Geomatica 77 (2025) 100063

Contents lists available at ScienceDirect

Geomatica

journal homepage: www.sciencedirect.com/journal/geomatica

Streetscape Analysis with Generative AI (SAGAI): Vision-language assessment and mapping of urban scenes

Joan Perez [a,*], Giovanni Fusco [b]

[a] Urban Geo Analytics, France
[b] Université Côte-Azur-CNRS-AMU-Avignon Université, ESPACE, France

ARTICLE INFO

Keywords:
Vision-language models
Street View imagery
Streetscape Analysis
Geospatial AI
Zero-shot inference

ABSTRACT

Streetscapes are an essential component of urban space. Their assessment is presently either limited to morphometric properties of their mass skeleton or requires labor-intensive qualitative evaluations of visually perceived qualities. This paper introduces SAGAI: Streetscape Analysis with Generative Artificial Intelligence, a modular workflow for scoring street-level urban scenes using open-access data and vision-language models. SAGAI integrates OpenStreetMap geometries, Google Street View imagery, and a lightweight version of the LLaVA model to generate structured spatial indicators from images via customizable natural language prompts. The pipeline includes an automated mapping module that aggregates visual scores at both the point and street levels, enabling direct cartographic interpretation. It operates without task-specific training or proprietary software dependencies, supporting scalable and interpretable analysis of urban environments. Two exploratory case studies in Nice and Vienna illustrate SAGAI's capacity to produce geospatial outputs from vision-language inference. The initial results show strong performance for binary urban–rural scene classification, moderate precision in commercial feature detection, and lower estimates, but still informative, of sidewalk width. Fully deployable by any user, SAGAI can be easily adapted to a wide range of urban research themes, such as walkability, safety, or urban design, through prompt modification alone.

1. Introduction

Assessing the qualities and functions of urban streetscapes is essential to understand walkability, safety, commercial vitality, and social life in cities [1–3]. However, traditional methods for observing and evaluating street-level conditions, such as field surveys, audits, and manual photo interpretation, remain time-consuming, labor-intensive, and difficult to scale beyond small pilot zones [2]. Geo-processing of vector models of the built environment allows the assessment of morphometric properties of the skeletal streetscape [4,5], that is, of the 3D masses defining the visual channel of the streetscape. However, more fine-grained aspects of the streetscape, including building façades, urban furniture, sidewalks, materials and texture of built-up elements, greenery, cleanliness, etc. can play a fundamental role for human perception and usage of the public space of the street. These elements are collectively referred to as the skin of the streetscape [4], and are rarely the object of automated assessment through street imagery. Clarke et al. [6] and Rundle et al. [7] are among the first to propose the use of streetscape images accessed online to produce a complete assessment of the streetscape, but they end up using traditional manual audit methods remotely instead of on site. While previous work has leveraged convolutional neural networks (CNNs) and object detection models to extract discrete features from streetscape imagery—such as sidewalks, trees, building façades, or cars—these approaches are generally task-specific. They require extensive annotated datasets and custom training for each visual variable of interest. As a result, they are difficult to adapt to new cities or scoring criteria, and struggle with capturing more abstract qualities like perceived safety, beauty, or walkability.

In contrast, vision-language models (VLMs) such as LLaVA (Large Language and Vision Assistant) offer a flexible alternative. These generative AI models combine a visual encoder with a large language model, allowing them to interpret street-level imagery and respond to natural language prompts. Instead of requiring pre-labeled data or retraining, a VLM like LLaVA can produce structured outputs—such as counts, classifications, or measurements—directly from descriptive instructions. For instance, it can be asked: "How wide is the sidewalk?" or "Are any commercial storefronts visible?", and respond with a scalar answer like 1.5 m or 2 storefronts.

* Corresponding author.
E-mail addresses: jperez@urbangeoanalytics.com (J. Perez), giovanni.fusco@univ-cotedazur.fr (G. Fusco).

https://doi.org/10.1016/j.geomat.2025.100063
Received 10 May 2025; Received in revised form 14 June 2025; Accepted 26 June 2025
Available online 28 August 2025

This capability is particularly valuable in street view auditing, where many important features are qualitative, spatially heterogeneous, or difficult to define in advance. By enabling zero-shot interpretation of visual content, generative VLMs dramatically reduce the barriers to scalable, customizable streetscape assessment. LLaVA stands out among current models due to its open-source nature, efficient architecture, and support for lightweight deployment—making it a practical tool for geospatial research and urban analytics.

The present research develops and operationalizes SAGAI (Streetscape Analysis with Generative Artificial Intelligence), a four-step workflow designed to automate the application of such protocols using open-access geospatial data and vision-language models. SAGAI enables reproducible and scalable image-based scoring on large urban areas. Built entirely in Python and deployable in free-tier Google Colab environments, the workflow links OpenStreetMap, Google Street View, and LLaVA to generate structured indicators of the built environment at street scale. Each of the four modules—point generation, image retrieval, model inference, and spatial aggregation—is designed to be modular, transparent, and adaptable to a wide range of urban research questions.

The remainder of this paper is organized as follows. Section 2 reviews related work in streetscape analysis, AI applications in urban contexts, and recent developments in vision-language models. Section 3 presents the SAGAI workflow, describing each of its four modules. Section 4 illustrates the pipeline through two case studies in Nice, France and Vienna, Austria. Section 5 reports accuracy results from manual validation and highlights interpretative gaps between AI predictions and human annotations. Section 6 discusses performance, limitations, and future extensions. We conclude in Section 7 by outlining implications for urban analysis and open-source tool development.

2. Related work

Research on streetscape analysis has long relied on computer vision to extract information from urban imagery, particularly from platforms like Google Street View. Before the rise of generative artificial intelligence (GenAI), urban studies predominantly employed discriminative models—algorithms designed to detect and classify specific visual features using pre-labeled datasets. A foundational example is Treepedia [8], which calculated a Green View Index through semantic segmentation of GSV panoramas to quantify urban tree canopy. Street-level imagery has since become a vital data source in urban research, enabling a range of applications: detecting sidewalks [9], measuring visual clutter from vegetation [10], classifying façade-level attributes such as fences and setbacks [11], and even inferring neighborhood demographics from the types of visible vehicles [12]. Recent works such as Urban Visual Intelligence [13] consolidate these approaches into a structured hierarchy, leveraging discriminative AI and deep learning techniques—such as semantic segmentation and scene classification—for large-scale physical and socioeconomic analysis of cities based on street-level images.

These studies typically relied on convolutional neural networks (CNNs) and object detectors such as R-CNN [14] or YOLO [15], trained on extensive labeled corpora via supervised or semi-supervised learning. While effective, such models are constrained by their architecture and training scope—optimized for specific tasks, rather than for contextual interpretation or semantic abstraction. Their generalizability depends heavily on dataset coverage and annotation quality, often requiring domain-specific training pipelines and hard-coded feature engineering.

Parallel to this, the field of geospatial artificial intelligence (GeoAI) is emerging as a broader paradigm integrating spatial data, machine learning, and advanced computational tools. GeoAI applications have proliferated across multiple domains of human geography—including urban morphology, mobility, health, and inequality—leveraging spatial imagery (satellite, aerial, or street-level) for classification, prediction, and simulation tasks [16]. Recent work further demonstrates GeoAI's capacity to map building footprints, classify land use, and model natural hazards using multimodal datasets [17]. However, despite these advancements, most GeoAI pipelines remain discriminative in nature—focused on identifying what is in a given image, rather than generating data, novel descriptions or simulating transformations from a given image.

A shift is currently underway with the emergence of vision-language models (VLMs)—a class of generative architectures that jointly process images and text within a unified framework. Unlike traditional discriminative models trained on object labels and bounding boxes, VLMs learn from large datasets of image–text pairs, enabling more flexible and contextual interpretation of visual scenes. These models typically consist of two core components: a vision encoder, which extracts visual features from images, and a large language model (LLM), which handles the textual input, performs reasoning, and generates responses.

Models such as LLaVA [18], BLIP-2 [19], GPT-4 with Vision [20], and Flamingo [21] all follow this general architecture. A widely adopted vision encoder in these systems is CLIP [22], an open-source model trained to align image and text representations in a shared embedding space via contrastive learning. While CLIP is commonly used due to its performance and accessibility, other encoders—such as Vision Transformers (ViT) [23] or Swin Transformers [24] —can also fulfill this role depending on the model.

In the case of LLaVA, the architecture explicitly wraps these components with an intermediate projection layer that aligns the image embeddings from the vision encoder with the token embeddings expected by the LLM. This enables multimodal inference, such as answering image-based questions, generating descriptions, or reasoning about spatial configurations, without task-specific retraining. Its open-source nature and efficient implementation—including quantized variants for deployment on consumer-grade hardware—make LLaVA especially suited for research and practical applications in geospatial and urban analytics.

Despite these advances, the potential of vision-language models (VLMs) for geospatial and urban research remains largely underexplored. A few recent studies have begun to experiment with approaches that resonate with the principles underlying SAGAI. For example, StreetViewLLM integrates street-level imagery with geographic metadata to infer location and semantic context, employing chain-of-thought prompting to enhance spatial reasoning [25]. Similarly, CityLLaVA adapts a LLaVA-style architecture to urban safety scenarios, combining structured prompts and bounding box inputs to interpret city environments [26]. Blečić et al. [27] propose a multimodal LLM-based workflow for assessing urban walkability from street view images, which augments visual scoring with textual interpretation to support planning decisions. In a complementary direction, Wei et al. [28] introduce GeoTool-GPT, a fine-tuned LLaMA-2 model designed to interface with GIS tools, while Beydokhti et al. [29] explore qualitative spatial reasoning for geospatial question answering. Additionally, Schmidt et al. [30] evaluate the spatial accuracy of several VLMs—including LLaVA1.6 and GPT-4o—in geocoding flood-related imagery.

Taken together, these studies reflect a growing interest in applying large multimodal and language models to urban domains. However, they also highlight a shared methodological challenge: how to evaluate generative models that do not rely on task-specific training. Unlike traditional object detection models, which are validated through comparisons with pixel- or label-level ground truth, vision-language models like LLaVA operate in a zero-shot setting and require alternative evaluation strategies. Most of the aforementioned works, including ours, rely on human-in-the-loop validation—comparing model predictions with expert assessments or manual annotations—to assess semantic plausibility and consistency. This paradigm emphasizes interpretability and generalization over task-bound accuracy, making it especially relevant for open-ended, perception-based urban questions.

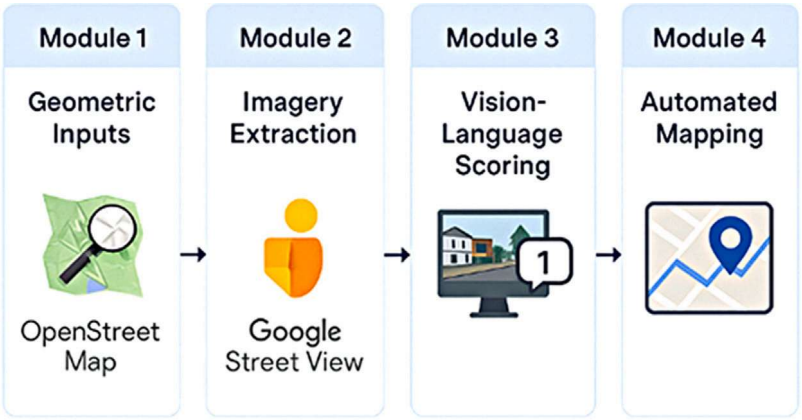


Fig. 1. The Four Modules of the SAGAI Workflow.

SAGAI contributes to this evolving field by combining generative vision-language reasoning with reproducible geospatial workflows, explicitly linking semantic image interpretation with open-access spatial datasets. SAGAI was developed within the emc2 research project [31], aiming to assess the potential of suburban areas for a human-centered 15 min city model. Within the project, urban patterns were identified within streetscapes of test areas, combining fieldwork data and ad hoc geoprocessing tools [32]. The need emerged to automate the workflow using street imagery as input and performing combined operations of object recognition, classification, counting and measurement to assess the presence/absence of the selected patterns, above all at the level of the streetscape skin. SAGAI is thus the first step towards automating complex operations of streetscape assessment. It is designed to generate interpretable numeric indicators—such as walkability, conviviality or safety scores—at scale, while preserving sensitivity to urban street network topology and human-centered planning theory [33,34]. By doing so, SAGAI bridges the gap between AI-powered image analysis and actionable urban insights, enabling customizable prompts and spatialized outputs across diverse urban themes.

3. Methodology: The SAGAI workflow

SAGAI is structured into four modules, each implemented in Python and designed to run in Google Colab. All scripts use Google Drive for reading inputs and saving outputs, requiring a one-time authentication step at the beginning of each session. The four modules are described below (see Fig. 1).

3.1. Module 1: OSM point generator

The first module of the SAGAI workflow is designed to generate sample points along the street network of a user-defined area. It operates without requiring any input data beyond a bounding box, as it automatically retrieves the relevant street geometries from OpenStreetMap. This ensures global applicability while minimizing the preparation burden on the user.

Once the network is downloaded, points are distributed along the streets based on two user-defined parameters: spacing and offset. The spacing determines the distance between points, while the offset prevents placing points too close to intersections. Each point is associated with its corresponding street segment via a unique identifier, enabling subsequent spatial aggregation steps.

The output consists of a single GeoPackage containing two layers: the cleaned street network and the set of generated points. These are projected to appropriate coordinate systems for accurate distance calculations and then reprojected to geographic coordinates for compatibility with later modules. This general-purpose module provides the spatial input for image retrieval and visual scoring tasks.

3.2. Module 2: Street view batch downloader

The second module of SAGAI automates the retrieval of street-level imagery from Google Street View based on the geographic points generated in the first module. For each location, the script sends requests to the Google Street View Static API to download images in four predefined directions—typically the cardinal angles (0°, 90°, 180°, and 270°). The number and orientation of views can be adjusted by the user, as can camera parameters such as pitch and field of view angles. To detect missing or invalid imagery, a filtering step is applied immediately after download: images dominated by a single pixel value—commonly used in Google's "no imagery" placeholders—are flagged accordingly.

The user must provide an active API key with access to the Street View Static API. The output is a structured dataset of georeferenced images that serves as input for the subsequent image-based scoring module. Designed for high-throughput execution in the Google Colab environment, the module can scale efficiently to thousands of locations with minimal user intervention.

3.3. Module 3: Scene assessment with LLaVA

The third module is the core of the SAGAI pipeline. It performs automated image-based scoring using the LLaVA v1.6 model with a Mistral-7B backbone. It applies structured prompts to each image to extract specific visual information—such as classification as urban scene, commercial presence, or sidewalk width—and returns a corresponding numerical score. The model is loaded in 4-bit quantized format to support efficient inference in memory-constrained environments. All components are open-access and retrieved from Hugging Face, a widely used repository for distributing pretrained machine learning models. By default, the script uses the publicly available checkpoint (a snapshot of the model's state at a particular point during training) liuhaotian/llava-v1.6-mistral-7b, which can be downloaded without authentication. If users wish to use private weights, a Hugging Face account and access token may be required.

A custom scoring function is implemented to structure the interaction between the image input and the LLaVA model. It follows LLaVA's conversation architecture, embedding the prompt with image tokens and formatting it using role-based templates. Images are preprocessed using CLIP-compatible vision encoders and inference is performed with low-temperature sampling and stopping criteria adapted to extract concise scalar outputs. Temperature sampling adjusts the randomness of the model's predictions—lower values make the output more deterministic, which is desirable for scoring tasks. Stopping criteria are used to terminate the generation process once a specific token or pattern

is detected, ensuring that only the relevant part of the response is returned. This controlled generation strategy helps enforce the expected numerical format and improves reproducibility across sessions.

Each image is processed independently using a user-defined prompt that encodes a scoring rule in natural language. The default configuration includes three pre-coded scoring tasks: (i) binary classification of urban versus rural scenes, (ii) shopfront counting, and (iii) estimation of sidewalk width. The desired task is selected through a task identifier (T1, T2, or T3), which activates the corresponding prompt template. These prompts can be freely modified or replaced by the user to implement other visual feature scoring strategies.

The script automatically reads the images retrieved by Module 2. Files previously marked as unavailable are skipped from visual analysis but still recorded in the output table to preserve traceability. A resuming mechanism allows the script to continue from the last processed image if an output file already exists. This feature is particularly useful when working with large-scale study areas, such as entire metropolitan regions, where long runtimes or interruptions are common.

The module is fully automated and requires minimal user input. The numerical scores produced in this module serve as direct inputs for the final stage of the pipeline, which aggregates and maps the results at the spatial level.

3.4. Module 4: Geospatial scoring aggregation and mapping

The fourth module of the SAGAI workflow aggregates the image-based scores and integrates them into a geospatial dataset. It takes as input the numerical outputs generated in Module 3 along with the spatial geometries produced in Module 1, and summarizes the values at both the point and street segment levels.

All four summary statistics are retained in the output GeoPackage for both points and streets. The user can then select either the mean or sum values for cartographic rendering, depending on the task and desired interpretation. Indeed, a visualization tool is integrated to render two thematic maps: one showing the values assigned at the point level and the other the scores at the street level. Streets and points without any valid data are displayed in gray.

The module requires three inputs: the case study name, the task identifier, and the aggregation mode for the maps. This final step of the pipeline allows users to translate raw visual assessments into interpretable spatial scores suitable for urban analysis, cartographic representation, or downstream statistical modeling. The outputs can serve as a foundation for further spatial analysis. In particular, the point-level indicators can be passed into complementary methods that retain the underlying street-based topology (e.g. ILINCS Yamada and Thill [35]) or to identify recurring spatial configurations.

3.5. Open-source availability and repository access

The full implementation of SAGAI (v1.0) is released as open-source software and publicly accessible at: https://github.com/perezjoan/SAGAI

The repository includes:

- Modular Python scripts for all four components of the pipeline;
- Predefined prompts for the three scoring tasks (urban classification, storefront counting, and sidewalk width estimation);
- Sample inputs and outputs for the Nice and Vienna case studies;
- Colab-ready notebooks enabling direct execution in the cloud;
- Detailed "NOTICE" files for each module, covering inputs, configuration, and expected outputs.

The platform is designed for zero-shot deployment on free-tier Google Colab environments. Only minimal configuration (bounding box coordinates, API key, and task identifier) is required to run the full pipeline. No local installation or specialized hardware is needed. The codebase is released under an open license; details on dependencies and usage rights are provided in the repository's LICENSE file. In addition to the current notebook-based version, future releases of SAGAI are planned to include an installable Python package.

4. Empirical evaluation: Two case studies in the peripheral areas of Nice and Vienna

To illustrate the capabilities of the SAGAI workflow, two contrasting urban environments were selected: the Paillon Valley in the north-eastern outskirts of Nice, France, and the western districts of Penzing and Wolfersberg in Vienna, Austria. Both areas present complex spatial structures but differ significantly in morphology, density, and land use. The Paillon Valley follows a narrow strip of flatland, linking the housing project of L'Ariane, in Nice, with the surrounding municipalities. It is characterized by linear urbanization, mixed land uses, and constrained topography. In contrast, the Penzing-Wolfersberg sector of Vienna lies at the foothills of the Wienerwald and features a low-density residential fabric with allotment gardens, winding roads, and forested patches. These case studies were chosen for their heterogeneity and the diversity of visual environments they offer, enabling the evaluation of SAGAI across distinct urban forms. Fig. 2 presents the spatial distribution of sampled points and their corresponding Street View image availability for the two case studies—Paillon Valley (Nice) on the left and Penzing–Wolfersberg (Vienna) on the right. In both areas, points were generated during module 1 along the OSM street network using a fixed offset of 15 m from intersections and a spacing interval of 40 m. The tabulated metrics show that despite covering a smaller bounding box (4.96 km^2), the Vienna case includes more street segments (1228 vs. 955) and a slightly longer total street length (154.84 km vs. 141.31 km), reflecting the denser layout of the street network in the area.

The Nice case study achieved near-complete coverage, with 1775 out of 1898 points successfully associated with four valid images. Only 123 points (6.5%) lacked coverage. In contrast, the one in Vienna shows significant coverage gaps: 449 of 1948 points (23%) returned no imagery. These gaps, visualized as red points in the figure, are especially concentrated along the southern and northeastern fringes of the Vienna study area. These discrepancies are not the result of the sampling procedure but rather reflect underlying differences in Street View coverage. The lack of imagery is often associated with areas that are poorly served by vehicular access, such as cul-de-sacs, pedestrian-only paths, or segments located within forested or semi-private residential zones. These features are more prevalent in the hilly and low-density morphology of Penzing and Wolfersberg, explaining the higher rate of missing images relative to the more continuous and linear urban form of Paillon Valley in Nice (see Table 1).

To evaluate the SAGAI workflow, all three predefined tasks of module 3 are deployed in both case studies. These include: (T1) a categorization task to classify scenes as either urban or rural; (T2) a counting task to detect the number of visible commercial storefronts; and (T3) a measuring task to estimate the visible width of sidewalks. Each task is implemented using a modular prompt architecture that guides the LLaVA vision-language model to return concise scalar values. Prompts are dynamically assembled from four components — *{role_description}*, *{theory_model}*, *{task}*, and *{response_format}*—allowing for

Table 1
Street network and imagery statistics for the nice and vienna case studies.

Metric	Nice (Paillon Valley)	Vienna (Penzing & Wolfersberg)
Number of street segments	955	1228
Total street length	141.31 km	154.84 km
Total number of points	1898	1948
Bounding box surface	6.18 km^2	4.96 km^2
Points with 4 images	1775	1499
Points with no coverage	123	449

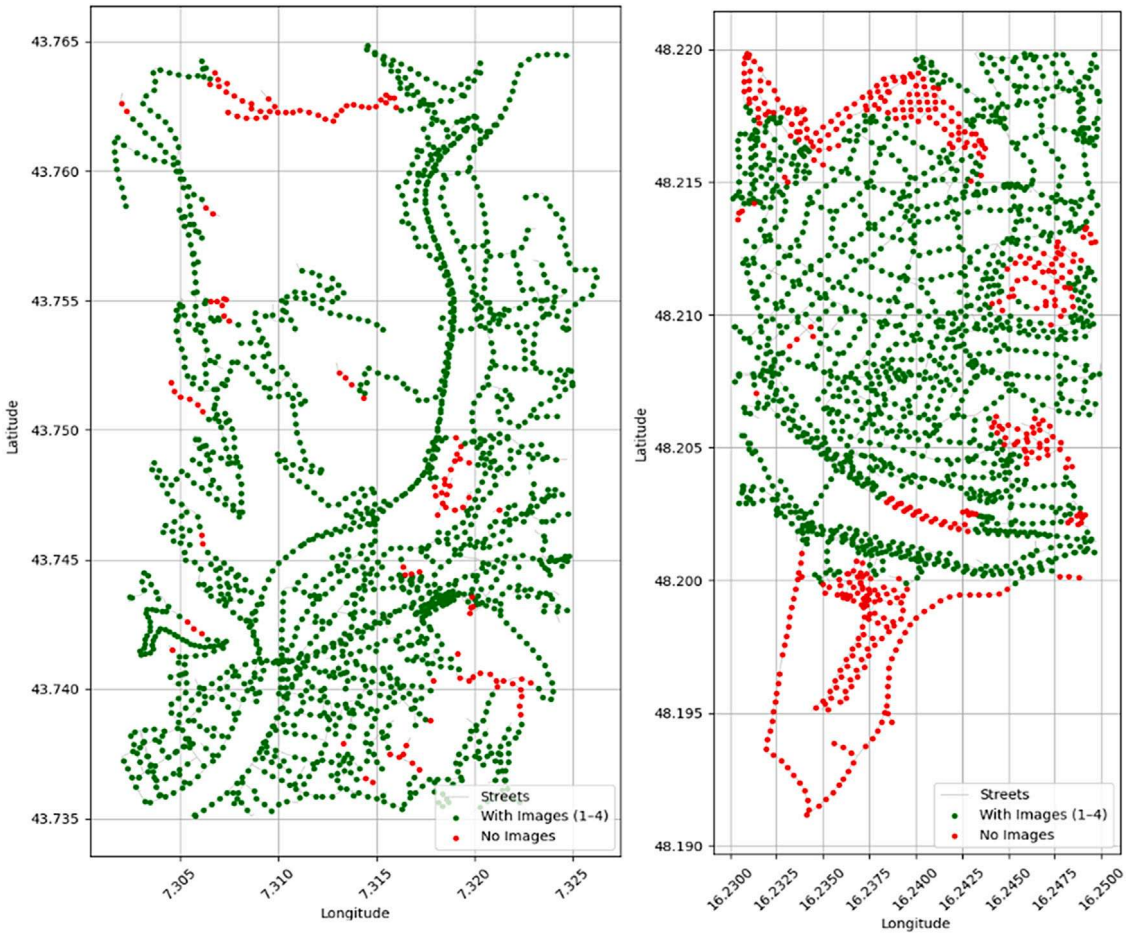


Fig. 2. Street View Image Availability by Sample Point in Nice (left) and Vienna (right).

flexible adaptation to different visual features or scoring schemes. The resulting outputs range from binary classifications (T1), to cardinal counts (T2), and continuous numeric estimates (T3). This triad of tasks demonstrates SAGAI's capacity to extract diverse visual information all relevant to the assessment of urban streetscapes—establishing a robust and extensible approach for comparative spatial analysis of streetscapes. Complete prompt examples used in version 1.0 of SAGAI are included in Appendix A.

The scoring outputs generated for each image are then aggregated spatially following the procedure described in Module 4. For every sampled point, the scores obtained across four directional images are summarized into a single point-level metric by averaging and summing valid values. These point-level results are further linked to their corresponding street segments to produce aggregated scores at the street level. The following section presents the results of this geospatial aggregation for each task, comparing distributions across the Nice and Vienna study areas.

5. Experimental results

To evaluate the reliability of SAGAI, a manual validation procedure was conducted across the three predefined scoring tasks—categorization (T1), counting (T2), and measuring (T3)—in both case study areas. A stratified random sample of 300 image-level predictions was selected and compared against human annotations. For each task, class-specific precision and total accuracy were calculated. Ambiguous cases—where the human annotator could not confidently assign a score due to unclear visual content, image truncation, or corrupted input—are marked as "NA". These are excluded from accuracy calculations but still counted in the reported sample sizes.

Table 2 summarizes the precision and overall accuracy results for each city and task. In Task 1 (urban vs. rural categorization), where the model assigns a score of 0 for rural and 1 for urban, SAGAI achieved its highest performance, with 92.73% total accuracy. In Task 2 (storefront counting), which uses 0 for no visible storefronts, 1 for one storefront, and 2+ for multiple, performance declined—particularly for intermediate categories. Task 3 (sidewalk width estimation), which accepts a range of values such as 0, 0.5, 1.0, 1.5, and so on, showed the lowest precision overall. This reflects the increased complexity and difficult objectification of fine-grained visual assessments.

Beyond the overall accuracy metrics, manual inspection revealed systematic differences in how the model interprets visual scenes compared to human annotators. First, we noted that ambiguity is not exclusive to AI since certain cases proved challenging even for human annotators. Fig. 3 illustrates several examples across the three tasks (T1–T3) where the model's predictions diverged from human judgment. Red boxes[1] illustrate features that were likely misinterpreted by the AI

[1] Red and green rectangles are not object detection bounding boxes, but visual annotations added manually to highlight interpretation differences between human and AI assessments.

Table 2
Accuracy and class-specific precision for each scoring task in nice and vienna (Random Samples, Human Validation)

Task	Location	Precision 0	Precision 0.5	Precision 1	Precision 1.5	Precision 2+	Overall Accuracy	NA Cases
T1: Categorization	Nice	95.83% (23/24)	–	84.00% (21/25)	–	–	89.80% (45/50)	1
	Vienna	91.30% (21/23)	–	91.67% (22/24)	–	–	91.49% (43/47)	3
	Total	–	–	–	–	–	91.67% (88/96)	4
T2: Counting	Nice	90.00% (18/20)	–	70.00% (14/20)	–	60.00% (12/20)	73.33% (44/60)	0
	Vienna	100.00% (20/20)	–	45.00% (9/20)	–	20.00% (4/20)	55.00% (33/60)	0
	Total	–	–	–	–	–	64.17% (77/120)	0
T3: Measuring	Nice	71.43% (10/14)	25.00% (3/12)	53.85% (7/13)	53.33% (8/15)	–	51.85% (28/54)	6
	Vienna	53.33% (8/15)	61.54% (8/13)	64.29% (9/14)	46.67% (7/15)	–	56.14% (32/57)	3
	Total	–	–	–	–	–	54.05% (60/111)	9

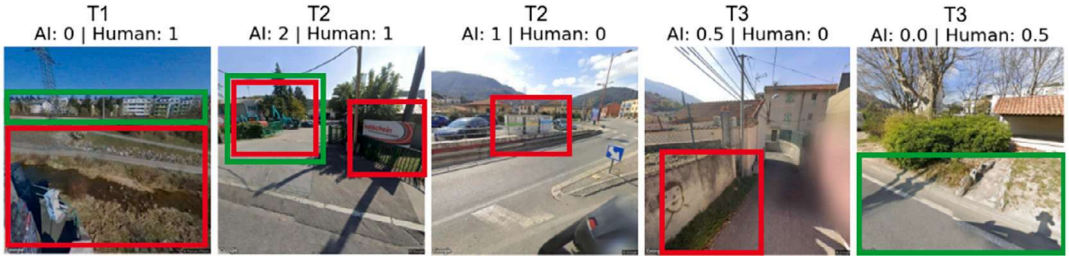


Fig. 3. Examples of Misclassified Images with Highlighted Interpretation Gaps between Human and AI Scoring.

while the green boxes highlight the elements that human annotators considered relevant when assigning their score.

For Task 1 (urban vs. rural classification), failures typically stemmed from hybrid landscapes where natural and built elements coexist. In the example shown, the model misclassified a clearly urban setting as rural, possibly due to the presence of large green spaces in the foreground. In Task 2 (storefront counting), we observed common confusions with non-commercial elements or with commercial elements which are not storefronts. The model often interpreted advertising signs, parked commercial vehicles, utility structures or clusters of waste containers as storefronts. Task 3 (sidewalk width estimation) presented the most diverse range of misinterpretations. One common issue was the consistent identification of grass strips—typical of North American residential streets—as sidewalks, even when they were clearly not usable pedestrian space. Painted edge lines serving as pedestrian sidewalks were also often not detected as sidewalks by the AI.

Another recurring issue arose when two sidewalks appeared in the same image: without instruction on which side to evaluate, the model made inconsistent or averaged estimations. While this ambiguity had limited impact when sidewalks were of similar widths on both sides, it remained a source of error in other configurations. Notably, the model never returned a width above 2 m, even when wider sidewalks were visible in the validation sample. Yet, not all mismatches should be interpreted as complete failures. For instance, detecting one storefront instead of two, or slightly underestimating a sidewalk width, is qualitatively better than returning no detection at all. That said, some failure patterns cannot be explained. For example, in certain images with no visible sidewalk at all, the model still returned a non-zero width estimate.

These examples illustrate how the LLaVA v1.6 Mistral-7B model, despite performing well in structured tasks, struggles with contextual reasoning in more complex urban scenes. They also highlight the value of integrating options for expressing uncertainty in the prompts. For the full set of annotated examples and corresponding scores, readers are referred to Appendix B.

Fig. 4 shows the spatial distribution of predicted urban scores from Module 4 – Task 1 (categorization) for the two case study areas: Nice (left) and Vienna (right). The left panel of each city map shows averaged point-based predictions, while the right panel displays averaged scores at the street segment level. Indeed, even point-based predictions have average values of the urban character of the scene, as 4 street-view images are normally associated with each point. A continuous color scale from light yellow (0: rural) to dark purple (1: urban) reflects the model's results, with gray elements representing locations where no prediction could be made due to missing imagery.

In both cities, the model clearly differentiates central urban cores from peripheral or more natural areas. In Nice, the densely urbanized housing project of L'Ariane and the valley floor stand out with high scores, while hillside streets and isolated paths along the Paillon River tend to receive lower urban character values. Similarly, in Vienna, the dense road network of Penzing is consistently classified as urban, while peripheral allotment zones and forest-adjacent roads receive lower scores. The aggregated street-level visualizations help reveal broader structural patterns and remove isolated point-level noise, highlighting the value of spatial smoothing.

Fig. 5 presents the summed outputs for Task 2 (Storefront Counting), visualized as total scores at both the point level and street-segment level (right subpanels) for Nice and Vienna. The underlying hypothesis for the sum statistics is that the 40 m spacing among the sample points is able to produce a good coverage of street segments while minimizing double counting of storefronts. In Nice, the model accurately highlights commercial corridors such as boulevard de l'Ariane and the central axis of Paillon Valley, the Auchan Shopping Mall in the east and the central street of the municipality of Drap in the north, —consistent with both observed commercial density and the high validation accuracy achieved in this case. Conversely, predicted scores drop sharply in low-density or peri-urban areas, aligning with expectations. In Vienna, although several high-score clusters align with known commercial zones—particularly in the southern sector and near key junctions—interpretation is less straightforward due to lower model accuracy in this area, especially for intermediate storefront counts, as confirmed by the manual validation results.

Fig. 6 presents the average predicted sidewalk width scores from Task 3 (Measuring), visualized at both the point level (left subpanels)

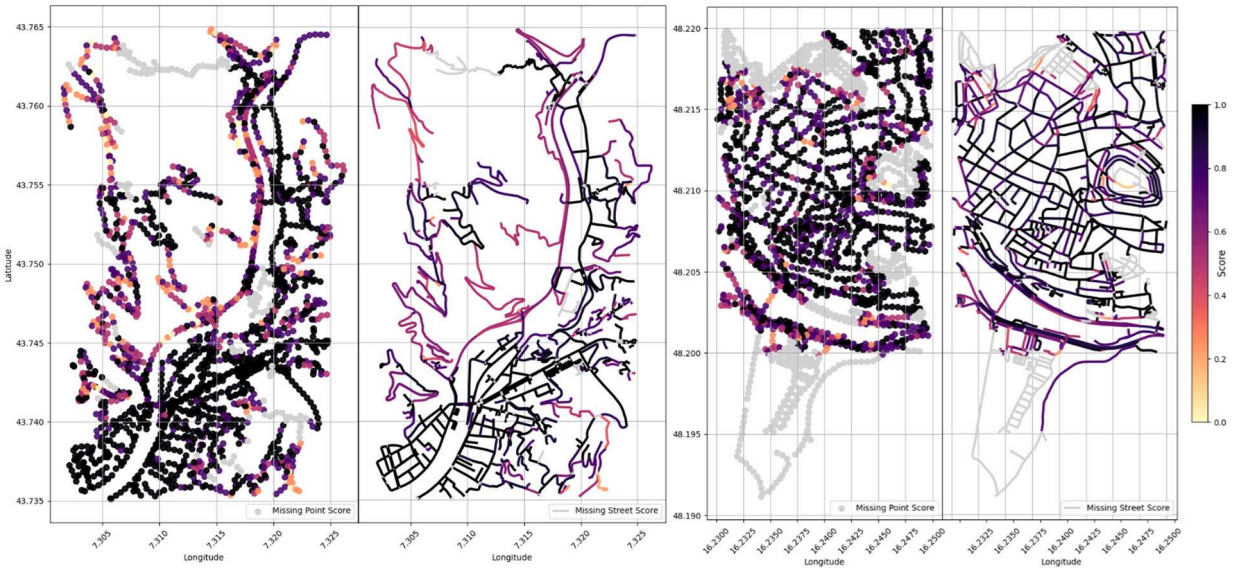


Fig. 4. Averaged Urban Character Scores from Module 4 – Task 1 (Categorization) for Nice (left) and Vienna (right), at Point and Street Segment Levels.

and as segment-level averages (right subpanels) for the Nice (top) and Vienna (bottom) study areas. Some spatial coherence can be observed in these results: for instance, wider thoroughfares such as Boulevard de l'Ariane in Nice tend to display higher predicted scores than adjacent narrower streets, which aligns with real-world expectations. However, other inconsistencies appear, including scattered high values in locations unlikely to feature wide sidewalks and abrupt variations between adjacent segments. In addition to the averaging of four separate images per location—which tends to dampen values—these patterns reflect the broader range of error types previously discussed for this task. As a result, predicted widths rarely exceed 1 m, even in visibly wider areas. While some encouraging spatial signals emerge, the overall uncertainty and variability of Task 3 predictions limit our ability to confidently interpret or comment on these results.

6. Discussion and perspectives

To ensure accessibility and broad usability, the current implementation of SAGAI (v1.0) deliberately adopts the lightest possible configuration. The workflow uses the official LLaVA v1.6 Mistral-7B checkpoint in 4-bit quantized format, allowing the full pipeline to run on free-tier environments in Google Colab. This design choice ensures that SAGAI is readily deployable by users without access to high-performance computing resources. Despite this minimal setup, the algorithm already demonstrates strong performance across some visual scoring tasks. Moreover, there is considerable potential to improve accuracy through lightweight strategies alone. For instance, prompt engineering—e.g. instructing the model to return "NA" when uncertain or to append qualifiers like "(unsure)"—can improve robustness without loading a heavier model. Another strategy would be multi-pass voting, where the model is run multiple times on the same image and only the majority result is kept. This helps retain consistent predictions and filter out unstable outputs. Rule-based validation filters, such as suppressing values below or above realistic thresholds or filtering predictions on visually degraded images, may further enhance output quality. These strategies offer a pragmatic way to improve performance while preserving SAGAI's lightweight and accessible nature.

From a research perspective, multiple extensions are planned to improve the accuracy and flexibility of SAGAI. The current version relies on Mistral-7B, a 7-billion-parameter language model, which balances accuracy and computational cost. Future versions may explore larger or more advanced backbones, including LLaMA-2 13B [36] or Mixtral 8 × 7B [37], which may capture finer visual nuances and improve interpretability in complex scenes. In parallel, experiments with higher-precision quantization formats (e.g., 8-bit or 16-bit) and newer LLaVA variants (e.g., LLaVA-Plus, LLaVA-NeXT [38]) will be explored, as these already integrate more advanced vision encoders beyond CLIP. These models may offer improved robustness in occluded, cluttered, or ambiguous urban scenes, while preserving computational efficiency. Proprietary multimodal systems such as Gemini or GPT-4V may also be evaluated in order to assess how commercial vision-language architectures compare to LLaVA in the context of SAGAI. Thus, in future versions, we plan to implement a modular inference backend that will support alternative model weights as well as API-based access to commercial VLMs. This will provide users with flexibility to balance performance, interpretability, and compute constraints, and enable experimentation with new architectures as they become available. While the present study focuses on LLaVA as an accessible and general-purpose VLM, future work should also include comparative evaluation with other model families (e.g. supervised CNN-based pipelines). Such comparisons would help quantify the trade-offs between interpretability, flexibility, and task-specific accuracy across modeling paradigms.

Importantly, the current implementation of SAGAI relies entirely on zero-shot learning: the vision-language model generates structured numerical outputs directly from natural language prompts, without any task-specific fine-tuning. This zero-shot generalization capability is one of the workflow's core strengths, enabling immediate deployment across different cities or scoring schemes, as long as prompts are clearly defined. However, this flexibility comes with limitations, particularly in visually ambiguous or culturally variable environments. To address this, future releases of SAGAI will include an optional few-shot learning module. This module will allow users to provide small labeled datasets—comprising only a few annotated examples—to locally calibrate the model. Such integration will make it possible to bridge generic prompting and task-specific customization, enabling

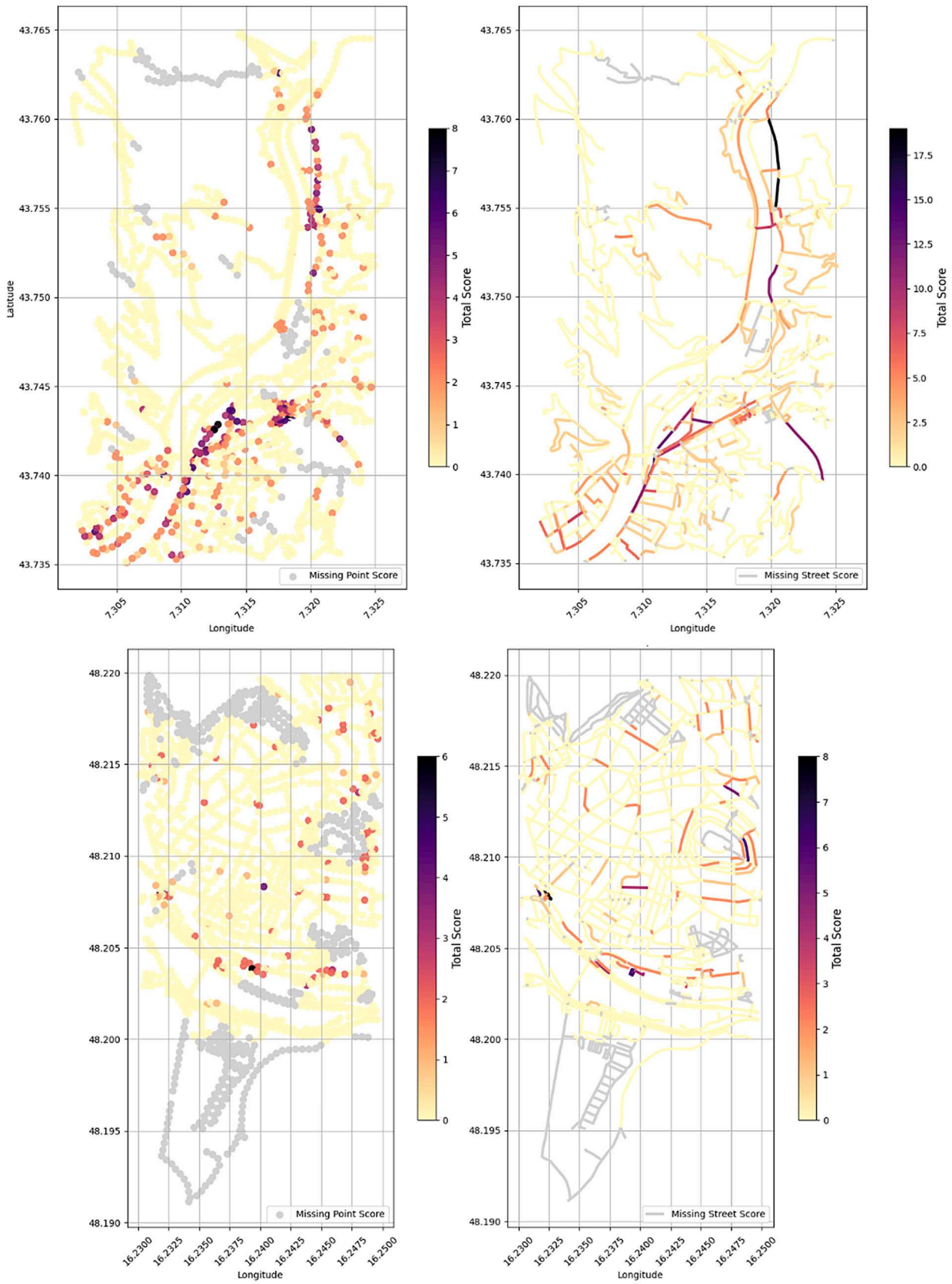


Fig. 5. Total predicted scores for storefront presence at the point level (left subpanels) and aggregated by street segment (right subpanels) from Module 4 – Task 2 (Counting) shown for Nice (top) and Vienna (bottom).

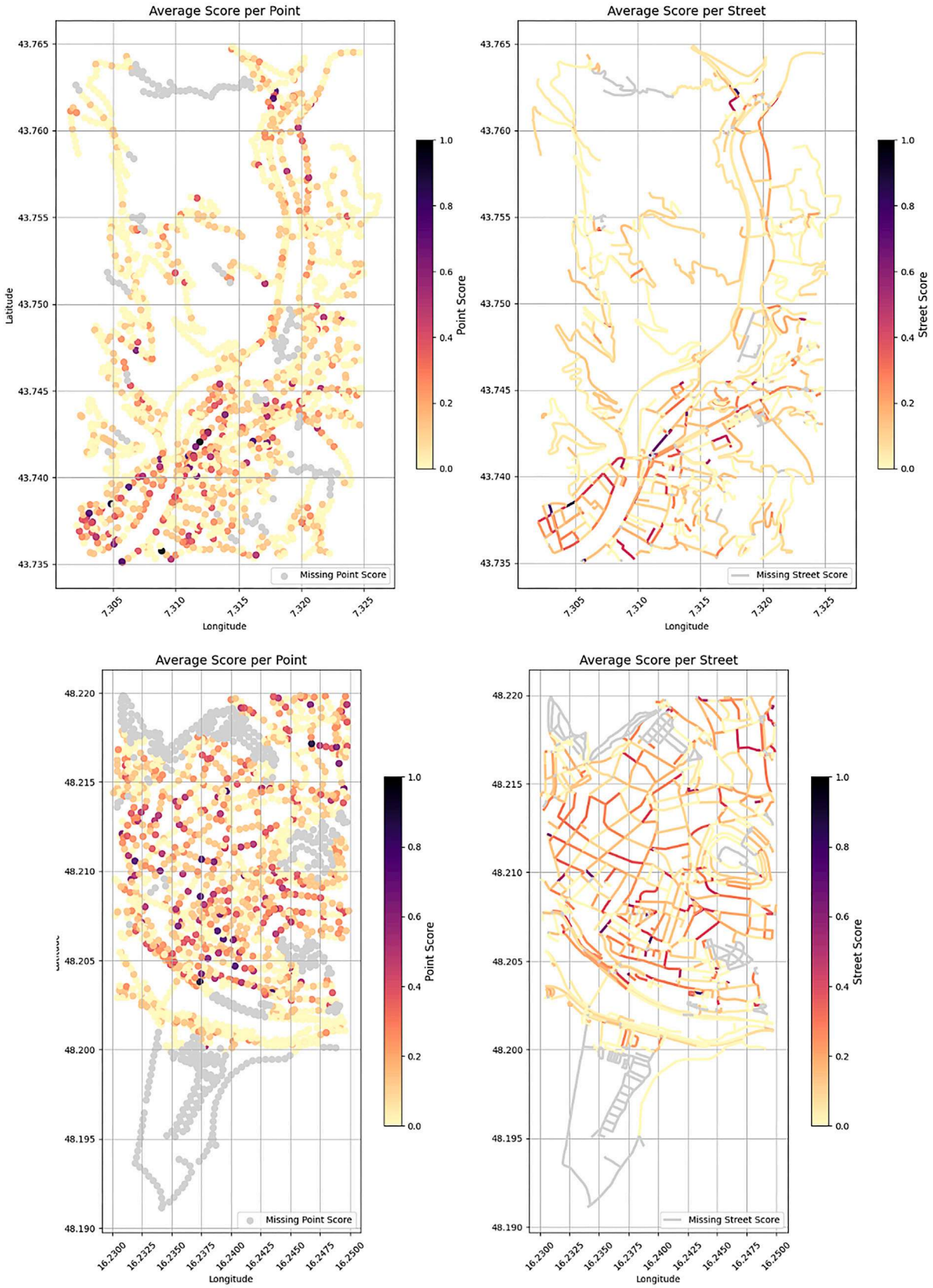


Fig. 6. Averaged predicted scores for sidewalk width at the point level (left subpanels) and aggregated/averaged at the street segment level (right subpanels) from Module 4 – Task 3 (Measuring) shown for Nice (top) and Vienna (bottom).

semi-supervised workflows that improve both precision and reliability in urban visual scoring tasks.

In addition to accuracy, on a standard Google Colab session (free tier), each case study—Nice and Vienna—required approximately 9,000 s (2.5 h) to process the full image set through the LLaVA model. This corresponds to an average throughput of around 1,200–1,300 images per hour, including image loading, preprocessing, and inference. The Street View Batch Downloader (Module 2) required approximately 80 min to retrieve and save all images for a single case study (7,000–8,000 images), depending on connection speed and server response times. While these figures already demonstrate reasonable throughput in a cloud-based environment, further acceleration is possible. Future versions may parallelize image downloads and model inference—such as issuing concurrent API requests or batching multiple images for scoring—which would enable SAGAI to scale more efficiently to larger datasets or multi-temporal applications.

Beyond the vision-language component, several additional enhancements could further strengthen the broader SAGAI pipeline. One practical limitation arises from the potential mismatch between the location of the sample points and the actual position from which the Street View images are captured. This discrepancy may introduce spatial noise, especially in dense or topographically complex environments. Future versions of SAGAI could leverage the metadata embedded in Street View imagery—such as precise camera coordinates or heading information—to reassign sample points to the true image capture location. Moreover, the current implementation relies on the Google Street View Static API, which does not support querying historical imagery. However, integrating alternative platforms such as Mapillary or Google's dynamic API endpoints could eventually enable temporal analyses, allowing researchers to track visual changes in the built environment over time. Finally, instead of capturing four fixed images at orthogonal directions, it may be more efficient to compute the local street orientation and only retrieve views aligned with the primary road axis—i.e., forward and backward along the street segment—thereby reducing redundancy and improving task relevance.

The modular prompt architecture of the current version of SAGAI already allows users to define new tasks by customizing natural language instructions. In addition, the three tasks implemented within SAGAI as examples—classifying, counting, and measuring—can be flexibly combined to support more specific protocols for streetscape assessment, such as identifying patterns related to streetscapes and public space components within pattern languages (e.g., 33,39). More broadly, this opens up the possibility of assessing visually observable urban scenes in relation to normative theoretical models. Thus, through the methodological improvements mentioned above—such as the use of heavier AI models, prompt engineering, few-shot learning, and code parallelization to scale up the spatial extent of urban analysis—SAGAI will open new perspectives for producing automated assessments of visually perceived built environments, integrating fine-grained qualities of the streetscape skin. Diachronic analysis of streetscape transformations over time also offers a promising avenue to assess the impact of underlying urban processes—such as gentrification, pauperization, cultural changes or the effects of specific policies and market trends—since the streetscape skin tends to evolve more rapidly than the skeletal built volumes that support it [40]. In addition to its current focus on street-level imagery, SAGAI's architecture can be extended to other imagery domains and problem settings. Relevant examples include work on crime risk classification using micro-built environment features and object-based building extraction from high-resolution aerial data [41,42].

7. Conclusion

This paper introduced SAGAI—Streetscape Analysis with Generative Artificial Intelligence—a lightweight, reproducible workflow that leverages open-access data and generative vision-language models to automate the scoring of street-level urban environments. Building upon recent advances in VLMs such as LLaVA, the pipeline demonstrates that complex visual tasks—including binary classification, object counting, and dimensional estimation—can now be performed consistently and with reasonable accuracy, without task-specific training or heavy computational infrastructure.

Empirical results across two distinct urban settings, Nice and Vienna, show that the system achieves high accuracy in scene classification (over 90%), moderate precision in storefront detection, and lower—but still informative—performance in estimating sidewalk width. These outcomes illustrate both the promise and the current limits of zero-shot VLMs in urban contexts. Importantly, many misclassifications arose not from outright model errors but from inherent ambiguities in the visual scenes or insufficiently precise prompt formulations. In several cases, the model's outputs were closer to plausible approximations than actual errors, suggesting that further gains may be achieved through prompt refinement rather than architectural overhaul.

Designed to run entirely in Python and deployable in free-tier Google Colab environments, SAGAI offers a reproducible and extensible workflow for researchers and practitioners alike. Its main contribution is to enrich streetscape analysis with micro-scale elements that are typically assessed visually, as they are absent from available geodatabases. This is achieved by leveraging online street imagery through generative AI algorithms. Its modular architecture makes it suitable for a range of urban applications, including walkability audits, streetscape benchmarking, infrastructure planning, and spatial equity assessments. The system can be readily adapted to new case studies, tasks, or scoring schemes simply by adjusting the prompts, underscoring its potential for broad reuse across geographic contexts and research questions. Yet, while SAGAI's zero-shot setup ensures wide applicability, localized fine-tuning may be required in settings with atypical streetscapes or culturally specific features.

The full codebase of SAGAI, including setup instructions and prompt templates, is available in the public GitHub repository described in Section 3.5. Future developments will focus on integrating few-shot tuning modules, supporting alternative imagery sources and vision-language models, and expanding the library of scoring prompts. As such, SAGAI provides both a methodological contribution and a practical foundation for the next generation of geospatial research using generative multimodal AI for streetscape analysis. A forthcoming development direction involves refactoring SAGAI into a Python package with modular architecture and a command-line interface, enabling tighter integration with analytical toolchains and automated processing workflows.

CRediT authorship contribution statement

Joan Perez: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Giovanni Fusco:** Writing – review & editing, Project administration, Methodology, Funding acquisition, Conceptualization.

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.

Acknowledgments

This work was supported by the emc2 project under the Driving Urban Transition Partnership, co-funded by ANR (France), FFG (Austria), MUR (Italy), and Vinnova (Sweden) with contributions from the European Commission.

Appendix A. Prompt templates used in SAGAI v1.0

Task T1—Categorization: Urban vs. Rural

{role_description} = You are an AI assistant designed to analyze street-level images. Your task is to determine whether the environment shown in the image is urban or rural.
{theory_model} = Classification Guide:
- 0: Rural area sparse built environment, natural surroundings, few or no buildings.
- 1: Urban area dense built environment, visible infrastructure, buildings.
{task} = Carefully observe the image and determine whether it depicts a rural or urban environment.
Use the classification guide above to assign a score.
Return only the classification (0 or 1). Do not explain your answer or add extra text.
{response_format} = Answer format: 0 or 1

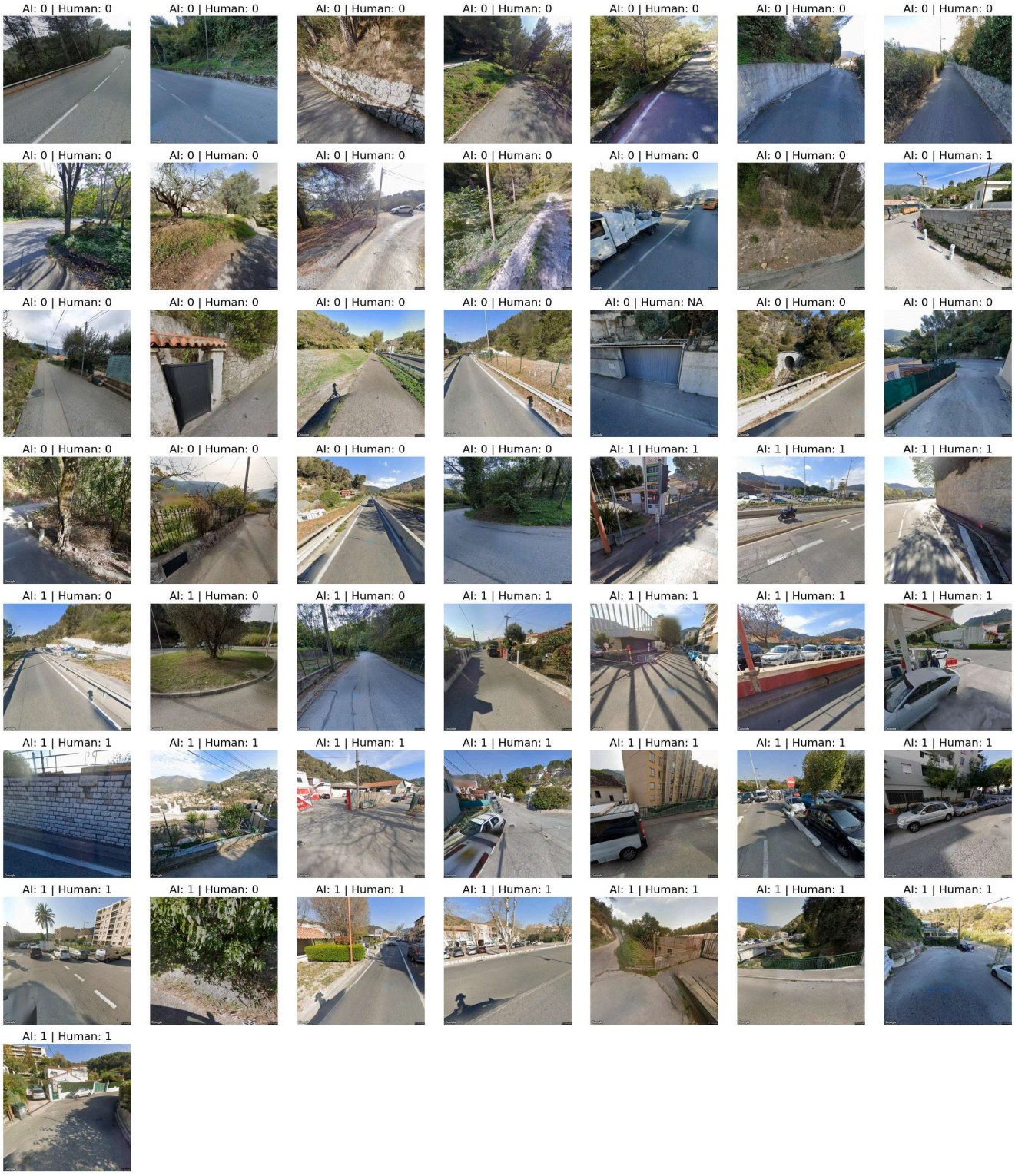
Task T2—Counting: Visible Shopfronts

{role_description} = You are an AI assistant designed to analyze street-level images.
Your job is to detect the presence of commercial storefronts, such as shops, restaurants, or businesses.
{theory_model} = Scoring Guide:
- 0: No visible shops or commercial storefronts.
- 1: One visible shop or storefront.
- 2: More than one shop or storefront is visible.
{task} = Look at the image carefully and apply the scoring guide above.
Return only the score (0, 1, or 2) based on how many shops are visible.
Do not explain your answer or add text. Only output the number.
{response_format} = Answer format: 0, 1, or 2

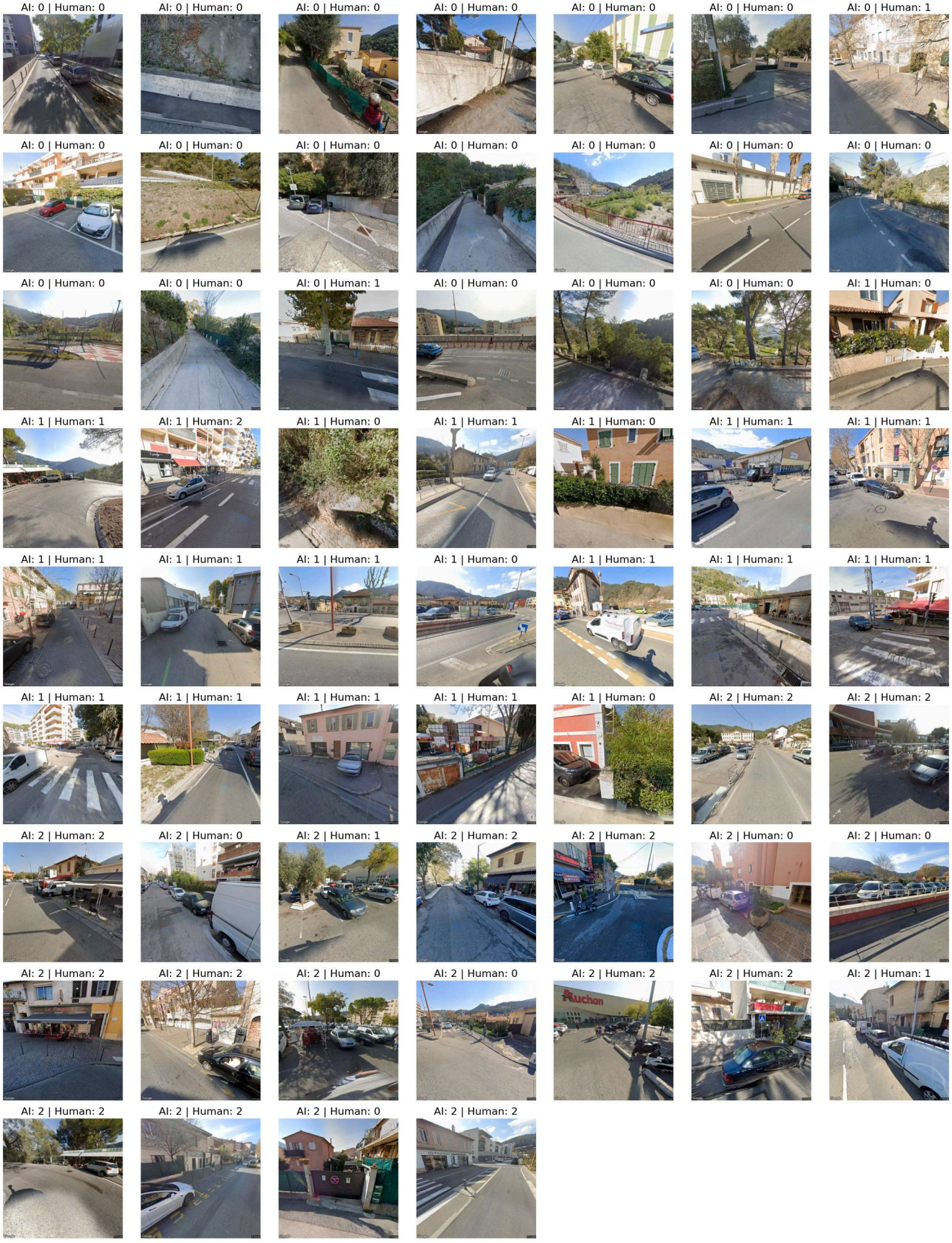
Task T3—Measuring: Estimated Sidewalk Width

{role_description} = You are an AI assistant designed to analyze street-level images. Your task is to estimate the visible width of a sidewalk.
{theory_model} = Scoring Guide:
- 0: No visible sidewalk or the sidewalk is not clearly identifiable.
- Otherwise: Return the estimated width of the sidewalk in meters, rounded to the nearest 0.5 (e.g., 1.0, 1.5, 2.0, 2.5, 3.0).
{task} = Look at the image carefully. If a sidewalk is visible, estimate its width in meters.
If no sidewalk is visible or it is unclear, return 0.
Do not explain your answer or add any text. Only output a single number.
{response_format} = Answer format: 0 or a number (e.g., 1.0, 1.5, 2.0, 2.5, 3.0)

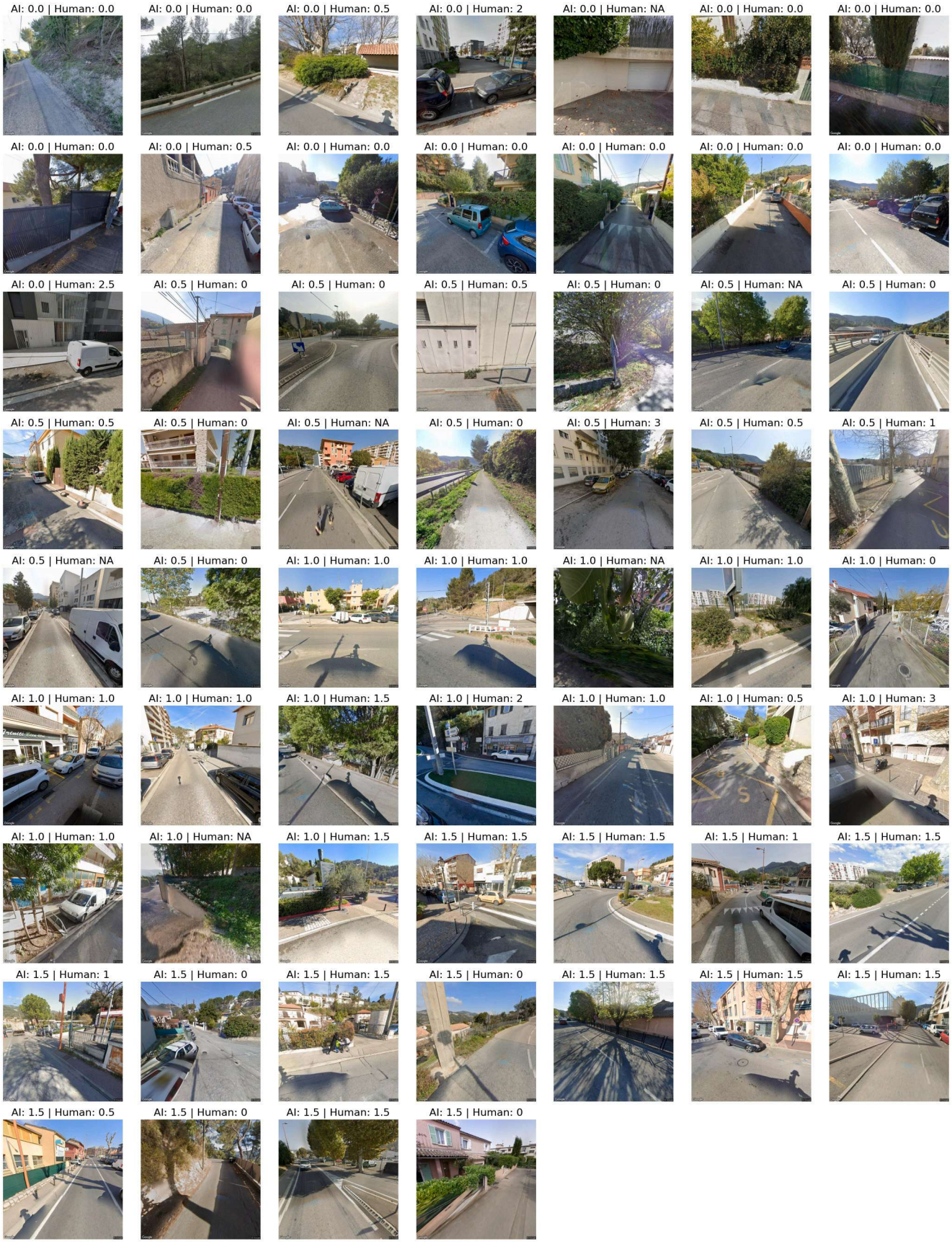
Appendix B. LLaVA predictions and human annotations



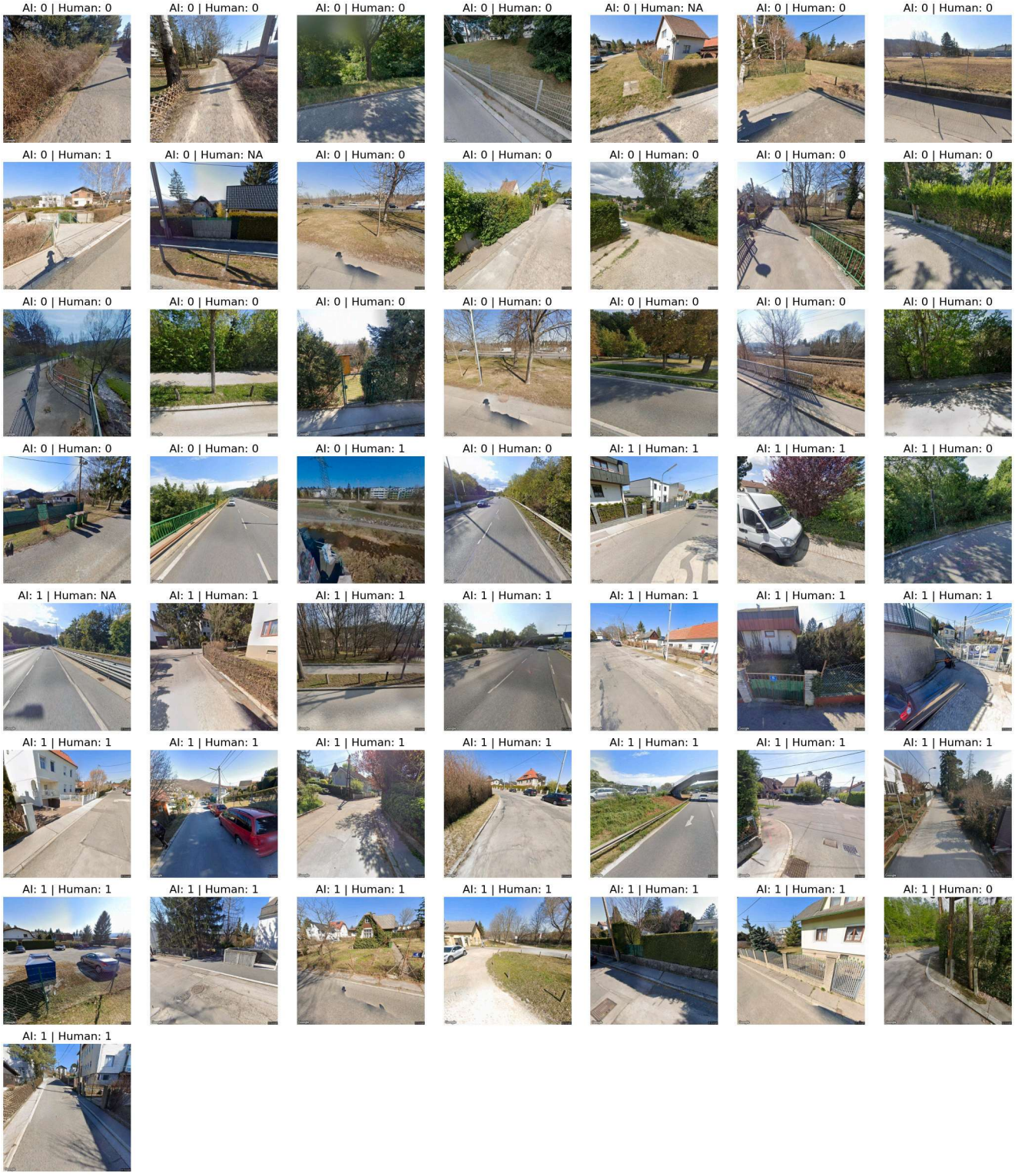
Task T1 – Nice: Comparison of LLaVA predictions and human annotations



Task T2 – Nice: Comparison of LLaVA predictions and human annotations



Task T3 – Nice: Comparison of LLaVA predictions and human annotations



Task T1 – Vienna: Comparison of LLaVA predictions and human annotations

Task T2 – Vienna: Comparison of LLaVA predictions and human annotations

Task T3 – Vienna: Comparison of LLaVA predictions and human annotations

Data availability

Data and code are available at the linked GitHub repository, as described in the paper.

References

[1] J. Gehl, B. Svarre, How to Study Public Life, Island Press, Washington, 2013.
[2] C. Harvey, L. Aultman-Hall, Measuring urban streetscapes for livability: A review of approaches, Prof. Geogr. 68 (1) (2016) 149–158.
[3] K. Dovey, E. Pafka, M. Ristic, Mapping Urbanities: Morphologies, Flows, Possibilities, Routledge, New York, 2018.
[4] C. Harvey, L. Aultman-Hall, A. Troy, S.E. Hurley, Streetscape skeleton measurement and classification, Environ. Plan. B: Urban Anal. City Sci. 44 (4) (2017) 668–692.
[5] A. Araldi, M. Fleischmann, G. Fusco, M. Novotný, Streetscape Morphometrics: Expanding Momepy to Analyze Urban Form from the Street Point of View, Tech. rep., SSRN Preprint, 2025, http://dx.doi.org/10.2139/ssrn.5051213, Available at SSRN.
[6] P. Clarke, J. Ailshire, R. Melendez, M. Bader, J. Morenoff, Using Google Earth to conduct a neighborhood audit: Reliability of a virtual audit instrument, Heal. Place 16 (6) (2011) 1224–1229.
[7] A. Rundle, M. Bader, C. Richards, K. Neckerman, J. Teitler, Using Google street view to audit neighborhood environments, Am. J. Prev. Med. 40 (1) (2011) 94–100.
[8] X. Li, C. Ratti, Mapping the spatial distribution of shade provision of street trees in boston using Google street view panoramas, Urban For. Urban Green. 31 (2018) 109–119.
[9] M. Hosseini, F. Miranda, J. Lin, C.T. Silva, CitySurfaces: City-scale semantic segmentation of sidewalk materials, Sustain. Cities Soc. 79 (2022) 103630.
[10] I. Seiferling, N. Naik, C. Ratti, R. Proulx, Green streets: Quantifying and mapping urban trees with street-level imagery and computer vision, Landsc. Urban Plan. 165 (2017) 93–101.
[11] S. Law, C.I. Seresinhe, Y. Shen, M. Gutierrez-Roig, Street-frontage-net: Urban image classification using deep convolutional neural networks, Int. J. Geogr. Inf. Sci. 34 (4) (2018) 681–707.
[12] T. Gebru, J. Krause, Y. Wang, D. Chen, J. Deng, E.L. Aiden, F.-F. Li, Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States, Proc. Natl. Acad. Sci. (PNAS) 114 (50) (2017) 13108–13113.
[13] F. Zhang, A. Salazar-Miranda, F. Duarte, L. Vale, G. Hack, M. Chen, D.A. Mantaras, F. de Souza, C. Ratti, Urban visual intelligence: Studying cities with artificial intelligence and Street-Level Imagery, Ann. Am. Assoc. Geogr. 114 (4) (2024) 876–897.
[14] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2014, pp. 580–587.
[15] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You Only Look Once: Unified, Real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 779–788.
[16] S. Wang, et al., Mapping the landscape and roadmap of geospatial AI in quantitative human geography, Int. J. Appl. Earth Obs. Geoinf. 128 (2024) 103734.
[17] Y. Song, et al., Advances in GeoAI for mapping, Int. J. Appl. Earth Obs. Geoinf. 120 (2023) 103300.
[18] H. Liu, Y. Zhang, Z. Xu, C. Li, J. Yang, J. Xiong, Y.J. Lee, Visual instruction tuning, 2023, arXiv preprint.
[19] J. Li, et al., BLIP-2: Bootstrapping language-image pre-training, 2023, arXiv preprint.
[20] OpenAI, GPT-4 Technical Report, Technical report, OpenAI, 2023, URL https://openai.com/research/gpt-4 (Technical report).
[21] J.-B. Alayrac, et al., Flamingo: a Visual Language Model for Few-Shot Learning, 2022, arXiv preprint.
[22] A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, I. Sutskever, Learning transferable visual models from natural language supervision, in: Proceedings of the International Conference on Machine Learning, ICML, 2021, pp. 8748–8763.
[23] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, 2020, arXiv preprint.
[24] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin Transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV, 2021, pp. 10012–10022.
[25] Z. Li, et al., StreetViewLLM, 2024, arXiv preprint.
[26] Z. Duan, et al., CityLLaVA, in: Proceedings of the CVPR Workshops, 2024, pp. 7180–7189.
[27] I. Blečić, V. Saiu, G.A. Trunfio, Enhancing urban walkability assessment with multimodal LLMs, in: ICCSA 2024 Workshops, LNCS 14819, Springer, 2024, pp. 394–411.
[28] C. Wei, et al., GeoTool-GPT, Int. J. Geogr. Inf. Sci. 39 (4) (2024) 707–731.
[29] M.K. Beydokhti, et al., Probabilistic qualitative spatial reasoning, Int. J. Geogr. Inf. Sci. 39 (4) (2024) 817–846.
[30] S. Schmidt, et al., Assessing spatial accuracy of geocoding flood-related imagery, Spat. Inf. Res. 33 (15) (2025).
[31] G. Fusco, M. Berghauser Pont, V. Cutini, A. Psenner, Guiding principles for the 15-minute city, in: AESOP 2024 Proceedings, 2024, pp. 690–707, URL https://hal.science/hal-04798781.
[32] G. Fusco, G. Picard, Assessing form patterns for the Suburban 15-Minute City: The case of Drap (France), in: ISUF 2025 Proceedings, 2025, in press.
[33] C. Alexander, S. Ishikawa, M. Silverstein, A Pattern Language, Oxford University Press, New York, 1977.
[34] J. Gehl, Cities for People, Island Press, Washington, 2010.
[35] I. Yamada, J.-C. Thill, Local indicators of network-constrained clusters, Ann. Am. Assoc. Geogr. 100 (2) (2010) 269–285.
[36] H. Liu, C. Li, Q. Wu, Y.J. Lee, LLaVA: Large language and Vision Assistant, 2023, arXiv preprint.
[37] A.Q. Jiang, A. Sablayrolles, A. Roux, A. Mensch, B. Savary, C. Bamford, D. Singh Chaplot, D. de las Casas, E.B. Hanna, F. Bressand, G. Lengyel, G. Bour, G. Lample, L.R. Lavaud, L. Saulnier, M.-A. Lachaux, P. Stock, S. Subramanian, S. Yang, S. Antoniak, T.L. Scao, T. Gervet, T. Lavril, T. Wang, T. Lacroix, W. El Sayed, Mixtral of experts, 2024, arXiv preprint.
[38] S. Liu, H. Cheng, H. Liu, H. Zhang, F. Li, T. Ren, X. Zou, J. Yang, H. Su, J. Zhu, L. Zhang, J. Gao, LLaVA-plus: Learning to use tools for creating multimodal agents, 2023, arXiv preprint.
[39] M. Mehaffy, Y. Kryasheva, A. Rudd, N. Salingaros, A New Pattern Language for Growing Regions: Places, Networks, Processes, Sustasis Press, Portland, OR, 2020.
[40] K. Dovey, S. Wood, Public/private urban interfaces: Type, adaptation, assemblage, J. Urban. 8 (1) (2015) 1–16.
[41] D.Y. Wu, J. Wang, Analyzing urban crime through Street View imagery: Insights from urban micro built environment and perceptions, Urban Sci. 8 (4) (2024) 247, http://dx.doi.org/10.3390/urbansci8040247.
[42] M.D. Hossain, D. Chen, Target-based building extraction from high-resolution RGB imagery using GEOBIA framework and tabular deep learning model, Geo-Spatial Inf. Sci. 100007 (2024) http://dx.doi.org/10.1016/j.geomat.2024.100007.