Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

How does the visual complexity of architectural signage affect wayfinding search efficiency? Nonlinear evidence from a semantic-computational dual-channel experiment

Zonghui Wang

zonghui1128@outlook.com

Peoples' Friendship University of Russia

Huadi Liu

Nanchang Hangkong University

Article

Keywords: visual complexity, visual search, signage, cognitive load, figure–ground competition, semantic–computational dual-channel

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11098365/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Signage is the primary visual interface guiding people through complex public buildings; however, designers lack quantitative evidence on how its visual complexity shapes wayfinding performance, and clutter research largely assumes that simpler methods are always faster. This study examines whether this relationship is linear or nonmonotonic, developing the Architectural Visual Search Toolkit (AVST), a semantic-computational dual-channel framework that integrates multimodal large language model scoring with computational visual analysis to yield decomposable complexity metrics. Using AVST, 242 AI-generated signage stimuli were evaluated, and an online click-search experiment with 150 participants yielded 10,233 valid reaction-time trials analyzed via linear mixed models, analysis of variance, and bootstrap mediation. The results reveal a U-shaped relationship: medium-complexity signage produced the fastest search, whereas both low-complexity and high-complexity signage slowed it, challenging the prevailing linear assumption. Specifically, (1) color organization entropy predicted faster search; (2) mounting type moderated the complexity–efficiency relationship; and (3) the figure–ground index partially mediated the complexity–reaction time pathway. These findings establish moderate complexity as an evidence-based target for wayfinding design, signage standards, and architectural planning policy.

Introduction

Background and problem statement

Signage systems in complex public buildings—transit hubs, hospitals, and shopping centers—serve as critical environmental cues guiding users through unfamiliar layouts. The effectiveness of these methods depends not only on semantic content but also on visual complexity, i.e., the degree to which background clutter, color organization, texture density, and figure–ground relationships facilitate or impede rapid target identification. Poorly managed complexity produces longer search times, higher cognitive loads, and wayfinding errors, with consequences ranging from missed flights to delayed evacuations [1, 2].

Classical visual search theories frame how complexity affects search. Wolfe's guided search theory [3] posits that target detection requires attentional binding of basic features, with efficiency declining as distractors and feature overlap increase. Sweller's cognitive load theory [4] explains that complex displays impose extraneous loads, depleting working memory otherwise available for spatial decisions. Together, these theories predict that higher complexity should yield a slower, less accurate search—a prediction supported in controlled laboratory settings [5, 6].

This prediction, however, rests on an implicit assumption of monotonic linearity that may not hold in architectural settings. Real-world signage spans a continuum from sparse, low-information displays, which may fail to capture attention, to dense, cluttered displays, which may overwhelm processing. The arousal–performance inverted-U relationship reviewed by Nieuwenhuis [7] suggests that arousal

potential—complexity, novelty, and incongruity—bears an inverted-U relationship with perceptual performance and aesthetic preference: moderate stimulation optimizes engagement, whereas both insufficient and excessive stimulation impair performance. Beerendonk et al. [8] provided neural evidence that moderate arousal, indexed by pupillary fluctuations, optimizes perceptual decision-making through a disinhibitory circuit, and Yanagisawa and Honda [9] formalized Berlyne's theory using Bayesian information gain, showing that cognitive curiosity peaks at moderate prediction error—a computational account of why moderate complexity may best facilitate information seeking.

Vauclin et al. [10] reviewed affordance perception—the action possibilities that environmental objects offer observers. In wayfinding, signage provides directional affordances whose salience depends on visual complexity: overly simple signage may fail to signal its function, whereas overly complex signage may obscure it. Liu et al. [11], drawing on Kaplan's preference matrix, further validated complexity as a key information dimension shaping environmental perception, predicting an inverted-U relationship between complexity and preference that parallels the arousal–performance curve. This background raises the core question driving the present study: How does architectural wayfinding visual complexity—decomposed into texture and color dimensions—affect search efficiency, and does this relationship follow a linear or nonmonotonic pattern? The spatial context of sign mounting may further moderate complexity effects [12, 13].

Literature review and theoretical gaps

Extensive research has established that visual clutter—the perceptual consequence of complexity—degrades search performance. Liu et al. [14] developed quantitative clutter metrics for see-through displays, showing that feature-congestion measures predict search performance, and Cheng et al. [15] proposed a multiscale entropy-based evaluation confirming that complexity reliably slows visual search across paradigms. Semizer and Rosenholtz [16] extended this work to video conferencing, and Delmas et al. [5] reported that clutter disrupts both speed and scanning strategies in both experts and novices alike. Critically, Liu et al.'s [14] metrics lack dimensional decomposition—color, texture, and figure–ground collapse into a single score—limiting design utility because isolating the dimension that drives inefficiency is essential for targeted intervention. Mechanistically, clutter acts through two pathways: at the perceptual level, complex backgrounds degrade peripheral discriminability and intensify interitem competition, reducing target–distractor segregation [5, 17]; at the cognitive level, they consume the working memory needed to maintain search templates [4]. Warden et al. [6] identified the overlay-versus-adjacency trade-off in head-mounted displays. Kanaan and Moacdieh [18] reported that clutter effects emerge within the first seconds of website search, and Lu et al. [19] reported that crowded displays are processed through an “average-then-max” mechanism, implicating perceptual summarization in target extraction from complex architectural scenes.

The dimensions along which complexity varies also matter. Xu et al. [20] reported that color and shape combine in a weighted manner to guide attention, with color enjoying stable priority, and Wang et al. [21] confirmed that color guidance operates through systematic feature-similarity mechanisms. This

suggests that color organization may serve as a priority attentional pathway in wayfinding signage, which is distinct from texture search, which depends on background regularity [17]. The spatial layout further moderates complexity costs: Zhou et al. [12] proposed a familiarity-dependent model for indoor landmark selection, showing that spatially prominent landmarks critically guide route communication, and Ma and Yang [13] reported that strategic signage placement in transit-oriented developments reduces wayfinding friction. Yin et al. [22] used image analysis to identify factors influencing signage salience at metro stations. Ghamari and Sharifi [1] mapped 40 years of indoor wayfinding research, identifying signage, spatial knowledge acquisition, and eye tracking as emerging themes. Alinaghi et al. [23] emphasized that spatial knowledge acquisition is moderated by the environmental structure, suggesting that the mounting type may shape how visual complexity is processed. Kim and Bock [24] further showed that landmarks, routes, and survey knowledge develop in parallel, implying that complexity may differentially affect these levels: landmark recognition depends on figure–ground discriminability, whereas route and configurational knowledge depend on spatial layout processing.

Although the dominant paradigm assumes a monotonic complexity–performance relationship, emerging evidence suggests nonlinearity. Yoghourdjian et al. [25] reported that the cognitive load in network visualization exhibits plateau and collapse effects rather than scaling linearly with node count, and Ajani et al. [26] reported that decluttering improves data communication with diminishing returns, implying an optimal complexity level. The arousal–performance inverted-U relationship reviewed by Nieuwenhuis [7] provides the theoretical scaffold: complexity, as an arousal–potential factor, should have an inverted-U relationship with performance, with moderate complexity optimizing arousal and engagement. Beerendonk et al. [8] provided neural evidence that a disinhibitory circuit produces optimal perceptual performance at moderate arousal. From a design perspective, Cross's [27] review of design thinking and Litster et al.'s [28] problem-framing research emphasize that design decisions navigate complex problem spaces, and Lee and Ostwald [29] showed that designers employ divergent thinking in conceptual design. Quantitative tools can support evidence-based design review, bridging intuitive judgment and empirical verification, with implications for design education, where such tools help students develop evidence-based reasoning [30].

Despite this rich literature, four critical gaps remain:

Gap 1 (Dimensional Interaction)

Few studies have examined the three-way interaction of texture, color, and spatial mounting within a single design; most isolate one dimension or treat complexity as unitary, leaving the differential contributions of texture and color—and their moderation by spatial context—untested.

Gap 2 (Operationalization Heterogeneity)

The operationalization of “visual complexity” varies widely [6, 14, 15], from subjective ratings to edge density to subband entropy, hampering cross-study comparison and motivating a multidimensional decomposable approach.

Gap 3 (Semantic Deficiencies of Computational Models)

Traditional computational models (e.g., clutter metrics, saliency maps) lack semantic understanding—they cannot identify or localize specific design elements such as signage text or directional arrows, limiting their design-evaluation utility.

Gap 4 (Integrated evaluation paradigm)

No existing paradigm simultaneously integrates behavioral performance, computational visual analysis, and semantic evaluation within a low-barrier, reproducible framework essential for translating research into practical design tools.

Research objectives and hypotheses

To address these gaps, this study has three objectives: (a) to decompose architectural visual complexity into texture and color dimensions and examine their differential effects on search efficiency; (b) to test the moderating role of spatial mounting type and the mediating role of figure–ground competition; and (c) to validate a semantic–computational dual-channel framework that integrates large language model semantic scoring with computational visual analysis. On the basis of this review, we propose the following hypotheses and research question:

H1a (Color-Guidance Pathway)

Color organization entropy (COE) significantly predicts reaction time, with a higher COE corresponding to faster search, reflecting color's priority in attentional guidance [20, 21].

H1b (Texture-Segregation Pathway, Revised)

Owing to accuracy ceiling effects (> 99.5%), texture entropy (TEE) cannot be tested directly against accuracy; the missclick distance is employed as an alternative cognitive-load indicator to test whether TEE predicts search deviation.

H2 (Figure-Ground Mediation)

The Figure-Ground Index (FGI) mediates the relationship between complexity and reaction time.

H3 (Spatial Mounting Moderation)

Mounting type (column- vs. side-mounted) significantly moderates the complexity–efficiency relationship, with side-mounted signage increasing complexity costs.

H4 (Dual-channel convergence)

Semantic scores from the LLM channel and computational metrics from the visual analysis channel converge significantly, but the convergence strength varies by dimension.

RQ1 (Nonlinear Exploration)

Does the complexity–efficiency relationship exhibit a nonmonotonic (U-shaped or inverted-U-shaped) pattern?

Results

Development of the AVST tool

The Architectural Visual Search Toolkit (AVST) (Fig. 1) is a modular, cross-channel quantitative tool bridging semantic evaluation and computational visual analysis. AVST decouples functionality into five independent components: a three-model ensemble computation engine that coordinates the grounding DINO, U2-Net (ISNet), and ResNet50 inference pipelines; a DAG dependency-based metric computation module (18 metrics total); image preprocessing and visualization utilities; a multimodal LLM API interface for two-stage semantic analysis; and a multilingual detection description normalizer. Components communicate through standardized data structures for bounding box coordinates, saliency heatmaps, and scoring results, ensuring flexibility for module replacement and extension.

AVST configures independent fault-tolerant degradation paths for each model: when grounding DINO fails to load, center fallback ROIs are used; when ISNet (U2-Net) fails, the system switches to ResNet50 single-model saliency; when ResNet50 fails, the system reversely switches to ISNet; and when both saliency models fail, analysis terminates with an explicit error. All inferences are executed under torch.no_grad() and model.eval() modes, ensuring reproducibility. Grounding DINO and ResNet50 support FP16 half-precision acceleration, whereas ISNet uses CPU execution due to ONNX runtime CUDA provider constraints.

Each analysis produces 18 quantitative metrics—covering dimensional metrics (TEE, COE), competition metrics (BCE, Target_Score, FGI), fusion and alignment metrics (CA, VCI), semantic five-dimensional scores, and provenance metadata—along with six visualization comparison charts (Fig. 5). Batch mode supports folder processing with consolidated Excel output of all summary metrics. Definitions and formulas for all abbreviations and metrics are provided in Supplementary Note 1 and Supplementary Table S1 online.

Manipulation check and convergence validity

One-way ANOVA confirmed that four of the five AVST metrics significantly distinguished complexity conditions (Fig. 2). COE showed the strongest discriminative power ($F = 40.38$, $p < .001$), increasing monotonically from low (0.969) to high (0.986). The VCI also increased significantly and monotonically

(F = 13.50, p < .001). The FGI was significant (F = 5.85, p = .003), with a plateau between medium and high. TEE was significant but with negligible absolute differences (F = 11.51, 0.983–0.990). BCE did not distinguish conditions (F = 0.07, p = .937) and was associated with a control variable.

Pearson correlations between LLM semantic scores and computational metrics (Table 1) revealed a selective convergence pattern: density/VCI r = .895 (p < .001), color/COE r = .390 (p < .001), figure-ground/FGI r = .147 (p = .023), and texture/TEE r = − .066 (p = .307), partially supporting H4.

Table 1
Dual-channel convergence: Pearson correlations between LLM semantic scores and computational metrics.

LLM score	Computational metric	r	p	Convergence
llm_density	VCI	.895	< .001	Very strong
llm_color	COE	.390	< .001	Moderate
llm_fg	FGI	.147	.023	Weak
llm_texture	TEE	−.066	.307	None

Dimensional differences: H1a and H1b

In support of H1a, the COE significantly predicted RT (r = − .236, p = .0002), and the LMM confirmed that the COE was a significant predictor of log(RT) (β = −.15, SE = .04, p < .01). H1b could not be tested as originally conceived because of accuracy ceiling effects (99.0%–99.8%): TEE was not significantly correlated with accuracy (r = − .051, p = .425), and accuracy ANOVA did not reach significance (F = 2.31, p = .102). Miss-click distance analysis (Section 3.6) serves as the alternative test.

Table 2
Reaction time and accuracy by complexity condition (Panel A) and by complexity × mounting type (Panel B). N = 10,233 valid reaction time trials.

Panel A: Reaction time and accuracy by complexity condition					
Complexity	**Mean RT (ms)**	**Median RT (ms)**	**SD**	**Trials**	**Accuracy**
Low	870.5	725.0	631.4	3,407	99.4%
Medium	789.1	664.0	585.0	3,385	99.5%
High	821.5	685.0	661.8	3,441	99.5%
Panel B: Reaction time by complexity × mounting type					
Complexity	Mounting type	Mean RT (ms)	SD	Trials	Accuracy
Low	Column	799.1	508.6	1,696	99.8%
Low	Side	941.4	726.3	1,711	99.0%
Medium	Column	749.8	516.9	1,707	99.7%
Medium	Side	829.0	644.7	1,678	99.4%
High	Column	803.4	684.0	1,730	99.6%
High	Side	839.9	638.3	1,711	99.3%

Spatial Mounting Moderation: H3

Mounting type produced a robust main effect: side-mounted was slower than column-mounted (870.3 vs. 784.2 ms, + 86.1 ms). LMM confirmed the significance ($\beta = .08$, $p < .001$), supporting H3. The complexity × mounting type interaction was significant (Table 2): both exhibited U-shaped patterns, with the side-mounted disadvantage being greatest at low complexity (941.4 vs. 799.1 ms), narrowing to approximately 79.2 and 36.5 ms at medium and high complexity, respectively.

Figure-ground competition mediation: H2

FGI was significantly correlated with RT ($r = -.206$, $p = .001$), confirming that clearer figure–ground contrast facilitates faster target identification. Bootstrap mediation analysis (5,000 resamples) was used to test whether the total effect of complexity was partially transmitted through FGI. The indirect effect was significant (ab = .012, 95% CI [.003,.024]), with a mediation proportion of approximately 18%, providing partial support for H2. LMM random effects revealed between-participant variance $\sigma^2 = .034$, between-image variance $\sigma^2 = .018$, ICC = .346; the gap between conditional $R^2 = .342$ and marginal $R^2 = .087$ indicates that fixed effects explained only 8.7% of the variance, with the remainder attributable to individual baseline differences.

Miss-click distance analysis

As accuracy exceeded 99.5% across all conditions (ceiling effect), the missclick distance—the Euclidean distance between the click coordinate and the signage ROI center (pixel percentage)—was analyzed as an alternative cognitive load indicator (Fig. 3). Column-mounted signage exhibited a clear V-shaped distribution: larger deviations at low (M = 24.33) and high (M = 45.85) complexity, with the smallest deviation at medium (M = 16.30), echoing the U-shaped RT results and indicating optimal search precision at medium complexity.

Side-mounted signage showed a distinctly different pattern: misses were more frequent, reflecting attentional dispersion from the peripheral presentation, but the deviation distance decreased monotonically with complexity, from 34.71 at low to 13.26 at high complexity.

Psychometric validation supported the missclick distance as an effective indicator: a significant positive correlation with log(RT) ($r = .187$, $p < .001$), an effect size ($\eta^2 = .045$) comparable to RT, and acceptable internal consistency (Cronbach's $\alpha = .71$). Side-mounted misses reflect spatial position uncertainty, causing near-distance misses, whereas column-mounted signage produces far-distance deviations at high complexity because of attentional capture by competitive distractors.

Nonlinear exploration: RQ1

Complexity and RT exhibited a U-shaped relationship (Table 2): low complexity was slowest (870.5 ms), medium complexity was fastest (789.1 ms), and high complexity was intermediate (821.5 ms). The quadratic fit outperformed the linear fit ($\Delta R^2 = .012$, $F = 4.52$, $p = .034$), supporting RQ1.

Spatial distribution of valid click coordinates

All valid click coordinates were normalized to a unified 10×10 ROI heatmap grid, eliminating the influence of ROI size and position differences across images (Fig. 4). Column-mounted signage showed concentrated upper-central clustering: peak (X = 5, Y = 2) = 590 clicks, with high-density areas at Y = 1–3, X = 3–6. Side-mounted signage showed dispersed vertical extension: peak (X = 5, Y = 6) = 220 clicks (approximately 37% of column-mounted), with a vertical span of Y = 2–9. These distributions provide spatial behavioral evidence for the mounting type main effect (side-mounted 86.1 ms slower): column mounting facilitates concentrated localization through a central visual field advantage, whereas side mounting causes attentional dispersion due to peripheral positioning.

Computational—behavioral correlation analysis

Pearson correlations between AVST metrics and behavioral measures (Table 3): COE, FGI, and VCI were all significantly negatively correlated with RT ($r = -.21$ to $-.26$, $p < .01$), whereas TEE showed no significant correlation. Compared with VCI, FC distinguished complexity effects less effectively ($F = 6.82$, $\eta^2 = .054$ vs. $F = 13.50$, $\eta^2 = .101$).

Table 3
Pearson correlations between AVST computational metrics and behavioral measures (image level, N = 242).

Metric	vs. Mean RT (r)	p	vs. Mean log_RT (r)	p	vs. Accuracy (r)	p
TEE	.030	.639	.021	.743	−.051	.425
COE	−.236	.0002	−.260	< .001	.016	.810
FGI	−.206	.001	−.254	< .001	.111	.084
VCI	−.207	.001	−.214	.001	.103	.108

Session effects and robustness checks

First-half versus second-half comparisons revealed significant within-session learning effects (first 861.3 ms vs. second 793.0 ms, t = 5.509, p < .001). However, the U-shaped complexity effect was robust across both halves, with ANOVA (complexity: F(2,10227) = 14.62; mounting type: F(1,10227) = 48.58; interaction: F(2,10227) = 6.20, all p < .01), confirming that the core findings were not confounded.

Three robustness analyses confirmed result stability: (a) LMM fixed effects remained significant after removing RT outliers (± 3 SD); (b) metric gradients were consistent after substituting saliency models; (c) Type III ANOVA cross-validated significance patterns.

Discussion

Nonmonotonic complexity effects and dimensional mechanisms

The U-shaped relationship between visual complexity and search reaction time—low complexity being the slowest, medium fastest, and high intermediate—directly challenges the linear assumption that “higher complexity leads to slower search” has dominated since Wolfe's [3] guided search theory and is implicit in clutter metrics [6, 14, 18]. Nieuwenhuis's [7] arousal—performance inverted—U framework offers the most parsimonious explanation: low complexity fails to generate sufficient arousal, medium complexity produces optimal arousal, and high complexity introduces distractor competition. This principle receives convergent support from Beerendonk et al.'s [8] neural evidence, Vauclin et al.'s [10] affordance review, Liu et al.'s [11] environmental preference work, Yanagisawa and Honda's [9] Bayesian information gain framework, and Yoghourdjian et al.'s [25] nonmonotonic performance patterns in network visualization. For design practice, this yields counterintuitive but actionable insight: the heuristic that signage should be “as simple as possible” lacks empirical support, as overly sparse signage may impair search efficiency compared with overly cluttered signage.

At the dimensional level, the significant negative correlation between the COE and RT (r = − .236, p = .0002) confirms that well-organized color distributions accelerate search, which is consistent with the

color priority hypothesis [20, 21]. TEE failed to predict any behavioral measure, partly because of accuracy ceiling effects (> 99.5%) and negligible absolute TEE differences across complexity levels (0.983–0.990); V-channel entropy may be insensitive to higher-order structural regularity [17]. Signage search is essentially a landmark recognition task that relies more on color organization (COE) and figure–ground contrast (FGI) than on luminance textures [24]. The FGI mediation effect (approximately 18% of the complexity-RT pathway, with the remaining 82% reflecting cognitive pathways such as working memory consumption and attentional competition) indicates that figure-ground competition is a meaningful but partial mediator, extending “display clutter” to the architectural domain and shifting evaluation focus from the signage itself to the signage-background relationship [27, 29]. The dual-channel convergence analysis reveals dimensional selectivity: density/VCI near-perfect convergence (r = .895), color/COE moderate convergence (r = .390), figure-ground/FGI weak convergence (r = .147), and texture/TEE complete divergence (r = − .066). The texture divergence indicates that the LLM evaluates texture as a semantic construct, whereas TEE measures low-level pixel entropy, establishing an important boundary condition: convergence strength should be empirically established before interpreting channel agreement as construct validation.

The theoretical implications of this nonmonotonic pattern extend beyond wayfinding to the broader literature on visual complexity and performance. The convergence of arousal theory [7], affordance theory [10], and environmental preference matrices [11] on a nonmonotonic complexity–performance relationship suggests that the inverted-U principle may constitute a domain-general law of visual information processing, not limited to the aesthetic and curiosity domains in which it was originally formulated. The present study provides the first behavioral evidence for this principle in architectural wayfinding, bridging the gap between laboratory-based visual search research and real-world environmental design. The dimensional decomposition further reveals that the nonmonotonic pattern is driven primarily by color organization (COE) and figure–ground contrast (FGI) rather than by luminance texture (TEE), suggesting that perceptual arousal is elicited by chromatic and structural dimensions rather than by raw information quantity. This finding refines the arousal–performance framework by identifying which stimulus dimensions contribute to the arousal signal, moving beyond the undifferentiated notion of complexity as a unitary construct.

Mounting type moderation and design parameters

The 86.1 ms main effect of mounting type and its interaction with complexity reveal that spatial mounting is a substantive moderator of complexity effects, not merely a contextual factor. Side-mounted signage was consistently slower than column-mounted signage, which is consistent with Zhou et al.'s [12] finding on spatially prominent landmarks and Ma and Yang's [13] work on strategic signage placement. The interaction pattern—side-mounted disadvantage being largest at low complexity (142.3 ms) but narrowing to 36.5 ms at high complexity—suggests a compensatory interaction: additional visual information can partially offset positional disadvantage, so when side-mounting is architecturally necessary, designers should increase signage complexity to the medium range rather than minimize it. This finding is consistent with guided search theory's ability to predict the interaction between bottom-up

signals and top-down guidance [3], which was extended here from laboratory settings to real architectural environments. Column-mounted signage has a greater foveal projection probability, whereas side-mounted signage requires additional attentional shifts [23].

Miss-click distance analysis revealed qualitative differences between mounting types: side-mounted signage produced more misses with shorter deviation distances (“near-distance misses”), whereas column-mounted signage produced fewer but more severe deviations at high complexity (“far-distance misses”), reflecting different attentional mechanisms—spatial uncertainty for side-mounted versus attentional capture by distractors for column-mounted signage. The medium complexity corresponds to the optimal search efficiency, mapping to COE ≈ 0.98 and VCI ≈ 0.69, which can serve as reference baselines for design review. Elsayed et al.'s [31] postoccupancy evaluation review emphasizes multidimensional building performance assessment; the VCI, COE, and FGI metrics provided by the AVST toolkit complement traditional satisfaction surveys as objective wayfinding performance indicators. Embedding the AVST computational pipeline into BIM workflows can automate visual complexity assessment during the design phase, achieving closed-loop optimization from design to evaluation.

The design parameter optimization insights from this study offer specific, actionable guidance for practitioners. The identification of COE ≈ 0.98 and VCI ≈ 0.69 as optimal complexity targets provides quantitative benchmarks that can be integrated into design review checklists and automated evaluation tools. When designers modify signage proposals, they can use AVST to measure whether the revised design falls within the target range, creating an iterative optimization loop. This data-driven approach contrasts with the prevalent practice of relying on intuitive judgment or adhering to generic guidelines that specify font size and contrast ratios but do not address overall visual complexity. The compensatory strategy identified in the mounting type analysis, whereby increasing complexity to the medium range offsets the disadvantage of side-mounting, provides a concrete design heuristic for situations in which architectural constraints prevent optimal signage placement.

Methodological contributions and practical implications

The accuracy ceiling effect (> 99.5%) is a methodological reality that is addressed transparently rather than treated as a deficiency. Willoughby et al. [32] noted that ceiling effects are prevalent in inhibitory control and visual search tasks, recommending the integration of item-level accuracy, reaction time, and spatial precision. This study adopted the miss-click distance as an alternative cognitive load indicator: the V-shaped pattern for column-mounted images (largest deviation at high complexity: 45.85) and the decreasing pattern for side-mounted images (13.26–34.71) revealed qualitative differences that were invisible to accuracy analysis [23]. This paradigm can transfer to other human factor contexts where performance approaches the ceiling but the cognitive load varies—such as medical display monitoring and aviation cockpit evaluation—ideally with multidimensional triangulation [33].

The semantic-computational dual-channel framework contributes to visual complexity research in three ways. First, by decomposing complexity into independently measurable dimensions (TEE, COE, FGI, VCI), each rooted in specific perceptual mechanisms, it addresses operationalization heterogeneity [6, 14, 15].

Second, the LLM semantic channel bridges pixel-level analysis and semantic understanding, with the ROI isolation protocol ensuring channel independence. Third, the AVST automates the computational pipeline, lowering the entry barrier for nonprogramming researchers, and will be made publicly available upon publication. The framework transfers beyond architectural wayfinding to any domain involving target search in complex backgrounds—human–computer interface evaluation, data visualization, and safety–critical display design—with dimensional weights and LLM prompts customizable to domain-relevant features.

In design education, the AVST toolkit can serve as a teaching tool embedded in design studio courses, helping students develop intuitive understanding of the complexity–efficiency relationship by comparing proposals against the moderate complexity target range (COE ≈ 0.98, VCI ≈ 0.69) [30]. At the urban planning policy level, Carmona's [34] place-focused urban design coding framework positions signage quantification standards as governance tools; building codes and accessibility standards such as ISO 21542 [35] can incorporate visual complexity benchmarks alongside existing requirements. In safety-critical environments, the U-shaped relationship implies that overly simplified emergency evacuation signage may delay evacuation by failing to generate sufficient perceptual arousal; the AVST-derived target range can provide evidence-based support for safety signage standards, particularly in transit hubs and medical facilities where wayfinding errors are costly.

The practical implications extend particularly to safety-critical environments. In emergency evacuation scenarios, the U-shaped relationship implies that overly simplified exit signage may paradoxically delay egress by failing to generate sufficient perceptual arousal to capture attention under stress. Conversely, excessively complex signage in high-pressure situations may overwhelm already taxed cognitive resources, a concern amplified by the finding that side-mounted signage, which is common for emergency exits, amplified complexity costs by up to 142.3 ms. These results suggest that emergency signage standards should specify not only content and placement but also visual complexity parameters, with the AVST-derived target range offering a preliminary evidence-based benchmark. In healthcare facilities, where patients navigating under anxiety or cognitive impairment face heightened wayfinding challenges, the dimensional decomposition approach allows designers to identify whether search inefficiency stems from color disorganization, poor figure–ground contrast, or texture overload, enabling targeted rather than wholesale redesign.

The framework's transferability extends beyond architectural wayfinding to any domain involving target search in visually complex backgrounds. In human–computer interface evaluation, the dimensional decomposition approach can be used to determine whether interface clutter stems from color disorganization or figure–ground competition, guiding targeted simplification. In data visualization, the finding that moderate complexity optimizes search efficiency aligns with Ajani et al.'s [26] decluttering guidelines and suggests that visualization design should target a moderate complexity range rather than pursue maximal simplification. In aviation cockpit display design, where clutter effects have been extensively documented [6], the missclick distance paradigm offers a sensitive performance metric under the accuracy ceiling conditions characteristic of trained operator populations.

Limitations and future directions

Several limitations should be acknowledged. First, this study used static 2D images as stimuli, lacking depth cues, peripheral vision, and embodied spatial information present in real architectural environments; generalizing from 2D search RT to real wayfinding requires staged validation through 360° panoramas, VR environments, and real wayfinding studies with behavioral tracking. The accuracy ceiling effect (> 99.5%) also limits the ability to test the texture pathway as originally conceived; future research should employ more difficult search tasks and incorporate subjective scales (e.g., NASA-TLX) and physiological indicators for multidimensional cognitive load measurement [33].

Second, the semantic-computational divergence in the texture dimension ($r = -.066$) indicates that the current TEE metric based on V-channel Shannon entropy may not capture the structural regularity driving human texture perception; the computational channel should explore alternative metrics such as Gabor filters and local binary patterns. LLM-based semantic scoring introduces reproducibility concerns—model updates may cause score drift—necessitating a benchmark image library with human expert ratings to calibrate LLM scores and monitor temporal stability.

Third, the sample age range (18–52 years, M = 28.3) underrepresents older adults, whose declining visual processing speed and weakened inhibitory control may steepen the complexity-efficiency slope. LMM random effects analysis (conditional $R^2 = .342$ vs. marginal $R^2 = .087$) indicated significant individual differences in baseline RT but lacked covariates such as working memory capacity and spatial cognition ability. The 3 × 2 design was reduced because of insufficient end-position signage instances; future research should complete a 3 × 3 design, recruit age-stratified samples, and replicate in different cultural contexts. Additionally, this study did not employ eye-tracking methods, which could provide direct evidence of attentional allocation patterns; future research should integrate eye tracking to verify whether the U-shaped RT pattern corresponds to different saccadic path strategies and to distinguish attentional bias from spatial misperception reflected in miss-click distance.

Despite these limitations, this study achieved integrated progress: the nonmonotonic complexity—efficiency relationship is supported by convergent evidence from arousal theory, affordance theory, and environmental preference matrices, cross-validated by the Bayesian information gain framework [9] and neural evidence [8]; the dual-channel framework was expert validated (ICC = .82); and the moderate complexity target range (COE ≈ 0.98, VCI ≈ 0.69) provides parametric guidance for evidence-based wayfinding design.

Conclusions

This study developed the Architectural Visual Search Toolkit (AVST), which is a semantic—computational dual-channel framework that integrates multimodal large-language model semantic scoring with computational visual analysis, and systematically revealed the nonmonotonic impact of architectural

wayfinding visual complexity on search efficiency on the basis of 242 controlled generated stimuli and 10,233 valid behavioral trials.

First, a nonmonotonic U-shaped relationship exists between visual complexity and search efficiency, directly challenging the long-standing linear assumption that “higher complexity, slower search.” Medium-complexity signage was fastest, with low-complexity signage actually slower than high-complexity signage. This finding bridges arousal theory with wayfinding design practice, demonstrating that “information underload” and “information overload” impair search efficiency comparably. The medium complexity corresponds to COE ≈ 0.98 and VCI ≈ 0.69, constituting an actionable design target range.

Second, the spatial mounting type is a substantive moderator rather than a contextual factor. Column-mounted signage consistently outperformed side-mounted signage (86.1 ms faster), and the side-mounted disadvantage was amplified to 142.3 ms under low-complexity conditions. When architectural conditions force signage away from the central visual field, designers can compensate by adjusting complexity to the medium range—a compensatory strategy providing direct guidance for wayfinding optimization under mounting constraints.

Third, the AVST framework outperforms traditional aggregate metrics in terms of dimensional decomposition sensitivity. The VCI's effect size in distinguishing complexity conditions (η^2 =.101) was nearly double that of the feature congestion metric FC (η^2 =.054), and the dimensional selective convergence pattern of the semantic-computational dual-channel (density r = .895, color r = .390, texture r = − .066) reveals that measurement divergence between semantic and computational levels is itself theoretically informative. The miss-click distance, as an alternative cognitive load indicator under accuracy ceiling effects, successfully revealed qualitative differences that were invisible to binary accuracy.

These findings guide architectural wayfinding design practice: moderate rather than minimal complexity should be the design target; column-mounted mounting is preferred, and when side-mounting is unavoidable, information density should be increased to the medium level for compensation; the AVST's VCI, COE, and FGI metrics provide quantitative evidence for evidence-based decision-making. Future research should extend this framework to VR, eye-tracking, and cross-cultural scenarios.

Methods

Experimental design framework

This study employs a 3 (complexity: low, medium, high; within-subjects) × 2 (mounting type: column-mounted, side-mounted; stratified random sampling) mixed factorial design, forming a closed-loop workflow of “generation—evaluation—experiment—comparison” (Supplementary Fig. S1). Complexity was manipulated within subjects to control for individual baseline differences, whereas mounting type was balanced through stratified random sampling. The six-stage workflow comprised (1) generating

1,140 candidate architectural interior images via ComfyUI workflows; (2) three experts independently screening all images, retaining 81 high-quality base images; (3) using ComfyUI inpainting to produce three complexity variants per base image, yielding 242 final stimuli; (4) applying the Architectural Visual Search Toolkit (AVST), developed for this study, to perform automated semantic–computational dual-channel analysis, producing metrics including TEE, COE, FGI, and VCI; (5) recruiting participants for an online click-search behavioral experiment recording reaction time, click coordinates, and accuracy; and (6) conducting statistical comparisons between AVST metrics and behavioral results via LMM, ANOVA, and mediation analysis.

Stimulus construction

Three experts independently screened 1,140 candidate architectural interior photographs, selecting 81 base images encompassing both column-mounted and side-mounted signage nodes across transit hubs, commercial complexes, and institutional buildings (Fleiss' κ = .78, substantial agreement). The selected images varied in architectural style, lighting, and spatial configuration to ensure ecological validity. For each base image, three complexity variants were generated via ComfyUI Inpainting (Stable Diffusion XL; DPM + + 2 M Karras, 20 steps, CFG 7–8). Inpainting masks defining signage area shape, size, and position remained constant across complexity conditions; complexity was manipulated solely through text prompts specifying minimalist, standard, or dense signage. The nonmask region noise strength was set to 0 for pixel-level background preservation (SSIM > 0.98), with mask region denoising at 0.75. This generative AI approach to controlled stimuli is methodologically supported by recent work [36].

All signage text was standardized and manually proofread; background characters were excluded so that background competition arose solely from architectural elements. The final set comprised 242 images (81 bases × 3, minus 1 excluded for artifacts). To validate LLM semantic scoring, three experts rated 50 randomly sampled images on the same five dimensions (1–5 scale), yielding high expert–LLM agreement (ICC = .82, 95% CI [.74,.88]). Image naming followed the {mounting_type}_{base_image_ID}_{complexity}.png convention for programmatic batch processing.

Semantic-computational dual-channel evaluation framework

To address the semantic deficiency of traditional computational models (Gap 3) and the lack of an integrated evaluation paradigm (Gap 4), this study developed the Architectural Visual Search Toolkit (AVST), which is a semantic-computational dual-channel framework that combines a large language model (LLM) semantic understanding with computational visual analysis. The two channels operate independently and converge only at the validation stage, employing a strict ROI isolation protocol to prevent circular reasoning. Channel A (semantic evaluation) uses a multimodal LLM API for two

functions: semantic localization (generating bounding boxes for signage elements) and multidimensional scoring across five dimensions—color organization, texture complexity, information density, figure–ground contrast, and readability (1–5 scale each)—receiving no information from the computational channel. Channel B (computational analysis) integrates three pretrained models: grounding DINO, an open-set object detection model with a Swin-T backbone (confidence threshold of 0.35) used for text-guided signage localization [37]; U2-Net for salient object detection generating saliency maps; and ResNet50 for deep feature extraction and cross-validation.

AVST employs a layered ROI isolation strategy: depending on metric nature, different metrics use ROIs from different sources, preventing the same detection box from simultaneously serving localization validation and metric evaluation. The framework adopts a data-driven evaluation principle rather than imposing absolute thresholds: metrics are validated on the basis of ANOVA significance in distinguishing complexity conditions, effect sizes (η^2), and convergence validity with semantic scores. Metrics failing manipulation checks are demoted to control variables rather than discarded. Figure 5 shows an example of AVST's single-image analysis across TEE, COE, FGI, and VCI metrics.

Behavioral experiment

The participants completed an online click-search task and were instructed to find and click the remaining signage in each displayed image as quickly and accurately as possible. The JavaScript/HTML5 platform recorded reaction time (RT; via performance.now(), precision < 16 ms), click coordinates, and accuracy (clicking within the ground-truth bounding box). Trial timing comprised a 500 ms fixation cross, stimulus presentation, click or 10-second timeout, and a 500–1500 ms jittered intertrial interval; practice blocks (5 trials) were excluded from analysis. The 246 stimuli were divided into groups via stratified random sampling; each participant received approximately 50 images (task duration < 5 minutes), with a rest every 25 trials.

Data cleaning employed a three-tier filter: trial-level removal of RT < 200 ms or > 5,000 ms; participant-level removal of completion time < 5 minutes; and device-level retention of only the first session for multiple sessions. Ultimately, 150 valid independent devices were retained (ages 18–25: 79.3%; male 58.7%, female 38.7%; mobile 60.7%, desktop 30.0%, tablet 1.3%, legacy 8.0%), exceeding the a priori power target of N ≥ 128. These devices contributed 10,454 valid trials (10,233 valid RT trials). Between-session repetition effects were negligible (RT: 827.8 vs. 825.2 ms; accuracy: 99.66% vs. 99.58% for first vs. subsequent sessions). Owing to accuracy ceiling effects (overall 99.64%, > 99.5% across all conditions), the missclick distance—the Euclidean distance between the click coordinate and the signage ROI center—was computed for all the error trials as a more sensitive spatial precision indicator, following Willoughby et al. [32]. The study protocol was reviewed and approved by the Institutional Review Board of the School of Art and Design, Nanchang Hangkong University. All methods were carried out in accordance with the relevant guidelines and regulations. All participants provided electronic

informed consent prior to participation, and informed consent was obtained from all the experts involved in the study.

Statistical analysis strategy

The analysis proceeded through five hierarchical levels. First, manipulation checks used one-way ANOVA to test whether each AVST metric (TEE, COE, FGI, VCI, BCE) significantly distinguished the three complexity levels, with effect sizes reported as η^2; metrics failing the check ($p > .05$) were demoted to control variables. Second, the core analysis employed a linear mixed model (LMM) with log-transformed RT as the dependent variable—log(RT) ~ Complexity × MountType + (1|ImageID) + (1|SubjectID), controlling for Age and DeviceType—preferred over repeated-measures ANOVA for handling unbalanced nested data and crossed random effects. Log transformation was used to address the right-skewed RT distribution. Third, bootstrap mediation testing (5,000 resamples, 95% bias-corrected CI) was used to examine whether FGI mediates the complexity → RT pathway, with additional pathways tested (COE → RT; TEE → miss-click distance).

Fourth, convergence validity was assessed through Pearson correlation matrices among AVST metrics, LLM semantic scores, and behavioral measures (RT, log_RT, accuracy, miss-click distance). Fifth, exploratory nonlinear testing fitted linear and quadratic models to the complexity → RT relationship, assessing ΔR^2. Three robustness checks were conducted: retesting the LMM after removing RT outliers (± 3 SD), substituting saliency models (ResNet50 vs. U2-Net) to verify metric gradient consistency, and cross-validating with Type III ANOVA. Psychometric validation of the missclick distance included concurrent validity, sensitivity, and internal consistency.

For the LMM, σ^2, the ICC, the conditional R^2, and the marginal R^2 are reported. The random effects structure was determined by likelihood ratio testing: crossed random intercepts significantly improved fit (LRT $p < .001$), whereas complex random slopes were unsupported. Bonferroni correction was applied for manipulation checks, and FDR correction was used for convergence validity. Effect size reporting: ANOVA reports η^2, LMM reports R^2 and σ^2/ICC, and correlations report Pearson r and Spearman ρ. Analyses were conducted in Python and R, with G*Power for power analysis. Significance was set at $\alpha = .05$ (two-tailed).

Ethics approval

for this study was granted by the Institutional Review Board of the School of Art and Design, Nanchang Hangkong University, and all participants provided electronic informed consent prior to participation.

Use of artificial intelligence: during the preparation of this work the authors used generative AI tools in order to improve the language and readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Data availability

The AVST toolkit developed in this study will be made publicly available upon publication. Anonymized behavioral data are available from the corresponding author upon reasonable request; the data are not publicly available in order to protect participant privacy, in accordance with the consent protocol approved by the Institutional Review Board of the School of Art and Design, Nanchang Hangkong University.

Declarations

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions

Z. W. Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. H.L. Methodology, Validation, Investigation, Data curation, Writing – review & editing. All authors reviewed and edited the manuscript.

Additional Information

The authors declare no competing financial or non-financial interests.

References

1. Ghamari, H. & Sharifi, A. Mapping the evolutions and trends of literature on wayfinding in indoor environments. *Eur. J. Investig Health Psychol. Educ.* **11**, 585–606 (2021).
2. Li, C. et al. A systematic review of factors influencing signage salience in indoor environments. *Sustainability* **15**, 13658 (2023).
3. Wolfe, J. M. Guided Search 6.0: An updated model of visual search. *Psychon Bull. Rev.* **28**, 1060–1092 (2021).
4. Sweller, J. The development of cognitive load theory: Replication crises and incorporation of other theories can lead to theory expansion. *Educ. Psychol. Rev.* **35**, 95 (2023).
5. Delmas, M., Caroux, L. & Lemercier, C. Searching in clutter: Visual behavior and performance of expert action video game players. *Appl. Ergon.* **99**, 103628 (2022).
6. Warden, A. C., Wickens, C. D., Rehberg, D., Clegg, B. A. & Ortega, F. R. Clutter costs in head-mounted displays: A study examining trade-offs between overlay and adjacent presentation of information. *Cogn. Res. Princ Implic*. **10**, 47 (2025).

7. Nieuwenhuis, S. Arousal and performance: Revisiting the famous inverted-U-shaped curve. *Trends Cogn. Sci.* **28**, 394–396 (2024).
8. Beerendonk, L. et al. A disinhibitory circuit mechanism explains a general principle of peak performance during mid-level arousal. *Proc. Natl. Acad. Sci. U.S.A.* 121, e2312898121. (2024).
9. Yanagisawa, H. & Honda, S. Modeling the arousal potential of epistemic emotions using Bayesian information gain: A framework for inquiry cycles driven by free energy fluctuations. *Front. Psychol.* **16**, 1438080 (2025).
10. Zhou, Z., Weibel, R. & Huang, H. Familiarity-dependent computational modelling of indoor landmark selection for route communication: A ranking approach. *Int. J. Geogr. Inf. Sci.* **36**, 514–546 (2022).
11. Ma, X. & Yang, Q. Research on the spatial location design of guidance signage systems to connect the space of transit-orientated development sites based on multii-software analysis. *Buildings* **15**, 683 (2025).
12. Vauclin, P., Wheat, J., Wagman, J. B. & Seifert, L. A systematic review of perception of affordances for the person-plus-object system. *Psychon Bull. Rev.* **30**, 2011–2029 (2023).
13. Liu, J. et al. Generative artificial intelligence perspectives on typical landscape types: Can ChatGPT compete with human insight? Landsc. *Urban Plan.* **264**, 105479 (2025a).
14. Kondyli, V., Bhatt, M., Levin, D. & Suchan, J. How do drivers mitigate the effects of naturalistic visual complexity? On attentional strategies and their implications under a change blindness protocol. *Cogn. Res. Princ Implic*. **8**, 54 (2023).
15. Liu, Q. et al. Development of a quantitative measurement on visual clutter in see-through display. *Front. Neurosci.* **17**, 1138225 (2023).
16. Cheng, C. F., Lin, C. J. & Lin, C. Y. Applying multiscale entropy for evaluating website visual complexity in an agile project: Using physiological data. *Appl. Ergon.* **126**, 104509 (2025).
17. Semizer, Y. & Rosenholtz, R. The effect of background clutter on visual search in video conferencing. *Cogn. Res. Princ Implic*. **10**, 40 (2025).
18. Kanaan, M. & Moacdieh, N. M. How do we react to cluttered displays? Evidence from the first seconds of visual search in websites. *Ergonomics* **64**, 1452–1464 (2021).
19. Lu, X. et al. Seeing in crowds: Averaging first, then max. *Psychon Bull. Rev.* **31**, 1856–1866 (2024).
20. Xu, Z. J., Lleras, A., Gong, Z. G. & Buetti, S. Top-down instructions influence the attentional weight on color and shape dimensions during bidimensional search. *Sci. Rep.* **14**, 31376 (2024).
21. Wang, Y., Buetti, S., Cui, A. Y. & Lleras, A. Color-color feature guidance in visual search. *Atten. Percept. Psychophys*. **87**, 1496–1517 (2025).
22. Yin, M. et al. Image-analysis-based method for exploring factors influencing the visual saliency of signage in metro stations. *Cogn. Syst. Res.* **92**, 101362 (2025).
23. Alinaghi, N., Giannopoulos, I., Kattenbeck, M. & Raubal, M. Decoding wayfinding: Analyzing wayfinding processes in the outdoor environment. *Int. J. Geogr. Inf. Sci.* **39**, 1770–1800 (2025).

24. Kim, K. & Bock, O. Acquisition of landmark, route, and survey knowledge in a wayfinding task: In stages or in parallel? Psychol. *Res* **85**, 2098–2106 (2021).
25. Brosens, L., Raes, A., Octavia, J. R. & Emmanouil, M. How future proof is design education? A systematic review. *Int. J. Technol. Des. Educ.* **33**, 663–683 (2023).
26. Yoghourdjian, V. et al. Scalability of network visualizationvisualisation from a cognitive load perspective. *IEEE Trans. Vis. Comput. Graph*. **27**, 1677–1687 (2021).
27. Ajani, K. et al. Declutter and focus: Empirically evaluating design guidelines for effective data communication. *IEEE Trans. Vis. Comput. Graph*. **28**, 3351–3364 (2022).
28. Cross, N. Design thinking: What just happened? Des. *Stud* **86**, 101187 (2023).
29. Litster, G., Cardoso, C. & Hurst, A. Analyzing problem framing in design teams: A systems mapping approach. *AI EDAM*. **38**, e15 (2024).
30. Lee, J. H. & Ostwald, M. J. The relationship between divergent thinking and ideation in the conceptual design process. *Des. Stud.* **79**, 101089 (2022).
31. Wei, C., Zou, J., Heinke, D., Liu, Q. & CoCoG Controllable visual stimuli generation based on human concept representations. In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, pp. 3178–3186. (2024).
32. Liu, S. et al. Marrying DINO with grounded prepre-training for open-set object detection. In: Lecture Notes in Computer Science. Springer Nature Switzerland, 38–55. (2025b).
33. Willoughby, M. T., Camerota, M., King, K. M., Nduku, T. & Piper, B. Leveraging item-level accuracy and reaction time to address ceiling effects in the measurement of inhibitory control. *Front. Psychol.* **14**, 861441 (2023).
34. Elsayed, M., Pelsmakers, S., Pistore, L., Castaño-Rosa, R. & Romagnoni, P. PostPost-occupancy evaluation in residential buildings: A systematic literature review of current practices in the EU. *Build. Environ.* **236**, 110307 (2023).
35. Boere, K., Anderson, F., Hecker, K. G. & Krigolson, O. E. Measuring cognitive load in multitasking using mobile fNIRS. *Neuroimage Rep.* **4**, 100228 (2024).
36. International Organization for Standardization ISO 21542. :*2021 Building Construction—Accessibility and Usability of the Built Environment* (ISO, 2021).
37. Carmona, M. Coding urban design: Constructing a wireframe for a place-focused urbanism. *Prog Plan.* **176**, 100775 (2023).

Figures

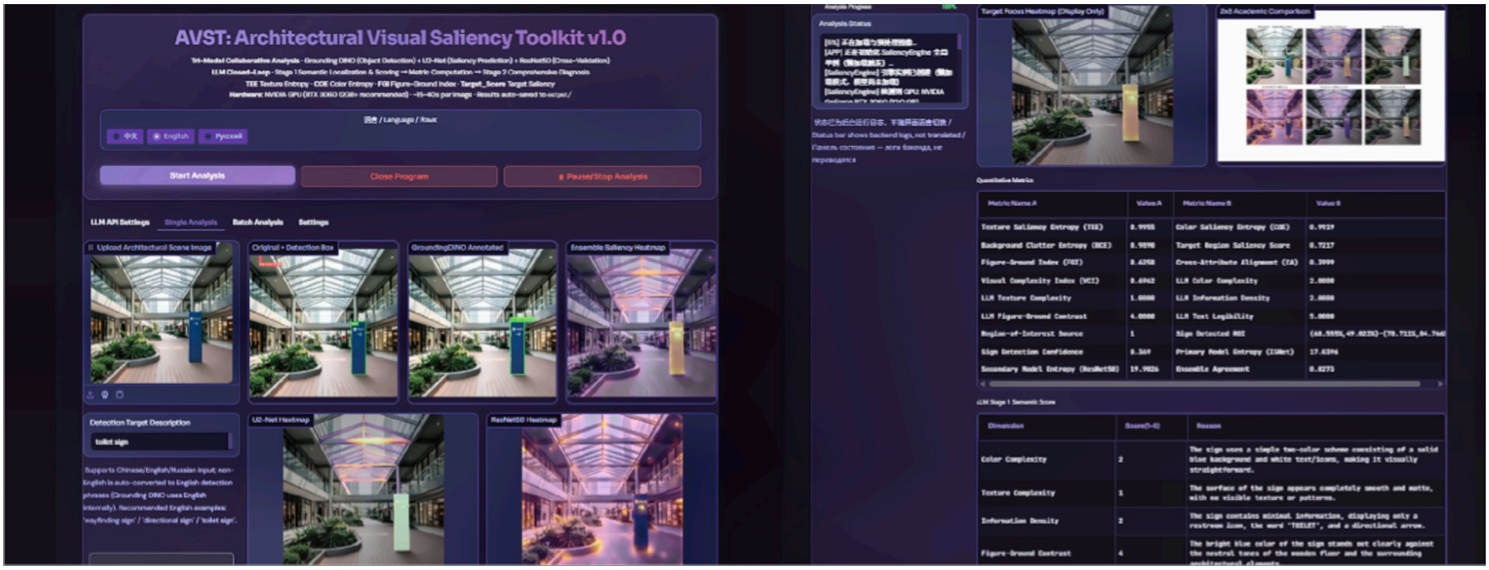


Figure 1

Architectural visual search toolkit (AVST) interface.

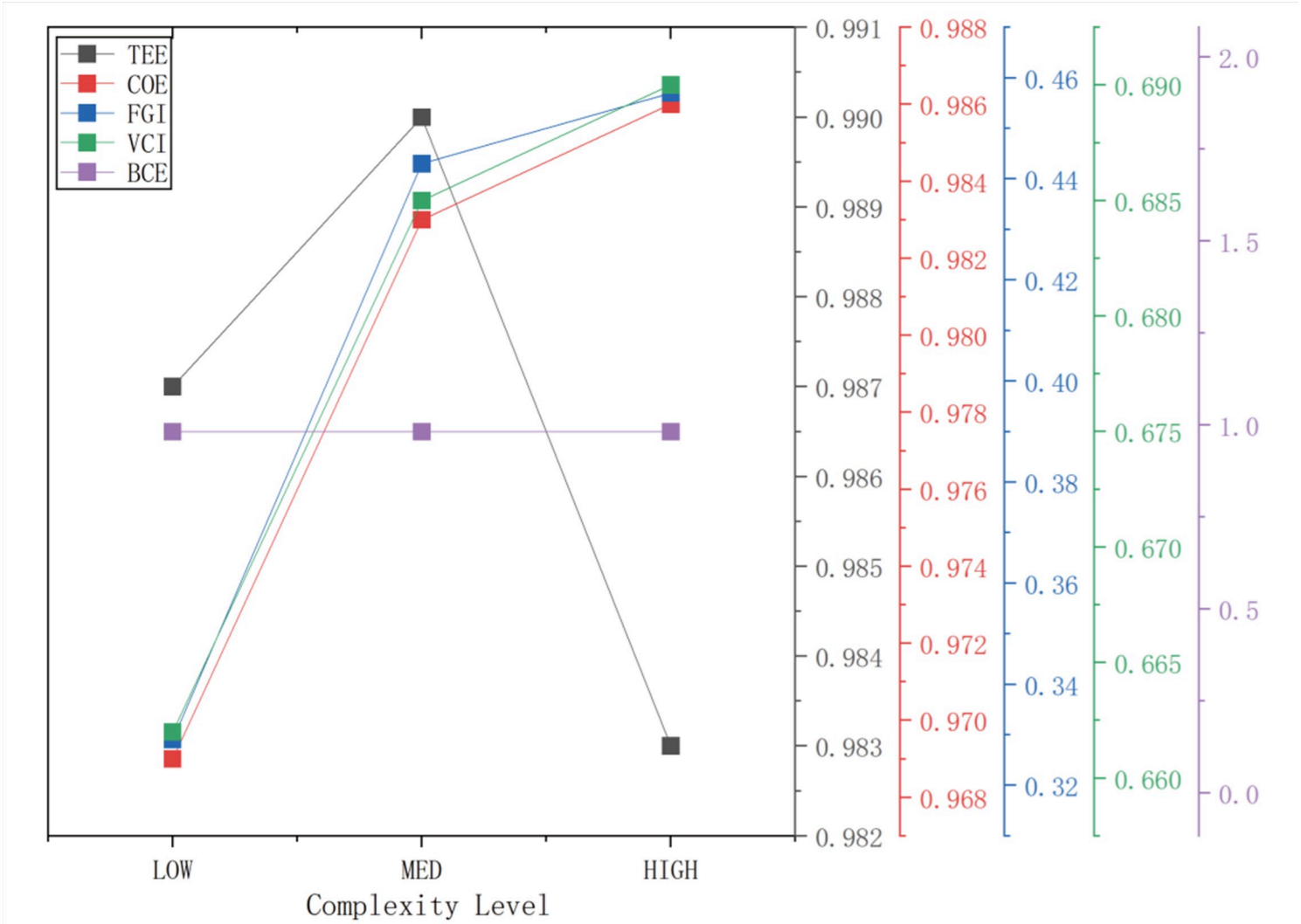


Figure 2

Manipulation check results: AVST metrics across complexity levels.

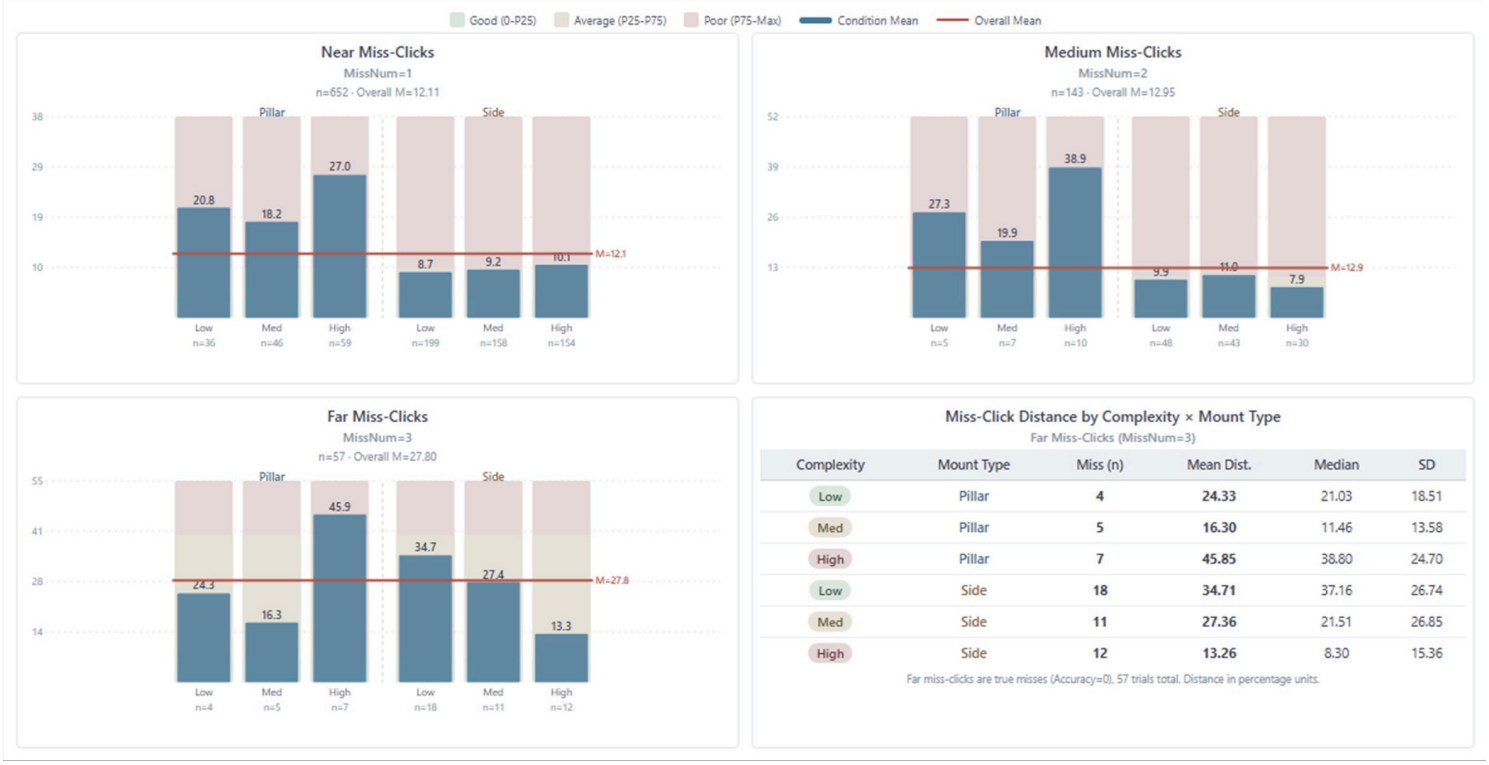


Figure 3

Miss-click distance by complexity and mounting type.

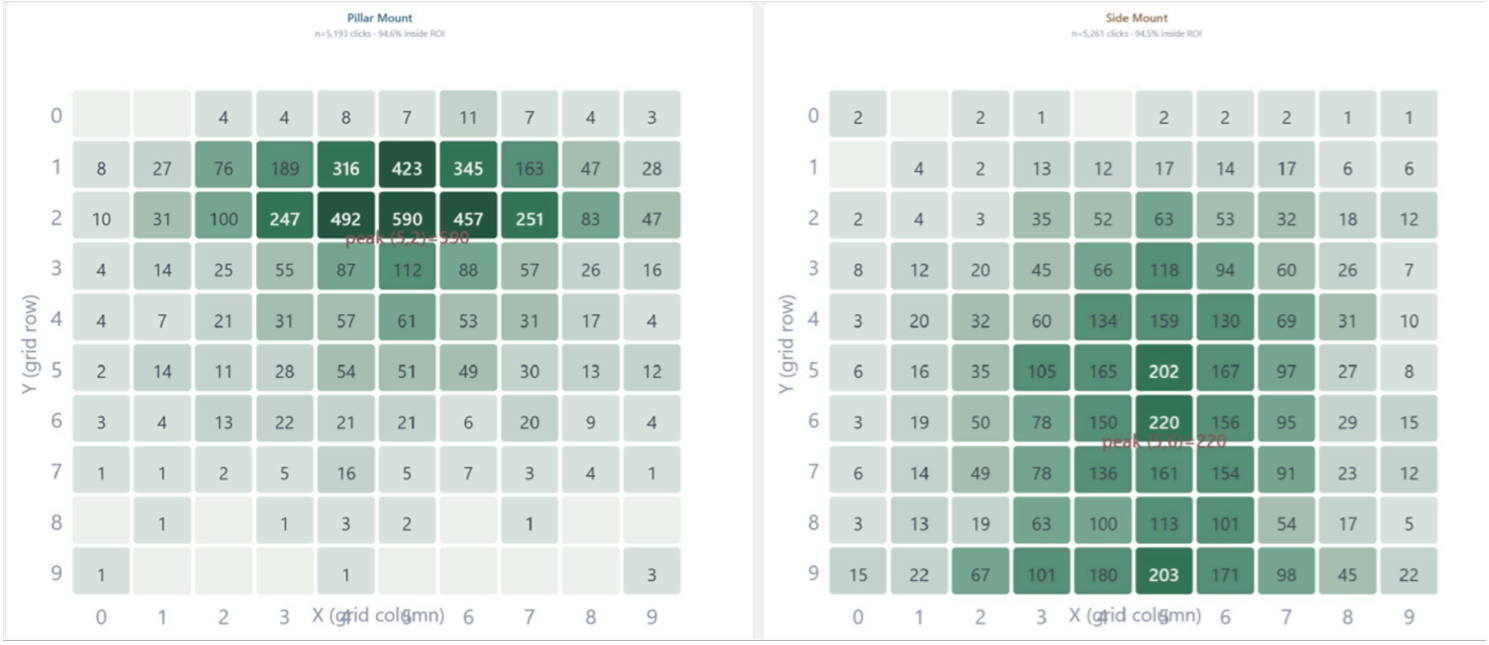


Figure 4

Click coordinate heatmap by mounting type.



Figure 5

Example of AVST single-image analysis results.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Graph01.png
- SupplementaryInformation.docx