Computers, Environment and Urban Systems 127 (2026) 102421

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Does weather matter? Examining measurement bias in street view image-based urban perception assessments

Donghwan Ki [a], Sungmin Lee [b], Chaeyeon Han [c], Youjung Kim [d], Bon Woo Koo [e,*], Uijeong Hwang [f]

[a] Department of City and Regional Planning, The Ohio State University, 275 West Woodruff Avenue, Columbus, OH 43210, USA
[b] Department of Landscape Architecture and Urban Planning, Texas A&M University, 789 Ross Street, College Station, TX 77843, USA
[c] Department of City and Regional Planning, Georgia Institute of Technology, 760 Spring Street NW Suite 217, Atlanta, GA 30332, USA
[d] School of Public Affairs, University of South Florida, 4202 E. Fowler Avenue, ALN 136, Tampa, FL 33620, USA
[e] Department of Urban Planning and Engineering, Yonsei Univerisity, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea
[f] Transportation Access & Mobility, Atlanta Regional Commission, 229 Peachtree ST NE, Suite 100, Atlanta, GA 30303, USA

ARTICLE INFO

Keywords:
Human perception
Measurement biases
Data quality
Google street view
Place Pulse 2.0 dataset

ABSTRACT

Recent advances in urban analytics, leveraging Google Street View (GSV), computer vision, and large-scale perception datasets, have enabled scalable measurement of how people perceive urban environments. While many studies posit that perception scores from this approach primarily reflect the built environment itself, these measures may be influenced by external conditions, especially weather. Such confounding effects may threaten the validity of measures, potentially biasing conclusions about built-environment influences. This study addresses two questions: (1) To what extent and in what contexts do weather-induced biases manifest in perception scores? (2) If such biases exist, what challenges do they pose for GSV data collection, particularly regarding spatial coverage loss under clear-only sampling? Using historical GSV imagery from Los Angeles (CA), Seattle (WA), and Portland (OR) for identical locations captured under varying weather conditions, we combine perception-prediction models trained on Place Pulse 2.0 with automated weather classification via the ZenSVI package. Our results indicate that perception scores are systematically lower in non-clear weather than in clear conditions, even when the built environment remains constant. The magnitude of these biases varies across perception attributes (safety, liveliness, and beauty), cities, and levels of perception scores. Restricting the sample to weather-consistent imagery reduces measurement bias but inevitably decreases spatial coverage, creating a trade-off between validity and representativeness. However, our results show that selecting city-specific dry periods with higher GSV availability can address this trade-off by minimizing coverage loss and improving measurement validity. Collectively, these findings underscore the importance of considering weather in perception assessment, offering methodological guidance for more valid analyses.

1. Introduction

A growing set of urban analytics tools, particularly street view imagery (SVI), computer vision models, and large-scale human-perception datasets such as Place Pulse 2.0, have transformed the study of urban perception. These resources enable researchers to capture how people perceive the built environment (e.g., perceived safety) across large geographic areas (Zhang et al., 2018). Building on these advances, scholars have explored links between such perception measures and societal outcomes (e.g., Miao et al., 2025; Zhou et al., 2022) and emphasized that understanding these perceptions is critical for urban design and management (Larkin et al., 2021; Rossetti et al., 2019). Compared to traditional, resource-intensive surveys, this novel approach substantially enhances the scalability in analyzing spatial variation in perceptions and their impacts.

As SVI-based built-environment analyses grow, studies increasingly highlight their sensitivities, particularly substantial variability in environmental measurements from differences in preprocessing procedures (Hou & Biljecki, 2022; Kim et al., 2021; Zhao et al., 2025). They underscore that insufficient control over SVI acquisition and preprocessing

* Corresponding author.
E-mail addresses: ki.17@osu.edu (D. Ki), sungminlee@tamu.edu (S. Lee), chan303@gatech.edu (C. Han), youjungkim@usf.edu (Y. Kim), bonwookoo@yonsei.ac.kr (B.W. Koo), uhwang@atlantaregional.org (U. Hwang).

https://doi.org/10.1016/j.compenvurbsys.2026.102421
Received 2 November 2025; Received in revised form 1 March 2026; Accepted 4 March 2026
Available online 16 March 2026

can introduce significant measurement biases, ultimately distorting inferences about built-environment effects. In urban planning, where empirical evidence guides investments in infrastructure and safety interventions, unreliable environmental measurements risk misleading policies and undermining the credibility of data-driven planning (Flyvbjerg, 2007).

While several studies have examined measurement biases associated with particular SVI preprocessing methods, such as seasonal bias in greenery assessments arising from the choice of temporal windows (Larkin et al., 2025; Zhao et al., 2025), weather-induced biases remain largely underexplored. Seasonal variation unfolds predictably, whereas weather conditions fluctuate abruptly and unevenly, sometimes within the same season or even the same day, making them essentially arbitrary. As a result, weather can introduce substantial and spatially inconsistent measurement bias in SVI-based perception assessment. For example, a location that appears safe based on its built-environment features may be perceived as less safe when depicted in imagery captured under cloudy and gloomy atmospheric conditions. Because these distortions are both arbitrary and highly localized, weather-induced bias has the potential to amplify errors in SVI-based perception measurement. Despite its methodological significance, this issue has so far been recognized only at a conceptual level (Biljecki & Ito, 2021; Ito et al., 2024; Lauko et al., 2020), with limited empirical examination.

To address this gap, we systematically examine the influence of weather on Google Street View (GSV)-derived perception measures (perceived safety, liveliness, and beauty) across three U.S. cities with distinct weather patterns: Los Angeles (CA), Seattle (WA), and Portland (OR). Specifically, we address the following research questions:

RQ1. To what extent and in what contexts do weather-induced biases manifest in perception scores?

RQ2. If such biases exist, what challenges do they create for GSV data collection, particularly regarding spatial coverage loss under clear-only sampling strategies?

RQ1 investigates whether perception scores vary with weather conditions under otherwise identical built-environment contexts and examines the contexts (e.g., score levels, cities) in which such biases are most pronounced. *RQ2* asks what challenges weather-induced biases pose for GSV data collection, with particular attention to spatial coverage loss under clear-only sampling.

Leveraging historical GSV imagery of identical locations captured under different weather conditions, this study quantifies weather-related biases in perception scores. We use perception-prediction models trained on the Place Pulse 2.0 dataset with automated weather classification from the ZenSVI package. In doing so, the study provides methodological guidance for mitigating measurement biases, thereby improving the validity of perception measures and strengthening the credibility of SVI-based urban analytics in planning practice.

2. Literature review

2.1. Street view imagery and human perception data

SVI has been increasingly used in urban studies on human perception of streetscapes due to its various advantages. SVI offers rich visual information with broad geographic coverage, offering feasible ways to quantify characteristics of urban streetscapes. Another value of SVI is its eye-level perspective, better representing the actual stimuli that pedestrians are exposed to (Biljecki & Ito, 2021; Koo et al., 2022). Studies have shown that eye-level perspective in SVI can be more effective in explaining behavioral outcomes than the traditional top-down satellite imagery (Lu et al., 2019).

A key resource enabling the scalability of SVI-based perception measurement is large-scale perception datasets with labeled scores, such as the Place Pulse 2.0 dataset (Dubey et al., 2016). As the most widely used dataset, Place Pulse 2.0 includes human perception annotations for 110,988 GSV images from 56 global cities across six perceptual dimensions (i.e., safety, liveliness, beauty, wealthiness, depression, and boredom), collected through pairwise comparisons. Researchers can train computer vision models on this dataset and apply the trained models to their own SVI data, thereby generating perception scores at scale (see Zhang et al., 2018). The use of this large-scale perception dataset enables planners to systematically capture how people perceive the built environment, providing valuable insights to guide the design, management, and planning of public spaces (Larkin et al., 2021; Rossetti et al., 2019).

The studies using Place Pulse-based computer vision models can be categorized into two broad groups: (1) those assessing what physical features determine the perception and (2) those examining how the perception shapes societal outcomes.

In the first category, studies analyze how objective visual attributes of images (e.g., greenery proportion and visual complexity) influence perception dimensions (Rossetti et al., 2019; Xu & Fan, 2025; Zhang et al., 2018). To operationalize image-level attributes, they commonly use semantic segmentation–derived measures and visual-complexity indicators, reporting that higher proportions of features such as greenery and sidewalks are associated with higher perceived safety, while visual complexity, as a proxy for informational richness, is linked particularly to perceived liveliness. Studies in the second category associate positive perceptions (e.g., safety and beauty) with desirable outcomes including longer time spent in parks (Zhou et al., 2022), less non-motorist crash risk (Miao et al., 2025), and greater bike share (Zhang et al., 2025).

Most research examining the spatial distribution of human perception and its relationship with societal outcomes across extensive geographic areas relies on the Place Pulse 2.0 dataset. The scarcity of training datasets that are comparably large, have global coverage, and are widely validated seems to have made the Place Pulse dataset a dominant choice of researchers across different domains. However, it is worth noting that some studies have highlighted the limitations of this dataset, such as the limited coverage in some countries and the lack of detailed sociodemographic profiles of survey participants (Danish et al., 2025; Lu & Chen, 2024; Qiu et al., 2022; Song et al., 2025). Therefore, they have addressed these gaps by developing their own surveys and using them in conjunction with computer vision models.

2.2. Measurement biases of street view imagery data

SVI data offer scalability for measuring the built environment, yet several studies have demonstrated the sensitivity of SVI-based environmental measurements to specific acquisition and processing procedures. Biljecki et al. (2023) analyze the sensitivity of common urban form metrics, such as green view and sky view index, to different panoramic settings (e.g., field of view) in SVI. Zhao et al. (2025) extend this experiment to assess seasonal bias (i.e., how seasonal changes at the same location alter greenery measurements) across 40 cities worldwide. Similarly, Kim et al. (2021) demonstrate that environmental measurement values can vary considerably depending on the image acquisition interval along road segments. Building on this concern, Fan et al. (2025) propose a method to determine the optimal SVI sampling intervals (i.e., the optimal n-meter spacing along road). Collectively, these studies underscore the need for carefully designed acquisition and analysis procedures to ensure reliable and valid SVI-based environmental measurements.

A major factor that amplifies such measurement sensitivity and potential biases, and thus must be carefully considered in SVI-based environmental assessments, is the heterogeneity of SVI data quality, such as inconsistencies in field of view and resolution (Biljecki & Ito, 2021; Ito et al., 2024; Li et al., 2022). Compared to crowdsourced SVI platforms like Mapillary, this issue is less pronounced in GSV due to its relatively standardized collection strategy (Huang et al., 2024).

Nevertheless, inconsistencies related to lighting, blurriness, weather conditions, and occlusion still remain (Biljecki & Ito, 2021; Lauko et al., 2020). Ito et al. (2024) imply that the heterogeneity of weather across SVI images may act as a potential confounding factor in measuring urban perception.

Such biases in built-environment measurement are not merely methodological; they can propagate directly into erroneous evidence-based planning decisions. In practice, reliance on flawed data can result in what Flyvbjerg (2007) terms the "survival of the unfittest," whereby projects that appear favorable on paper are prioritized over those that genuinely address community needs. From a research standpoint, such measurement errors also risk leading scholars to misinterpret relationships between built-environment characteristics and their targeted behavioral or health outcomes (Hooper et al., 2013).

Taken together, prior studies acknowledge the susceptibility of SVI-based measures to acquisition and processing procedures, yet few have specifically addressed weather-induced biases. Zhao et al. (2025) examine seasonal variability in greenery assessments, thereby highlighting measurement biases. However, unlike season, weather is more arbitrary and localized, requiring investigation from a different analytical angle. Accordingly, our study fills this gap by empirically examining how weather conditions systematically bias SVI-based perception measures, offering new insights into the validity of environmental assessments.

3. Methods

3.1. Study areas

This study examines weather-induced measurement biases using three U.S. case cities characterized as having distinct weather patterns, including Los Angeles (CA), Seattle (WA), and Portland (OR) (Fig. 1). Los Angeles experiences relatively few cloudy days annually, whereas Portland and Seattle are at the cloudier end of the distribution. This contrast enables analysis of measurement biases across diverse weather patterns, thereby strengthening the generalizability of the findings. Because Portland and Seattle are considerably smaller than Los Angeles in both land area and population, including two cloudier cities helps maintain a more balanced basis for cross-city comparison.

3.2. Analysis frameworks

In this study, measurement bias in SVI-based human perceptions is defined as differences in perception scores that arise not from the built environment itself but from variations in weather conditions (see Fig. 3). These differences can be observed by comparing perception scores (e.g., perceived safety) for two GSV images captured at the same location but under varying weather conditions, and can conceptually be expressed as:

$$\Delta Human\ Perception\ Scores = f(\Delta Weather, \Delta Built\ Environment, \Delta Location) \quad (1)$$

In practice, such differences (Δ Scores) can result from (1) changes in weather (e.g., clear vs. cloudy), (2) changes in the built environment,[1] and (3) minor positional offsets between two historical GSV images at each location. To isolate the effect of weather, this study designs the dataset to explicitly control the latter two factors (Fig. 2).

3.2.1. Step 1 (acquiring historical GSVs)

Sampling points are generated at 100-m intervals along each city's road network (see the *Location ID* column in Fig. 2), and all available historical GSVs and metadata (e.g., coordinates, capture dates) are retrieved from these sampling points. Next, images captured more than 5 m away from the designated sampling point are excluded to eliminate the effect of minor positional offsets among historical images (see Eq. (1)). Finally, to eliminate potential effects associated with changes in GSV camera equipment, only historical GSVs captured after 2017 are retained (see Section 3.3).

3.2.2. Step 2 (weather classification)

Weather conditions for the GSVs obtained in Step 1 are classified using the ZenSVI package (Ito et al., 2025). We retain only those *Location IDs* exhibiting within-location weather variation (i.e., at least one clear image and at least one non-clear image per location).

3.2.3. Step 3 (perception inference)

Perceived safety, liveliness, and beauty scores are inferred for each GSV using computer-vision models trained on the Place Pulse 2.0 dataset. Accordingly, each image is assigned predicted scores that reflect how people perceive the built environment.

3.2.4. Step 4 (semantic segmentation and visual complexity)

Semantic segmentation and visual complexity measures are computed for each image to derive built-environment attributes. These measures are included as control variables in the subsequent regression models to account for built-environment differences between historical GSVs at each location (Eq. (1)).

Through these processes, we compile for each location only those images that (1) are captured within 5 m of the designated sampling point, (2) exhibit weather variation across historical GSVs at each location, and (3) were taken after 2017.

3.3. Google street view

This study retrieves historical GSVs at sampling points generated at 100-m intervals along each city's road network. Although there is no single agreed-upon standard for defining sampling intervals in GSV-based built environment measurement, a 100-m spacing has been commonly used in prior studies (e.g., Li & Ratti, 2019; Zhou et al., 2024). In addition, given the spatial extent of our study areas, this interval provides a practical balance between spatial coverage and computational feasibility. For the image type, we use panoramic GSVs, which are stitched together from four directional static images.[2] In addition, metadata such as capture date (year and month) and geographic coordinates are obtained for each image.

We restrict the dataset to GSVs captured after 2017, when Google upgraded its Street View system to higher-resolution cameras (Simonite, 2017). This restriction ensures that only GSVs captured with the same equipment are analyzed. To the best of our knowledge, Google has not reported any further camera updates since then.

[1] Here, changes in the built environment refer not only to large-scale transformations, such as new construction, but also to small-scale alterations, such as increases in parked vehicles.

[2] Although the computer vision models used in this study (see Section 3.4) are trained on static images rather than stitched panoramic images, we use panoramic images as input to improve computational efficiency in large-scale analysis. Using a stitched panoramic image reduces the number of required inferences from four to one, substantially reducing processing time. To verify that this choice does not materially affect inference outcomes, we compare predictions for 500 sampled locations by averaging scores from four static images and contrasting them with scores from a single panoramic image. The resulting mean absolute error (MAE) is low (0.170), representing the largest observed discrepancy across the three perception dimensions. Even in this worst-case comparison, the correlation between the two approaches remains very high ($r = 0.973$, $p < 0.001$). Agreement in weather classification, implemented using the ZenSVI package, is also high, with a Cohen's kappa of 0.857, indicating almost perfect agreement (McHugh, 2012). These results indicate that inference outcomes from the computer vision models remain robust when using a stitched single panoramic image instead of multiple static images.

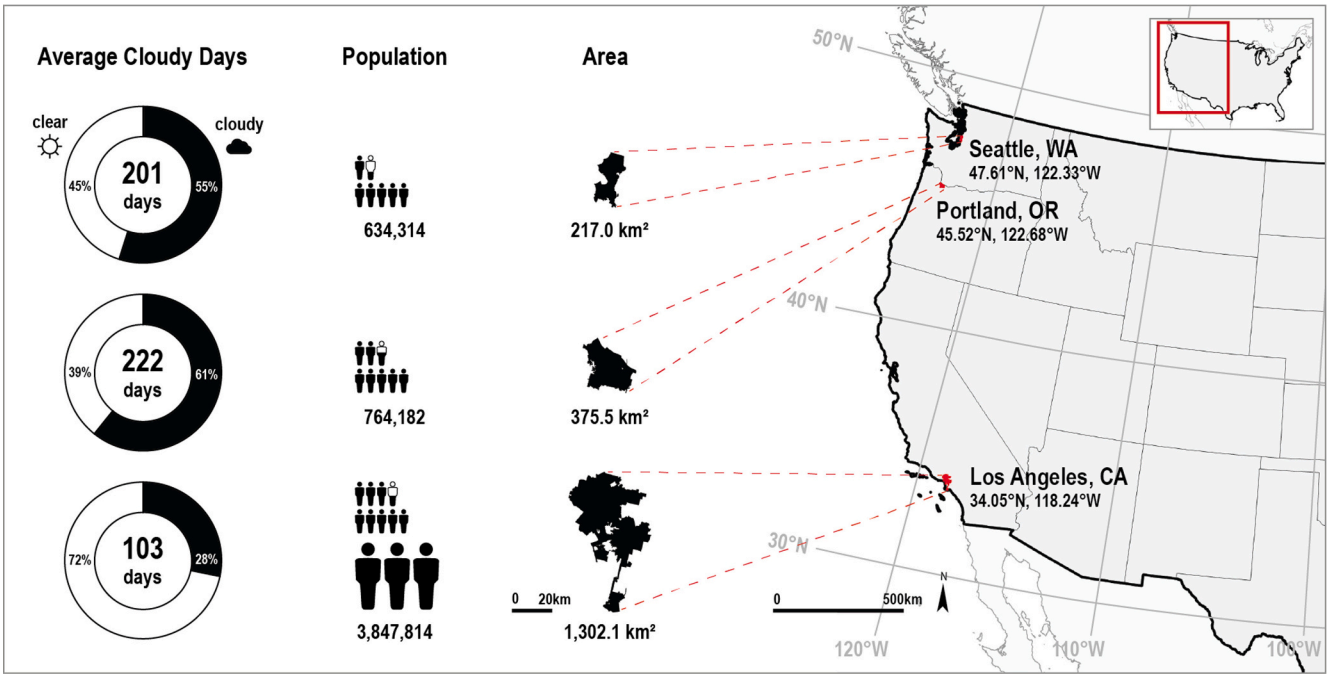

Fig. 1. Study areas.

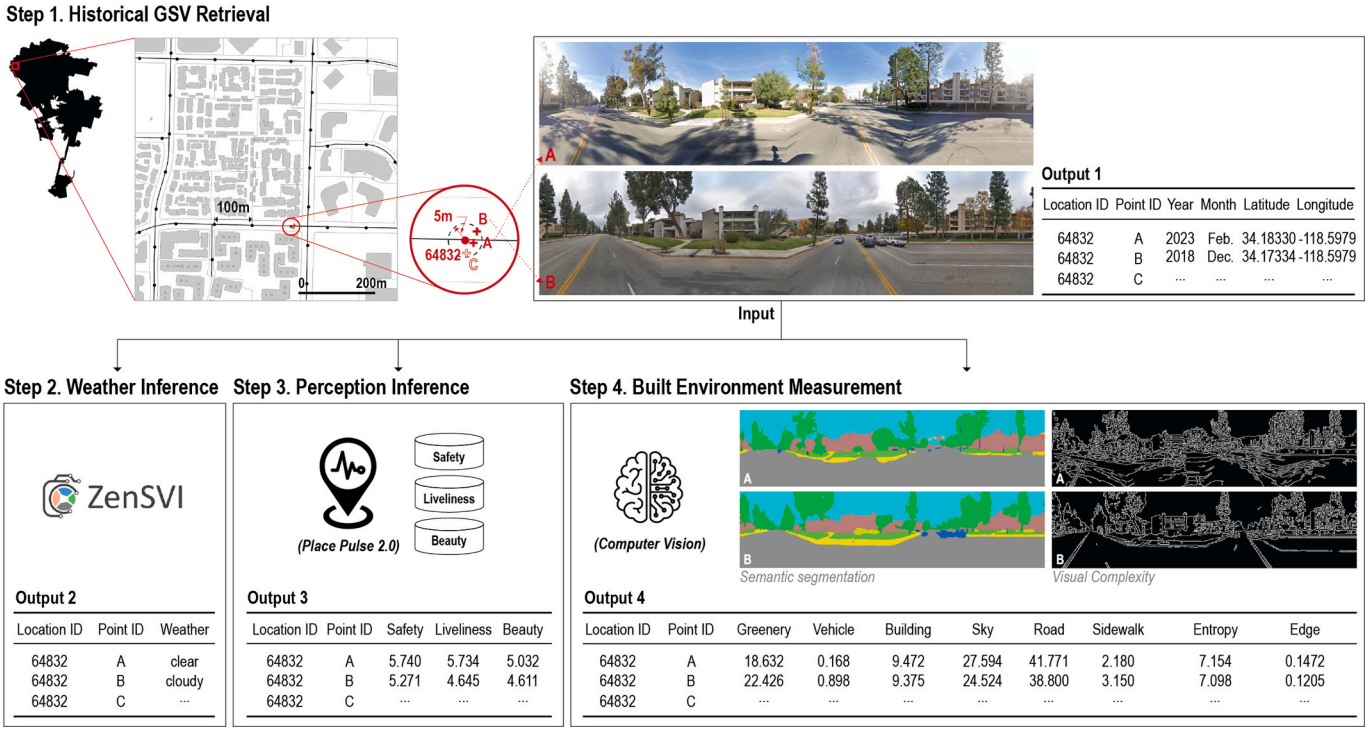

Fig. 2. Dataset construction workflow.

3.4. Computer vision models

This study employs four computer-vision models: (1) perception-prediction models trained on the Place Pulse 2.0 dataset, (2) a weather-classification model embedded in the ZenSVI package, (3) a semantic-segmentation model, and (4) visual complexity algorithms used to quantify image-level built-environment compositions.

3.4.1. Place Pulse 2.0 dataset

This study employs the Place Pulse 2.0 dataset, a large-scale crowdsourced resource for quantifying human perceptions of urban environments. Because the dataset was originally collected through pairwise comparisons, we apply the TrueSkill algorithm (Microsoft, 2026) to convert these comparisons into continuous scores for each image. We then normalize the perception scores to range from 0 to 10. Through this process, we obtain continuous scores for 110,988 images across six perceptual dimensions: safety, liveliness, beauty, wealthiness,

depression, and boredom.

A key consideration in selecting training samples from the TrueSkill-based perception scores is the varying reliability of the scores. Some images received only a small number of pairwise comparisons (i.e., votes), which directly affects the reliability of the resulting perceptual scores. To address this issue, Larkin et al. (2021) restricted their analysis to images with vote counts in the top quartile. Following their approach, we include only images with vote counts above the top quartile for each perceptual dimension. Of these, 80% are randomly sampled for training, and the remaining 20% are reserved for validation. To train the models, we use the MaxVit Multi-Axis Vision Transformer (Tu et al., 2022), pre-trained on ImageNet 12- k and fine-tuned on ImageNet 1- k for classification tasks.

We build individual prediction models for each of the six perceptual dimensions. However, model performance on the test dataset (the 20% not used for training) varies by perceptual attributes, as measured by the correlation between ground truth and predicted values. Specifically, safety achieves a relatively reliable correlation of 0.66, while boredom exhibits a lower correlation of 0.36. This variation can be attributed to the cut-off criteria for vote counts and the resulting differences in the number of training samples. For this reason, we restrict subsequent analyses to the three perceptual attributes with more stable prediction accuracy: safety (0.66), liveliness (0.57), and beauty (0.59). The inference results from our models are presented in Fig. 3.

3.4.2. ZenSVI package

The ZenSVI package provides comprehensive and standardized pre-processing tools for SVI-based analyses (Ito et al., 2025). We use only its weather-classification function.[3] The weather-classification model, originally developed by Hou et al. (2024), is integrated into the ZenSVI package. To train this model, the developers constructed a human-labeled dataset in which annotators manually labeled weather conditions from SVI images (training $n = 27{,}771$; test $n = 8068$) (Hou et al., 2024). Accordingly, the classifier's weather labels are intended to approximate how humans categorize SVI scenes based on overall visual appearance, reflecting a composite outcome in which multiple visual characteristics (e.g., brightness, contrast, and saturation) are likely to be jointly expressed.

The model is based on a MaxViT architecture and achieves 75.5% accuracy on the developers' test set when classifying images into five weather categories: clear, cloudy, foggy, rainy, and snowy (Hou et al., 2024). Because foggy, rainy, and snowy samples are relatively few and visually difficult to distinguish from cloudy, we combine these four categories into one "non-clear" class. Fig. 3 and Fig. S1 provide examples of the classification results (see *Supplementary Document A*).

3.4.3. Semantic segmentation

Semantic segmentation is a computer vision technique that classifies each image pixel into a predefined semantic category (e.g., sidewalks). We employ SegFormer pretrained on the ADE20K dataset (Xie et al., 2021; Fig. S2), which provides fast and reliable inference suitable for large-scale GSV processing. We compute the proportion of pixels belonging to each object class to derive image-level built-environment attributes. In particular, we focus on attributes that the model segments reliably and that are theoretically related to human perception: vehicles, sidewalks, greenery, sky, buildings, and roads. These pixel-based measures provide an objective representation of the built environment for each image and are incorporated as control variables in the subsequent analyses.

[3] The ZenSVI package also includes a lighting-condition vision model that classifies images as day, night, or dusk/dawn. We use this model solely for sample filtering, retaining only images captured during the day.

3.4.4. Visual complexity

To capture subtle environmental characteristics that extend beyond those represented by semantic segmentation-derived measures, we quantify two image-level measures of visual complexity: textural heterogeneity and edge density. Prior research suggests that, beyond the composition of semantic elements, the overall visual complexity of streetscapes can shape how people perceive and evaluate scenes. Drawing on Kaplan's preference matrix (Kaplan, 1979), complexity reflects the amount and variety of visual information in a scene, which can invite exploration, sustain attention, and influence affective responses. Consistent with this perspective, several studies have used direct (e.g., Abkar et al., 2011; Kawshalya et al., 2022; Xu & Fan, 2025) or indirect (e.g., Harvey et al., 2015) measures of visual complexity to demonstrate associations with preferences for urban environments.

Textural heterogeneity serves as a proxy for the diversity of visual patterns (i.e., textures) within an image. It is operationalized by converting each GSV image into a 256-level grayscale format and calculating the Shannon entropy of the pixel intensity distribution. Higher entropy values indicate informational heterogeneity, representing scenes with diverse textures and unpredictable patterns, whereas lower values characterize homogeneous scenes, such as monotone walls or clear skies. Separately, edge density captures structural complexity by quantifying the proportion of pixels identified as edges. Using the Canny edge detection algorithm, we calculate the ratio of edge pixels to the total number of pixels. This metric reflects the abundance of object boundaries and geometric details, such as architectural ornaments or intricate vegetation. Examples of these measures are shown in Fig. S3.

3.5. Analysis strategies

The aims of this study are twofold. First, we demonstrate to what extent and in which contexts weather-induced measurement bias manifests in SVI-based perception scores (*Analysis 1; RQ1*). Second, we examine the challenges of accounting for such biases in SVI data collection, particularly coverage loss from restricting to weather-consistent imagery (*Analysis 2; RQ2*). *Analysis 2* extends beyond demonstrating the existence of weather-induced biases to assess the practical feasibility of accounting for them in SVI data collection. The analytical design is structured to address the objective of each analysis.

The two analyses rely on different datasets. For *Analysis 1*, we retain only locations with at least one clear and one non-clear image, enabling within-location comparisons across weather conditions (see Section 3.2). For *Analysis 2,* we include all sampling points from the pre-filter dataset described in Section 3.2.2, since the goal is to assess spatial coverage loss under consistent-weather acquisition strategies.

3.5.1. Analysis 1: quantifying weather-induced measurement bias

To examine weather-induced biases, we employ both descriptive analysis and mixed-effect OLS models. All analyses are conducted across three cities and three perception dimensions: perceived safety, liveliness, and beauty. For each location, we select two historical GSVs captured under different weather conditions (clear vs. non-clear) with the smallest temporal gap.

The final stage of *Analysis 1* uses mixed-effects models to quantify the effect of weather on perception while controlling for built-environment features (see Eq. (1)). Given the hierarchical structure of the dataset, with two GSVs nested within each location, the model includes a random intercept at the location level. Control variables include built-environment attributes derived from semantic segmentation (e.g., percentage of greenery) and visual complexity algorithms, and image capture date (year and month) to account for seasonality, while the key predictors are the binary weather indicator (clear vs. non-clear) and its interaction with city dummies (city × weather).

	Year	Month	Weather	Safety	Liveliness	Beauty
(A)	2025	May	Cloudy	5.436	5.473	4.790
(B)	2023	February	Clear	6.024	6.116	4.975

Fig. 3. Perception scores for the same location under different weather conditions.

3.5.2. Analysis 2: coverage-loss analyses

Analysis 2 examines spatial coverage loss arising from using only images captured under consistent weather conditions to reduce weather-induced biases. Specifically, when restricting all sampling points to clear-weather images, locations with no imagery meeting this condition (i.e., all images captured under non-clear weather) are excluded from analysis. While this strategy improves measurement validity by reducing weather-related biases, it inevitably cuts spatial coverage, introducing a trade-off between measurement validity and spatial representativeness. In this analysis, the unit of interest is the sampling point (location) rather than individual images.

We quantify this trade-off by evaluating how limiting analysis to weather-consistent imagery affects the spatial coverage of GSV sampling points under two scenarios: (1) imposing only a weather-consistency constraint regardless of capture month and (2) applying both restrictions on weather consistency and specific temporal window (e.g., May–August). The second scenario reflects a more realistic design inspired by prior SVI-based studies, which often limit image retrieval to fixed temporal windows to reduce potential seasonal biases (e.g., Zhao et al., 2025). By comparing these two scenarios, we assess how increasingly restrictive yet practically relevant sampling conditions influence spatial coverage.

4. Results

4.1. Descriptive statistics for weather conditions

Table 1 presents the number of sampling points (locations) and corresponding GSVs by weather condition (clear vs. non-clear) for each city across the two datasets. It is worth reiterating that *Dataset 1* is used to examine weather-induced measurement biases and therefore excludes sampling points lacking variation in weather conditions (e.g., locations containing only clear images). As a result, *Dataset 1* contains fewer sampling points and GSVs than *Dataset 2*.

Due to differences in size, Los Angeles has substantially more sampling points and GSVs than Seattle and Portland. The distribution of weather conditions, inferred using the ZenSVI package, also differs across cities. As expected, Seattle and Portland have a higher proportion of images captured under non-clear conditions than Los Angeles; nevertheless, even in these comparatively cloudy cities, clear-weather images remain more prevalent than non-clear ones.

4.2. Analysis 1: quantifying weather-induced measurement biases

4.2.1. Descriptive pattern of measurement biases

Fig. 4 presents histograms of perception score differences between

Table 1
Distribution of GSV images and weather conditions across cities.

	City	Sampling Points	GSVs	GSVs with Clear Weather	GSVs with Non-Clear Weather
Dataset 1[a]	Los Angeles	73,921	375,631	239,823 (63.85%)	135,808 (36.15%)
	Seattle	18,086	64,576	35,518 (55.00%)	29,058 (45.00%)
	Portland	26,972	106,525	60,674 (56.96%)	45,851 (43.04%)
Dataset 2	Los Angeles	123,799	503,017	353,847 (70.34%)	149,170 (29.66%)
	Seattle	32,672	91,314	55,234 (60.49%)	36,080 (39.51%)
	Portland	41,374	135,366	82,797 (61.17%)	52,569 (38.83%)

[a] As noted in Section 3.5, we retain only two GSVs per location with different weather conditions and the smallest temporal gap for *Analysis 1*. Here, however, we report the total number of images prior to applying this filtering criterion to describe the overall distribution of weather conditions.

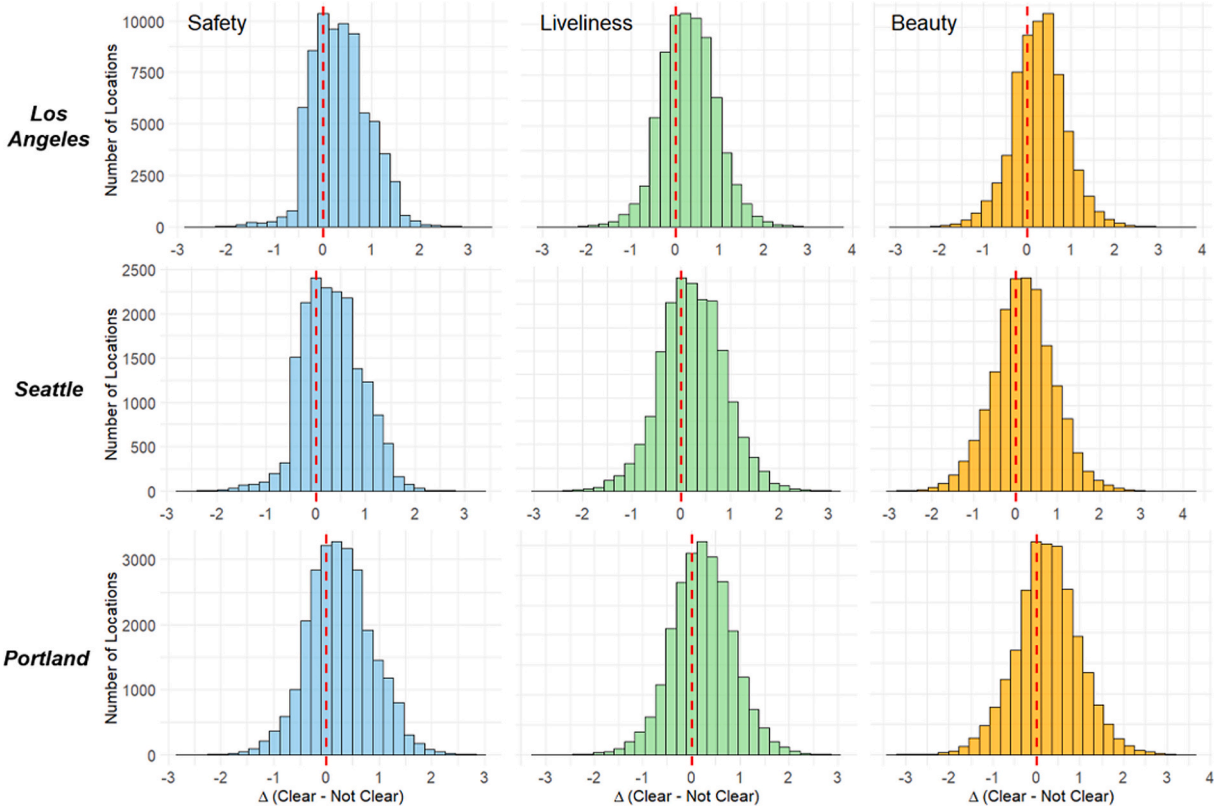

Fig. 4. Histograms of measurement bias ($Perception_{clear} - Perception_{non-clear}$).

two GSVs at each location captured under distinct weather conditions ($Perception_{clear} - Perception_{non-clear}$). Overall, the distributions are skewed to the right of $x = 0$, indicating that perception scores are generally higher under clear weather across all perception attributes and cities. In all cases, two-sample *t*-tests comparing scores between clear and non-clear conditions reveal significant differences ($p < 0.01$).

Some observations that are located to the left of $x = 0$ exhibit higher scores under non-clear conditions than under clear ones. These cases are partially attributable to changes in the built environment between the two historical GSVs or to misclassification errors in the weather inference models (see *Supplementary Document B*).

Table 2 summarizes descriptive statistics for the measurement biases. Overall, scores under non-clear weather conditions range from approximately 0.940 to 0.974 of those under clear weather conditions. In addition, cases with large biases (i.e., $\Delta \geq 1$) account for roughly 12.62% to 19.13% of all observations. Regarding intercity variation, Los Angeles demonstrates larger measurement biases overall compared to Seattle and Portland. Although tentative, this may reflect the exceptionally bright and vivid appearance of Los Angeles under clear weather, which amplifies visual contrast and increases measurement bias.

Table 2
Descriptive statistics for measurement biases.

City		MeanΔ	Mean |Δ|	Mean (Non-Clear / Clear)	Δ ≥ 1 (%)
Los Angeles	Safety	0.351	0.524	0.940	15.931
(n = 73,921)	Liveliness	0.303	0.535	0.947	13.646
	Beauty	0.269	0.517	0.951	12.621
Seattle	Safety	0.319	0.530	0.947	16.077
(n = 18,086)	Liveliness	0.227	0.545	0.964	14.255
	Beauty	0.193	0.589	0.974	17.955
Portland	Safety	0.263	0.522	0.959	14.501
(n = 26,972)	Liveliness	0.225	0.523	0.964	12.880
	Beauty	0.250	0.609	0.964	19.130

Note: $\Delta = Perception_{clear} - Perception_{non-clear}$

4.2.2. Contextual moderators of measurement biases

Fig. 5 presents scatter plots with clear-weather perception scores on the x-axis and the measurement bias on the y-axis. These plots examine whether measurement bias varies with the level of the perception score.

Across all cities and perception attributes, measurement bias increases as clear-weather perception scores rise. For samples with relatively low perception scores (approximately 4 or lower), differences between clear and non-clear conditions are minimal. By contrast, as perception scores increase, the magnitude of measurement bias grows substantially.

Another potential driver of measurement bias is the proportion of sky visible in GSV imagery, as greater sky exposure could make weather conditions more apparent. To test this, we examine the relationship between measurement bias and the mean proportion of sky across the two images used to compute the bias (see Fig. 6). The proportion of sky is derived through semantic segmentation, with the average value plotted on the x-axis and the corresponding measurement bias plotted on the y-axis.

Contrary to expectations, the relationship between sky proportion and measurement bias is not statistically significant. A possible explanation is that even when the sky is only minimally visible, other visual cues, such as lighting, color saturation, and contrast, may sufficiently convey differences between clear and non-clear weather conditions. Consequently, weather can still be fully perceived without a large share of visible sky, thereby limiting the moderating role of sky proportion.

4.2.3. Mixed-effects regression estimating weather effects

Table 3 presents the results of mixed-effects OLS models, with locations specified as Level 2 and the two GSVs per location (clear vs. non-clear) as Level 1. The models estimate the effect of weather conditions

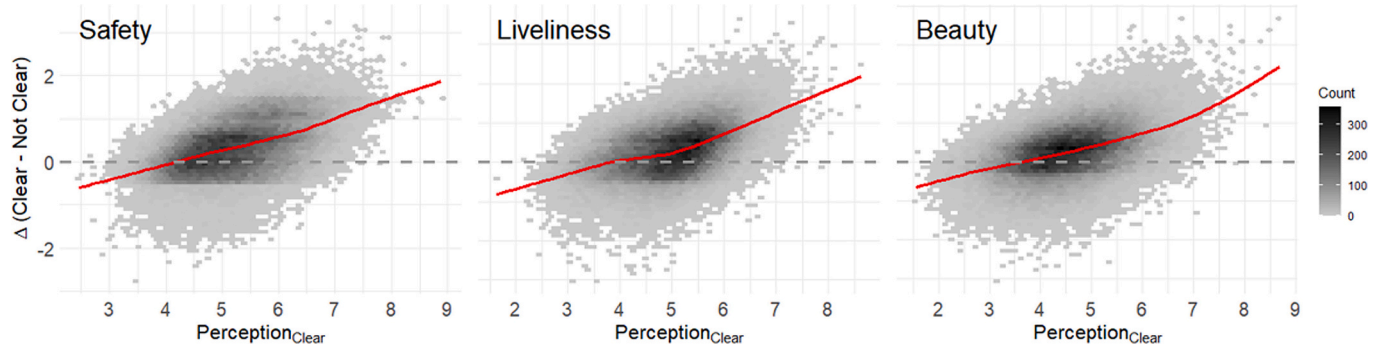

Fig. 5. Relationship between clear-weather scores and measurement biases (Los Angeles).[a]
[a]The relationship exhibits a similar pattern across all three cities. Therefore, for reasons of space, we report results for Los Angeles only in the manuscript. The same applies to Fig. 6.

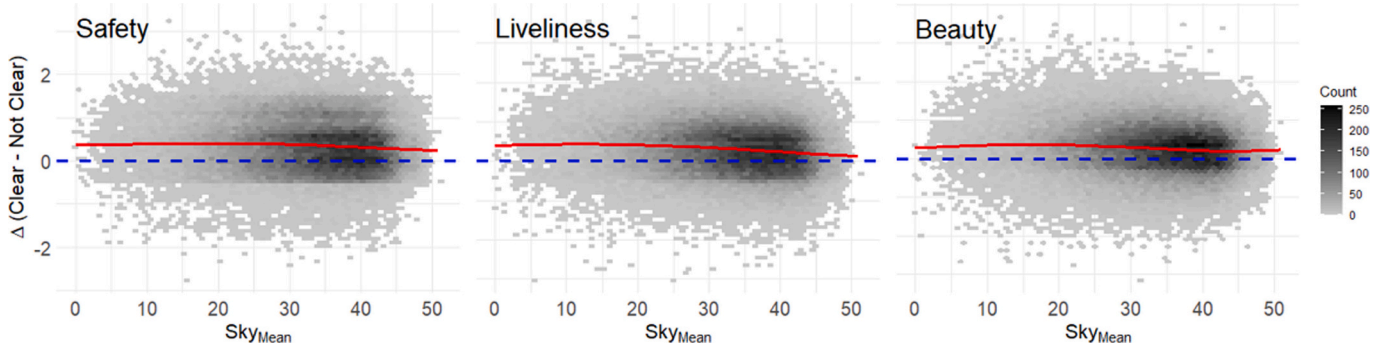

Fig. 6. Relationship between sky ratio and measurement biases (Los Angeles).

Table 3
Mixed-effects models result (pooled models across cities).

Variables		Safety		Liveliness		Beauty	
		Coef.	t value	Coef.	t value	Coef.	t value
Target Variables							
Weather (Ref.: Clear)	Non-Clear	−0.369	−152.12	−0.319	−120.64	−0.259	−95.85
City (Ref.: Los Angeles)	Seattle	0.048	7.59	0.029	5.95	0.243	41.63
	Portland	0.103	22.46	0.077	11.99	0.413	78.85
Non-Clear x Seattle		0.040	8.08	0.079	14.84	0.136	25.24
Non-Clear x Portland		0.110	25.45	0.095	20.98	0.078	16.71
Control Variables							
Greenery		0.018	86.89	0.019	69.96	0.037	169.70
Vehicle		0.029	66.30	0.042	102.48	−0.006	(−1.20)
Building		0.028	91.04	0.042	152.31	−0.000	(−0.95)
Road		0.001	6.37	0.006	26.75	0.010	39.97
Sidewalk		0.031	72.53	0.037	83.85	0.016	34.62
Textual Heterogeneity		0.057	8.87	0.047	7.16	0.157	22.77
Edge Density		2.222	28.83	3.849	49.25	4.033	48.99
Intercept		3.995	78.37	3.385	64.90	2.139	39.09
N	Level 1 (Image)			237,958			
	Level 2 (Location)			118,979			
ICC (Level 2)		0.598		0.521		0.555	
AIC		413,353.8		415,545.0		441,305.7	
BIC		413,509.2		415,700.4		441,461.1	
Log Likelihood		−206,661.9		−207,757.5		−220,637.8	

Note 1. Except for the vehicle and building variables in the beauty model (with those t-values shown in parentheses), all other variables across the models are significant at the 1% level.

on perception scores while controlling for built-environment features derived from semantic segmentation and visual complexity algorithms. During model specification, we exclude the semantic segmentation share of *sky* due to high multicollinearity with other segmentation variables; after this exclusion, variance inflation factors (VIFs) for all predictors remained below 5. To account for seasonality, all models additionally include fixed effects for image capture year and month; these coefficients are not reported for brevity.

Across all perception attributes, scores are consistently lower under non-clear weather compared to clear weather. The magnitude of weather-induced biases is greatest for safety, followed by liveliness and beauty. Cross-city comparisons show stronger weather-induced biases in

Table 4
Mixed-effects models fit.

Model	With Weather		Without Weather	
	Marginal R^2	Conditional R^2	Marginal R^2	Conditional R^2
Safety	0.297	0.729	0.245	0.642
Liveliness	0.358	0.697	0.290	0.631
Beauty	0.525	0.789	0.481	0.751

Los Angeles than in the two cloudier cities, as reflected in the positive coefficients of the interaction terms. For safety, perception scores decrease by 0.369, 0.329 (−0.369 + 0.040), and 0.259 (−0.369 + 0.110) points under non-clear conditions in Los Angeles, Seattle, and Portland, respectively.

In terms of control variables, the image-level segmentation measures (e.g., greenery) are largely consistent with prior studies (see Zhang et al., 2018). Visual complexity has often been reported to exhibit non-linear patterns in some contexts (Kawshalya et al., 2022), but it is generally positively associated with human perceptions of scenes (Abkar et al., 2011; Harvey et al., 2015; Xu & Fan, 2025).

Table 4 reports the R^2 values for the mixed-effects models, with particular attention to comparisons between models with and without the weather variable. When accounting for the random intercept at the location level, the conditional R^2 ranges from 0.697 to 0.789 with the weather variable included, while the explanatory power of the independent variables alone (marginal R^2) ranges from 0.297 to 0.525. As expected, the inclusion of the weather variable improves model fit across all perception attributes.

4.3. Analysis 2: scenario-based analysis of coverage loss

4.3.1. Spatial coverage loss under weather constraints

Table 5 presents the proportion of GSV sampling points lost when requiring weather-consistent imagery at all points. When considering only clear-weather images, Los Angeles shows the smallest coverage loss (6.11%) relative to the other two cities, reflecting the predominance of clear-weather imagery (see Table 1). In this case, 6.11% of the lost points are locations for which all historical GSVs were classified as non-clear. Across all cities, the use of clear-weather images results in smaller coverage loss than the use of non-clear images.

4.3.2. Spatial coverage loss under weather and temporal window constraints

Fig. 7 illustrates, for each city, the monthly distribution of GSV weather conditions (clear vs. non-clear), precipitation levels,[4] and the share of GSV images captured by month. A notable yet intuitively expected pattern is that months with lower precipitation generally correspond to substantially higher proportions of clear-weather imagery, a trend particularly pronounced in Seattle and Portland. For example, in Seattle, 73.76% of images captured in July, the driest month, are clear, whereas 87.14% of images captured in January, a month with high precipitation, are classified as non-clear. In Los Angeles, however, clear-weather imagery remains dominant even during wetter months (January–March). While this may initially appear counterintuitive, the city's comparatively low precipitation, even in its wetter months with an average of around 2.5 in., likely explains this pattern.

Another salient pattern is that in Portland and Seattle, more GSVs were captured during months with lower precipitation (Fig. 7, right), that is, months with higher proportions of clear-weather imagery. Specifically, during the mid-year dry season, the proportion of GSVs captured increases markedly. In Los Angeles, by contrast, the largest share of images was captured in February, the wettest month. However, as noted earlier, the absolute level of precipitation in Los Angeles is low, and the proportion of clear-weather imagery remains high as a result.

[4] Precipitation data were obtained from the U.S. National Centers for Environmental Information (Source: https://www.ncei.noaa.gov/access/us-climate-normals/)

Table 5
Sampling points lost under consistent weather conditions.

City	Sampling Points	Sampling Points Lost (Clear-Only)	Sampling Points Lost (Non-Clear-Only)
Los Angeles	123,799	7569 (6.11%)	42,309 (34.18%)
Seattle	32,672	4287 (10.36%)	10,115 (24.45%)
Portland	41,374	4431 (13.56%)	10,155 (31.08%)

Table 6 extends the coverage loss analysis by restricting it to clear-weather images within specific temporal windows. Unlike Table 5, Table 6 retains a sampling point only if it has a clear-weather image within the specified time window. As expected, imposing additional temporal constraints results in greater overall coverage loss than in Table 5. The most favorable windows, defined by the smallest losses, vary across cities: in Los Angeles, January–April yields the lowest loss (21.70%), whereas in Seattle (27.53%) and Portland (25.78%), May–August is most favorable. These patterns suggest that during these periods, both the share of clear-weather imagery and the number of available GSVs are higher, offering researchers a practical strategy to minimize coverage loss while ensuring valid perception measurement under consistent weather conditions.

5. Discussions

5.1. Key findings

This study demonstrates weather-induced measurement biases (i.e., $Perception_{clear} - Perception_{non-clear}$) when using GSV and computer vision models to assess human perceptions of the built environment. The key findings are as follows.

First, perception scores are systematically lower under non-clear weather than under clear conditions, by approximately 0.259–0.369 points in the case of Los Angeles (see Table 3). Importantly, this effect emerges even after controlling for semantic segmentation and visual complexity variables that proxy built-environment features. Non-clear weather GSVs are also prevalent, comprising approximately 29.66%–39.51% of images across the cities (see Dataset 2 in Table 1), indicating that such biases are likely to occur frequently. Although these differences may appear modest given that the computer vision models trained on the Place Pulse 2.0 dataset operate on a 10-point scale, their implications are nontrivial. For safety, liveliness, and beauty, 79.52%, 76.86%, and 64.73% of GSV images, respectively, fall within a narrow range of 4–6 points across the full sample. In this context, a bias of 0.369 points corresponds to approximately a 7.38% change relative to a mid-range score (score of 5), which is quite substantial. Furthermore, more than 10% of images exhibit a bias exceeding one full point (see Table 2).

These weather-related variations in human perception can be interpreted through the lens of environmental psychology, particularly the concept of perceptual fluency. This mechanism suggests that stimuli that are easier to process visually tend to elicit more positive evaluations even when objective environmental content remains unchanged (Alter & Oppenheimer, 2009; Reber et al., 2004). Clear weather conditions are typically associated with higher brightness, sharper visual clarity, and greater color saturation. Such visual characteristics can facilitate information processing and increase perceptual fluency, which in turn may contribute to higher perceived levels of safety, liveliness, and beauty. From this perspective, weather operates as a contextual modifier of visual experience.

Second, the magnitude of weather-induced biases varies systematically across contexts. Across cities, biases are strongest for safety, followed by liveliness and beauty, and are more pronounced in Los Angeles than in Seattle or Portland (Table 3, interaction term). This may reflect

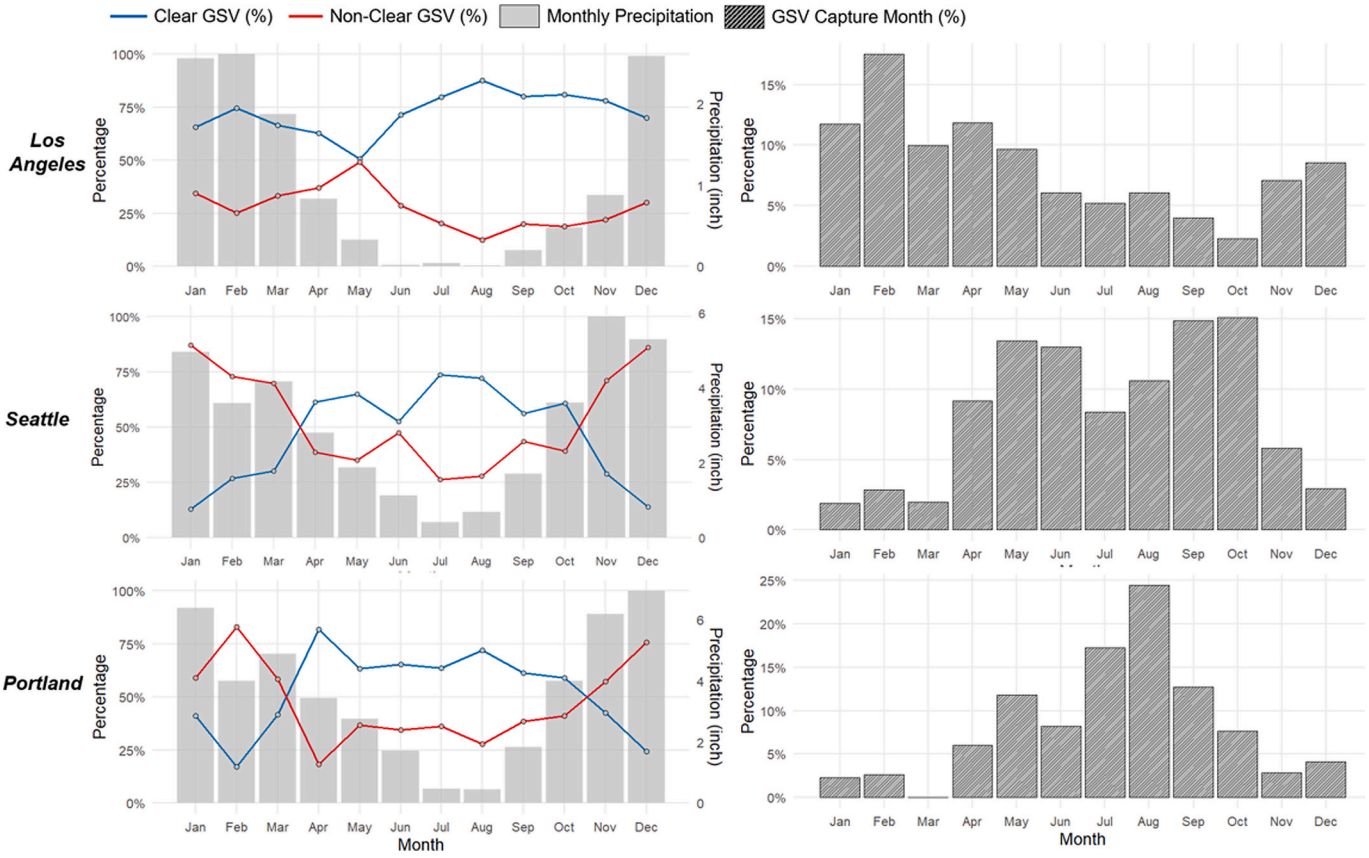

Fig. 7. Monthly distribution of weather conditions, precipitation, and GSV acquisition dates.

Table 6
Sampling point lost under clear-only conditions across temporal windows.

City	Sampling Points Lost (Clear-Only)		
	Jan. – Apr.	May - Aug.	Sep. - Dec.
Los Angeles	26,862 (21.70%)	53,014 (42.82%)	58,546 (47.29%)
Seattle	27,780 (67.14%)	11,390 (27.53%)	20,862 (50.42%)
Portland	28,059 (85.88%)	8422 (25.78%)	21,892 (67.01%)

Los Angeles distinctive visual conditions, where clear weather appears more vivid. As perception scores under clear-weather conditions increase, biases grow consistently across all attributes and cities, approaching zero when scores fall below four but exceeding one point when they rise above six (Fig. 5). This suggests that environments perceived as highly unsafe, unlively, or unaesthetic due to evident built-environment disamenities are relatively insensitive to weather conditions.

In environments with low perception scores, salient negative physical cues, such as visual disorder or infrastructural deficiencies, tend to anchor perception and dominate evaluative judgments regardless of situational context, including weather conditions (Nasar, 1990). As a result, shifts from clear to non-clear weather exert limited additional influence on human perception in such settings. By contrast, in environments with fewer dominant negative cues, evaluative judgments are less strongly anchored by physical environments, allowing weather-related visual conditions to play a more influential role in shaping perceived safety, liveliness, and beauty.

Third, efforts to attenuate the measurement biases by using only GSV images with homogeneous weather conditions inevitably involve a trade-off between measurement validity and the spatial coverage. In Los Angeles, where clear-weather images are dominant (70.34%), restricting all sampling points to clear-weather condition results in only a 6.11% coverage loss, whereas the loss is larger in Seattle (10.36%) and Portland (13.56%) (Table 5). Moreover, applying additional restrictions, such as a temporal window, further increases coverage loss, underscoring the methodological challenge of balancing valid measurement with sufficient spatial coverage.

Finally, although the trade-off between spatial coverage and measurement validity cannot be fully resolved, monthly patterns in GSV availability and clear-weather imagery provide researchers with a strategic avenue to reduce weather-induced biases while minimizing coverage loss. In months with low precipitation, clear-weather images are more prevalent, a pattern particularly pronounced in Seattle and Portland. In these two cities, overall GSV availability is also considerably higher during these drier months (see Fig. 7). While Google does not disclose detailed criteria for GSV collection, its official statement confirms that weather is taken into account: *"We pay close attention to many factors, including the weather … to determine when and where we can collect the best possible imagery"* (Google, 2026). Lauko et al. (2020) similarly note that although efforts are not fully sufficient, Google attempts to secure adequate weather conditions when collecting images. Taken together, these findings imply a favorable circumstance: because GSV availability is higher in months with more clear-weather images, researchers can strategically prioritize such periods in collecting GSVs.

5.2. Implications

This study offers several important implications for future research employing GSV-based measures of human perception. First, our findings call for greater attention to the fact that perception scores derived from GSVs do not solely reflect built-environment features but are also systematically influenced by weather conditions. This finding is particularly critical for studies that use such perception scores as proxies for the built environment and link them to societal outcomes. Biased

measurements of the built environment are not merely methodological issues; they can distort evidence-based planning decisions and lead to a misunderstanding of how built-environment conditions shape societal outcomes (Flyvbjerg, 2007; Hooper et al., 2013).

Insights from environmental psychology suggest that the observed weather-induced bias reflects a fundamental property of human perception. GSV-based perception measures capture not only evaluations of the built environment itself but also momentary experiences shaped by situational factors such as lighting and atmospheric conditions. Accordingly, while these measures are well suited for studying lived experience and environmental exposure, they require careful control and interpretation when employed as proxies for stable indicators of underlying built-environment features.

To our knowledge, no previous study has explicitly accounted for weather inconsistencies when measuring perception scores from SVIs, although our analyses suggest that such discrepancies may occur frequently and can produce substantial measurement differences. While prior research has pointed out broader data-quality heterogeneity in SVIs and its potential to introduce measurement bias (e.g., Zhao et al., 2025), weather-induced bias has been discussed only as a possibility (Ito et al., 2024; Lauko et al., 2020). Our study makes a novel contribution by providing direct empirical evidence of such effects.

Second, our findings provide researchers with concrete guidance on the circumstances under which weather-induced bias is most likely to occur, underscoring the need for greater caution in both study design and interpretation. In particular, the perception of safety is especially sensitive to weather conditions, and high-scoring areas may be more vulnerable to bias. Although further validation is needed for broader generalization, sunnier regions such as Los Angeles tend to exhibit stronger weather-induced biases. These insights imply that researchers should apply stricter weather-consistency filters or conduct sensitivity tests when working in such contexts.

Third, we move beyond merely demonstrating bias to examine its methodological implications for real-world GSV acquisition practices, particularly the trade-off between measurement validity and spatial coverage. We show that this trade-off can be mitigated by selecting city-specific temporal windows in which both the proportion of clear-weather images and the availability of GSVs are high. This strategy provides a practical means of achieving more valid measurement while maintaining adequate coverage.

Fourth, we acknowledge the practical limitations of ideal solutions. Verifying weather conditions for every GSV and restricting analysis to homogeneous imagery would be optimal but is resource-intensive even with automated tools such as the ZenSVI package. As a pragmatic alternative, researchers may instead prioritize GSVs captured during drier months. For example, in Seattle and Portland, drier months (i.e., higher proportion of clear-weather images) coincide with greater image availability. As a result, prioritizing imagery from these periods can help preserve spatial coverage while improving measurement validity. Although it remains unclear whether this pattern reflects Google's systematic acquisition strategy or is specific to the study context, it suggests that the approach is operationally feasible in the case areas.

Fifth, although our analysis is conducted using the Place Pulse 2.0 dataset, the aforementioned implications extend well beyond this particular source. This dataset has been widely adopted for its scale, rich perceptual attributes, and broad global coverage (e.g., Zhang et al., 2018). Yet its limitations, including the lack of respondent sociodemographic information, have motivated the development of alternative SVI-based perception surveys (Danish et al., 2025; Lu & Chen, 2024; Qiu et al., 2022; Song et al., 2025). Our findings highlight that any such effort should explicitly account for weather heterogeneity. In this way, our study offers methodological guidance for the design and application of next-generation SVI-based perception datasets.

5.3. Caveats for interpretation

We caution that the weather-induced differences in perception scores observed in this study should not be interpreted as solely reflecting changes in human perception. Although our perception prediction models are trained on the Place Pulse 2.0 dataset, constructed from human evaluations, the training process of convolutional neural networks (CNNs) offers no guarantee that inferences rely on the same high-level cues used by humans. Instead, CNNs are known to adopt *shortcut learning* strategies, leveraging low-level visual cues that correlate with labels in the training data rather than semantically meaningful features (Geirhos et al., 2020). In our context, this implies that the models may infer lower safety not because of subtle built-environment cues or meaningful atmospheric differences, but simply because darker or lower-saturation images statistically co-occur with unsafe locations in the training data. Consequently, the measurement biases identified here may primarily reflect genuine shifts in human perception under different weather conditions, although there is also some potential for model-specific biases introduced through shortcut learning to amplify these differences. While disentangling these components is beyond the scope of the present study, this methodological limitation is essential to acknowledge when interpreting our findings.

While this study focuses on a binary clear vs. non-clear weather framework and the resulting differences in perception scores, it is important to note that these differences may reflect, at least in part, weather-related differences in visual properties rather than weather as a purely meteorological construct. Prior research shows that degraded weather conditions often appear as reductions in luminance and contrast (Narasimhan & Nayar, 2003), and that image brightness and chromaticity are significantly associated with weather conditions (Granzier & Valsecchi, 2014). This interpretation is consistent with the evidence in *Supplementary Document C*, which shows that manipulating brightness, contrast, and saturation within the same image can produce measurable shifts in perception scores. At the same time, clear vs. non-clear scenes should not be interpreted as outcomes of any single visual property. Rather, they reflect an overall appearance shaped by multiple visual properties that co-occur and are often subtly blended.

The sampling strategy proposed in this study prioritizes city-specific temporal windows to reduce measurement bias while minimizing the loss of spatial coverage. This approach is intended to improve the spatial consistency of perception measures, not to capture season-specific perceptions or residents' year-round evaluations. By restricting image collection to comparable observational conditions across the study area, such as predominantly clear-weather imagery, the strategy helps avoid spatially inconsistent scores in which some neighborhoods are inflated by clear summer images while others are deflated by non-clear or winter scenes. Nevertheless, it may be inappropriate when the research objective is to characterize year-round perceptions or when measures are required for periods outside the proposed temporal windows.

5.4. Limitations

It is important to acknowledge several limitations of this study. First, we employ four computer vision models, and the potential for misclassification must be considered. For example, the ZenSVI weather classifier reports an accuracy of 75.5%, and such errors can influence our results (see *Supplementary Document B*). Second, while weather conditions can be represented along a continuum (e.g., varying degrees of cloudiness), this study adopts a binary clear vs. non-clear framework to enhance clarity and interpretability. In principle, the weather classifier could be extended to produce continuous probability scores (e.g., by applying a softmax function to the model's logits), which would treat weather as a graded construct and could capture more nuanced variation (e.g., "slightly cloudy" vs. "heavily overcast"). However, the binary specification is advantageous for RQ2 because it operationalizes clear-only sampling as an unambiguous include-or-exclude rule for

estimating spatial coverage loss.

Third, caution is warranted regarding the generalizability of our findings. Although we select three cities with distinct weather patterns to strengthen external validity, further exploration across broader climatic contexts is needed. In particular, incorporating additional samples from regions with more diverse climatic conditions to examine how weather-induced measurement biases vary accordingly represents a worthwhile direction for future research. In addition, it is important for future research to examine whether the observed pattern that months with higher proportions of clear images also have greater image availability holds more broadly, given its significant methodological implications. Expanding to additional cities, however, would necessarily require acquiring and processing larger volumes of GSVs, imposing considerable technical demands.

6. Conclusions

This study provides the first direct empirical evidence that weather systematically biases GSV-based perception scores, with clear images producing higher evaluations than non-clear images. These biases are most pronounced for perception of safety and in positively rated environments, underscoring the need for careful consideration of weather effects. Although restricting analyses to homogeneous weather imagery reduces bias, it also entails a trade-off with spatial coverage, which can be partially mitigated by prioritizing drier months when clear-weather images are more abundant. Overall, our findings establish weather as a critical yet overlooked dimension of data quality in GSV-based research and offer important methodological guidance for producing more valid perception measures.

CRediT authorship contribution statement

Donghwan Ki: Writing – review & editing, Writing – original draft, Validation, Software, Investigation, Formal analysis, Data curation, Conceptualization. **Sungmin Lee:** Writing – original draft, Methodology, Investigation. **Chaeyeon Han:** Writing – original draft, Methodology, Investigation. **Youjung Kim:** Writing – original draft, Visualization, Methodology, Investigation. **Bon Woo Koo:** Writing – original draft, Software, Methodology, Data curation. **Uijeong Hwang:** Writing – original draft, Software, Methodology, Data curation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compenvurbsys.2026.102421.

Data availability

Data will be made available on request.

References

Abkar, M., Kamal, M., Maulan, S., & Davoodi, S. R. (2011). Determining the visual preference of urban landscapes. *Scientific Research and Essays, 6*(9), 1991–1997.

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235.

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

Biljecki, F., Zhao, T., Liang, X., & Hou, Y. (2023). Sensitivity of measuring the urban form and greenery using street-level imagery: A comparative study of approaches and visual perspectives. *International Journal of Applied Earth Observation and Geoinformation, 122*, Article 103385.

Danish, M., Labib, S. M., Ricker, B., & Helbich, M. (2025). A citizen science toolkit to collect human perceptions of urban environments using open street view images. *Computers, Environment and Urban Systems, 116*, Article 102207.

Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016, September). Deep learning the city: Quantifying urban perception at a global scale. In *European Conference on Computer Vision* (pp. 196–212). Cham: Springer International Publishing.

Fan, Z., Feng, C. C., & Biljecki, F. (2025). Coverage and bias of street view imagery in mapping the urban environment. *Computers. Environment and Urban Systems, 117*, 102253.

Flyvbjerg, B. (2007). Policy and planning for large-infrastructure projects: Problems, causes, cures. *Environment and Planning B: Planning and Design, 34*(4), 578–597.

Geirhos, R., Jacobsen, J. H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., & Wichmann, F. A. (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence, 2*(11), 665–673.

Google. (2026). Discover When, Where, and How We Collect 360 Imagery. Retrieved from https://www.google.com/intl/en_ch/streetview/how-it-works/.

Granzier, J. J., & Valsecchi, M. (2014). Variations in daylight as a contextual cue for estimating season, time of day, and weather conditions. *Journal of Vision, 14*(1), 22.

Harvey, C., Aultman-Hall, L., Hurley, S. E., & Troy, A. (2015). Effects of skeletal streetscape design on perceived safety. *Landscape and Urban Planning, 142*, 18–28.

Hooper, P. L., Middleton, N., Knuiman, M., & Giles-Corti, B. (2013). Measurement error in studies of the built environment: Validating commercial data as objective measures of neighborhood destinations. *Journal of Physical Activity and Health, 10*(6), 792–804.

Hou, Y., & Biljecki, F. (2022). A comprehensive framework for evaluating the quality of street view imagery. *International Journal of Applied Earth Observation and Geoinformation, 115*, Article 103094.

Hou, Y., Quintana, M., Khomiakov, M., Yap, W., Ouyang, J., Ito, K., … Biljecki, F. (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, X., Wang, S., Yang, D., Hu, T., Chen, M., Zhang, M., … Hohl, A. (2024). Crowdsourcing geospatial data for earth and human observations: A review. *Journal of Remote Sensing, 4*, 0105.

Ito, K., Kang, Y., Zhang, Y., Zhang, F., & Biljecki, F. (2024). Understanding urban perception with visual data: A systematic review. *Cities, 152*, Article 105169.

Ito, K., Zhu, Y., Abdelrahman, M., Liang, X., Fan, Z., Hou, Y., … Biljecki, F. (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*, Article 102283.

Kaplan, S. (1979, April). Perception and landscape: Conceptions and misconceptions[1]. In *, Vol. 35. Proceedings of Our National Landscape: A Conference on Applied Techniques for Analysis and Management of the Visual Resource, April 23-25, 1979, Incline Village, Nevada* (p. 241). Pacific Southwest Forest and Range Experiment Station, Forest Service, US Department of Agriculture.

Kawshalya, L. W. G., Weerasinghe, U. G. D., & Chandrasekara, D. P. (2022). The impact of visual complexity on perceived safety and comfort of the users: A study on urban streetscape of Sri Lanka. *PLoS One, 17*(8), Article e0272074.

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

Koo, B. W., Guhathakurta, S., & Botchwey, N. (2022). How are neighborhood and street-level walkability factors associated with walking behaviors? A big data approach using street view images. *Environment and Behavior, 54*(1), 211–241.

Larkin, A., Gu, X., Chen, L., & Hystad, P. (2021). Predicting perceptions of the built environment using GIS, satellite and street view image approaches. *Landscape and Urban Planning, 216*, Article 104257.

Larkin, A., Huang, T., Chen, L., Lin, P. I. D., Hart, J. E., Zhang, W., … Hystad, P. (2025). Developing nationwide estimates of built environment quality characteristics using street-view imagery and computer vision. *Environmental Science & Technology, 59*(27).

Lauko, I. G., Honts, A., Beihoff, J., & Rupprecht, S. (2020). Local color and morphological image feature based vegetation identification and its application to human environment street view vegetation mapping, or how green is our county? *Geospatial Information Science, 23*(3), 222–236.

Li, X., & Ratti, C. (2019). Mapping the spatio-temporal distribution of solar radiation within street canyons of Boston using Google street view panoramas and building height model. *Landscape and Urban Planning, 191*, Article 103387.

Li, Y., Peng, L., Wu, C., & Zhang, J. (2022). Street view imagery (SVI) in the built environment: A theoretical and systematic review. *Buildings, 12*(8), 1167.

Lu, Y., & Chen, H. M. (2024). Using google street view to reveal environmental justice: Assessing public perceived walkability in macroscale city. *Landscape and Urban Planning, 244*, Article 104995.

Lu, Y., Yang, Y., Sun, G., & Gou, Z. (2019). Associations between overhead-view and eye-level urban greenness and cycling behaviors. *Cities, 88*, 10–18.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282.

Miao, C., Chen, X., & Zhang, C. (2025). Perceived built environment and non-motorist crashes: An exploration with street view imagery. *Journal of Transport Geography, 128*, Article 104363.

Microsoft. (2026). TrueSkill™ ranking system. Retrieved from https://www.microsoft.com/en-us/research/project/trueskill-ranking-system.

Narasimhan, S. G., & Nayar, S. K. (2003). Contrast restoration of weather degraded images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*(6), 713–724.

Nasar, J. L. (1990). The evaluative image of the city. *Journal of the American Planning Association, 56*(1), 41–53.

Qiu, W., Zhang, Z., Liu, X., Li, W., Li, X., Xu, X., & Huang, X. (2022). Subjective or objective measures of street environment, which are more effective in explaining housing prices? *Landscape and Urban Planning, 221*, Article 104358.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review, 8*(4), 364–382.

Rossetti, T., Lobel, H., Rocco, V., & Hurtubia, R. (2019). Explaining subjective perceptions of public spaces as a function of the built environment: A massive data approach. *Landscape and Urban Planning, 181*, 169–178.

Simonite, T. (2017). *Google's new street view cameras will help algorithms index the real world*. WIRED. Retrieved from https://www.wired.com/story/googles-new-street-view-cameras-will-help-algorithms-index-the-real-world/.

Song, Q., Fang, Y., Li, M., van Ameijde, J., & Qiu, W. (2025). Exploring the coherence and divergence between the objective and subjective measurement of streetscape perceptions at the neighborhood level: A case study in Shanghai. *Environment and Planning B: Urban Analytics and City Science, 52*(5), 1231–1251.

Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., & Li, Y. (2022, October). Maxvit: Multi-axis vision transformer. In *European Conference on Computer Vision* (pp. 459–479). Cham: Springer Nature Switzerland.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems, 34*, 12077–12090.

Xu, Y., & Fan, X. (2025). Street view image-based emotional perception modeling of old residential communities: An explainable framework integrating random Forest and SHAP. *ISPRS International Journal of Geo-Information, 14*(12), 471.

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., & Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning, 180*, 148–160.

Zhang, K., Song, Q., Ma, H., Qiu, W., Li, M., & Kim, I. (2025). Synergistic role of audio-visual perceptions in promoting bikeshare for active travel. *Transportation Research Part D: Transport and Environment, 145*, Article 104806.

Zhao, T., Liang, X., Biljecki, F., Tu, W., Cao, J., Li, X., & Yi, S. (2025). Quantifying seasonal bias in street view imagery for urban form assessment: A global analysis of 40 cities. *Computers, Environment and Urban Systems, 120*, Article 102302.

Zhou, H., Wang, J., Widener, M., & Wilson, K. (2024). Examining the relationship between active transport and exposure to streetscape diversity during travel: A study using GPS data and street view imagery. *Computers, Environment and Urban Systems, 110*, Article 102105.

Zhou, H., Wang, J., & Wilson, K. (2022). Impacts of perceived safety and beauty of park environments on time spent in parks: Examining the potential of street view imagery and phone-based GPS data. *International Journal of Applied Earth Observation and Geoinformation, 115*, Article 103078.