Segment-level slope and relief estimation

Each road segment consists of multiple sampling points. To calculate the segment-level road slope, we aggregated the point-level slope estimates within each segment.

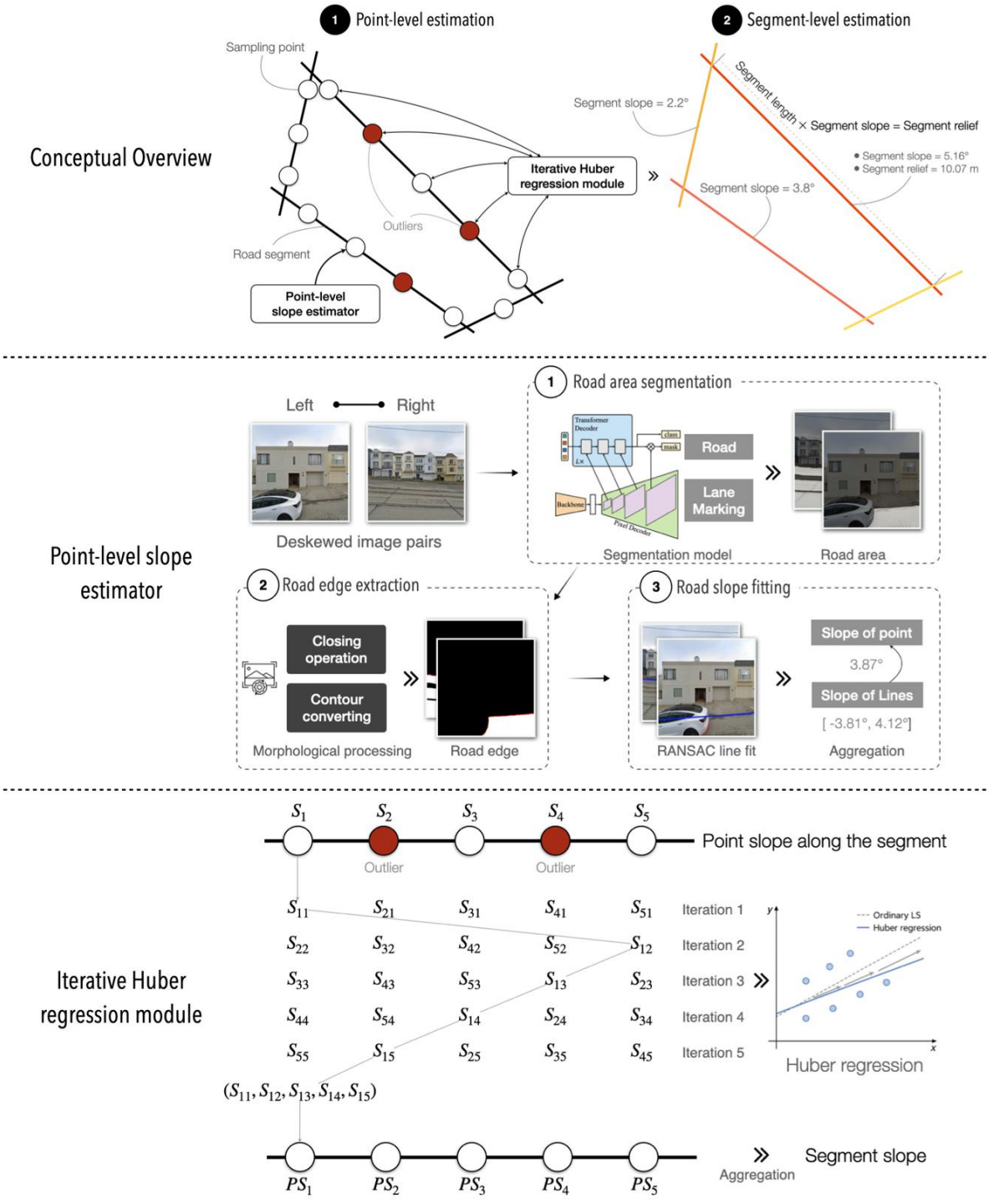

Figure 5. Workflow of two-level road slope estimation. Source of imagery: Google Street View.

Rather than directly using a simple arithmetic mean, which is highly sensitive to extreme values, the segment-level slope was derived from aggregated point-level slopes with explicit treatment of outliers. Although a median-based estimator could provide robustness against extreme outliers, it treats all retained observations equally and ignores systematic variations among valid point-level estimates caused by local occlusions, partial visibility of road edges, or segmentation uncertainty. Outliers in point-level estimates may bias the segment-level slope and relief. To mitigate this effect while preserving information from valid but heterogeneous observations, we employed an iterative Huber regression to filter point-level slopes. Huber regression is a robust method that effectively handles outliers by combining the advantages of ordinary least squares (OLS) and least absolute deviations (LAD) (Huber 1983). It behaves like OLS for small residuals and like LAD for large residuals.

Considering the uncertainty and potential positional bias of outliers along a street segment, a single robust fit may still be influenced by clustered local artifacts. Therefore, we applied Huber regression iteratively. In each iteration, a sliding window shifted the first point to the end of the sequence, resembling a stacking operation, so that each point occupied different relative positions during regression. The maximum number of iterations was set to the number of points within the segment. After all iterations, the final slope of each point was obtained as the average prediction across iterations. The segment-level slope was then derived as the mean of the filtered point-level slopes along the segment.

The segment-level relief was defined as the elevation difference between the two endpoints of a road segment. Beyond serving as a simple geometric descriptor, the segment-level relief represents the cumulative vertical variation experienced along the road, encapsulating how local slope fluctuations integrate over distance. It provides an intuitive and comprehensive indicator of the terrain's vertical structure, bridging the gap between local slope and the broader topographic context of the street network. The segment-level relief was computed by multiplying the segment-level slope by the segment length calculated by geometry in OSM road data.

4. Experiments and results

4.1. *Experiment setting and data source*

In this study, we selected San Francisco (SF) and Manhattan in New York City (NYC) as our study areas due to their contrasting topography conditions, different urban morphology, and extensive availability of SVI. SF is characterized by its hilly terrain, with numerous steep roads and varying elevations, making it an ideal location to test the effectiveness of our framework. NYC's Manhattan presents a contrasting urban landscape with relatively flat terrain but dense urban infrastructure.

For data sources, we used four main datasets: road networks, building footprints, DEMs, and SVI. The road networks and building footprints in both cities were obtained from OSM. Road networks were used to sample SVI locations and aggregate point-level slope estimates to segment-level metrics. Building footprints were used to calculate the distance from the observation point to the nearest building as a sensitive factor in the accuracy analysis. Regarding DEMs, we selected two resolution DEMs in our study. The availability of high-resolution DEMs (1-meter) from the 3DEP project of the United States Geological Survey in both cities allows for accurate validation of our road slope estimations. Specifically, we sampled elevations at buffered locations

centered on available SVI points spaced at 30 m intervals along the road direction, computed elevation differences between adjacent samples, and derived point-level slopes as the ground-truth reference. It is important to note that this 30 m value refers to the sampling interval, rather than the spatial resolution of the DEM, which remains at meter-level detail. Meanwhile, we calculated the mean values of the point slope within each segment as the reference for ther the segment slope. We accessed this DEM data using Opentopography (OpenTopography 2021). Moreover, we used the global 30 m DEM, Copernicus DEM[7], to compare the performance difference between SVI-based and medium-resolution DEM-based road slope. For the SVI data source, we utilized GSV, which offers extensive coverage, consistent acquisition protocols, and high-quality panoramic views of street scenes.

Regarding the detailed model configuration, we employed the Swin-L backbone of the Mask2Former architecture, pretrained on the Mapillary Vistas dataset comprising over 60 urban scene classes (Neuhold *et al.* 2017). This model was selected for its proven capability in capturing complex street-level semantics, achieving a mean Intersection over Union of 63.2, and for its public availability on the Hugging Face platform. To ensure the stability of geometric feature extraction, the parameters of PPHT were tuned to effectively suppress outlier lines. Specifically, the minimum number of votes and minimum line length were both set to 50, while the maximum gap between line segments was adjusted to maintain consistent local angle estimation. For the subsequent RANSAC-based line fitting, we adopted 1,000 iterations and a residual threshold of 1 to enhance robustness against noise and local irregularities. Meanwhile, considering the uncertainty of some short road edge boundaries with a single point due to parked cars or other elements, we set a minimum node threshold of 10 to ensure that each fitted line segment contains a sufficient number of valid edge points. All experiments were executed on a workstation equipped with an NVIDIA RTX 4090 GPU and 128 GB of system memory, providing sufficient computational capacity for efficient model inference and large-scale evaluation.

4.2. *Evaluation metrics*

We assessed the accuracy of road slope estimation at both point and segment levels. Two types of metrics were used: residual difference and absolute difference. The residual difference is defined as the ground truth minus the estimated value, reflecting systematic bias. The absolute difference is the absolute value of the residual, reflecting overall accuracy. For residual differences, we reported the median error (ME) and skewness to describe error distribution. For evaluating absolute differences, we used Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and symmetric Mean Absolute Percentage Error (sMAPE).

4.3. *Spatial map of estimated road slope*

In total, we collected data for 31,897 and 28,110 points, and 10,740 and 6,720 segments in SF and NYC, respectively. Figure 6 shows the spatial distribution of road slopes in SF, providing a multi-scale view from point-level estimation to aggregated segment-level measures. Panel (a) presents the point-level slope map, revealing fine-grained variations across the urban terrain, with steeper slopes concentrated in several hilly areas. Panel (b) displays the segment-level slope map, where local slope estimates are

[7]https://collections.sentinel-hub.com/copernicus-dem/

smoothed into a more continuous pattern that captures the structural slope characteristics of the street network. Panel (c) illustrates the segment-level relief, defined as the elevation difference between the endpoints of each road segment, which closely corresponds to the city's overall topographic relief. The zoomed-in maps demonstrate that aggregating slopes from point to segment effectively reduces noise while preserving meaningful elevation gradients at the block scale. The histograms below summarize the statistical distributions of slope and relief, with medians of 1.91°, 1.98°, and 4.06 m, respectively.

Figure B1 highlights the relatively flat road slope of NYC. Panel (a) reveals subtle variations mainly concentrated along the northern and southern edges of the island where terrain relief occurs. Panel (b) clearly delineates slightly elevated corridors in Upper Manhattan. Panel (c) mirrors the city's gentle but structured relief pattern. The histograms at the bottom depict the statistical distributions of slope and relief values, with median values of 1.12°, 1.27°, and 2.42 m, respectively, indicating that most roads are nearly flat with modest elevation differences.

4.4. *Performance assessment*

We evaluated a large number of diverse samples in the two cities. In SF, the assessment included 8,893 sampling points and 1,896 road segments, while in NYC, 7,103 sampling points and 1,820 road segments were analyzed. The data distributions of the estimation results and their deviations from ground truth are shown in Figure 7 and Figure B2.

4.4.1. Results in San Francisco

In SF, the results show strong consistency with the reference data at both point and segment levels. Regression analyses yield slope estimates around 0.8 and Pearson's r above 0.9 (Figure 7), confirming high overall accuracy in this hilly environment. The residuals are slightly right-skewed, with small mean errors (ME $< 0.25°$ or 0.25 m), indicating a mild underestimation. Stratified by slope intervals, the model slightly overestimates gentle slopes (0–2°) but underestimates steeper ones (¿ 5°), consistent with shifts in residual skewness. As summarized in Table 2, the RMSE and MAE values (approximately 1° and 2–3 m) demonstrate strong agreement with ground truth, while sMAPE values near 38% indicate reliable performance across varying terrains. Although absolute errors increase for steeper slopes, the relative accuracy improves, suggesting good robustness in complex topography. Additionally, we examined the effect of filtering on estimation performance in SF (Figure A1) and analyzed the relationship between segment-level slope and the coefficient of variation of slope estimates (Figure A2). Overall, these supplementary analyses confirm that the proposed filtering strategy improves within-segment slope consistency and that estimation variability systematically decreases with increasing segment-level slope.

4.4.2. Results in New York City

In contrast, the results in NYC show moderate but consistent accuracy, reflecting challenges posed by its flat terrain and dense built environment. Correlations (Pearson's r = 0.53–0.74) remain reasonable across scales, with the strongest agreement observed in segment-level relief (Figure B2). Residuals are nearly symmetric for point-level slope but left-skewed for segment estimates, with negative MEs suggesting a tendency toward overestimation. This bias diminishes as slope estimates increase, indicating that

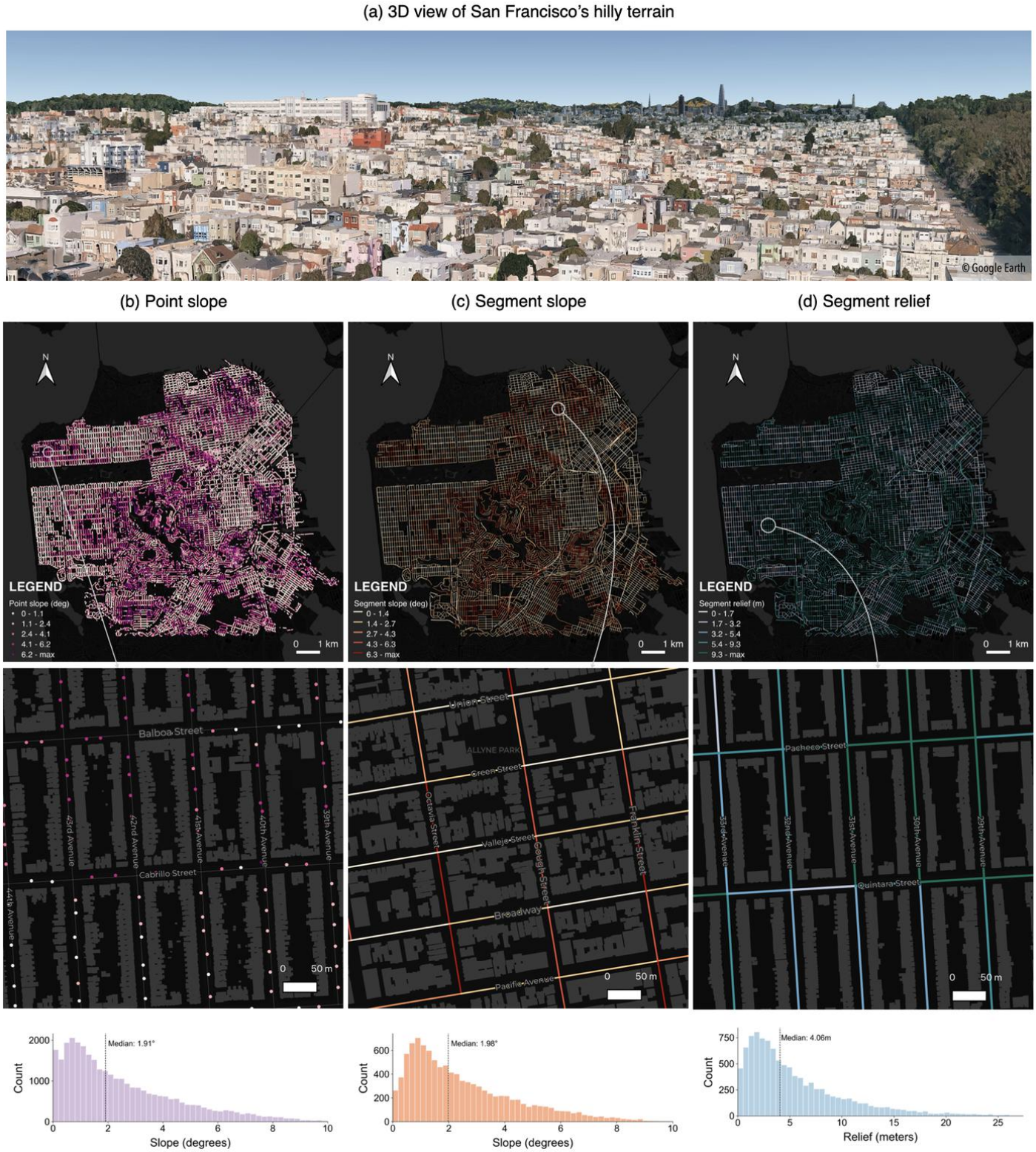

Figure 6. (a) A 3D scene illustrating the hilliness in SF (b) Point-level road slope estimation, (c) Segment-level slope map. (d) Segment-level relief map. Basemap: CARTO and OpenStreetMap contributors.

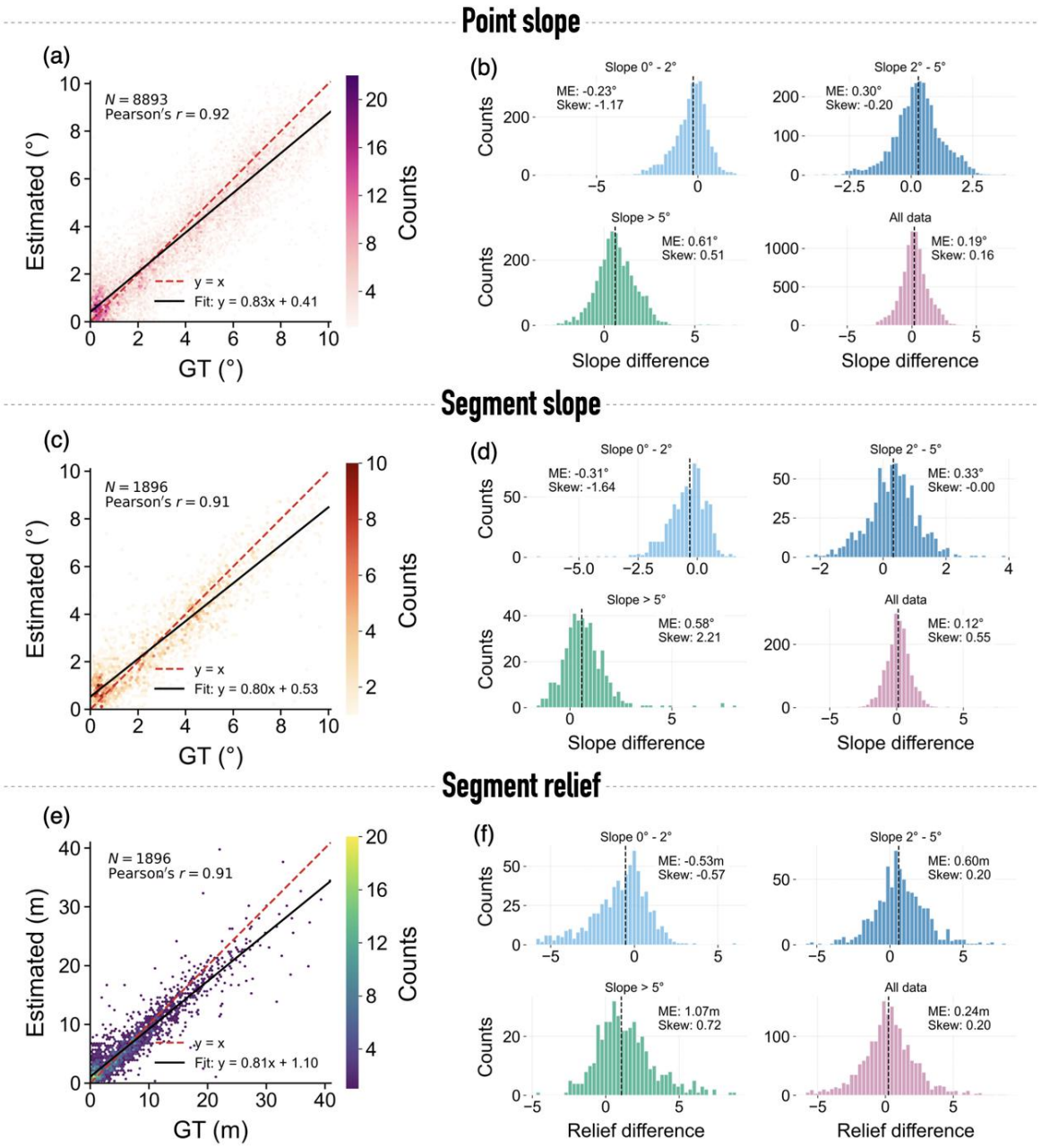

Figure 7. Comparison of point-based and segment-based slope and relief estimations against ground truth (GT) in San Francisco. (a, c, e) Scatter plots between estimated and GT values with regression lines (black) and 1:1 lines (red). Color scales represent point density. (b, d, f) Histograms of estimation differences grouped by slope or relief intervals: 0–2°, 2–5°, > 5°, and all data. Median and skewness values are shown for each distribution.

Table 2. Absolute error of slope and relief estimation in San Francisco.

Metric	Range	RMSE	MAE	sMAPE (%)	N
Point slope (°)	All	1.09	0.81	38.63	8893
	0–2°	0.91	0.65	74.30	2832
	2–5°	1.01	0.78	27.70	3017
	> 5°	1.30	1.00	16.29	3044
Segment slope (°)	All	1.01	0.72	38.49	1896
	0–2°	0.94	0.66	69.27	702
	2–5°	0.85	0.66	22.95	761
	> 5°	1.33	0.93	15.91	433
Segment relief (m)	All	2.86	1.70	37.36	1896
	0–2°	3.16	1.63	66.41	702
	2–5°	2.28	1.53	22.74	761
	> 5°	3.24	2.11	15.95	433

the framework performs better in relatively steeper urban areas. As shown in Table B1, RMSE and MAE values are generally below 1° for slope and around 2 m for relief. Although the sMAPE values are higher due to NYC's dominance of gentle slopes (this point is discussed in Section 5.2), the relative accuracy improves notably in higher slope intervals.

4.5. *Conditional bias analysis*

The accuracy of road slope estimation can be influenced by multiple factors, such as road characteristics and street openness. To systematically assess these influences, we conducted a conditional bias analysis by stratifying estimation errors according to key variables. Sampling points in SF were used for this analysis due to the city's diverse topography and overall strong model performance.

Figure 8 illustrates the effects of road pixel count and car pixel count on point-level slope estimation. The road pixel count represents the visible road area in SVI, which is critical for semantic segmentation and road edge detection. The car pixel count reflects the extent of vehicle presence, which may occlude the road surface and hinder slope estimation. Both variables were divided into ten deciles, and the MAE was calculated for each combination. The results indicate that higher car pixel counts are associated with increased MAE, suggesting that vehicle occlusion degrades estimation accuracy (Figure 8(a)). Interestingly, very low car pixel counts also correspond to higher MAE (Figure 8(b)), likely because these cases often occur on narrow or partially occluded roads where the visible road area is limited. In contrast, higher road pixel counts generally lead to lower MAE, implying that larger visible road areas enhance the accuracy of slope estimation.

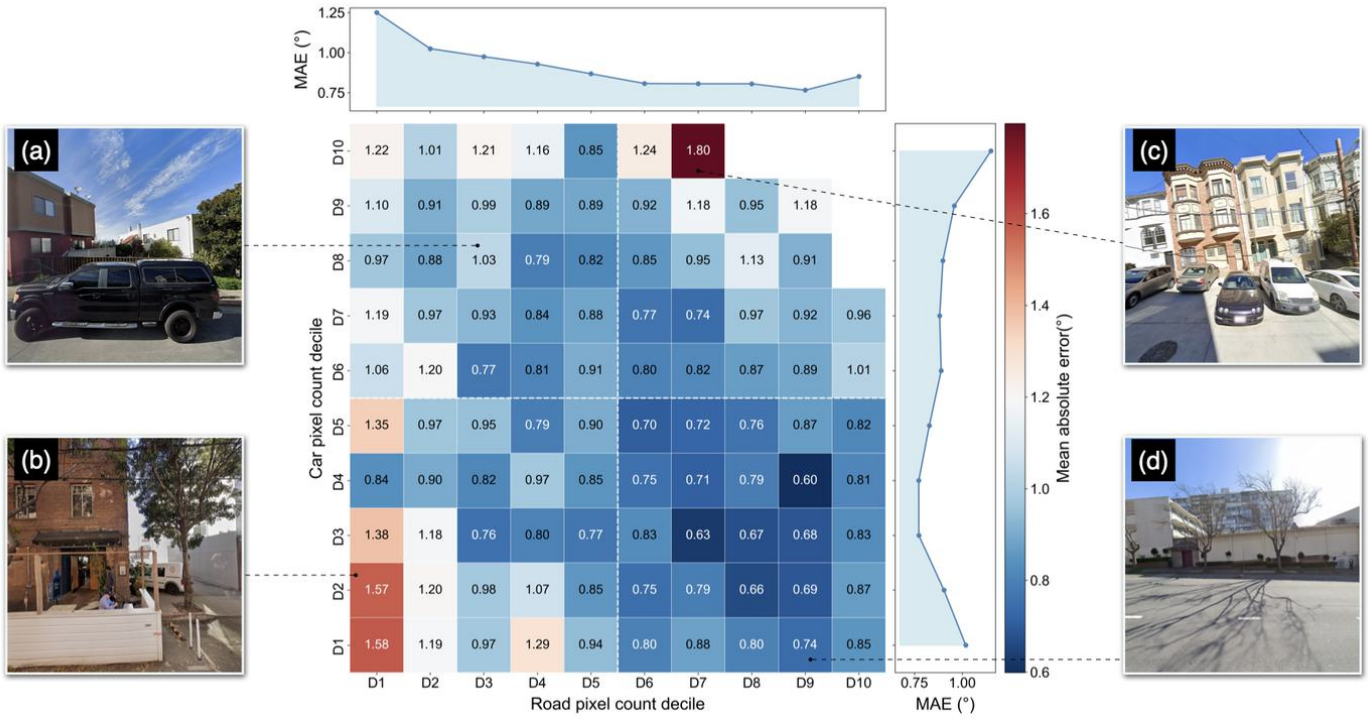

Figure 8. Effect of road pixel count and car pixel count on road slope estimation. MAE is shown across deciles of road pixel count (horizontal) and car pixel count (vertical). Blue indicates lower error; red indicates higher. Marginal MAE trends for each variable appear in the top and right plots. Sample images (a–d) illustrate representative cells. Source of imagery: Google Street View.

The interaction between these two factors is also notable. The lowest MAE occurs when the visible road area is large and vehicle presence is minimal (Figure 8(d)). Conversely, both the upper-right corner (high road pixel count and high car pixel count) and the lower-left corner (low road pixel count and low car pixel count) of the response surface exhibit higher MAE. These patterns suggest that insufficiently visible road edges whether due to vehicle obstruction or limited road exposure, can substantially increase estimation error. Figure 8(c) exemplifies such a case, where cars obstruct portions of the visible road surface despite a relatively large overall road area. Deskewing is a critical step in our framework, as it corrects tilt distortions in SVI images that can significantly affect slope estimation. Two key factors influence the deskewing performance: the number of structural lines detected and the distance from the viewpoint to the nearest building. The structural line count reflects the availability of vertical features used to estimate the tilt angle, while the building distance indicates the proximity of stable reference structures. To quantitatively analyze their effects, we divided both factors into deciles and calculated the MAE for each combination. The results reveal that both excessively close and distant buildings lead to higher MAE. In contrast, a greater number of structural lines generally corresponds to lower MAE, indicating that abundant vertical features enhance tilt angle estimation and, consequently, slope accuracy.

As illustrated in Figure 9(a) and (c), well-ordered building facades with sufficient structural lines yield more stable corrections. Conversely, scenarios with few detectable

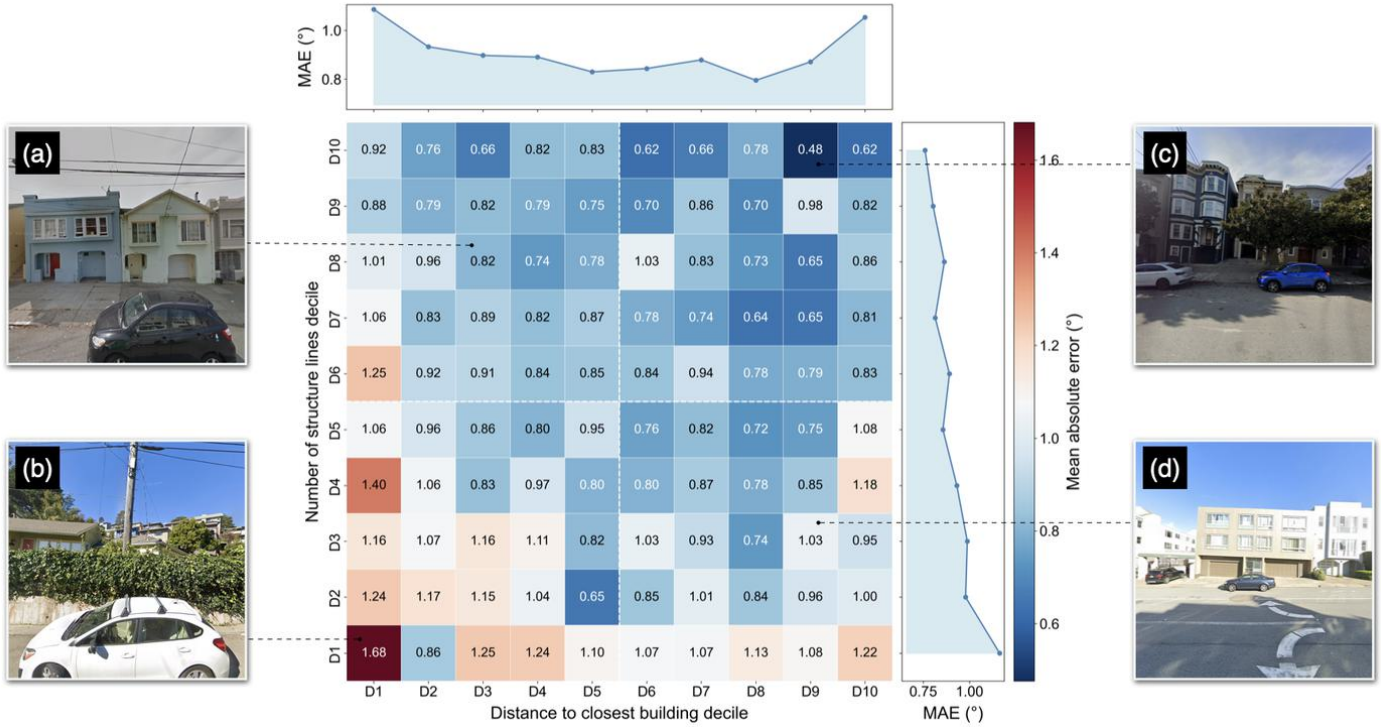

Figure 9. Effect of building distance and structure line count on road slope estimation. Source of imagery: Google Street View.

lines exhibit higher MAE for multiple reasons. When viewpoints are too close to irregular or vegetated facades, the scarcity of clear vertical features increases uncertainty, as shown in Figure 9(b). When viewpoints are too far and vertical lines become short or indistinct, tilt angle estimation also becomes unreliable, as shown in Figure 9(d). For other intersections of built-environment factors (between building proximity with car area and road area), we conducted similar stratified analyses and observed consistent patterns in Figure C1 and Figure C2, with the lowest estimation errors concentrated in moderate spatial contexts and higher errors emerging primarily at extreme combinations of building proximity and surrounding area characteristics.

4.6. *Comparative analysis with DEM-derived road slope*

To benchmark the performance of the Vision2Slope framework, we compared its results with road slope estimates derived from the Copernicus DEM (30 m resolution), a widely used medium-resolution dataset. The comparison adopted the same point-level and segment-level metrics as in our primary evaluation. DEM-derived slopes were obtained by extracting elevation values at the sampling points and along road segments, followed by calculating slope and relief. Both quantitative metrics and qualitative terrain profile comparisons were used to evaluate performance differences.

As summarized in Table 3, Vision2Slope substantially outperformed the 30 m DEM in both cities. In SF, point-level slope estimation achieved an RMSE reduction from 2.34° to 1.09° (53.23%) and an MAE reduction from 1.42° to 0.81° (42.80%). At the segment level, RMSE decreased from 1.47° to 1.01° (31.48%) and the relief RMSE from 4.23 m to 2.86 m (32.48%). The MAE also improved by 16.71% for slope and 13.34%

for relief, confirming the robustness of Vision2Slope in extreme cases.

In New York City, the improvement was even more pronounced. The point-level slope RMSE dropped from 8.67° to 0.71° (91.86%) and MAE from 5.67° to 0.57° (89.95%). Segment-level slope and relief RMSEs decreased from 7.40° to 1.00° (86.54%) and from 19.17 m to 2.36 m (87.67%), respectively. MAE and sMAPE reductions followed similar trends. These results demonstrate the superior accuracy and adaptability of Vision2Slope, particularly in dense, high-rise urban environments where medium-resolution DEMs often suffer from data gaps and interpolation artifacts.

Table 3. Performance comparison with 30 m DEM results in SF and NYC.

City	Level	Metric	Vision2Slope	30 m DEM	Improvement
SF	Point slope	RMSE	**1.09**	2.34	53.23%
		MAE	**0.81**	1.42	42.68%
	Segment slope	RMSE	**1.01**	1.47	31.48%
		MAE	**0.72**	0.85	14.32%
	Segment relief	RMSE	**2.86**	4.23	32.48%
		MAE	**1.70**	1.96	13.34%
NYC	Point slope	RMSE	**0.71**	8.67	91.86%
		MAE	**0.57**	5.67	89.95%
	Segment slope	RMSE	**1.00**	7.40	86.54%
		MAE	**0.78**	5.52	85.86%
	Segment relief	RMSE	**2.36**	19.17	87.67%
		MAE	**1.66**	12.57	86.81%

The visual comparisons further demonstrate the superiority of our framework. Figures 10 and 11 present the comparison between our estimation results and the 30 m DEM-derived values for point slope, segment slope, and relief under different urban contexts. In Figure 10, we compare our results with the 30 m DEM and reference data (1 m DEM) in two representative gentle-slope scenarios, (a) streets lined with high-rise buildings and (b) streets partially covered by tree canopies. In both cases, the slope variation along the road is plotted against distance, showing that our estimates (orange line) closely follow the reference slopes (green line), whereas the 30 m DEM (blue line) consistently overestimates due to coarse spatial resolution and vertical smoothing. In the high-rise buildings area (a), the estimated segment slope (0.36°) closely matches the reference (0.38°), while the 30 m DEM produces a much higher value (9.42°). The corresponding segment reliefs further confirm this agreement. In the tree-canopy case (b), our result yields a slope of 0.70°, again close to the reference (0.49°) and far lower than the DEM-derived 2.01°. Although the relief values remain within a reasonable range, the differences are substantially smaller than those from the 30 m DEM. These examples illustrate that our method has better performance in dense urban contexts because of the advantages of observation perspective. In SF (Figure 11), our results also outperform the 30 m DEM-derived values for open streets

with varying slopes. For the moderate-slope street (a), the estimated segment slope (2.71°) aligns well with the reference (3.31°), while the DEM overestimates it at 4.62°. The corresponding reliefs (6.89 m estimated vs. 8.34 m reference) show a consistent magnitude. In the steep-slope case (b), Vision2Slope maintains close agreement with the reference (6.21° vs. 6.25°), while the DEM slightly underestimates (6.04°). The relief values (9.90 m estimated vs. 9.73 m reference) further validate the robustness of our model.

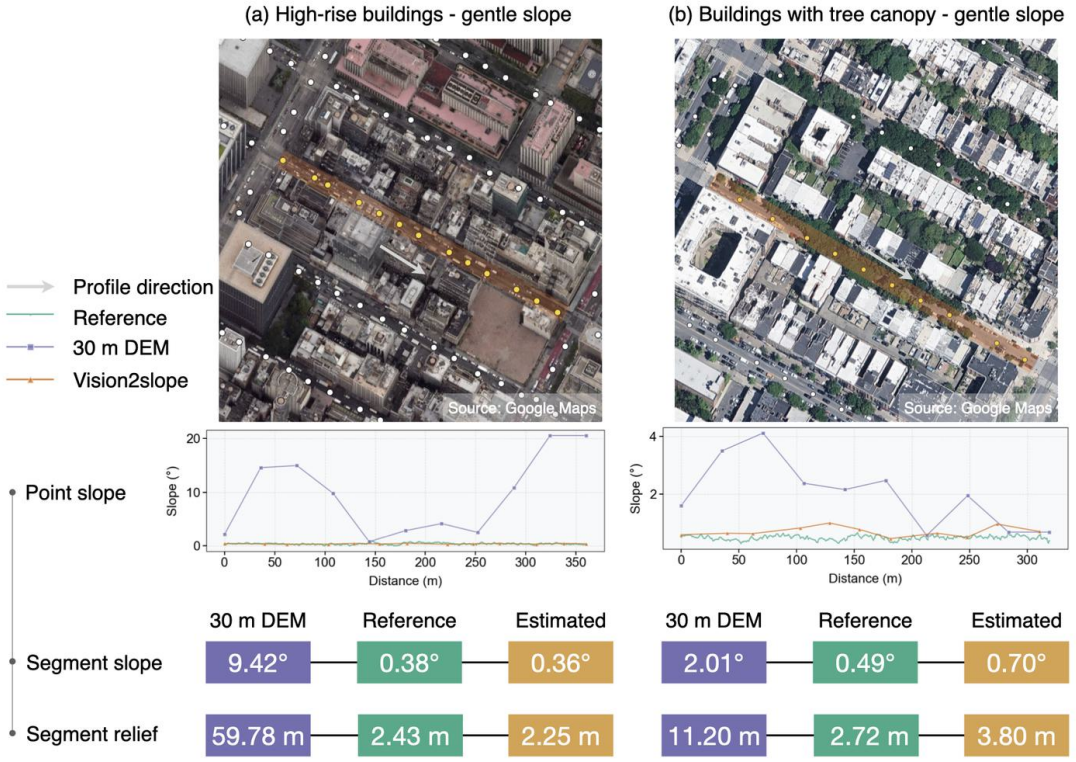

Figure 10. Comparison of slope estimation under different urban conditions. (a) High-rise buildings, gentle slope. (b) Tree canopies, gentle slope. The top panels show profile directions overlaid on Google Maps, with slope profiles plotted below, comparing Vision2Slope estimates (orange), reference measurements (green), and 30 m DEM-derived values (blue).

4.7. *Sensitivity and uncertainty assessment*

The quality of SVI data can substantially influence the accuracy of road slope estimation. Factors such as camera mounting position, lighting conditions, and image ratio may introduce distortions and noise into the imagery. In particular, crowdsourced SVI datasets often exhibit large variations in image quality due to differences in cameras, weather conditions, and data collection methods (Biljecki *et al.* 2023).

To evaluate the robustness of our framework, we conducted a sensitivity simulation to examine how various image degradation scenarios affect slope estimation accuracy. Four types of degradation were considered: pitch variation, roll variation, exposure variation, and FOV variation. For each type, the corresponding parameter was sys-

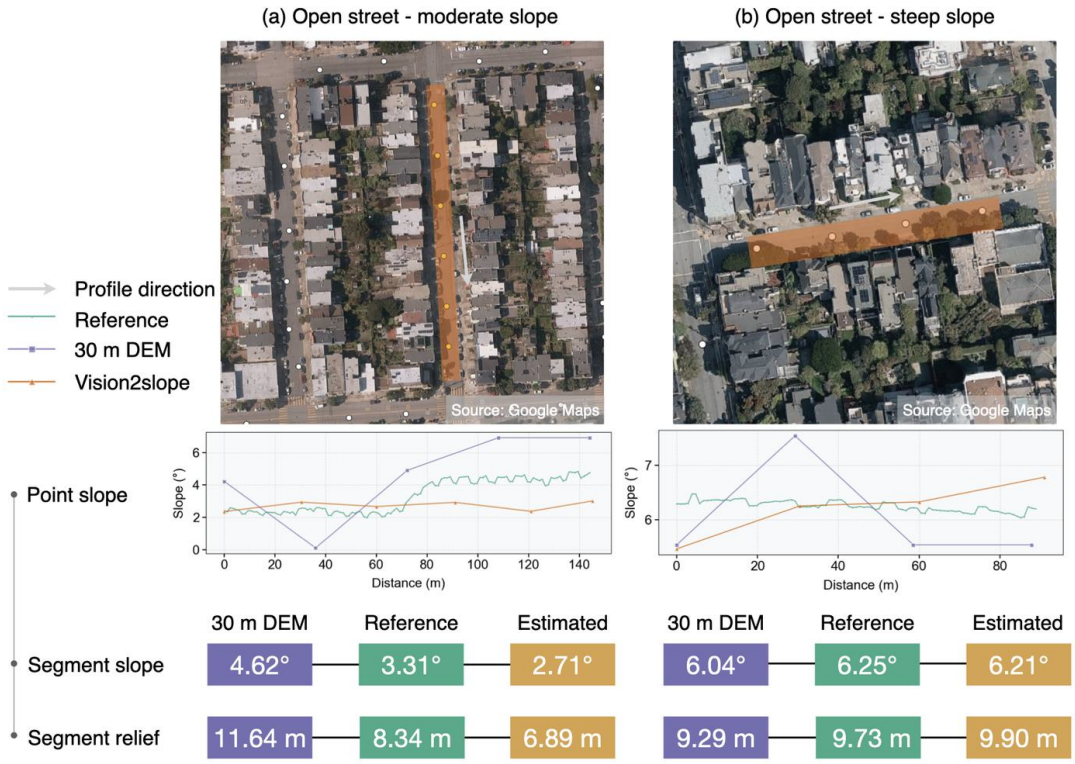

Figure 11. Comparison of slope estimation under different urban conditions. (a) Open street, moderate slope. (b) Open street, steep slope.

tematically adjusted within a realistic range, and the resulting changes in estimation accuracy were assessed using 1,000 high-quality sample points.

For geometric degradations, namely pitch and roll, the angles were varied from -30° to 30° in increments of 5°. These variations reflect realistic changes in road conditions (e.g., slope and curvature) and camera mounting positions, which can alter the visible proportions of buildings and road surfaces in the imagery. To simulate these effects, we modified the transformation matrix parameters during the perspective projection of panoramic images. The matrix is defined as Eq. 1. In this process, θ_x, θ_y, and θ_z correspond to the rotation angles around the camera's principal axes, controlling pitch, yaw, and roll, respectively. By varying θ_x and θ_z within realistic ranges, we simulated different viewing angles to assess the robustness of slope estimation under non-ideal imaging conditions.

$$\mathbf{R} = R_x R_y R_z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix} \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

Figure 12 illustrates the results under different roll angle conditions. A negative roll angle indicates that the side view image contains more road area, whereas a positive roll angle corresponds to more building area. As the roll angle increases, slope estimation errors also rise. Although the median error remains relatively stable, the

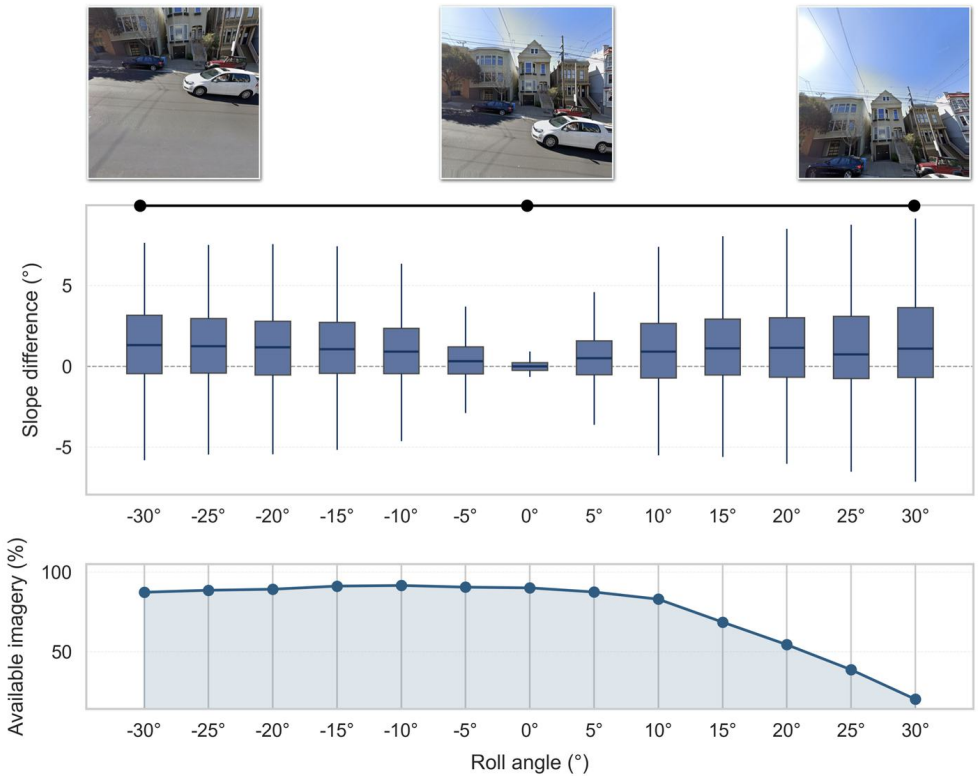

Figure 12. Sensitivity analysis of slope estimation under varying roll angle conditions. Boxplots show slope estimation errors, and line plots the proportion of available imagery. Source of imagery: Google Street View.

interquartile range (IQR) and overall error distribution widen substantially, indicating greater variability and uncertainty in the estimates. Furthermore, the proportion of usable imagery decreases with increasing roll angle, suggesting that extreme roll angles often lead to slope estimation failures due to limited road visibility. In contrast, negative roll angles do not markedly affect estimation availability, despite a moderate increase in errors.

Figure 13 shows the sensitivity analysis under different pitch angle conditions. Large positive or negative pitch angles correspond to more distorted images. The results indicate that pitch variation has a limited effect on slope estimation errors and data availability in our framework. The median error, IQR, and overall error distribution remain stable across the tested range, and the proportion of usable imagery exhibits no consistent trend with pitch changes. At extreme pitch angles ($> 20°$ or $< -20°$), errors increase slightly and availability decreases marginally, though such cases are uncommon in real-world SVI datasets. Overall, these findings suggest that our framework is robust to pitch variations, likely due to the effectiveness of the image deskewing and road edge detection steps in mitigating geometric distortions.

For exposure variation, the brightness of SVI images can be influenced by lighting conditions such as time of day and weather. Poor illumination often results in low contrast and increased noise, which may degrade the performance of semantic segmentation and road edge detection. To simulate exposure variation, we adjusted image brightness by scaling pixel values across five exposure levels. For FOV variation, the difference mainly arises from the transformation settings used in processing panoramic

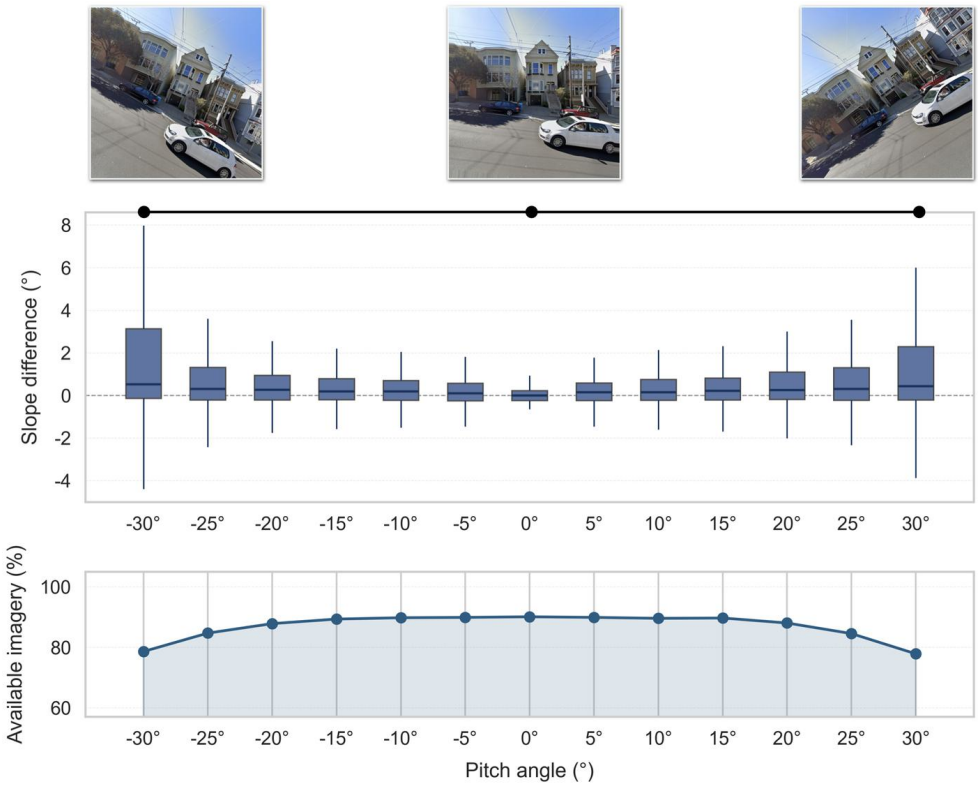

Figure 13. Sensitivity analysis of slope estimation under varying pitch angle conditions. Source of imagery: Google Street View.

SVIs. To simulate FOV effects, we varied the FOV from 60° to 150° in increments of 10° during the perspective projection of panoramic images.

As shown in Figure 14 and Figure 15, slope estimation remains stable across moderate changes in both factors. For FOV, the method maintains high accuracy and data availability within the 60°–120° range, while wider angles introduce slight performance degradation due to lower effective resolution in the road region. For exposure, the framework is robust to moderate brightness perturbations: performance decreases only marginally under underexposed or overexposed conditions, though extreme underexposure noticeably reduces data availability due to missing road surface features. These results indicate that the Vision2Slope framework is resilient to common SVI quality variations.

5. Discussions

5.1. *Unlocking the power of SVI in road slope estimation and filling data gaps*

Although satellite remote sensing continues to make great strides in extracting road planar boundaries (Zhou *et al.* 2025), 3D road attributes are still largely neglected. Our framework demonstrates how SVI can be converted into quantitative estimates of road slope at both point and segment levels. Beyond methodological innovation, this study illustrates how slope estimation can be scaled to entire cities, offering a cost-effective alternative in regions where high-resolution DEMs or LiDAR data are unavailable or

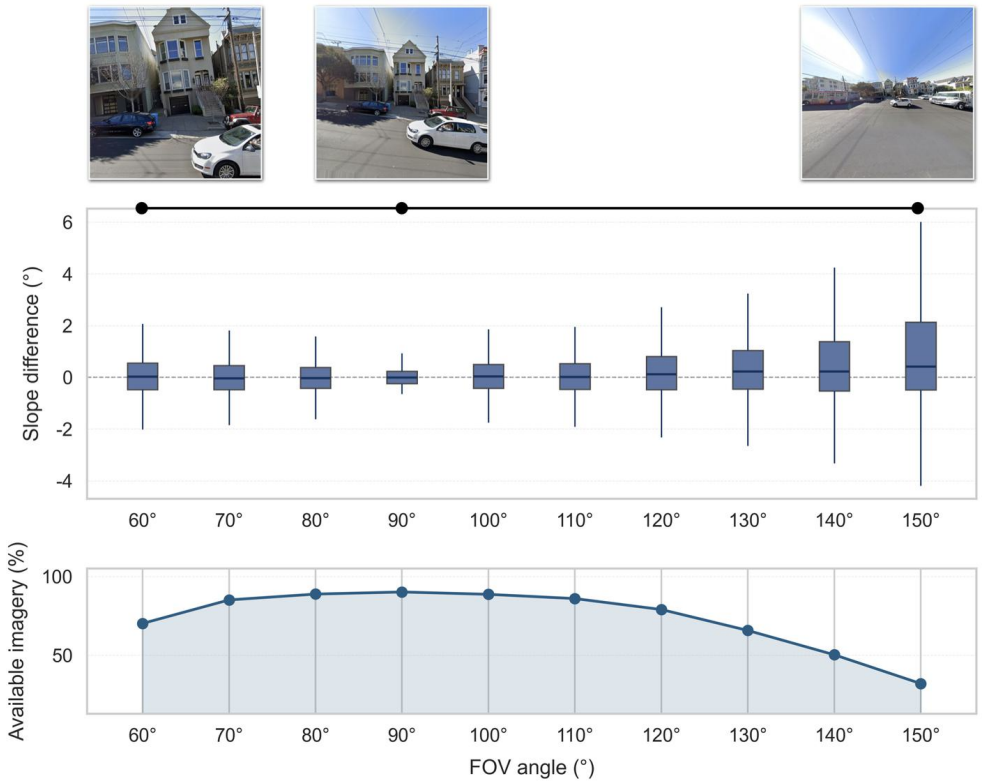

Figure 14. Sensitivity analysis of slope estimation under varying FOV angle conditions. Source of imagery: Google Street View.

prohibitively expensive. The comprehensive validation across two contrasting urban environments further underscores the robustness and generalizability of SVI-based slope estimation. Moreover, although some commercial SVI providers (e.g., Google, Tencent) restrict data redistribution, the proposed approach still enables researchers and individual users to obtain refined 3D road information based on our framework. In contrast, some open SVI platforms in data-rich regions are already available and well-maintained, thereby supporting the extension of road network data enrichment and 3D urban analytics.

More broadly, harnessing the potential of SVI for road slope extraction provides meaningful implications for smart land surveying in the era of artificial intelligence. This approach bridges traditional geodetic surveying with data-driven urban analytics, enabling scalable, low-cost, and human-centered characterization of urban terrain. It is especially beneficial for (i) regions with scarce or outdated high-resolution topographic data, particularly in developing contexts (e.g., La Paz and Nairobi), and (ii) regions where access to high-resolution terrain data is restricted due to licensing, commercial, or security constraints (e.g., Singapore and Seoul), where traditional data sources are unavailable or difficult to obtain (Figure 16). In addition, SVI-based approaches are advantageous in dense urban cores with complex vertical structures or heavy vegetation cover, where traditional remote sensing methods often perform poorly (Osama *et al.* 2023). For instance, using our pipeline, the slope at a given location can be estimated within a few seconds anywhere in the world where suitable imagery is available, including many data-scarce regions.

Notably, in some less developed regions, informal or low-level roads are more preva-

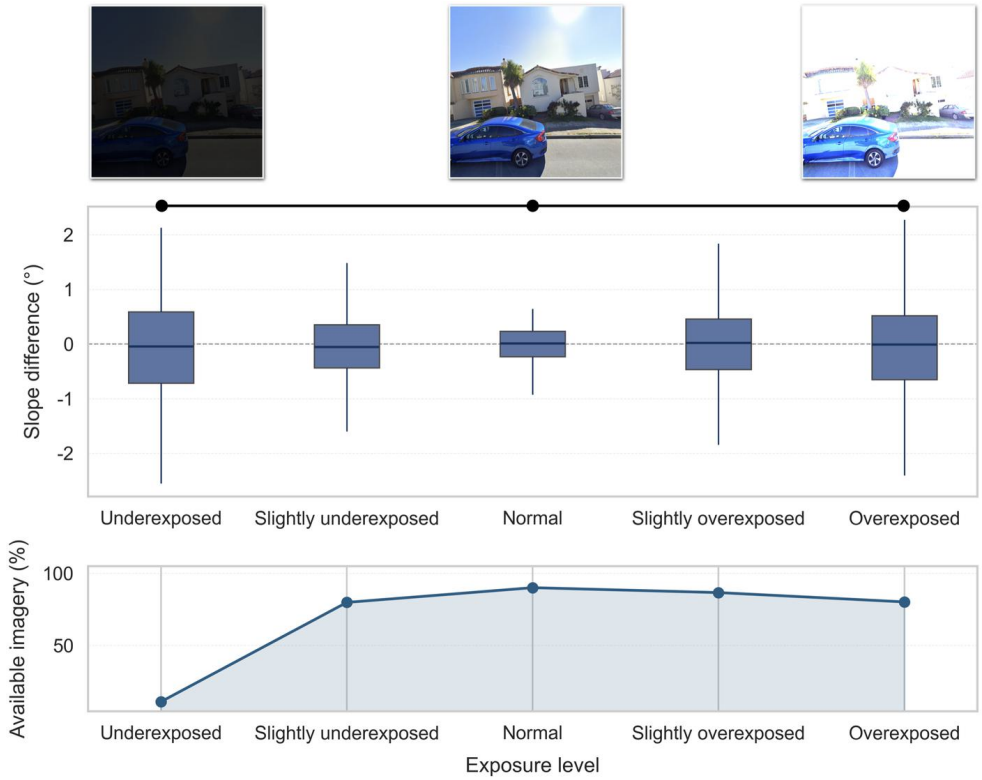

Figure 15. Sensitivity analysis of slope estimation under varying light exposure conditions. Source of imagery: Google Street View.

lent, which may degrade the performance of our method due to limited visibility in narrow and dense environments, as well as misclassifications of road surfaces and curbs. For narrow streets, we conducted a preliminary test in Hong Kong and found that adjusting the effective FOV can substantially increase the visible extent of road surfaces and curbs (Figure D1), partially mitigating occlusion effects and improving the robustness of slope estimation. Moreover, we systematically examined the sensitivity of the method to different FOV settings in Section 4.7 and found that estimation performance is largely insensitive to reasonable variations in FOV. Although recent studies have made progress in unpaved road classification using remote sensing and SVI (Randhawa *et al.* 2025, Liu *et al.* 2026), accurately delineating road area and edges on unpaved roads remains challenging and represents an important direction for future research. Meanwhile, the availability of high-quality SVI remains a key constraint in less developed regions, even though SVI-based processing can offer a low-cost alternative to traditional topographic access. Accordingly, the applicability of Vision2Slope is inherently bounded by the availability of high-quality street view imagery and high-resolution DEMs and should not be overgeneralized to data-scarce regions. Nevertheless, it provides a methodological foundation that is well positioned to unlock the power of the increasingly accessible SVI in 3D cities.

The framework also offers a pathway for enriching global geospatial road datasets, such as OSM or the Global Roads Inventory Project (Meijer *et al.* 2018). With SVI-derived metrics, road attributes can extend beyond conventional two-dimensional representations toward richer three-dimensional descriptors to provide a more complete depiction of urban mobility infrastructure.

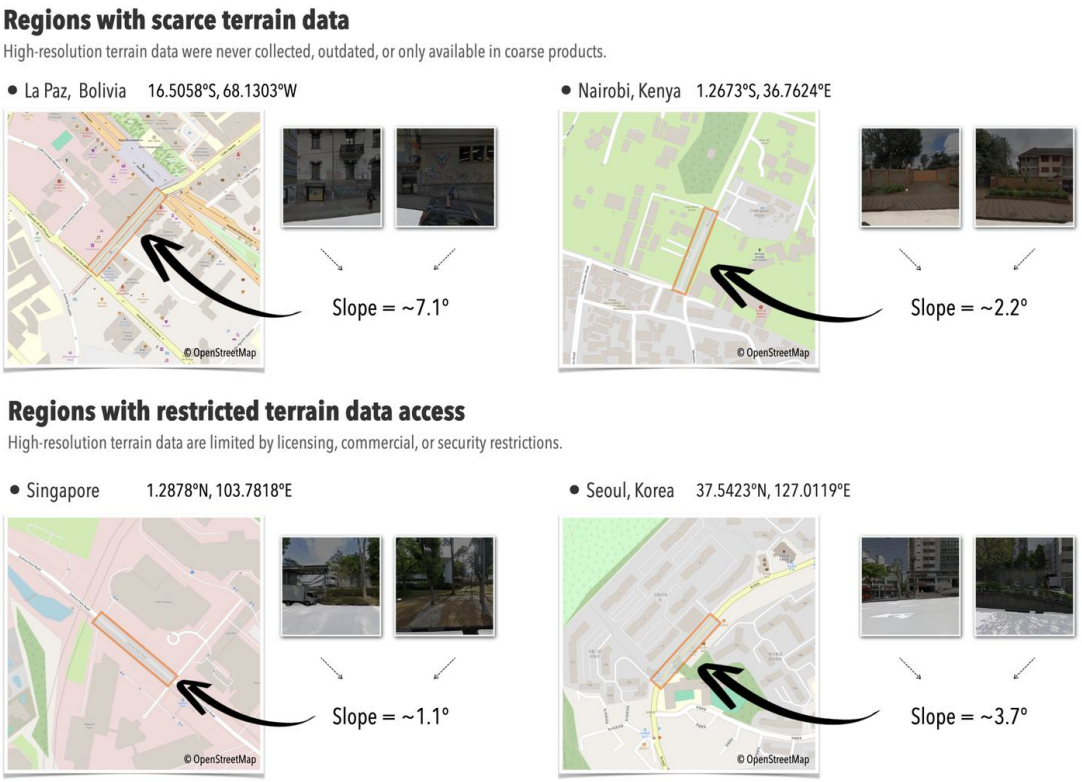

Figure 16. Estimation examples in cities in less developed and developed countries where high-resolution terrain data remain unavailable or inaccessible, demonstrating the versatility and global applicability of our work. Source of imagery: Google Street View.

5.2. *Limitations in gentle slopes and crowded streets*

Although our framework provides an automatic workflow for urban road slope estimation, the relative error remains high in areas with gentle slopes, particularly in NYC, as shown in Table B1. The main reasons for this limitation are summarized as follows.

On the one hand, the distinctive characteristics of NYC's street environments, both densely built-up and relatively flat, pose major challenges. As illustrated in Figure 17, streets in NYC are typically crowded, and almost every SVI contains parked cars along the roadside. These occlusions obscure road edges, which are key features in our slope estimation process. This issue is especially severe in NYC, where most streets are lined with vehicles, leading to an overestimation of slope in flat areas. Meanwhile, some abnormal bumps the of road surface and other urban infrastructure may alter the shape of the road edge, which causes overestimation in flat regions. Although our method effectively reduces outliers in detected road edges through robust regression, some unavoidable errors still persist.

Additionally, the general flatness of NYC's terrain further amplifies relative errors. Because true slopes are small, even minor deviations can result in seemingly large percentage biases. It is worth noting that Manhattan, in particular, has undergone extensive human alteration to flatten its natural terrain. Despite the less satisfactory accuracy in gentle slopes, our results in moderate and steep areas remain robust and perform promisingly.

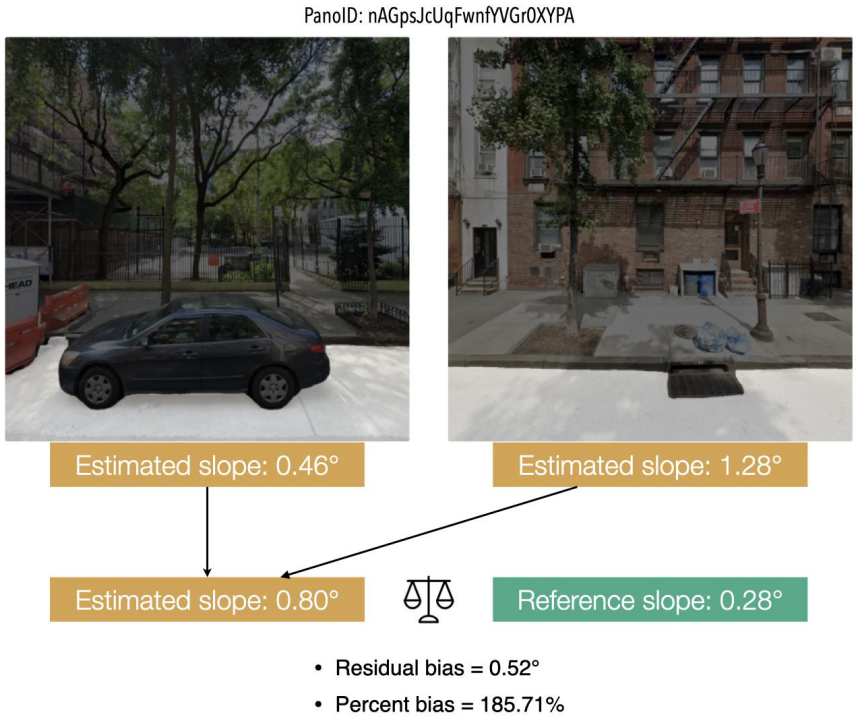

Figure 17. Example of slope estimation error in NYC. Road surface by semantic segmentation is highlighted. Source of imagery: Google Street View.

On the other hand, uncertainties rooted in the SVI data itself also influence the accuracy of slope estimation. Variations in image quality, camera conditions, and the reliability of segmentation models all introduce additional noise (Hou *et al.* 2024). These issues tend to be amplified in crowdsourced imagery, where data are captured using different devices, under inconsistent lighting and weather conditions, and without standardized protocols. As emphasized in Section 4.7, within our framework, only in a small number of situations do low-quality images impair geometric inference, and even state-of-the-art segmentation models can introduce their own systematic biases.

5.3. *Future directions*

Future research could advance both technical development and practical application. For the technical aspects, first, incorporating time-series SVI data could reduce random errors and enhance slope estimation accuracy through temporal averaging. Time-series imagery captures variations in road conditions and image quality, including moments with clear visibility and optimal exposure, thus improving the reliability of slope estimates. Second, although an advanced semantic segmentation model was used, further improvements may be achieved by fine-tuning existing models or adopting Vision–Language Models (VLMs) to better recognize road features in complex urban scenes (Lu *et al.* 2025, Liang *et al.* 2025). However, the regression components of VLMs still introduce uncertainty, which may affect estimation accuracy (Xue and Zhu 2025). Third, developing ensemble predictors that combine SVI-based estimates with other urban structural variables has the potential to further enhance performance. As discussed in Section 4.5, urban features such as building and road areas significantly

influence estimation accuracy. Integrating these variables into a unified model could help correct systematic biases and improve performance. Meanwhile, as discussed in Appendix D, increasing the effective FOV can improve slope estimation on narrow streets by increasing the visible road area, but this adjustment is currently applied manually. Future ensemble predictors could explicitly encode FOV-related visibility as a model component, enabling automatic adaptation to varying street widths and reducing reliance on human intervention.

For the applications, our results demonstrate that SVI can provide reliable road slope information in urban environments, offering broad potential for urban digital twins and sustainable city practices. Integrating slope outputs into hydrological models may strengthen flood risk assessments and improve resilience to increasingly frequent extreme weather events (Rong *et al.* 2020) and disaster response (Li *et al.* 2025). Road slope data can support environmental risk management, particularly in assessing landslides and gully erosion (Zhou 2025). Since slope strongly affects runoff and soil stability, especially in unpaved or peri-urban zones (Mawe *et al.* 2025), incorporating it into geohazard models can improve the detection of vulnerable areas and guide preventive measures to protect infrastructure and communities. Road slope information is also essential for transportation planning and management (Chen *et al.* 2026). Because slope affects vehicle performance, energy consumption, and routing efficiency (Wang *et al.* 2015, Zhang *et al.* 2024), incorporating it into navigation systems can support low-emission travel and optimize transport energy use.

Beyond mobility efficiency, slope data carries important implications for social equity and inclusive accessibility (Rhoads *et al.* 2023). Steep road segments can substantially constrain independent movement, disproportionately affecting wheelchair users, older adults, and individuals with mobility impairments, for whom even small slope increases may translate into significant physical effort, safety risks, or complete route infeasibility (Ng *et al.* 2025). When such constraints are systematically overlooked, conventional accessibility assessments tend to overestimate effective access to services and opportunities for these populations, masking latent spatial inequalities. Integrating slope information into accessibility models enables the identification of problematic areas and the design of targeted interventions such as ramps, elevators, or alternative relatively flat routes (Liu *et al.* 2025). In this sense, slope-aware accessibility modeling not only improves analytical accuracy but also provides a practical foundation for advancing equitable urban design and mobility justice.

6. Conclusion

This study advances the scarcely investigated potential of street view imagery to estimate slopes of roads in cities, by presenting an integrated framework to accomplish that using single panoramic images, through the fusion of semantic segmentation, geometric correction, and robust regression. By leveraging panoramic data and computer vision methods, our approach — Vision2Slope — provides a scalable and cost-effective solution to a problem traditionally constrained by expensive or low-resolution datasets. Validation in two contrasting urban settings, hilly SF and flat NYC, demonstrates that the framework performs consistently across diverse topographies, outperforming medium-resolution DEMs. In response to the proposed research questions, we find that: (1) street-level visual environments encode substantial and exploitable information about road slope, enabling reliable large-scale estimation from SVI; (2) image-related uncertainties, particularly perspective distortion and visual ambiguity,

introduce systematic errors, but can be effectively mitigated through semantic-guided geometric correction and robust modeling strategies; (3) urban morphological characteristics, such as street canyon structure and occlusion by vehicles, systematically contribute to estimation biases, highlighting that slope inference is intrinsically shaped by the built environment. Despite these advances, challenges remain in extremely flat or visually congested streets, where reliable visual cues are limited or heavily occluded. Future research could incorporate time-series SVI and multimodal data sources, such as LiDAR, GPS, or vision–language models, to further improve estimation accuracy and generalizability. Extending Vision2Slope to global datasets has the potential to support large-scale 3D city modeling, urban drainage and flood risk simulation, and the analysis of vertical accessibility in sustainable urban design.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT 5 to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Acknowledgments

We are grateful to the members of the NUS Urban Analytics Lab for their insightful discussions and to Shuhui Wu for her assistance with the illustration. We also appreciate the reviewers for their efforts and constructive feedback on this work. This work was supported by the National Key Research and Development Program of China (Grant No. 2024YFB3909001), National Natural Science Foundation of China (Grant Nos.42171402, 42571479), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (Grant No. 164320H116). This research is part of the project Large-scale 3D Geospatial Data for Urban Analytics, which is supported by the National University of Singapore under the Start Up Grant R-295-000-171-133. This research is part of the project Multi-scale Digital Twins for the Urban Environment: From Heartbeats to Cities, which is supported by the Singapore Ministry of Education Academic Research Fund Tier 1. The first author is supported by the China Scholarship Council. The second author is supported by the National University of Singapore under the President's Graduate Fellowship.

Notes on contributors

Yang Chen is a PhD candidate at the School of Geography, Nanjing Normal University, China. His research focuses on 3D urban analysis and urban morphology.
Zicheng Fan is a PhD candidate at the Urban Analytics Lab, National University of Singapore. His research interests include urban morphology, spatial perception, and machine learning.
Hao Li is a Lecturer at the Department of Geography, National University of Singapore. His research interests include GIScience, GeoAI, spatial computing, and volunteered geographic information.

Wufan Zhao is an Assistant Professor at the Urban Governance and Design Thrust, The Hong Kong University of Science and Technology (Guangzhou), where he leads the AI4DCity Lab. His research focuses on AI-based remote sensing image understanding, 3D city modelling, multi-modal spatiotemporal data fusion, and their methodological and applied studies.
Xin Yang is a Professor at the School of Geography, Nanjing Normal University, China. Her research interests include digital terrain analysis, spatial analysis, and urban morphology.
Guoan Tang is a Professor at the School of Geography, Nanjing Normal University, China. His research focuses on digital terrain analysis, and geographic information science.
Filip Biljecki is an Assistant Professor at the National University of Singapore and the Principal Investigator of the NUS Urban Analytics Lab. He holds an MSc in Geomatics and a PhD in 3D GIS from the Delft University of Technology in the Netherlands.

Data and codes availability statement

Due to commercial licensing restrictions of the Google Street View imagery used in this study, the full dataset cannot be shared. Instead, we provide a subset of the data together with the end-to-end analytical pipelines and codes for re-creating tables and figures in the paper at https://github.com/CubicsYang/Vision2Slope. This subset is sufficient to reproduce the complete analysis workflow, including all data processing, estimation, and visualization steps.

Disclosure statement

The authors report there are no competing interests to declare.

References

Bacharidis, K., Sarri, F., and Ragia, L., 2020. 3D Building Façade Reconstruction Using Deep Learning. *ISPRS International Journal of Geo-Information*, 9 (5), 322.

Biljecki, F. and Ito, K., 2021. Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*, 215, 104217.

Biljecki, F., *et al.*, 2023. Sensitivity of measuring the urban form and greenery using street-level imagery: A comparative study of approaches and visual perspectives. *International Journal of Applied Earth Observation and Geoinformation*, 122, 103385.

Boeing, G., 2025. Modeling and Analyzing Urban Networks and Amenities With OSMnx. *Geographical Analysis*, 57 (4), 567–577.

Chen, Y., *et al.*, 2026. Vertical 15-minute city: Modeling urban density and functional mix with multi-source geospatial data. *Cities*, 169, 106516.

Chen, Y., *et al.*, 2022. An Automatic Approach to Extracting Large-Scale Three-Dimensional Road Networks Using Open-Source Data. *Remote Sensing*, 14 (22), 5746.

Cheng, B., *et al.*, 2022. Masked-attention mask transformer for universal image segmentation. *In*: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1290–1299.

Danish, M., *et al.*, 2025. A citizen science toolkit to collect human perceptions of urban environments using open street view images. *Computers, Environment and Urban Systems*, 116, 102207.

Du, Y., *et al.*, 2016. Application of Vehicle Mounted Accelerometers to Measure Pavement Roughness. *International Journal of Distributed Sensor Networks*, 12 (6), 8413146.
El Masri, O. and Bigazzi, A.Y., 2019. Road grade estimates for bicycle travel analysis on a street network. *Transportation Research Part C: Emerging Technologies*, 104, 158–171.
Fan, K., *et al.*, 2024. Pano2Geo: An efficient and robust building height estimation model using street-view panoramas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 215, 177–191.
Fan, P., *et al.*, 2022. Road grade estimation based on Large-scale fuel consumption data of connected vehicles. *Transportation Research Part D: Transport and Environment*, 106, 103262.
Fan, Z., *et al.*, 2023. Urban visual intelligence: Uncovering hidden city profiles with street view images. *Proceedings of the National Academy of Sciences*, 120 (27), e2220417120.
Fan, Z., Law, S., and Biljecki, F., 2025. Three Dimensional Street Scene Representation Learning for Street Frontage Classification. *In*: *The 19th International Conference on Computational Urban Planning and Urban Management*, London.
Fang, F., *et al.*, 2022. Spatial context-aware method for urban land use classification using street view images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 192, 1–12.
Faria, M.V., *et al.*, 2019. How do road grade, road type and driving aggressiveness impact vehicle fuel consumption? Assessing potential fuel savings in Lisbon, Portugal. *Transportation Research Part D: Transport and Environment*, 72, 148–161.
Gao, L., *et al.*, 2021. Novel Framework for 3D Road Extraction Based on Airborne LiDAR and High-Resolution Remote Sensing Imagery. *Remote Sensing*, 13 (23), 4766.
Gebru, T., *et al.*, 2017. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences*, 114 (50), 13108–13113.
Gupta, A., *et al.*, 2020. Road grade estimation using crowd-sourced smartphone data. *In*: *2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 313–324.
Higgins, C.D., 2019. A 4D spatio-temporal approach to modelling land value uplift from rapid transit in high density and topographically-rich cities. *Landscape and Urban Planning*, 185, 68–82.
Hosseini, R., *et al.*, 2024. A specialized inclusive road dataset with elevation profiles for realistic pedestrian navigation using open geospatial data and deep learning. *Computers, Environment and Urban Systems*, 114, 102199.
Hou, Y., *et al.*, 2024. Global Streetscapes — A comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics. *ISPRS Journal of Photogrammetry and Remote Sensing*, 215, 216–238.
Huang, W., Wang, J., and Cong, G., 2024. Zero-shot urban function inference with street view images through prompting a pretrained vision-language model. *International Journal of Geographical Information Science*, 38 (7), 1414—-1442.
Huber, P.J., 1983. Minimax Aspects of Bounded-Influence Regression. *Journal of the American Statistical Association*, 78 (381), 66–72.
Ito, K., *et al.*, 2025. ZenSVI: An open-source software for the integrated acquisition, processing and analysis of street view imagery towards scalable urban science. *Computers, Environment and Urban Systems*, 119, 102283.
Jauch, J., *et al.*, 2018. Road Grade Estimation With Vehicle-Based Inertial Measurement Unit and Orientation Filter. *IEEE Sensors Journal*, 18 (2), 781–789.
Jiang, S., *et al.*, 2025. Using multi-source data to evaluate walkability in urban and rural Hong Kong: A multidimensional framework. *Information Geography*, 1 (2), 100029.
John, S., *et al.*, 2017. Deriving incline values for street networks from voluntarily collected GPS traces. *Cartography and Geographic Information Science*, 44 (2), 152–169.
Larrañaga, A.M., *et al.*, 2016. The influence of built environment and travel attitudes on walking: A case study of Porto Alegre, Brazil. *International Journal of Sustainable Transportation*, 10 (4), 332–342.
Li, H., *et al.*, 2025. Cross-view geolocalization and disaster mapping with street-view and vhr

satellite imagery: A case study of hurricane ian. *ISPRS Journal of Photogrammetry and Remote Sensing*, 220, 841–854.
Liang, X., *et al.*, 2025. OpenFACADES: An open framework for architectural caption and attribute data enrichment via street view imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 230, 918–942.
Liu, D., *et al.*, 2025. Assessment of the 15-minute park accessibility accounting for fine-grained 3D pedestrian network and subjective quality perception among the public housing residents. *Urban Forestry & Urban Greening*, 114, 129165.
Liu, H., *et al.*, 2018. Development of road grade data using the United States geological survey digital elevation model. *Transportation Research Part C: Emerging Technologies*, 92, 243–257.
Liu, Z., *et al.*, 2026. The first road surface type dataset for 50 African countries and regions. *Earth System Science Data*, 18 (1), 267–285.
Lu, K., *et al.*, 2025. StreetSenser: a novel approach to sensing street view via a fine-tuned multimodal large language model. *International Journal of Geographical Information Science*, 0 (0), 1–29.
Lu, Y. and Karimi, H.A., 2015. Real-Time Sidewalk Slope Calculation through Integration of GPS Trajectory and Image Data to Assist People with Disabilities in Navigation. *ISPRS International Journal of Geo-Information*, 4 (2), 741–753.
Ma, Y., Li, Y., and Long, Y., 2025. Measuring Temporal Evolution of Nationwide Urban Physical Disorder: An Approach Combining Time-Series Street View Imagery with Deep Learning. *Annals of the American Association of Geographers*, 115 (4), 923–948.
Matas, J., Galambos, C., and Kittler, J., 2000. Robust Detection of Lines Using the Progressive Probabilistic Hough Transform. *Computer Vision and Image Understanding*, 78 (1), 119–137.
Mawe, G.I., *et al.*, 2025. Mapping urban gullies in the Democratic Republic of the Congo. *Nature*, 644 (8078), 952–959.
McKenzie, G. and Janowicz, K., 2017. ISED: Constructing a high-resolution elevation road dataset from massive, low-quality in-situ observations derived from geosocial fitness tracking data. *PLOS ONE*, 12 (10), e0186474.
Meijer, J.R., *et al.*, 2018. Global patterns of current and future road infrastructure. *Environmental Research Letters*, 13 (6), 064006.
Micusik, B. and Kosecka, J., 2009. Piecewise planar city 3D modeling from street view panoramic sequences. *In*: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2906–2912.
Neuhold, G., *et al.*, 2017. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes. 4990–4999.
Ng, K.Y., *et al.*, 2025. 3D walking accessibility in practice: exploring the imperfections from data, method, and assumptions of human-space interaction. *International Journal of Geographical Information Science*, 40 (2), 1–21.
Ning, H., *et al.*, 2022a. Converting street view images to land cover maps for metric mapping: A case study on sidewalk network extraction for the wheelchair users. *Computers, Environment and Urban Systems*, 95, 101808.
Ning, H., *et al.*, 2022b. Exploring the vertical dimension of street view image based on deep learning: a case study on lowest floor elevation estimation. *International Journal of Geographical Information Science*, 36 (7), 1317–1342.
OpenTopography, 2021. USGS 1 meter Digital Elevation Model. Available from: https://opentopography.org/meta/OT.012021.4269.3.
Osama, N., Shao, Z., and Freeshah, M., 2023. The FABDEM Outperforms the Global DEMs in Representing Bare Terrain Heights. *Photogrammetric Engineering & Remote Sensing*, 89 (10), 613–624.
Pang, H.E. and Biljecki, F., 2022. 3D building reconstruction from single street view images using deep learning. *International Journal of Applied Earth Observation and Geoinformation*, 112, 102859.

Qin, T., *et al.*, 2024. Crowd-sourced nerf: Collecting data from production vehicles for 3d street view reconstruction. *IEEE Transactions on Intelligent Transportation Systems*, 25 (11), 16145–16156. Publisher: IEEE.

Randhawa, S., *et al.*, 2025. Paved or unpaved? A deep learning derived road surface global dataset from mapillary street-view imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 223, 362–374.

Rhoads, D., Solé-Ribalta, A., and Borge-Holthoefer, J., 2023. The inclusive 15-minute city: Walkability analysis with sidewalk networks. *Computers, Environment and Urban Systems*, 100, 101936.

Rong, Y., *et al.*, 2020. Three-dimensional urban flood inundation simulation based on digital aerial photogrammetry. *Journal of Hydrology*, 584, 124308.

Rosero, F., *et al.*, 2021. Effects of passenger load, road grade, and congestion level on real-world fuel consumption and emissions from compressed natural gas and diesel urban buses. *Applied Energy*, 282, 116195.

Salazar Miranda, A., *et al.*, 2022. Favelas 4D: Scalable methods for morphology analysis of informal settlements using terrestrial laser scanning data. *Environment and Planning B: Urban Analytics and City Science*, 49 (9), 2345–2362.

Schönberger, J.L. and Frahm, J.M., 2016. Structure-from-Motion Revisited. *In*: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June. 4104–4113.

Seidl, D.E., Jankowski, P., and Tsou, M.H., 2016. Privacy and spatial pattern preservation in masked GPS trajectory data. *International Journal of Geographical Information Science*, 30 (4), 785–800.

Shi, K., *et al.*, 2025. Extensive terrestrial biodiversity threats from global hillside urban expansion. *Nature Cities*, 2 (10), 937–947.

Sun, G., Webster, C., and Zhang, X., 2019. Connecting the city: A three-dimensional pedestrian network of Hong Kong. *Environment and Planning B Urban Analytics and City Science*, 48, 1–16.

Szcześniak, J.T., *et al.*, 2022. A method for using street view imagery to auto-extract window-to-wall ratios and its relevance for urban-level daylighting and energy simulations. *Building and Environment*, 207, 108108.

Torii, A., Havlena, M., and Pajdla, T., 2009. From Google Street View to 3D city models. *In*: *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*. 2188–2195.

van Ginkel, K.C.H., *et al.*, 2021. Flood risk assessment of the European road network. *Natural Hazards and Earth System Sciences*, 21 (3), 1011–1027.

Wang, H., *et al.*, 2021. Automatic generation of large-scale 3D road networks based on GIS data. *Computers & Graphics*, 96, 71–81.

Wang, Y., *et al.*, 2015. Impacts of Road Grade on Fuel Consumption of Light Vehicles by Use of Google Earth DEM. *In*: *2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*. 360–363.

Xiong, L., *et al.*, 2025. Towards the next generation of digital terrain modelling and analysis: New value-added perspective for geoscience. *Information Geography*, 1 (1), 100017.

Xue, J., *et al.*, 2022. Quantifying the spatial homogeneity of urban road networks via graph neural networks. *Nature Machine Intelligence*, 4 (3), 246–257.

Xue, X. and Zhu, X.X., 2025. Regression in Earth Observation: Are vision–language models up to the challenge? *IEEE Geoscience and Remote Sensing Magazine*, 2–31.

Yadav, M., Singh, A.K., and Lohani, B., 2017. Extraction of road surface from mobile LiDAR data of complex road environment. *International Journal of Remote Sensing*, 38 (16), 4655–4682.

Yan, Y. and Huang, B., 2022. Estimation of building height using a single street view image via deep neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 192, 83–98.

Yan, Y., *et al.*, 2025. Towards vertical urban geometry extraction: occlusion-reduced estimation from street view images using diffusion models. *International Journal of Digital Earth*, 18 (1), 2520475.

Yang, B., Fang, L., and Li, J., 2013. Semi-automated extraction and delineation of 3D roads of street scene from mobile laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 79, 80–93.

Yang, C., *et al.*, 2022. Human expansion into Asian highlands in the 21st Century and its effects. *Nature Communications*, 13 (1), 4955.

Yang, X., *et al.*, 2023. Three-dimensional structure determination of grade-separated road intersections from crowdsourced trajectories. *International Journal of Applied Earth Observation and Geoinformation*, 125, 103598.

Yazdani Boroujeni, B. and Frey, H.C., 2014. Road grade quantification based on global positioning system data obtained from real-world vehicle fuel use and emissions measurements. *Atmospheric Environment*, 85, 179–186.

Ying, Y., *et al.*, 2020. Urban 3d modelling methods: A state-of-the-art review. *In*: *XXIVth ISPRS Congress 2020*. International Society for Photogrammetry and Remote Sensing (ISPRS), 699–706.

Zhang, C., Fan, H., and Kong, G., 2021. VGI3D: an Interactive and Low-Cost Solution for 3D Building Modelling from Street-Level VGI Images. *Journal of Geovisualization and Spatial Analysis*, 5 (2), 18.

Zhang, H., *et al.*, 2024. Three-dimensional road network empowered traffic emission estimation via trajectory data. *Journal of Cleaner Production*, 475, 143669.

Zhao, W., Persello, C., and Stein, A., 2023. Semantic-aware unsupervised domain adaptation for height estimation from single-view aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196, 372–385.

Zhou, C., 2025. Exploring future GIS visions in the era of the scientific and technological revolution. *Information Geography*, 1 (1), 100007.

Zhou, H., *et al.*, 2025. A Comparative Study of Deep Learning Methods for Automated Road Network Extraction from High-Spatial-Resolution Remotely Sensed Imagery. *Photogrammetric Engineering & Remote Sensing*, 91 (3), 163–174.

Appendix A. Supplementary analysis of within-segment slope variability

Figure A1 compares the distribution of the coefficient of variation (CV) of point-level slope estimates within each road segment before and after applying the iterative Huber regression. When using the original point-level slopes, the CV distribution is wide and strongly right-skewed, indicating substantial within-segment variability and the presence of local outliers in many segments. After robust aggregation, CV values are markedly reduced and concentrate within a narrow low-CV range, demonstrating substantially improved internal consistency of point-level slope estimates. Figure A2

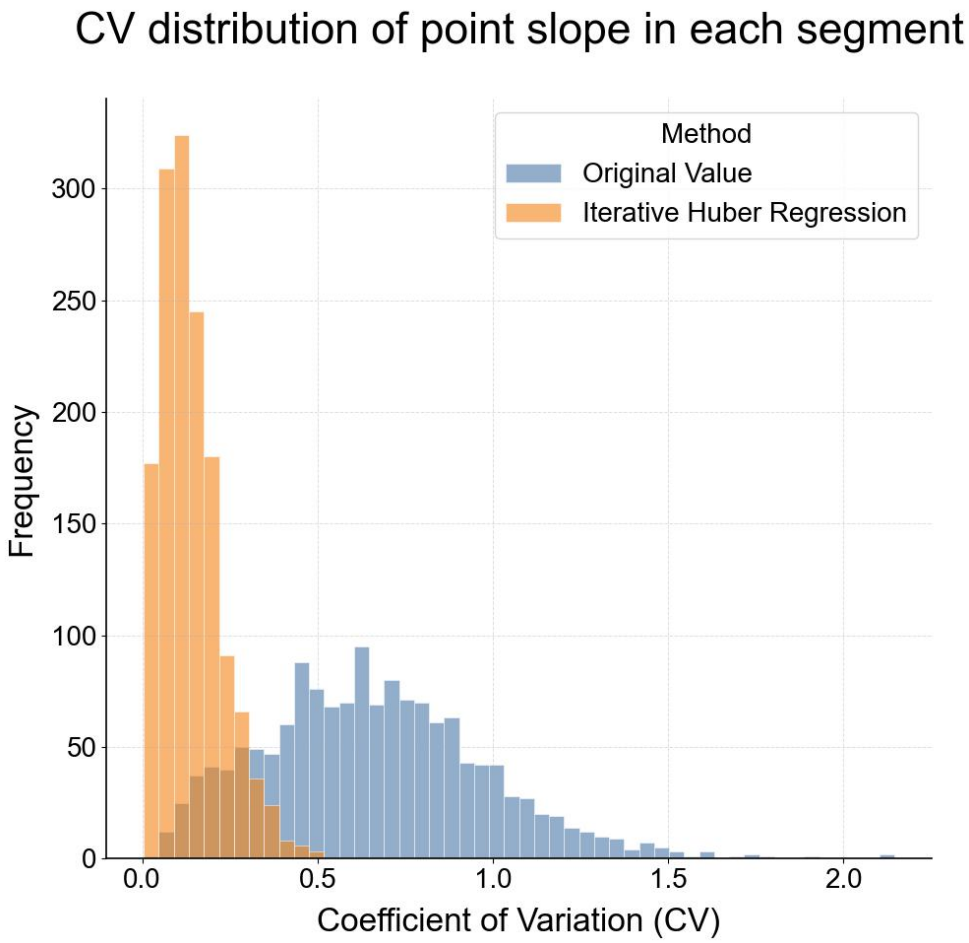

Figure A1. Distribution of the coefficient of variation (CV) of point-level slope estimates within road segments. The blue histogram shows CV values computed from the original point-level slope estimates, while the orange histogram corresponds to results after applying the iterative Huber regression. Lower CV values indicate more uniform slope estimates within a segment.

shows an inverse relationship between segment-level slope and CV of point-level slopes. Flatter segments exhibit higher relative variability due to greater sensitivity to local noise, while steeper segments display more uniform slope estimates. Figure A3 further demonstrates the spatial heterogeneity in slope variability within segments, providing a potential reference for future research.

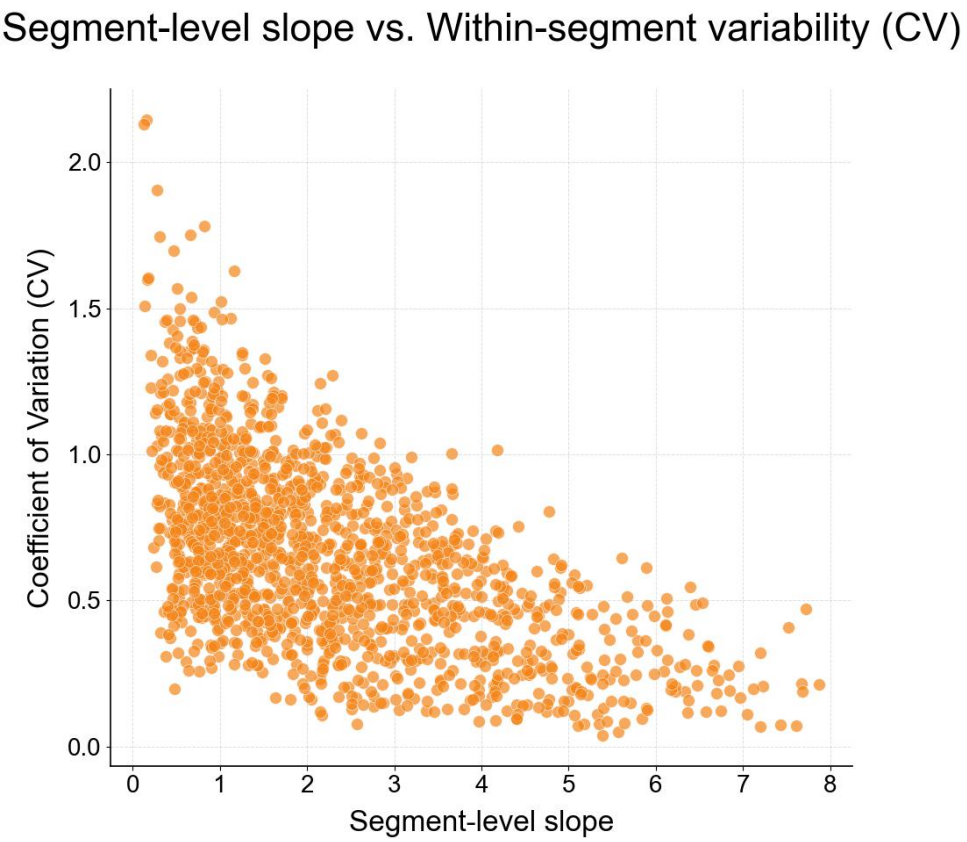

Figure A2. Relationship between segment-level road slope and within-segment slope variability. Each dot denotes a road segment, showing the relationship between segment-level slope and the coefficient of variation (CV) of point-level slope estimates. Lower CV values correspond to more uniform within-segment slopes.

Appendix B. NYC estimation results

This appendix presents the detailed estimation results for New York City, provided to complement the main experiments and to illustrate the spatial patterns and quantitative performance of the proposed method in a large, complex urban environment.

Figure B1 visualizes the spatial distribution of the estimated road slopes at both the point and segment levels, together with the derived segment-level relief, highlighting how local point estimations aggregate into coherent segment-scale patterns across the city. Figure B2 further evaluates the estimation accuracy by comparing the predicted slopes and relief against ground-truth values, using scatter plots and error distributions stratified by slope magnitude.

Table B1 summarizes the quantitative error metrics for point-level slope, segment-level slope, and segment-level relief estimations, reported across different slope ranges to provide a more nuanced view of performance under varying terrain conditions.

Figure A3. Spatial distribution of within-segment variability in slope across the road network. Lighter colors indicate lower variability, while darker purple denotes higher variability.

Appendix C. Additional condition bias

Figures examine the sensitivity of road slope estimation accuracy across joint deciles of distance to the closest building and surrounding spatial context, represented by car area (Figure C1) and road area (Figure C2), respectively. In both figures, estimation errors are lowest in intermediate deciles, forming a clear U-shaped pattern along the building-distance dimension, with higher MAE observed for roads that are either very close to buildings (D1–D2) or located in more open environments (D9–D10). This trend is also reflected in the marginal curves, indicating that moderate building proximity provides more stable visual cues for slope estimation. Along the y-axis, errors generally decrease from lower to middle deciles of car or road area and increase again at the extremes, suggesting that both highly constrained and highly exposed road environments pose challenges. Notably, the lowest MAE values concentrate around the central deciles in both dimensions (approximately D4–D7), whereas the largest errors occur at combinations of minimal building distance with very low or very high car/road area, highlighting compounded effects of visual occlusion and scene sparsity on estimation performance.

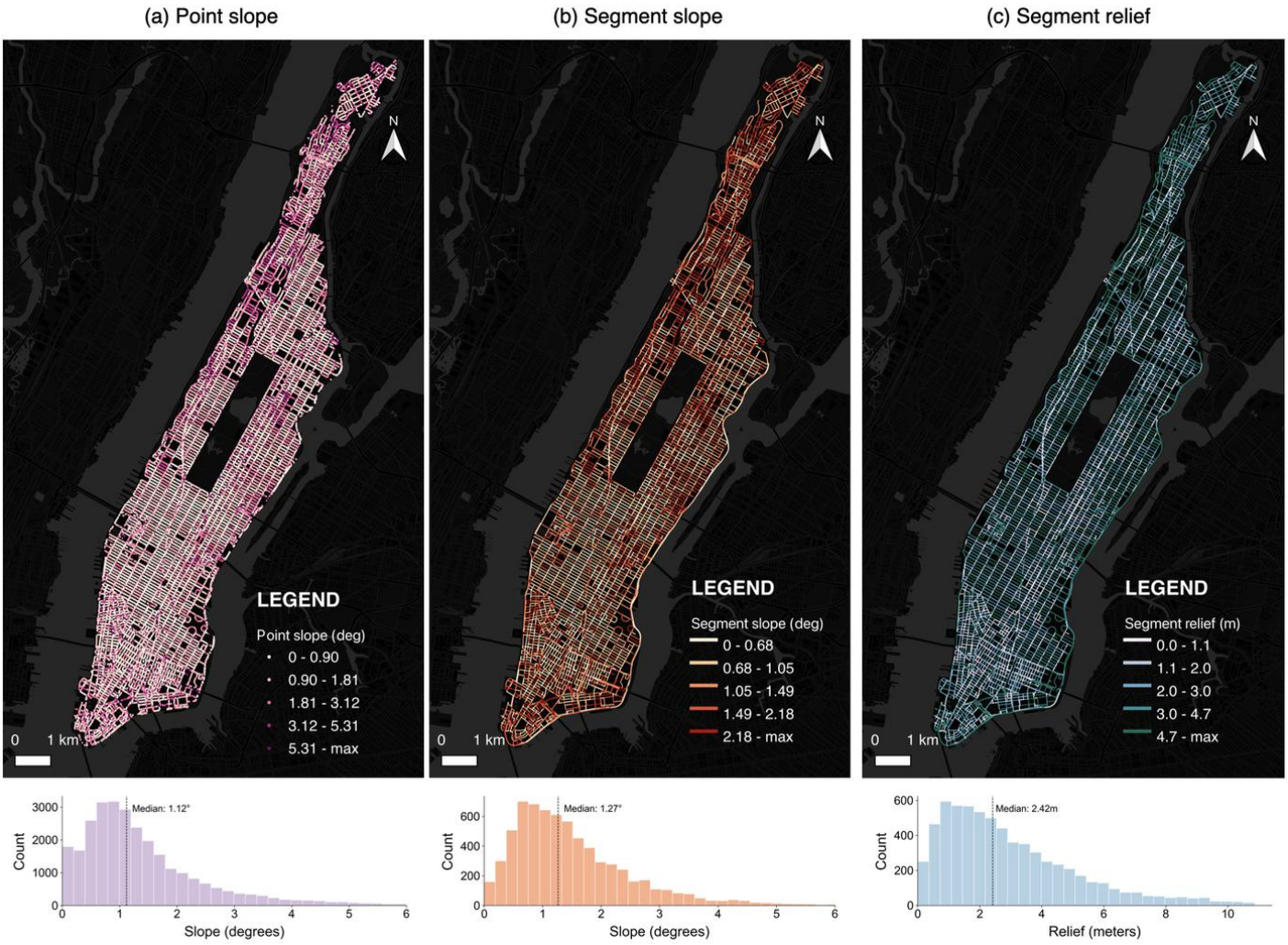

Figure B1. (a) Point-level road slope estimation across NYC, (b) Segment-level slope map aggregated from point estimations. (c) Segment-level relief map showing elevation difference along each road segment. Basemap: CARTO and OpenStreetMap contributors.

Appendix D. Narrow street FOV adjustment example

Figure D1 demonstrates that, even in narrow streets typical of dense urban environments, increasing the effective FOV can substantially enhance the visibility of road surfaces and curbs, thereby mitigating occlusion effects and improving the robustness of slope estimation from single-view SVI.

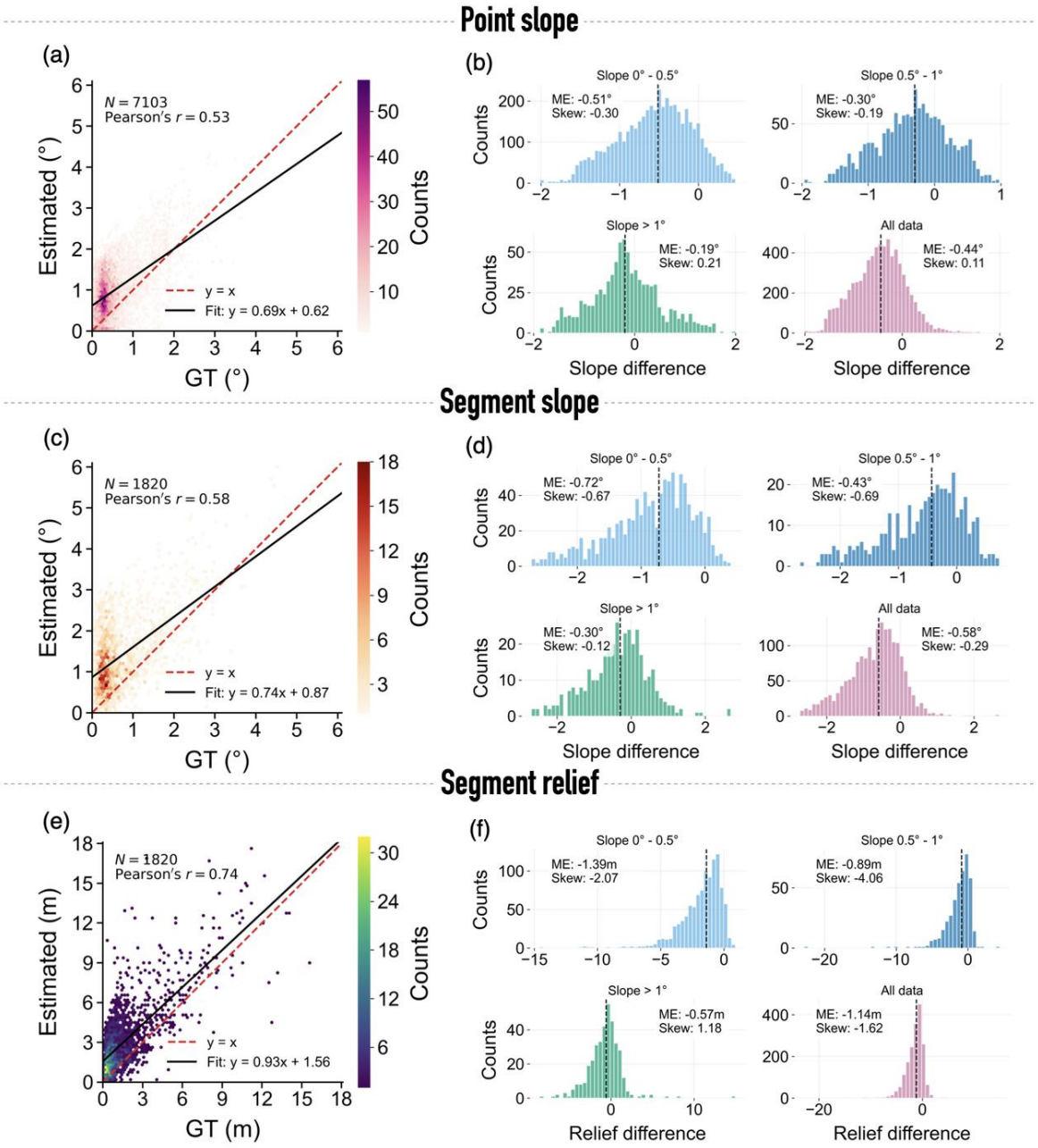

Figure B2. Comparison of point-based and segment-based slope and relief estimations against ground truth (GT) in NYC (a, c, e) Scatter plots between estimated and GT values with regression lines (black) and 1:1 lines (red). Color scales represent point density. (b, d, f) Histograms of estimation differences grouped by slope or relief intervals: 0–0.5°, 0.5–1°, > 1° , and all data.

Table B1. Absolute error of slope and relief estimation in New York City.

Metric	**Range**	**RMSE**	**MAE**	**sMAPE (%)**	**N**
Point-level slope (°)	All	0.71	0.57	85.09	7103
	0–0.5°	0.73	0.59	101.58	4763
	0.5–1°	0.64	0.51	59.83	1478
	> 1°	0.70	0.55	37.29	862
Segment-level slope (°)	All	1.00	0.78	82.93	1820
	0–0.5°	1.07	0.86	108.54	1043
	0.5–1°	0.90	0.67	60.90	406
	> 1°	0.89	0.67	35.04	371
Segment-level relief (m)	All	2.36	1.66	82.94	1820
	0–0.5°	2.40	1.78	108.55	1043
	0.5–1°	2.46	1.51	60.91	406
	> 1°	2.14	1.47	35.07	371

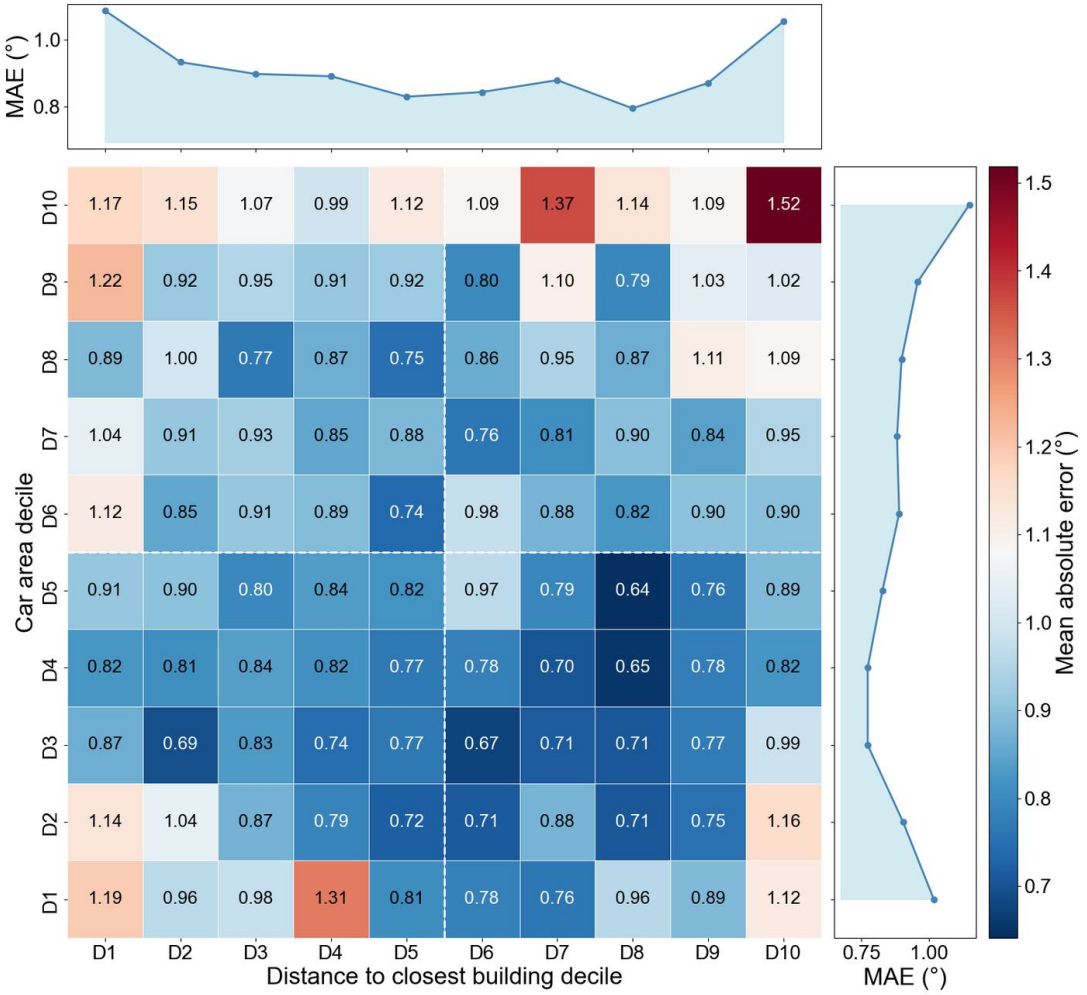

Figure C1. Sensitivity of road slope estimation accuracy stratified by distance to the closest building decile (x-axis) and car area decile (y-axis).

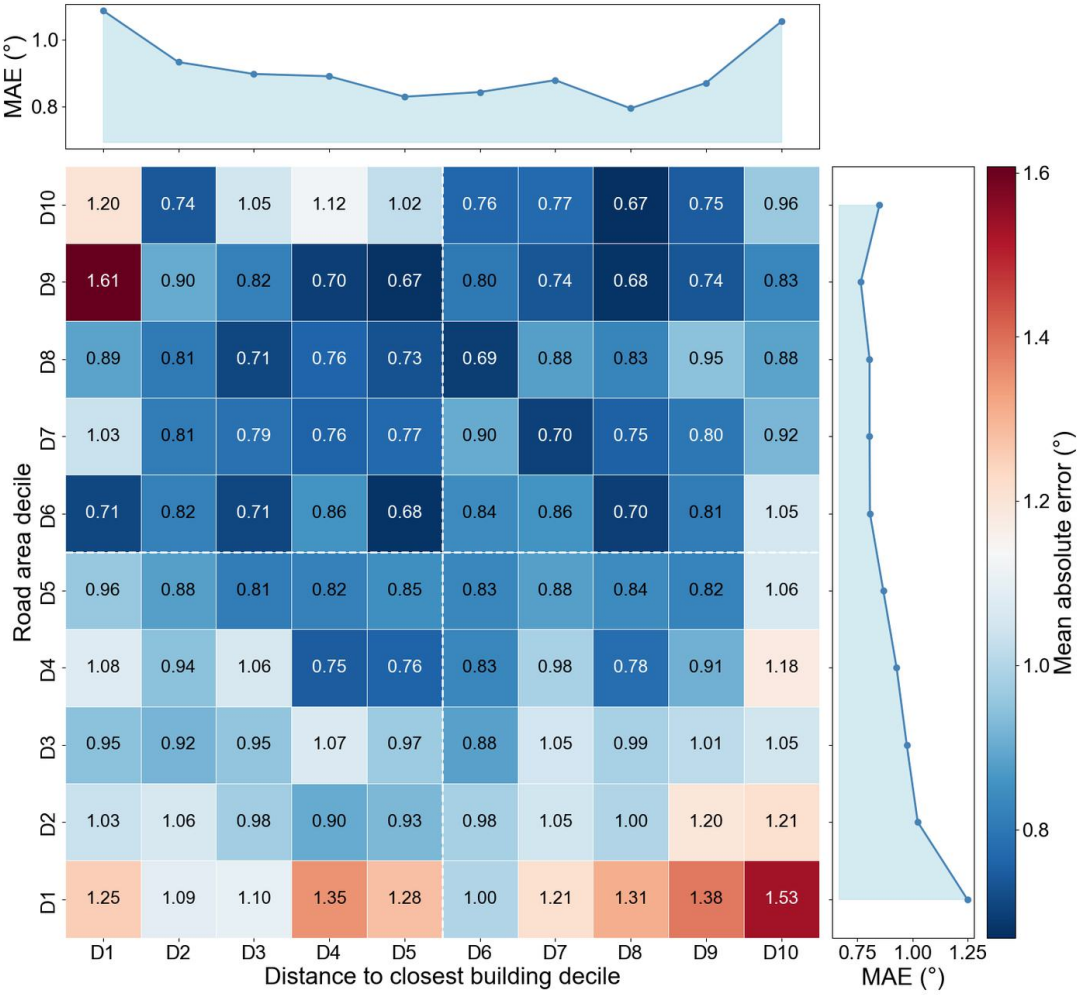

Figure C2. Sensitivity of road slope estimation accuracy stratified by distance to the closest building decile (x-axis) and road area decile (y-axis).

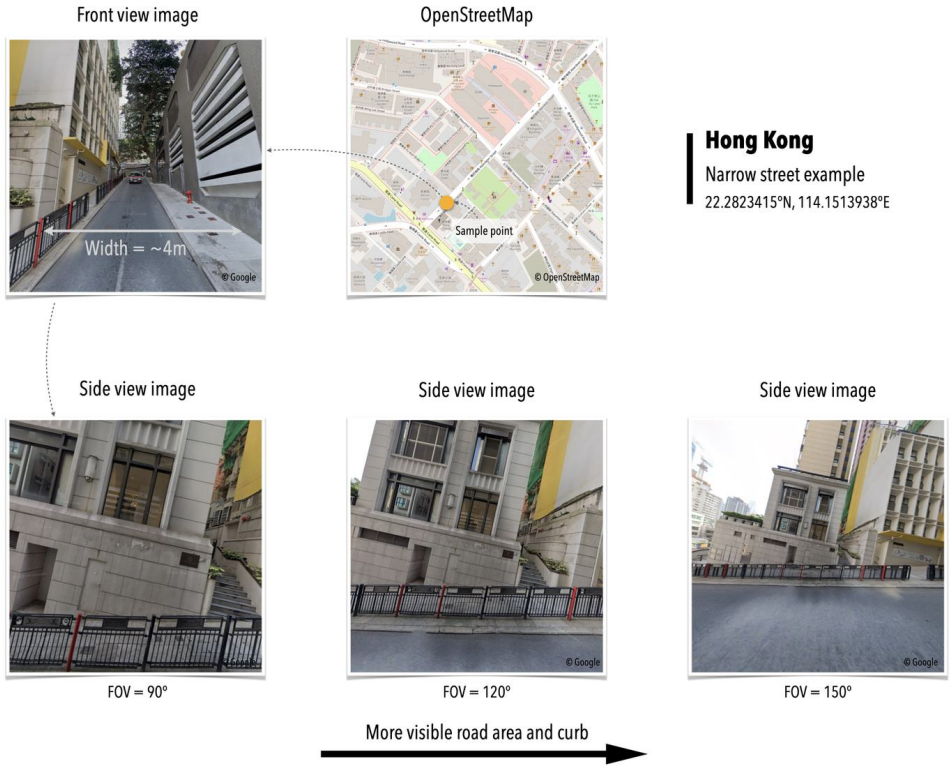

Figure D1. Narrow street example in Hong Kong showing how increasing the effective FOV from 90°, 120° and 150° progressively reveals more road surface and curb through panorama re-projection.