

Urban Complexity through Vision Intelligence: Variance, Gradients, and Correlations across Six Italian Cities

Mirko Degli Esposti^{1,*} Armando Bazzani¹ Chiara Dellacasa³
Matteo Falcioni¹ Mario Massimo¹
Martino Pietropoli²

¹ Department of Physics and Astronomy, University of Bologna, Italy

² Department of Architecture, University of Bologna, Italy

³ CINECA, Bologna, Italy

* Corresponding author: mirko.degliesposti@unibo.it

Abstract

This paper introduces a scalable methodology for the objective analysis of quality metrics across six major Italian metropolitan areas: Rome, Bologna, Florence, Milan, Naples, and Palermo. Leveraging georeferenced Street View imagery and an advanced Urban Vision Intelligence system, we systematically classify the visual environment, focusing on key metrics such as the Pavement Condition Index (PCI) and the Façade Degradation Score (FDS). The findings quantify *Structural Heterogeneity (Spatial Variance)*, revealing significant quality dispersion (e.g., Milan $\sigma_{\text{PCI}}^2 = 1.52$), and confirm that the classical *Urban Gradient*—quality variation as a function of distance from the core—is consistently weak across all sampled cities ($R^2 < 0.03$), suggesting a complex, polycentric, and fragmented morphology. In addition, a *Cross-Metric Correlation Analysis* highlights stable but modest interdependencies among visual dimensions, most notably a consistent positive association between façade quality and greenery ($\rho \approx 0.35$), demonstrating that structural and contextual urban qualities co-vary in weak yet interpretable ways. Together, these results underscore the diagnostic potential of Vision Intelligence for capturing the integrated spatial and morphological structure of Italian cities and motivate a large national-scale analysis.

Keywords urban vision intelligence; street view imagery; spatial variance; urban gradient; pavement condition; façade quality; greenery; Italian cities; multimodal AI; agentic systems

1 Introduction: Vision Intelligence and the Quantifiable City

The visual dimension of cities has long been central to urban studies, from Lynch’s theory of *imageability* to contemporary research on perception and aesthetics [1]. The convergence

of large-scale geospatial imagery—particularly Google Street View (GSV)—and recent advances in Artificial Intelligence (AI) has enabled a new paradigm of *Urban Visual Intelligence* (UVI) [1, 2]. This framework allows the systematic observation and quantification of the built environment at the human scale, overcoming the limitations of traditional field surveys.

Recent research extends this paradigm through multimodal large language models (LLMs) and agentic reasoning for geospatial understanding. Models such as *StreetViewLLM* [3], *StreetLens* [4], and *UrbanSense* [5] demonstrate how vision-language architectures can infer semantic and perceptual attributes directly from street-level imagery, while *SAGAI* [6] explores their generative potential to reconstruct and interpret urban scenes. In parallel, the new *momepy.streetscape* module [7] provides a systematic framework for pedestrian-scale morphometrics. Collectively, these approaches signal the convergence of AI-based perception and quantitative morphology, forming the foundations of an emerging science of vision intelligence in the quantifiable city.

Beyond visual analysis, a complementary research trajectory envisions cities as ecosystems of interacting AI agents capable of reasoning, simulation, and goal-oriented behavior. Theoretical contributions such as *Conceptualising the Emergence of Agentic Urban AI* [8] and *Towards Urban Planning AI Agent in the Age of Agentic AI* [9] outline this shift from automation to agency, proposing frameworks in which autonomous agents participate in planning, governance, and adaptive decision-making. These perspectives extend the scope of urban AI from perception and representation to deliberation and action.

Building on these advances, this paper introduces a high-resolution geospatial analysis aimed at quantifying infrastructure quality and spatial heterogeneity across six major Italian metropolitan areas—Rome, Bologna, Florence, Milan, Naples, and Palermo. The data were generated through *UrbIA*, a multimodal agentic system that integrates language, vision, and spatial reasoning. Within this framework, 500 virtual agents per city—referred to as *Humarels* (Human-scale Relational Agents)—simulate distributed visual observations of the urban environment. The complete architecture and operational deployment of the *UrbIA* system are detailed in a forthcoming paper [10].

Each *Humarel* samples multiple panoramic viewpoints through the GSV Static API, using controlled camera parameters to capture targeted urban metrics. For example, the *Pavement Condition Index* (PCI) is obtained by setting the camera pitch between -35° and -45° to optimize the visibility of road surfaces. The imagery is then processed by a multimodal vision-language model performing automated visual assessment and scoring, producing georeferenced visual indicators such as surface quality, façade condition, and greenery presence.

The resulting dataset¹ forms a consistent visual census for each city, from which we derive two primary indicators: the *Spatial Variance*, capturing intra-urban heterogeneity, and the *Urban Gradient*, describing the variation of visual quality with distance from the historical core. A further *Cross-Metric Correlation Analysis* explores interdependencies among perceptual and infrastructural attributes, such as the relationship between pavement condition, greenery, and façade quality.

These layers of analysis—variance, gradient, and correlation—provide an integrated picture of urban quality, showing how structural and perceptual dimensions interact within complex, polycentric city structures. The results highlight the potential of vision-based intelligence systems for large-scale, comparative urban studies and support a broader research agenda toward explainable, AI-mediated urban analytics.

¹Available upon request.

The remainder of the paper is organized as follows. Section 2 details the data acquisition and vision analysis pipeline. Section 3 presents comparative results across cities, focusing on spatial variance and urban gradients. Section 4 introduces the cross-metric correlation analysis, and Section 5 discusses city-specific patterns. Section 6 concludes with reflections on the implications of vision intelligence for urban analysis and planning.

2 Materials and Methods

Our approach integrates systematic data acquisition using the Google Maps Platform (GMP) APIs with a custom Vision Intelligence pipeline to produce a standardized, quantifiable representation of the urban environment.

The study encompasses six major Italian cities: *Rome, Bologna, Florence, Milan, Naples, and Palermo*. These cities were selected to represent a diverse array of urban forms, population densities, and geographical contexts across the Italian peninsula.

In the initial development phase of the UrbIA system, we explored two distinct sampling strategies for the deployment of the virtual agents (Humarels):

1. *Path-Based Sensing*: Utilizing the Directions API to define a continuous, high-density route (e.g., a critical urban corridor) and generating image captures every a 20 meters along the computed polyline. This method is effective for *linear analysis* and capturing gradients along specific axes.
2. *Random Area Sampling (Adopted)*: Selecting a fixed number of agents ($N = 500$) based on random coordinates within the city’s administrative area or a defined bounding box. This approach prioritizes spatial coverage and statistical representativeness of the entire urban fabric, avoiding bias towards predefined routes.

For the comparative analysis of the *Urban Gradient* and *Spatial Variance* presented in this paper, the *Random Area Sampling* method was utilized to ensure a statistically robust and unbiased assessment of the overall heterogeneity within each city’s core region.

The area of study for each city was defined by its *Administrative Bounding Box* (BB), typically encompassing the municipal or metropolitan boundary. Within this BB, we ensured the presence of available Street View imagery before selecting the $N = 500$ random coordinates. The final sample size and the geographic coordinates of the study regions are provided in Table 1.

Table 1: Study Areas Defined by Administrative Bounding Box

City	BB Area (km ²)	Latitude Min/Max (°)	Longitude Min/Max (°)
Bologna	140.73	44.40/44.57	11.23/11.44
Florence	102.41	43.71/43.83	11.12/11.31
Milan	181.76	45.40/45.54	9.10/9.27
Naples	117.27	40.79/40.89	14.18/14.34
Palermo	158.9	38.03/38.20	13.24/13.43
Rome	1285.31	41.76/42.04	12.33/12.68

Each of the randomly selected Humarel locations serves as the point of origin for the image capture process. This step is critical as it transforms a single coordinate into a rich, multi-perspective visual dataset, maximizing the information gain from a static location.

Since the Humarel coordinates are randomly sampled and not path-dependent, the concept of a constant `travel_heading` is irrelevant. Instead, we employ a *Rotational Sampling* strategy to capture the full 360-degree environment visible from each point. For metrics requiring a full panoramic understanding (e.g., street width, context), multiple images are generated at fixed angular increments (e.g., 0° , 90° , 180° , 270° for a 4-view capture). This systematic rotation of the camera’s *Heading* ensures that all adjacent facades, street corners, and contextual elements are documented.

The Street View Static API allows for precise control over the camera’s orientation via the *Pitch* (vertical angle) and *Heading* (horizontal angle) parameters. This feature is exploited to perform *Targeted Vision Sensing*, isolating the specific urban feature required for the analysis.

The final step in our methodology involves transforming the raw, targeted visual data generated by the Humarels into quantifiable urban metrics. For the analysis presented in this paper, we adopt a direct prompt-based scoring approach using the GPT-4 Vision multimodal language model.

For each metric—Pavement Condition Index (PCI), Façade Degradation Score (FDS), Green Presence, Graffiti Index, and Urban Canyon Index—we designed specific text prompts that instruct the model to evaluate the street-level image and return a numerical score within a predefined range (e.g., 1–5 for PCI and FDS). This prompt engineering strategy leverages the zero-shot and few-shot reasoning capabilities of vision-language models to produce direct quality assessments without requiring explicit pixel-level segmentation or custom-trained classifiers.

We acknowledge that this approach, while straightforward and scalable, represents a preliminary implementation of vision-based urban sensing. The reliance on holistic image-level scoring rather than fine-grained spatial analysis introduces limitations in interpretability and geometric precision. However, the consistency of results across cities and the statistical robustness of derived indicators suggest that prompt-based scoring captures meaningful urban quality patterns at scale.²

The output of this comprehensive pipeline is a georeferenced database containing the objective visual score for each Humarel location across the six cities, which forms the empirical basis for the gradient and variance analysis presented in the following sections.

2.1 Illustrative Humarel Observations

To provide a qualitative overview of the visual content processed by the Humarel agents, this subsection presents a selection of representative Street View frames acquired during the data collection phase. Each Humarel captures multiple perspectives of its surroundings by varying the camera’s heading and pitch, as described in Section 2. The images shown here were intentionally selected to illustrate clear examples of the visual features used to compute the metrics introduced in this paper.

Figure 1 shows several representative examples of the targeted vision sensing used to derive both structural and contextual indicators. The top row includes cases of *pavement degradation* and *façade deterioration*, corresponding respectively to low Pavement

²Our broader research program (UrbIA [10]) integrates dedicated geometric segmentation models (e.g., SA2VA [11]) for precise pixel-level analysis, quantifying the percentage of road surface, greenery, sky, graffiti, and defects. These more sophisticated vision architectures will be deployed in future large-scale studies. The results presented here serve to establish baseline metrics and validate the feasibility of vision intelligence for comparative urban analysis.

Condition Index (PCI) and low Façade Degradation Score (FDS) values. The bottom row presents additional examples illustrating *graffiti presence*, *urban greenery*, and the *urban canyon effect*, where tall façades and narrow street sections define highly enclosed spatial configurations.

While the examples in Figure 1 were manually chosen for clarity, in practice the dataset also contains a fraction of frames that are visually ambiguous or partially unusable due to occlusions, strong shadows, or camera artefacts. A preliminary heuristic inspection suggests that such anomalous or low-quality images account for approximately 15% of the total sample. This issue is intrinsic to large-scale image-based urban sensing and will require further refinement of our filtering and quality-control procedures. Nevertheless, we emphasize that the statistical indicators introduced in this study—including the variance, gradient, and correlation measures—are designed to be robust to this level of noise, as confirmed by their stability and consistency across all six metropolitan areas.

Table 2: Street View Camera Parameters

Metric Target	Pitch (V.)	Heading (H.)	Analytical Focus
Pavement Condition (PCI)	-35° or -45°	Aligned to street segment	Road surface wear, material defects (<i>Manto</i>).
Façade Degradation (FDS)	0° (Horizon)	Rotational Sampling	Building quality and deterioration (<i>Facade</i>).
Green Presence Index	0° (Horizon)	Rotational Sampling	Visible vegetation density and canopy cover (<i>Verde</i>).
Urban Canyon Index	0° (Horizon)	Rotational Sampling	Street enclosure, SVF surrogate (<i>Canyon</i>).
Graffiti Index	0° (Horizon)	Rotational Sampling	Presence and extent of visible graffiti (<i>Graffiti</i>).
Sky View Factor (SVF)	$+90^\circ$ (Zenith)	Irrelevant	Microclimatic analysis, visible sky measurement.

3 Results: Quantifying Urban Disparities

The analysis of the Humarel data across the six metropolitan areas yielded a rich dataset of street-level metrics. We emphasize that the results presented in this section are preliminary findings derived from the initial $N = 500$ random samples per city. These results serve primarily to demonstrate the predictive power and scalability of the UrbIA Vision Intelligence approach. While the trends observed are statistically significant within the sampled population, future work will involve refining the Vision AI models and expanding the sample size to validate these findings against established socioeconomic and municipal infrastructure data.

The urban quality metrics extracted by the *UrbIA Multimodal Vision Agents* were analyzed by grouping them into two categories: *Structural Quality Metrics* (Pavement Condition Index — PCI, and Façade Degradation), for which both the mean value and the Spatial Variance (σ , σ^2) are key indicators, and *Contextual Metrics* (Aesthetic and Morphological). Table 3 reports the updated statistics computed from the full Humarel dataset for the six metropolitan areas.

The comparison of average scores reveals distinct urban profiles across the six cities:



Figure 1: Representative examples of the visual content captured by Humarel agents across the six cities. The top row illustrates structural conditions, including pavement and façade degradation, while the bottom row shows contextual and morphological features such as graffiti, greenery, and the urban canyon effect. Images are included solely for qualitative illustration of the UrbIA vision sensing process.

Table 3: Structural Quality Metrics: Mean Scores, Standard Deviation (σ), and Variance (σ^2) computed on 500 Humarel points per city.

City	Mean PCI (1–5)	PCI σ	PCI σ^2	Mean Façade (1–5)	Façade σ	Façade σ^2
Bologna	3.37	0.99	0.99	3.55	1.03	1.07
Florence	3.21	0.97	0.94	4.26	1.00	1.00
Milan	3.22	1.23	1.52	3.63	1.09	1.19
Naples	3.18	0.82	0.68	2.92	0.94	0.89
Palermo	2.93	0.69	0.48	3.07	0.77	0.60
Rome	3.08	1.00	1.01	3.44	1.01	1.03

Note: Bold values indicate extreme values (highest or lowest) within each column.

Table 4: Contextual Metrics: Average Scores (Aesthetic and Morphological)

City	Graffiti Index (0-2)	Green Presence (0-5)	Urban Canyon (0-2)
Bologna	0.48	1.99	0.84
Florence	0.13	2.75	1.30
Milan	0.36	2.30	0.90
Naples	0.45	1.72	1.08
Palermo	0.30	2.40	1.43
Rome	0.48	2.73	1.20

- *Pavement Condition (PCI)*: **Bologna** and **Milan** exhibit the highest average pavement scores (**3.37** and **3.22**, respectively), indicating comparatively better road surface conditions. **Palermo**, with the lowest mean PCI (**2.93**), confirms the greatest overall need for maintenance intervention. In terms of spatial variance, **Milan** shows the highest dispersion ($\sigma^2 = 1.52$), pointing to pronounced heterogeneity between well-maintained and deteriorated areas, while **Palermo** displays the lowest variance ($\sigma^2 = 0.48$), suggesting more uniformly modest conditions across the urban fabric.
- *Façade Condition (FDS)*: **Florence** stands out with the highest average façade score (**4.26**), reflecting stronger preservation practices and visual consistency of its built environment. Conversely, **Naples** records the lowest mean façade score (**2.92**), indicating widespread degradation. Regarding variance, **Milan** again exhibits the highest dispersion ($\sigma^2 = 1.19$), consistent with a visually diverse building stock, whereas **Palermo** shows the lowest façade variance ($\sigma^2 = 0.60$), highlighting more homogeneous, though moderately degraded, façades.
- *Green Presence*: **Florence** (**2.75**) and **Rome** (**2.73**) exhibit the highest average scores for visible green presence. **Naples** recorded the lowest score (**1.72**), indicating a low density of visible vegetation in the Humarels' viewpoints.
- *Graffiti Index*: The average presence of visible graffiti is highest in **Bologna** and **Rome** (**0.48**). Notably, **Florence** recorded the lowest average index (**0.13**), suggesting effective cleaning or a lower incidence within its core study area.
- *Urban Canyon Index*: This index, measuring the perceived enclosure of the street space, is highest in **Palermo** (**1.43**), indicating a prevalence of extremely narrow streets and dense building morphology. **Bologna** shows the lowest index (**0.84**), reflecting a sampling dominated by more open avenues or boulevards.

The analysis of simple average scores, while useful for macro-level comparison (Section 3.2), fails to capture the defining characteristic of the Italian urban experience: *extreme spatial heterogeneity*. Unlike many modern cities where quality often follows predictable sectorial or radial patterns, historic Italian cities exhibit high contrast and rapid transitions in urban quality. It is common to find beautifully preserved areas—high-scoring tourist centers—immediately adjacent to segments displaying significant degradation in terms of pavement condition or façade maintenance, often mere meters away (the "degrado dietro l'angolo" effect). This phenomenon of juxtaposition of excellence and deficit is a critical dimension of urban resilience and planning efficiency.

3.1 Spatial Variance (Heterogeneity)

Spatial Variance quantifies the degree of disparity or non-uniformity of a given metric within the boundaries of a single city. A high variance in a score (e.g., Pavement Condition Index) indicates that the city is deeply heterogeneous, suggesting a significant difference in infrastructure quality between the best-maintained and the worst-maintained areas. This can be calculated using the standard deviation (σ) or the coefficient of variation (CV) of the 500 Humarels' scores for each metric.

Figure 2 visualizes the empirical distributions of the Pavement Condition Index (PCI) and the Façade Degradation Score (FDS) for the six metropolitan areas. Each violin

shows the density of Humarel scores, with median and interquartile markers. The figure complements Table 3 by illustrating the internal shape of spatial heterogeneity: Milan and Bologna display broad PCI distributions with visible bimodality, while Palermo and Naples show narrower, lower-centered profiles consistent with their lower means. In façade quality, Florence presents a compact, upper-shifted distribution (homogeneously high-quality façades), whereas Naples exhibits a long lower tail, indicating spatially clustered degradation. These distributions confirm that visual quality is not evenly distributed within cities, providing the empirical rationale for the Urban Gradient analysis discussed in the following section.

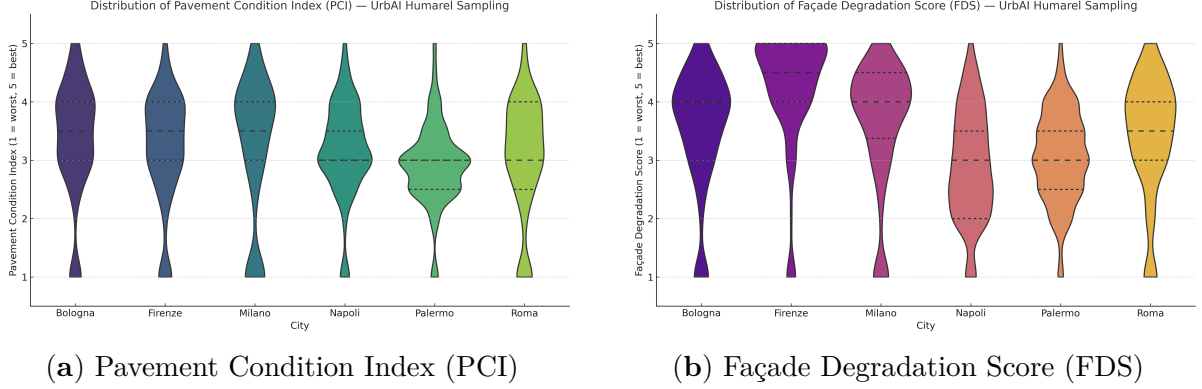


Figure 2: Empirical distributions of structural visual quality scores across six Italian metropolitan areas (Humarel sampling, $N=500$ per city). Violin plots display the full density of scores (1–5), with medians and extrema. PCI and FDS share the same y-axis range for comparability. The plots complement Table 3 by revealing the internal shape of spatial heterogeneity (e.g., skewness, multimodality), thereby motivating the Urban Gradient analysis introduced in the next section.

3.2 the Urban Gradient: Center Selection

The *Urban Gradient* quantifies how visual quality varies as a function of distance from the historical core of the city. For each Humarel observation i , the gradient is modeled as the relationship between the measured visual score (e.g., the Pavement Condition Index, PCI, or the Façade Degradation Score, FDS) and the *Distance from the Historical Center* (DHC), expressed in kilometers.

A critical methodological step is the definition of the anchor point for DHC. Italian cities are often polycentric or possess layered historical centers that complicate the notion of a single “city center.” To ensure comparability and reproducibility, we selected for each metropolitan area a landmark that is both historically and functionally central—typically the main cathedral square or administrative nucleus.

Table 5 lists the selected anchor points used for calculating DHC for all 500 Humarel observation points per city. Geographic distances were computed using the Haversine formula on latitude–longitude coordinates, providing great-circle distances in kilometers.

The Urban Gradient is operationally defined as the slope β_1 in a linear model of the form:

$$S_i = \beta_0 + \beta_1 DHC_i + \varepsilon_i,$$

where S_i represents either PCI or FDS for observation i , and DHC_i is the corresponding

Table 5: Anchor Points for Distance-from-Historical-Center (DHC) Calculation. Coordinates are expressed in decimal degrees.

City	Center Location Used	Latitude	Longitude
Bologna	Piazza Maggiore	44.4938° N	11.3426° E
Florence	Piazza del Duomo	43.7730° N	11.2561° E
Milan	Piazza del Duomo	45.4642° N	9.1900° E
Naples	Piazza del Plebiscito	40.8384° N	14.2494° E
Palermo	Quattro Canti	38.1147° N	13.3619° E
Rome	Piazza del Campidoglio	41.8933° N	12.4828° E

distance from the city’s historical center. Negative values of β_1 indicate a decline in visual quality with increasing distance from the historical core.

By fitting regression models across the six cities, the gradient analysis enables us to:

1. Evaluate whether the classical concentric urban model holds—i.e., if visual quality decays radially with distance.
2. Identify non-linear or segmented trends in the quality–distance relationship, characteristic of polycentric, industrial, or postmodern urban morphologies [1].

To visualize the spatial coverage and distribution of the 500 Humarel sampling points across the study areas, Figure 3 presents a static map for each city, with the points plotted relative to the Historical Center (Hollow Circle Marker), reflecting the shape of the city.

3.3 Urban Gradient Results and Interpretation

Table 7 reports the linear regression coefficients of the Urban Gradient model for both the Pavement Condition Index (PCI) and the Façade Degradation Score (FDS). Across the six Italian metropolitan areas, the slopes (β_1) are generally small, with $R^2 < 0.03$, indicating that radial distance from the historical center explains only a limited portion of the spatial variance. **Bologna**, **Milan**, and **Palermo** exhibit mildly negative PCI gradients, suggesting a gradual decline in pavement quality outward from the core, while **Florence**, **Bologna**, and **Palermo** show positive FDS gradients, reflecting newer and better-maintained façades in suburban districts. **Naples** and **Rome** display weak or contrasting patterns, consistent with their polycentric and historically layered structure. These results confirm that radial distance alone explains only a minor share of the spatial variance in visual quality ($R^2 < 0.03$). The Urban Gradient analysis reveals weak or city-specific trends—negative PCI gradients in **Bologna**, **Milan**, and **Palermo**, and positive FDS gradients in **Florence**, **Bologna**, and **Palermo**—highlighting the complexity and polycentric nature of Italian urban morphologies. Such differentiated and metric-dependent patterns demonstrate the diagnostic value of the Urban Gradient framework and motivate its extension to a large-scale, data-intensive analysis.

Although the regression coefficients in Table 7 quantify the general direction and strength of the Urban Gradient, their small magnitudes do not fully convey the spatial structure of the underlying data. Figure 4 visualizes representative scatter plots of Pavement Condition Index (PCI) and Façade Degradation Score (FDS) against the Distance from Historical Center (DHC). The plots confirm that while weak negative or positive trends exist, the dispersion of points is high and city-specific patterns diverge

Distribution of Humarel Sampling Points in Six Italian Cities

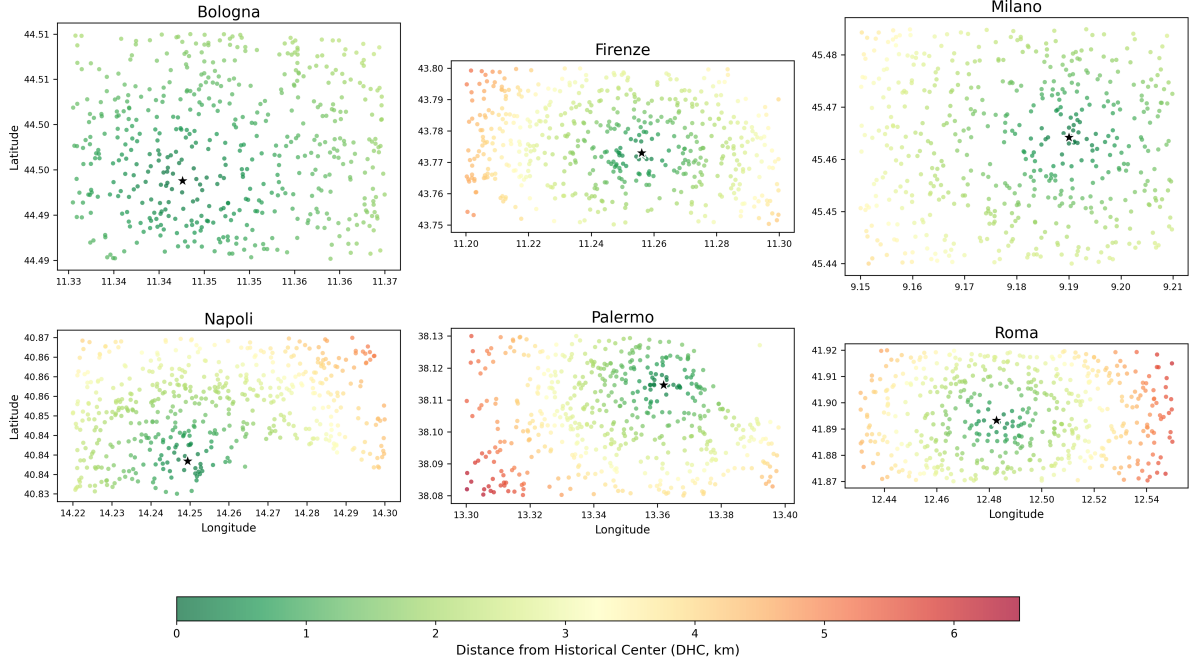


Figure 3: Distribution of the 500 Humarel Sampling Points in the six study cities. The location of the Historical Center (anchor point for DHC) is marked with a distinctive symbol.

Table 6: Summary of Urban Gradient regressions: slope (β_1), determination coefficient (R^2), and statistical significance (p). Interpretations highlight the direction and strength of each metric’s relationship with distance from the historical center (DHC).

City	Metric	β_1	R^2	p	Interpretation
Bologna	PCI	-0.10	0.003	0.23	Very weak negative slope; not significant.
	FDS	+0.19	0.009	0.03	Significant positive trend; façades improve outward.
Florence	PCI	-0.02	0.001	0.56	Flat; no measurable gradient.
	FDS	+0.14	0.025	< 0.001	Strong positive; façade quality increases outward.
Milan	PCI	-0.09	0.004	0.17	Weak decline; not significant.
	FDS	+0.02	< 0.001	0.79	Flat; no correlation.
Naples	PCI	0.00	< 0.001	0.99	No gradient; uniform quality.
	FDS	-0.11	0.016	0.004	Negative slope; façade degradation increases outward.
Palermo	PCI	-0.07	0.022	< 0.001	Significant negative; quality declines outward.
	FDS	+0.03	0.003	0.22	Weakly positive; not significant.
Rome	PCI	+0.02	0.001	0.53	No meaningful gradient.
	FDS	-0.06	0.006	0.08	Slight decline; marginally significant.

Table 7: Linear regression coefficients of the Urban Gradient model: $S_i = \beta_0 + \beta_1 DHC_i + \varepsilon_i$, for Pavement Condition Index (PCI) and Faade Degradation Score (FDS).

City	Metric	β_0	β_1	R^2	p -value	N
Bologna	PCI	3.49	−0.10	0.003	0.23	500
Bologna	FDS	3.33	+0.19	0.009	0.031	500
Florence	PCI	3.27	−0.02	0.001	0.56	500
Florence	FDS	3.90	+0.14	0.025	< 0.001	500
Milan	PCI	3.39	−0.09	0.004	0.17	500
Milan	FDS	3.60	+0.02	< 0.001	0.79	500
Naples	PCI	3.17	0.00	< 0.001	0.99	500
Naples	FDS	3.18	−0.11	0.016	0.004	500
Palermo	PCI	3.13	−0.07	0.022	< 0.001	500
Palermo	FDS	2.99	+0.03	0.003	0.22	500
Rome	PCI	3.01	+0.02	0.001	0.53	500
Rome	FDS	3.62	−0.06	0.006	0.08	500

markedly. These visualizations illustrate the limited explanatory power of purely radial models and motivate the extension of this framework to non-linear and multi-center gradient analysis using large-scale computational modeling.

3.4 Cross-Metric Correlation Analysis

To complement the spatial and radial analyses presented in the previous sections, we examined the internal relationships among the structural and contextual visual metrics extracted by the UrbIA Multimodal Vision Agents. The goal of this correlation analysis is to determine whether specific aspects of the urban scene—such as pavement quality, faade condition, greenery, or graffiti presence—tend to co-vary across the 3,000 Humarel observation points collected from the six metropolitan areas.

Table 9 reports the pairwise **Spearman correlation coefficients** among the main visual metrics aggregated over all cities. Correlations were computed on standardized 1–5 scores after excluding missing or invalid entries. The resulting coefficients highlight the structural coupling and independence between key dimensions of urban visual quality.

Table 8: Spearman correlation coefficients among structural and contextual visual metrics (all cities combined, $N=3000$).

	Pavement	Faade	Greenery	Graffiti	Canyon	Material
Pavement (PCI)	1.00	0.22	−0.08	−0.03	−0.05	0.02
Faade (FDS)	0.22	1.00	0.35	−0.05	0.13	−0.00
Greenery	−0.08	0.35	1.00	0.04	0.30	−0.04
Graffiti	−0.03	−0.05	0.04	1.00	0.00	0.04
Canyon	−0.05	0.13	0.30	0.00	1.00	−0.06
Material	0.02	−0.00	−0.04	0.04	−0.06	1.00

The correlation structure reveals several interesting patterns:

- *Pavement and Faade Quality* ($\rho = 0.22$): A weak positive correlation indicates

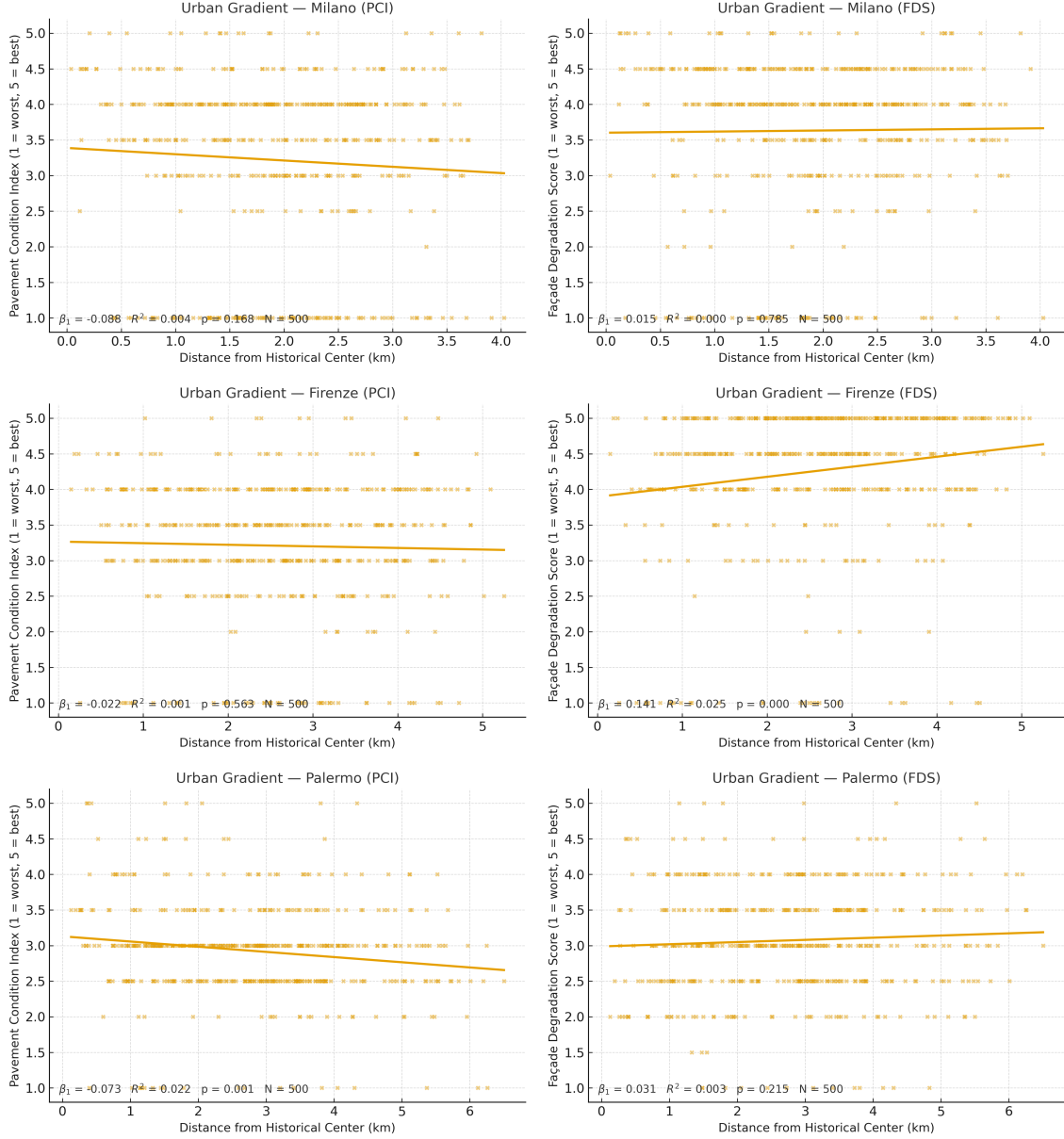


Figure 4: Representative Urban Gradient plots for three Italian metropolitan areas (Milan, Florence, and Palermo). Each scatter plot shows individual Humarel observations ($n=500$ per city) with fitted regression lines for both Pavement Condition Index (PCI, left column) and Façade Degradation Score (FDS, right column). Despite mild positive or negative slopes, the overall dispersion remains high, confirming that distance from the historical core explains only a small share of the observed spatial variance.

that better road conditions are often found alongside better-maintained façades, suggesting that maintenance efforts may cluster spatially.

- *Façade Quality and Greenery* ($\rho = 0.35$): The strongest, but still moderate, observed correlation highlights that greener streetscapes are generally associated with higher façade quality, reinforcing the link between environmental and visual well-being.
- *Façade Quality and Graffiti* ($\rho = -0.05$): The expected negative relationship suggests that graffiti presence slightly co-varies with facade degradation, although the effect remains very weak, almost negligible.
- *Urban Canyon and Greenery* ($\rho = 0.30$): Denser built-up areas tend to include more vegetation in visible frames, possibly due to the presence of trees along narrow streets, a configuration typical of Mediterranean city cores.
- *Material and Pavement Quality* ($\rho = 0.02$): The absence of a clear relationship confirms that the type of pavement material is not directly predictive of its visual state at the scale of analysis considered.

Overall, the correlation magnitudes remain quite modest ($|\rho| < 0.4$), confirming that the various visual metrics capture *distinct and complementary dimensions* of urban quality. While Pavement and Façade scores show some shared structure, contextual variables such as Greenery, Graffiti, and Urban Canyon contribute independent information to the city’s morphological and perceptual profile.

Figure 5 provides a visual overview of these relationships, illustrating the moderate positive association between Façade and Greenery scores and the weak coupling of the remaining indicators. The relative independence of these metrics supports their combined use in a multivariate modeling framework for large-scale urban analysis to process expanded datasets across multiple European cities.

Table 9: Spearman correlation coefficients among structural and contextual visual metrics (all cities combined, $N=3000$).

	Pavement	Façade	Greenery	Graffiti	Canyon	Material
Pavement (PCI)	1.00	0.22	−0.08	−0.03	−0.05	0.02
Façade (FDS)	0.22	1.00	0.35	−0.05	0.13	−0.00
Greenery	−0.08	0.35	1.00	0.04	0.30	−0.04
Graffiti	−0.03	−0.05	0.04	1.00	0.00	0.04
Canyon	−0.05	0.13	0.30	0.00	1.00	−0.06
Material	0.02	−0.00	−0.04	0.04	−0.06	1.00

3.5 City-Level Correlation Patterns

To verify whether the relationships observed in the global correlation matrix are consistent across different urban morphologies, we computed the pairwise Spearman coefficients for each of the six metropolitan areas (Table 10). The analysis focuses on four key relationships: Pavement–Façade, Façade–Greenery, Façade–Graffiti, and Greenery–Canyon.

The results confirm that the overall structure of correlations is robust across cities, with all coefficients remaining within the mild-to-moderate range ($|\rho| < 0.5$). The *Façade–Greenery*



Figure 5: Spearman correlation heatmap of visual metrics across all cities. Colors range from blue (negative correlation) to red (positive correlation). The strongest relationship appears between Façade and Greenery, while other pairs show weaker, complementary dependencies.

Table 10: Spearman correlation coefficients for key metric pairs computed separately for each city.

City	$\rho(\text{Pavement, Façade})$	$\rho(\text{Façade, Greenery})$	$\rho(\text{Façade, Graffiti})$	$\rho(\text{Greenery, Canyon})$
Bologna	0.10	0.27	-0.05	0.45
Florence	-0.04	0.45	-0.07	0.35
Milan	0.27	0.21	-0.09	0.30
Naples	0.26	0.44	-0.13	0.30
Palermo	0.20	0.42	-0.01	0.45
Rome	0.24	0.24	-0.05	0.30

relation emerges as the most stable and substantial, reaching $\rho \approx 0.45$ in Florence and Naples, while *Pavement-Façade* correlations are weaker ($\rho \approx 0.10-0.27$) but consistently positive. *Façade-Graffiti* correlations remain negative in all cases, indicating that higher graffiti presence is associated with lower façade quality, with Naples showing the strongest effect ($\rho = -0.13$).

Taken together, these city-level results confirm the general trends observed in the aggregated analysis while revealing distinct local signatures. The persistence of consistent correlation signs across diverse morphologies underscores the stability of the UrbIA visual metrics, while the magnitude variations highlight how local planning histories modulate the coupling between infrastructural and contextual dimensions of urban quality.

4 Discussion and Conclusions

The results presented in this paper provide a coherent, data-driven portrait of the visual and infrastructural complexity of six major Italian cities. Across all metrics, the analysis reveals a high degree of spatial heterogeneity, confirming that visual quality is not evenly distributed but rather fragmented into a patchwork of well-maintained and degraded areas. The weak or statistically insignificant *Urban Gradients* observed (typically $R^2 < 0.03$) indicate that distance from the historical core alone no longer explains variations in infrastructure quality, supporting the notion that contemporary Italian cities are intrinsically polycentric and spatially stratified. The *Cross-Metric Correlation Analysis* complements this picture by showing modest but consistent associations among visual indicators, particularly the positive link between façade quality and greenery ($\rho \approx 0.35$), suggesting that structural and environmental qualities co-vary in weak yet interpretable ways.

From a morphological perspective, these findings imply that Italian metropolitan areas exhibit a high level of internal differentiation: the coexistence of well-preserved heritage districts, recently renewed zones, and neglected interstitial spaces. Such complexity challenges traditional models of concentric urban decay and supports a view of the city as a multi-scalar, heterogeneous system shaped by successive waves of construction, conservation, and infrastructural maintenance. The strong intra-urban contrast revealed by the *Spatial Variance* measures may thus be interpreted as an empirical signature of this historical layering and functional diversification.

Methodologically, this study demonstrates the viability of large-scale, image-based urban sensing using multimodal Vision Intelligence. Despite the presence of approximately 15% visually ambiguous or low-quality images—an inherent limitation of Street View sampling—the indicators defined here prove statistically robust, producing stable distributions and correlations across all cities. Future refinements will focus on improving the *Humarel* agents through better tuning of visual recognition parameters, adaptive frame selection, and automated filtering of anomalous or non-representative images. Enhanced calibration and multi-view consistency checks will allow the extraction of more precise and semantically consistent visual metrics.

The next phase of this work, currently under development, involves scaling the analysis to national and European levels using cloud-based GPU infrastructure for large-scale image processing and vision model inference. This expansion will enable the processing of millions of Humarel observations and the integration of additional contextual layers, such as socioeconomic indicators, mobility patterns, energy performance, and land-use

statistics. By correlating visual metrics with these external data sources, we aim to explore how physical appearance, environmental quality, and social conditions intertwine within complex urban systems. Ultimately, the project will transition from descriptive analysis to predictive modeling, employing statistical learning and simulation to anticipate spatial dynamics of degradation, maintenance, and renewal.

Beyond its empirical findings, this study illustrates how *Urban Vision Intelligence* can serve as a foundational tool for quantitative morphology and policy-oriented diagnostics, bridging visual perception, infrastructure assessment, and computational urban science.

Author Contributions

Conceptualization, M.D.E.; methodology, M.D.E. and M.F.; software, M.D.E., M.F., and M.M.; writing—original draft preparation, M.D.E.; writing—review and editing, M.D.E., A.B., M.P., M.M., M.F., and C.D.C. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The datasets generated and analyzed during the current study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
FDS	Façade Degradation Score
GMP	Google Maps Platform
GSV	Google Street View
LLM	Large Language Model
PCI	Pavement Condition Index
UVI	Urban Visual Intelligence
ViT	Vision Transformer

References

- [1] Zhang, F.; Salazar-Miranda, A.; Duarte, F.; Vale, L.; Hack, G.; Chen, M.; Liu, Y.; Batty, M.; Ratti, C. *Urban Visual Intelligence: Studying Cities with Artificial Intelligence and Street-Level Imagery*. Preprint, arXiv:2301.00580, **2025**.
- [2] Fan, Z.; Zhang, F.; Loo, B.P.Y.; Ratti, C. Urban visual intelligence: Uncovering hidden city profiles with street view images. *Proc. Natl. Acad. Sci. USA*, **2023**, *120*, e2220417120.
- [3] Li, Z.; Xu, J.; Wang, S.; Wu, Y.; Li, H. StreetviewLLM: Extracting Geographic Information Using a Chain-of-Thought Multimodal Large Language Model. Preprint, arXiv:2411.14476, **2024**.
- [4] Kim, J.; Jang, L.; Chiang, Y.-Y.; Wang, G.; Pasco, M.C. StreetLens: Enabling Human-Centered AI Agents for Neighborhood Assessment from Street View Imagery. Preprint, arXiv:2506.14670, **2025**.
- [5] Yin, J.; Zhong, J.; Li, P.; Pan, R.; Zeng, P.; Zhang, M.; Lu, S. UrbanSense: A Framework for Quantitative Analysis of Urban Streetscapes Leveraging Vision-Language Models. Preprint, arXiv:2506.10342, **2025**.
- [6] Pérez, J.; Fusco, G. Streetscape Analysis with Generative AI (SAGAI): Vision-Language Assessment and Mapping of Urban Scenes. Preprint, arXiv:2504.16538, **2025**.
- [7] Araldi, A.; Fleischmann, M.; Fusco, G.; Novotný, M. Streetscape Morphometrics: Expanding Momepy to Analyze Urban Form from the Street Point of View. *SoftwareX*, **2025**, *31*, 102242.
- [8] Tiwari, A. Conceptualising the Emergence of Agentic Urban AI: From Automation to Agency. *Urban Informatics*, **2025**, *4*, 1–16. <https://doi.org/10.1007/s44212-025-00079-7>.
- [9] Fu, Y.; Wang, D. Towards Urban Planning AI Agent in the Age of Agentic AI. Preprint, arXiv:2507.14730, **2025**.
- [10] Degli Esposti, M.; Falcioni, M.; Massimo, M.; Pietropoli, M.; Dalla Casa, C. UrbIA: Toward an Integrated Framework for Urban Questioning and Vision-Based Intelligence. University of Bologna, Preprint, **2025**.

- [11] Yuan, H.; Li, X.; Zhang, T.; Huang, Z.; Xu, S.; Ji, S.; Tong, Y.; Qi, L.; Feng, J.; Yang, M.-H. Sa2VA: Marrying SAM2 with LLaVA for Dense Grounded Understanding of Images and Videos. Preprint, arXiv:2501.04001, **2025**.