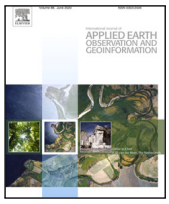
International Journal of Applied Earth Observation and Geoinformation 152 (2026) 105484

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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



Beyond fidelity: Rethinking realism through human behavior in VR city modeling LOD design ☆

Su Yunlei, Yang Guosheng, Zhao Wufan *, Wu Cai

The Hong Kong University of Science and Technology (Guangzhou), Nansha District, Guangzhou, China

ARTICLE INFO

Dataset link: https://github.com/suyunlei/Perception-LOD

Keywords:
Virtual city modeling
Level of detail
Human perception
Eye-tracking
Urban digital twins

ABSTRACT

This study redefines realism in virtual reality (VR) city modeling by proposing a perception-driven Level of Detail (LOD) framework informed by human gaze behavior. Challenging the conventional assumption that higher fidelity equates to greater realism, the study systematically investigates the interplay of Feature Complexity (FC), Appearance (AP), and Semantics (SE) in achieving "perceptual realism". Using eye-tracking technology integrated with VR, we conducted experiments in Guangzhou's Xianfeng Community to identify LOD configurations that most closely align with real-world gaze distribution. Results reveal that the High Feature Complexity, Mid Appearance, Low Semantics (H-M-L) combination optimizes both global gaze patterns and perceptual coherence, outperforming higher semantic density configurations. The study also introduces an open-source tool for analyzing gaze behavior in urban VR environments, providing a scalable solution for immersive urban design research. These findings establish a foundation for advancing human-centric design in VR city modeling, emphasizing perceptual realism over technical fidelity to better align virtual environments with real-world human cognition. This work further positions VR-based perception-driven modeling as a key enabler for Urban Digital Twins, supporting the transition from static 3D models to dynamic, multi-source-data decision-making platforms.

1. Introduction

Digital Twins are driving the transition of cities from static 3D models to dynamic, multi-data-integrated decision-making platforms (Luo et al., 2025; Jeddoub et al., 2023). As an essential component for immersive visualization and interaction, Virtual Reality (VR) technology offers new opportunities for real-time human–computer interaction and cognitive research, enabling high-fidelity and immersive simulations of human behavior within complex urban contexts such as neighborhood regeneration and disaster emergency response training (Fu et al., 2023; Sermet and Demir, 2019).

Human visual attention, as captured by eye-tracking technology, offers a direct window into perceptual prioritization during urban exploration (Dupont et al., 2017; Zhou et al., 2023). For instance, empirical evidence demonstrates that gaze distribution correlates strongly with spatial decision-making and environmental cognition in VR settings (Orquin et al., 2021). By analyzing fixation patterns, researchers can identify critical urban features (e.g., signage, building facades, green spaces) that dominate attentional resources, independent of their geometric or semantic complexity (Naghibi, 2024; Guo et al., 2025).

While virtual reality city modeling offers an innovative approach to studying environmental cognition (Zhang et al., 2018), the pervasive assumption that 'higher complexity equates to higher realism' lacks robust empirical validation. Although tools like the Visual Realism Complexity Scale (VRCS) assess 3D fidelity subjectively (Schmied-Kowarzik et al., 2024), objective behavioral evidence — especially gaze-based perceptual metrics — is still lacking. Recent studies suggest that excessive semantic richness or visual detail may introduce visual clutter, disrupting natural attentional patterns and diminishing the perceptual validity of VR simulations. This aligns with recognition–interference theory, which suggests that excessive stimuli can interfere with attentional focus rather than enhance clarity (Osth and Dennis, 2015; Bearwald, 1976; Sari et al., 2024).

Moreover, current level-of-detail (LOD) frameworks in VR-based urban modeling are predominantly inherited from 3D computer graphics principles and web-based platform standards, which prioritize technical rendering efficiency over perceptual plausibility in human–environment interactions. For example, traditional LOD strategies in VR urban modeling predominantly prioritize technical fidelity—such

☆ This article is part of a Special issue entitled: 'Urban Digital Twins' published in International Journal of Applied Earth Observation and Geoinformation.

* Corresponding author.
E-mail addresses: ysu186@connect.hkust-gz.edu.cn (Y. Su), gyang559@connect.hkust-gz.edu.cn (G. Yang), wufanzhao@hkust-gz.edu.cn (W. Zhao), caiwu@hkust-gz.edu.cn (C. Wu).

https://doi.org/10.1016/j.jag.2026.105484
Received 19 August 2025; Received in revised form 8 June 2026; Accepted 14 July 2026
Available online 28 July 2026

as geometric detail, textures (Graham et al., 2019), whereas web-based frameworks like CityGML define LOD hierarchies according to multi-resolution modeling standards and rendering efficiency constraints (Gröger and Plümer, 2012). Hence, these technical paradigms overlook the cognitive and perceptual mechanisms governing urban engagement. As a result, LOD design in existing VR urban models is not optimized for collecting perception-based gaze data, resulting in mismatches between technical rendering and human-centric perception.

To address this gap, this study proposes a **perception-driven LOD framework** for VR city modeling, grounded in empirical analysis of gaze behavior. We critically examine the prevailing "more is better" assumption by systematically investigating how the interplay of feature complexity (FC), appearance (AP), and semantic information (SE) shapes perceived authenticity (Biljecki et al., 2014). Through a case study of Guangzhou's Xianfeng Community — a rapidly urbanizing neighborhood — we demonstrate that optimal realism is achieved not through maximal complexity, but by strategically balancing these key dimensions.

This study makes the following key contributions: First, it proposes a perception-driven LOD framework that leverages gaze distribution to guide VR city modeling. Second, it develops and releases an open-source eye-tracking analysis tool compatible with commercial VR devices, enabling systematic evaluation of how feature complexity, appearance, and semantics shape gaze behavior. Third, through a case study in Guangzhou's Xianfeng Community, it reveals how LOD configurations affect visual attention and perceived realism. The findings offer insights for perception-aligned VR urban simulations and demonstrate how planners' perceptual preferences can inform participatory urban design in the metaverse.

2. Related work

This section synthesizes the theoretical and methodological foundations of our perception-driven LOD framework for VR city modeling, focusing on three aspects: First, conventional LOD strategies remain static and geometry-oriented, prioritizing efficiency over perceptual or semantic dimensions, with limited human-centric exploration. Second, while eye-tracking has been applied in urban studies to analyze gaze and visual attention, its integration into LOD design is still nascent. Third, insights from cognitive science — such as semantic clutter, cognitive overload, and recognition–interference — underscore the need to balance complexity and semantics to enhance perceptual clarity in VR environments.

2.1. Level of detail in city modeling

The concept of Level of Detail (LOD) originated in 3D computer graphics (Clark, 1976; Heok and Daman, 2004), offering solutions for efficient rendering of complex scenes (Hoppe, 1998; C. Li et al., 2020). LOD techniques dynamically adjust geometric complexity based on object distance or importance, using mesh simplification and approximation methods (Zhu et al., 2010). A variety of advanced methods achieve this by employing mesh simplification or approximation techniques to control geometric complexity (Garland and Heckbert, 1997; Gao et al., 2022). By representing objects with reduced polygon counts or simplified textures when they are far away or less significant, these methods ensure smooth real-time rendering and resource optimization (Biljecki et al., 2015; Takikawa et al., 2021).

Traditional city modeling adopted this principle, emphasizing geometric representations at different scales (Verdie et al., 2015; Pan et al., 2025). Standards like CityGML define several hierarchical LODs, ranging from basic block models to highly detailed architectural forms (Gröger and Plümer, 2012), aimed at managing data and rendering workloads (Biljecki et al., 2015). However, these LOD schemes are typically static, with each urban feature pre-assigned a level of geometric detail that does not adapt to user interaction or specific perceptual needs.

When applied to VR, these static, geometry-centered LODs persist (Astheimer and Pöche, 1994; Reddy, 1995), supporting technical performance but overlooking experiential qualities like immersion and presence (Shin, 2019). The fidelity of geometry and textures remains the primary focus, and the models continue to prioritize rendering efficiency over the nuances of human perception (Shi et al., 2022; Kim et al., 2019).

Recognizing these limitations, recent discussions have begun to address the importance of semantic information in city modeling (Biljecki et al., 2016; Hu, 2008). Semantic LOD refers to the incorporation of meaningful attributes and contextual relationships, such as the function of a building, the presence of signage, or the nature of urban green spaces. By enriching geometric models with semantic content, city models can better support applications that require more than just visual representation, enabling advanced analysis, navigation, or simulation tasks (Chen, 2018; Colucci et al., 2018).

Yet immersive visualization must ultimately be human-centered (Zhang et al., 2002; Lei et al., 2023). The evolution of city modeling in VR increasingly points towards adaptive LOD frameworks that respond to perception and cognition (Çöltekin et al., 2019; Pasalar and Hallowell, 2020). Rather than relying solely on static, geometric, or even semantic hierarchies, emerging approaches advocate for adaptive LOD frameworks that respond to human perceptual and cognitive processes. In this view, detail levels are guided by attention, gaze behavior, or task context, highlighting relevant information while minimizing unnecessary complexity.

2.2. VR eye-tracking and visual attention analysis in urban space

The integration of eye-tracking with VR has opened new possibilities for studying human perception and behavior in urban environments (Zhang and Zhang, 2021; J. Li et al., 2020). By capturing precise gaze data, researchers can quantify visual attention and reveal how people interpret and navigate complex city spaces (Yu et al., 2018; Van Veen et al., 1998). In recent years, VR-based eye-tracking studies have been increasingly applied to a wide range of urban contexts, including green space design, architectural evaluation, signage effectiveness, and wayfinding analysis (Xing et al., 2025; Yu et al., 2023).

In green spaces, VR eye-tracking highlights which elements—such as vegetation or water features—draw attention and shape environmental quality perceptions (Yu et al., 2018; J. Li et al., 2020). In architecture and urban design, it helps assess façades, public art, and street views, improving spatial legibility and aesthetics (Xing et al., 2025). For navigation, mapping gaze patterns identifies bottlenecks, signage clarity, and cognitive processes behind spatial decisions, offering insights for accessibility and safety (Yu et al., 2023; Ye et al., 2022).

Importantly, VR-based eye-tracking provides objective, real-time measures of attention and cognitive load, complementing subjective feedback (Wang et al., 2014). This is crucial for balancing realism and cognitive demands in VR city models, where excessive detail risks clutter and overload (Xing et al., 2024). Yet, whether VR attention patterns fully replicate real-world behavior remains debated: some studies find close alignment (Drettakis et al., 2007), while others note differences due to absent sensory cues (Kroeck et al., 1989). Further research is needed to clarify this relationship.

Recent studies have shown that the signal quality of commercial HMD eye-trackers varies systematically with calibration quality, headset fit, measurement latency, and viewing eccentricity, with precision generally highest in the central visual field and declining towards the periphery (Schuetz and Fiehler, 2022; Stein et al., 2021; Hou et al., 2024; Adhanom et al., 2023; Moreno-Arjonilla et al., 2024). These findings directly motivate the multi-point calibration protocol, per-frame synchronization, and SDK validity-flag filtering we adopted

in our workflow (Appendix A). In addition to these signal-quality controls, binocular VR eye tracking also involves a distinct geometric issue: gaze rays from the left and right eyes are often skew rather than exactly intersecting, so existing work treats point-of-regard estimation through gaze ray casting, midpoint-of-shortest-segment, or least-squares/weighted closest-approach methods (Lamb et al., 2022; Soans et al., 2023). Beyond spatial gaze metrics, VR/XR research increasingly combines eye-tracking with pupillometry to infer cognitive load, perceptual preference, and behavioral responses in immersive settings (Eckert et al., 2021; Lee et al., 2023; Stark et al., 2023; Xie and Zhang, 2024); while the PICO 4 Pro does not expose pupil-diameter data via its SDK, these studies highlight a clear direction for future extensions of our framework. Finally, recent open-source Unity toolchains demonstrate that standardized acquisition and analysis of gaze data in consumer headsets is becoming increasingly accessible and reproducible (Hosp and Wahl, 2023; Kasowski and Beyeler, 2023; Klotzsche et al., 2025), providing the software ecosystem within which our dual-sphere ray-casting pipeline is situated.

2.3. Cognitive and perceptual frameworks in VR city modeling

Understanding how users process information in virtual urban environments requires insights from cognitive science and perceptual psychology. A key concern is visual clutter, the overload of signage, symbols, or annotations that compete for attention and obscure relevant cues (Rosenholtz et al., 2007; Baldassi et al., 2006). In VR city modeling, visual clutter emerges as a critical concern as designers strive to enhance environments with both realistic and informative content (McCabe-Bennett et al., 2020). Although the potential of abstract representations in VR city modeling has been discussed in the literature (Glander and Döllner, 2009; Engel et al., 2012), the balance between visual richness and clarity remains underexplored for urban applications.

Equally important is cognitive load, which reflects the limits of working memory when faced with excessive or complex information (Schnotz and Kürschner, 2007). In immersive urban simulations such as simulations in virtual reality, the drive for realism and detail can easily overwhelm users, leading to reduced comprehension, navigational errors, and diminished engagement (Albus et al., 2021). Visual clutter and excessive information may disrupt attentional processing and reduce perceptual clarity. This can impair task performance by overwhelming attentional resources, underscoring the need for balanced complexity to support focused gaze behavior (Souchet et al., 2022). Additionally, the interference in visual recognition principle highlights the presence of too many competing or ambiguous stimuli can lead to interference—distracting attention and disrupting users' ability to recognize and interpret key elements of the scene (Bruner and Potter, 1964). This insight underscores the importance of selectivity and clarity in presenting urban features, rather than indiscriminate complexity.

Recent research increasingly applies cognitive frameworks to city modeling, focusing on spatial cognition and navigation in large-scale environments (Manley et al., 2021; Lynch, 2023). Yet VR simulations often center on small-scale spaces—streets, buildings, or blocks—where perception is shaped by fine-grained detail (Freiwald et al., 2020; Nakamura, 2021). These contexts demand careful calibration of feature complexity, appearance, and semantics to balance realism with cognitive accessibility (Biljecki et al., 2016). Achieving fidelity thus requires managing clutter, load, and interference to optimize user experience.

3. Methodology

This study employs a systematic methodology to explore how varying Levels of Detail (LOD)—Feature Complexity (FC), Appearance (AP), and Semantics (SE)—affect human gaze behavior in virtual urban environments. Using a VR application integrated with eye-tracking technology, we conducted experiments comparing real-world and virtual street views of Guangzhou's Xianfeng Community.

As shown in Fig. 1, the workflow includes the design of a Unity-based VR program for gaze data collection, the creation of virtual environments with different LODs using SketchUp, and experiments involving participants exploring these scenes. Attention patterns were analyzed using distance-based and similarity-based metrics, complemented by Pearson correlation analysis to investigate the relationship between gaze behavior and scene details. This approach quantitatively identifies the LOD configurations that best replicate real-world visual perception, offering insights for urban design and immersive VR studies.

3.1. VR application design

We developed a Unity application (Unity Engine version $2021.3.33f1c1$) to ensure compatibility and the stability of eye-tracking data collection. After extensive testing with multiple versions, we confirmed that this version offered the most stable performance. The application is based on the PICO Software Development Toolkit (version v3.0.5).

As shown in Fig. 2, we implemented a dual-sphere algorithm to acquire eye-tracking data. Our goal is to project eye-tracking data onto both real and virtual panoramic images. The intuitive approach involves placing the XR Origin (which also corresponds to the Unity Camera position) at the center of the scene, then rendering the panoramic image onto the interior of a sphere. By using a raycast from the user's eye, we detect its intersection with the sphere's interior collider. After obtaining the three-dimensional (3D) coordinates of the intersection point, we can convert it into two-dimensional (2D) coordinates.

However, Unity's built-in sphere does not support direct mapping of UV coordinates on the sphere's interior surface (which would correspond to the 2D coordinates on the panoramic street image). To address this issue, we employed Unity's Sky Dome to render the street texture. Since the Sky Dome is placed at a sufficient distance from the XR Origin, it minimizes distortion in the user's immersive viewing experience.

Next, we used $Cinema4D(C4D)$ to create an inward-facing normal sphere ($SphereR$) to capture the intersection point ($HitA$) between the eye ray and $SphereR$. Along the extension of the eye gaze ray, we placed an empty GameObject that always aligns with the extended eye ray. Near the outside of $SphereR$, we positioned a standard Unity sphere ($SphereN$) with outward-facing normals and an exterior collider. This $SphereN$ captures the reverse ray cast from the GameObject along the eye's gaze direction. The intersection point ($HitB$) on $SphereN$ is then transformed into 2D coordinates using Unity's *RaycastHit.textureCoord* function.

Because $SphereR$, $SphereN$, and the Sky Dome share the same sphere center, the 2D coordinates of $HitB$ on $SphereN$ align precisely with the coordinates on the Sky Dome, thus providing the required 2D coordinates on the street image. Here, the term "dual-sphere" refers to the panoramic UV readout after the intersection between the gaze ray and the panoramic sphere; the gaze ray used for this projection is the SDK-provided combined gaze direction, not an independently fused pair of left eye and right eye rays. Detailed geometry, projection, and calibration are shown in Appendix A.

Overall, the Unity program (as illustrated in Fig. 2) operates as follows: When the user presses the Start button using the Joy-Con controller, they can explore the panoramic street view image. The program captures the user's left and right eye positions, the gaze vector, as well as the 3D world position and 2D UV position of $HitB$. Scenes were presented in randomized order; each session began with a neutral 15 s acclimation panorama, and we inserted a brief interstimulus reset image (5 s) at a random head-centered azimuth/elevation to decorrelate initial gaze/head pose for the next scene. Short rest breaks were added between different sets of LOD scenes to reduce fatigue and temporal entrainment. To make this explicit, participants were instructed to observe each scene naturally without being required to complete a

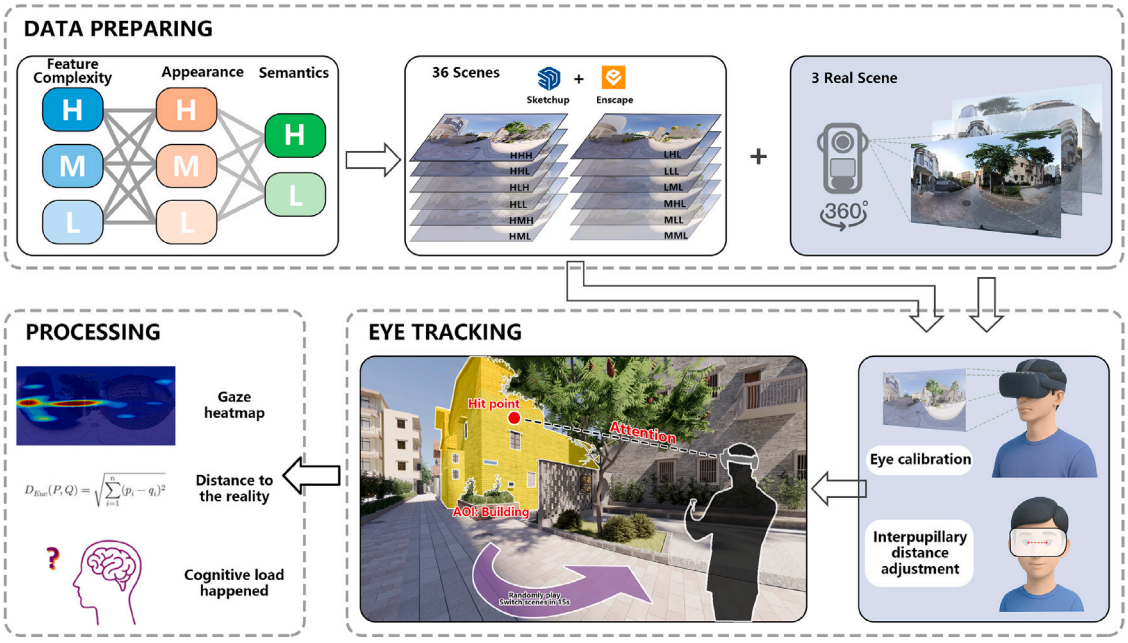


Fig. 1. Workflow of experimental methodology: from lod scene preparation to eye-tracking data processing. This figure illustrates the systematic process used in the study, including the design of a Unity-based VR program for gaze data collection, the creation of virtual urban environments with varying Levels of Detail (LOD) using SketchUp, and the experimental procedure where participants explore these scenes. Attention patterns were analyzed using distance-based and similarity-based metrics, and Pearson correlation analysis was employed to explore the relationship between gaze behavior and scene details. This approach identifies the LOD configurations that most closely replicate real-world visual perception, providing valuable insights for urban design and immersive VR research.

specific search, navigation, decision-making, or memory task. This standardized instruction was intended to provide a common observational frame across participants while preserving free-viewing behavior, consistent with recent VR eye-tracking studies that use immersive free-viewing paradigms to characterize attentional allocation in naturalistic and urban virtual environments (Haskins et al., 2020; Xie and Zhang, 2024). All panoramas were rendered in SketchUp+Enscape with fixed physically based parameters (shown in Appendix B); for each base scene we matched exposure *at render time* across its LOD variants to keep global luminance/contrast consistent, and Unity presented the resulting textures on a sky dome using Unlit/Texture with identical display settings. After the final street image is displayed, a "Finish" button appears. Upon pressing the "Finish" button with the Joy-Con, the eye-tracking data are saved as a CSV file. We apply Kernel Density Estimation (KDE) to generate heatmaps from the recorded data, by computing a *time-weighted gaze occupancy heatmap* by accumulating dwell time over the equirectangular UV domain and smoothing with a Gaussian KDE (bandwidth matched to the device precision, ~1° at the equator). AOIs are defined as polygons in UV (or on the unit sphere); each per-frame gaze sample is assigned via point-in-polygon.

In addition to supporting pre-rendered panoramic images, the VR application was designed with the capability to ingest and visualize geospatial and building information data. This integration enables the platform to serve not only as an experimental environment for perception-driven rendering, but also as a visual interface within Urban Digital Twins frameworks, allowing dynamic updates and multi-source data fusion for real-time decision support.

3.2. Different LOD scenes design

We created a set of urban scene models using SketchUp, systematically varying three key Level of Detail (LOD) attributes: Feature Complexity (FC), Appearance (AP), and Semantics (SE). These dimensions capture the structural, visual, and cognitive qualities of the virtual environment (Fig. 3).

Feature Complexity (FC) refers to the intricacy of object geometry. High FC models exhibit detailed architectural features and realistic forms. Mid FC simplifies these elements, while Low FC reduces them to abstract primitives like cubes and spheres.

Appearance (AP) pertains to surface texture quality. High AP includes high-resolution, photorealistic materials; Mid AP uses moderate-resolution textures; and Low AP omits texturing entirely, rendering plain surfaces.

Semantics (SE) indicates the clarity of meaning conveyed by visual elements. High SE features readable text and recognizable signage, aiding interpretation. Due to geometric and rendering constraints, High SE is only feasible when FC is high; all other configurations default to Low SE, omitting such semantic cues.

In total, we generated **12 unique scene** combinations for each scene, each representing a distinct LOD configuration. These models were rendered using Enscape, a real-time plugin for SketchUp, under controlled conditions for lighting, material reflectivity, and environmental consistency. The resulting panoramic views were displayed in a head-mounted VR application during user testing. This setup enabled immersive visual exploration while capturing how variations in FC, AP, and SE influence human perception, particularly in the legibility of features like greenery and signage. The comparison highlights that perceptual clarity depends less on sheer technical detail and more on a thoughtful balance across LOD dimensions.

To facilitate integration with standard urban modeling frameworks such as CityGML, Table 1 provides an approximate correspondence between our perception-driven LOD dimensions and CityGML 3.0 LOD levels and semantic classes, noting that CityGML emphasizes geometric abstraction while our framework prioritizes human gaze behavior.

3.3. Data processing and metrics

To evaluate the perceptual similarity between virtual scenes with different LOD configurations and their real-world counterpart, we conducted a multi-step analysis of the collected eye-tracking data. This included the extraction of gaze behavior, calculation of gaze dwell-time across semantically labeled Areas of Interest (AOIs), and statistical comparison of gaze distributions using multiple complementary metrics. These AOIs were manually delineated per panorama (by VGG

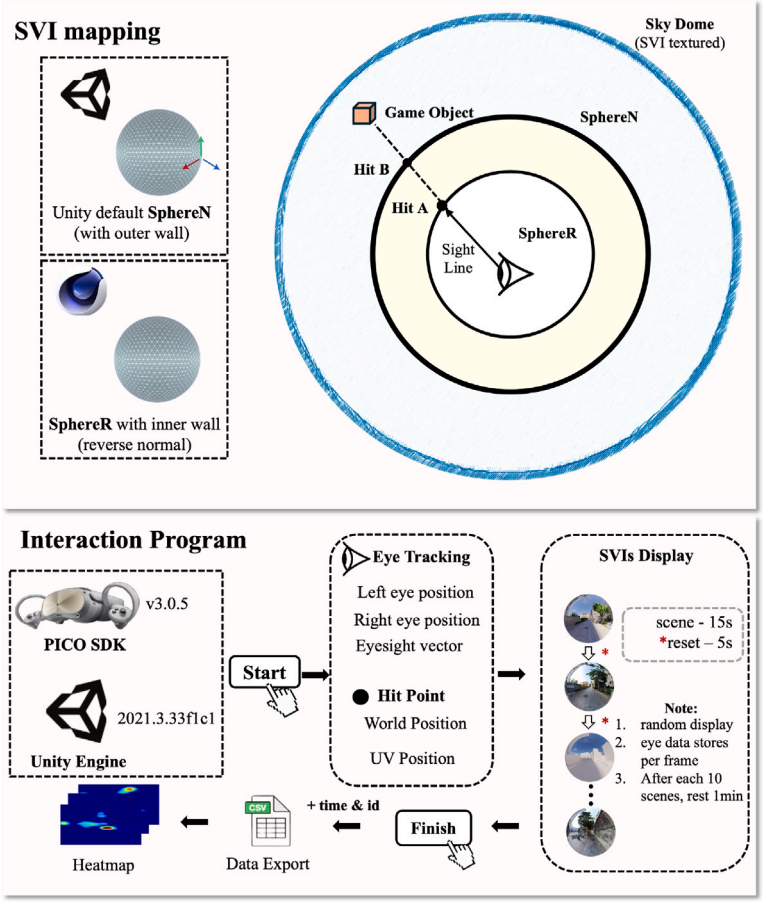


Fig. 2. Unity VR application design for eye-tracking data collection. This figure illustrates the Unity-based application used to project eye-tracking data onto both real and virtual panoramic images. The system uses a dual-sphere algorithm to detect intersections between the user's gaze and the sphere's interior, converting 3D coordinates into 2D UV coordinates. A Sky Dome is used for accurate texture rendering, minimizing distortion. The application captures both the 3D and 2D gaze positions, switching panoramic textures every 15 s and saving the collected data for further analysis. The Kernel Density Estimation (KDE) algorithm is applied to generate heatmaps from the recorded gaze data.

Table 1
Correspondence between proposed LOD dimensions and CityGML 3.0 levels.

Dimension	Low	Mid	High
Feature complexity	LOD0: Abstracted geometry (points, footprints)	LOD1: Prismatic block models	LOD2-3: Detailed architectural forms
Appearance	LOD0-1: No/basic textures	LOD1-2: Moderate textures	LOD2-3: High-res photorealistic
Semantics	Limited (basic labels)	N/A (omitted)	LOD0-3: Rich classes (function, type, surfaces, openings)

Image Annotator; 36 virtual and 3 real equirectangular images, unified resolution), with overlapping AOI labels retained where applicable; per-frame HitUV samples were assigned via point-in-polygon on the corresponding scene's mask. Because LOD edits can remove or merge structures, AOI masks were created and used independently within each condition (Real and each LOD); this study used a single trained annotator, so inter-annotator agreement is not applicable. By integrating both distance-based and similarity-based indicators, as well as correlation analysis, we aimed to identify which LOD composition best replicates real-world visual attention patterns. The following subsections detail each component of the data processing pipeline.

3.3.1. Gaze behavior

We began by analyzing participants' overall gaze behavior within the VR building environment, focusing on key indicators such as gaze dwell-time. Due to limitations in hardware performance and rapid head or eye movements by some users, portions of the raw eye-tracking data were found to be noisy or imprecise. To ensure data quality and reliability, a preprocessing step was conducted to remove invalid or corrupted gaze samples prior to analysis. This initial step established a clean and consistent dataset, enabling a more accurate assessment of visual engagement and attentional load across different LOD conditions, and laying the groundwork for subsequent AOI-based analysis.

3.3.2. AOI proportions and semantic dimensions

To explore how participants distributed their visual attention across different semantic elements, we calculated the gaze dwell-time within each AOI category (e.g., building, green space, sign). These AOIs were predefined and consistent across all virtual and real scenes. This analysis allowed us to examine the influence of feature complexity, appearance, and semantics on gaze allocation.

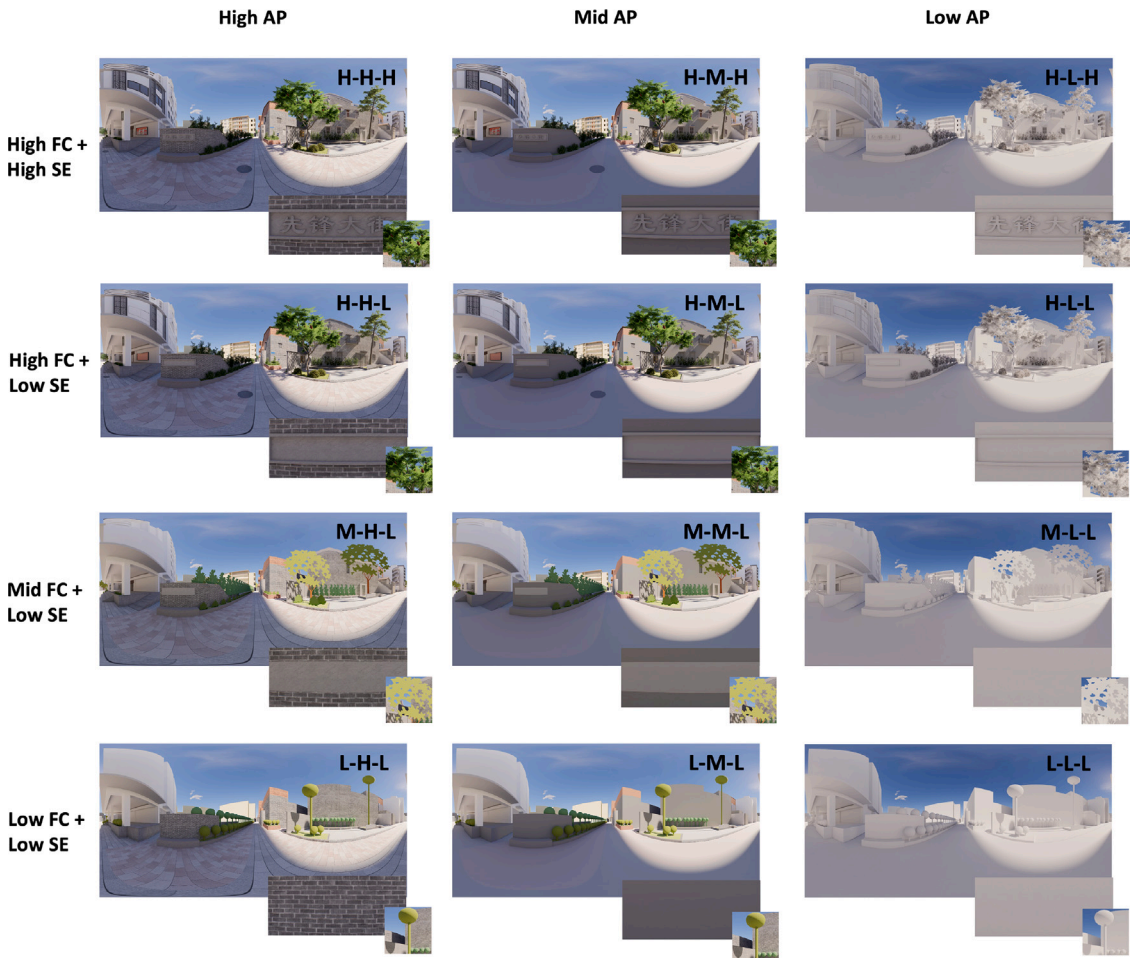


Fig. 3. Virtual urban scenes systematically constructed by varying Feature Complexity (FC), Appearance (AP), and Semantics (SE). Scenes were modeled in SketchUp and rendered in Enscape under consistent visual settings. High Semantics (SE) is only present in configurations where FC is high, as semantically rich features — such as readable signage — require detailed geometry. The figure illustrates how different LOD combinations affect visual expressiveness and perceptual clarity in VR environments.

3.3.3. Gaze distribution patterns

To quantitatively compare gaze distributions across virtual LOD configurations and the real-world street view, we employed five complementary metrics: Euclidean Distance, Cosine Distance, KL Divergence, Average Relative Error (ARE), and Proportional Similarity (see Table 2) (Kim et al., 2021; Distefano et al., 2025). All metrics were normalized to ensure comparability, and their equal-weight sum was computed as a comprehensive similarity score. This multi-metric evaluation enabled us to identify which LOD variants most accurately approximate real-world visual attention. In addition, Pearson correlation analysis was used to examine the relationship between gaze behavior and specific scene features.

Table 2
Metrics for systematic evaluation of gaze distributions across AOI categories.

Metric	Description
Euclidean Distance	Quantifies the mean squared differences in AOI proportions between distributions.
Cosine Distance	Measures the angular similarity between gaze distributions.
KL Divergence	Evaluates the divergence in probability distributions between real and virtual AOIs.
Average Relative Error (ARE)	Computes the mean relative error of AOI proportions.
Proportional Similarity	Assesses the degree of overlap between two gaze distributions.
Comprehensive Score	Aggregates normalized results of all metrics into a single score.

4. Case study

We selected a typical urban renewal neighborhood, Xianfeng Community, located in Guangzhou, China. This area is characterized by distinctive Lingnan architectural features, making it an ideal site to study urban transformation in a context of rich cultural heritage. Within this community, we identified a highly representative urban renewal scene that effectively reflects the diversity of the built environment. We captured high-fidelity panoramic street views of this scene using an Insta360 camera.

The selected scene includes a variety of functional buildings, decorative walls, urban greenery, signage, and detailed architectural elements such as windows, doors, and façade decorations made of different materials. The environment also contains rich background features, including roads, the sky, and manhole covers, which together enhance the realism and complexity of the scene.

Based on these real-world street views, we used SketchUp to construct corresponding virtual environments at different Levels of Detail (LOD). Each scene was then systematically segmented into Areas of Interest (AOIs), encompassing the relevant built environment features.

For the experimental sample, we recruited 30 participants with prior VR experience and without susceptibility to VR motion sickness. The participants group enhances diversity across professional backgrounds, educational levels, and expertise, ensuring broader generalizability of our findings on gaze behavior in VR city modeling. This composition ensures a mix of expert and non-expert perspectives, enhancing the generalizability of our results. Participants wore the PICO 4 Pro VR headset to explore and observe both the virtual and real-world built environments. After cleaning the data (e.g., removing samples with closed eyes or other recording errors), we collected $1,263,900$ valid eye-tracking data points.

4.1. Analysis of gaze behavior and statistical validation in VR Urban environments

After preprocessing the eye-tracking data, we calculated the standard deviation along both the X-axis and Y-axis to identify differences in participants' gaze behavior. The results revealed that the standard deviation along the X-axis ranged from 0.23 to 0.3, while the standard deviation along the Y-axis was approximately 0.1. These findings indicate that, during the VR-based urban perception experiment, participants exhibited a tendency to explore the urban renewal scene more extensively along the X-axis. Specifically, they were more likely to turn their heads or move their eyes horizontally, covering a broader range of observation. Conversely, the limited standard deviation along the Y-axis suggests relatively less vertical movement, indicating that participants were less inclined to look upward or downward during the experiment.

Furthermore, a multivariate analysis of variance (MANOVA) was performed on the eye-tracking dataset to assess the influence of Level of Detail (LOD) on gaze distribution. The analysis yielded Wilks' Lambda $F = 162$, $p < 0.001$, confirming that LOD significantly impacts the spatial distribution of gaze points.

4.2. Impact of AOI proportions and semantic dimensions on gaze behavior in VR urban environments

We analyzed the gaze dwell-time within designated Areas of Interest (AOIs) based on the eye-tracking data collected during the experiment. As illustrated in Fig. 4, across the 36 virtual scenes with varying Levels of Detail (LOD) combinations and 3 real-world scenes, participants consistently directed a higher proportion of gaze points towards buildings compared to background elements such as the sky or roads. This observation highlights the visual prominence of buildings in urban environments and their ability to attract attention in both virtual and real-world contexts. Moreover, the standard error associated with gaze proportions in building AOIs was notably larger compared to that of background AOIs across all scenes. This suggests greater variability in participants' gaze behavior when observing buildings, potentially reflecting individual differences in how architectural features are perceived or prioritized.

The semantic dimension encompasses the identification and classification of urban elements based on their functional roles and contextual meanings within the built environment. In our experimental scenes, specific architectural features — such as doors, windows, decorative elements, and signage — are regarded as semantically significant entities due to their ability to convey distinct attributes and information. These detailed elements play a crucial role in shaping the perception and understanding of urban spaces.

To quantify the impact of semantic information, we categorized three dimensions — Feature Complexity, Appearance, and Semantics — into High (H), Medium (M), and Low (L) levels, assigning numerical values of 3, 2, and 1, respectively. Using Pearson correlation coefficients, we then analyzed the relationship between the dwell-time within Areas of Interest (AOIs) and these three dimensions. Notably, the architectural components classified as detailed elements (i.e., doors, windows, decorative elements, and signage) exhibited a strong correlation with the semantic dimension, with a Pearson correlation coefficient reaching 0.73. This result highlights the significant influence of semantic information in guiding participants' visual attention towards detailed urban features, underscoring its pivotal role in human perception within virtual environments.

4.3. Analyzing gaze distribution patterns across virtual and real-world street views

Using kernel density estimation, we generated heatmaps for virtual street view under different LOD combinations as well as for real-world street view, as illustrated in Fig. 5. The heatmaps highlight regions of concentrated gaze density, with red and yellow areas indicating the highest density of participant attention. These regions predominantly correspond to critical architectural details, such as doors, windows, decorative elements, and signage. This observation suggests that areas with rich semantic information are particularly effective in capturing visual attention within both virtual and real-world environments.

The heatmaps reveal a clear trend: gaze fixation is distributed more broadly along the horizontal axis (X-axis) and is relatively constrained along the vertical axis (Y-axis). This indicates that participants tend to focus on the lateral structures of buildings, such as facades, rather than exploring vertical elements like height or elevation. Furthermore, participants demonstrated a preference for observing details of buildings located farther away from their immediate position, rather than focusing on nearby structures. This behavior could be attributed to the spatial layout or visual prominence of distant elements in the street view. And across the gaze-density heatmaps of all LODs, H–M–L exhibits the closest hotspot pattern to the Real condition.

Additionally, we observed a stark contrast between buildings and background elements such as the sky and roads. Across all experimental scenes, buildings consistently attracted more gaze dwell-time compared to background elements, underscoring their visual dominance and functional importance in urban perception tasks. These findings emphasize the role of architectural details and semantic richness in guiding human attention, highlighting the critical influence of LOD design in virtual urban modeling.

We conducted a comprehensive assessment using five distinct metrics, each addressing a different dimension of comparison: absolute feature differences (Euclidean distance), feature direction consistency (Cosine distance), data distribution alignment (KL divergence), localized error (Average relative error), and proportional relationships (Proportional similarity) (Distefano et al., 2025; Yamamoto et al., 2025; Spertus et al., 2005). Formal definitions and implementation details are given in Appendix C. The evaluation of differences in gaze distributions between virtual LOD combinations and the real-world street view, along with the results derived from distance-based and similarity-based metrics, are presented in Fig. 6.

These metrics collectively provide a multi-dimensional framework for evaluating the similarity between gaze distributions in virtual and real-world environments. The analysis reveals that the High Feature Complexity, Mid Appearance, and Low Semantics (H-M-L) combination emerged as the configuration with the smallest disparity in gaze distribution when compared to the real-world street view. The lowest score for this combination indicates that human gaze behaviors in this virtual LOD combination are most similar to those observed in real-world environments.

5. Discussion

5.1. Perception-driven LOD framework

This study challenges the entrenched assumption that higher graphical fidelity guarantees greater realism in VR city models. Instead of prioritizing technical detail, we emphasize perceptual realism —

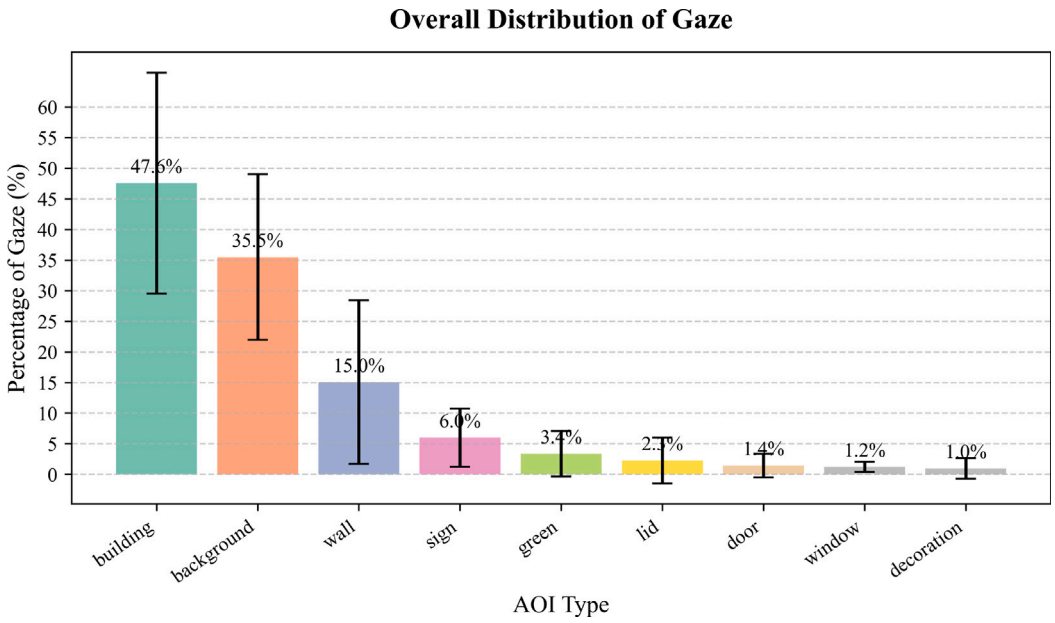


Fig. 4. Distribution of eye-tracking data across Areas of Interest (AOIs) in virtual scenes with varying LODs. Participants consistently focused more on buildings than background elements and other detail elements.

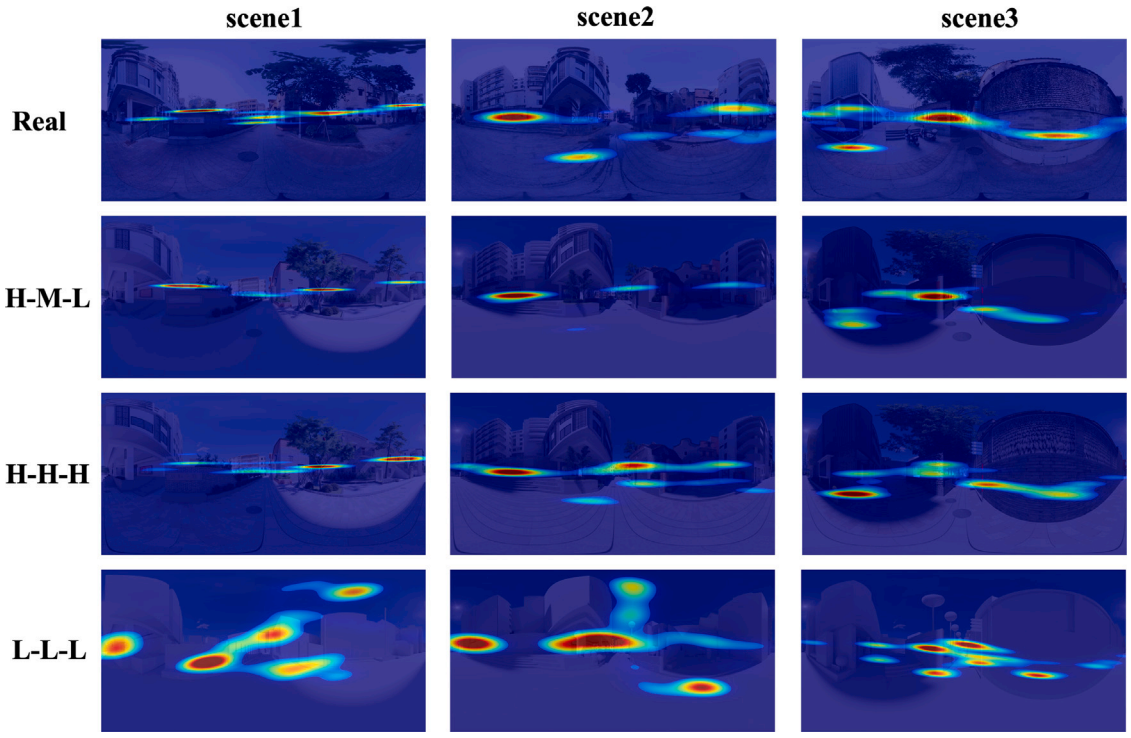


Fig. 5. Gaze density heatmaps on specific LOD configurations and real-world street views, generated via kernel density estimation. Warmer colors represent regions with higher visual attention, primarily concentrating on semantically rich architectural features (e.g., doors, windows, signage). Results reveal a horizontal bias in gaze distribution and a preference for distant buildings, underscoring the visual dominance and perceptual importance of built structures in both virtual and real urban scenes.

defined by alignment with natural gaze behavior — thereby reframing realism as "attentional coherence" rather than pixel accuracy. Throughout this paper, the term *perceptual realism* refers specifically to the degree of alignment between gaze distributions in a virtual LOD scene and its real-world counterpart, as measured by the five complementary metrics described in Section 3.3.3. This is an *attentional* definition of realism: it captures how closely a virtual scene reproduces the natural visual attention patterns elicited by the corresponding real environment, and does not extend to task-based cognition, functional performance, or goal-directed behavior.

Our results show that perceptual realism — how closely a model reflects natural human attention patterns — does not correlate directly with increased detail in the environment. The configuration that performed best, High Feature Complexity, Mid Appearance, and

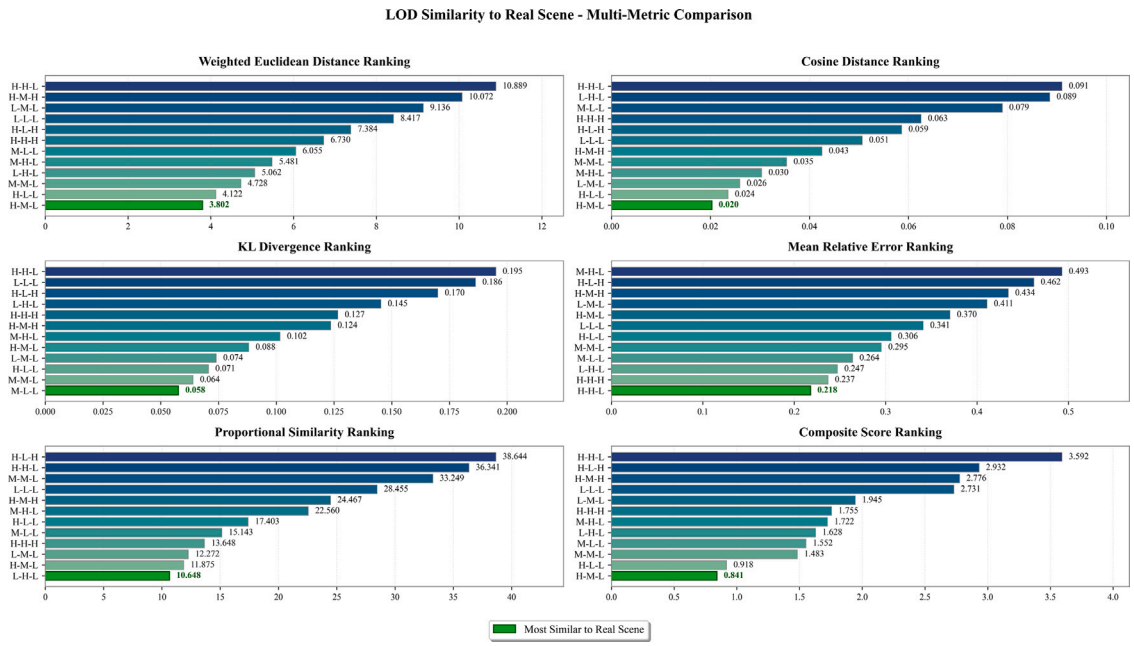


Fig. 6. Comparison of gaze distribution similarity between various virtual LOD combinations and real-world street views using five quantitative metrics: Euclidean distance, Cosine distance, KL divergence, Average relative error, and Proportional similarity. The results demonstrate that the LOD configuration with High Feature Complexity, Mid Appearance, and Low Semantics (H-M-L) most closely aligns with real-world gaze patterns across most metrics, with a comprehensive similarity score of 0.841. This suggests that the H-M-L combination best replicates human visual behavior observed in real street environments.

Low Semantics (H-M-L), strategically balances visual richness with attentional economy. This balance maximizes both global gaze coherence and focused attention, challenging the long-standing approach of increasing detail to increase realism in city modeling (Döllner and Buchholz, 2005).

This human-centered approach advances behavior-driven LOD frameworks, aligning with recent work in digital twins and immersive urban modeling that prioritize perceptual effectiveness over sheer detail (Lei et al., 2023; Konis, 2013; Kuru, 2023). By embedding gaze data into VR city models, we lay the foundation for adaptive simulations that respond dynamically to user cognition, offering more intuitive and personalized urban experiences. The implications extend to participatory planning and sustainable urban development, where perception-informed models can enhance perceptual focus, decision-making, and the realism of urban digital twins (Shakibamanesh et al., 2025).

5.2. Semantic recognition and visual clutter theory

Our findings reveal a nuanced relationship between semantic richness and attentional clarity in VR urban environments. While dense configurations such as H-H-H (High Feature Complexity, High Appearance, High Semantics) may appear more realistic, they often introduce visual clutter that fragments gaze patterns and disrupts holistic scene comprehension. This supports the recognition–interference framework (Osth and Dennis, 2015), which argues that excessive stimuli can fragment gaze patterns and disrupt attentional coherence.

By contrast, the H-M-L (High Feature Complexity, Mid Appearance, Low Semantics) configuration demonstrates how reducing non-essential semantics fosters perceptual coherence and more natural gaze flows. Informal post-exposure responses suggested that participants found information-dense scenes, particularly those with abundant signage, harder to process, an impression directionally consistent with the eye-tracking finding that High-Semantics configurations showed the largest divergence from real-scene gaze distributions. However, as the questionnaire was not formally analyzed, this observation is treated as supplementary context rather than independent evidence. These results contribute to the growing discourse on attention-aware abstraction and minimalist information design in immersive systems, where the goal is not to maximize informational density but to strategically curate visual and semantic input to support human cognitive processes. This perspective is gaining traction in both urban simulation and cognitive visualization, where excessive realism is being re-evaluated in favor of functional and perceptually optimized environments.

Empirically, our study shows that semantic simplification, when paired with sufficient geometric and visual detail, enhances perceptual clarity and supports cognitively sustainable VR city models. This has broader implications for accessible urban simulations, ensuring perceptual accessibility across diverse user groups. Moreover, by linking eye-tracking with adaptive LOD strategies, we provide evidence for gaze-driven simulation frameworks that prioritize both computational efficiency and user-centered realism—an essential step as immersive technologies become integral to participatory planning and digital twin development.

5.3. Gaze-informed modeling tools and the "experiment-as-a-service" paradigm

A key methodological contribution of this study is the development of an open-source, modular tool that integrates eye-tracking analytics into VR city modeling. Following the "Experiment-as-a-Service" paradigm (Amaxilatis et al., 2018), it provides a flexible platform where researchers can import datasets, define custom Areas of Interest (AOIs), and conduct real-time or post-hoc gaze analysis without advanced VR expertise.

By embedding perceptual analytics into the LOD pipeline, the tool enables data-driven evaluation of how visual, semantic, and geometric complexity shape attention and spatial understanding. Unlike conventional LOD methods focused on rendering efficiency, it adds a human-centric layer that captures actual perceptual attention and gaze behavior—crucial for wayfinding, attention allocation, and spatial learning in immersive simulations.

This approach addresses the growing demand for reproducible and behavior-informed modeling tools in digital twin development, participatory planning, and urban cognitive simulation. It aligns with computational behavioral urbanism by bridging spatial computation and cognition, ensuring immersive platforms are not only technically functional but also perceptually coherent.

Beyond research, the tool supports professional practice by allowing designers and planners to test visual accessibility, landmark legibility, or demographic-specific navigation patterns. In doing so, it advances inclusive, adaptive, and human-centered approaches to VR city modeling.

5.4. Limitations and future directions

Despite its contributions, this study has several limitations that should be acknowledged. First, the analysis was confined to static urban scenes, excluding dynamic elements such as moving vehicles, pedestrians, or interactive features. Future research should incorporate these components to simulate more realistic urban scenarios and examine their impact on gaze behavior. In addition, future work could incorporate multimodal metrics such as pupillometry to explore proxies for cognitive load, provided that current hardware limitations can be addressed.

All AOI masks in this study were produced by a single trained annotator; inter-annotator agreement was accordingly not measured. As the dataset scales to larger or more diverse urban scenes, adopting dual independent annotation with formal agreement reporting will be an important quality-control step for future research (Cordts et al., 2016).

Also, while we attribute the observed horizontal bias in gaze distribution primarily to the strong horizontal structure of the urban streetscape (e.g., building façades, road lines, skyline), and while scene randomization and reset images were used to reduce sequential dependence, we acknowledge that the optical characteristics of the HMD, such as its wider horizontal field of view may also contribute to this pattern. The present study does not formally disentangle device-induced effects from scene-structure-driven exploration; future work using matched real-world and HMD viewing conditions, or cross-device replication, would help clarify this distinction (Hou et al., 2024; Sipatchin et al., 2021).

Furthermore, while the PICO SDK's multi-point calibration, per-frame validity flags, and static panoramic presentation were used to ensure data quality, no eccentricity-based uncertainty model was applied to account for the reduction in gaze-tracking precision that typically occurs towards the edges of the visual field (Holmqvist et al., 2011; Nyström and Holmqvist, 2010). Future work, particularly studies conducted in full 3D VR environments where participants make larger head movements, would benefit from incorporating such a model to produce more reliable gaze measurements across the full viewing range.

A further set of limitations concerns the experimental design and the interpretation of participant responses. The post-exposure questionnaire employed in this study was designed primarily to orient participants towards the key perceptual dimensions of the experiment and to provide informal supplementary support for the eye-tracking findings, rather than to constitute a formal analytical instrument. Developing a more rigorous, psychologically grounded questionnaire, drawing on established subjective measurement frameworks in environmental perception and cognitive psychology, and integrating the resulting data with objective gaze metrics in a systematic mixed-methods analysis represents an important direction for future work.

A further limitation of the present study is that it employed a free-viewing paradigm rather than task-based conditions. Consequently, the AOI-based results should be interpreted as reflecting spontaneous visual exploration and attentional allocation, rather than goal-directed functional behavior such as search, navigation, or decision performance. We note, however, that introducing an explicit task would also have imposed task-dependent biases on gaze behavior, as widely reported in the scene-viewing literature (Castelhano et al., 2009; Cronin et al., 2020); the present design was therefore deliberately chosen to establish a clean baseline of attentional allocation across LOD conditions. Future work incorporating task-based conditions, such as navigation or semantic identification tasks, would allow the LOD framework to be validated against a wider range of perceptual and cognitive behaviors.

Beyond these methodological considerations, the generalizability of the findings should also be further examined. Additionally, cross-cultural studies could investigate the applicability of LOD combinations across diverse urban contexts, offering insights into universal and culture-specific aspects of urban perception. Expanding beyond eye-tracking, future work could incorporate multimodal metrics such as EEG, auditory cues, and emotional responses to further validate perceptual realism and extend the application of the proposed LOD framework.

Finally, this study used pre-rendered equirectangular panoramas rather than real-time full 3D scenes. This was a deliberate design choice: panoramic imagery provides consistent ray–image intersections, removes variability from mesh-collider resolution, and ensures stable rendering performance across all LOD conditions. However, it also means that binocular vergence-based depth estimation was not possible, and that the findings may not fully generalize to interactive 3D VR environments. Future studies using interactive full-3D geometry or gaze-depth measures should retain per-eye gaze rays and evaluate explicit skew-ray fusion strategies, such as midpoint-of-shortest-segment or weighted closest-approach estimation, together with careful synchronization of head- and eye-tracking coordinate frames. Full 3D modeling combined with binocular convergence analysis remains an important direction for future work (Holmqvist et al., 2011).

6. Conclusion

This study presents and empirically validates a perception-driven Level of Detail (LOD) framework for VR-based city modeling, grounded in eye-tracking analytics and cognitive-perceptual theory. Moving beyond conventional assumptions that equate higher graphical fidelity with increased realism, we demonstrate that perceptual realism — defined as alignment with human attention and information processing patterns — is more effectively achieved through a strategic balance of feature complexity (FC), visual appearance (AP), and semantic content (SE).

Through a case study of Guangzhou's Xianfeng Community, we identified the H-M-L configuration (High Feature Complexity, Mid Appearance, Low Semantics) as the most perceptually coherent among the tested LOD variants. This configuration closely mirrored real-world gaze distributions and supported smoother attentional flow, suggesting that reducing semantic overload while preserving key structural and visual cues can significantly improve attentional clarity in immersive environments. These findings challenge prevailing LOD design logics in 3D urban modeling and contribute to a broader shift towards human-centered, gaze-informed simulation paradigms.

Moreover, we developed an open-source, modular tool for conducting gaze-based LOD evaluation, offering a reproducible infrastructure that can be readily adapted to other urban contexts, demographic groups, or research questions. This tool operationalizes the concept of "Experiment-as-a-Service" and provides a scalable pipeline for integrating perceptual data into urban simulation, evaluation, and design workflows.

Ultimately, this research contributes to the growing discourse on computational behavioral urbanism, where immersive technologies are leveraged not merely as visualization platforms but as "perceptual instruments for informing the design of human-centered urban VR environments". As VR becomes increasingly embedded in urban planning, participatory design, and public consultation processes, it will be essential to incorporate real behavioral data — such as gaze patterns — into

virtual environment construction. Doing so not only enhances technical feasibility and computational efficiency, but also ensures that virtual city models remain cognitively accessible, experientially authentic, and responsive to the perceptual capacities of diverse users.

Looking ahead, this perception-driven LOD framework lays the groundwork for next-generation immersive urban tools that prioritize attentional clarity, inclusiveness, and perceptual economy—core principles in the design of meaningful digital urban futures.

CRediT authorship contribution statement

Yunlei Su: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Guosheng Yang:** Writing – original draft, Methodology. **Wufan Zhao:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Cai Wu:** Writing – review & editing, Supervision, Project administration.

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

During the preparation of this work the authors used ChatGPT, a large language model developed by OpenAI, in order to polish and correct the English expressions. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding and acknowledgements

The authors would like to thank all participants who took part in the experiments. This work was supported by the Nansha District Metaverse Innovation Project: A Public Experience and Interaction Platform for the National Games (Grant No. 2024YYZ001), Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 42401567), the Guangzhou Municipal Science and Technology Bureau Program (Grant No. 2025A03J3640), and the Foundation of Shanxi Key Laboratory of Watershed Built Environment with Locality (Grant No. WaBEL-KFYB-2025-01).

Appendix A. Eye-tracking geometry, projection, error assessment, and calibration

A.1. Coordinate frames

The sky dome is a sphere centered at $\mathbf{C} = (0, 0, 0)$ of radius R (Unity Y-up, Z-forward, X-right). Left and right eye poses are denoted by $\mathbf{O}_L$ and $\mathbf{O}_R$. For panoramic projection, the ray origin $\mathbf{O}$ was defined as the midpoint $(\mathbf{O}_L + \mathbf{O}_R)/2$ when both eye poses were available, and the ray direction $\mathbf{d}$ was the PICO SDK-provided combined gaze direction transformed into the Unity world coordinate frame.

A.2. Binocular fusion and skew-ray scope

In real HMD eye tracking, the gaze rays from the left and right eyes are generally skew and do not necessarily intersect at a single 3D point. Existing VR eye-tracking methods therefore estimate binocular point of regard using direction averaging, midpoint-of-shortest-segment estimation, or least-squares/weighted closest-approach methods (Lamb et al., 2022; Soans et al., 2023).

The present study does not reconstruct vergence depth from separate left eye and right eye rays. Because our stimuli are static equirectangular panoramas, the analysis uses projected UV/AOI gaze distributions rather than 3D fixation depth. Accordingly, the dual-sphere procedure refers to panoramic UV readout after intersecting the combined gaze ray with the panoramic sphere, not to an independent binocular ray-fusion algorithm.

A.3. Ray–sphere intersection

With $\mathbf{r}(t) = \mathbf{O} + t\,\mathbf{d}$ and $\|\mathbf{d}\| = 1$,

$$\|\mathbf{O}-\mathbf{C}+t\mathbf{d}\|^2 = R^2 \Rightarrow t = -\mathbf{d}(\mathbf{O}-\mathbf{C}) \pm \sqrt{\left(\mathbf{d}\cdot(\mathbf{O}-\mathbf{C})\right)^2 - \left(\|\mathbf{O}-\mathbf{C}\|^2 - R^2\right)}.$$

A.4. Equirectangular mapping

Let $\mathbf{n} = (\mathbf{P}-\mathbf{C})/\|\mathbf{P}-\mathbf{C}\|$. Define

$$\lambda = \operatorname{atan2}(n_x, n_z), \qquad \varphi = \arcsin(n_y).$$

Then

$$u = \frac{\lambda + \pi}{2\pi}, \qquad v = \frac{\frac{\pi}{2} - \varphi}{\pi}, \qquad x = uW, \quad y = vH.$$

A.5. Error propagation (Sky dome finite R)

If the eye offset satisfies $\|\Delta\| \ll R$, then

$$\delta\theta_{\text{parallax}} \lesssim \arcsin\left(\frac{\|\Delta\|}{R}\right) \approx \frac{\|\Delta\|}{R}.$$

A.6. Validation and data quality

We cross-check analytic UVs with Unity's `RaycastHit.textureCoord` (agreement away from the seam) and summarize per-session angular/UV errors after applying the invalid-sample filtering described in Section 3.3.1.

A.7. Calibration

Device fit and IPD (pre-session). Participants completed PICO system–level IPD adjustment and facial fit. They then fixated a target at ~2–3 m to confirm alignment and minimize eyelid occlusion.

Eye-tracker calibration (re-run as needed). We executed the SDK's calibrate eye tracking multi-point routine, as shown in Fig. 7, until the on-device quality indicator reported *Pass* (else: *"Recalibrate"*). To avoid drift, we repeated calibration after each scene group or when visible drift was observed.

During-session quality control (per-frame). Head pose, eye samples, the 3D intersection point, and the 2D (u, v) coordinates were logged within the *same frame* to avoid cross-frame offsets. When both eyes were valid, we used the SDK-provided combined gaze direction for panoramic projection. Invalid samples (blinks, occlusions, tracking loss) were removed using SDK validity flags before analysis.

Table 3
Controlled parameters used in SketchUp+Enscape+Unity (fixed across scenes/LODs unless otherwise noted).

Category	Parameter	Example value	Notes
Renderer	Toolchain	SketchUp 2019 + Enscape 3.5	Export equirectangular panoramas (2:1)
Lighting	Sky/Sun model	Clear sky; Sun alt. = 45°, az. = 120°; int. = 1.0	Same time-of-day for all renders
Post	Vignette/DOF/Bloom/FX	**OFF**	No post-processing
Output	Panorama res./format	8192 × 4096; PNG or JPEG (Q = 95)	Same codec/resolution across LODs
Materials	Diffuse reflectance (ρ)	Asphalt $\rho = 0.12$; Concrete $\rho = 0.50$; Vegetation $\rho = 0.25$	Constant across LODs
Materials	Metallic/Roughness	Metallic = 0.0; Roughness = 0.50 (typical)	PBR coefficients fixed
Materials	Glass	IOR = 1.50; Trans. = 0.90; $F_0 = 0.04$	Same glazing settings
Unity	Shader/pipeline	Unlit/Texture; Linear space; sRGB sampling	No dynamic lights; no post
Unity	Sky dome	Radius = 50 m; seam behind participant	Same geometry/projection
HMD	Device/display	PICO 4 Pro; brightness = 80% (auto OFF)	Same device/settings for all participants

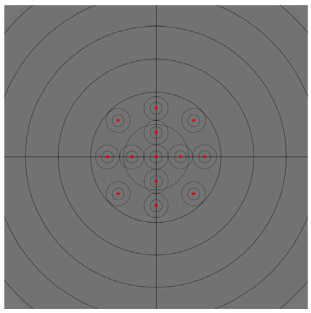


Fig. 7. Calibration interface (SDK multi-point eye calibration targets) used in our sessions (PICO 4 Pro).

Appendix B. Controlled rendering parameters and render-time exposure matching

B.1. What is fixed vs. What is manipulated

Fixed across LODs (rendering stage): Sky/solar model and orientation, tone mapping curve, white balance, exposure *mode* (Auto OFF), post effects (OFF), material physical coefficients (diffuse reflectance ρ, metallic, roughness, glass IOR/transmittance), camera projection and panorama resolution/format, as shown in Table 3.

Manipulated by LOD: Geometric detail and presence/simplification of textures.

Presentation stage (Unity): Unlit/Texture, Linear color space with sRGB texture sampling, no dynamic lights, no fog/post, identical sky-dome radius/projection and identical HMD display settings.

B.2. Render-time exposure matching (Per base scene)

For each base scene, we choose the H-H-H scene as the reference and compute linear-domain mean luminance μ^*. We then adjust Enscape exposure EV for the rest of LOD configurations to satisfy $|\mu - \mu^*| \le 2\%$.

Appendix C. Metrics for comparing gaze AOI distributions

C.1. Data representation and preprocessing

For each scene we form an AOI percentage vector

$$\mathbf{p} = (p_1, \dots, p_K) \in \mathbb{R}^K, \qquad \sum_{i=1}^{K} p_i = 100,$$

where K is the number of AOIs (e.g., `building`, `background`, `wall`, `green`, `window`, `sign`, ...). The Real scene provides $\mathbf{q} = (q_1, \dots, q_K)$. These percentages are computed from per–frame gaze samples using the same AOI definitions across conditions.

For probability–based metrics we use $p_i^{\text{prob}} = p_i/100$, $q_i^{\text{prob}} = q_i/100$ and apply a small floor $\epsilon = 10^{-10}$ followed by renormalization to ensure numerical stability:

$$\tilde{p}_i = \frac{\max(p_i^{\text{prob}}, \epsilon)}{\sum_j \max(p_j^{\text{prob}}, \epsilon)}, \quad \tilde{q}_i = \frac{\max(q_i^{\text{prob}}, \epsilon)}{\sum_j \max(q_j^{\text{prob}}, \epsilon)}.$$

All metrics are computed on the same AOI set and vectors for a fair comparison, and all are oriented as "smaller = more similar".

C.2. Metric definitions

(1) Euclidean distance.

We emphasize perceptually salient AOIs using fixed a priori weights w_i (here: $w_{\texttt{building}}$=0.35, $w_{\texttt{background}}$=0.25, $w_{\texttt{wall}}$=0.20, $w_{\texttt{green}}$=0.10, $w_{\texttt{window}}$=0.05, $w_{\texttt{sign}}$=0.05). The distance is

$$d_{\text{Euc}}(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^{K} w_i \, (p_i - q_i)^2}.$$

This provides an intuitive baseline of overall absolute deviation while allowing domain–informed emphasis.

(2) Cosine distance (pattern consistency).

Cosine similarity tests pattern (relative weighting) consistency; we use cosine *distance*

$$d_{\text{Cos}}(\mathbf{p}, \mathbf{q}) = 1 - \frac{\mathbf{p} \cdot \mathbf{q}}{\|\mathbf{p}\|_2 \, \|\mathbf{q}\|_2},$$

which remains small when the AOI pattern matches even if overall scale differs.

(3) KL divergence (probability allocation mismatch).

We anchor KL at the Real distribution to penalize allocating mass to AOIs that are unlikely under Real:

$$d_{\text{KL}}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{K} \tilde{q}_i \log \frac{\tilde{q}_i}{\tilde{p}_i}.$$

(4) Average relative error (ARE; percentage–scale deviation).

To report percentage–scale errors while avoiding instability in tiny AOIs, we average only over AOIs with $q_i > 1\%$:

$$d_{\text{ARE}}(\mathbf{p}, \mathbf{q}) = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \frac{|p_i - q_i|}{q_i}, \qquad \mathcal{I} = \{i : \; q_i > 1\}.$$

(5) Proportional similarity (ratio–structure deviation over primary AOIs).

To capture the *relative share structure* among primary AOIs, we compare all unordered pairs (i, j) within $\mathcal{S}$ = {`building`, `background`, `wall`, `green`} using pairwise ratios with a small floor τ=1% to avoid division by very small denominators:

$$R_{\mathbf{p}}(i,j) = \frac{p_i}{\max(p_j, \tau)}, \quad R_{\mathbf{q}}(i,j) = \frac{q_i}{\max(q_j, \tau)}, \quad i < j, \; i, j \in \mathcal{S}.$$

Our proportional similarity score is the total ratio–structure deviation

$$d_{\text{PS}}(\mathbf{p}, \mathbf{q}) = \sum_{i<j, \, i,j \in \mathcal{S}} \Big| R_{\mathbf{p}}(i,j) - R_{\mathbf{q}}(i,j) \Big|.$$

C.3. Standardization and comprehensive score

For each metric $m \in \{\text{Euc}, \text{Cos}, \text{KL}, \text{ARE}, \text{PS}\}$ and each LOD ℓ, we apply min–max normalization across all LODs under comparison:

$$\hat{d}_m(\ell) = \frac{d_m(\ell) - \min_{\ell'} d_m(\ell')}{\max_{\ell'} d_m(\ell') - \min_{\ell'} d_m(\ell') + \varepsilon_n}, \quad \varepsilon_n = 10^{-12}.$$

The *comprehensive score* is the equal–weight sum

$$\text{Comp}(\ell) = \sum_m \hat{d}_m(\ell),$$

so that a smaller value denotes a distribution closer to Real in a balanced, multi–view sense.

Data availability

Data and code (Unity project, scripts, and de-identified sample data) are openly available at https://github.com/suyunlei/Perception-LOD.

References

Adhanom, I.B., MacNeilage, P., Folmer, E., 2023. Eye tracking in virtual reality: a broad review of applications and challenges. Virtual Real. 27 (2), 1481–1505.

Albus, P., Vogt, A., Seufert, T., 2021. Signaling in virtual reality influences learning outcome and cognitive load. Comput. Educ. 166, 104154.

Amaxilatis, D., Boldt, D., Choque, J., Diez, L., Gandrille, E., Kartakis, S., Mylonas, G., Vestergaard, L.S., 2018. Advancing experimentation-as-a-service through urban IoT experiments. IEEE Internet Things J. 6 (2), 2563–2572.

Astheimer, P., Pöche, M.-L., 1994. Level-of-detail generation and its application in virtual reality. In: Virtual Reality Software and Technology. World Scientific, pp. 299–309.

Baldassi, S., Megna, N., Burr, D.C., 2006. Visual clutter causes high-magnitude errors. PLoS Biol. 4 (3), e56.

Bearwald, R.R., 1976. The Effects of Stimulus Complexity and Perceptual Deprivation or Stimulus Overload on Attention to Visual Stimuli. Indiana University.

Biljecki, F., Ledoux, H., Stoter, J., 2016. An improved LOD specification for 3D building models. Comput. Environ. Urban Syst. 59, 25–37.

Biljecki, F., Ledoux, H., Stoter, J., Zhao, J., 2014. Formalisation of the level of detail in 3D city modelling. Comput. Environ. Urban Syst. 48, 1–15.

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

Bruner, J.S., Potter, M.C., 1964. Interference in visual recognition. Science 144 (3617), 424–425.

Castelhano, M.S., Mack, M.L., Henderson, J.M., 2009. Viewing task influences eye movement control during active scene perception. J. Vis. 9 (3), 6–6.

Chen, J., 2018. Defining semantic levels of detail for indoor maps. ISPRS Ann. Photogramm. Remote. Sens. Spat. Inf. Sci. 4, 27–34.

Clark, J.H., 1976. Hierarchical geometric models for visible surface algorithms. Commun. ACM 19 (10), 547–554.

Çöltekin, A., Oprean, D., Wallgrün, J.O., Klippel, A., 2019. Where are we now? Revisiting the digital earth through human-centered virtual and augmented reality geovisualization environments.

Colucci, E., Noardo, F., Matrone, F., Spanò, A., Lingua, A., 2018. High-level-of-detail semantic 3D GIS for risk and damage representation of architectural heritage. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 42, 107–114.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3213–3223.

Cronin, D.A., Hall, E.H., Goold, J.E., Hayes, T.R., Henderson, J.M., 2020. Eye movements in real-world scene photographs: General characteristics and effects of viewing task. Front. Psychol. 10, 2915.

Distefano, J.P., Manjunatha, H., Thammineni, C., Dalland, K., Esfahani, E.T., 2025. Gaze-guided contrastive unsupervised representations learning. Neural Comput. Appl. 37 (28), 23381–23393.

Döllner, J., Buchholz, H., 2005. Continuous level-of-detail modeling of buildings in 3D city models. In: Proceedings of the 13th Annual ACM International Workshop on Geographic Information Systems. pp. 173–181.

Drettakis, G., Roussou, M., Reche, A., Tsingos, N., 2007. Design and evaluation of a real-world virtual environment for architecture and urban planning. Presence: Teleoperators Virtual Environ. 16 (3), 318–332.

Dupont, L., Ooms, K., Duchowski, A.T., Antrop, M., Van Eetvelde, V., 2017. Investigating the visual exploration of the rural-urban gradient using eye-tracking. Spat. Cogn. Comput. 17 (1–2), 65–88.

Eckert, M., Habets, E.A., Rummukainen, O.S., 2021. Cognitive load estimation based on pupillometry in virtual reality with uncontrolled scene lighting. In: 2021 13th International Conference on Quality of Multimedia Experience. Qomex, IEEE, pp. 73–76.

Engel, J., Pasewaldt, S., Trapp, M., Döllner, J., 2012. An immersive visualization system for virtual 3d city models. In: 2012 20th International Conference on Geoinformatics. IEEE, pp. 1–7.

Freiwald, J.P., Ariza, O., Janeh, O., Steinicke, F., 2020. Walking by cycling: A novel in-place locomotion user interface for seated virtual reality experiences. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–12.

Fu, X., Qian, Q.K., Liu, G., Zhuang, T., Visscher, H.J., Huang, R., 2023. Overcoming inertia for sustainable urban development: understanding the role of stimuli in shaping residents' participation behaviors in neighborhood regeneration projects in China. Environ. Impact Assess. Rev. 103, 107252.

Gao, X., Wu, K., Pan, Z., 2022. Low-poly mesh generation for building models. In: ACM SIGGRAPH 2022 Conference Proceedings. pp. 1–9.

Garland, M., Heckbert, P.S., 1997. Surface simplification using quadric error metrics. In: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques. pp. 209–216.

Glander, T., Döllner, J., 2009. Abstract representations for interactive visualization of virtual 3D city models. Comput. Environ. Urban Syst. 33 (5), 375–387.

Graham, K., Chow, L., Fai, S., 2019. From BIM to VR: defining a level of detail to guide virtual reality narratives. J. Inf. Technol. Constr. 24, 553–568.

Gröger, G., Plümer, L., 2012. CityGML–Interoperable semantic 3D city models. ISPRS J. Photogramm. Remote Sens. 71, 12–33.

Guo, S., Jang, K.M., Duarte, F., Kang, Y., Ratti, C., 2025. Urban visual uniqueness: A landmark-free framework to quantify city's identity and distinctiveness from everyday scenes. Comput. Environ. Urban Syst. 122, 102351.

Haskins, A.J., Mentch, J., Botch, T.L., Robertson, C.E., 2020. Active vision in immersive, 360 real-world environments. Sci. Rep. 10 (1), 14304.

Heok, T.K., Daman, D., 2004. A review on level of detail. In: Proceedings. International Conference on Computer Graphics, Imaging and Visualization, 2004. CGIV 2004, IEEE, pp. 70–75.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J., 2011. Eye Tracking: A Comprehensive Guide to Methods and Measures. oup Oxford.

Hoppe, H., 1998. Smooth view-dependent level-of-detail control and its application to terrain rendering. In: Proceedings Visualization'98 (Cat. No. 98CB36276). IEEE, pp. 35–42.

Hosp, B.W., Wahl, S., 2023. Zero: A generic open-source extended reality eye-tracking controller interface for scientists. In: Proceedings of the 2023 Symposium on Eye Tracking Research and Applications. pp. 1–4.

Hou, B.J., Abdrabou, Y., Weidner, F., Gellersen, H., 2024. Unveiling variations: A comparative study of vr headsets regarding eye tracking volume, gaze accuracy, and precision. In: 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops. VRW, IEEE, pp. 650–655.

Hu, M.Y., 2008. Semantic based LoD models of 3D house property. In: Proceedings of Commission II, ISPRS Congress Beijing. Citeseer.

Jeddoub, I., Nys, G.-A., Hajji, R., Billen, R., 2023. Digital twins for cities: Analyzing the gap between concepts and current implementations with a specific focus on data integration. Int. J. Appl. Earth Obs. Geoinf. 122, 103440.

Kasowski, J., Beyeler, M., 2023. SimpleXR: An open-source unity toolbox for simplifying vision research using augmented and virtual reality. J. Vis. 23 (9), 5518–5518.

Kim, H.G., Lim, H.-T., Ro, Y.M., 2019. Deep virtual reality image quality assessment with human perception guider for omnidirectional image. IEEE Trans. Circuits Syst. Video Technol. 30 (4), 917–928.

Kim, T., Oh, J., Kim, N., Cho, S., Yun, S.-Y., 2021. Comparing kullback-leibler divergence and mean squared error loss in knowledge distillation. arXiv preprint arXiv:2105.08919.

Klotzsche, F., de Mooij, J., Ohl, S., Gaebler, M., 2025. EDIA: An open-source toolbox for virtual reality-based eye tracking research using unity. J. Vis. 25 (9), 2485–2485.

Konis, K.S., 2013. Leveraging ubiquitous computing as a platform for collecting real-time occupant feedback in buildings. Intell. Build. Int. 5 (3), 150–161.

Kroeck, K.G., Kirs, P.J., Fiedler, A.M., 1989. Cognitive biasing effects in information systems: implications for linking real world information with human judgment. In: Proceedings of the Twenty-Second Annual Hawaii International Conference on System Sciences. Volume III: Decision Support and Knowledge Based Systems Track. vol. 3, IEEE Computer Society, pp. 517–518.

Kuru, K., 2023. Metaomnicity: Toward immersive urban metaverse cyberspaces using smart city digital twins. IEEE Access 11, 43844–43868.

Lamb, M., Brundin, M., Perez Luque, E., Billing, E., 2022. Eye-tracking beyond peripersonal space in virtual reality: validation and best practices. Front. Virtual Real. 3, 864653.

Lee, J.Y., de Jong, N., Donkers, J., Jarodzka, H., van Merriënboer, J.J., 2023. Measuring cognitive load in virtual reality training via pupillometry. IEEE Trans. Learn. Technol. 17, 704–710.

Lei, B., Su, Y., Biljecki, F., 2023. Humans as sensors in urban digital twins. In: International 3D GeoInfo Conference. Springer, pp. 693–706.

Li, J., Jin, Y., Lu, S., Wu, W., Wang, P., 2020. Building environment information and human perceptual feedback collected through a combined virtual reality (VR) and electroencephalogram (EEG) method. Energy Build. 224, 110259.
Li, C., Kang, F., Hou, X., 2020. Novel hierarchical level of detail combinatorial optimization method for large-scale 3D scene. J. Syst. Simul. 29 (9), 2073–2080.
Li, J., Zhang, Z., Jing, F., Gao, J., Ma, J., Shao, G., Noel, S., 2020. An evaluation of urban green space in Shanghai, China, using eye tracking. Urban For. Urban Green. 56, 126903.
Luo, J., Liu, P., Kong, X., Shen, J., Wu, Q., Xu, D., 2025. Urban digital twins for citizen-centric planning: A systematic review of built environment perception and public participation. Int. J. Appl. Earth Obs. Geoinf. 143, 104746.
Lynch, K., 2023. The image of the city (1960). In: Anthologie Zum Städtebau. Band III: Vom Wiederaufbau Nach dem Zweiten Weltkrieg Bis Zur Zeitgenössischen Stadt. Gebr. Mann Verlag, pp. 481–488.
Manley, E., Filomena, G., Mavros, P., 2021. A spatial model of cognitive distance in cities. Int. J. Geogr. Inf. Sci. 35 (11), 2316–2338.
McCabe-Bennett, H., Lachman, R., Girard, T.A., Antony, M.M., 2020. A virtual reality study of the relationships between hoarding, clutter, and claustrophobia. Cyberpsychology Behav. Soc. Netw. 23 (2), 83–89.
Moreno-Arjonilla, J., López-Ruiz, A., Jiménez-Pérez, J.R., Callejas-Aguilera, J.E., Jurado, J.M., 2024. Eye-tracking on virtual reality: a survey. Virtual Real. 28 (1), 38.
Naghibi, M., 2024. Rethinking small vacant lands in urban resilience: Decoding cognitive and emotional responses to cityscapes. Cities 151, 105167.
Nakamura, K., 2021. Experimental analysis of walkability evaluation using virtual reality application. Environ. Plan. B 48 (8), 2481–2496.
Nyström, M., Holmqvist, K., 2010. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. Behav. Res. Methods 42 (1), 188–204.
Orquin, J.L., Lahm, E.S., Stojić, H., 2021. The visual environment and attention in decision making. Psychol. Bull. 147 (6), 597.
Osth, A.F., Dennis, S., 2015. Sources of interference in item and associative recognition memory. Psychol Rev 122 (2), 260.
Pan, S., Zhang, R., Liu, Y., Gong, M., Huang, H., 2025. Building LOD representation for 3D urban scenes. ISPRS J. Photogramm. Remote Sens. 226, 16–32.
Pasalar, C., Hallowell, G., 2020. Toward human-centered smart cities: understanding emerging technologies. Crit. Pr. Archit.: Unexam. 317.
Reddy, M., 1995. A survey of level of detail support in current virtual reality solutions. Virtual Real. 1, 95–98.
Rosenholtz, R., Li, Y., Nakano, L., 2007. Measuring visual clutter. J. Vis. 7 (2), 17–17.
Sari, R.C., Pranesti, A., Solikhatun, I., Nurbaiti, N., Yuniarti, N., 2024. Cognitive overload in immersive virtual reality in education: More presence but less learnt? Educ. Inf. Technol. 29 (10), 12887–12909.
Schmied-Kowarzik, R., Kaschub, L., Keser, T., Rodeck, R., Wende, G., 2024. Does it look real? Visual realism complexity scale for 3D objects in VR. In: International Conference on Human-Computer Interaction. Springer, pp. 73–92.
Schnotz, W., Kürschner, C., 2007. A reconsideration of cognitive load theory. Educ. Psychol. Rev. 19, 469–508.
Schuetz, I., Fiehler, K., 2022. Eye tracking in virtual reality: Vive pro eye spatial accuracy, precision, and calibration reliability. J. Eye Mov. Res. 15 (3), 15.
Sermet, Y., Demir, I., 2019. Flood action VR: a virtual reality framework for disaster awareness and emergency response training. In: ACM SIGGRAPH 2019 Posters. pp. 1–2.
Shakibamanesh, A., Ghorbanian, M., Izadi, S., Riahi, A., Zeifodini, P., 2025. Utilizing virtual reality in the participatory urban policy-making process: A step toward facilitating effective citizen engagement. Int. J. Hum. Cap. Urban Manag. 10 (2).
Shi, Y., Boffi, M., Piga, B.E., Mussone, L., Caruso, G., 2022. Perception of driving simulations: Can the level of detail of virtual scenarios affect the driver's behavior and emotions? IEEE Trans. Veh. Technol. 71 (4), 3429–3442.
Shin, D., 2019. How does immersion work in augmented reality games? A user-centric view of immersion and engagement. Inf. Commun. Soc. 22 (9), 1212–1229.
Sipatchin, A., Wahl, S., Rifai, K., 2021. Eye-tracking for clinical ophthalmology with virtual reality (vr): A case study of the htc vive pro eye's usability. In: Healthcare. vol. 9, MDPI, p. 180, 2.
Soans, R.S., Renken, R.J., Saxena, R., Tandon, R., Cornelissen, F.W., Gandhi, T.K., 2023. A framework for the continuous evaluation of 3D motion perception in virtual reality. IEEE Trans. Biomed. Eng. 70 (10), 2933–2942.
Souchet, A.D., Philippe, S., Lourdeaux, D., Leroy, L., 2022. Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review. Int. J. Hum.–Comput. Interact. 38 (9), 801–824.
Spertus, E., Sahami, M., Buyukkokten, O., 2005. Evaluating similarity measures: a large-scale study in the orkut social network. In: Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining. pp. 678–684.
Stark, P., Appel, T., Olbrich, M.J., Kasneci, E., 2023. Pupil diameter during counting tasks as potential baseline for virtual reality experiments. In: Proceedings of the 2023 Symposium on Eye Tracking Research and Applications. pp. 1–7.
Stein, N., Niehorster, D.C., Watson, T., Steinicke, F., Rifai, K., Wahl, S., Lappe, M., 2021. A comparison of eye tracking latencies among several commercial head-mounted displays. I-Perception 12 (1), 2041669520983338.
Takikawa, T., Litalien, J., Yin, K., Kreis, K., Loop, C., Nowrouzezahrai, D., Jacobson, A., McGuire, M., Fidler, S., 2021. Neural geometric level of detail: Real-time rendering with implicit 3d shapes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11358–11367.
Van Veen, H.A., Distler, H.K., Braun, S.J., Bülthoff, H.H., 1998. Navigating through a virtual city: Using virtual reality technology to study human action and perception. Future Gener. Comput. Syst. 14 (3–4), 231–242.
Verdie, Y., Lafarge, F., Alliez, P., 2015. LOD generation for urban scenes. ACM Trans. Graph. 34 (3), 30.
Wang, Q., Yang, S., Liu, M., Cao, Z., Ma, Q., 2014. An eye-tracking study of website complexity from cognitive load perspective. Decis. Support Syst. 62, 1–10.
Xie, Q., Zhang, L., 2024. Entropy-based guidance and predictive modelling of pedestrians' visual attention in urban environment. In: Building Simulation. vol. 17, Springer, pp. 1659–1674, 10.
Xing, Y., Leng, J., Zhou, H., 2025. Cognitive preferences for architectural renovation strategies in traditional villages combining subjective evaluation and eye tracking. Front. Archit. Res.
Xing, Z., Zhao, R., Guo, W., 2024. From traditional to digital contexts: new characteristics of the public's spatial perception of urban streets in the age of technology. Humanit. Soc. Sci. Commun. 11 (1), 1–16.
Yamamoto, T., Akahoshi, H., Kitazawa, S., 2025. Emergence of human-like attention and distinct head clusters in self-supervised vision transformers: A comparative eye-tracking study. Neural Netw. 107595.
Ye, Y., Shi, Y., Xia, P., Kang, J., Tyagi, O., Mehta, R.K., Du, J., 2022. Cognitive characteristics in firefighter wayfinding tasks: An eye-tracking analysis. Adv. Eng. Inform. 53, 101668.
Yu, C.-P., Lee, H.-Y., Luo, X.-Y., 2018. The effect of virtual reality forest and urban environments on physiological and psychological responses. Urban For. Urban Green. 35, 106–114.
Yu, L., Wang, Z., Chen, F., Li, Y., Wang, W., 2023. Subway passengers' wayfinding behaviors when exposed to signage: An experimental study in virtual reality with eye-tracker. Saf. Sci. 162, 106096.
Zhang, F., Hu, M., Che, W., Lin, H., Fang, C., 2018. Framework for virtual cognitive experiment in virtual geographic environments. ISPRS Int. J. Geo-Inf. 7 (1), 36.
Zhang, J., Johnson, K.A., Malin, J.T., Smith, J.W., 2002. Human-centered information visualization. In: International Workshop on Dynamic Visualizations and Learning, Tubingen, Germany.
Zhang, R.-X., Zhang, L.-M., 2021. Panoramic visual perception and identification of architectural cityscape elements in a virtual-reality environment. Future Gener. Comput. Syst. 118, 107–117.
Zhou, X., Cen, Q., Qiu, H., 2023. Effects of urban waterfront park landscape elements on visual behavior and public preference: Evidence from eye-tracking experiments. Urban For. Urban Green. 82, 127889.
Zhu, Q., Zhao, J., Du, Z., Zhang, Y., 2010. Quantitative analysis of discrete 3D geometrical detail levels based on perceptual metric. Comput. Graph. 34 (1), 55–65.