ArtScape: Detecting, Characterizing, and Interactively Exploring Urban Art at Scale Using Vision Language Models

Lucy Jiang*
Human Centered Design and Engineering
University of Washington
Seattle, Washington, USA
lucjia@uw.edu

Jared Hwang*
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, Washington, USA
jaredhwa@cs.washington.edu

Jon E. Froehlich
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, Washington, USA
jonf@cs.uw.edu

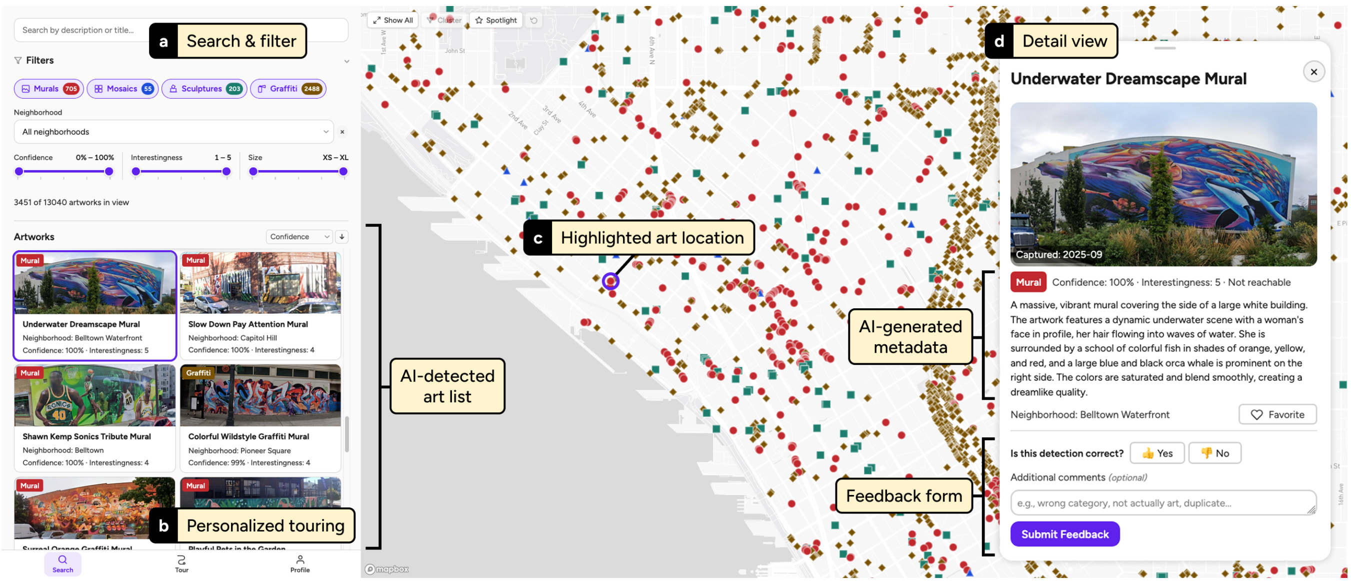

Figure 1: We introduce *ArtScape*, an automated, large-scale artwork cataloging and exploration tool that detects, locates, and describes urban artwork in four categories, then interactively presents it. Users can (a) search and filter, (b) generate personalized tours, (c) see artwork across entire regions, and (d) engage with AI-generated metadata such as descriptions, titles, and alt text, and submit per-artwork feedback. Refer to video demo for further details.

Abstract

Urban art brings cultural expression, community identity, and aesthetic enrichment to cities. Currently, there are no systematic ways of exploring urban art at scale. We introduce *ArtScape*, a VLM-based interactive tool that lets users view, search, filter, and tour four categories of art (*murals*, *mosaics*, *sculptures*, *graffiti*). To comprehensively and automatically detect urban artwork throughout a region, we designed a custom five-stage pipeline that uses VLMs to detect and describe urban artwork within 360° panoramic Street View images, achieving a micro-F1 of 0.83 on an urban art dataset we labeled. ArtScape categorizes, describes, and scores the reachability and interestingness of each artwork while deduplicating based on proximity and similarity. We evaluated this tool with 12 participants spanning urban planners, artists, sighted users, and blind and low vision users. Our findings show that ArtScape's at-scale AI artwork curation enables granular and aggregate urban art discovery, thereby helping people learn about their communities and city-wide trends, motivating in-person urban exploration, and expanding perceptions of what constitutes urban art.

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**; **Accessibility systems and tools**.

Keywords

urban art, vision language model, urban informatics, accessibility

ACM Reference Format:
Lucy Jiang, Jared Hwang, and Jon E. Froehlich. 2026. ArtScape: Detecting, Characterizing, and Interactively Exploring Urban Art at Scale Using Vision Language Models. In *The 39th Annual ACM Symposium on User Interface Software and Technology (UIST '26), November 02–05, 2026, Detroit, MI, USA.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3830398.3830557

*Both authors contributed equally to this research.

UIST '26, Detroit, MI, USA

ACM ISBN 979-8-4007-2856-3/26/11
https://doi.org/10.1145/3830398.3830557

1 Introduction

Urban artwork is an enriching feature of cities, providing social, political, and aesthetic value and fostering community identity. Despite the ubiquity of urban art, its discoverability is bottlenecked by real-world interaction—there is currently no large-scale urban artwork dataset with accessible artwork information. What *does* exist is limited to manually-collected catalogs, which are labor-intensive to create, inaccessible, and biased towards "sanctioned" art forms, ignoring graffiti and other ephemeral work. This discoverability problem is particularly acute for blind and low vision (BLV) people, given the widespread inaccessibility of visual art both in-person and online [46, 54].

Prior work has yet to catalog urban artwork at scale due to challenges with defining, and therefore detecting, urban art. Urban artwork lacks a uniform definition—some scholars emphasize social engagement [51] and environmental integration [39], while existing databases like the Public Art Archive [82] enumerate urban artwork categories such as murals, mosaics, and performance art but leave out graffiti [65]. On the technical side, existing work in computer vision has identified and described a variety of features such as signage [55], bus stops [49], and building facades [56] in urban streetscapes by training/fine-tuning models or using zero-shot models such as Vision Language Models (VLM). However, these benefit from a *constrained* search space with existing datasets or concrete rubrics. Prior work on detecting urban art is largely limited to graffiti, relying on pre-existing, manually curated datasets [22, 52]. Collectively, the disparate definitions and absence of *any* urban artwork datasets means there is currently no way to comprehensively detect, curate, and catalog urban artwork with prior methods.

We introduce *ArtScape*: the first automated, large-scale art exploration tool that **detects, locates, and describes urban artwork** in four categories (*murals*, *mosaics*, *sculptures*, *graffiti*) using Google Street View (GSV), and **interactively presents** these detections via an accessible interface that augments online and real-world art exploration. The automated pipeline identifies urban artwork in GSV images using criteria we synthesized from prior work [46, 51, 65, 73]. We pass each GSV panorama through a five-stage pipeline (Figure 2) to adhere to these criteria and produce rich descriptive information: (1) image preprocessing to reduce context-window and token burden; (2-3) two VLM passes with concrete rubrics to exhaustively detect, characterize, and describe artworks; (4) geolocation using DA360 [44], a monocular depth estimation model; and (5) artwork deduplication using DINOv3 [80] embeddings. We then designed an accessible tool for interactively and meaningfully exploring these thousands of artworks. ArtScape presents artwork in both list and map modalities, supports filtration by type, size, confidence, interestingness, description, and neighborhood, and includes a built-in tour functionality that highlights art on a user-specified path. Through our system, we emulate the spontaneous discovery of artwork in the real-world while also achieving what in-person exploration cannot: city-wide pattern analysis, neighborhood-level trend understanding, and cross-artwork comparison at scale.

To evaluate ArtScape, we collected and hand-labeled a Street View Image (SVI) dataset of 704 images, and achieve a micro-F1 of 0.83 to detect art in our test set. We find that ArtScape's AI-generated descriptions are highly *descriptive* and *accurate*, via manual scoring of 204 randomly sampled art pieces by two researchers. We also conducted a user study with 12 participants spanning artists, urban planners, and sighted and BLV individuals. Our findings show that ArtScape unlocks multi-level insights and complements in-person exploration, helping users appreciate known works and discover unfamiliar ones. AI curation also provided unique opportunities by lowering the barrier for participants to actively and critically interact with artwork, suggesting that transparent AI usage can invite broader participation in cultural curation.

In summary, we contribute (1) a **new algorithmic pipeline** that detects, locates, and characterizes size-, shape-, and style-agnostic urban artwork at scale, (2) an **interactive, accessible urban artwork exploration and engagement tool** that enables city-wide trend understanding and discovery, (3) **empirical findings** demonstrating how AI-curated urban art enables multi-level discovery and deepens critical engagement, and (4) a **benchmark dataset** of four categories of urban artwork in 704 Street View images. To support open science and enable others to build upon our work, we release both our annotated dataset[1] and all code.[2]

2 Related Work

We first provide background on urban artwork typologies, then situate our work in research on urban feature detection, artwork description, and urban art exploration tools.

2.1 Urban Art Typology

There is no single definition or typology of urban artwork. While scholars emphasize attributes such as social engagement [51], non-commerciality [11], site-specificity [51], and environmental integration [39], for this work, we adopt a broader definition: urban art encompasses any outdoor and publicly accessible artwork, commissioned or not. Public art, a subset of urban art, refers specifically to publicly funded artwork commissioned via a public process [73, 93]. Existing public art repositories (*e.g.*, the Public Art Archive [82], Project for Public Spaces [73]) enumerate categories for murals, mosaics, sculptures, performance art, and other art types distinguished by material or topic (*e.g.*, fiber art, environmental art). Our typology builds on Jiang *et al.*'s urban artwork taxonomy consisting of murals (2D, on flat surfaces), mosaics (2.5D, also on flat surfaces), and sculptures (3D, typically freestanding or on pedestals) [45, 46]. We additionally surface uncommissioned graffiti artwork (spray painted tags or lettering) as a fourth category of our typology due to its ubiquity and cultural meaning in urban spaces [10, 65].

2.2 Detecting Urban Features with AI

ArtScape's VLM pipeline is unique in that it centers on detecting *scale*-, *shape*-, *style*-, and *location*-variable artwork in streetscape imagery. While there is a large body of existing work in streetscape object detection and characterization, it largely focuses on either ranking the holistic features of a street (*e.g.*, roads [42], foliage [25], walkability [19]), or detecting objects with already known rough positioning (*e.g.*, curb ramps [70], building facades [56, 58]) or

[1] https://huggingface.co/datasets/makeabilitylab/urbanartwork
[2] https://github.com/makeabilitylab/artscape

constrained shapes (*e.g.*, traffic signage [55], sidewalks [75], bus stops [49], parking [43]). For example, Liang *et al.* train an R-CNN with a traffic sign dataset to detect traffic signage in images [55]. Without model training, Cai *et al.* use VLMs to look for the existence of specific features in streetscape imagery to score *walkability* [19]. In these cases, through either a pre-existing dataset of objects, *or* a pre-defined, constrained search space, existing work has accurately detected or described objects in streetscape imagery. However, urban artwork has *no existing* dataset of artwork in its varying types, shapes, and styles—a dataset that this paper contributes.

In cases where existing work has detected artwork in streetscapes, it primarily only focuses on a single *subtype* of artwork, thereby constraining the detection space. One popular area of interest is graffiti detection, aimed not at surfacing cultural relevance or artistic value but automated vandalism detection [7, 15, 22, 35, 52, 69, 84, 85, 88]. Other work has examined murals [7, 35, 89] and sculptures [41, 76]. For example, Bomfim *et al.* [15], Wang *et al.* [88], and Lavi *et al.* [52] all train detection models on custom curated graffiti datasets (*e.g.*, [22]) or web-scraped graffiti datasets (*e.g.*, [52]), and Alzate *et al.* do the same for murals [7]. However, these works use convolutional neural networks (*e.g.*, [15, 41]), or transformer based models (*e.g.*, [88]) that all require prior training or fine-tuning. ArtScape is the first to utilize a state-of-the-art VLM to detect multiple categories of urban art of varying artistic styles and shapes *simultaneously*, throughout a region and without prior training—while also describing, rating, and locating each art piece.

2.3 Describing Artwork with AI

A key feature of ArtScape is automatically generating titles, descriptions, and alt text for detected artwork, drawing on prior work [8, 21, 23, 29, 33, 46, 90]. While a variety of models have been used for describing artwork, (*e.g.*, CLIP [21], BERT [29]), recent VLM/LLM advances have significantly improved art description capabilities [23, 90]. For example, Chheda-Kothary *et al.* use LLMs to generate descriptions of children's artwork for mixed-ability families [23], Jiang *et al.* explore BLV users' perspectives on using LLMs for generating urban art descriptions [46], and Arita *et al.* investigate whether LLM generated descriptions can be distinguished from human-written descriptions [8]. Inspired by these works, ArtScape uses a VLM to generate descriptions of each piece's contents and color, revealing potentially subtle details while supporting accessibility for BLV users.

However, Jiang *et al.* found that participants often wanted more cultural information in descriptions [46]. Some work addresses this on traditional artwork (*e.g.*, Renaissance paintings) via retrieval-augmented generation (RAG), augmenting the LLM's knowledge with external sources [33, 90]—for instance, providing additional context (*e.g.*, artist) and querying a pre-constructed knowledge graph [90]. However, in these cases, the input is assumed to be *known* artwork with a clean image for inference. As ArtScape operates on urban art that is more transient and may lack documented history or a known artist, we consider RAG-supported description generation out of scope but a promising direction for future work.

2.4 Urban Art Exploration Tools

The increasing popularity and ubiquity of urban artwork has given rise to urban artwork tourism, also known as "mural tourism" [20, 71, 81]. In the U.S. alone, cities such as Detroit [72, 87] and Seattle [12, 77] offer guided public art tours. In recognition of this interest in urban art, researchers have leveraged augmented reality to enhance in-person street art engagement by overlaying artwork meaning, history, and artist information on murals [36] and providing animations and interactive mini-games [13]. Databases such as the Public Art Archive [82] and Google Arts & Culture's Street Art collection [38] also document and map out urban art. However, prior work has been limited to a small set of manually curated artworks and does not support participant-led exploration; with ArtScape, we combine automated, at-scale detection of urban artwork with interactive, personalized exploration and touring capabilities.

3 The ArtScape Pipeline

Grounded in prior literature, we built and evaluated the ArtScape prototype, the first accessible interactive urban art exploration tool.

3.1 Pipeline Implementation

To comprehensively catalog scale-, shape-, style-, and location-variant urban artwork across a wide region, we introduce a novel, custom VLM pipeline that detects, geolocates, and characterizes art in widely-available Street View imagery. State-of-the-art VLMs are remarkable at making zero- or few-shot inferences to identify features within images without domain-specific fine-tuning [74], making them perfectly suited for our task of detecting artwork with no prior training set. However, they are not omniscient—and unlike objects such as trees, sidewalks, or signage, there is no uniform understanding of what *exactly* is urban artwork. Grounded in prior literature [46], we target four categories of urban artwork: *murals*, *mosaics*, *sculptures*, and *graffiti*. While some prior work [45, 46] does not include graffiti, we consider it as a distinct artwork category, recognizing its cultural and political value [65], its association with socioeconomic trends [52, 66], and its prominence within prior HCI research for detection [52, 69, 88].

The key challenge in applying VLMs to the task of exhaustive art detection is (1) communicating in a prompt the line between "artwork" and "urban feature" that we intuitively understand as humans and synthesize from prior work, and (2) filtering out extraneous detections that do not conform to the guidelines we enumerate. Our pipeline consists of five stages to maximize performance: image pre-processing, exhaustive artwork detection using a VLM, detection filtration and characterization using a second VLM pass, detection enrichment via geolocation using monocular depth estimation, then finally artwork deduplication. This pipeline forms the backbone of ArtScape's ability to enable city-scale artwork exploration and regional trend understanding. Artificial Intelligence was used to assist with implementation and we carefully reviewed all code. See Figure 2 for the pipeline components, which we describe below.

3.1.1 Street View Images: Sampling a Region. We use Google Street View (GSV) [5] for its highly comprehensive coverage of cities from a *ground-level* view, allowing us to traverse streets and search for urban artwork akin to in-person exploration. A critical decision,

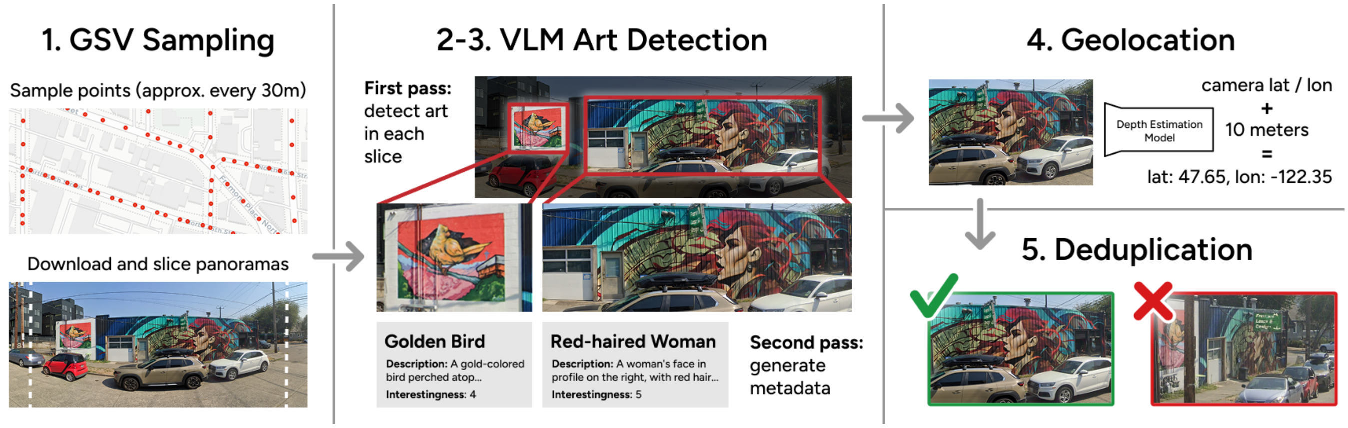

Figure 2: ArtScape is comprised of a five-stage pipeline: (1) sampling, downloading, and preprocessing panoramas from Google Street View; (2) detecting urban artwork in the panoramas with a VLM; (3) filtering false detections and characterizing each artwork with descriptions, titles, and interestingness, with a second VLM pass *per artwork*; (4) geolocating artwork to a position in the real world using DA360 [44]; and (5) deduplicating instances of the same artwork, prioritizing better viewing angles.

however, lies in choosing which images to fetch from GSV. Using OpenStreetMap's [3] road network, we sample points at 30 meter intervals along each network edge (and the midpoint for those shorter than 30m) to balance between having complete line-of-sight to all points on a given street and resource consumption. We return to this tradeoff in Section 6.1. This produces 51,251 street edge sample points in Seattle. Then, for each sample point, we query and download the closest available SVI within 10 meters using the GSV API [5], resulting in 49,859 total unique images in Seattle which we pass to our artwork detection pipeline.

3.1.2 VLM Passes. To effectively communicate our conception of urban artwork to a VLM for object detection and description, we must address two key challenges: (1) panoramic images are both large in filesize and cover an expansive area, increasing the burden in tokens and context for a single VLM call, and (2) by targeting exhaustive detection, the VLM may both misinterpret and overzealously detect objects that do not belong in our categorizations, exacerbated further by limited thinking budget as a result of (1). As such, we take three steps to augment and refine the VLM output: image preprocessing to maximize tokens-per-image, a first pass over entire panoramic images targeting *exhaustive* artwork detection, and a second pass to *filter* extraneous output from the first pass while also generating high-quality descriptions. For all VLM calls, we use Gemini 3 Pro [4] for its state-of-the-art object identification and description capabilities, as well as its ability to generate bounding boxes around objects without additional model head or output layers. We describe each step in detail below.

Preprocessing. In informal experiments, we found that when passed the entire 360° panorama (6,656 pixels × 3,328 pixels), the VLM both inadequately adhered to our guidelines and failed to consistently detect small or occluded objects. We hypothesized that the primary constraint was the limited context window and thinking budget needed for a large image. As downsampling panoramas may lead to small artworks or details becoming indistinguishable, we opt to instead split the panoramas into six slices and call the VLM on each slice individually. In this way, the VLM operates on a much smaller image, improving detection quality. Each slice is 120°, sharing overlap with neighboring slices, to circumvent having to recombine objects split at seams. For each slice, we only accept detections whose centroid exists in the center 60°, ensuring objects are detected in their entirety. Note that our preprocessing strategy lowers tokens spent *per API call*, of which there are six per panorama. It does not lower tokens spent per panorama.

First VLM Pass. We use the first VLM pass as an *exhaustive* detection pass—passing each panorama slice to the VLM and asking it to find all instances of art in the image while also categorizing and describing them. This task is not trivial; as artwork is inherently subjective, we include several concrete rubrics and guidelines to indicate to the VLM what art is of interest. For the full prompt, see Section A.2. In summary, we provide category descriptions and expectations to set the stage (*e.g.*, *"**Mural**: Paint on walls, ground, infrastructure, Pictorial/large-scale designs."*) Second, we describe inclusion and exclusion locations, targeting only permanently installed, non-transient artwork: *"**Include:** art on utility boxes, dumpsters..."*, *"**Do not include:** architecture, advertisements..."* Third, we include a checklist of questions, where any "yes" response would disqualify a piece of art: *"Is this a **road or ground marking** (crosswalks, utility markers...)?"* Fourth, we highlight easily missed locations: *"Utility infrastructure, ground surfaces and overhead structures..."* Last, we provide expectations for each descriptive field (*i.e.*, confidence, description, alt text, reachability): *"include **Subject matter:** what is depicted, **Colors & palette**, **Style & technique**..."* Through our prompt, we specify precise guidelines for the VLM to follow, translating *subjective* ideas about artwork and creativity into grounded and specific features.

Second VLM Pass. The first pass alone produces a large number of detections that do not adhere to our guidelines (*i.e.*, false positives). Thus, a second VLM pass runs only on individual cropped detections, targeting (1) filtering out false positives and (2) auditing the first pass by fixing misclassifications and generating detailed descriptions, a title, and an "interestingness" score. In prompting, we take a similar approach as before and reiterate the same guidelines: first by including art categories and definitions, second by including the same checklist (*e.g.*, *"Is this a road/ground marking?"*), and third by explicitly asking if the artwork appears to be commercial. In addition to verifying that a detection is artwork, we ask the VLM to audit the description generated by the first pass: *"Does the* ***description*** *accurately capture subject, colors, style material... condition? Does the* ***alt text*** *concisely summarize the artwork?"* We also ask for an "interestingness" score: *"Rate how likely this piece would be to appear on a neighborhood urban art walking tour: 1: Barely notable... 3: a respectable piece a tour guide would point out in passing... 5: a landmark piece people would seek out specifically."* Lastly, to support accessibility for BLV people who may wish to visit in-person, we generate a reachability rating, inspired by prior work on tactile access for artwork (*e.g.*, [17, 24, 46]). See Section A.3 for the full prompt.

3.1.3 Distance Filtering and Geolocation. The VLM pipeline yields artwork detections in an SVI and its approximate (camera) location in the real-world. However, we do not know *where* on the street it is precisely located, nor if our view of the artwork is ideal. We encounter three potential issues: (1) as we prompt the VLM for exhaustive detection, there may be many detections that are far away, small, and with poor viewing angles; (2) by detecting all objects that the camera can see, there will be significant duplicate detections from SVIs that are nearby; and (3) we have no fine-grained understanding of the real-world location of each artwork beyond camera position. To address all three issues, we estimate the distance of each artwork's centroid from the camera and filter out artwork farther than 25 meters away. We choose 25m as it is just under our 30m sampling interval, allowing for some error buffer in our depth estimation.

To estimate distance, we use *DA360* [44], a state-of-the-art monocular depth estimation model that accepts panoramic images directly and predicts scale-invariant depth. Because DA360 does not provide absolute depth (*i.e.*, meters), rather, depth scaled by a per-image scaling factor, *we* must determine each SVI's scaling factor. Unfortunately, there is no uniform reference point in SVIs to derive this scaling factor. However, as all SVIs are taken in a similar environment (*i.e.*, outdoors streetscapes), we assume this scale will be relatively similar across all SVIs. We randomly sampled 20 Seattle SVIs and used objects with known lengths (*e.g.*, cars, signage, cones) to calculate the scaling factor for each image, averaging to 2.54. We apply this factor uniformly across all SVIs for depth estimation. Finally, with the estimated real-world distance of each artwork, we derive the latitude and longitude of each object from the heading of the image, position of the object in the panorama, and distance from the camera.

3.1.4 Deduplication. To further ensure that our detections are of high quality and at good viewing angles, we run a deduplication step to filter out detections of the same object and choose the one that occupies the largest FOV in its panorama (*i.e.*, was closest to the camera). This has the added benefit of removing duplicate objects due to variance in the distance filtering or images not being precisely 30m apart. To do so, for each detection crop, we generate embedding vectors using *DINOv3* (ViT-B/16) [80], a state-of-the-art image understanding model. Then, to determine if two detections

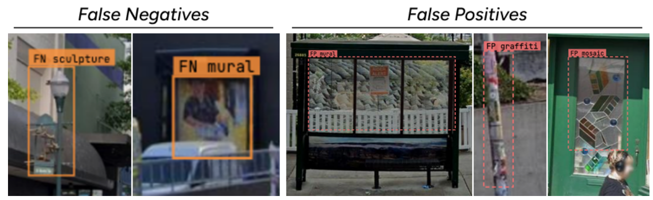

Figure 3: Common false negatives include distant and off-angle pieces. Common false positives include objects we do not consider urban artwork (Section 5.2.3), *e.g.*, bus stop decoration, posters, or store-front decoration.

are the same, we take the cosine similarity of their embedding vectors: if above 0.88, they are determined to be the same, if below 0.7, they are determined to be different. Between 0.7 and 0.88, we use *SuperPoint* [27] to extract key features and *LightGlue* [57] to match feature points as the deciding factor: if the two detections have greater than 30 matching points, they are considered duplicate. All thresholds were determined experimentally. To compare all detections across Seattle, we run a grid search, whereby we assume that our geolocation is accurate within 50m. We segment the map into a 50m grid and compare each artwork to every artwork in its grid cell and the eight neighboring cells. Sets are constructed of matching objects, and all but the largest of the set is discarded.

3.2 Evaluation

We evaluate our VLM pipeline, examining our ability to (1) detect objects in panoramic Street View images compared to a hand-labeled image dataset and (2) generate appropriate and accurate descriptions according to two human judges.

3.2.1 Dataset. To generate the first-ever urban artwork training dataset from streetscape images, we created and hand-labeled a 704 image dataset composed of GSV images from artwork-dense regions in six cities: Seattle, San Francisco, San Diego, New York, Philadelphia, and Detroit. One researcher manually labeled every instance of urban art in four categories (*murals* = 581, *mosaics* = 73, *sculptures* = 157, and *graffiti* = 494, total = 1,305) using polygonal masks for larger objects and rectangular bounding boxes for smaller objects (*e.g.*, small graffiti tags). For ArtScape, we split the dataset into 50/50 validation/test splits, using the validation set for iterating on our custom prompts and the test split for final evaluation. This is the *first* dataset that captures a wide variety of urban art from a street-level viewpoint; we release it for open science.

3.2.2 Object Detection. To evaluate our VLM pipeline's artwork detection ability, we first run our pipeline on the test set (*# images* = 352, *# objects* = 673) without deduplication, which is only relevant when cataloging a region. We match detected objects to ground truth using a fuzzy approach, as we prioritize whether the artwork will be captured in the final detection list (*e.g.*, if there is a graffiti piece on top of a mural) over boundary accuracy. First, we use the *Hungarian* method for bipartite matching [48], with an *Intersection Over Union* (IoU) threshold exceeding 0.5. Matched objects are removed from both ground truth and predicted pools. Second, if any remaining ground truth objects fully envelop any predicted labels (or vice versa), they are considered matched and removed from the pool. Third, if any remaining detections are within 300 pixels of a ground truth detection, they are matched (as Gemini's bounding box detection can be offset by some amount). Therefore, if a ground truth object was matched, it was captured in the detections, though it may not have its own individual label.

For all aggregated performance statistics (precision, recall, and F1), we use the micro-average to capture holistic performance of the model across classes. Misclassifications are considered true positives as we are primarily concerned with detection rate. We encourage readers to observe the per-class statistics in Table 1. Overall, our pipeline performs well, with a precision of 0.77, recall of 0.88, and F1 of 0.83. We consider **recall** to be the most valuable metric (and we optimized our performance for it), as it is easier to correct a false positive than find a false negative. See Figure 3 for example false negatives and positives, and Figure 5 and Section A.1 for successful detections.

3.2.3 Distance Scaling Factor. To evaluate potential variance introduced by assuming a uniform scaling factor for estimating distance, we take the average and standard deviation of our 20 derived scaling factors. We find the average is 2.54 (used in the pipeline), with a standard deviation of 0.28 (11% of the mean). Assuming the distribution holds over all Seattle panoramas, this means that a majority of distances we estimate are accurate to within ~11%. As artwork heading is more critical for geolocating on the map, this error in distance estimation is acceptable for our purposes.

3.2.4 Deduplication. To evaluate the deduplication error rate, we randomly sampled 100 matched objects. We found that 27% were incorrect matches, 24/27 of which were graffiti. This indicates that deduplication works well for distinct art pieces but poorly for more simple graffiti pieces, suggesting that DINO's pretraining is incapable of distinguishing between similarly looking graffiti artwork. Ultimately, this means ArtScape will rarely falsely deduplicate murals, mosaics, and sculptures, but may spuriously remove graffiti.

3.2.5 Description Generation. We evaluate our pipeline's ability to generate both *descriptive* and *accurate* descriptions of each art piece to reaffirm existing work [8, 46] on VLMs' ability to describe artwork. We randomly sampled 204 detections and had two researchers, one with a high level (R1) and one with a low level (R2) of image description writing experience, score each detection's description on a Likert-type scale from 1-5 on descriptiveness and accuracy (matching the user study protocol in Section 4.1). On artworks where both scorers agreed that the detection was artwork

Table 1: Per-class results of our object detection pipeline. Note *recall* is the highest amongst most categories—our goal.

Category	TP	FN	FP	Precision	Recall	F1
Mural	126	14	41	0.75	0.90	0.82
Mosaic	21	0	20	0.51	1.00	0.68
Sculpture	19	6	5	0.79	0.76	0.78
Graffiti	191	27	38	0.83	0.88	0.85
Total	357	47	104	0.77	**0.88**	0.83

Figure 4: Edge cases we excluded from our definition of urban artwork, including exclusively commercial artwork, architecture, and urban furniture (*e.g.*, ground tiling, lamps, bollards).

(187/204), *both* R1 & R2's median scores were 4 for descriptiveness and 5 for accuracy (for breakdown by class, see Section A.4). This indicates that our pipeline produced highly descriptive and accurate artwork descriptions, affirming prior work and verifying the quality of our art catalog.

3.2.6 VLM Cost-Reduction Strategy. To reduce total VLM API cost and time burden, we experimented with a version of our pipeline where we pass slices to the VLM in one batched call, rather than independently. This reduced API cost and time significantly, but recall suffered (0.62). Given the savings, we use the *batched* VLM pipeline for the version of ArtScape used in the user study, as full comprehensive data is not required for understanding participant reactions and interactions in a proof of concept. We ran this pipeline over 49,859 SVIs, which cost ~US$2,000. See Section A.5 for results of the batched object detection evaluation, run with the same methodology as described above.

3.3 User Interface Design and Implementation

ArtScape enables users to interactively explore urban artwork at scale. The primary UI (Figure 1) consists of four key components: (1) a **search panel** for searching, filtering, and viewing the artwork in cards as part of a structured list, (2) a **tour panel** for generating routes that pass by artworks of interest, (3) a **large map view** with all urban artwork detections marked at their approximate location, and (4) **detail modals** that pop up once a user clicks on an artwork card. We provide key details and rationale below. See video figure for additional details.

3.3.1 Exploring Urban Artwork Online. To help users make sense of thousands of artworks, we follow Shneiderman's information-seeking mantra of *"overview first, zoom and filter, then details on demand"* [79] via the map view, the search panel, and the detail modals. This pattern is used widely by commercial geovisualization systems (e.g., Google Maps [37], Zillow [94], Airbnb [6]) and artwork-related research (e.g., [31, 50, 60, 68, 92]).

Unlike museum works contained within a single venue, urban artwork is site-specific and distributed across a city. To maintain the saliency of location, we surface the artwork's neighborhood and link the map and artwork list views, supporting coordinated exploration from an artwork-first or map-first perspective. We present metadata in a detail pane with the artwork title, a zoomable image with alt text, the description, the SVI's capture date, artwork type, confidence, interestingness, reachability, and neighborhood. The detail pane also includes a button to favorite the artwork and a feedback form regarding AI-generated data quality.

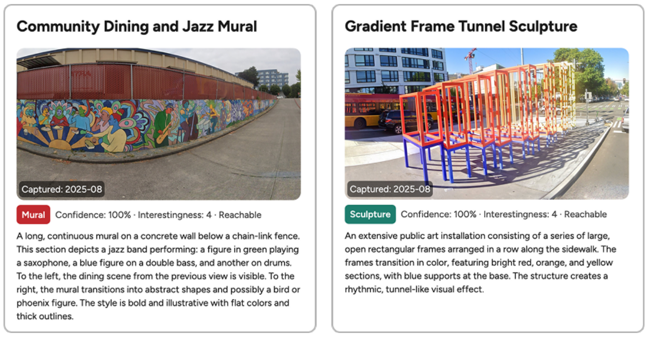

Figure 5: Examples of ArtScape's detections across different art types; overall, descriptions were rated highly for descriptiveness and accuracy by two researchers (Section 3.2.5). See Section A.1 for additional detection examples.

However, we encounter a unique challenge: cataloging the artwork inherently removes some of the spontaneity of discovering urban artwork in the wild. We approximate this experience via the tour option, where users can generate an urban art walking tour, and a "spotlight" feature, where ArtScape highlights two random works with high confidence and high interestingness scores. Despite this, we acknowledge this tension between structured organization and serendipitous discovery and return to it in Section 6.3.

3.3.2 Generating Urban Artwork Tours. Inspired by curated public art tours, we enable users to personalize walking tours by specifying origin and destination (using the *Mapbox* Search Box API [64]), distance or duration, categories of interest, number of stops, and artwork themes (Figure 6c). If no end location is set, the tour defaults to a round trip. The system scores each artwork, weighing confidence (30%), interestingness (30%), and theme relevance via keyword matching (40%, full score if unspecified). Artworks are greedily added in score-ranked order, excluding works beyond a radius of half the distance or duration limit for round trips, and works requiring a detour exceeding half the remaining distance or time after subtracting the direct path between endpoints for one-way trips. To avoid clustering, we set a minimum gap of 80 meters between tour stops, waived if two nearby works have both high confidence and high interestingness. The selected stops are passed to the *Mapbox* Optimization API [63] to determine route order. If routes exceed a user's specified duration or distance, we iteratively remove the stop which contributes the most to total route length and re-optimize the route until the constraint is satisfied or only two stops remain. As walking or rolling speed can vary, speed is adjustable via settings.

3.3.3 Screen Reader Accessibility. A key priority of our system is ensuring screen reader access. As described above, we generate detailed artwork descriptions, guided by prior work (e.g., [9, 30, 46]). We generate alt text separate from our detailed descriptions to ensure that BLV people can use a shorter description for cursory browsing and avoid the "screen reader stutter" effect [18].

3.3.4 Implementation. ArtScape is a static web app implemented with *Mapbox* [62] for the map and routing visualizations. We use the *Mapbox* Geocoding API [61] to determine which neighborhood each artwork is within. Rather than fetching artwork images from GSV at runtime, we preload artwork images to our web server for speed, consistency, and longevity. Artificial Intelligence was used to assist with implementation and we carefully reviewed all code.

4 User Study

We evaluated ArtScape with 12 participants and describe key details about our procedure, including study tasks.

4.1 Participants and Procedure

We recruited 12 participants via word-of-mouth, including three BLV participants, three artists, and two urban planners (refer to Section A.6 for demographic details). Participants were all above the age of 18 and live in the greater Seattle area. Interviews were conducted via Zoom and lasted between 60-90 minutes; participants were compensated US$25/hr. Interviews consisted of four parts, described below.

Part 1: Preliminary Interview. We first asked participants about their experiences with discovering and engaging with urban artwork (*e.g.*, current discovery practices). For BLV participants, we briefly asked about the accessibility of exploring urban art.

Part 2: Prototype Evaluation. Then, we asked participants to complete five tasks using our prototype, listed below. Participants were encouraged to "think aloud" and ask questions. We provided all participants with a summary of system features prior to sharing the link to our prototype and asking participants to share their screen. **Task 1** was to identify city-wide trends of urban artwork using search and filters. **Task 2** was to explore specific artworks and engage with the metadata (*e.g.*, descriptions, confidence, interestingness, reachability), naturalistically opening opportunities to discuss AI accuracy. **Task 3** was to identify urban art themes across neighborhoods of varying degrees of familiarity. We asked participants whether and how they were able to gain a better understanding of each neighborhood. **Task 4** was to plan an urban art walking tour, inspired by existing art tourism (*e.g.*, [12, 87]). Participants generated a tour based on location, distance or duration, number of stops, and potential interests (*e.g.*, animals, colorful artwork). Lastly, **Task 5** was to assess AI-generated information accuracy for two detections (Figure 5 left, Figure 7). We asked participants follow up questions about how to address errors that might arise when using AI.

Part 3: Reflection. Next, we engaged participants in a reflection on their experience with using the prototype. We focused on the human-AI interaction components of the system with Likert-type questions regarding the descriptiveness and accuracy (for BLV participants, perceived accuracy) of the descriptions and open-ended questions on the usage of AI for artistic applications.

Part 4: Artwork Edge Cases. We concluded with open-ended questions regarding the artistic value of additional visual urban elements—commercial artwork, architecture, and urban furniture, all of which ArtScape identified as art in earlier iterations (Figure 4)—and whether those should be presented as part of the tool.

4.2 Data and Analysis

We recorded all interviews. We developed a codebook using a hybrid inductive-deductive coding method [34]. Two researchers first coded a subset of interviews and discussed the codebook before splitting the remaining transcripts for analysis [16]. We then performed a thematic analysis. Participants' system interactions, logged locally and sent to the study team, were analyzed to support the qualitative findings.

5 User Study Findings

Participants were excited about how ArtScape helped them explore artwork at a *micro*-, *meso*-, and *macro*-scale, and inspired future in-person urban art visits—as opposed to merely passive discovery. Below, we describe key findings related to multi-scale artwork and urban discovery, human-AI interaction, and ArtScape's impact.

5.1 Multi-Scale Discovery

5.1.1 Micro-Exploration of Individual Artworks. Participants enjoyed that ArtScape enabled them to explore a large amount of familiar and unknown urban art. During the study, as measured by detail views opened, participants engaged with 464 unique artworks in total (3.56% of the entire dataset)—206 murals, 154 sculptures, 72 graffiti works, and 32 mosaics. Each participant engaged with an average of 44.9 unique artworks (*median* = 38, *range* = 3–118). A key benefit of ArtScape was being able to engage with artwork that participants typically could not: *"I drive by this at night, so it's cool that I'm using this tool because I can look at it in more detail"* (S2). Some who searched for unknown pieces narrowed their search by filtering for high-confidence and high-interestingness artworks, while others filtered specifically to find low-confidence artwork: *"I like artwork that is satirical or subversive"* (U10).

The artwork descriptions were helpful, especially to BLV participants, but limited in their artistic context. B6 noted that evocative adjectives such as *"dreamlike quality"* were *"fanciful enough to appeal to [her] imagination,"* and appreciated that the descriptions included stylistic qualities such as an *"Indigenous style"* and the artwork's *"spatial association"* with its surroundings. For U7, a sighted urban planning student, the description directed her attention to details partially obscured by a tree (Figure 1): *"it says that there's a woman's face [...] and I hadn't even noticed her at all."* However, the majority of sighted participants (S2, S3, S5, S9, A8, A11, U10) lamented the limited cultural context provided by visual-only descriptions, desiring additional information about the artist, history, materials, or even *"fun facts"* (S9) but acknowledging that these details could be difficult to determine solely with AI.

5.1.2 Meso-Exploration of Themes Within Neighborhoods. Participants highlighted their ability to explore neighborhoods as a whole, both familiar and unfamiliar. As an architect and urban planner, U10 noted that *"having this meta-analysis helps a lot with understanding the visual identity [of a neighborhood],"* referencing *"ex-industrial spaces that are being redeveloped [...] [with a new] level of visual interest."* Others mentioned that this could be useful for encouraging them to explore neighborhoods, either as a resident or a visitor. B6, who was moving, was excited that ArtScape provided *"information about the artwork [in her new neighborhood]"* to help her build her familiarity and understanding. Others found that the tool could dispel stereotypes: *"my perception was just that old people live there, but now I'm seeing, no, there's cool places I can visit"* (S9).

5.1.3 Macro-Exploration of City-Wide Trends. Lastly, the tool enabled people to engage in macro-level exploration to understand their city as a whole, beyond neighborhoods. People were often surprised by the number of artworks in Seattle, commenting that it was *"more than I would have thought"* (B1), and were able to identify patterns relating to neighborhood density and along thoroughfares. For example, S5 noticed that the graffiti was *"almost perfectly tracing out the main roads or freeways"* (Figure 6a). BLV participants used the neighborhood filters instead of the map markers for inferences: *"I had no clue there were that many graffiti artworks [in Capitol Hill], but I'm not at all surprised"* (B6). Having a broader view of the city also led to wider cultural and societal inferences: *"there's not as much of a concentration [of graffiti] in areas that might be considered more wealthy"* (A8) and *"in richer neighborhoods, [I'm seeing] a lot more of the little libraries"* (S5).

5.2 Human-AI Interaction

Participants found the usage of AI for ArtScape positive, but identified key limitations with the system's accuracy; however, these limitations opened up opportunities for critique and engagement, ultimately increasing participation in artwork curation.

5.2.1 Participants' Reaction to AI. All participants were enthusiastic about the use of AI for scalably detecting and characterizing urban artwork for exploration. Given the ubiquity of AI integrations, which S9 noted could cause people *"immediate disgust"* due to it being *"forced in [their] face,"* he felt this was *"a really cool, tangible, and actually helpful way"* to use AI. S3 also commented that *"there's a lot of controversy around AI-generated art, but this is not generating anything, it's more like categorizing,"* which she thought was a positive application.

5.2.2 Limitations of AI for Art Detection and Description. Though participants were positive about ArtScape's usage of AI, many felt the detections were overinclusive, capturing small graffiti tags, false positives, and artificially segmented art. Even after applying filters (*e.g.,* for interestingness), several participants found the volume of detections distracting: *"the repetitive images added a lot of visual noise"* (S9). In addition, the presence of false positives and negatives reduced confidence. For example, U10 noticed a large mural on the side of a building was not detected, but its reflection in the windows across the street was (Figure 6b). Despite this, participants still felt reasonably confident that ArtScape captured most artwork—sighted participants rated their confidence in the system's comprehensiveness as 4.28/5, reinforced by the large number of artworks ArtScape *did* find, and more (*i.e.,* false positives): *"it's better to have a wider view and allow people to refine the selections"* (A12) through in-system feedback. BLV participants were more skeptical, rating their confidence in ArtScape's comprehensiveness 3.17/5 on average due to prior experiences with fallible AI visual identification systems and limited methods for verifying detection or description accuracy compared to sighted users: *"I don't have any reason to feel confident"* (B6). While some participants acknowledged that GSV images could be out-of-date, with A11 mentioning

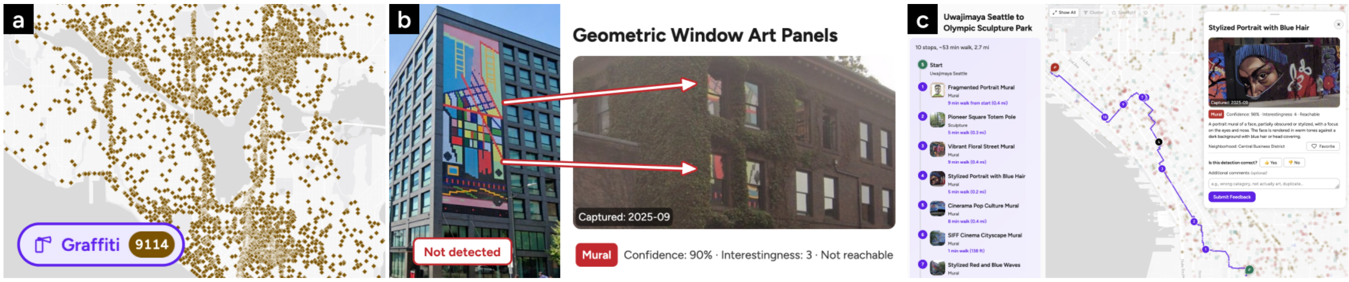

Figure 6: (a) Map of all graffiti detections in Seattle, showing graffiti lining major thoroughfares. (b) Left: mural known by U10 but not detected by the tool. Right: detection from ArtScape which falsely identifies the mural's reflection as art (false positive). (c) ArtScape's *tour* functionality, designed to augment the experience of offline art discovery.

she would be disappointed if she visited an artwork on ArtScape that was no longer there in person, U10 viewed this positively, almost like the system was surfacing *"historical documents."*

Participants rated the AI-generated descriptions as very descriptive (4.38/5), but the descriptions were still imperfect—participants gave lower ratings for accuracy (3.71/5) and cultural competency (2.83/5). For example, Task 5's *"Green Tree Climber Plush"* was in reality a neon green plastic safety sign [83] strapped to a tree (Figure 7). BLV participants also picked up on some inaccuracies, with B4 using her residual vision to check the image and B6 questioning the use of a plush material for an outdoor sculpture. For large artworks, the descriptions occasionally referenced parts of the art that existed in reality but were not visible in the current thumbnail, rendering the descriptions less accurate. For example, when looking at a sculpture of a human figure, U10 noticed that *"it refer[red] to a second human figure"* that was out of view. Regarding "cultural awareness," two participants wished details about the artist were presented more reliably and consistently. U10 mentioned that *"if someone is trying to [see] all the [artworks from a local artist], they would miss this one [...] because the AI did a bad job choosing a thumbnail"* and could not read the artist's distinctive signature. Others mentioned that information like *"who commissioned this and why, what the story is"* (A8) was also missing. Interestingly, BLV participants gave higher cultural competency ratings (4.67/5); this may reflect differing expectations for cultural detail, with sighted participants more frequently seeking contextual information beyond what a solely visual analysis could provide.

5.2.3 Interrogating Urban Art through AI's Inclusions and Edge Cases. ArtScape's AI curation gave rise to a unique phenomenon: it simultaneously lowered the barrier to criticism and increased artwork engagement by allowing participants to interrogate what they considered urban art. For example, for unconventional detections such as the *"Green Tree Climber Plush"* (Figure 7), A8 thought its inclusion in the system was legitimizing: *"it's almost an artwork because the AI found it and saw it as an artwork."* Others found these detections humorous and thought provoking: *"it's fun to see what it considers urban art... [because] it brings up the debate"* (U7).

The nuance of the major design challenge for ArtScape's pipeline, *what constitutes artwork*, was underscored when we showed participants edge cases that we had excluded when refining our detection

Figure 7: An ArtScape detection that inaccurately describes a plastic caution sign as a green plush; we used this to spark discussion about artwork inclusion and description accuracy.

guidelines. Most participants thought commercial artwork and architecture could provide aesthetic value to public spaces; however, they also recognized that urban art is generally more *"distinct intentional works"* (A8) and that it *"has to be contextual"* (U10). Despite this distinction, most participants still valued including edge cases in the system as urban art is *"nebulous"* (A12) and should be defined by its audience, not decided without their input.

5.3 Impact of the Tool

Following the prototype exploration, all participants shared their excitement for this tool, and five proactively asked if they could save the link for future use or to share with friends. Multiple participants mentioned that their level of interest in urban art increased. Post-study, participants reported high interest (4.29/5) but only moderate knowledge (2.79/5) of urban artwork; some noted that *"there's more than I thought there was, [so] I would say that my level of knowledge has gone down"* (U10). ArtScape also increased the accessibility of artwork discovery and exploration for BLV participants, all of whose ratings went from 1 to 5: *"on my own, it would be a 1, [but] using this tool, I would give it a 5"* (B6).

Participants felt that ArtScape motivated in-person urban exploration, especially given its personalized tour functionality (Figure 6c). They thought these features could actualize their interest in public art and encouraged active engagement beyond the screen: *"if I fall in love with a piece, I [might] go and visit it"* (B4), and U10 noted how ArtScape surfaced art he had never noticed on his regular commute. U7 mentioned, *"there's a big rise in urban hiking [...] this would make walking even more fun,"* while S2 suggested recreational, semi-competitive "art collection" to *"check off the artwork that you saw."* Furthermore, as an urban planning student, U7 explained the educational applications of ArtScape for visualizing how artwork, walkability, and environment design are intertwined.

6 Discussion

We introduced and evaluated *ArtScape*, the first urban art exploration tool that (1) detects, characterizes, and describes urban artwork at scale and (2) enables interactive and personalized urban art exploration. Despite widespread controversy surrounding generative AI and artwork (e.g., [47, 59]), all participants were positive about the usage of AI for cataloging art, especially for urban settings where other options are not comprehensive or personalizable. In this section, we discuss key technical challenges and opportunities, effects and implications of using AI for artwork detection, the accessibility impacts of the tool, and limitations of our work.

6.1 Technical Challenges and Opportunities

We use Google Street View images as the basis for comprehensively detecting, locating, and characterizing urban art; while powerful, this raises questions about the limits and potential of using GSV.

Data Sources. ArtScape's detections are contingent on the quality of its data sources, currently GSV for streetscape images. However, GSV's reliability and currency is variable [32, 91], and artwork may be occluded by trees, buildings, or even sunlight. To address this, our pipeline could easily be expanded to use a wider range of data sources and modalities. Additional Street View platforms (*e.g.*, Mapillary [2], Apple Look Around [1]) could address gaps in spatial and temporal coverage. Beyond Street View, geotagged photographs from the internet and social media (*e.g.*, [69]), low-altitude drone surveying, or even delivery robots would all enrich our data pool, with the latter two offering distinct vantage points. We consider GSV as merely the entryway to more comprehensive cataloging and characterization of urban street features.

Improving Performance. ArtScape relies on Gemini 3 Pro [4], DA360 [44], and DINOv3 [80], all state-of-the-art image understanding models, to maximize detection performance. As these models continue to advance, so too will ArtScape's performance. First, VLM performance markedly improves with fewer input tokens, which we leveraged by using image slices; this suggests that as thinking budgets and context-windows increase, detection performance will too. Second, using local models to first highlight visually salient or color contrasting regions may point VLMs to candidate locations, further decreasing thinking burden. Third, our dataset facilitates few-shot prompt injection, shown to improve performance [74].

Temporality. Beyond limitations arising from our data being snapshots in time, this temporality also suggests exciting opportunities for *across-time tracking*—granularly understanding the art landscape at a given place, throughout time [78]. ArtScape used the most up-to-date available SVIs; however, by extending our analysis back through historical images, we can detect past artwork and identify changes through the present. As new, more current SVIs are uploaded, we can track new artwork, and even see if and when old art pieces were removed. This means we could automatically create a comprehensive, four-dimensional art catalog allowing for historical, location-sensitive art trend tracking and understanding.

Comprehensiveness. We sampled points on the road network at 30 meter intervals, potentially leaving gaps in city coverage—however, denser sampling intervals are not always strictly better. Sampling more densely will increase recall by virtue of having clearer and more numerous lines of sight on a given artwork. However, this will also increase the number of duplicate detections and false positives, reducing precision and placing an even greater reliance on a robust deduplication step. While this creates the opportunity to surface more angles of a single art piece in a single entry (which we did not do for ArtScape), some failure cases are inevitable. This precision-recall tradeoff is inherent to streetscape object detection regardless of data source, and any pipeline will necessitate thresholding to balance the two.

Performance on Long Pieces. In cases where one piece of artwork is not fully contained in one SVI, ArtScape detects the artwork in every SVI it appears in, assigns each instance its own latitude and longitude, and deduplicates based on similarity and proximity. This means one piece could appear as multiple entries in ArtScape, or be cut off in a single entry. Future work could augment the system with deeper contextual understanding of art pieces and aggregate detections of a contiguous piece.

Beyond Art. We demonstrated that current VLMs can detect artwork—varying in shape, size, and style—throughout cities with no prior training or fine-tuning. While we focused on art, our approach could be equally effective on other urban features (*e.g.*, detecting parking, bike racks, or signage). VLMs could augment and improve existing computer vision pipelines to detect curb ramps [70] or sidewalk faults [75] or to understand subjective qualities of streetscapes [67, 86]. VLMs present a remarkable opportunity to bolster visibility, both in terms of *scale* and *detail*, into a multitude of urban features previously only accessible via manual surveying or large curated datasets.

6.2 AI for Artwork Detection

Using AI for detecting artwork at scale can foster new relationships between people and the art that surrounds them. We discuss insights on AI curation as revealed through our study and consider longer-term implications.

Lowering the Barrier to Artwork Curation. As mentioned in Section 5.2.3, the question of "what is urban art?"—already difficult to answer—was further complicated with ArtScape's AI curation. Participants had varied opinions about what was art in the first place, as evidenced by their reactions to potentially unintentional art (*e.g.*, the *"Green Tree Climber Plush"*) and smaller, less artistically interesting pieces (*e.g.*, faded graffiti tags). Even if they did not initially consider it art, some found a detection's presence in the system to be validating: *"it's almost an artwork because the AI found it and saw it as an artwork"* (A8), echoing prior theories that museum artwork is legitimized due to its curation [26, 28] and extending this idea to AI-driven systems. Interestingly, participants viewed AI as a fairly reliable curator, but one they could openly criticize and correct. Their willingness to critique artwork detections and descriptions suggests that AI-based cataloging lowered the barrier to urban artwork curation, allowing users to not only enjoy what was exhibited but also participate in defining urban art through error correction and community contributions.

Broader Implications. Whether artwork is included in ArtScape can also have downstream consequences. Over time, false positives may reframe what people consider art by bringing awareness to less mainstream cultural expressions in currently overlooked districts. However, while our AI-driven approach can expand community artwork understanding beyond manual curation, we acknowledge that systemic biases due to data and algorithm performance constraints also risk reinforcing inequities. Inconsistent false negatives could lead to unequal resource allocation or cultural acknowledgment. For example, places with less GSV coverage (*e.g.*, parks) are more likely to have missed detections, and some types of artwork were shown to be harder to successfully deduplicate (*e.g.*, graffiti). At scale, this could lead to inaccurate representations of artwork density, ultimately inspiring artwork tourism in some areas but not others and leading to unequal downstream economic outcomes. We recognize that these benefits and risks both stem from the curatorial power given to AI as part of our system. AI classification systems in other domains may similarly shape what users consider valid or noteworthy—as a result, systems should be designed with this legitimizing effect of AI curation in mind and should clearly communicate that absence from an AI-curated catalog should not be mistaken for absence in the real world.

6.3 Improving Access to Urban Artwork

ArtScape improves access to urban artwork to both disabled and nondisabled people. Creating a digital catalog of urban artwork makes it possible to explore these pieces from home. While prior work explored how digital catalogs can increase access for museum artwork (*e.g.*, [30, 40, 50]), benefits of virtual exploration are more pronounced for urban art—it is spread out across the city [10], making it infeasible to visit certain artworks, especially those in inaccessible or dangerous locations (*e.g.*, front yards, freeways). Furthermore, ArtScape's use of GSV imagery preserves a record, or *"historical documents"* (U10), of transient works that have since been removed or defaced, allowing for access to artwork of streetscapes past. Though this system does not fully replicate the spontaneity of in-person discovery [46], participants emphasized their appreciation of being able to discover artworks they would not have otherwise found and preview a neighborhood before visiting.

Another major objective of ArtScape was providing visual descriptions of urban artwork, building on a need identified in prior work [45, 46]. AI-generated descriptions fill this gap at scale, as existing repositories lack alt text for BLV users. All BLV participants' accessibility ratings of urban art exploration rose from 1 to 5, demonstrating the value of descriptions in ArtScape. As an example of the "curb cut effect" [14], some sighted participants appreciated how the descriptions helped identify occluded elements, and AI-generated descriptions also enabled object-based search. However, our descriptions were primarily visual and lacked detail about artists or artwork provenance that most sighted participants wished to know. Future work could enrich descriptions via RAG, tailor content to users' information preferences, and incorporate generative AI (*e.g.*, [53]) to help BLV users observe regional trends and artwork density more easily.

6.4 Limitations

Our work has three primary limitations. First, our study focused on only one U.S. city and twelve participants. While we deliberately sought out participants with diverse experiences and interests (artists, urban planners, BLV people, sighted people), future work should expand beyond a U.S. focus. Second, our user study was entirely virtual, with participants only interacting with ArtScape on their computers. However, many findings revolve around participants' experience or future interest in seeing urban art *in person* or "trusting" that the detections are comprehensive. Future work should examine ArtScape's impact via a deployment study. Third, we did not compare ArtScape's detections against a fully manually labeled region (*e.g.*, by walking through and cataloging all art). Even then, we acknowledge that using temporally-lagged data limits detection comprehensiveness compared to real-life, which we discussed in Section 6.1.

7 Conclusion

ArtScape marks a fundamental advance in enhancing the discovery and exploration experience for urban artwork. Our novel five-stage VLM-based pipeline accurately detects, characterizes, geolocates, and describes urban artwork across four categories (*murals*, *mosaics*, *sculptures*, *graffiti*) from Street View imagery, enabling interactive, accessible exploration at scale. Our work contributes (1) a new algorithmic pipeline, (2) an urban art exploration tool, (3) empirical findings from a study with 12 diverse participants highlighting the implications of AI curation, and (4) a benchmark dataset of urban artwork, released for open science. We showcase the benefit of such tools and demonstrate feasibility at a large scale, with implications ranging far beyond urban artwork.

Acknowledgments

We thank all study participants for their time. We also thank Sophia Hwang and Daniel Zhu for their feedback. This work was supported by NSF grant #2411222 and partially supported by Google Cloud Credits from Google.org.

References

[1] 2012. Maps. https://www.apple.com/maps/.
[2] 2013. Mapillary. https://www.mapillary.com/.
[3] 2022. API - OpenStreetMap Wiki. https://wiki.openstreetmap.org/wiki/API.
[4] 2025. Gemini 3.1 Pro. https://deepmind.google/models/gemini/pro/.
[5] 2025. Google Maps Platform Documentation | Street View Static API. https://developers.google.com/maps/documentation/streetview.
[6] Airbnb, Inc. 2026. Airbnb. https://www.airbnb.com
[7] Javier Rozo Alzate, Marta S. Tabares, and Paola Vallejo. 2021. Graffiti and Government in Smart Cities: A Deep Learning Approach Applied to Medellín City, Colombia. In *International Conference on Data Science, E-learning and Information Systems 2021 (DATA'21)*. Association for Computing Machinery, New York, NY, USA, 160–165. doi:10.1145/3460620.3460749
[8] Takaya Arita, Wenxian Zheng, Reiji Suzuki, and Fuminori Akiba. 2025. Assessing LLMs in Art Contexts: Critique Generation and Theory of Mind Evaluation. arXiv:2504.12805 [cs] doi:10.48550/arXiv.2504.12805
[9] Elisabeth Salzhauer Axel, Virginia Hooper, Teresa Kardoulias, Sarah Stephenson Keyes, and Francesca Rosenberg. 1996. *ABS's Guidelines for Verbal Description.* https://www.artbeyondsight.org/handbook/acs-guidelines.shtml
[10] María Laura Guerrero Balarezo and Kayvan Karimi. 2018. Urban art and place. Spatial patterns of urban art and their contribution to urban regeneration. In *24TH ISUF International Conference: City and territory in the Globalization Age.* Universitat Politècnica de València, 587–597. http://doi.org/10.4995/ISUF2017.2017.6069
[11] Andrea Lorenzo Baldini. 2022. What is street art? *Estetika: The European Journal of Aesthetics* 59, 1 (2022), 1–21. https://doi.org/10.33134/eeja.234

[12] Ethan Bancroft. 2024. *Marvelous Murals in the City: Take a Self-Guided Walking Tour or Join the Belltown Art Walk on September 13*. https://sdotblog.seattle.gov/2024/09/12/marvelous-murals-seattle/

[13] Axel Bauer, Juergen Hagler, and Martin Kocur. 2025. Interactive Augmented Reality Experiences for Urban Art Spaces Using Markerless Tracking. In *Proceedings of the 18th International Symposium on Visual Information Communication and Interaction*. 1–6. https://doi.org/10.1145/3769534.3769544

[14] Angela Glover Blackwell. 2017. *The Curb-Cut Effect*. https://ssir.org/articles/entry/the_curb_cut_effect

[15] Tácio Souza Bomfim, Éldman de Oliveira Nunes, A. Belén Moreno, and Ángel Sánchez. 2025. Graffiti and Covered Material Detections in Urban Images Using Deep Learning. In *2025 32nd International Conference on Systems, Signals and Image Processing (IWSSIP)*. 1–6. doi:10.1109/IWSSIP66997.2025.11151881

[16] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[17] Matthew Butler, Erica J Tandori, Vince Dziekan, Kirsten Ellis, Jenna Hall, Leona M Holloway, Ruth G Nagassa, and Kim Marriott. 2023. A Gallery In My Hand: A Multi-Exhibition Investigation of Accessible and Inclusive Gallery Experiences for Blind and Low Vision Visitors. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–15. https://doi.org/10.1145/3597638.3608391

[18] Sheri Byrne-Haber. 2026. The Screen Reader Stutter: Why Your Content Repeats Itself and How to Fix It. https://www.sheribyrnehaber.com/the-screen-reader-stutter-why-your-content-repeats-itself-and-how-to-fix-it/

[19] Chenyi Cai, Kosuke Kuriyama, Youlong Gu, Filip Biljecki, and Pieter Herthogs. 2025. Can a Large Language Model Assess Urban Design Quality? Evaluating Walkability Metrics Across Expertise Levels. arXiv:2504.21040 [cs] doi:10.48550/arXiv.2504.21040

[20] Ricardo Campos and Ágata Sequeira. 2020. Urban art touristification: the case of Lisbon. *Tourist Studies* 20, 2 (2020), 182–202. https://doi.org/10.1177/1468797619873108

[21] Eva Cetinic. 2021. Towards Generating and Evaluating Iconographic Image Captions of Artworks. *Journal of Imaging* 7, 8 (Aug. 2021), 123. doi:10.3390/jimaging7080123

[22] Patrikakis Charalampos, Kasnesis Panagiotis, Toumanidis Lazaros, and Tzitamidis Alexandros. 2019. STORM Graffiti/Tagging Detection Dataset. doi:10.5281/zenodo.3238357

[23] Arnavi Chheda-Kothary, Ritesh Kanchi, Chris Sanders, Kevin Xiao, Aditya Sengupta, Melanie Kneitmix, Jacob O. Wobbrock, and Jon E. Froehlich. 2025. ArtInsight: Enabling AI-Powered Artwork Engagement for Mixed Visual-Ability Families. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*. Association for Computing Machinery, New York, NY, USA, 190–210. doi:10.1145/3708359.3712082

[24] Jun Dong Cho, Luis Cavazos Quero, Jorge Iranzo Bartolome, Do Won Lee, Uran Oh, and Inae Lee. 2021. Tactile colour pictogram to improve artwork appreciation of people with visual impairments. *Color Research & Application* 46, 1 (2021), 103–116. https://doi.org/10.1002/col.22567

[25] Kwanghun Choi, Wontaek Lim, Byungwoo Chang, Jinah Jeong, Inyoo Kim, Chan-Ryul Park, and Dongwook W. Ko. 2022. An Automatic Approach for Tree Species Detection and Profile Estimation of Urban Street Trees Using Deep Learning and Google Street View Images. *ISPRS Journal of Photogrammetry and Remote Sensing* 190 (Aug. 2022), 165–180. doi:10.1016/j.isprsjprs.2022.06.004

[26] Arthur Danto. 1964. The artworld. *The journal of philosophy* 61, 19 (1964), 571–584. https://doi.org/10.2307/2022937

[27] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. 2018. SuperPoint: Self-Supervised Interest Point Detection and Description. arXiv:1712.07629 [cs] doi:10.48550/arXiv.1712.07629

[28] George Dickie. 2004. The new institutional theory of art. *Aesthetics and the philosophy of art: the analytic tradition: an anthology* (2004), 47–54. https://cinescontemporaneos.wordpress.com/wp-content/uploads/2016/06/new-institutional-thoery-of-art-dickie.pdf

[29] Pierre Dognin, Igor Melnyk, Youssef Mroueh, Inkit Padhi, Mattia Rigotti, Jarret Ross, Yair Schiff, Richard A. Young, and Brian Belgodere. 2022. Image Captioning as an Assistive Technology: Lessons Learned from VizWiz 2020 Challenge. *J. Artif. Int. Res.* 73 (May 2022). doi:10.1613/jair.1.13113

[30] Stacy A Doore, David Istrati, Chenchang Xu, Yixuan Qiu, Anais Sarrazin, and Nicholas A Giudice. 2024. Images, Words, and Imagination: Accessible Descriptions to Support Blind and Low Vision Art Exploration and Engagement. *Journal of Imaging* 10, 1 (2024), 26. https://doi.org/10.3390/jimaging10010026

[31] Bruno Dumas, Bram Moerman, Sandra Trullemans, and Beat Signer. 2014. ArtVis: combining advanced visualisation and tangible interaction for the exploration, analysis and browsing of digital artwork collections. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. 65–72. http://dx.doi.org/10.1145/2598153.2598159

[32] Zicheng Fan, Chen-Chieh Feng, and Filip Biljecki. 2025. Coverage and Bias of Street View Imagery in Mapping the Urban Environment. *Computers, Environment and Urban Systems* 117 (April 2025), 102253. doi:10.1016/j.compenvurbsys.2025.102253

[33] Nicola Fanelli, Gennaro Vessio, and Giovanna Castellano. 2025. ArtSeek: Deep Artwork Understanding via Multimodal in-Context Reasoning and Late Interaction Retrieval. arXiv:2507.21917 [cs] doi:10.48550/arXiv.2507.21917

[34] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods* 5, 1 (2006), 80–92. https://doi.org/10.1177/160940690600500107

[35] Joana Fogaça, Tomás Brandão, and João C. Ferreira. 2023. Deep Learning-Based Graffiti Detection: A Study Using Images from the Streets of Lisbon. *Applied Sciences* 13, 4 (Jan. 2023), 2249. doi:10.3390/app13042249

[36] Juan Garzón, Sebastián Ceballos, Esteban Ocampo, and Maryam Correa. 2023. Augmented reality-based application to explore street art: development and implementation. In *International Conference on Extended Reality*. Springer, 179–193. https://doi.org/10.1007/978-3-031-43404-4_12

[37] Google. 2026. Google Maps. https://maps.google.com

[38] Google Arts & Culture. 2026. *Street Art*. https://artsandculture.google.com/project/street-art

[39] Hilde Hein. 2006. *Public art: Thinking museums differently*. Bloomsbury Publishing PLC.

[40] Leona Holloway, Kim Marriott, Matthew Butler, and Alan Borning. 2019. Making sense of art: Access for gallery visitors with vision impairments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3290605.3300250

[41] Dajeong Hong and Jongweon Kim. 2017. Sculpture Detection Method Using the Convolution Neural Network. In *Proceedings of the 2017 International Conference on Information Technology (ICIT '17)*. Association for Computing Machinery, New York, NY, USA, 148–151. doi:10.1145/3176653.3176727

[42] Maryam Hosseini, Andres Sevtsuk, Fabio Miranda, Roberto M. Cesar, and Claudio T. Silva. 2023. Mapping the Walk: A Scalable Computer Vision Approach for Generating Sidewalk Network Datasets from Aerial Imagery. *Computers, Environment and Urban Systems* 101 (April 2023), 101950. doi:10.1016/j.compenvurbsys.2023.101950

[43] Jared Hwang, Chu Li, Hanbyul Kang, Maryam Hosseini, and Jon E. Froehlich. 2025. "Where Can I Park?" Understanding Human Perspectives and Scalably Detecting Disability Parking from Aerial Imagery. In *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '25)*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3663547.3746377

[44] Hualie Jiang, Ziyang Song, Zhiqiang Lou, Rui Xu, and Minglang Tan. 2025. Depth Anything in 360°: Towards Scale Invariance in the Wild. arXiv:2512.22819 [cs] doi:10.48550/arXiv.2512.22819

[45] Lucy Jiang, Jon E Froehlich, and Leah Findlater. 2024. Making Urban Art Accessible: Current Art Access Techniques, Design Considerations, and the Role of AI. *arXiv preprint arXiv:2410.20571* (2024). https://doi.org/10.48550/arXiv.2410.20571

[46] Lucy Jiang, Amy Seunghyun Lee, Jon E. Froehlich, and Leah Findlater. 2026. Unseen City Canvases: Exploring Blind and Low Vision People's Perspectives on Urban and Public Art Accessibility. https://doi.org/10.48550/arXiv.2603.26909

[47] Reishiro Kawakami and Sukrit Venkatagiri. 2024. The impact of generative ai on artists. In *Proceedings of the 16th Conference on Creativity & Cognition*. 79–82. https://doi.org/10.1145/3635636.3664263

[48] H. W. Kuhn. 1955. The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly* 2, 1-2 (1955), 83–97. doi:10.1002/nav.3800020109

[49] Minchu Kulkarni, Chu Li, Jaye Jungmin Ahn, Katrina Oi Yau Ma, Zhihan Zhang, Michael Saugstad, Kevin Wu, Yochai Eisenberg, Valerie Novack, Brent Chamberlain, and Jon E. Froehlich. 2023. BusStopCV: A Real-time AI Assistant for Labeling Bus Stop Accessibility Features in Streetscape Imagery. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '23)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3597638.3614481

[50] Nahyun Kwon, Yunjung Lee, and Uran Oh. 2022. Supporting a crowd-powered accessible online art gallery for people with visual impairments: a feasibility study. *Universal Access in the Information Society* 21, 4 (2022), 967–982. https://doi.org/10.1007/s10209-021-00814-2

[51] Suzanne Lacy. 1995. Mapping the terrain: New genre public art. (1995). https://monoskop.org/images/7/7c/Lacy_Suzanne_ed_Mapping_the_Terrain_New_Genre_Public_Art_1995.pdf

[52] Bahram Lavi, Eric K. Tokuda, Felipe Moreno-Vera, Luis Gustavo Nonato, Claudio T. Silva, and Jorge Poco. 2022. 17K-Graffiti: Spatial and Crime Data Assessments in São Paulo City. In *17th International Conference on Computer Vision Theory and Applications*. 968–975.

[53] Chu Li, Rock Yuren Pang, Arnavi Chheda-Kothary, Ather Sharif, Henok Assalif, Jeffrey Heer, and Jon E. Froehlich. 2026. GeoVisA11y: An AI-based Geovisualization Question-Answering System for Screen-Reader Users. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. 1–17.

https://doi.org/10.1145/3772318.3790334

[54] Franklin Mingzhe Li, Lotus Zhang, Maryam Bandukda, Abigale Stangl, Kristen Shinohara, Leah Findlater, and Patrick Carrington. 2023. Understanding Visual Arts Experiences of Blind People. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21. https://doi.org/10.1145/3544548.3580941

[55] Tianjiao Liang, Hong Bao, Weiguo Pan, and Feng Pan. 2022. Traffic Sign Detection via Improved Sparse R-CNN for Autonomous Vehicles. *Journal of Advanced Transportation* 2022, 1 (2022), 3825532. doi:10.1155/2022/3825532

[56] Xiucheng Liang, Jinheng Xie, Tianhong Zhao, Rudi Stouffs, and Filip Biljecki. 2025. OpenFACADES: An Open Framework for Architectural Caption and Attribute Data Enrichment via Street View Imagery. arXiv:2504.02866 [cs] doi:10.48550/arXiv.2504.02866

[57] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. 2023. LightGlue: Local Feature Matching at Light Speed. arXiv:2306.13643 [cs] doi:10.48550/arXiv.2306.13643

[58] Lun Liu, Elisabete A. Silva, Chunyang Wu, and Hui Wang. 2017. A Machine Learning-Based Method for the Large-Scale Evaluation of the Qualities of the Urban Environment. *Computers, Environment and Urban Systems* 65 (Sept. 2017), 113–125. doi:10.1016/j.compenvurbsys.2017.06.003

[59] Juniper Lovato, Julia Witte Zimmerman, Isabelle Smith, Peter Dodds, and Jennifer L Karson. 2024. Foregrounding artist opinions: A survey study on transparency, ownership, and fairness in AI generative art. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 7. 905–916. https://doi.org/10.1609/aies.v7i1.31691

[60] Lev Manovich. 2020. *Cultural analytics*. MIT Press.

[61] Mapbox. 2026. Geocoding v5 API. https://docs.mapbox.com/api/search/geocoding-v5/

[62] Mapbox. 2026. Mapbox. https://www.mapbox.com/

[63] Mapbox. 2026. Optimization API v1. https://docs.mapbox.com/api/navigation/optimization-v1/

[64] Mapbox. 2026. Search Box API. https://docs.mapbox.com/api/search/search-box/

[65] Cameron McAuliffe. 2012. Graffiti or street art? Negotiating the moral geographies of the creative city. *Journal of urban affairs* 34, 2 (2012), 189–206. https://doi.org/10.1111/j.1467-9906.2012.00610.x

[66] Veronika Megler, David Banis, and Heejun Chang. 2014. Spatial analysis of graffiti in San Francisco. *Applied Geography* 54 (2014), 63–73. https://doi.org/10.1016/j.apgeog.2014.06.031

[67] Lieze Mertens, Delfien Van Dyck, Ariane Ghekiere, Ilse De Bourdeaudhuij, Benedicte Deforche, Nico Van de Weghe, and Jelle Van Cauwenberg. 2016. Which Environmental Factors Most Strongly Influence a Street's Appeal for Bicycle Transport among Adults? A Conjoint Study Using Manipulated Photographs. *International Journal of Health Geographics* 15, 1 (Sept. 2016), 31. doi:10.1186/s12942-016-0058-4

[68] Louie Meyer, Johanne Engel Aaen, Anitamalina Regitse Tranberg, Peter Kun, Matthias Freiberger, Sebastian Risi, and Anders Sundnes Løvlie. 2024. Algorithmic ways of seeing: Using object detection to facilitate art exploration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18. https://doi.org/10.1145/3613904.3642157

[69] Tessio Novack, Leonard Vorbeck, Heinrich Lorei, and Alexander Zipf. 2020. Towards Detecting Building Facades with Graffiti Artwork Based on Street View Images. *ISPRS International Journal of Geo-Information* 9, 2 (Feb. 2020), 98. doi:10.3390/ijgi9020098

[70] John S. O'Meara, Jared Hwang, Zeyu Wang, Michael Saugstad, and Jon E. Froehlich. 2025. RampNet: A Two-Stage Pipeline for Bootstrapping Curb Ramp Detection in Streetscape Images from Open Government Metadata. https://arxiv.org/abs/2508.09415v1.

[71] Punartha Jeevana Perera. 2019. Urban art scene in Madrid: How can contemporary art be used for tourism? *Enlightening Tourism. A Pathmaking Journal* 9, 1 (2019), 1–37. https://doi.org/10.33776/et.v9i1.3552

[72] Preservation Detroit. 2026. *Public Art in the Cultural Center Tour*. https://preservationdetroit.org/public-art-in-the-cultural-center-tour

[73] Project for Public Spaces. 2010. *Public Art: An Introduction*. https://www.pps.org/article/pubart-intro

[74] Peter Robicheaux, Matvei Popov, Anish Madan, Isaac Robinson, Joseph Nelson, Deva Ramanan, and Neehar Peri. 2025. Roboflow100-VL: A Multi-Domain Object Detection Benchmark for Vision-Language Models. arXiv:2505.20612 [cs] doi:10.48550/arXiv.2505.20612

[75] Manaswi Saha, Michael Saugstad, Hanuma Teja Maddali, Aileen Zeng, Ryan Holland, Steven Bower, Aditya Dash, Sage Chen, Anthony Li, Kotaro Hara, and Jon Froehlich. 2019. Project Sidewalk: A Web-based Crowdsourcing Tool for Collecting Sidewalk Accessibility Data At Scale. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300292

[76] Mubarak Auwalu Saleh, Zubaida Said Ameen, Chadi Altrjman, and Fadi Al-Turjman. 2022. Computer-Vision-Based Statue Detection with Gaussian Smoothing Filter and EfficientDet. *Sustainability* 14, 18 (Jan. 2022), 11413. doi:10.3390/su141811413

[77] Seattle Center. 2026. *Public Art*. https://www.seattlecenter.com/explore/arts/public-art

[78] Ather Sharif, Paari Gopal, Mikey Saugstad, Shiven Bhatt, Raymond Fok, Galen Weld, Kavi Dey, and Jon E. Froehlich. 2021. Experimental Crowd+AI Approaches to Track Accessibility Features in Sidewalk Intersections Over Time. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual conference). 6 pages. doi:10.1145/3441852.3476549

[79] Ben Shneiderman. 2003. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *The Craft of Information Visualization*, BENJAMIN B. BEDERSON and BEN SHNEIDERMAN (Eds.). Morgan Kaufmann, San Francisco, 364–371. doi:10.1016/B978-155860915-0/50046-9

[80] Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. 2025. DINOv3. arXiv:2508.10104 [cs] doi:10.48550/arXiv.2508.10104

[81] Jonathan Skinner and Lee Jolliffe. 2017. 'Wall-to-wall coverage': An introduction to murals tourism. *Murals and tourism* (2017), 1–24. https://doi.org/10.4324/9781315547978-1

[82] Ryan Soares and Shehar Dewan. 2022. *Art in Public Spaces, an exploration of public art collections*. https://explore.publicartarchive.org/2022/02/09/art-in-public-spaces-an-exploration-of-public-art-collections/

[83] Step2. 2026. KidAlert! V.W.S.™. https://www.step2.com/products/kidalert-v-w-s

[84] Eric K. Tokuda, Roberto M. Cesar, and Claudio T. Silva. 2019. Quantifying the Presence of Graffiti in Urban Environments. In *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*. 1–4. doi:10.1109/BIGCOMP.2019.8679113

[85] Andrea Vallebueno and Yong Suk Lee. 2023. Measuring Urban Quality and Change through the Detection of Physical Attributes of Decay. *Scientific Reports* 13, 1 (Oct. 2023), 17316. doi:10.1038/s41598-023-44551-3

[86] Veerle Van Holle, Jelle Van Cauwenberg, Benedicte Deforche, Liesbet Goubert, Lea Maes, Jack Nasar, Nico Van de Weghe, Jo Salmon, and Ilse De Bourdeaudhuij. 2014. Environmental Invitingness for Transport-Related Cycling in Middle-Aged Adults: A Proof of Concept Study Using Photographs. *Transportation Research Part A: Policy and Practice* 69 (Nov. 2014), 432–446. doi:10.1016/j.tra.2014.09.009

[87] Visit Detroit. 2026. *Detroit Mural Guide*. https://visitdetroit.com/inside-the-d/detroit-mural-guide/

[88] Bingshu Wang, Qianchen Mao, Aifei Liu, Long Chen, and C. L. Philip Chen. 2025. DIG: Improved DINO for Graffiti Detection. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 55, 5 (May 2025), 3557–3569. doi:10.1109/TSMC.2025.3541795

[89] Penglei Wang, Xin Fan, Qimeng Yang, Shengwei Tian, and Long Yu. 2025. Object Detection of Mural Images Based on Improved YOLOv8. *Multimedia Systems* 31, 2 (Feb. 2025), 93. doi:10.1007/s00530-025-01687-8

[90] Shuai Wang, Ivona Najdenkoska, Hongyi Zhu, Stevan Rudinac, Monika Kackovic, Nachoem Wijnberg, and Marcel Worring. 2025. ArtRAG: Retrieval-Augmented Generation with Structured Context for Visual Art Understanding. In *Proceedings of the 33rd ACM International Conference on Multimedia*. ACM, Dublin Ireland, 6700–6709. doi:10.1145/3746027.3755673

[91] Zeyu Wang, Yingchao Jian, Adam Visokay, Don MacKenzie, and Jon Froehlich. 2025. Street View for Whom? An Initial Examination of Google Street View's Urban Coverage and Socioeconomic Indicators in the US. In *Proceedings of the 4th ACM SIGSPATIAL International Workshop on Spatial Big Data and AI for Industrial Applications*. ACM, The Graduate Hotel Minneapolis Minneapolis MN USA, 128–131. doi:10.1145/3764919.3770883

[92] Mitchell Whitelaw. 2015. Generous interfaces for digital cultural collections. (2015). https://dhq.digitalhumanities.org/vol/9/1/000205/000205.html

[93] K. M. Williamson. 2018. *What is Public Art - Definition*. https://www.publicartinpublicplaces.info/home/what-is-not-public-art

[94] Zillow Group. 2026. Zillow. https://www.zillow.com

A Appendix

A.1 ArtScape Examples

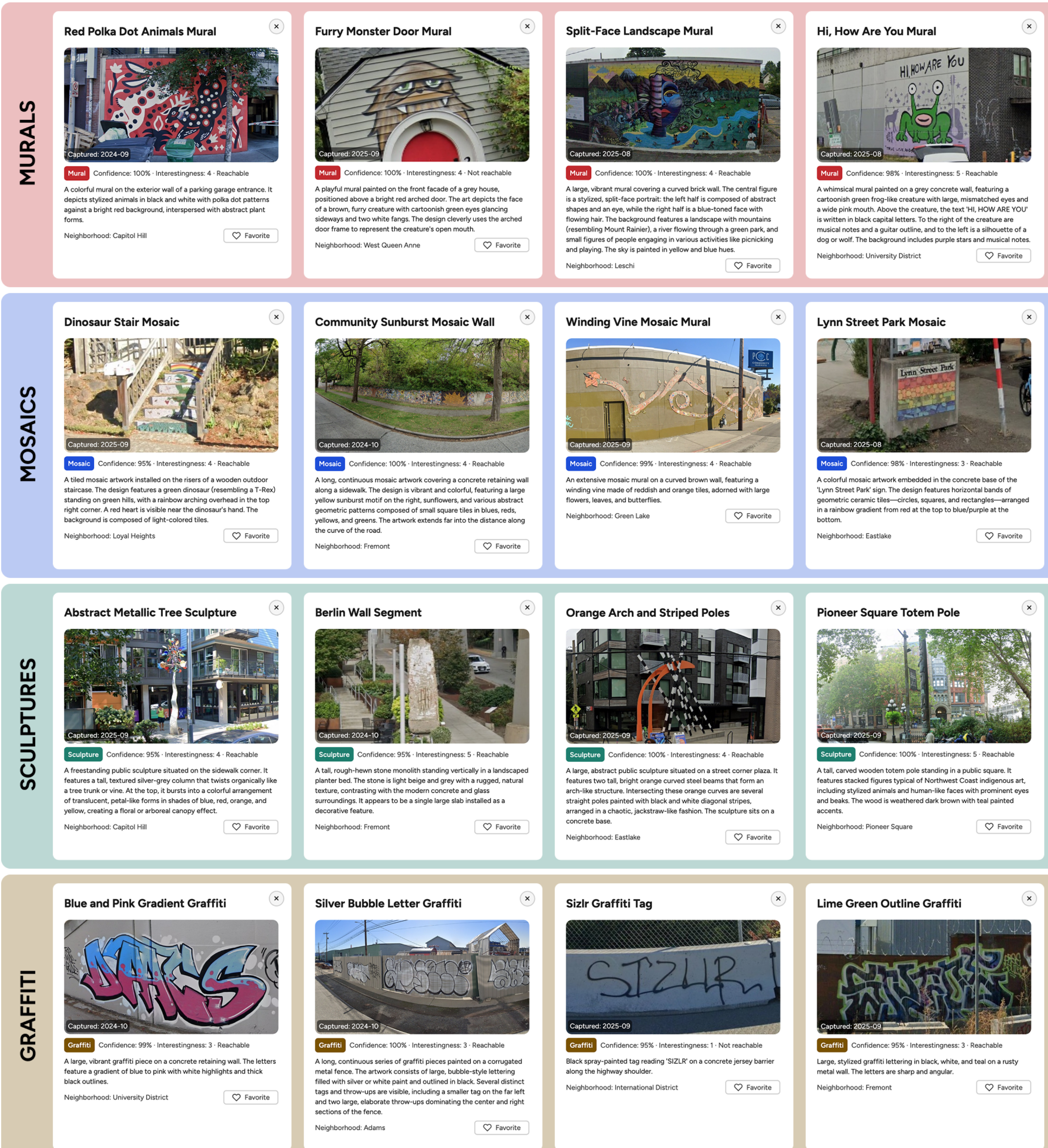

Figure 8: Examples of murals, mosaics, sculptures, and graffiti detected by ArtScape. All metadata is AI-generated and curated.

A.2 VLM First Pass Prompt

```
# Urban Art Detection in Street View Images
**Role:** You are an expert urban art surveyor. Analyze the provided street view image to identify, locate, categorize, and describe every instance of permanent urban art.
**Primary directive:** Maximize recall. Include uncertain instances with lower confidence scores rather than omitting them. There is no minimum size — distant, small, or partially obscured art should be included.
## Image Format & Bounding Boxes
The image is a roughly 120-degree slice of a streetscape.
**Detection Standard:**
For every artwork found, provide a bounding box in normalized coordinates: ‘[ymin, xmin, ymax, xmax]‘.
- The values must be integers between **0 and 1000**.
- (0,0) is the top-left corner; (1000, 1000) is the bottom-right corner.
- The box should be a tight axis-aligned rectangle enclosing the visible portion of the art.
## Art Categories
| Category | Dimensions | Materials | Key Distinction |
|—|—|—|—|
| **Mural** | 2D, painted | Paint on walls, ground, infrastructure | Pictorial/large-scale designs (not text-based tagging) |
| **Mosaic** | 3D surface (slight protrusion), tactile | Stone, tile, glass, clay | Composed of small pieces/tiles |
| **Sculpture** | Fully 3D, freestanding or on pedestal | Stone, clay, wood, concrete, metal | Observable from multiple angles. Includes statues, artistic bollards, figurines, monoliths |
| **Graffiti** | 2D, painted | Spray paint | Tags, bubble letters, spray-paint art. Group contiguous or closely clustered pieces on the same wall as one instance; separate groups are divided by significant blank space or architectural features |
| **Architecture** | 3D, Integrated into building structure | Stone, metal, concrete, ceramic | **Only** distinct sculptural building features that are clearly *additive* to the building’s baseline style: gargoyles, figurative carvings, globe ornaments, decorative figureheads. **Explicitly excluded:** cornices, keystones, dentil molding, pilasters, decorative lintels, window surrounds, ornate railings, arched doorframes, column capitals, and any ornamentation that is conventional for the building’s architectural period/style. **When in doubt, it is architecture, not art.** |
**Included surfaces:** Art on semi-permanent street infrastructure — utility/electrical boxes, dumpsters, bollards, bus stops, plastic lane dividers — should be included.
## Exclusions
Do **not** include:
- **Standard architecture** — ordinary building features, facades, decorative molding, cornices, generic trim (exception: distinct sculptural features like gargoyles or globe ornaments → classify as Architecture)
- **Performance** — people performing or temporary installations
- **Commercials/advertisements** — art made to sell a product, including promotional murals
- **Business signs** — logos, shop names, neon signs, sandwich boards
- **Vehicle art** — art on cars, trucks, vans
## Pre-Classification Gate
**Before logging any detection**, run through these questions in order. If **ANY** answer is YES, **do not include** the item:
1. Is this a **road or ground marking** (lane lines, crosswalks, bike symbols, "STOP," utility markers, arrows)?
2. Is this a **plain bollard, column, or sphere** with no artistic embellishment?
3. Is this a **standard lamp post, clock, or wayfinding sign** without decorative sculptural elements?
4. Is this a **business sign, logo, or shop name** in plain text/standard fonts with no decorative painting or illustration?
5. Is this a **plain advertisement** (product photo, undecorated banner, promotional poster with no original artistic design)?
6. Is this **ordinary architecture** — facades, molding, cornices, window frames, decorative trim, railings, or generic ornamentation that is standard for the building’s architectural style? Only distinct *figurative or sculptural* features (gargoyles, figurative carvings, globe ornaments) qualify as art.
7. Is this on a **moving vehicle** (car, bus, bike)?
8. Is this a **temporary performance or ephemeral installation** (not a physical artwork on a fixed surface)?
> **Warning: The architecture question (#6) deserves extra scrutiny.** Many buildings have ornate but *standard* features — cornices, keystones, dentil molding, pilasters, decorative lintels, column capitals, ornate railings, arched doorframes, and window surrounds. These are architectural convention, not urban art, even when they are visually elaborate. Only include a building feature if it is a distinct figurative or sculptural element that goes well beyond the building’s baseline decorative style.
## Scanning Guidance
Check these commonly missed locations:
- Distant background (horizon, far buildings)
- Utility infrastructure (electrical boxes, lamp posts, trash cans, bollards)
- Ground surfaces and overhead structures
- Small tags on posts, fences, walls
- Building rooflines and upper facades (gargoyles, sculptural ornaments)
**Common false positives to avoid:**
```

```
- Ornate but standard architectural trim (cornices, keystones, window surrounds, column capitals) — these are conventional building features, not urban art
- Decorative railings, balustrades, and ironwork that are part of the building’s original design
- Awnings, canopies, and structural elements with minor decorative touches
- Patterned brickwork or tiling that is part of the building’s construction rather than an applied artwork
## Description Guidelines
Each description must be fully self-contained and independently understandable — do not reference other entries (e.g., avoid phrases like "similar to the previous entry," "same style as #3," or "as described above"), even if the object is the same. A reader should be able to understand any single description without seeing any other entry. Each description should be a vivid passage that allows both sighted and blind/low-vision readers to understand the artwork. Include:
- **Subject matter**: What is depicted (figures, animals, abstract forms, text/lettering, patterns)
- **Colors & palette**: Dominant and accent colors
- **Style & technique**: e.g., realistic, abstract, cartoonish, stencil, freehand
- **Material & surface**: What it appears to be made of or painted on (e.g., "painted on a concrete electrical box," "bronze statue on a stone pedestal")
- **Scale & context**: Approximate size relative to surroundings and where it sits in the scene
- **Condition**: Any visible damage, fading, or weathering
**Exemplar description:**
*"A colorful mural on a roughly five-foot tall electrical box, with the center of the main panel featuring a stylized large yellow-green sea monster with purple eyes and yellow teeth and a pink tongue in its open mouth. There is a small humanoid figure wearing a red shirt and carrying a small fishing net sitting atop the monster’s head. The sea monster is surrounded by a turquoise background with white waves and there are small orange rocky formations on grassy islands in each of the four corners of the electrical box. On the side panel, there is a pink and yellow sea dragon and additional islands and waves."*
## Alt Text Guidelines
In addition to the full description, provide a short ‘alt_text‘ field for each artwork. This should follow accessibility best practices:
- **Concise**: One to two sentences, ideally under 125 characters when possible (hard max: 250 characters).
- **Descriptive, not interpretive**: State what is visually present — subject, dominant colors, medium — without subjective judgment or artistic interpretation.
- **Self-contained**: Understandable without any other context.
- **No redundant phrasing**: Do not begin with "Image of," "Photo of," or "Art of." Start directly with the content.
- **Prioritize the most important visual information**: If space is limited, lead with what makes this artwork identifiable.
**Exemplar alt text:**
*"Colorful sea monster mural on an electrical box, featuring a yellow-green creature with an open mouth surrounded by turquoise waves."*
## Reachability Guidelines
For each artwork, assess whether a person standing at street level could feasibly reach out and touch the artwork. Provide a ‘reachability‘ field with one of three values and a brief ‘reachability_reason‘ explanation:
| Value | Meaning |
|-|-|
| ‘"reachable"‘ | A person of average height standing on the ground could touch the artwork without assistance. |
| ‘"not_reachable"‘ | The artwork is too high, too far, behind a barrier, or otherwise physically inaccessible to touch from street level. |
| ‘"uncertain"‘ | Reachability cannot be confidently determined from the image (e.g., ambiguous height, unclear barriers). |
**Examples:**
- A statue on a low pedestal at sidewalk level → ‘"reachable"‘, ‘"Freestanding statue on a low pedestal at ground level along the sidewalk."‘
- A mural painted on the third story of a building → ‘"not_reachable"‘, ‘"Painted high on the building’s third-story wall, well above arm’s reach."‘
- A mural on a wall starting from ground level up to about 8 feet → ‘"reachable"‘, ‘"Lower portion of the mural is at arm’s reach from the sidewalk."‘
- A gargoyle on a roofline → ‘"not_reachable"‘, ‘"Mounted on the building’s roofline, far above street level."‘
- Art on a utility box → ‘"reachable"‘, ‘"Painted on a waist-height electrical box on the sidewalk."‘
## Output Format
Return a JSON array. If no art is found, return ‘[]‘.
Each object:
“‘json
{
    "id": 1,
    "category": "Mural",
    "location": [ymin, xmin, ymax, xmax],
    "description": "A vivid description following the guidelines above.",
```

```
      "alt_text": "Concise alt text following accessibility guidelines.",
      "reachability": "reachable",
      "reachability_reason": "Brief explanation of why the artwork is or is not touchable from street level.",
      "confidence": 0.85
}
“‘
**Field notes:**
- ‘id‘: Unique integer per artwork
- ‘category‘: One of ‘"Mural"‘, ‘"Mosaic"‘, ‘"Sculpture"‘, ‘"Graffiti"‘, or ‘"Architecture"‘
- ‘alt_text‘: Short accessible description (1–2 sentences, ideally under 125 characters, max 250)
- ‘reachability‘: One of ‘"reachable"‘, ‘"not_reachable"‘, or ‘"uncertain"‘
- ‘reachability_reason‘: Brief explanation of the reachability assessment
- ‘confidence‘: 0.0–1.0. Use 0.9+ for clear art, 0.6–0.9 for likely art with some uncertainty, 0.4–0.6 for possible art,
    0.2–0.4 for uncertain but plausible
Now analyze the image. Identify every instance of urban art, being thorough with distant, small, and partially visible items.
    Apply the Pre-Classification Gate to each candidate before including it in your output.
```

A.3 VLM Second Pass Prompt

```
You are validating a candidate urban art detection from a street view panorama. A red-outlined bounding box marks the detection
    within a cropped region (2× buffer). The box may not be perfectly aligned—use it along with the description to identify
    what was detected. If the art is near but not perfectly inside the bounding box, assume it is what the detection refers to.
**First-pass output:**
- Category: $category
- Description: $description
- Alt text: $alt_text
- Reachability: $reachability — $reachability_reason
—
### DECISION 1: Art or Not Art?
Ask yourself these questions in order. If any answer is YES, the detection is NOT art:
1. Is this a road/ground marking (lane lines, crosswalks, bike symbols, "STOP," utility markers, arrows)?
2. Is this a plain bollard, column, or sphere with no artistic embellishment?
3. Is this a standard lamp post, clock, or wayfinding sign without decorative sculptural elements?
4. Is this a business sign, logo, or shop name with no artistic intention (plain text, standard fonts, no decorative painting
    or illustration)?
5. Is this a plain advertisement — a product photo, undecorated banner, or promotional poster with no original artistic design?
6. Is this ordinary architecture (facades, molding, cornices, generic trim) without distinct sculptural features?
7. Is this on a moving vehicle (car, bus, bike)?
8. Is this a temporary performance or ephemeral event installation (not a physical artwork on a fixed surface)?
**Important clarifications:**
- **Graffiti is art.** If graffiti is visible in or near the bounding box, classify it as art even if the first-pass category
    was wrong. Correct the category and description accordingly.
- **Art on infrastructure is still art**, even if the surface is temporary (construction scaffolding, hoarding, plastic
    barriers, utility boxes). The "temporary" rule targets performances or vehicle-mounted displays, not painted/applied
    artwork on fixed surfaces.
- **Bounding box is a guide, not a hard boundary.** If the box highlights a plain area but art matching the description is
    clearly visible nearby, treat the detection as valid and describe the nearby art.
- **Artistic commercial murals are art.** A colorful, decoratively painted mural for a business (e.g., a market, restaurant) is
    still art if it shows artistic intention — i.e., it required a dedicated artist. Only flag purely commercial items: plain
    ads, product photos, or undecorated signage.
If NOT art → determine whether the detection is commercial in nature (e.g., business sign, logo, branded advertisement,
    product promotion), then return:
“‘json
{
    "is_not_art": true,
    "not_art_rationale": "Which rule above applies and why.",
    "commercial": true | false,
    "commercial_rationale": "Why this is or isn't commercial, e.g. 'Business sign for a restaurant' or 'Plain utility bollard
        with no commercial purpose'.",
    "title": null,
    "interestingness": null
}
“‘
```

```
If it IS art → continue.
### DECISION 2: Category Check
Valid categories and their defining traits:
| Category | Key Trait |
|−|−|
| Mural | 2D painted pictorial work on surfaces |
| Mosaic | Composed of small tiles/stones/glass |
| Sculpture | Fully 3D, freestanding or on pedestal |
| Graffiti | Spray-paint: tags, letters, freehand |
| Architecture | Distinct sculptural building features only (gargoyles, ornaments, figureheads) |
Agreement with first-pass category: "Agree", "Neutral", or "Disagree".
If not "Agree," suggest the correct category.
### DECISION 3: Description Quality
Evaluate description and alt_text against what you see in the image.
- Does the **description** accurately capture subject, colors, style, material, scale, and condition?
- Does the **alt text** concisely and accurately summarize the artwork (under 250 chars, no "Image of..." prefix)?
Score each as "Agree", "Neutral", or "Disagree". Provide replacements for anything not "Agree."
### DECISION 4: Commercial Purpose
Is this artwork a plain advertisement for a business, product, or service — with no original artistic design? A mural or
    illustration with artistic intention (decorative painting, stylized imagery, work that would require a dedicated artist) is
    **not commercial** even if it references a business. Only flag items that are purely promotional: product photos, plain
    banners, undecorated signage.
### DECISION 5: Title
Generate a short descriptive title (2–6 words, title case) as if labeling it in a neighborhood art walking-tour pamphlet.
Examples: "Blue Geometric Wall Mural", "Bronze Horse Statue", "Abstract Faces Graffiti", "Tiled Mosaic Archway"
### DECISION 6: Interestingness
Rate how likely this piece would be to appear on a neighborhood urban art walking tour, on a scale of 1–5:
| Score | Meaning |
|−|−|
| 1 | Barely notable — small tag, minor marking, or heavily degraded piece most people would walk past |
| 2 | Mildly interesting — competent but modest in scale or ambition; worth a glance |
| 3 | Solid inclusion — a respectable piece a tour guide would point out in passing |
| 4 | A highlight — eye-catching, skillful, or culturally resonant; worth stopping for |
| 5 | A destination — landmark-quality piece people would seek out specifically |
Use your judgment holistically. Consider scale, visual impact, craftsmanship, uniqueness, and condition.
−
### Output (JSON only):
“‘json
{
    "is_not_art": false,
    "not_art_rationale": null,
    "category_agreement": "Agree|Neutral|Disagree",
    "suggested_category": null,
    "description_agreement": "Agree|Neutral|Disagree",
    "suggested_description": null,
    "alt_text_agreement": "Agree|Neutral|Disagree",
    "suggested_alt_text": null,
    "commercial": false,
    "commercial_rationale": null,
    "title": "Your Generated Title",
    "interestingness": 3
}
“‘
Analyze the image now. Return JSON only.
```

A.4 Description Scoring

Table 2: The median and mean description ratings assigned by R1 and R2 for both descriptiveness and accuracy across all four artwork categories. Both researchers rated the descriptions to be very descriptive and accurate, with descriptiveness seeing slightly lower scores compared to accuracy.

	Median				Mean			
	Descriptiveness		Accuracy		Descriptiveness		Accuracy	
Category	R1	R2	R1	R2	R1	R2	R1	R2
Mural	4	5	5	5	4.29	4.49	4.63	4.90
Mosaic	4	4	5	5	4.30	4.38	4.57	4.79
Sculpture	4	4	5	5	4.29	4.42	4.56	4.85
Graffiti	4	4	5	5	4.39	4.16	4.53	4.53
Total	4	4	5	5	4.32	4.35	4.57	4.76

A.5 Batched VLM Results

Table 3: Per-class results of our object detection pipeline when passing images in batches (resource-saving method). The sum of TPs, FNs, and FPs does not match Table 1 due to our fuzzy matching algorithm.

Category	TP	FN	FP	Precision	Recall	F1
Mural	85	48	8	0.91	0.64	0.75
Mosaic	13	8	2	0.87	0.62	0.72
Sculpture	9	16	1	0.90	0.36	0.51
Graffiti	141	79	16	0.90	0.64	0.75
Total	248	151	27	0.90	0.62	0.74

A.6 Participant Demographics

Table 4: Participant demographics. S# = sighted participants, B# = BLV participants, A# = artists, and U# = urban planners.

ID	Age	Gender	Group
S2	25–34	Woman	Sighted
S3	18–24	Woman	Sighted
S5	18–24	Man	Sighted
S9	25–34	Man	Sighted
B1	65+	Man	Blind
B4	55–64	Woman	Low vision
B6	65+	Woman	Blind
A8	35–44	Woman	Urban artist (sighted)
A11	25–34	Woman	Artist (sighted)
A12	45–54	Man	Urban artist (sighted)
U7	25–34	Woman	Urban planning student (sighted)
U10	25–34	Man	Architect, urban planner (sighted)